MAPPING THE COUNTRYSIDE: INFORMATION FOR POLICY AND MANAGEMENT

Dominic Alan Shand Tantram

Submitted as partial fulfilment of the requirements of the higher degree of Doctor of Philosophy, University of Leicester.

University College Northampton

June 2001

MAPPING THE COUNTRYSIDE: INFORMATION FOR POLICY AND MANAGEMENT

Dominic Alan Shand Tantram

There is an increasing demand for information for the rational assessment and reporting of the state the environment, to detect change and to assess the effectiveness of policy or management measures.

The research investigated the use of information by conservation organisations through case studies in the Statutory Nature Conservation Agencies and the North York Moors National Park. The results highlighted a number of key problems in the organisational use of information and in the content and utility of the data available. These included the lack of an organisational culture of information use, imperfect knowledge and utilisation of available data, the need to meet changing information demands and the requirement to produce comparable local, regional and national habitat stock estimates.

Many of the data deficiencies highlighted would appear to be met by the Countryside Survey (CS) initiative. Despite offering potentially suitable data, with a combination of an environmental stratification (the ITE land class system), field survey and remotely sensed data, this source was little used. Thus, the study sought to assess the scope for comparing CS data with other habitat estimates and for improving the accuracy of these data through the use of Geographical Information Systems (GIS). Three main techniques were employed, modified areal weighting, modified areal weighting with control zones and 'intelligent weighting' a hybrid approach in which Land Cover Map of Great Britain (LCMGB) data were employed to redistribute Countryside Survey 1990 (CS90) totals within ITE land classes.

The research found that sub-land class estimates from CS90 data could be improved in some circumstances. In most cases, LCMGB provided better estimates of habitat location and quantity than CS90. In a few cases, the intelligent weighting method improved the interpolation of CS90 estimates. It is suggested that regional habitat estimates may be improved further through greater within-land class differentiation, an increase in within-land class sampling intensity or stratification and the further development of the LCMGB.

The problems faced in integrating, analysing and using available geographic data are considered and conclusions presented.

Including contents pages, text, references and appendices the thesis is approximately 99,000 words in length.

ACKNOWLEDGEMENTS

Thanks are due to a range of people for advice and encouragement:

Prof. David Briggs for providing expert supervision above and beyond the call of duty, to Dr Duncan McCollin and Dr Tom Dargie for additional supervisory support and to Dr Ian Livingstone.

Thanks are due to the following for permission to use material in the case studies and analysis:

Susan Davies (Formerly Joint Nature Conservation Committee, now Scottish Natural Heritage);

Rosie Simpson (Countryside Agency);

Mike Welbourn (North York Moors National Park);

Dr Timothy Moffat (Formerly Environmental Information Centre, Centre for Ecology and Hydrology);

Prof. Martin Parry (Jackson Environment Institute, University College London).

Part funding of the case studies by the Joint Nature Conservation Committee and the Countryside Agency and Association of National Park Authorities is gratefully acknowledged.

Thanks are due to the following individuals for providing data, information and specific guidance:

Dr Nick Michael, (Formerly Heathland Ecologist, English Nature);

Justin Gilbert (Forestry Commission);

Prof. Sylvia Richardson (Chair in Biostatistics, Imperial College);

Geoff Radley (Coastal Ecologist, Joint Nature Conservation Committee);

Kathy Duncan (Coastal Ecologist, Scottish Natural Heritage);

Margaret Hill (Formerly botanist, Joint Nature Conservation Committee);

Dr Keith Kirby and Christene Reid (Woodland ecologists, English Nature);

Dr Ian Bradley (Soil Survey and Research Centre, Cranfield University).

A very special thanks to Elaine for putting up with me during the write-up and for providing much needed threats together with psychological and practical support; and to my family for encouragement and nagging over the years.

Chapter 2 of the thesis contains some material produced in collaboration with the CLAUDE (**Co**-ordinating **La**nd **U**se and cover **D**ata in **E**urope) Concerted Action Project for DG XII of the European Commission, funding from which provided partial sponsorship of this research. CLAUDE aimed to develop an internally consistent Europe-wide plan for land cover monitoring, land use monitoring and research and to link with other relevant international programmes. Work for the UK inventory was conducted by the author, as part of this thesis, and comprised two main strands: the collection and review of published literature and contract reports, and the conduct of a number of project interviews with key data suppliers and users.

The SNCO case study in Chapter 3 of the thesis contains some material produced in collaboration with the Joint Nature Conservation Committee who provided partial sponsorship for this research. Thanks are due to Prof. Martin Kent (University of Plymouth) for assistance in development of the questionnaire, running workshops and for comments on the findings and interpretation of the results.

The National Park case study in Chapter 3 of the thesis contains some material produced in collaboration with the North York Moors National Park, the Countryside Agency and the Association of National Park Authorities who provided partial sponsorship of this research as part of the Parks Information and Monitoring Development Project (PIMS). Thanks are due to Peter Scott (Peter Scott Planning Services) for assistance in running the workshops.

CONTENTS

FIGURES

TABLES

ACRONYMS

ADAS	Agricultural Development Advisory Service (Formerly part of MAFF, now privatised)
ANOB	Area of Outstanding National Beauty
ANPA	Association of National Park Authorities
AOD	Above Ordnance Datum
AWI	Ancient Woodland Inventory
BAP	Biodiversity Action Plan
BRC	Biological Records Centre (CEH, Monkswood)
BSBI	Botanical Society of the British Isles
BTO	British Trust for Ornithology
CA	Countryside Agency
CASI	Compact Airborne Spectrographic Imager
CBD	(UN) Convention on Biological Diversity
CC	Countryside Commission (Now part of CA)
CCW	Countryside Council for Wales
CEH	Centre for Ecology and Hydrology (part of NERC)
CIS	Countryside Information System
CLA	Country Landowners Association
COREDATA	NCC Legacy information system
CPRE	Council for the Protection of Rural England
CS	Countryside Survey
CWT	Census of Woodlands and Trees
DEFRA	Department of Environment, Food and Rural Affairs (Established June 2001)
DETR	Department of the Environment, Transport and the Regions
DNF	Digital National Framework (OS)
DOE	Department of the Environment (Now part of DETR – DEFRA from June 2001)
EA	Environment Agency
EIA	Environmental Impact Assessment
EN	English Nature
EPA	Environmental Protection Act (1990)
ESA	Environmentally Sensitive Area
FA	Forestry Authority
FC	Forestry Commission
FE	Forest Enterprise
FGDC	Federal Geographic Data Committee
FRCA	Farming and Rural Conservation Agency (Ceded back to MAFF 01/04/2001 and then to DEFRA
GDP	Gross Domestic Product
HMIP	Her Majesty's Inspectorate of Pollution (Now part of EA)
IFE	Institute of Freshwater Ecology (Now within CEH)
IOH	Institute of Hydrology (Now within CEH)
ISA	Indicator Species Analysis
ISR	Invertebrate Site Register
ITE	Institute of Terrestrial Ecology (Now within CEH)
IUCN	International Union for the Conservation of Nature
JNCC	Joint Nature Conservation Committee
LANDSAT TM	LANDSAT Thematic Mapper (RS observation tool)
LBAP	Local Biodiversity Action Plan
LCMGB	Land Cover Map of Great Britain
LCS88	Land Cover Scotland 1988

LIDAR	Laser Induced Direction and Range
LRC	Local Record Centre
LUCID	Land Use Classification Information and Documentation (Computer software)
LUCS	Land Use Change Statistics
MAFF	Ministry of Agriculture, Fisheries and Food (DEFRA from June 2001)
MHWS	Mean High Water Spring tide
MLC	Monitoring Landscape Change
MLCNP	Monitoring Landscape Change in National Parks
MLURI	Macaulay Land Use Research Institute
MLWS	Mean Low Water Spring tide
MNCR	Marine Nature Conservation Review
NAW	National Assembly for Wales
NBN	National Biodiversity Network
NCC	Nature Conservancy Council (Split into EN, CCW, SNH, JNCC & ITE)
NCMS	National Countryside Monitoring Scheme
NCR	Nature Conservation Review
NERC	Natural Environment Research Council
NFU	National Farmers Union
NGDF	National Geospatial Data Framework
NGO	Non-Governmental Organisations
NIWT	National Inventory of Woodland and Trees
NLUC	National Land Use Classification
NLUSS	National Land Use Stock System
NNR	National Nature Reserve
NP	National Park
NRA	National Rivers Authority (Now part of EA)
NSA	Nitrate Sensitive Area
NTD	National Topographic Database (OS)
NUTS	Nomenclature des Unites Territoriales Statistiques
NVC	National Vegetation Classification
NYMNP	North York Moors National Park
OECD	Organisation for Economic Co-operation and Development
OPAC	Online Public Access Catalogue
OS	Ordnance Survey
PIMS	Parks Information Management System
PPG	Planning Policy Guideline (Issued by DETR)
RQO	River Quality Objective
RS	Remote Sensing
RSNC	Royal Society for Nature Conservation (the Wildlife Trusts Partnership)
RSPB	Royal Society for the Protection of Birds
SAC	Special Area for Conservation
SAP	Species Action Plan
SAR	Synthetic Aperture Radar
SNCI	Site of Nature Conservation Interest (Second tier site)
SNCO	Statutory Nature Conservation Agency
SNH	Scottish Natural Heritage
SPA	Special Protection Area
SPANS	Spatial ANalysis System, GIS software product from PCI Geomatics
SQL	Structured Query Language
SSI	Site of Scientific Interest (Second tier site)

SSSI	Site of Special Scientific Interest
Strategi	OS digital map product
UN	United Nations
UNEP	United Nations Environment Programme
WCMC	World Conservation Monitoring Centre (Now part of UNEP)
WWF	World Wide Fund for Nature
WWT	Wildfowl and Wetlands Trust

1 INTRODUCTION

1.1 BACKGROUND

During the second half of the last century, collective attitudes to the environment have changed greatly. The public perception of the environment has moved from one related to resource exploitation to the recognition of a complex and finite resource. Both public and political attitudes have evolved and continue to do so (HM Government 1990). It is only relatively recently that environmental policy has moved from a largely reactive position towards a more planned or precautionary framework. Public and political attitudes change quickly; a good example of this is global climate change. Fifteen years ago the concept was highly disputed, but now, despite the fact that the scientific case is still young and largely unproven, the combination of increasing public awareness and growing empirical evidence has led to the general acceptance that climate change is with us (Chapin *et al.* 2000, Kaiser 2001, Kerr 2001). Alongside the concept of climate change, issues such as the depletion of the ozone layer, destruction of rainforests and the conservation of biodiversity have entered common parlance. Against this global background, developments in sustainability and the UN (United Nations) Convention on Biological Diversity have helped bring international environmental concerns into a national context. What were once seen as rather narrow and sectoral concerns about the countryside have now achieved a higher profile of wider relevance to a greater constituency of people. This change in direction has been followed by changing policy and legislative measures. This, in turn, has highlighted the need for information, both to guide and monitor policy and to support management decisions and practice.

The countryside comprises a complex mixture of natural and manmade features. In Britain, the vast majority of the countryside has been modified by thousands of years of human influence. There are few truly natural habitats remaining. However, some of the most important habitats are those shaped by human intervention through continuity of traditional farming or husbandry practices. In many cases these are extensive systems that helped to produce and maintain semi-natural habitats – those influenced by human intervention, but retaining a range of 'natural' characteristics. The countryside is composed of a range of land cover types ranging from relatively unmodified habitats through semi-natural ones to highly modified and intensive agricultural land use systems. In biodiversity and conservation terms, the more natural and semi-natural habitats retain the most interest and value (Green 1985, Warren and Goldsmith 1974).

Biodiversity is a relatively new concept, but one that builds upon earlier conservation ideas and values. It is argued that biodiversity should be retained for a variety of reasons including: *moral obligation* - that species and habitats enrich our lives, are intertwined with our culture and should be passed onto the next generation; *stewardship* - humans have determinative power over species and species that are lost cannot be regained; *benefit to society* – retaining natural processes protects our way of life and potentially valuable natural resources; *economic value* – genetic material should be retained for agricultural, health and environmental purposes, natural features provide the varied landscapes that attract visitors and underpin the tourism industry (HMSO 1995a). Thus habitats are important as conservation and biodiversity resources in their own right and they are also valuable as amenity resources and in providing the fabric of our natural heritage.

Despite this importance, there has been a long-term decline in both the quantity and quality of habitats in Great Britain. Whilst long-term trends, such as climate modification, have undoubtedly contributed to these changes, by far the most important driving force in most cases has been change in land use or management. In agriculture, for example, technological innovations and the impacts of the Common Agricultural Policy have had far-reaching effects on the quality, range and distribution of habitats (Milton 1994, NCC 1984, Stoate 1996). In addition, urbanisation and infrastructural development, have led to the loss of important habitats and contributed to the wider pollution of the rural environment (Marrs 1993, Rice 1994, Royal Commission on Environmental Pollution 1994).

1.2 WHY COUNTRYSIDE INFORMATION IS NEEDED

1.2.1 Habitat Stock and Change

Despite ongoing research into habitat change and loss, there is still a paucity of information about habitat resources and distribution. Habitat change as a result of development and other pressures is both diverse and subtle. Impacts may be represented in both the extent and quality of habitats. In each case, change may also be expressed in different ways. Changes in extent, for example, may be evidenced by gross loss in the area of the habitats concerned: given relationships which often exist between area and quality, such changes may have profound effects on habitat quality (Barr *et al.* 1993, Kirby 1995). More subtly, however, habitats may undergo processes of fragmentation or isolation with little change in their overall area, but with important effects for habitat diversity and stability (e.g., Kirby 1995, Lescourret and Genard 1994, Webb and Hopkins 1984). Equally, changes in management or physical conditions (e.g. soil fertility, salinity, pollution load) may result in changes in species richness or abundance, in species distribution, or species composition, independently of any detectable variation in extent (Cooper *et al.* 1995). Thus measures of habitat stock are needed, both to assess our resource base and as a reference point for the assessment of change.

Habitat information is needed for management and monitoring purposes. If it is to help assess the state of a valued resource and to evaluate the effectiveness of management measures monitoring should be intimately linked to some kind of target or value system (Hellawell 1992, IUCN 1994). However, evaluation is not necessarily straightforward; how do we express improvement or deterioration and what should we use as a reference point? Rowell and Reed (1994) suggested the explicit connection of feature-based objectives to the management planning process by utilising management objectives as monitoring standards. If suitable information is available this is a valuable step forward for work in managed areas, but further development is required to apply the same concepts to area-wide monitoring systems in counties or regions that do not possess coherent or logical management planning frameworks. Indeed, even within recognised planning units, such as National Parks, objective-based management systems are often poorly developed. To turn the story full circle, this is often because of a lack of suitable data to underpin management objectives (Briggs *et al.* 1996).

Information on the magnitude of habitat change remains limited. Much of the information that does exist is derived from local, intensive studies - there have been few comprehensive national habitat resource surveys. Local studies generate important data on changes in specific habitats and key sites, and thus play a vital role in conservation. In terms of countryside management and planning, however, the need for information extends beyond these priority areas. Information is needed, equally, on the state of the wider countryside, which whilst, at first, not appearing to be of such intrinsic value for nature conservation is vitally important to concepts of sustainability and the conservation of biodiversity (Adams 1996). Information is also needed at broader scales, to support strategic management of the countryside, and to monitor the impacts of broader policies. In this context, detailed, site-specific information has limited value, unless methods exist by which it can be extrapolated to the wider countryside. Routine, area-wide measurement or monitoring of the countryside, however, is as yet only partially developed. The Co-ordinating Commission for Biological Recording (CCBR, Burnett *et al.* 1995) suggested that few nation-wide projects were sufficiently well organised to provide a framework for future monitoring.

Routine data on land use are collected by the Department of Environment, Farming and Rural Affairs (DEFRA – previously Ministry of Agriculture, Fisheries and Food [MAFF]), as part of the annual agricultural returns, but these give little direct information on habitats, and cover only the utilised agricultural area. A number of other broad-based surveys have also been undertaken to examine land use or land cover change (Wyatt *et al.* 1993). Although these projects have provided an immense wealth of data on land cover and land use, data on semi-natural habitats or vegetation are not so widespread or detailed. Additionally, there are a number of limitations such as those of spatial scale, survey frequency and classification detail that mean, in many cases, existing surveys or classifications cannot provide reliable estimation of habitat stock or change at the resolution required for countryside management and planning (Sinclair 1992). For example, the Countryside Survey 1990 comparison of land cover definitions compared estimates of land cover produced by different surveys; the comparison of Countryside Survey 1990 field survey *vs.* the ITE Land Cover Map produced a correspondence of only 46-54% (Wyatt *et al.* 1993). While this was thought to be reasonable within the scope of that study, these figures suggest serious problems for the comparison of such data for different applications. Indeed the work by Wyatt *et al.* reports possible survey error in the region of 25%. Such problems with land cover based approaches have historically led ecologists to be wary of their results and to depend upon more detailed and therefore more expensive field-by-field ground surveys such as Phase 1 (NCC 1990) or Phase 2 surveys. Phase 1 survey has provided the basis for designated site identification at local and regional scales since the early 1980s, but was conducted in different areas with different qualities and with some variation in methodology (Wyatt 1991). In many counties, Phase 1 is now fifteen years old and most organisations are unable to afford new survey work to meet new requirements. The approach adopted by the (then) Department of the Environment's (DOE) Countryside Survey was to follow a hybrid system of sample-based field survey data combined with satellite imagery (e.g. Barr *et al.* 1993, Stott 1994). Thus, although there are many surveys, the data produced by these are often explicitly land cover based, difficult to compare or of insufficient resolution to be of specific use for habitat monitoring. Consequently, we still know little about habitat stock and processes of change in the countryside. A better understanding of these changes is required to underpin policy and management to meet objectives for nature conservation, biodiversity and sustainability.

1.2.2 Policy context

Information on the wider countryside is being increasingly required for policy making and management decisions. In terms of policy, planning and management, habitat-related data is needed for three main reasons. Firstly it is required for the rational assessment and reporting of the state of a resource; secondly, it is required to assess the effectiveness of policy or management measures; and thirdly it is needed to detect incipient change. Therefore the demand for information has increased in line with increasing concern over the environment and in conjunction with increasing financial constraints faced by conservation agencies and land managers.

Information is often lacking on the effect of policy and management and their outcomes. Various applications including environmental auditing, state of the environment reporting, strategic environmental assessments and integrated environmental assessment, require area-wide environmental data which are often deficient or non-existent. Data are often not available due to the great time and expense of acquiring and maintaining geographically wide data sets, because data are often collected for other specific requirements, and because data collection has been slow to catch up with policy development and management requirements (Sinclair 1992).

Numerous policy and legislative processes specifically require information on the countryside and habitats within it. For example, the Government is a signatory to the International

Convention on Biodiversity and as such is obliged to identify, monitor and protect a variety of habitats and species. Similar requirements are imposed by European-based legislation and directives on the conservation of both species and habitats. The United Kingdom's sustainability strategy (DETR 1999a) identifies the importance of our countryside for its natural, historical and cultural significance. At the national level information is also required on the state of the countryside and how it is changing. This is used by Government bodies such as the Countryside Agency to promote sustainability, countryside management and recreation. It has also used by the Department of the Environment Transport and the Regions for 'Quality of Life Counts' indicators.

Many of these applications require the supply of area-wide information, for example the Convention on Biological Diversity. The British Government's response, expressed initially through *Biodiversity: The UK Action Plan* (DOE 1994), proposed what is probably the most significant policy for monitoring wildlife and countryside. The plan set out a range of actions that the Government and its agencies would undertake and provided for the establishment of targets for both species and habitats. In addition, Britain has agreed to produce Local Biodiversity Action Plans. To do this, information is required on species and habitat distribution, state, extent and threats. This is often available for selected and protected sites, but often scarce or missing in the wider countryside. Other recent initiatives have increased the demand for habitat-related data. The 1995 Environment Act placed new requirements upon National Parks that included reporting upon the state of National Parks and setting targets for evaluating the effectiveness and performance of National Park Authorities. The Countryside and Rights of Way Act 2000 (HMSO 2001, Phillips and Huggett 2001) made provision for the identification and mapping of areas of open land for public access. The Act included mountain, moorland, downland and commons. With the exception of the last category, none of these landscape types are currently mapped and the Countryside Agency has therefore begun a complicated and expensive process to capture suitable data to map these areas (Clark *et al.* 2000).

Together, these policy developments have a number of practical implications. Policy makers and land managers now have to consider wider geographical areas in more detail for different specific purposes. In addition to these new information needs, there are stronger monitoring requirements for setting targets and maintaining standards. Management now has to be shown to be both effective and cost-efficient. Information is needed to make decisions based upon knowledge of natural heritage resources and is needed on a wide range of characteristics including habitat extent, quality, status, change and threat.

1.3 THESIS RATIONALE

The aims of the thesis are to investigate the use of information by conservation organisations, to look at why they need information and what use is made of it. The research explores how key organisations source data to meet their requirements and the data problems they encounter.

The study then looks in detail at two important data sources and explores the scope for improving the accuracy and utility of these data through the use of Geographical Information System (GIS) techniques. The problems faced in integrating, analysing and using available habitat-related data are investigated and conclusions made on the implications drawn.

Specifically the research aims are to:

- assess the use of habitat-related data for environmental management in the UK;
- examine the availability and utility of existing data in relation to different information needs;
- investigate methods for improving capabilities to estimate and map habitat stock.

The research objectives are to:

- review the range of available data sources for habitat estimation and mapping in the wider countryside;
- assess the information requirements of the Statutory Conservation Agencies and the National Parks through two case studies;
- identify the technical, methodological and organisational requirements for habitat-related information;
- identify possible technical and methodological improvements for integrating data sources to improve results and improve habitat estimation at the sub-regional scale.

1.3.1 Structure of the Thesis

The research comprises a number of different sections and research techniques. Chapter Two describes the organisational and policy background to countryside management and habitat-related information use in Great Britain. It also examines the conceptual background to habitat identification and classification and the development of different data sources. A range of potentially relevant data sources are described and reviewed based upon published sources and a brief component of interview-based research amongst relevant bodies to identify data issues for further examination. To assess more fully the role and use of information by conservation bodies, Chapter Three describes two case studies. These focus firstly upon the work conducted by the statutory nature conservation agencies and secondly upon that of a National Park. The case studies were conducted by a variety of interview and seminar based techniques to assess the use of information by individuals and their organisations and identified a range of key information issues based upon data use and utility for specific applications. Specifically, the widespread need to produce stock estimates for different habitats, at different scales was identified. Thus, chapters Four and Five assess the scope for improving habitat estimates. Countryside Survey data and other geographic data sets are used to evaluate the potential for producing and mapping accurate habitat estimates for three target habitats. The study uses GIS-based techniques and produces quantitative results for assessment at different scales. The discussion (Chapter 6) reviews the range of issues faced by habitat-related data providers and users, and a model of data utilisation is proposed. The data processing methods used in the thesis are critically reviewed and placed within the context of the wider research findings. The thesis concludes with a final chapter presenting a summary of the thesis and presenting summary conclusions and implications.

2 INFORMATION AND COUNTRYSIDE MANAGEMENT

2.1 THE ROLE OF INFORMATION

Environmental information is used for a wide range of policy and management purposes, at a variety of scales and by many different organisations. Over the past thirty years, public and governmental interest in the environment has grown almost exponentially (Pepper 1989). At the same time the scientific disciplines of ecology, environmental science and biogeography have further developed and been supplemented by related disciplines such as landscape ecology. In Great Britain, there has long been a tradition for natural history recording, often originating from Victorian natural history societies (Burnett *et al.* 1995). These developments have resulted in the creation of vast amounts of environmental data, often for very specific purposes. These data are used to different extents by both the originators of the data and by others. Information is used to further scientific knowledge, to support operational requirements, to meet legal obligations, to lobby for political or policy changes or simply for aesthetic or recreational purposes.

The growth in environmental studies has coincided with industrial development and change and information is sought to help understand the causes and impacts of environmental change. Over the last century the combined impacts of industrial development, concomitant social change and development of mechanised agricultural practices has led to the loss of many natural and semi-natural habitats and a decline in numbers and distribution of the species associated with them. As environmental issues entered the political mainstream in the 1980s, information on environmental change and impact has been increasingly sought by policy makers and environmental managers. The term 'policy makers' is used here in its widest possible sense, to include individuals and organisations making decisions at a local level up to those developing strategy at the international level.

In most cases, a rational policy maker would seek to develop policy from an informed position. However, in many cases decisions are made on the basis of sketchy or incomplete information, either because suitable information does not exist, or because it is not known about. In various attempts to address these situations, different environmental information, survey, and collation projects have been conducted over the years. These have resulted in a variety of different information sources, usually created to address specific needs at specific times. Collectively, they have helped form the pool of information available to policy makers and environmental managers. Nevertheless, the existence of information sources does not mean that they are accessible. To be accessible users need to be aware of the existence of information sources and their content. The formats in which data are stored are also crucial. A data source that exists in the form of hand written coded record cards in a shoe box in someone's loft can be just as inaccessible as one on a 5.25" floppy disk held in a bespoke machine code format! The rapid development of information technology over the past thirty years has helped support the collation and management of large quantities of environmental data. In the 1970s and 1980s, there was a vogue for developing large database systems. In the 1990s this was overtaken by the technology-supported trend for Geographical Information Systems (Burrough 1986). However, technology has also played a part in restricting the accessibility and use of information. Large-scale and over ambitious database projects often failed to deliver promised outputs and disillusioned many people in their utility. It became clear that environmental data bases would not solve everyone's information problems and that they were simply tools that could provide suitable functionality under a range of circumstances limited by their original design and their data content.

This chapter examines the legislative and policy context underpinning the need for environmental information and the roles of the different 'countryside' actors. It then goes on to review the range and content of existing information sources and the potential for technology to address some of the problems of information management and analysis.

2.1.1 The UK's legislative context

The UK Government is subject to a number of international and European conventions and legal agreements that are relevant to habitat conservation and protection. Over the last thirty years, environmental legislation has been driven largely by European Policy (Haigh 1991). After the original signing of the Treaty of Rome in 1957, EC member states had appeared to have little interest in the environment. However, the original treaty, as amended by the Single European Act to integrate environmental considerations into all EC policy, provided for the development of environmental policy. The Maastricht Treaty gave the European Union the additional objective of promoting sustainable growth. Beginning in 1972 a series of environmental action programmes had been established, each running between three and five years. The fifth environment action Programme (1992-2000) provided measures for progress towards sustainable development. The sixth environment action programme (European Commission 2001) for the European Union for 2001-2010 identifies four priority areas: climate change, nature and biodiversity, environment and health, and natural resources and waste. To achieve progress in each of these areas five key approaches are proposed to:

- ensure the implementation of existing environmental legislation;
- integrate environmental concerns into all relevant policy areas;
- work closely with business and consumers to identify solutions;
- ensure better and more accessible information on the environment for citizens;
- develop a more environmentally conscious attitude towards land use.

These approaches mirror earlier ones from previous action programmes, but the emphasis on the provision of environmental information to the public is strengthened, as is the co-operation with industry.

Despite such policies and accompanying legislation (some 200 directives and 280 items of legislation by mid 1991 [Young 1993]) the 1992 report (European Commission 1992) on the state of the European environment revealed a marked deterioration in environmental quality. This included measures such as increased levels of atmospheric and aquatic pollution, increased levels of waste products, deterioration of soil resources and urban environments and threats to nature conservation.

International treaties are generally enforced in the UK through government issued policy guidance and domestic legal instruments. The major obligations and conventions relating to habitats are outlined below to help describe the legislative and policy framework that nature-conservation and countryside organisations work within in the UK. The descriptions below are not exhaustive, but instead are summaries of major influences and drivers for habitat-related information requirements. The focus of our interest here is Great Britain which, obviously, is covered by UK-level legislation, so only that dealing with Great Britain is described. Similar but subtly different arrangements exist for Northern Ireland and dependent territories.

UN Convention on Biological Diversity, 1992

The United Nations' Convention on Biological Diversity was one of the key outcomes of the 'Earth Summit' held in Rio de Janeiro in 1992 (United Nations 2001). The UK was one of 150 signatories who recognised that action should be taken to try and stem the world-wide decline in animal and plant species and their associated genetic resources. They agreed to develop national strategies, programmes and plans to conserve biological resources and diversity within their jurisdiction.

As a signatory to the International Convention on Biological Diversity, the UK government is obliged to undertake measures for safeguarding biodiversity and co-ordinating the collection and maintenance of supporting information and data (under Article 7).

The Government's response to the Convention was *Biodiversity: The UK Action Plan* (HMSO 1994a) that set out broad goals for conservation and made important statements on the integration of biodiversity conservation in government programmes, policy and action and made provisions for the establishment of the UK Biodiversity Steering Group. The Government's response was forestalled somewhat by the publication of *Biodiversity Challenge* in 1993 by the major environmental NGOs (Wynne *et al.* 1993) who then consolidated their position (Wynne *et al.* 1995). In contrast to the Government's initial offering, this document emphasised the need for action plans and target-setting. The UK Biodiversity Steering Group were charged with developing a comprehensive plan of action and, together with input from the NGOs, worked on a series of detailed action plans. These plans were published in two UK Steering Group Reports

(HMSO 1995a, b) and were endorsed by government in 1996. The UK Biodiversity Action Plan has the major goal:

> *"To conserve and enhance biological diversity within the UK, and to contribute to the conservation of global biodiversity through all appropriate mechanisms."* (HMSO 1995a).

The starting point for this process was the development of action plans for species and habitats. Species were selected for action plans using the following criteria:

- "their numbers or range have declined substantially in recent years; or
- they are endemic; or
- they are under a high degree of international threat; or
- they are covered by relevant Conventions, Directives or legislation." (HMSO 1995a).

This process resulted in a provisional list of 1250 species. Species Action Plans (SAPs) were prepared for 116 of the most threatened species and a further 286 were prioritised.

A similar process was adopted for habitat conservation. Initially 37 broad habitat types were identified to provide a basic framework for the entire land surface of the UK. These 'biodiversity habitat categories' were based upon mixtures of broad habitat types and their physical characteristics in what was termed a 'biotope complex' approach. Within these broad habitat types key habitats were selected for action plans by using the following criteria:

- "for which the UK has international obligations; or
- which are at risk, such as those with a high rate of decline especially over the last 20 years, or which are rare; or
- which may be functionally critical; or
- which are important for key species." (HMSO 1995a).

Habitat Action Plans (HAPs) were initially prepared for 14 key habitats and a further 24 were proposed. Most, but not all, key habitats are listed as Annex 1 types in the EU Habitats Directive.

The Biodiversity Convention requires information to be gathered on a range of components related to biodiversity, the starting point is Article 7 of the convention (Figure 2.1).

Figure 2.1 Article 7, Convention on Biological Diversity.

Article 7 states that each contracting party shall, as far as possible and as appropriate, for the purposes of *in situ* conservation, *ex situ* conservation, and the sustainable use of components of biological diversity undertake the following:

- identify the components of biological diversity important for conservation and sustainable use;
- monitor through sampling and other techniques the components of biological diversity paying particular attention to those requiring urgent conservation measures and those offering the greatest potential for sustainable use;
- identify processes and categories of activities which have, or are likely to have, a significant adverse impact on the conservation and sustainable use of biological diversity and monitor their effects through sampling and other techniques;
- maintain and organise by any mechanism, data derived from identification and monitoring activities relevant to the above measures.

In the identification of biological diversity regard should be given to:

- ecosystems and habitats:
  - containing high diversity, large numbers of endemic or threatened species, or wilderness;
  - required by migratory species;
  - of social, economic, cultural or scientific importance;
  - that are representative, unique or associated with key evolutionary or other biological processes;
- species and communities that are:
  - threatened;
  - wild relatives of domesticated or cultivated species;
  - of medicinal, agricultural or other economic value;
  - of social, scientific or cultural importance;
  - important for research into the conservation and sustainable use of biodiversity, such as indicator species;
- described genomes and genes of social, scientific or economic importance.

After HMSO 1995a.

The Ramsar Convention, 1971

The Ramsar Convention was agreed by contracting parties in Iraq in 1971 (Peck 1996). Signatories agreed to designate wetland areas of international importance according to internationally agreed criteria. There are currently 128 classified Ramsar sites in the UK covering an area of 551,191 ha (JNCC 2001a); they are protected through designation as Sites of Special Scientific Interest.

The Bern Convention, 1979

The Bern Convention requires the protection of endangered and vulnerable species and their habitats in Europe. Like other instruments, various appendices list relevant species. In the UK the Convention was ratified in 1982; legal implementation has been through the Wildlife and Countryside Act 1981/85. There is a requirement to gather information on threatened habitats.

The Bonn Convention, 1979 & 1994

The Bonn Convention requires the protection of migratory animals, from the Arctic to Africa. As with the Bern Convention, after ratification in the UK in 1985, legal implementation has been through the Wildlife and Countryside Act 1981/85.

Natura 2000

There are two major European Directives that impact on nature conservation policy and legislation: the 'Birds' and 'Habitats' Directives. Together they have resulted in the identification of Special Protection Areas (SPAs) and Special Areas for Conservation (SACs) respectively. Collectively SACs and SPAs form a network of protected areas across Europe known as Natura 2000. The UK Government's chosen legal instrument for the protection of SPAs and SACs was the notification of land as Sites of Special Scientific Interest (SSSIs), but SPAs and SACs are afforded some additional protection through European Law.

Birds Directive

The European Community's Directive on the Conservation of Wild Birds (79/409/EEC) requires member states to safeguard the habitats of migratory birds and concentrates on specified threatened species. The Directive was ratified under UK law under the Wildlife and Countryside Act 1981/85 and the Conservation (Natural Habitats, &c.) regulations 1994 (HMSO 1994b) and committed the Government to taking measures to preserve, maintain and re-establish a sufficient diversity and areas of habitat for all species of naturally occurring birds in the wild state. This objective includes the designation of SPAs that are designed to protect wild birds and to provide sufficient diversity of habitats to maintain populations at a viable and sustainable level. There are currently 187 classified SPAs in the UK covering an area in total of 764,198 hectares (JNCC 2001a).

Habitats Directive

The 'Habitats Directive' is implemented in the UK through the Conservation (Natural Habitats, & c.) Regulations (HMSO 1994b) and stemmed from the European Union Council Directive on the conservation of natural habitats and of wild fauna and flora (92/43/EEC) which entered into force in June 1994. This directive requires member states to identify natural habitats and species designated as being of 'Community interest' that occur within their geographic jurisdiction (DOE 1995). Community interest is specified in a range of technical annexes and species or habitats considered to be under particular threat within the EU were identified as 'priorities' and afforded stronger protection. Priority Habitats are listed in Annex 1 of the Habitats Directive which provides a description and classification of relevant habitats. Annex 1 also affords some protection to rare species other than birds and therefore extended the provisions of the Birds

Directive to other (non-avian) species. Once identified, rare species or habitats (in the case of species the land that they inhabit) are required to be designated as Special Areas for Conservation (SACs) and member states are obliged to maintain, or achieve in these, a 'Favourable Conservation Status'. The identification and monitoring of Favourable Conservation Status created a significant information requirement for UK nature conservation agencies. Not only were they required to identify and propose SACs, but they were also required to have knowledge of their condition and the trends for change in that condition. The priority habitats were described along phytosociological lines according to the long-standing tradition on the continent (e.g. Braun-Blanquet 1932). These required translation into National Vegetation Classification NVC communities (Rodwell 1991a, b, 1992, 1995, 2000) and provided an instant requirement for habitat information that was not necessarily readily available.

NNRs and SSSIs

The origins of NNRs and SSSIs were in the 1949 National Parks and Access to the Countryside Act (Green 1985). One of the results of this Act was the establishment of the Nature Conservancy that was mandated to identify and notify a suite of National Nature Reserves (NNR). The 1949 Act gave the Nature Conservancy powers to buy or lease land, or enter into management agreements with land owners to manage land as NNRs. In addition to these powers, provision was also made to enable the Nature Conservancy to notify special areas of land. After notification, local planning authorities were obliged to inform the Nature Conservancy of any forthcoming planning applications for land development. Under the 1981 Wildlife and Countryside Act (later amended in 1985), these special land areas became known as Sites of Special Scientific Interest (SSSI). NNRs were intended to represent the best examples of different habitat types across the country. They differ from SSSIs in that they have to be managed appropriately to retain their status. To date, SSSIs have been protected only by a presumption against development. Appropriate management could not be enforced by the Nature Conservancy and its successor bodies, and only be encouraged by voluntary management agreements.

SSSIs are designated for either biological or geological interests. The biological SSSI series was intended to form a national network of areas to conserve the total national 'special interest' of a representative range of habitats and their associated flora and fauna in Great Britain. The effectiveness of the SSSI network has been the subject of much debate (Porter 1985, Rice 1994, Rowell 1991) and the subject of much political and policy controversy because of the relatively weak sanctions available for the protection of their primary features of interest.

The guidelines for SSSI selection are complex and inter-related (NCC 1989, JNCC 1994a, 1996) and are based upon selection rules developed for the Nature Conservation Review (Ratcliffe 1977). Primary criteria include; size, diversity, naturalness, rarity, fragility and typicalness; secondary criteria include; recorded history, position in an ecological or geographical unit, potential value and intrinsic appeal. At the time of writing, 6,235 SSSIs have been identified in Great Britain (JNCC 2001b).

The 1999 Countryside and Rights of Way Bill came into force in December 2000 and placed a statutory duty upon public bodies to manage SSSIs to conserve and enhance their value. In contrast to the 1981 Act, English Nature and the Countryside Council for Wales will have the power to impose a management scheme upon owners/occupiers of SSSIs to enforce appropriate stewardship. Additionally, the legislation provides heavier and more wide-ranging penalties for those who damage SSSIs than was included in the 1981 Act and 1985 amendments.

Sites of Importance for Nature Conservation

Local Planning Authorities are responsible for identifying second tier sites, generally known as Sites of Importance for Nature Conservation (SINCs), but also known as Sites of Scientific Interest (SSIs) and by other names on a local basis. These are similar in concept to SSSIs, but are intended to represent local characteristics and variation. There is a presumption against development in these areas, but no further legal protection.

Related information requirements

Changing policy emphasis or reporting requirements can mean that information requirements change. The new requirements of the Convention on Biodiversity and the EC Habitats Directive are good examples. Almost overnight, these two initiatives imposed new reporting requirements on the Statutory Conservation Organisations (SNCOs) and other public bodies. In the case of the latter, the SNCOs were required to identify Priority Habitats and report to the EC. In many cases, these Priority Habitats were already represented in the existing SSSI series, but, under the new definitions and priorities, many areas with relevant habitat were not.

Table 2.1 shows the summary information requirements listed in the UK biodiversity Steering Group report. In the Biodiversity Convention and other conventions and directives the relationship between species and habitats is not clear-cut. Species both live in habitats and are part of habitats. The protection of species cannot be isolated from their particular habitat requirements. The Convention on Biological Diversity differed from previous conservation initiatives in that it considers biological diversity wherever it occurs. Other conservation measures apply only within protected areas. The Convention on Biological Diversity has helped change the emphasis of conservation effort to the wider countryside, outside protected areas as part of a growing recognition that biological diversity simply cannot be protected within small isolated patches of land. This change of emphasis has increased the demand for area-wide habitat data. In the past inventory type surveys had been conducted by bodies such as the NCC to identify habitat resources and to locate important areas with a view to protecting them. The motivation for identification is still area protection but within a wider network of ecosystem and landscape conservation.

Table 2.1 Information requirements under International Conventions and EC Directives.

Information requirement	EC Birds Directive	EC Habitats Directive	Ramsar Convention	Bonn Convention	Bern Convention	Biodiversity Convention
Gather information on	Wild birds	Habitats & species	Wetlands & species	Migratory species	Threatened habitats	All components of biodiversity
Maintain and organise data	Yes	Yes	Yes	Yes	Yes	Yes
Monitor	Bird population levels	Habitats & species	Wetland & species	Migratory species	Threatened habitats	All components of biodiversity
Collect information on designated sites	SPAs	SACs	Wetland sites	-	-	-

After HMSO 1995a – Table 2, p. 28.

2.1.2 'Countryside' organisations

There is an almost bewildering array of different countryside and environment related organisations, many of which are providers and/or users of countryside information. A brief synopsis is provided here to describe the main actors in Great Britain, their major responsibilities and activities.

Department of the Environment Transport and the Regions

The Department of the Environment Transport and the Regions[1] (DETR) is the Government Department with overall responsibility for environmental protection and wildlife conservation. Its other major responsibilities include transport, planning and local government. In Scotland, this role is conducted by the Scottish Executive and in Wales by the National Assembly for Wales. Responsibility for agriculture is held by the Ministry for Agriculture, Fisheries and Food (MAFF). DETR is aided by a number of executive agencies (e.g. English Nature, Countryside Council for Wales) which officially are constituted as non-departmental government bodies. The Countryside Agency is an executive agency formed from a merger of the Countryside Commission and the Rural Development Commission in April 1999, which provides advice on the countryside to DETR and government. Within the DETR the Wildlife and Countryside Directorate is responsible for nature-conservation related matters. It manages a countryside research programme, intended to supplement current knowledge and address specific issues and manages the Natural Environment Research Council (NERC). DETR publishes the Digest of Environmental Statistics; this includes sections on land use and land cover and on wildlife. Information is drawn from the Countryside Survey and data provided by the SNCOs. DETR provides other departments and ministers with guidance on environmental and nature conservation issues. At a policy level, the Department ensures that the Government is meeting international and European obligations.

[1] After the general election of June 2001, the new government reorganised the DETR, splitting transport from environment and amalgamating it with the old MAFF thereby creating a new Department for Environment, Food and Rural Affairs. The implications of this are as yet (June 2001) unknown, but the broad responsibilities in relation to the countryside are likely to remain unchanged.

DETR provides local planning authorities with planning guidance and publishes Planning Policy Guidelines (PPGs). The current guidance on nature conservation and planning issues is the now rather dated PPG9. However, the Government's recent White Paper *Our Countryside: the future* (DETR 2000) included a commitment to issue revised planning guidance.

Ministry of Agriculture Fisheries and Food, Farming and Rural Conservation Agency

The Ministry of Agriculture Fisheries and Food[2] (MAFF) is the Government Department responsible for agriculture and related industries. It does not directly formulate nature conservation policy, but arguably has a great impact upon it. Together with the Farming and Rural Conservation Agency (FRCA), MAFF provides farmers with guidance on agricultural practice and administers price support policies for the industry. As agricultural price support policies move towards emphasising environmental protection rather than production, the role of MAFF and FRCA will become increasingly related to conservation issues. FRCA is an Executive Agency of MAFF and the National Assembly for Wales (NAW) and administers grant schemes, many of which are designed to promote conservation. To administer these schemes various countryside designations have been introduced. Environmentally Sensitive Areas (ESAs) have been defined to designate areas where traditional agricultural husbandry has created and sustained particular landscapes and habitats. Within these areas, grants are available to farmers and land managers who enter into management agreements. These agreements limit the range and types of inputs allowed in return for financial support. Monitoring of these agreements and their effects upon the landscape and habitats is conducted by FRCA specifically for ESAs. FRCA have developed a GIS system for use by their policy staff and regional project officers. The system includes boundaries for all relevant designations (e.g. ESA, National Park, SSSI) tenure information, holdings within land management agreements and where available habitat/vegetation data. This system is used by staff to administer and manage grant applications and to check for eligibility and relevant constraints. This application is currently probably the most advanced use of GIS by a government agency for supporting day-to-day policy and management implementation. DETR has recognised the importance of integration of environmental geographic information and in its 2000/2001 Countryside Research Programme requested expressions of interest for a large-scale project, Multi-Agency Geographic Information for the Countryside (MAGIC), to develop an inter-agency GIS to manage and share information on respective agency interests in countryside features, schemes and designations.

The Statutory Nature Conservation Agencies

The Nature Conservancy was established and given powers and duties as a result of the National Parks and Access to the Countryside Act 1949. Established by Royal Charter, the Nature Conservancy was an advisory and land holding governing council (Green 1985, NCC 1984). The Nature Conservancy established National Nature Reserves and under Section 23 of the 1949 Act identified other protected land 'areas' that were the precursors of Sites of Special Scientific Interest. Included in the original charter was a research function and the Nature Conservancy established a number of research stations around the country to conduct ecological and hydrological research. In 1965 the Government established two national research institutes, the Social Sciences Research Council and the Natural Environment Research Council (NERC), and brought the Nature Conservancy within the remit of the NERC. Because of its executive and land holding functions it was thought that the Nature Conservancy did not fit well within the NERC and in 1973 the Institute of Terrestrial Ecology (ITE) and the Nature Conservancy Council (NCC) were established under the Nature Conservancy Council Act. ITE took on the research functions and remained within the NERC. The NCC continued to provide advice, scientific guidance and conduct executive functions across Great Britain with a strengthened autonomy (NCC 1984).

The 1990 Environmental Protection Act (HMSO 1990) included provision for the division of the NCC which was seen by many interests as being overly effective in its opposition to development proposals (Young 1993). At the same time, the Countryside Commission was also sub-divided. Three new successor bodies were created and began operation in April 1991. In England, the NCC's functions were taken on by English Nature, fulfilling a similar role to the old Nature Conservancy. The Countryside Commission was reduced to working in England. In Wales, the remit was taken up by the Countryside Council for Wales (CCW) that also incorporated the previous functions of the Countryside Commission in Wales. The same model was used in Scotland where Scottish Natural Heritage (SNH) assumed responsibility for the NCC's and Countryside Commission's previous activities. There was much opposition to the break-up of a national nature conservation body and in response to claims that nature conservation would lack scientific co-ordination and standards, the Government established the Joint Nature Conservation Committee to promote national standards and to oversee UK-wide issues.

In Great Britain, the Statutory Nature Conservation Organisations (English Nature, Countryside Council for Wales and Scottish Natural Heritage) have similar core functions, they provide government and other public bodies with advice on nature conservation and natural features, manage NNRs and administer the SSSI network, provide grants for environmental conservation and undertake a limited amount of research. In addition to these core functions, CCW and SNH have responsibility for functions inherited from the Countryside Commission including recreational activities, landscape planning and public access.

---

[2] From June 2001 MAFF ceased to exist as a separate department and became part of the new Department of Environment, Farming and Rural Affairs (DEFRA).

Natural Environment Research Council

The Natural Environment Research Council (NERC) was established in 1965 (as described above). In 1973 its applied research functions were adopted by the Institute of Terrestrial Ecology (ITE), the Institute of Hydrology (IOH) and the Institute of Freshwater Ecology (IFE). In 2000, the ITE and IFE were merged to form the Centre for Ecology and Hydrology (CEH) still working within the NERC. CEH's role is to conduct ecological research on behalf of the NERC. CEH (and previously ITE) have been responsible for major research and survey work on ecology and habitat distribution, type and change. They have conducted a series of nationwide countryside surveys to proved data on habitat distribution and change. These are further discussed below.

Countryside Agency

The Countryside Agency was established in 1999 from the merger of the Countryside Commission and the Rural Development Commission. The Countryside Agency works towards the conservation and enhancement of the countryside, promotion of social equity and economic opportunity and to help all people enjoy the countryside. The Countryside Agency publishes indicators on the state of the countryside (e.g. Countryside Agency 1999) and together with English Nature and English Heritage, developed a set of countryside character areas (Natural Areas for EN) which divided England into 159 distinct areas based upon landscape elements. Following the launch of *Access to the countryside in England and Wales: The Government's Framework for action* (DETR 1999b) and the Countryside and Rights of Way Bill 2000, the Agency is working on defining and mapping areas of mountain, moorland, downland, heath and registered common land for the proposed new public right for open access. This will be a major task as there are few available data sources that adequately define and map the chosen landscape categories (Clark *et al.* 2000). The Countryside Agency uses habitat data to monitor the effectiveness of National Parks (MLCNP), to assess the state of the environment and to manage and monitor recreational access provision.

National Park Authorities, Association of National Park Authorities

The origins of National Parks in England and Wales lie with public pressure for access to the countryside in the 1930s, the formulation of firm proposals from the Addison Committee in 1931 and the report produced by John Dower in 1945. Dower's report suggested National Parks should be:

an extensive area of beautiful and relatively wild country in which, for the national benefit and by appropriate national decision and action,

i. the characteristic landscape beauty is strictly preserved,

ii. access and facilities for public open-air enjoyment are amply provided,

iii. wildlife and buildings and places of architectural and historic interest are suitably protected, while

iv. established farming use is effectively maintained. (after Smith 1995).

Dower proposed ten National Parks and these were ratified by the Hobhouse Committee that also suggested that the South Downs and the Broads should be included. The 1949 National Parks and Access to the Countryside Act provided for the establishment of the ten Parks suggested by Dower during the 1950s. In 1989, the Broads effectively became a National Park in all but name and there are current proposals to add the New Forest too. The Lake District and Peak District National Parks were established as independent planning boards; the other eight parks were constituted as special planning committees administered by constituent local authorities. The Association of National Park Authorities (ANPA) was established in 1992 as a result of the Edwards Report (Countryside Commission 1991a), produced by the National Parks Review Panel. This body was proposed to give National Parks greater autonomy and an independent voice at a national and international level. There are currently no National Parks in Scotland, the closest equivalent being National Scenic Areas. The National Parks were set up primarily on the basis of landscape and scenic quality but by defining land using such measures encompassed some of the most environmentally valuable land in England and Wales. Much of the land within National Parks is designated under other schemes such as SSSI and SAC.

Forestry Commission

The Forestry Commission is the government department responsible for forestry in Great Britain. The Commission is composed of a board of commissioners appointed by the Queen with powers and duties established by statute. The Forestry Commission works separately in England, Wales and Scotland but represents Great Britain after consultation. The work of the commissioners is directed by MAFF and conducted by two executive agencies, Forest Enterprise and Forest Research. The role of the Forestry Commission has evolved since its establishment early last century. The aims of its work have developed from a purely resource management and production role to include greater emphasis on environmental and recreation protection and provision. The Commission's role is now stated to: protect and expand Britain's forests and woodlands and increase their value to society and the environment.

The Forestry Commission's objectives are to:

- protect Britain's forests and woodlands;
- expand Britain's forest area;
- enhance the economic value of our forest resources;
- conserve and improve the biodiversity, landscape and cultural heritage of our forests and woodlands;

- develop opportunities for woodland recreation;
- increase public understanding and community participation.

(Forestry Commission 2001).

The change in emphasis in the Commission's role means that they now require more information on the ecological value of their holdings and the impacts caused by forest creation and management. The Forestry Commission produces large quantities of data, mainly for internal purposes. Many of these data are forest mensuration statistics, but nationwide woodland censuses have also been conducted. Like other statutory agencies, the Commission is required to report on the condition of SPAs and SACs and contribute to the Biodiversity Action Plan (BAP) process.

Environment Agency

The Environment Agency was established in 1996 from the National Rivers Authority and Her Majesty's Inspectorate of Pollution (HMIP). The Environment Agency is mainly responsible for flood defence and pollution controls and enforcement. As well as using land cover, land use and habitat data for flood planning and prediction the Environment Agency are producers of habitat related data. They produce river surveys based upon NCC Phase 1 methodology and use a variety of remote sensing techniques (e.g. CASI and LIDAR see Section 2.2.3) to capture data on rivers, estuaries and coastal areas.

Local Authorities

Local Authorities use habitat-related information in many ways, often dependent upon the type of local authority. Local government have undergone a number of re-organisations over the last thirty years which has resulted in different hierarchical structures in different areas of the country (i.e. county, district, parish and unitary). The administrative detail of the different arrangements is not relevant here, but it should be noted that statutory functions such as planning regulation are conducted at different levels in local government. Local Authorities require habitat related information to manage land holdings, to formulate countryside management strategies, to provide input to Local Biodiversity Action Plans (LBAPs), to support state of the environment reporting and support sustainability initiatives. Habitat data are also required when commenting on development proposals: development control is dealt with by district and unitary councils and mineral planning regulations are dealt with by county and unitary authorities. The impact assessment of development control proposals is usually based upon local site (SINC) registers and referral to local team staff from the SNCOs. Where formal environmental impact statements are required, these are usually commissioned from consultants who rely upon a mixture of existing sources and their own field survey.

National Trust

The National Trust is unique in its role and constitution and most land owned by the Trust is unalienable and held in perpetuity on behalf of the nation. There is a separate National Trust for Scotland with similar powers and constitution. The Trust is a major land-holding organisation for historic properties and countryside estates, land holdings and coast. The National Trust employs a large number of countryside staff including ecologists and countryside managers. The Trust conducts biological surveys of its properties using its own staff and shares information with the relevant SNCOs.

Non-Governmental Organisations

In the latter half of the twentieth century non-Governmental Organisations (NGOs) have grown enormously in popularity, membership and influence (Pepper 1989). This category includes a variety of different organisations ranging from those with an emphasis on campaigning to those more closely involved in land management and a number who encompass both roles. Recently, even campaigning groups have attained a degree of 'insider' status politically (e.g. Royal Society for the Protection of Birds [RSPB], Friends of the Earth) as part of the Government's Biodiversity Steering Group. The NGOs' requirements and uses of habitat-related data consequently vary according to their roles and also their sectoral interests. Sectoral interests range from broad-scale environmental campaigning (e.g. Friends of the Earth, World Wide Fund for Nature) to specific species-related interest (e.g. Herpetological Conservation Trust, Butterfly Conservation). Campaigning organisations such as Friends of the Earth and Plantlife use data on habitat occurrence, distribution and change to support their various campaigns and initiatives. Collective groupings of NGOs such as the Environment Challenge Group have used a range of data to develop environmental indicators (e.g. MacGillivray 1994). More narrowly targeted interest groups such as the Herpetological Conservation Trust are interested in the habitats that support the species they are concerned with (i.e. lowland heathland). Each of these different NGOs uses habitat-related information in different ways. Little, if any, work appears to have been conducted on NGO's use of information, but the author's experience suggests that information is used to support site acquisition and management decisions in land-holding organisations and to support research and policy work in the larger NGOs such as the World Wide Fund for Nature (WWF) and the RSPB. The requirement for good quality information has been great enough to have prompted two organisations to conduct their own reviews of available information sources. The Council for the Protection of Rural England (CPRE) and the RSPB have both conducted reviews of available habitat, land cover and land-use information (Gilbert and Gibbons 1996, Sinclair 1992). Some NGOs are data providers in their own right. To date the major providers have been species-related organisations, specifically ornithological groups. The RSPB, the Wildfowl and Wetlands Trust (WWT) and the British Trust for Ornithology (BTO) are all major providers of species data. Few data on habitat distribution appear to be produced by NGOs, probably due to the prohibitive costs of doing so. Habitat maps and data are produced by some organisations such as the RSPB but these are restricted to nature reserves and land managed for conservation. Hence, most NGOs are likely to be net users rather than suppliers of habitat data.

The Wildlife Trusts Partnership, managed by the Royal Society for Nature Conservation (RSNC), are both campaigning and land holding/managing groups. In the early 1980s the RSNC developed a classification for rapid field-by-field habitat survey and mapping which later became, with slight modification, the Phase 1 classification (NCC 1990). This methodology is discussed in more detail in the following section. The Wildlife Trusts are primarily concerned with the identification and management of reserves but in addition to this function conduct an active campaigning role.

National Biodiversity Network

The National Biodiversity Network (NBN), a consortium of conservation organisations including the JNCC, the SNCOs, the Natural History Museum and various NGOs, was formed to co-ordinate the collection and dissemination of information in recognition of the fact that effective action on biodiversity should be underpinned by relevant and accurate information. It aims to address the problems created by the dispersed location of data sources and the lack of consistent recording and management standards. The NBN is attempting to address these issues through the Linking Local Record Centres initiative and the NBN Gateway project. At the time of writing the Gateway project is at an experimental stage. Its aim is to provide data access by means of the provision of discovery metadata through the Index of Biodiversity Information Sources and direct hyperlinks to data through the Gateway. The Gateway is intended to provide a seamless portal to different databases. However, the technology is yet untested and the Gateway pilot simply serves data amalgamated in a single database. While the NBN Gateway may provide some technological fixes in providing access to disparate data sources, the main problems still appear to lie in the supply of data. The concept behind the NBN model is to produce detailed data at a local level, according to nationally agreed common standards. These data will then be made available through Local Record Centres and will be available for aggregation at regional and national levels to meet reporting requirements at those scales. However, despite some commonality, the strategic information requirements for users at these different scales may well differ. In biodiversity reporting terms, national plan makers require information of habitat extent, location and status across the entire country. Local managers will require the same information, albeit in more detail, at a local level but will also need national information to provide context and quantification.

2.1.3 Summary

- There are a range of organisations who act as data suppliers and/or users. In many cases these roles are poorly defined.

- The variety of organisations, coupled with the complexity of their responsibilities and activities, means that there is a broad range of information needs. These requirements are related to a range of different purposes, including policy, management and information provision; each of these uses is likely to imply rather different requirements in terms of data characteristics.
- The relative lack of homogeneity between the purposes and objectives of different information suppliers and users is also likely to lead to important (though perhaps often subtle) differences in the character and ultimate utility of the available data. Consequently, there is not necessarily a straightforward relationship between data needs and data supply.

2.2 CHARACTERISTICS OF HABITAT-RELATED DATA

2.2.1 Habitat – the concept and measurable characteristics

Strictly speaking the term habitat relates to a place or environment and comprises the sum of abiotic elements: the interaction of climatic, edaphic, biotic and anthropogenic factors. However, the term habitat has become common parlance for referral to vegetation types or characteristic groups of communities – more formally referred to as biotopes – the composition of which is determined by the interaction of abiotic and biotic factors. Because of the widespread adoption of this more common definition of the term habitat, this will be used throughout the thesis, except where distinctions between the two concepts need to be made.

The concept of a plant 'community' is an abstract one and is really a human construct. A plant community is a collection of species within a circumscribed area. The size and location of this area and the way in which species are recorded within it pose key issues for ecologists. Historically there have been two schools of thought on the 'organisation' of plant communities. Clements (1928) proposed an essentially deterministic model where the plant community was in effect a super-organism that had a complex but predictable series of possible life stages that ultimately resulted in the achievement of a 'climax' state. This theory led to the concept of vegetational succession, which Clements thought was mainly determined by climatic factors. In contrast, Gleason (1927) described plant communities as random assemblages of species. Gleason recognised the existence of plant associations but suggested that species interactions were not determined but part of a process of continual casual processes more strongly influenced by abiotic conditions. Both hypotheses recognised the potential for equilibrium states, but that these were rarely stable. The modern view of community process is more closely aligned to Gleason's position. However, the concept of vegetation succession has become enshrined in ecological thinking, albeit with greater emphasis on the effects of abiotic factors and enshrining a less deterministic and more individualist view of plant species.

If the composition of plant communities is primarily based upon abiotic characteristics, how can these be used to explore the concept of habitat? Various theories and principles have been suggested to describe or model the distribution and abundance of plant species. These include the competitive exclusion principle, the niche concept and various community diversity measures. These different models differ in their assumptions, mainly in terms of the importance they confer upon equilibria, stability and scale. Just as the description and modelling of plant community composition is complex at a fine scale, the same issues are faced at larger scales when considering the distribution, abundance and interaction of the communities themselves at a 'habitat' level. It is clear that the distinctions drawn are inevitably somewhat arbitrary and should always be considered in terms of scale and the reasons behind their formulation.

Krebs (1978) identifies five internal characteristics of communities that can be measured. These properties are not relevant at the species level, only the community level:

i. species diversity – a measure of species richness or diversity;
ii. growth form and structure – describing community type by morphology;
iii. dominance – species successful in determining the nature of the community by virtue of their size, abundance or characteristics;
iv. relative abundance – the relative numbers of species;
v. trophic structure – the characteristics of energy and matter flows from plants to herbivores to carnivores.

These characteristics are important for the assessment of habitat condition and quality, but less important for habitat identification. However, discrimination of these factors is vital in classification systems, the description of communities or habitats and thus their identification and quantification.

The habitats of nature conservation interest are a product of complex and long-term interactions between 'natural' ecological factors and exogenous anthropogenic influences. Environmental conditions such as soil chemistry and hydrological properties, climatic characteristics and herbivory rates help determine the range of plants and plant communities that can potentially inhabit a specific locality. Anthropogenic influences operate through actions such as control of grazing intensity, disturbance through cutting or felling, tillage or burning and nutrient inputs. Because of such actions, natural succession can be modified, halted or arrested and this can influence the composition and distribution of different habitat types. The dynamic nature of habitats was described by Watt (1947) as their 'pattern and process' and consideration of habitat data should always be tempered with the fact that we are considering a dynamic construct.

The concept of habitat must therefore be considered both in terms of land cover (physical/objective composition) and land use (a more anthropocentric concept based upon productive systems and types). Thus, the acquisition of habitat-related information is intricately bound with both land cover and land use concepts. Traditionally the focus of different surveys has been divided by these concepts (though in reality many land cover classifications tend to include elements of land use, and vice versa). If our aim for policy and management purposes is to identify and quantify habitats and the ways in which they change, what properties should be considered? Data relating to habitats can usefully be considered to refer to four fundamental properties of habitats – their typology, location, quantity (extent/frequency) and quality. Each of these habitat properties is examined below.

a) ***Typology*** refers to the common group or class to which any habitat is assumed to belong. This grouping is normally defined through the use of typological definitions or classifications. The nomenclature used in specifying these typologies also helps to provide a common language for communication, and thus to ensure that the same feature, characteristic or form is measured and described in a consistent way. By their very nature, classifications represent the imposition of a generalised or simplified view of a system or continuum of characteristics present in reality. The form and nature of the generalisation or simplification has in turn implications for the appropriate application of that classification and therefore they are often use-specific. Although the use of a definition or classification inevitably introduces some types of bias and generalisation into data, its usefulness as a common language for communication tends to outweigh the disadvantages. Habitats are described by a variety of classifications (see later discussion) based upon different semantic models. Some relate primarily to land cover and are based upon categories identifiable using remote sensing observation tools (e.g. Land Cover Map of Great Britain). Others, such as the Phase 1 habitat classification, use a mixture of land cover and land use characterises for categorical definition. Phytosociological approaches, such as the National Vegetation Classification (NVC), depend upon detailed floristic and frequency occurrence data recorded from field survey. The typology employed will influence the suitability of a data set in different applications. The implications of any typology for monitoring change need to be recognised. In particular, the accuracy of any typology (i.e. the proportion of sites or cases that are correctly classified) tends to be inversely proportional to the number of classes. Thus, typologies with many classes are likely to have higher error rates when applied to any area (Causton 1988); broad typologies, with few classes, tend to cause fewer misclassifications. On the other hand, changes within any class (i.e. changes which do not cause the habitat to move from one class to another) are unlikely to be detected or reported because a typology will only detect change in the characteristics measured and expressed. Thus, the sensitivity of any typology to change depends – amongst other things – on the number and narrowness of its classes. Broad typologies, are likely to be less sensitive to change; more detailed typologies tend to be more sensitive. A logical consequence of these characteristics is that typologies suitable for expressing measures of stock may be insensitive indicators of habitat change.

b) ***Location*** refers to actual place of a habitat in space at a point in time. Although most environmental phenomena are generically three dimensional, location is generally represented by points, lines or areas in two dimensions or by attribution/relationship to an otherwise described feature. Location data are most useful for legal purposes (e.g. to define ownership or responsibilities) and for referential context, and to assess movement or change. In many cases, small amounts of locational change are critical in monitoring applications and the assessment of geographical covariation. Location is important in research terms to help understand ecosystem function and process – i.e. to investigate the relationships between plant communities and environmental gradients – and in applied terms for management and for policy and reporting. Locational information is needed for management purposes to calculate areas under management, accessibility and resource inputs, and also to direct management activities. Absolute or relative locations are needed in policy and strategy areas to investigate the inclusivity/performance of protected areas and proximity to potential threats. It is important to note that in many applications the accuracy of relative location or topology can be more important than absolute location. Topology characterises the logical spatial relationships between objects. Clearly, this is important in situations where it may be critical to know whether a protected species or habitat is inside or outside a protected area. In the case of nested hierarchical representations of habitat areas it is important that data are spatially consistent with the logic implied by the classification employed. This factors are especially important when considering change through time-series data, if absolute or topological consistency is not suitably maintained, monitoring can be confounded by errors of omission and commission in addition to any actual change in reality.

c) ***Quantity*** refers to extent (length, area, volume) or frequency (number) of habitat patches. Area/extent is important ecologically, though there has been much debate over patch size and species occurrence interactions (e.g. Crawley 1987, Greig-Smith 1983, Williamson 1981), which are complex and operate over long periods. It does, however, seem clear that fragmentation of once more widespread habitats can affect their stability, species composition and long-term viability. The loss of semi-natural habitat in Great Britain, due to causes such as changes in farming practice, urban development and hydrological management, has led to the fragmentation of habitats. In a physical sense smaller areas are potentially more susceptible to influence by surrounding land uses, in part because the potential reservoir of individuals available for genetic mixing and population replacement is also reduced. Insufficiently mobile species reliant on specific habitats or habitat components can find themselves 'marooned' physically and genetically within remnant habitat patches. The way in which quantity is recorded will have a direct bearing upon what information can be derived from the data, or any definition or classification that uses those data. It is therefore important to measure quantities in a form relevant to the intended use for the data. For example, if edge features are important in a specific habitat, more appropriate data might be obtained by measuring perimeter instead of area. As with typological systems, the resolution at which quantity data are measured can have profound implications for the utility of the data when comparing change over time. This is very important when monitoring change, since the resolution of the data must be finer than the smallest significant change, if change is to be reliably detected (Birnie *et al.* 1995). For example, if a minimum area threshold is applied to features, this may cause marked under-estimation where the size distribution of features is skewed and lead to an information 'black hole' relating to features below this threshold limit. A good example of this relates to woodlands in West Yorkshire. The author conducted a woodland survey for the local authority in a 100km^2 area (Tantram 1994). The mean woodland size for broadleaf woodlands in the Colne Valley management area was found to be 2.88ha and the modal size was 0.5ha; 48 broadleaf woodlands were identified making a combined total of 140ha. The Forestry Commission's Woodland Census uses a minimum woodland size of 2ha and consequently the vast majority of woodlands in the Colne Valley would be omitted. Such a threshold therefore fails to describe the nature of the habitat resource and would also fail to detect any change in the resource over time.

d) ***Quality*** – habitat quality is a rather subjective concept that is related to anthropogenic interpretations of natural or semi-natural characteristics. Ratcliffe (1977) described the primary criteria for quality assessment of habitat as size, diversity, naturalness, rarity, fragility and typicalness. Secondary criteria included recorded history, position in an ecological or geographical unit, potential value and intrinsic appeal. After their use in the Nature Conservation Review, these criteria became enshrined in British nature conservation assessment. Many of the ecological characteristics important for conservation are quality-related and these are inherently difficult to describe, especially when change in habitat quality is important. To describe quality, specific and identifiable features must be chosen, which are based upon accepted definitions and classifications. These are often taken one step further with the attachment of values to different quality classifications. These values can form an ordinal scale (e.g. river water quality classes) or can be expressed as thresholds over or under which different states are reached, using these methods change can be expressed in relative terms. Such thresholds are sometimes based upon clearly identifiable or observable phenomena, or, more often, against a negotiated or derived value that is often notional. This method is employed by the statutory conservation agencies in common standards for monitoring change and is embodied in the concept of 'Favourable Condition' used for assessing change in SSSIs and SACs (Rowell and Reed 1994). Thresholds can be used to trigger a response, either of more detailed monitoring, or management action (as with the 'Alert Limits' and 'Limits of acceptable change' principles).

In addition to these basic characteristics, the measurement and detection of habitat change requires the ability to identify differences in these characteristics between two different points in time. This is often difficult because:

i. all surveys are subject to sampling and measurement error and thus subtle changes often cannot be distinguished against this background of uncertainty – many apparent changes are in reality noise in the survey data;

ii. survey methods and classifications change so that direct comparisons often cannot be made;

iii. many changes are subtle and are not detected by existing survey methods (e.g. because of the spectral/spatial resolution of satellite data);

iv. gross changes in stock/extent in one direction in one location are often counter-balanced by changes in other in another direction elsewhere. If this happens at a sub-detection unit level data on stock at different times will mask these transfers and changes.

2.2.2 Data collection methods

Just as there are different habitat characteristics that can be measured, there are a variety of techniques for data collection. At the topmost level, surveys can be categorised as either samples, where a proportion of the population is recorded, or as censuses where the entire population is measured (Cochran 1977). In either case, survey data may be obtained from a variety of sources including field surveys, remote sensing and secondary sources (e.g. cadastres).

The original purpose of a data set will determine what was recorded, where and how and will largely determine the data collection methods employed. Where detailed information on the composition or quality of a habitat is sought a sample-based approach may offer the best means for data collection, particularly if the distribution of a habitat resource is relatively well known. However, in many cases resource inventories are needed. For example, the Lowland Heathland Inventory (EN/RSPB 1994/5/6) used existing sources to identify, map and quantify the lowland heathland resource. Similarly, the Lowland Wet Grassland Resource Survey (Dargie *et al.* 1995) sought first to identify a particular habitat resource, and then to seek additional information related to the geographical areas identified. In such cases, where the size and distribution of the population (habitat resource) under investigation is unknown, census-based recording methods can provide a suitable approach to data gathering.

There are both advantages and disadvantages to sample and census-based approaches. The advantages of sampling (Cochran 1977) can include reduced costs and greater speed because of the smaller number of measurements. Cochran also argues that sample approaches can offer greater accuracy because a smaller number of highly trained personnel can produce results that are more accurate.

Table 2.2 Advantages and disadvantages of census and sampling approaches.

Type	Advantages	Disadvantages
Census	Complete inventory recorded Potentially zero levels of sampling error Can be used for distribution mapping Less complex statistics are required to interpret results	Time-consuming Expensive Long turn-around times Potential increase in measurement error/consistency
Sample	Quicker to complete Less expensive than census Can produce accurate estimations of stock and distribution Shorter turn-around times Potentially lower levels of measurement error	Bias can be introduced by sampling framework Cannot be used for accurate distribution mapping Often requires complex statistical methods to analyse results and variance

However, there are also drawbacks to the sampling approach. In particular, in some circumstances, survey or measurements are so specialised that there are only a limited number of personnel who can conduct the survey; in these cases a census approach may not be possible and a sample approach is the only way forward. Samples must also be representative of the population and, particularly when little is known about the population, this can be problematic. Statistical methods must also be used to extrapolate from the sample data to give overall estimates of quantities or trends in the wider population and these are inevitably based upon a further set of assumptions about the distributional characteristics of the population.

Censuses are often adopted where there is little information available about the population. The SNCOs have produced a number of such inventories over the last fifteen years, notably for ancient woodland (Spencer and Kirby 1992, Thomas and Phillips 1994), saltmarsh (Burd 1989), lowland heathland (EN/RSPB 1994/5/6) and lowland wet grassland (Dargie *et al.* 1995). Such inventories tend to employ a broadly inclusive habitat definition and aim to identify and delineate target habitat areas. They are used primarily for resource identification and quantification. They tend to provide little detail on the quality and composition of habitats. By their very nature, quality assessments must be more closely linked to objectives. Resource constraints play a large part in the selection of recording and survey methods. In the past, habitat inventories have adopted a broad-brush approach because of the prohibitive costs of conducting nation-wide detailed field surveys. The highest quality and most detailed results may be obtained by intensive field-based survey, but this is labour-intensive and hence expensive. For this reason, field survey is often limited to areas of known interest or priority or employed as part of a sample-based approach. Thus, the decision to adopt a sample or census-based approach must be considered in terms of information requirements, time available and resource availability.

To date, the SNCOs have followed census-based methods to produce habitat-based inventories. This is because censuses offer the advantages of greater knowledge about the overall population and are helpful where distributions or quantities do not follow known or easily predictable distributions. The disadvantages of censuses are that the measurement of the entire population can be a time-consuming and slow process and consequently tends to be a very expensive one. Large surveys over long periods are difficult to manage and it is difficult to maintain common survey and recording standards. A good example of the problems of census survey quality is illustrated by the variability in Phase 1 survey quality between different counties and often within the same county (Wyatt 1991). In Shropshire two survey teams recorded semi-improved/improved grassland with different levels of rigour/detail and where their mapping met displayed inconsistencies. The problem of observer bias in ecological survey, and Phase 1 in particular, has not been fully investigated. In a limited study comparing six different Phase 1 surveyors, however, Cherrill and McClean (1999) found agreement between pairs of maps of only 25.6%: these tentative results suggest that further work is required on the sensitivity of census data to observer bias. Because of their time and expense, censuses tend to gather only general information. The Ancient Woodland Inventory (AWI) records ancient woodland sites and where these are still vegetated, felled or replanted. Because keeping such a large data source up-to-date is so resource intensive, the inventory is updated on an *ad hoc* basis as new information is received and verified. In recognition of this it is permanently accorded the status of being 'provisional' and consequently cannot be used to measure resource change or loss. The Lowland Heathland Inventory, similarly, records only a few variables and is based upon existing (and therefore heterogeneous) survey, as was the Lowland Wet Grassland Resource Survey.

Despite their shortcomings, these inventories are very valuable for nature conservation purposes and the SNCOs continue to develop similar information sources for other habitats. In conjunction with others, the author has worked on the Lowland West Grassland Resource Inventory for English Nature, the Floodplain Grassland and Coastal Grazing Marsh inventory for CCW and a Maritime Cliff and Slope Habitat Inventory for English Nature. These information sources are generally based upon existing but dispersed and heterogeneous data and are required to service operational and reporting requirements of the SNCOs. They are needed because the SNCOs have no alternative information sources and are required to react to evolving reporting requirements. But despite this need, many more habitat types remain undocumented than those that are documented. The situation for those that have inventories is not entirely rosy, as there are often problems with data sets. These mainly revolve around confidence in the data, lack of content and current relevance. The lowland heathland and saltmarsh inventories, for example, are predominantly based on data over ten years old.

If censuses are problematic due to their scale and unwieldy nature, can sampling approaches offer a more manageable and useful alternative? Notwithstanding the work of the SNCOs, the largest single source of habitat-based data in Great Britain is the Countryside Survey Series (e.g. Barr *et al.* 1993). The Countryside Survey Field Survey is based upon a stratified random sample of 1 km squares across Great Britain and uses extrapolation to predict vegetation distribution and quantification across the country on the basis of the field samples. If extrapolation is effective, it offers a method for resource estimation that has the advantages of a sample-based approach and provides inventory-like results. For such a scheme to be effectual and produce reliable and comparable results, both the stratification and extrapolation methods must effectively represent the type and range of diversity in the variables to be recorded and modelled. The Countryside Survey is discussed in more detail in Section 2.3.2, below.

2.2.3 Observation methods

There is a range of survey methodologies available for habitat-related survey, whether sample- or census-related. The particular observation tools employed in surveys are often a reflection of the original purpose of the survey. The observation tools employed to gather land cover and land use information in Great Britain can be divided into three main categories: field survey, remote sensing and record-based/cadastral.

Field Survey is often used to gather primary data and to validate remote sensing or air photo data. It can be more accurate than these particularly for the determination of vegetation composition and the accurate identification of land use. The two main problems with field survey are data quality and expense. If large numbers of surveyors are used there is often variability in the way they identify, attribute and record observations. The field mapping of habitats can be highly variable due to variability in operator behaviour and methods of defining and mapping habitat boundaries (e.g. Angold *et al.* 1996, Cherrill and McClean 1999). Field survey is, by definition, labour-intensive. Training and logistical requirements mean that data acquisition costs are therefore high. Field survey techniques are used extensively by the Ordnance Survey (OS) for topographical survey providing the national mapping base for the UK. Field survey is also used extensively for applications where detailed land use is required or where other techniques can not offer sufficient detail or resolution. The Land Use Change Statistics (LUCS) are based upon OS survey revision returns and provide accurate interpretations of urban land use which is difficult to achieve by other means, but provide little timely information on change in more rural areas. Most vegetation surveys for nature conservation purposes are based upon field survey in order to identify plant communities in sufficient detail to meet conservation objectives and reporting requirements.

Remote sensing data are most often based upon satellite data. The most common sources used in Great Britain are Landsat Thematic Mapper, SPOT multispectral and panchromatic imagery. The most significant use of satellite data has been the application of Landsat TM imagery for the ITE Land Cover Map of Great Britain (Fuller *et al.* 1994), and its successor the Land Cover Map 2000 (Haines-Young *et al.* 2000). The use of Monitoring Landscape Change (MLC) project (Hunting Technical Services 1986) was one of the first to pilot the use of satellite imagery. The accuracy of landcover classification is affected by the time of year that the imagery was captured, the MLC project found problems in acquiring cloud-free imagery, resulting in the use of seasonal scenes. The overall classification accuracy for a Thematic Mapper (TM) scene ranged from 39 to 80%. One method for overcoming this problem is to classify the same area from scenes derived from the same area at different times of year to overcome seasonal differences. This technique was employed for the ITE Landcover Map of GB where both winter and summer scenes were used in a multi-temporal approach that also employed images from a number of years to acquire cloud-free imagery. The Environment Agency has made much use of the Compact Airborne Spectrographic Imager (CASI) for applications such as land classification, monitoring, pollution detection and estimation of aquatic chlorophyll levels and suspended solids loads. The CASI is installed in a small aircraft and can be configured to record a full swath width instantly. CASI can detect up to 288 spectral wavebands from the visible to near infra-red regions of the spectrum and spatial resolution can be varied between 1 and 10m. By using geo-positioning equipment in the aircraft, the digital output from the CASI can be processed and mapped in a matter of hours. The Environment Agency has also been conducting feasibility studies on the use of Synthetic Aperture Radar (SAR) and Laser Induced Direction and Range (LIDAR) for digital elevation measurement. SAR can be used to generate ground imagery day or night in varied weather conditions. SAR generates an image from ground-reflected radar energy which can be two-dimensional or three dimensional by measuring the phase difference of signal returns. LIDAR is an airborne terrain mapping system which uses laser technology to measure aircraft to ground distance. Vertical accuracies of 10-15cm can be achieved with a horizontal resolution of 1-4 metres. The equipment is being tested by the Environment Agency for the measurement of coastal profiles, coastal erosion and flood risk.

Aerial photography is another remote sensing tool most often used at local and regional scales, but has also been used for national sample surveys such as MLC and for the census-based Land Cover of Scotland project. Traditionally, high-resolution (e.g. 1:10,000 scale) 3-D monochromatic photography has been used for habitat mapping, but in recent years greater use has been made of colour photography. Air photo interpretation provides the basis for landcover monitoring in designated areas such as National Parks and Environmentally Sensitive Areas (ESA) because it can provide detailed information (often resolving features not discriminated by satellite imagery) for relatively large areas. Additionally, it has been used by the Forestry Commission for Forest Census. Air photo interpretation is the suggested rural data source for the proposed National Land Use Stock System (NLUSS), (Dorothy Salathiel, pers. comm. 1997). The Monitoring Landscape Change in National Parks (MLCNP) project (Countryside Commission 1991b) reported air photo interpretation agreement with field survey data at 87%. The MLC project reported an overall accuracy of interpretation of 87% based upon some control sites and field data (Hunting Technical Services 1986). These figures correspond well with the figure of between 80 and 90% quoted by Rhind and Hudson (1980).

Record-based/cadastral sources can generally be considered as secondary or statistical data sources, often based upon a variety of source data and often containing added interpretation. In many cases, data are reported by landowners or land holders, either under statutory obligations (e.g. the MAFF 'June returns') or through voluntary questionnaire surveys. These sources are most useful for providing additional information on land use or historical land use (e.g. Ancient Woodland). There are a number of problems with these sources such as the lack of consistency between different classifications and geographical areas and a consequent lack of comparability. Because returns are made by non-experts, misclassification may also be high. For reasons of confidentiality, such data are also often made available only in an aggregated and anonymised form. These various issues were highlighted by Sinclair (1992) in work for the CPRE where attempts were made to reconcile various land cover and land use survey statistics.

Because it is often difficult or impossible to verify records for accuracy or to deconstruct aggregated totals for geographic areas, statistical sources have limited uses.

2.2.4 Nomenclature

We have touched upon the issue of typology at the theoretical level, but the classification and recording of habitat type, land cover or land use has practical ramifications for the application and utility of the resulting data. Classifications must yield categorisations suitable for the intended purpose. The type of classification or nomenclature employed will affect the results obtained and the comparability with other data sources using different classifications. While observation tools govern the size and/or shape of features that can be identified, the classification into which they are placed affects reporting outputs. The nomenclature must delimit issues such as the definition of a particular type or characteristic, for example 'woodland' or 'industrial', and the units with which these are recorded. It must also set out how mixtures and mosaics are dealt with at different scales. In order to address these issues, classification schemes have been developed in a number of ways (Causton 1988):

i. Hierarchical or reticulate.

   Hierarchical classifications are employed in species taxonomy and are common in the categorisation of plant communities. In most cases, the hierarchy implied is largely semantic and not reflected in any real-world relationships. In reticulate classification, groups are identified separately as members of a network and there is no vertical structure implied. In reality, hierarchical relationships are often scale-dependent. At fine scales some habitat elements may be present in different classes and this can compromise the integrity of the hierarchy. For example, the habitat heathland actually comprises a range of biotopes including ericoid heaths, acid grassland mosaics, scrub, valley mire and bog. Together they are generally recognised as lowland heathland, but, at a finer scale, specific areas can be classified as, for example, bog. Such bog types are related by many characteristics to bogs that may also be recognised at the (higher) 'habitat' level. In this case, the relationship is more reticulate than truly hierarchical.

ii. Divisive or agglomerative.

   These options are derived from two main approaches to classification, discriminant or divisive and *a priori* or agglomerative. The former uses statistical methods of discriminant analysis to identify groups within measured data on the basis of similarity or dissimilarity (e.g. ITE land classification or Brabyn 1996). Such techniques are statistically more robust than *a priori* methods and support a hierarchical approach but are susceptible to producing statistical artefacts where categories are separated by only one or two subtle variables. *A priori* or agglomerative approaches seek to develop categories from generic principles and observable characteristics, and to fit measurements into these by combining recognisable groups to make classes. They can display logical inconsistencies, poor normalisation and are less amenable to the imposition of hierarchies.

iii. Monothetic or polythetic.

In monothetic classifications single attributes are used to distinguish related groups. By contrast, polythetic classifications employ multiple attributes to determine class membership.

These characteristics provide the theoretical framework for possible classification systems, but we also need to consider the practical application in data capture. Rhind and Hudson (1980) described a 'desiderata' for a classification system:

i. the classes must be mutually exclusive, i.e. any geographical individual can only fall into one class;

ii. it has to meet the detailed needs of the primary user, who may have paid for the survey;

iii. it has to meet as many of the needs of secondary users of the data as is possible, concomitant with (ii);

iv. it has to be easily understood and applied;

v. it has to produce repeatable results with use by different surveyors using the same survey technology. Both this and (iv) imply the need for the classification to be explicit, i.e. well-documented;

vi. it has to be exhaustive in that all 'geographical individuals' under consideration must be classifiable – even if only in an 'other uses dustbin';

vii. it has to be hierarchical, to cope with surveys at differing levels of resolution in differing areas;

viii. it has to be structured in such a fashion that, if different survey technologies are ever used, the results from both can be compared (this overlaps with (v), but is not identical to it);

ix. it has to be sufficiently stable for surveys carried out at different moments in time to be compared;

x. it has to be sufficiently flexible for new interests and tasks to be met from a modified, rather than a completely new, classification;

xi. it must incorporate some recognition of seasonal or other cyclic changes so that aliasing is not embalmed in the results;

xii. wherever possible, it must be based upon quantitative criteria (thus contributing to (v)).

Achieving each and all of these criteria is virtually impossible in reality. As previously mentioned, classifications and therefore surveys remain heavily influenced by their primary purpose or emphasis. There is often a rural/urban split and land cover or land use types that are of peripheral interest to the main purpose are often excluded or conflated into general categories – Rhind and Hudson's 'other uses' dustbin. Rhind and Hudson imply, but do not make explicit, the fact that good classifications should be internally consistent. If they are not then the scope exists for misclassification or over-use of the 'other uses' category.

Table 2.3 Major habitat-related classifications in use in Great Britain.

Title/Ref.	Description	Structure	Review
Phase 1 (NCC 1990)	A field survey based classification for identification and mapping of land cover classes. Some elements, such as *amenity grassland* and *improved grassland* introduce land use criteria. Placement of vegetation in some categories requires multiple criteria and complex judgements.	Hierarchical, but recognises mosaics which are reticulate in nature but included in hierarchical sub-divisions.	The system was in active use and development until 1990. No formal review procedures appear to exist from this point.
NVC (Rodwell 1991a *et seq.)*	The NVC is a phytosociological classification based upon divisions made firstly according to functional, abiotic and morphological characteristics and secondly by purely floristic criteria.	Hierarchical.	No formal process.
Annex 1 (HMSO 1994b)	Produced to support the Habitats Directive, Annex 1 describes communities of European importance. Navigating the hierarchy is complex and requires multiple decision criteria.	Hierarchical, reflecting origins in CORINE.	Driven by EU legislation, may be related to update of CORINE.
BAP (HMSO 1995a)	A classification developed to describe important habitat types for biodiversity reporting and to make links with the Annex 1 system while also maintaining consistency with the Phase 1 classification.	Hierarchical, but conformance with Annex 1, key habitats and Phase 1 categories has introduced anomalies and some reticulate relationships.	The UK Biodiversity Action Plan (HMSO 1995a) refers to a review policy.
ITE Land Classification (Bunce *et al.* 1996a, b, c)	A complete classification of all 1km grid squares in GB into one of 32 land classes. The classes were based upon the statistical analysis of environmental characteristics from existing map data	Hierarchical.	No explicit review policy.
CS Reporting Classes 1990 (Barr *et al.* 1993)	A complete classification incorporating vegetation, land use and land cover criteria. Includes explicit quantitative definitions and thresholds.	Employs a basic hierarchy to avoid reticulate relationships. Complex decision criteria for classification at three levels.	No explicit review policy.

In Great Britain there are a variety of sectorally-led habitat, land cover and land use surveys that, in most cases, employ different classifications; these are outlined in Table 2.3. This issue is probably the greatest barrier to comparability between different sources. To help overcome these problems the former Department of the Environment (DOE) commissioned (the former) Institute of Terrestrial Ecology (ITE) to produce a comparison of landcover definitions which also resulted in a computer programme (LUCID) to aid comparison of land cover classifications in the UK. LUCID is distributed with the published report (Wyatt *et al.* 1993) and as part of the Countryside Information System (CIS), a software system that displays and analyses land cover data. The Land Cover Definitions work established a 'baseline' hierarchical classification with which other land cover and land use classifications could be compared. The creation of the baseline classification greatly simplified comparisons of different classifications as it provided the baseline classification as a reference against which to compare other systems. It therefore reduced the number of comparisons that would need to be made. Inevitably, the agreement between different sources that employ different observation tools, scales and nomenclature cannot reach 100% and great care should be taken when considering definitions used in different classification schemes. In many ways, the work by Wyatt *et al.* (1993) offers the nearest equivalent to an integrated classification scheme, but it has not been adopted by any other practitioners.

Other land classification schemes have also been used in Great Britain. The National Land Use Classification (NLUC), for example, represented an attempt by local authorities to standardise classifications developed from the 1940s onwards so that standardised reporting could be made to the DOE from 1974. The scheme was abandoned some years later due to lack of support. The DOE's Land Use Change Statistics (LUCS) are derived from a principally urban hierarchical classification based upon 24 essentially object-based categories. LUCS are derived from OS field survey where surveyors allocate a feature (from 651 OS land use features) to one of the 24 LUCS categories. An object-based approach such as this, used in conjunction with object structured digital maps, offers two principal advantages (Dunn and Harrison 1994):

- an 'atomistic' approach – rather than allocating a land area to broad categories, all features can be classified according to a detailed framework which provides flexibility in that specified attributes can be grouped into user-defined classification schemes.
- spatial flexibility – used in conjunction with a digital map, land use coded objects offer spatial flexibility for combination and classification of both attribute characteristics and spatial characteristics.

A similar version of this basic approach is being piloted by English Nature in the South West of England to develop an atomistic object-based GIS model for BAP Broad Habitats (Keith Porter pers. comm. 2001). Such an approach would clearly offer many benefits in a number of applications, but the accuracy with which new classifications could be built still depends entirely upon the nature and scale of constituent objects that the classification scheme recognises at its most basic level. To date, no widely adopted habitat-based classifications have used this approach. The development and application of different classification schemes has been *ad hoc* and sectoral. This has resulted in a range of different schemes for different purposes or at different times. In assessing different information sources, it is necessary to compare data from different classifications. As outlined above, different approaches have been made to address this issue. Because different classifications employ different methods for defining class boundaries, and therefore class membership, comparing categories across different classifications is not straightforward because the relationship between even superficially similar categories from two different classifications are rarely one-to-one. In reality, up to five possible relationships are identifiable (for categories (a_1, b_1, c_1...a_2, b_2, b_3...):

i. one-to-one (1:1) – where a category in classification a is directly comparable to one in b ($a_1 \equiv b_2$);

ii. one-to-many (1:m) – where a category in classification a is equivalent to one or more entire categories in classification b ($a_1 \equiv b_2$ or $a_1 \equiv c_2$) or ($a_1 \equiv b_2 + c_2$);

iii. many-to-one (m:1) – the reverse of ii;

iv. many-to-many (m:m) – where two or more categories from scheme a can be related to two or more categories from scheme b ($a_1 \equiv b_2$ and $a_1 \equiv c_2$) or ($b_2 \equiv c_1$ and $b_2 \equiv a_1$);

v. fuzzy – where categorical relationships cannot be clearly identified.

The ease and value of comparison between categories from different classifications depends upon these relationships. Any ambiguity in the relationship diminishes the value of comparison along with the degree of reliability and certainty in the transaction. Because different classifications often relate to different basic entities, there is not necessarily any logical level of equivalence in the relationship. This may be because hierarchical levels are not comparable or because selection criteria are not equivalent. For example, the Phase 1 classification includes categories for improved grasslands, one of which, amenity grassland, is also land use derived. In contrast, the Countryside Survey 1990 uses quantitative criteria such as the percentage of *Lolium* present in the sward to identify agricultural improvement. In such cases, the categories are clearly related, but only in a fuzzy relationship, because the entities, in addition to their descriptions, are not equivalent. It should be emphasised in this context that nominative equivalence (i.e. comparability in the naming or definition of classes) may be misleading, if the methods by which these definitions are applied vary from one data source to another (e.g. if the survey methods, sampling framework or selection criteria vary).

In contrast to hierarchical classifications, a substantially different approach is offered by the former Countryside Commission's and English Nature's Countryside Character and Natural Areas initiative. This divided England into a set of Countryside Character areas on the basis of a nominal characterisation with the purpose of producing a map of England depicting the natural and cultural dimensions of the landscape (Brooke 1994). Natural Areas are used by English Nature, which requires a less detailed landscape characterisation, and were formed from aggregated Character Areas. Each Character Area has its own unique landscape character as defined by its ecological, historical, cultural and landscape features. The map was based upon a large set of land use, land cover and cultural variables that were collated for all land 1x1 km squares in England. These were used to produce a characterisation; so in classification terms the relationship between character (or Natural) areas is truly reticulate. The Character Map is used by both the Countryside Agency and English Nature as a strategic framework for policy development. It is intended that the initiative will help identify opportunities for agri-environment schemes and in the siting and composition of new woodland planting within the overall framework of retaining and strengthening local landscape character. One of the main strengths of the project was the creation of supporting data sets; however, no metadata were generated for these data and the copyright position for different data sets is complex and generally confused.

National Vegetation Classification (NVC)

The NVC is a quantitative phytosociological classification for assigning vegetation types to plant communities (Rodwell 1991a *et seq.*). Although primarily designed as a classification of vegetation composition, it is also used for vegetation mapping. The classification was based upon the analysis of c. 35,000 vegetation composition and abundance samples taken from quadrats located within 'homogenous' stands of vegetation. These data were analysed using multivariate methods and divisive polythetic analysis with Indicator Species Analysis (Hill 1979, Hill *et al.* 1975). Only floristic, and not environmental, data were used and emphasis was placed upon obtaining realistic and observable end-groups rather than strict adherence to statistical outputs.

The NVC provides a field survey method; quadrats (of different sizes related to the vegetation type) are located in homogenous vegetation stands. The presence and abundance of plant species are recorded using the DOMIN scale and the results are tabulated in constancy tables. These tables can then be used in conjunction with keys, descriptions and tables provided in the published NVC volumes (Rodwell 1991a *et seq.*) to assess similarity with recognised communities or sub-communities. The NVC has become the *de facto* standard for Phase 2 survey by the SNCOs. It is used as in the selection of biological SSSIs (NCC 1989) and is widely used for descriptive and reporting purposes.

BAP Broad Habitats

Broad habitat categories were developed by the Biodiversity Steering Group for reporting purposes (HMSO 1995a). Broad habitats were developed to identify important habitat types within a general framework consistent with the Phase 1 habitat classification (NCC 1990). The classification was also intended to be consistent with habitats as described by Annex 1 of the 'Habitats Directive' so a second tier (below Broad habitats) was introduced, known as 'Key' habitats. Membership of either tier is derived from a range of variables including environmental, vegetation physiognomy, floristic composition, land cover and land use. Broad Habitats are used by the statutory nature conservation agencies for reporting requirements and are used by all agencies involved in the BAP process. In recognition of the importance of this, the Countryside Survey 2000 (Haines-Young *et al.* 2000) has used Broad Habitats as reporting categories as opposed to the main reporting classes used in 1990. This has necessitated the re-engineering of the CS90 (and earlier) data into the BAP classes for reporting change in CS2000. It is anticipated that further detail will be made available as the results for CS2000 are published.

Countryside Survey Main Reporting Classes (1990)

The Countryside Survey (Barr *et al.* 1993) is a comprehensive and ongoing survey of the British countryside (described in further detail in Section 2.3.2). The Countryside Survey 1990 (CS90) used its own 'Main Reporting Classes' for recording land cover for mapped land or landscape features. The field recording system for CS90 was essentially open-ended and allowed surveyors to pick terms from a large predetermined list of coded primary terms and secondary qualifiers (Barr *et al.* 1993, Appendix 2). Surveyors used primary terms to provide general descriptions of features and secondary terms to provide greater detail. Where more than one primary code was used the first listed reflected the dominant land cover/use.

Table 2.4 CS90 main reporting classes.

No.	Category name	No.	Category name
1	Wheat	31	Unmanaged lowland grassland and tall herbs
2	Barley	32	Dense heath
3	Oats	33	Open-canopy heath
4	Other cereals	34	Berry-bush .heath
5	Maize	35	Drier northern bogs
6	Turnips/swedes	36	Saturated bogs
7	Kale	37	Conifer woodland
8	Oil-seed rape	38	Mixed woodland
9	Other crucifer crops	39	Broadleaved woodland
10	Peas	40	Shrub
11	Field beans	41	Felled woodland
12	Other legumes	42	Inland rocks and screes
13	Sugar beet	43	Still water
14	Potatoes	44	Running water
15	Other roots and beets	45	Wetland
16	Other non-horticultural field crops	46	Inter-tidal soft coast without vegetation
17	Horticulture	47	Saltmarsh
18	Non-cropped arable	48	Dune
19	Woody perennial crops	49	Unvegetated hard coast
20	Non-agricultural mown grass	50	Maritime vegetation
21	Recently sown grass	51	Railway
22	Established perennial rye-grass swards	52	Road
23	Well-managed perennial rye-grass mixtures	53	Agricultural buildings
24	Weedy swards with 25-50% perennial rye-grass	54	Residential buildings
25	Permanent non-intensive grass	55	Continuously built land
26	Semi-natural calcareous grass	56	Vegetated waste land, derelict land and allotments
27	Upland grass	57	Hard areas without buildings
28	Bracken (>50% cover)	58	Quarries and other extractive industries
29	Molinia moorland	59	Sea/estuary
30	Non-Molinia moorland and mountain grass		

After Wyatt *et al.* 1993

These data were processed and the combinations of codes that describe land use and land cover features were aggregated into 58 exclusive categories – the CS1990 Reporting Classes; the exact detail of the aggregation process is not described in the published literature. The 58 classes were essentially the same as the 59 categories used by the Land Cover Definitions work (Wyatt *et al.* 1993) and differ only in the sub-division of built and coppice woodland categories. In the majority of cases, the 59 reporting classes are referred to because the work by Wyatt *et* *al.* defined these as a finite set of categories to which CS90 records could be clearly related with little ambiguity. The main reporting classes (Table 2.4) were designed to be a simple hierarchy that avoided reticulate relationships. Considerable attention was given to the explicit quantitative definitions of particular classes (e.g. percentage of *Lolium* in improved swards) to try to avoid subjective class attributions. The CS90 field survey data as published in the Countryside Information System (CIS) makes exclusive use of the CS90 reporting classes and these must therefore be used for any analysis or reporting of CS90 using CIS.

2.2.5 Data quality

Data are usually gathered for specific purposes and each purpose is likely to imply different data needs and consequently different characteristics. Because of this, data gathered for one purpose may not be suitable for another. Data quality is therefore not an absolute characteristic but a contingent one that is application-related. Information on data quality should thus be an integral component of the description of the data (Goodchild 1993). Different types of data may be collected in different ways: which is most effective depends upon the use for which the data are intended. In an ideal world data could be gathered specially every time they were needed. However, data collection is expensive and consequently data are often re-used to help mitigate or recover costs. For example, Great Britain did not produce a CORINE land cover map in the same way as other European countries; instead, ITE's Land Cover Map of Great Britain was re-classified to produce CORINE compatible categories (Fuller and Brown 1996, Seabra 1996). In an increasingly market-driven environment, where data are also a commodity, many organisations seek to re-sell data for a variety of uses and applications. The use of data in different applications, particularly where there is a divergence from the original purpose, can create tensions between data characteristics and data requirements. Overall, therefore, the utility, quality and cost-effectiveness of any data depend primarily on their fitness-for-purpose. The assessment of 'fitness' is thus specific to the chosen application and is a relative rather than absolute concept: it must be seen as the product of the data needs of the user, on the one hand, and data characteristics, on the other.

The quality of habitat-related data can be considered in terms of a number of different properties of the data. The specific properties determining the quality of any individual data set (and their relative importance) will inevitably depend on the particular use for which the data are intended, as noted above, but the following factors are generally applicable:

i. Thematic relevance – the appropriateness of the variables measured to the intended use;

ii. Temporal relevance – do the measures relate to an appropriate timescale?

iii. Spatial relevance – are the data related to the intended geographical area?

iv. Resolution – do the data measure variables at a suitable spatial resolution?

v. Accuracy – do the reported data give a true measure of reality within acceptable tolerances? It should be noted that accuracy is inherently multi-dimensional: it refers not only to simple accuracy of quantification, but also the reliability of feature identification, logical consistency, spatial consistency (e.g. topology) and repeatability.

These data characteristics are affected by events throughout the complete lifecycle of any data: from the initial data creation (recording) stage, through subsequent data processing and reporting. At any stage, errors can occur and interpretation or generalisation may be introduced.

At all stages of data creation and handling, standards are needed to ensure consistency, repeatability and correct interpretation for all these data types and their relevance to different applications. To aid this process metadata[3] are required. The development of metadata in ecologically related applications has been relatively slow. In Great Britain, applicable standards are being developed by the National Biodiversity Network (NBN 2001) and the National Geospatial Data Framework (NGDF 2000). In the United States, the Federal Geographic Data Committee (FGDC 1997) has also produced standards for metadata. In each case, the standards pertain to discovery metadata. Discovery metadata (Morgan 1999) relate to records primarily designed to provide an overview of an information resource's scope, content, format and availability. As such 'discovery metadata' catalogues (e.g. the NBN Index of Biodiversity Information sources [NBN 2001]) are akin to library catalogues (Online Public Access Catalogues – OPACs). 'Creation' metadata, in contrast, provide lineage information describing the creation, processing and management of a data resource or the characterisation of sources, transformations and input/output relationships between products as described by Lanter (1990). To date, little work on creation metadata standards appears to have been completed. The adoption of metadata standards and publishing of metadata for data sources by data providers is rare. If metadata standards were widely adopted, and metadata routinely published, then the task of assessing relevant data sources for providing habitat information for policy makers and environmental managers would be relatively straightforward. Unfortunately, this is not generally the case. The identification and assessment of possible data sources for use in this project, for example, generally had to be completed by reference to secondary published sources such as other reviews of data sources, by interviewing data providers and users and by the author's personal knowledge.

Access to data is clearly crucial in determining their utility. Data access, however, depends on a number of factors. First and foremost, of course, is the question of whether relevant data actually exist. Non-existence of data is clearly fundamental, but it may not be definitive. If a strong need exists, and the capability to provide those data is available, then in time the data are likely to be generated. If the data cannot be provided, users will typically resort to

[3] Metadata are essentially 'data about data' (Morgan 1999).

alternative methods – to proxy data or to other ways of assessment (e.g. modelling or estimation). The second factor is knowledge about the existence of the data sources that do exist. If data sources are not catalogued or published, and if good discovery metadata do not exist, references to their existence may be few and far between and largely based upon the personal knowledge of a few professionals. The problem is less acute for habitat-based data, which are usually held and disseminated by well-established organisations, but species records are often held by a handful of specialists and their knowledge often dies with them (Burnett *et al.* 1995). Thirdly, potential users must have knowledge about the characteristics of the data that are available – the name of the data source, its content and relevance and where it can be obtained. Accessibility of data is also constrained by the format(s) available. Many data sources have not been digitised or are available only in obscure digital formats. A more subtle constraint lies in reporting outputs. Raw data are often generalised into an attenuated or generalised form, thereby reducing their utility for some applications. The opposite situation can also present access problems: if only raw data are available, the size and complexity of the data source may preclude their use. Categorisation and generalisation for reporting can also reduce the utility or relevance of data sources. A fourth aspect is related to cost. Again, this is a practical but very real restraint. Many have argued that the recent pricing policies of the Ordnance Survey have held back the widespread adoption and development of GIS (Robinson 1999). Individual data users interviewed for this project also felt that the pricing policies of the (then) ITE precluded wider use and peer-review of data resources.

2.2.6 Summary

- The concept of plant communities is a fluid one; consequently so are the related concepts of biotopes and habitats.
- It is possible to identify and measure various characteristics of habitats to categorise and quantify those of interest.
- The nature of habitat-related data and their wider utility are governed by the nomenclature, observation methods and techniques used to collect them. A variety of methods have been developed to measure and record habitat data; each represents/imposes a particular view of 'reality' upon the results.
- The use of different classification methodologies and different nomenclature can cause major comparability problems between different data sources.
- Data quality is an important criterion for the value assessment of data sources. However, there are few absolutes in this assessment; data quality is intimately related to fitness of purpose. Data quality relates to both data content and data format. Adherence to standards and transparency throughout the data creation process are essential.

2.3 DATA SOURCES – QUALITY, ORIGINS AND RELEVANCE

There are a wide range of data sources relevant to providing habitat-related and countryside information. In addition to habitat-based or vegetation surveys, land cover and land use sources can also provide relevant information. Although many land cover and land use surveys may not resolve detailed characteristics of vegetation, they can often provide useful information for inventories of for targeting further field survey. Many different organisations and agencies compile vegetation, land cover and land use information by using various techniques. Survey methodologies vary greatly according to the requirements of the commissioning organisations and the nature of the information needed. The specific focus of different surveys thus varies; they may provide information on the distribution of a specific habitat, or land cover in rural areas or map specific land use activities.

To date the integration of land classification systems beyond immediate policy purposes has been fragmentary (Birnie *et al.* 1995 and Wyatt *et al.* 1993). Specific sectoral or regional studies have been conducted which offer greater levels of detail on specific themes or geographical areas. In general terms, the availability of data is determined by the range and strength of policy interests and the costs relating to collecting them. This development pattern has led to sectoral development that in most cases meets narrow policy needs, but in many cases can prove divisive and hinder wider integration. This section examines the specific characteristics of data sources commonly used for habitat resource estimation and related activities.

2.3.1 Methods

The review of data sources presented here was conducted in collaboration with the CLAUDE (Co-ordinating **La**nd **U**se and cover **D**ata in **E**urope) Concerted Action Project for DG XII of the European Commission, funding from which provided partial sponsorship of this research. CLAUDE aimed to develop an internally consistent Europe-wide plan for land cover monitoring, land use monitoring and research and to link with other relevant international programmes. Work for the UK inventory was conducted by the author, as part of this study, and comprised two main strands: the collection and review of published literature and contract reports, and the conduct of a number of project interviews with key data suppliers and users.

Background and generic information was sought from the accounts by Sinclair (1992), Gilbert and Gibbons (1996) and Dunn and Harrison (1994). These were supplemented by 'grey literature' and notes supplied by data suppliers and users. Together they were used to generate data source profiles describing the nature, coverage, type and availability of data sources. In addition to these desk-based reviews, feedback from data users and suppliers was sought. In conjunction with the CLAUDE project, key data suppliers and users were identified to provide a sample of the major data-related issues encountered. Because a limited amount of time was available, effort was concentrated on senior staff that were experienced in the use and/or supply of data in their respective organisations. Organisations were chosen to cover different areas of work and a variety of applications. The review was not intended to be either quantitative or comprehensive, but to help identify data issues of relevance. Individual responses were not identified to enable issues to be discussed frankly. The staff selected were members of the (then) Department of the Environment (staff working on land use change, countryside change and countryside survey), English Nature (staff working on habitat reporting, habitat inventory and geographic information management, the (then) Institute of Terrestrial Ecology (staff working on countryside survey), and the (then) Countryside Commission (staff working on countryside change and indicators). In total eight staff were separately interviewed, at their organisation's offices. During the interviews respondents were asked to complete a form documenting the details of data sources supplied and/or used and to answer specific questions relating to data supply and/or use for individual sources. The proformas for data suppliers and users are included in Appendix 1. The results from the interviews are interpreted qualitatively here and included in the user's assessments for each data source where the strengths and weaknesses of data sources are outlined based upon comments made by respondents.

2.3.2 Description and review of available data sources

Data sources are used by a variety of organisations. In many cases, the specific information users require is only available from a limited number of sources. Although many information suppliers are also the main users, there is typically little vertical integration between suppliers and users so that many of the users of a data source have little influence over the manner in which the sources are collected, reported or distributed. This is mainly due to the sectoral development pattern followed in GB but also reflects, to some extent, an apparent lack of communication and co-operation between suppliers and users.

It is difficult to categorise habitat, land cover and land use data sources. Some authors categorise according to policy imperative, others by data collection method and others by intended use. In conjunction with the fact that there is some variation in the titles given to data sources it is also difficult to ensure exhaustiveness when comparing different sources of information. Table 2.5 shows a list of significant habitat, land cover and land use data sources for GB. At this stage, the distinction between habitat-based, land cover and land use has not been made, as this would introduce somewhat arbitrary and artificial divisions. The characteristics of the major data sources existing in GB and a brief summary of each follow in this section. Although many sources are listed in Table 2.5, relatively few of these sources are widely used or generally available.

Table 2.5 Major land use, land cover and habitat-related data sources in Great Britain.

Survey	Thematic coverage	Geographical coverage	Frequency (latest)	Sample or census	Data collection (major source)
Countryside Survey 1990	Comprehensive	GB (rural)	1978, 1984, 1990, 1998	Sample	Field survey
Land Cover Map of Great Britain	Comprehensive	GB	1990, 1998	Census	Satellite
Land Cover of Scotland	Comprehensive	Scotland (rural)	1988	Census	Air photo
Phase 1 Survey	Comprehensive	GB	Sporadic (by county)	Census	Field survey
Phase 2 Survey	Vegetation	UK	Continuous (non-systematic)	Sample (census of surveyed areas)	Field survey
Ancient Woodland Inventory	Ancient woodland	GB	Continuous	Census (2ha threshold)	Field survey/air photo/historical
Lowland Heathland Inventory	Lowland heathland	England	1996	Census	Existing sources
Saltmarsh Survey of Great Britain	Saltmarsh	GB	1989	Census	Field survey
Woodland Inventory (Census of Woodland and Trees)	Woodland	GB	Periodic (Current)	Census	Air photo
Land Use Change Statistics	Comprehensive	England	Continuous	Census (time lags)	Field/air photo
OS Land-Line.93	Comprehensive	GB	Continuous	Census (time lags)	Field survey/air photo
Monitoring Landscape Change Survey	Comprehensive	England & Wales	1951, 1971, 1981	Sample	Air photo
National Countryside Monitoring Scheme	Comprehensive	Scotland	1947, 1973, 1989	Sample	Air photo
Agricultural & Horticultural Census	Agriculture	UK	Annual	Census	Postal questionnaire
Monitoring Landscape Change in National Parks	Comprehensive	National Parks in England and Wales	1970, 1980	Census	Air photo
Environmentally Sensitive Areas Monitoring	Comprehensive	ESAs in GB	1986-88, 1989-91	Census	Air photo

Countryside Survey 1990

The purpose of this scheme is to provide "independent and scientifically reliable regional and national estimates of land cover, land use, linear features, vegetation, freshwater faunas and soils in a wide range of habitats in Great Britain. To determine change by comparison with surveys in 1978 and 1984. To add to the baseline data against which future habitat changes can be measured" (W.S. Atkins 1996). All of Great Britain was sampled excluding urban areas. The Countryside Survey (CS) was started in 1978 and repeated in 1984 and 1990 with larger sample areas on each occasion. Since the review was conducted, CS2000 has been carried out. This was based on survey data from 1998. At the time of writing, the full results have not been published (although a summary report is available – Haines-Young *et al.* 2000), but CS2000 will provide further time-series data. CS2000 sought to address many of the problems users reported with CS90 but it is too early to tell whether this will meet the needs of the user community, for example by reporting data in Broad Habitat categories.

Field survey sampling and the extrapolation of results and reporting of CS data relies upon the ITE Land Classification System. The ITE Land Classification System (e.g. Bunce *et al.* 1996a, b, c) had its origins in work conducted in Cumbria and the Lake District (Bunce *et al.* 1975). In these studies, multivariate statistics were used to provide classifications of grid squares to develop the basis of a land classification for field sampling. The rationale behind the classification is to divide the land surface into exclusive environment units and to classify these according to measured environmental characteristics. Such a classification should then be able to provide a sampling framework for ecological and landscape research, one that is related to ecological and landscape characteristics but that represents a relatively static and objective framework. The classification sought to describe elements of the British landscape and partition their variation into relatively homogenous classes. By this rationale land squares of the same class should have similar ecological and landscape characteristics and between class variation should be greater than within class variation. In developing the classification, Bunce *et al.* (1996b, p.41) made four major assumptions:

i. There is a 'natural' structure to the environmental factors which underlie the GB countryside which is reflected in the inter-correlations between the variables. These features can be measured and analysed to formalise the relationships. However, the structure has fuzzy, not crisp, boundaries because of the continuous nature of the variation.

ii. The physical environmental parameters are correlated with the composition of elements of the countryside, such as land use and habitats, directly and indirectly.

iii. An appropriate classification technique applied to measured map parameters can define strata and identify the main environmental trends and structure of the GB environment.

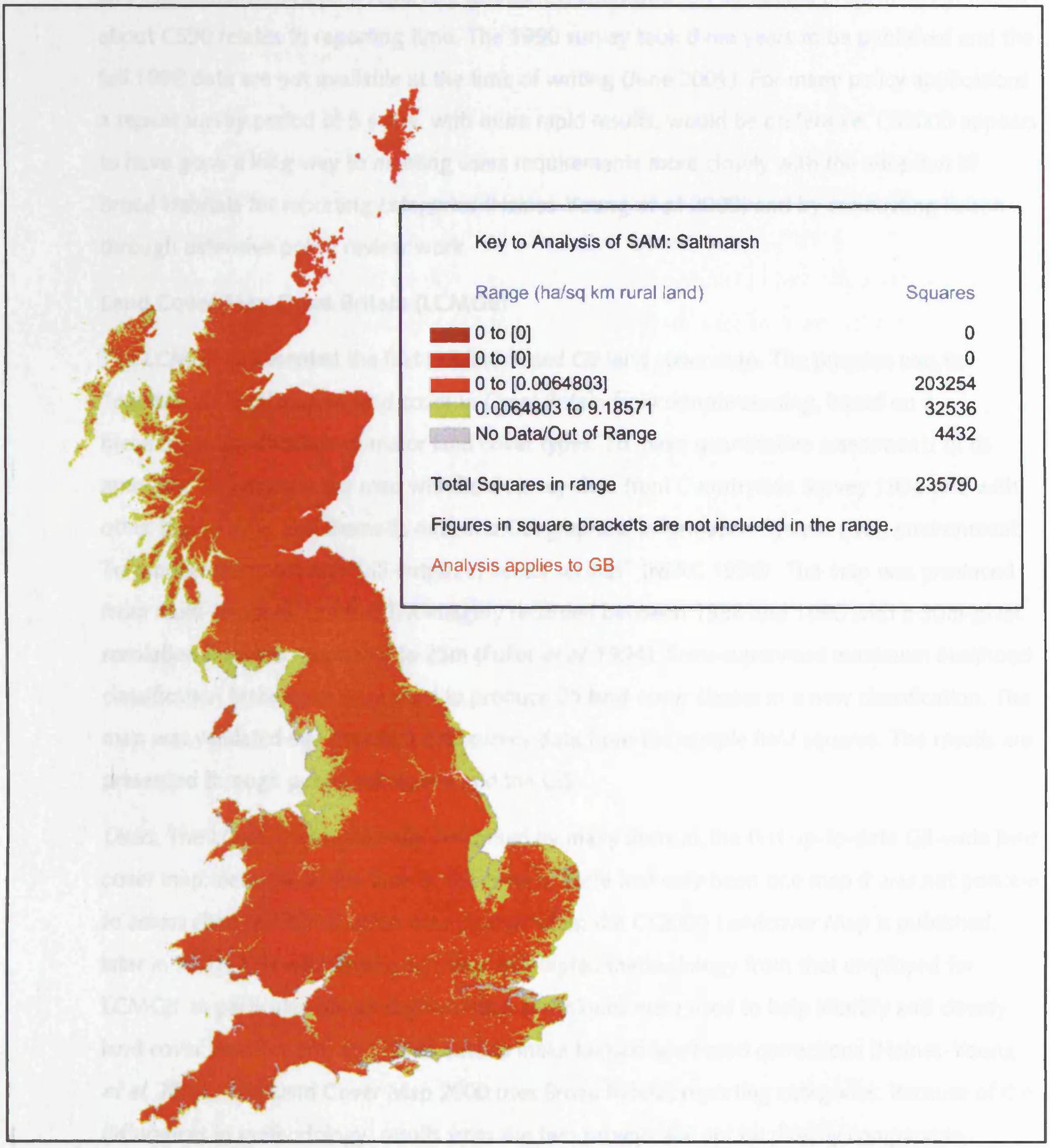
iv. The land features of interest may be predicted with statistical measures of accuracy depending upon their degree of correlation with the strata, and hence the strata may be used to improve estimates of the areal cover of the features.

An initial classification was developed in 1977 based upon 1212 1km squares located at the intersections of a 15km grid over GB. Forty environmental variables were recorded from four groups: climate, topography, human geography and geology and drift. The methodology used was complex and not completely transparent. Some variables were pre-selected using Principal Component Analysis 'to prevent distortion of the final classification' (Bunce *et al.* 1996b) although the thinking behind this is not elaborated. Because the data were to be analysed by use of Indicator Species Analysis (ISA) (Hill *et al.* 1975) and later TWINSPAN (Hill 1979) they were converted into a binary form. Because some variables were continuous (and therefore not binary) they were ranked and split evenly into four 'pseudo' attributes. ISA and TWINSPAN use essentially the same ordination and classification algorithms, and produce divisive polythetic classifications. They were primarily designed for analysing quadrat data (TWINSPAN was used in the development of the NVC) which tends to contain many zero values and was therefore considered suitable for analysing landscape variables. In total 281 attributes were analysed using ISA and the classification was run to five levels of binary division, producing 32 land classes. The resulting ISA classification was based upon 76 variables which could be used to 'key out' squares for assignment to a land class. In later years, it became evident that the classification should be extended to all GB. Bunce *et al.* (1996b) calculated that it would take 20 person years to key out the squares based upon the original 76 variables. Hence, a simplified set of variables was developed by transforming the original data using logistic discrimination and linear discriminant functions (Bunce *et al.* 1996b). Tests against the original 1212 1 km squares showed that the new method allocated 62% of the squares to the correct land class and the others to mainly similar land classes. The revised classification was used to allocate land classes to each of the 240,000 1km squares in GB using machine readable variables.

The land class system has been used as the basis for stratifying samples for the Countryside Survey (CS) series of surveys and is used as the basis for interpolating results from the CS through the Countryside Information System (CIS) software.

Users. Responses provided by participants suggested that the CS was widely used by policy makers at the national level. It was commissioned by DOE (now DEFRA) to assess stock and rates of change of land cover and vegetation in the countryside. Information is made available through published reports and the CIS at a standardised resolution of 1x1km^2. It is mainly used to report headline figures for land cover and land cover change at national and regional scales. It is generally considered not to be suitable for use for areas below around 3000km^2. Its basis upon the ITE 32 class land classification system is seen as both its major strength and its major weakness by many users. The use of the system has meant that the sample is randomly stratified; however, it also means that it is difficult to link to other data and reported figures represent an interpolated mean likelihood rather than a mapped distribution. The subtlety of these distinctions appears to escape many users. To display field sample data for the entire country, the system interpolates the sample data by using the 32 ITE land classes. This produces reliable national estimates for different habitat or land use categories, but can produce misleading results for resource distribution. Figure 2.2 shows the CS estimate for Saltmarsh based upon the sample for saltmarsh. The figure shows a predominantly coastal distribution, as would be expected, but also shows locational artefacts: saltmarsh can be seen in inland areas. This situation arises where land classes with a predominantly, but not exclusively, coastal distribution receive a mean saltmarsh area estimate based upon a field sample from coastal square representatives of the land class.

Figure 2.2 CS1990a.csf estimate for saltmarsh from data mapped by the CIS.



A number of policy makers indicated that they would find the data more useful if they incorporated urban areas using a compatible classification. Although the CS offers perhaps the most comprehensive set of time-lapse vegetation data sampled across GB, the data are rarely used by the statutory conservation agencies. Their arguments for this limited use include too high levels of error, at the scale required, and the fact that the survey uses non-standard reporting categories. The statutory conservation agencies now use the National Vegetation Classification (NVC) as a *de facto* standard for vegetation survey and reporting. Their reporting requirements for the Habitats Directive are largely fulfilled through the use of NVC-based data. In published form, SNCO users consider that the CS offers insufficient detail on habitats for monitoring purposes and its comparability is severely limited by the underlying sampling system and the nomenclature (and reporting categories) employed. An additional problem often raised about CS90 relates to reporting time. The 1990 survey took three years to be published and the full 1998 data are not available at the time of writing (June 2001). For many policy applications a repeat survey period of 5 years, with more rapid results, would be preferable. CS2000 appears to have gone a long way to meeting users requirements more closely with the adoption of Broad Habitats for reporting categories (Haines-Young *et al.* 2000) and by conducting liaison through extensive policy review work.

Land Cover Map Great Britain (LCMGB)

The LCMGB represented the first satellite-based GB land cover map. The purpose was to "compile a digital map of land cover in Great Britain from remote sensing, based on a hierarchical classification of major land cover types. To make quantitative assessments of its accuracy. To integrate the map with field survey data from Countryside Survey 1990 and with other topographic and thematic data in a Geographical Information System (GIS) environment. To produce demonstrator GIS output in vector format" (NERC 1996). The map was produced from multi-temporal Landsat TM imagery recorded between 1988 and 1990 with a 30m-pixel resolution that was resampled to 25m (Fuller *et al.* 1994). Semi-supervised maximum likelihood classification techniques were used to produce 25 land cover classes in a new classification. The map was validated against CS90 field survey data from the sample field squares. The results are presented through published reports and the CIS.

Users. The LCMGB was generally welcomed by many users as the first up-to-date GB-wide land cover map. Because, at the time of this review, there had only been one map it was not possible to assess change. This situation may change when the CS2000 Landcover Map is published, later in 2001. This was developed using an adapted methodology from that employed for LCMGB. In particular, image segmentation techniques were used to help identify and classify land cover parcels using additional data to make knowledge-based corrections (Haines-Young *et al.* 2000). The Land Cover Map 2000 uses Broad habitat reporting categories. Because of the differences in methodology, results from the two surveys will not be directly comparable, making estimates of land cover change difficult. As this indicates, one of the problems in using land cover data derived from remote sensing has been the lack of a standard methodology in what has been a young and fast-developing technology.

Because the LCMGB represents a census and is measured, not interpolated, it can be used for resource estimation and mapping. According to user's responses, the data appear to have some geographical inconsistencies. Some users are unhappy with the accuracy and content of the classification, which is not directly related to other classifications in common use. Additionally, there is some noise in the data introduced through the use of multiple satellite scenes captured at different times of the year. Land cover classes can be compared with the help of LUCID, software for the comparison of land cover definitions (Wyatt *et al.* 1993).

Land Cover of Scotland (LCS88)

The purpose of LCS88 was to "provide baseline information on land cover for the whole of Scotland to enable both prospective and retrospective studies of land cover/environmental change, and to provide a basic land cover inventory" (CIS EC 1996). The survey was based upon 1:24,000 aerial photography flown in 1988 and validated with ground survey in 702 1km x 1km squares using stratified random sampling. The scheme uses a hierarchical land cover classification based upon 6 principal features and 129 classes, including two for mosaics. The data were digitised and analysed on a SPANS GIS system. The scheme has a variable spatial resolution, depending upon land cover type, of between 2 and 10ha.

Users. LCS88 is used for research and policy applications and to provide forest inventory data. The LCS88 methodology has been proposed as a basis for a National Land use Stock System (NLUSS) in England and Wales.

Nature Conservancy Council Phase 1 Survey

Although Phase 1 represents a thorough vegetation/land cover survey methodology and classification system, the data generated from it are more piecemeal. The purpose of Phase 1 is to provide a record of semi-natural vegetation and wildlife habitat over large areas of countryside. The methodology is based upon field-by-field ground survey. The major outputs are coloured thematic maps and a survey report containing summary and statistical information. Phase 1 has many advantages for conservation management objectives because it maps real distribution on the ground and can distinguish between features and characteristics that are not discernible through remote sensing techniques. However, Phase 1 data suffers from a number of problems. Surveys were conducted on a county basis by different organisations. Many surveys were conducted in the late 1980s and few more recently. Surveys were carried out by a range of different organisations, often County Wildlife Trusts, with different levels of expertise. These factors mean that, across the country, Phase 1 data are of greatly differing ages and quality and many counties now have very old data (Wyatt 1991). An exception to the general trend is Wales where the Countryside Council for Wales has recently completed a Phase 1 survey for the entire country.

Users. Phase 1 surveys are widely used, but applications vary in specific counties according to the age and quality of survey data in that area. The maps are used by statutory and voluntary conservation bodies to target surveys and to identify protected sites. The maps are also used in planning casework and for environmental impact assessments (EIAs).

Phase 2 and NVC survey

Phase 1 surveys were intended to map and identify potential sites of interest. Phase 2 surveys are more detailed follow-up surveys that produce floristic and environmental data for sites. Since the early 1980s Phase 2 survey has employed NVC methodology. Over the last twenty years vast quantities of NVC data have been collected. NVC survey across the country is not part of any co-ordinated exercise, but is driven by specific user needs. To-date these data have not been centrally listed or catalogued and it remains difficult to estimate both the size and coverage of NVC data holdings in the UK (Tantram *et al.* 1997). Attempts have been made to combine and catalogue NVC-based data sets, for example for woodland (Hall 1996) but the overall resource remains fragmented and uncoordinated.

Users. The NVC was developed by the University of Lancaster and has become the *de facto* standard for ecological surveying in the national conservation agencies and most charities and NGOs. Its uptake by the research community has been fragmentary. Because it is based upon detailed ground survey, the NVC is an expensive method for covering large land areas. However, for nature conservation purposes detailed floristic data are needed in many cases to assess conservation status. In most cases, it is not possible to obtain the detail of information required through traditional land cover-based techniques and land cover-based data sources are of little use for conservation purposes because they fail to reflect the nuances of semi-natural vegetation that are important for assessing the type, state and quality of vegetation communities.

Woodland Inventory

The purpose of the Woodland Inventory (previously the Census of Woodland and Trees) is to provide statistics on the stock of woodland and trees for policy on forestry and timber management and production. The earlier Census of Woodlands and Trees was carried out in 1947, 1965 and 1980 with slight differences in methodology, the minimum woodland area recognised in each case being 2ha, 0.4ha and 0.25ha respectively. The latest full survey for which results are available is the 1980 census (Locke 1987). Woodland areas were taken from OS 1:50,000 maps and placed in size categories which were used for stratified random sampling within counties. Samples for woodland area were calculated from air photo interpretation and used to calculate the relationship between OS map sheet areas and measured areas to result in re-calculated woodland areas. Additionally, sub-samples were selected for ground survey. The methodologies used in Scotland vary in their detail. The Woodland Inventory superseded the Census in 1994 and is a rolling programme expected to run until 2001. It is based upon 1:25,000 aerial photography in England and Wales and the Landcover Map of Scotland (LCS88) in Scotland. All woodlands over 2ha in size are included (Wright 1998). Woodlands are divided into three size categories (2-100ha, 100-500ha and >500ha) and a 1% sample of each category is selected using a cluster sample process. Woodlands below

2ha in area and groups of trees are treated differently and are recorded using a mixture of aerial survey and field survey, again employing a 1% sample for GB.

Users. The Woodland Census figures are used by the Forestry Commission and its daughter bodies for policy purposes and are also used by government departments and other countryside and conservation agencies. One respondent in this review used these data: the only limitation mentioned was that data are not available for small areas. It is notable that, unlike the CS and LCMGB data, these data are not available through the CIS.

Ancient Woodland Inventory

The AWI was conducted to provide an improved estimate of the extent and location of Ancient Woodland in GB (Spencer and Kirby 1992). It was also intended to help analyse the rates of loss of the resource and to assist in the selection of Sites of Special Scientific Interest (SSSIs). The AWI categorises different types of woodland, over 2ha in area, present on continuously wooded sites since c.1600 and therefore relates to historical land use. The AWI is updated on a continuous although sporadic basis – as and when new information becomes available. Because it consists of data of different ages, and is updated only sporadically, it is not possible to use AWI data to assess rates of change. The inventory does include an assessment of loss of Ancient Woodland since 1920.

Users. The AWI is used by the SNCOs to review the SSSI site series and plan woodland conservation management. It is also used by other conservation-related bodies as a statistical source.

English Nature/RSPB Lowland Heathland Inventory

The Lowland Heathland Inventory was jointly conducted by English Nature and the RSPB. Like most other habitat inventories it comprised a compilation of existing data, which in many cases were up to ten years old. Where possible, site boundaries were mapped onto 1:25,000 OS map bases. However, the boundaries recorded were mainly SSSI boundaries and not necessarily heathland vegetation boundaries. Data were gathered from English Nature's site files, RSPB local offices, heathland projects and county wildlife trusts, and hence the quality and coverage of information was variable. The original project was conducted in 1993 with subsequent revisions in 1994/5 to include all English counties.

Users: The Inventory has been mainly used by SNCO and RSPB staff for policy and planning casework. ITE (now CEH) used the Inventory in their Key Habitats project (Barr 1997) but found a poor correspondence between the inventory and the heathland 'mask' developed by ITE.

Saltmarsh Survey of Great Britain

The Saltmarsh Survey of Great Britain (Burd 1989) was carried out by the former Nature Conservancy Council to review the status of saltmarsh habitats in Great Britain. Saltmarshes were selected for the first major coastal habitat inventory by the Chief Scientist's Directorate because of their rarity and the lack of knowledge about the overall resource. The survey comprised three main stages: the assessment of existing surveys, field survey to cover underrepresented areas and analysis of the data. The time and resources available for the survey precluded the use of NVC methodology and a simplified approach was adopted, grouping NVC communities into larger groups easily identifiable in the field without recourse to quadrat data. The survey was conducted between 1981 and 1989 and resulted in sketch maps (no attempt was made to use an OS or other scaled base), record cards and species records.

Users: Use appears to be restricted to SNCO coastal staff. Only a few copies of the original field data appear to have been made and the location of these appears unclear.

Land Use Change Statistics (LUCS)

These statistics are based upon the Ordnance Survey's map revision process (e.g. DOE 1993). Their purpose is to monitor land use change especially to developed uses. Recording is part of the OS's map revision programme of field survey and has been continuous since 1985. Between 1985 and 1992 the scheme covered GB; from 1993 onwards it has been confined to England. The LUCS uses its own land use classification system of 24 categories that are derived from over 350 features that can be described by surveyors. It is important to note that the statistics recognise change only at the time of revision and not at the actual time of change. The comparability of change statistics also varies with geographic area as rural areas are re-surveyed less often than more urban areas. Change is only measured when sufficient change has occurred to make it an economic exercise; thus updates do not relate to true changes in time for all categories in the whole survey.

Users. The LUCS are commissioned and mainly used by the DOE (now DEFRA) for policy analysis and formulation and to provide information to other departments. LUCS data are compared to other sources and are used for monitoring land use change to urban uses and monitoring house building targets. The major limitation that users quoted for the data is introduced by the time-lags in updating, thus absolute change between specific dates cannot be measured and changes only reflect the time of survey, not the time of the change.

Ordnance Survey Land-Line.93

Land-Line.93 is perhaps the major digital topographic map product of the Ordnance Survey, the UK's national mapping agency. Land-Line is a digital product available at a charge from the OS. It contains some implicit land use and land cover information, although it is intended to be a topographic product rather than a land use or land cover information source. Layers include buildings, roads, rivers, and some height information. Land-Line is derived from three scales of mapping: 1:1250 (urban), 1:2500 (rural) and 1:10,000 (mountain and moorland areas). Land-Line is used extensively for context and for locating features and areas. Land-Line93.Plus adds 27 additional layers including vegetation characteristics and limits. Currently, OS are developing a National Topographic Database (NTD) to support the Digital National Framework (DNF) (OS

2001) and are beginning to use an object-oriented methodology for referencing and storing map data. These changes are mainly relevant to the use of the data in GIS. Data will be structured with a referencing layer and a topographic definition. This object-oriented approach will allow the creation of associations and relationships and will allow for the propagation of changes throughout relationship hierarchies.

Monitoring Landscape Change Survey (MLC)

The purpose of this study was to survey rural landscape features of major policy importance in England and Wales. The study used mainly monochrome air photos within sample areas stratified within 188 strata based upon major soil groups. Three target dates of 1951, 1971 and 1981 were aimed for, although there was some variability around these dates. Feature boundaries within sample sites were digitised. A hierarchical classification was used comprising 7 major groups, 31 categories and 23 sub-categories. The survey also tested the used of Landsat TM satellite data. The survey provided a broad framework for assessing land cover change. The results are subject to a significant margin of error as they are based upon a small (2.4%) sample size. Additionally, any survey of this kind is subject to interpretative and measurement errors.

Users. The MLC data were used by the Countryside Commission and the DOE (now DEFRA) for policy purposes and producing statistics. They appear to be little used currently due to their age and some respondents questioned the reliability of the results.

National Countryside Monitoring Scheme (NCMS)

This scheme was developed by the Nature Conservancy Council (now the SNCOs) and was originally conceived as a GB-wide initiative. A few counties in England and Wales were surveyed and the scheme was taken forward by Scottish Natural Heritage (SNH) in Scotland. The purpose was to provide data on the extent and distribution of a number of defined landscape components. The survey had three target dates of 1947, 1973 and 1989 and was based upon air photo interpretation and some field validation. The results for 1947-73 were published in 1994.

Users. The NCMS data appear to have been used mainly within SNH for policy development.

Agricultural and Horticultural Census

The Annual Agricultural and Horticultural Census originated in 1866 and is now conducted by MAFF (now DEFRA) every June. Data are gathered on livestock numbers and crop area from registered agricultural holdings. The data are considered confidential and are aggregated to the parish level in England and Wales and the agricultural parish level in Scotland. To preserve confidentiality, parishes are sometimes combined for reporting purposes. That part of the countryside not registered as an agricultural holding is not included; this category will also include small part-time farming enterprises. The Census has undergone a multitude of incremental methodological changes over the years necessitating great care in the comparison of data from different periods.

Users. The Census has been extensively used within MAFF for policy purposes and is also heavily used by other government departments and organisations concerned with farming (e.g. FWAG, NFU and CLA). Agencies such as the Countryside Agency and the Farming and Rural Conservation Agency (FRCA) use the data to monitor changes in agricultural activity and analyse agri-environment schemes. The data are broadly compatible with other sources though there are some reported problems with grassland categories. For many potential users, the main limitation to these data has been the problems of access due to confidentiality restrictions.

Monitoring Landscape Change in National Parks (MLCNP)

The purpose of this study was to monitor the past and current distribution and extent of major land cover categories within National Parks. In the UK, National Parks *per se* are currently restricted to England and Wales. Two target dates of 1970 and 1980 were chosen for air photo interpretation of 1:10,000 imagery. Within each National Park a 2% random sample was taken for ground verification. The resulting information was digitised and then converted to SPANS Quadtree format at a 20m resolution.

Users. The data appear to have been used exclusively by the National Park Authorities. The level and type of use varies across the parks. Section 43 of the Wildlife and Countryside Act 1981 and amendments in Section 3 of the Widlife and Countryside (Amendment) Act 1985 required National Park Authorities to map moorland, heathland, mountain, woodland and coastal features of conservation importance. The National Parks were required to update these every five years, but many failed to do so. However, the MLCNP data, used in conjunction with GIS, enabled some parks to help automate the process. The Lake District National Park has used MLCNP data to aid countryside management and policy making, in conjunction with extensive use of GIS (Fishwick and Clayson 1995).

Environmentally Sensitive Areas (ESA) Monitoring

This monitoring programme was set in place to help ensure that ESA schemes were being effective. They were intended to establish a baseline record of the extent, distribution and condition of biological, landscape and heritage elements and to monitor changes over time. The monitoring activities vary specifically according to the environmental objectives and characteristics of each ESA. The methodology also varies by ESA but most monitoring is based upon aerial photo interpretation and field survey.

Users. The data are primarily used by FRCA (now RDS) for policy support and monitoring and do not appear to be made routinely valuable to outside agencies.

2.3.3 Problems and limitations with data sources

The users included in this review encompassed a range of different agencies information applications. There was, however, a high degree of consensus amongst these users about the quality and utility of the available data sources. The problems they encountered can be grouped into four main categories: nomenclature, spatial scale, temporal frequency and data availability/accessibility.

Most users found that the use of different classification systems in different surveys made it difficult to compare results. The effect that different nomenclatures have on survey outputs and applications is complex and interrelated. The possible nature of these relationships has been discussed above, but in most cases, unless one-to-one relationships are identifiable, comparability between different nomenclatures is poor.

The spatial scale of features that can be observed and the level at which results are reported can have a great effect upon the value and application of data sources. In many cases, users require data at a greater resolution for smaller area applications. Many of the users who were interested in CS90, for example, were local authorities who wanted state of environment data and a means for placing their county or district in a regional and national context. At present, however, these data are not considered to be reliable enough to be used at these scales in the form they are made available in the CIS.

Temporal frequency was also seen by many users as a major restraint on data applicability. Many data sources are based upon information that is either very old or includes data of variable age. The LUCS data include time-lags of up to 40 years and the 'continuous' recording of data in sources such as the Ancient Woodland Inventory means that change statistics cannot be compiled. Again CS90 was singled out for criticism due to the long periods between data capture. Many other data sets are not maintained at all, so that they provide no more than a single 'snapshot' of habitat conditions. Because of differences in survey methodology and classification system, change cannot be inferred by comparing results from different surveys; even where surveys have been repeated, changes in methodology often make comparison difficult (e.g. the agricultural census data and LCM1990-LCM2000 data).

Availability and accessibility problems proved to be one of the most common reasons for the prevention of data use. Data sources were not used by respondents for a number of reasons:

- *Lack of knowledge* – potentially interested users are often unaware of available data, either because of lack of suitable metadata or poor communications between potential users and data suppliers.
- *Confidentiality* – many data are subject to restrictions due to commercial, military or economic interests (e.g. MAFF agricultural census data).

- *Cost* – in many cases users find data are too expensive. For example, many users in the UK have restricted their use of OS digital data as it has been considered prohibitively expensive. This is especially a problem where users need the data for exploratory reasons (e.g. to evaluate the data for a particular purpose) or where the potential value of the data is uncertain (e.g. where the apparent problem to be addressed may not be real). Improved metadata would, in many cases, help to resolve these problems, as would opportunities for obtaining data on a use-or-return or trial basis. Cost issues have, in many cases, been made more severe by cost-recovery strategies of public and pseudo-public data suppliers (e.g. the Ordnance Survey). However, cost deterrents relate not only to data purchase, but also to any in-house data processing that may be necessary.

- *Data quality* –issues of data quality relate to many different aspects of the data, including both their locational accuracy and the detail and content of their attribute data. Problems derive from both the survey methodologies used (including sample design and methods of data collection), and the classification systems applied. These problems translate into doubts about the accuracy of the data and/or difficulties in interpreting the data within the context of the specific application. These problems are especially acute for statutory organisations, for which validity and verifiability of any results or interpretations are paramount. Problems of data quality are often compounded by lack or transparency about the way the data have been gathered, analysed and reported – in part due to inadequate metadata. As this implies, issues of data quality may, in some cases, be perceptual rather than real, but for users even the possibility of error, or uncertainties about data quality, may act as a significant deterrent to their use. This suggests the need for more rigorous testing and validation of data, and reporting of these results. Even then, issues of data quality are likely to arise where data are being used for a purpose – and in a situation – for which they were not originally intended. This implies the need for users to check and validate the data in the context in which they are using them – something many user organisations cannot easily do. It also requires the availability of feedback mechanisms from users to data suppliers, so that experience can be shared, and organisational learning about the characteristics of the data can take place.

- *Data quantity* – problems occur both due to too much, and too little data. Data overload occurs where the available data are not adequately synthesised and aggregated to meet the needs of users; in these cases, lack of technical capability, or costs of data processing, may be major deterrents to use. Lack of data typically takes one of three forms: incomplete geographical coverage; lack of data for the time period of interest (see below); or inadequate attribute data (i.e. the available data do not provide information on the specific features of interest). Data gaps also occur for several different reasons, including lack of initial survey (e.g. due to cost, lack of awareness about data demands), survey failure (e.g. cloud cover limiting availability of satellite/aerial survey data), analytical problems (e.g. lack of data processing) and reporting failure (lack of publication or dissemination of survey results).

- *Timeliness* – most data sets are no more than snapshots in time; others are temporal averages, compiled over long periods; some (e.g. AWI) are amalgamations of data collected at different times. In each case, the timeliness of the data may pose problems for users, for a number of reasons: because the data are out-of-date; because they do not relate to the period of interest; or because their timeliness is ambiguous.
- *Technical*–many data are not available in a digital form, or are available in formats which users find difficult to translate. Data are often held in legacy systems, proprietary formats or those that cannot be easily queried (e.g. Recorder 3.x). A lack of adequate metadata again compound these problems.

2.3.4 Data management

Knowledge of, accessibility to and ease of use of data are intimately linked to data management practices and the use of technology. Despite the potentially vast quantities of data available or needed for environmental applications, data management practice has often been fragmentary and lacking in standards (Burnett *et al.* 1995). The lack of a consistent approach has often led to the poor implementation of technology, in turn resulting in inferior accessibility to the available data and poor quality of derived outputs. Knowledge about data availability can be provided by metadata and the improved use of information technology can help address data management problems. In particular, the development of Geographic Information Systems offers powerful means for the management of habitat-related data.

Metadata

Many of the problems described above can be addressed by the development of data catalogues and metadata. In response to the CCBR report (Burnett *et al.* 1995) the National Biodiversity Network (NBN) developed a pilot Internet-based catalogue of biodiversity information sources – the NBN Index (NBN 2001). Metadata such as these can direct users to the most appropriate sources and can also provide data lineage information that can help with the appropriate use of data. However, the data users interviewed here tended to be aware of available data sources and in some cases were using the best available sources, even if these were not necessarily completely appropriate for their intended applications. The problems appeared to revolve around access to suitable data rather than identification of suitable data. However, the lack of standardised metadata allows some confusion about the technical details of data sources to continue and in some cases, such as the Countryside Character Initiative, has led to the apparent loss of raw data.

Geographical Information Systems

Geographical Information Systems (GIS) have their origins in environmental applications (Burrough 1986) and have developed from being primarily a research tool to widespread use for policy analysis and data management (e.g. Aspinall 1995, Briggs and Tantram 1997, Fishwick and Clayson 1995, Harrison *et al.* 1991). The power of GIS lies in the capability to query data by virtue of their spatial location and relationships. Because habitat-related data virtually always include spatial components GIS can offer a range of useful and powerful tools for their creation, analysis, management and reporting. Aspinall (1995) suggested that a geographical approach has three main advantages:

i. geographic data provide a powerful reference base for data collection and storage;

ii. geography provides the context for analysis and the examination of relationships;

iii. maps provide a effective medium for communicating information.

GIS enables a geographical approach by providing database management systems that can store, manage, query and visualise both attribute and graphical data. Thus, analyses and statistics can be produced from data based upon both numerical and spatial characteristics and outputs can be produced in mapped, tabular and graphical forms. Over the last ten years GIS has moved further into mainstream computing. With the advent of powerful personal desktop computers and the targeting by GIS application vendors of more mainstream business markets, 'desktop GIS' has become a reality. The lower cost and relative ease of use offered by desktop GIS has led to higher levels of uptake by countryside-related policy and management organisations. English Nature has begun to use desktop GIS in local team offices, but uptake by the other SNCOs has been slow. Many Wildlife Trusts and local record centres also now use desktop GIS and their use is now ubiquitous in National Parks (Briggs and Tantram 1997). More complex and powerful applications, such as ESRI's[4] ArcInfo product are widely used for their power and versatility and are still used in research applications.

Fully functional GIS products allow for the integration of different kinds of data. Data integration is one of the most powerful features of GIS, and in most cases, data are combined in GIS using common spatial frameworks (co-ordinate systems and projections). All GIS include two-dimensional capabilities and most allow the use of both vector and raster-based data. Vector-based data relate to points and linework (lines and/or polygons) and are used to represent features or regions. Raster data, such as digitised aerial photographs and satellite images, are inherently less flexible in that they are essentially static 'pictures', albeit spatially referenced, but become powerful tools when overlaid with other coverages, such a designated

[4] (Environmental Systems Research Institute Inc.)

areas or vegetation boundaries. More sophisticated GIS offer three-dimensional modelling capabilities through the creation of grids. Grids represent two-dimensional arrays of data points with one or more additional Z values (in the third dimension). By use of interpolation and modelling techniques, GIS software creates virtual 'data surfaces' from the Z values which can be queried and analysed by integration with other two-dimensional vector data. Such three-dimensional modelling is used for a variety of applications such as intervisibility analysis for landscape planning, the quantification and analysis of mineral deposits and catchment modelling.

GIS provide powerful means for the integration of data and assessing relationships between them. For example, Bird *et al.* (1994) and Peccol *et al.* (1996) combined data on landscape designations, planning policies and land cover to assess land cover change and the relationship with countryside designations. In this way, GIS can be used to combine data within a common geographic framework and conduct analyses on the basis of spatial attributes. Such uses of GIS make them especially effective tools for decision-makers (Harrison *et al.* 1991) and GIS are often incorporated into decision support systems. GIS are often used to combine vector and raster data sources, such as overlaying topographic map features with aerial photography or satellite imagery. A further development is to use the data together intelligently to help classify or re-classify remotely sensed data. Long and Skewes (1996) used Landsat TM imagery, spatial topographic data (elevation, distance from water) and ecologically-based principles (vegetation stand perimeter/area ratios) to derive simple topological rules for the identification and mapping of mangroves. By these means, remotely sensed data can be differentiated further than may be possible using image properties alone. This technique, known as knowledge-based re-classification, was successfully employed by Adinarayana *et al.* (1994) to map land use patterns in India. Two seasonal sets of satellite imagery were combined in a GIS and classified. However, individually each classification was insufficient to discriminate between all necessary classes because of similar spectral signatures. Therefore, logical combinations were employed to combine the two satellite classifications and to combine these with other data. For example, Adinarayana *et al.* used combinations such as:

> If SpectralClass=Vegetation AND DrainageDensity <10% AND Height >600m THEN Class=Forest.

Such rules were based upon knowledge of local land use. If similar data do not already exist, threshold levels (i.e. DrainageDensity <10%) have either to be inferred (from knowledge of the target resource), or based upon sub-samples of the total data set (Lees and Ritman 1991).

The logical extension of these techniques is to combine various data to build species or habitat occurrence models. Austin *et al.* (1996) used GIS to produce a predictive model for the distribution of buzzard (*Buteo buteo*). The relationship between various environmental features and buzzard distribution was used with logistic regression and discriminant function analyses to predict with a high level of success, the possible distribution of buzzard nesting areas. Such methods can provide information to inform policy or simply to target field survey effort more effectively. Lavers and Haines-Young (1996) turned this concept around and used species abundance models to predict the possible impact of conservation policies in the Flow Country. They identified some sites reserved for potential peat extraction as being of equal or greater importance to dunlin populations than some sites already designated as SSSIs. Work such as this confirms firstly the analytical power of GIS for scientific and policy purposes, and secondly the lack of a rigorous and spatially consistent scientific approach to site selection.

Countryside Information System (CIS)

The success of GIS has raised the profile of geographic information and demonstrated the uses to which it can be put. Geographical functionality is now included in many purpose developed systems (including experts systems). One such system, which potentially provides an important source of habitat-related data, is the Countryside Information System (CIS); this is not a true GIS, but rather a spatially enabled database. The CIS was developed to provide policy users with easy access to strategic information about the countryside (NERC 1996, Howard *et al.* 1994, Howard and Bunce 1996, Ireland and Maycock 1996). CIS uses the ITE Land Classification system to display and report Countryside Survey data, simple analyses can also be conducted. The CIS also provides the means for the display, combination and analysis of other data made available in the appropriate format. The CIS handles data on a 1 km^2 basis, with a single attribute representing the value of the particular data type expressed in hectares. The potential value of the CIS is that it provides a common basis for publishing, accessing and using different countryside data. The Environmental Catalogue, which is packaged with the CIS, provides metadata on the data sources available in CIS format.

2.3.5 Summary

- There are a wide range of data sources for habitat and habitat-related data, created by different organisations, using a range of methods and for a variety of purposes. This situation reflects both the subtly contrasting requirements of different organisations and also a possible lack of integration or co-operation.
- Many data sources have a tendency to be short-lived; relatively few are maintained over any length of time and many have changes in methodology between repeat surveys.
- There is a lack of comparability in survey methods, classification and reporting frameworks.
- A consistent approach to data source provision has been lacking with independent sectoral development in many areas.
- Although there are many different data sources relatively few appear to be actively used or are made easily available.

- Data users have problems with available data sources; these can be grouped into those related to nomenclature, spatial scale, temporal frequency and data availability/accessibility.
- Data management also appears to be uncoordinated and lacking in standards.
- There is a lack of metadata, more widespread use of metadata would help improve users' knowledge of data sources, accessibility and some data quality issues.
- GIS can provide powerful tools for the integration, management, analysis and reporting of data. Most habitat-related data have geographical components and GIS can provide useful approaches to many data-related issues.

2.4 CONCLUSIONS

The development of policy in the countryside is conducted at many levels and by many different organisations and the resulting information 'environment' is complex. Information is needed to inform policy development and to meet an ever-expanding range of reporting requirements. The brief scoping review conducted for this chapter identified a range of problems in the use of habitat-related data sources. These also appeared to vary by application. The linkage between the creation and supply of data and the demand for data by users appears to be weak and users encounter a number of problems with available data sources. Some of these relate to data accessibility issues, but others relate to data content and quality. These problems appear to be present despite the apparently large number of potential data sources and processing techniques. Indeed, the heterogeneous nature of such data sources may be creating some of the problems encountered. The problems appear to be both on the supply side – suitable data does not appear to be available, and on the demand side – there are problems with optimal data utilisation relating to knowledge and expertise. Further research was clearly required to gain a better insight into the organisational utilisation of information. It was therefore decided to look in detail at information use within specific conservation organisations, in order to investigate a number of information-related questions such as:

- What are the information demands of different conservation organisations and how are they identified?
- What data sources are used for what applications and how well do these meet requirements?
- What characteristics and behaviour mediate data utilisation; do organisations use the best available data sources for a particular application?
- How might existing data sources be used more effectively so that existing data supply may better service current demand?

Statutory conservation organisations are vitally important because they develop and maintain the protected sites networks which are one of the major constituents of 'natural' environment protection in Great Britain. The protected sites networks are intended to represent and protect the best and most important examples of habitats and related species. Thus, the first case study examines the major activities of the statutory conservation agencies. It was decided to consider all of the statutory agencies (in Great Britain) because they each had subtly different remits. The Joint Nature Conservation Committee was included because it provides scientific co-ordination and support to the agencies and fulfils a co-ordinating role. Although the protected sites network provides a crucial backbone to conservation effort, countryside management is also important for sustaining conservation across wider areas. Outside the relatively restricted areas of protected sites, most active countryside management is conducted by the National Parks and, on a smaller scale, by local authorities. Established over fifty years ago, National Parks contain some of the most important conservation and landscape resources in the country and in many ways can be considered as the flagships of countryside management in Great Britain. It was therefore decided to choose a National Park as the focus for a second case study. Each National Park is unique and has its own particular characteristics and issues and therefore no single park could be considered entirely representative of all the others. However, they also have much in common, they have the same top-level policy and management objectives and work within the same policy and funding frameworks. Discussions were held with the (then) Countryside Commission in order to develop a suitable approach. It was decided to concentrate effort on one park in depth and to look for a park beginning a review of its old National Park Plan – the first stage in the development of National Park Management Plans. The North York Moors National Park was one of the first National Parks to begin this process and the Countryside Commission felt that lessons learnt from their planning process would provide valuable insights into the needs of other parks. Because the North York Moors were beginning their review process their staff were also able to commit time to the assessment of their information requirements and data use for this study as would also help inform the new National Park Plan.

3 CASE STUDIES: INFORMATION USE

Two case studies on the use of habitat-related information are presented here. The first deals with data use for activities conducted at national (UK) and country levels by the statutory conservation agencies. This case study was part funded by the Joint Nature Conservation Committee. The second case study focuses on information use within the North York Moors National Park. This case study was part funded by the (then) Countryside Commission as part of the Parks Information Management System (PIMS) initiative.

Similar basic approaches were used in both case studies, although the details of each study vary to accommodate the different roles and activities of the organisations concerned (as further described below). The first step was the collation and review of available literature about the work activities and planning of the relevant organisations. Where they were available, documents relating to information management were also examined. Collectively, these documents helped to provide background information on the policy and management of the case study organisations. For each organisation, detail was required on the information requirements for different work activities, the information sources used and how well they met requirements. Because the majority of these questions relate to the working practices of staff and their expert knowledge, a questionnaire-based approach was followed. Because of the complex and inter-related nature of many of the questions, a solely questionnaire-based approach was unlikely to yield a high return rate or good quality responses. It was therefore decided to follow a combined approach of using workshops, where groups of similar staff were brought together to consider and discuss issues in a group setting, and make use of a questionnaire to provide specific and structured responses. In both cases, the relevant organisations were only able to commit a limited amount of staff time; therefore this needed to be used carefully and in a targeted manner to include a thorough review of activities. Meetings were held with senior management to select groups of staff from the relevant organisations. These were structured to include a cross-section of different subject areas and responsibilities and to include staff of different levels of seniority.

3.1 NATIONAL DATA SOURCES AND USE - SNCOs

The statutory nature conservation agencies in Great Britain comprise English Nature, The Countryside Council for Wales, Scottish Natural Heritage and the Joint Nature Conservation Committee (EN, CCW, SNH and JNCC). Their remits have been discussed earlier in Chapter 2. As the Government's advisors on nature conservation, they are amongst the major users of habitat-related information and it is essential that the information available to them meets their needs. It is only with suitable information that reporting requirements can be accurately met and that conservation policy and management measures are based upon best available knowledge. Historically, particularly before the break-up of the Nature Conservancy Council (NCC), the agencies have also been major data producers. The term 'national data sources' is potentially misleading. As national organisations, the SNCOs work within international, national and regional remits. Many of the data sources they rely upon comprise field surveys conducted at small scales, but provide information for reporting on issues of national or international importance.

The three statutory conservation agencies provide advice on nature conservation to the Government under the 1990 Environmental Protection Act. The agencies work together through the JNCC which provides a platform for the delivery of statutory responsibilities within Great Britain and at an international level. These special responsibilities are known as 'special functions' and are intended to contribute to sustaining and enhancing biological diversity, geological features and natural systems (JNCC 1995). The purpose of this case study was to assess the activities of the statutory conservation agencies and their use of information. The case study was conducted in conjunction with the JNCC and staff from the three country agencies with the intention of addressing four main objectives, namely to:

i. identify the 'key business activities' (KBAs) of the conservation agencies (national and international) that require information on habitats;
ii. review the current sources of information available to, and used by, the country agencies to fulfil their 'key business activities', through highlighting their strengths and weaknesses;
iii. identify known sources of habitat data of potential value to the 'key business activities' but which are currently inaccessible;
iv. identify what quality of habitat information is required to deliver GB-wide functions.

Part of the goal of this case study was to examine the similarities and differences in the nature, use and provision of habitat-related data between the three country agencies and between them and the JNCC. Information flows are notoriously difficult to analyse within organisations. In terms of the statutory conservation organisations the task is further complicated by the different geographic and management-related structures of the organisations. The three country agencies are charged with similar statutory 'general functions', while the Support Unit of the JNCC undertakes a co-ordinating, collaborative and integrative role.

3.1.1 Methods

The methods employed were developed in conjunction with the JNCC Support Unit and were based upon four main approaches:

i. Review and evaluation of policy documents, internal and published reports relating to the acquisition, storage and utilisation of habitat-related data in relation to 'key business activities'. These reports were primarily authored or commissioned by the statutory agencies themselves or by related bodies such as the Department of the Environment Transport and the Regions (DETR, previously the DOE and now DEFRA) or the Biological Records Centre (BRC).

ii. Workshops with key staff from both JNCC and the country agencies using structured questionnaires to gather information on 'key business activities', staff use, requirements and evaluation of habitat-related data. Additionally, commentary notes were taken to provide background and contextual information.

iii. Interviews with specific staff within the country agencies. These concentrated on IT staff and systems developers to gather background on current and planned information strategies within each organisation.

iv. Telephone-based interviews with agency staff and external organisations to provide specific details and wider context and to cross-check results from the workshop-based questionnaires.

3.1.2 Workshops and structured questionnaires

Three workshops were conducted, two at JNCC headquarters in Peterborough, attended by staff from JNCC, EN and SNH, and the third at Bangor with CCW headquarters staff.

Prior to the workshops, a questionnaire was produced to provide a framework for responses and discussion and to ensure that the same topics were covered in each workshop. The questionnaire was circulated to participants before the workshops and is included in Appendix 2. It was structured in six sections, each tackling separate but related topics and following a developing line of enquiry:

i. The nature of habitat information: what are habitat-related data - definitions.

ii. Identification of 'key business activities', both of the organisation and those conducted by responders.

iii. The use of and demand for habitat-related data according to 'key business activities'.

iv. Sources of habitat-related data and evaluation of their utility, strengths and weaknesses.

v. Potential sources of habitat-related data and reasons why they may not be used.

vi. The management of habitat-related data and the affect this has upon data utilisation.

In some cases, open-ended questions were used to try to elicit free responses to specific issues. As far as possible, the questionnaire was designed to avoid leading questions and to give respondents the best possible chance of providing accurate responses.

The specific interpretation of questions within the questionnaire was deliberately left open to allow participants to discuss the questions presented and the issues raised by them. This was done because the object of the exercise was to gather ideas and information from agency staff and not to lead them or present a biased agenda. This approach proved appropriate, since it immediately became apparent that staff held widely varied views on almost all of the subject matter contained within the questionnaire. To have tried to produce a tightly worded and defined questionnaire would have been difficult, given the breadth of the subject matter, and counter-productive, since many staff would have been unable to answer many of the questions.

Each of the workshops lasted up to four hours and began with a background introductory session, explaining the purpose of the project and the workshop, followed by a detailed group examination of the questionnaire with participants. This process enabled all issues to be discussed by the group and ensured that all participants understood the complex questionnaire. It also represented another valuable source of information on the use and attitudes to habitat-related data, and notes were taken and summarised. Once the questionnaire had been discussed, participants were asked to individually complete it as fully as possible, clarifying additional points as they arose. This stage took an average of two hours. Once the questionnaires had been completed, a final group debriefing session was conducted. These sessions helped with the development of concepts for both the research and for the participants.

A total of fifteen questionnaires were completed, six responses from JNCC staff and three responses each from EN, CCW and SNH. Most of the staff were from headquarters positions within their respective agencies, although two staff from English Nature were from Local Teams.

Given the nature of this study, with the small overall sample of staff completing the questionnaire, only a qualitative analysis of responses was possible. Nevertheless, a wide range of opinion and information was obtained from which it was possible to structure the results presented in the following sections.

3.1.3 The nature of habitat-related data

Definitions

In the first part of the questionnaire, participants were asked to give examples of habitat-related data and to provide their own definitions. They were also asked to provide an example of use of habitat-related data from their own area of work. The responses provided illustrated a number of interesting points:

i. There were a number of differences in the initial responses to what constitutes habitat-related data. The most common responses related to biological and ecological characteristics and the measurement of them.

ii. It was clear that many interviewees did not necessarily regard contextual (i.e. environmental) information as important or core components of 'habitat-related data'.

iii. Discussions were often tied up with detailed issues of definitions. All groups interviewed discussed the issue of defining the term 'habitat'. The consensus seemed to be that a 'habitat' is a place or environment and refers to inanimate factors or abiota, and that a 'biotope' was the combination of abiota and biota. It was also agreed that many regard the word 'habitat' as synonymous with the term 'biotope'.

iv. Debate and written responses revolved around the discussion of various types of interpreted or derived data. No reference was made to raw data.

v. While differences in the perception of definitions and information are most likely to be explained by the result of different job functions, and by the different activities and processes that habitat-related data are utilised for, it is unclear whether this has any operational impact on the agencies concerned. There is room for confusion or misinterpretation, but the extent to which this may occur is difficult to quantify.

A working definition of 'habitat-related data' was developed from the combination of responses received:

> *Data on the extent, distribution, composition and change in natural, semi-natural and artificial habitats, together with the related environmental and land-use data required to understand, manage and conserve those habitats and their dependent flora and fauna.*

There was a high degree of agreement with this definition from the responses given. A number of issues are raised by the definition which are explored later when considering the key business activities of the agencies and the data sources used by them. If the agencies work on a fully information-driven basis all the areas or issues detailed in this definition should be fully satisfied by data sources their staff utilise.

3.1.4 Key components of habitat-related data

Participants were asked to add to a list of the key components of habitat-related data, on the questionnaire, which resulted in the following list:

i. habitat descriptions/classifications (quantitative or qualitative);
ii. species distribution and abundance;
iii. soils, geomorphology, hydrology;
iv. associated environmental data e.g. nitrogen deposition, air quality, water quality;
v. data on conservation designations, land use, land use restrictions;
vi. data on the nature and magnitude of threats and impacts;
vii. data on historical ecology and land use history;
viii. data on management history;
ix. data on current management practice and any site monitoring;
x. information on the costs of current management of a habitat;

xi. distribution/spatial data on habitat - including information on mosaics and habitat juxtaposition - habitat boundaries and transitions/ecotones;

xii. habitat restoration potential;

xiii. data relating to ecological function and processes operating within the habitat.

Table 3.1 documents the key components of habitat-related data derived from the working definition outlined above. This definition implies the consideration of not simply the processes within a habitat or ecosystem but also the wider environmental system they interact with.

Table 3.1 IPSEMA, the key components of habitat-related data.

Category	Data component
Influence	the nature and magnitude of existing (and potential) threats and impacts
State	species/habitat distribution and abundance soils, geomorphology, hydrology associated variables e.g. nitrogen deposition, air quality, water quality historical ecology and land use history management history distribution/spatial data on habitat - including information on mosaics and habitat juxtaposition - habitat boundaries and transitions/ecotones habitat restoration potential
Process	ecological function and processes operating within the habitat e.g. population dynamics, successional dynamics, critical loads, causal links
Effects	species/habitat: loss, modification e.g. change in distribution, populations, composition
Management Actions	conservation designations, land use, land use restrictions current and historical management practice and any site monitoring costs of current management

The key components have been divided into five functional categories: Influence, State, Process, Effects and Management Actions (ISPEMA) derived from a development and expansion of the 'pressure-state-response' model (OECD 1991). Information on all these components is necessary to fulfil the definition of habitat-related data and to ensure that conservation goals are met.

Discussion

The introductory sessions produced a number of interesting points indicating the role that habitat-related information has in the statutory conservation agencies. There was concentrated discussion about whether data on conservation designations were a component of habitat-related data, suggesting that they normally represented an end product of habitat conservation. While conservation designation is a function of the agencies and has a bearing upon the habitat information requirements for a site or area, it was not viewed as habitat-related data by some staff. There are two likely reasons for this: firstly, that the different roles of staff are driving their personal information requirements and hence outlook, or secondly, that staff are oriented more closely to the subject matter than the function of their work. This issue is explored further in the following section. Other related points emerged from the initial interviews including that:

i. there are no institutional or national/international standards or definitions for the word 'habitat';

ii. no general agreement exists over the relationships between habitat and species data and on the differences between the two. However, it was agreed that there is a reciprocal relationship between them with important differences in emphasis and scale;

iii. there is a measure of agreement about the nature of habitat-related data and of their central importance in the work of the statutory conservation agencies.

The perception of what constitutes 'habitat-related data' varies in detail according to the role and position of a staff member in their organisation. Not surprisingly, specialists would be expected to require information on the ecology of a habitat and be less interested in information such as the protection status of a site that would be important to a site manager or policy maker. The extent to which this issue presents a problem to an organisation is difficult to evaluate. However, it does raise a number of information-related issues. If staff do not value data that they consider are ancillary to their main purpose, those data may not be recorded or may be under-recorded. If staff do not take 'ownership' of data, then they will take less care in recording, managing and reporting them. Although this information may be regarded as peripheral by individual staff, it may nevertheless be of great interest to the organisation. To function efficiently an organisation needs to have a clear delineation of what information is required for specific staff and functions and the overall information resources required by the organisation. A coherent corporate information strategy, that has the ownership of the staff, may help people to conceptualise where their information fits into the wider system and the contribution it makes. It would also help explain the value of 'ancillary' data.

Conclusions

i. There is no coherent corporate view amongst country agency staff on the role and use of habitat conservation and related data sources within the statutory agencies.

ii. While the nature and goals of species conservation and the data necessary to support it appear to be clearly identified by all staff, the situation with habitat conservation, and the nature of habitat-related data, is more nebulous.

iii. The lack of a corporate view and the disjunctive perception of habitat-related data may affect the organisation's ability to record, manage and report information accurately and efficiently.

3.1.5 The 'key business activities' of JNCC and the country agencies

'Key Business Activities' (KBAs) are defined as those functions and activities that are performed by staff of the Support Unit and the country agencies that are essential to meet their objectives and 'goals for conservation' as laid out in the various corporate plans. The term 'key business activities' was introduced, for the purpose of this project, by the JNCC.

'Key business activities' identified from corporate documents

For each agency the following documents were consulted:

JNCC - Annual Report 1993-1994, Work Plan 1997-98, Corporate Plan 1997-98 – 2000 (JNCC 1994b, 1995, 1997).

EN - Third Report 1st April 1993-31st March 1994, Corporate Plan 1996-1999 (EN 1994, 1995).

CCW - Annual Report 1994-1995, Corporate Plan 1996-97 - 1998-99 (CCW 1994, 1995).

SNH - Annual Report 1992-1993, Plans and Progress 1995-96 (SNH 1994, 1995).

The key work activities identified from the published reports of the agencies are shown in Table 3.2.

Table 3.2 'Key business activities' identified from corporate documents.

Key Business Activity	JNCC	EN	CCW	SNH
International Designations/Conventions	✓	✓	✓	✓
European Designations	✓	✓	✓	✓
Nature Conservation Designations	✓	✓	✓	✓
Landscape Designations	×	×	✓	✓
Recreation Designations	×	×	✓	✓
Statutory Casework	×	✓	✓	✓
Non-Statutory Casework	×	✓	✓	✓
Agri-environment	✓	✓	✓	✓
Grant Aid	×	✓	✓	✓
Licensing	×	✓	✓	✓
Formal Consultation	×	✓	✓	✓
Research Projects	✓	✓	✓	✓
Promotion and Promulgation	✓	✓	✓	✓

While Table 3.2 shows apparent similarities in the business activities of the statutory agencies, there are actually marked differences in terms of the functions they carry out in relation to these 'business activities'. For example, JNCC has a responsibility to develop standard guidelines for the selection of SSSIs but it is each of the country agency's responsibility to identify sites that meet these guidelines and to carry out the notification process. EN has a different remit from CCW and SNH, in that it is not responsible for landscape and recreation issues. Although these are related to nature conservation purposes and in themselves will require some use of habitat-related information, this study has not concentrated on these aspects, as they are outside the remit of the JNCC. However, these fall within the KBAs of CCW and SNH and, in CCW's case, the new purposes introduced by the 1995 Environment Act for National Parks bring a greater prominence to wildlife conservation (see Chapter 2 and below).

Table 3.2 therefore masks the relative importance of different KBAs to the agencies. As conservation of species and habitats is the ultimate goal of all four agencies, the identification, gathering, maintenance and management of data related to conservation value could be argued to be the primary business activity. However, as illustrated above, each agency has a different role to play in the delivery of the KBAs and consequently will place a different emphasis on the importance of maintaining data from different stages in the process.

Table 3.3 outlines the types of site designation and the levels at which each conservation organisation delivers different functions.

Table 3.3 Functions related to site designation.

Key:		
	✓	Responsible for
	✗	Secondary responsibility
		Little or no involvement
	N2K	Natura 2000 site series

JNCC Support Unit	**Site type**			
Function	**Ramsar**	**N2K**	**SSSI**	**MNR**
Identification[1]	✗	✗	✗	✓
Notification				
Management				
Monitoring	✗	✗	✗	✓
Audit	✓	✓	✓	✓
Reporting	✓	✓	✓	✓
Standards[2]	✓	✓	✓	✓

1 Actual identification of individual sites based upon the application of standard guidelines/criteria which contribute to an overall listing as either/or Ramsar, Natura 2000, SSSI etc.

2 The Support Unit's role in developing standards is in research and monitoring for nature conservation, (HMSO 1990) and setting guidelines and standards for site selection.

Country Agencies	**Site type**			
Function	**Ramsar**	**N2K**	**SSSI**	**MNR**
Identification[1]	✓	✓	✓	✓
Notification	✗	✗	✓	✓
Management	✓	✓	✓	✓
Monitoring	✓	✓	✓	✓
Audit			✓	
Reporting	✓	✓	✓	✓
Standards[3]	✓	✓	✓	✓

3 The Agencies are responsible for the operational implementation of standards e.g. for site monitoring.

It is apparent that the functions undertaken by each of the agencies form a continuum towards a single end-product and thus it is very difficult clearly to categorise or separate the different functions in relation to each KBA. Conversely, because it is a continuum, it should be possible to create a holistic picture of joint information needs. This should, in turn, enable each agency to view and appreciate why there is a different emphasis placed upon the importance of maintaining information from different stages in the process. Following this model, it should be possible to develop 'conservation process' models for all major activities including agreements on the levels of investment required by whom and for what purpose at each stage in the process.

'Key business activities' of the country agencies identified from workshops

The original workshop questionnaire identified four major categories of KBAs:

i. legal requirements;
ii. advice;
iii. internal liaison/use;
iv. knowledge provision.

During the workshops, these areas were reworked and re-organised into six new categories:

i. legal requirements and international conventions and agreements;
ii. advice;
iii. management and monitoring;
iv. information provision and knowledge development;
v. education, promotion and publicity, community involvement;
vi. landscape and recreation (SNH and CCW).

Table 3.4 represents the combination of KBA categories from published agency documents and questionnaire responses. The numbers refer to the frequency with which staff referred to each KBA and therefore give a representation of the activities conducted by staff in the sample, but not necessarily of the importance of the KBA in each agency. The blank cells correspond to KBAs conducted by the conservation agencies but not referred to by staff in the workshop sample. The frequency of responses therefore gives an indication of the representation at the workshops of different types of staff from the agencies. There is a bias within country agency representation, with regional staff under-represented in the sample. The single geographic location, smaller number of total staff and less diverse key functions of the Support Unit meant that its staff sample was more representative of the organisation.

The most significant difference in KBAs between the three agencies is clearly the wider brief of the Countryside Council for Wales and Scottish Natural Heritage, which have additional responsibility for recreational activity, access and landscape planning.

From the questionnaires, it is also clear that the three country agencies differ in the manner in which they implement similar KBAs through emphasising different activities. This is partly a reflection of the different habitats and pressures on these, which occur in the different geographical areas they cover. However, it also relates to different organisational structures and initiatives: for example, EN appeared to be further developed with implementation of landscape unit characterisation with its Natural Areas/Countryside Character programme than either CCW or SNH.

Table 3.4 'Key business activities' in the statutory conservation agencies.

Key Business Activity	Frequency	Specific components
a) Legal requirements & Conventions		
International Conventions & Directives	8	RAMSAR, Biosphere & Biogenetic Reserves, World Heritage Sites
UK Biodiversity Action Plan	29	Information, Reporting, Favourable Conservation Status, Species
UK Sustainable Development Strategy		
European Directives	8	Birds Directive, Habitats Directive
Special Protection Areas	13	Identification, Designation, Reporting, Favourable conservation status
Special Areas for Conservation	32	Identification, Designation, Reporting, Favourable conservation status
National UK		
Quinquennial Review	22	Species status listing
SSSI Site Selection/Notification	67	Identification, Notification, Review, Survey, Including NNRs
Marine Nature Reserves	2	identification, Notification, Review, Survey
Second Tier Sites (e.g. SNCIs)	2	Identification, Notification, Review, Survey
b) Advice		
International Conventions & Directives	3	
European	7	DGXI, European Environment Agency, ETC Nature Conservation
Other Government Agencies and Departments	36	
Parliament	8	
NGOs	9	
Country Agencies	32	(66% of which is intra-agency)
Externally	31	Landowners and managers, the public and individuals
Advocacy	5	
c) Management and Monitoring		
Management of designated sites	37	SPAs, SACs, NNRs, SSSIs, MNRs
Monitoring of designated sites	22	Time-series, SPAs, SACs, NNRs, SSSIs
National state reporting	2	
State of habitat/resource	3	(Sub-national scale)
Species recovery	13	
Species translocation	1	
Restoration	4	
Audit	6	
Nature Conservation Orders		
Management Agreements		
Local Nature Reserves		
Grant aid		e.g. Wildlife and reserve enhancement schemes, Tir Cymen, ESAs
Licensing		
Planning casework	10	

Table 3.4 (Contd.)

Key Business Activity	Frequency	Specific components
d) Information Provision and Knowledge Development		
Information to inform and run all functions	4	
Area-based characterisation	15	Natural Areas, Natural Heritage Zones
Providing contextual information	7	
Assessing conservation status	4	
Survey		
Inventory	19	
Baseline	1	
Setting conservation priorities	12	
Data collection	8	
Data computerisation	6	
Conduct and influence research	9	
Developing information systems	3	
Establishment and development of GIS		
Developing evaluation techniques	9	
Setting/maintaining common/quality standards	22	Habitat definition and classification, Relating habitats
e) Education, promotion & publicity, community action		
Education	13	Landowners & managers, wider public, dissemination of ideas
Promotion & Publicity		
Partnership	2	
Liaison - internal	6	
Liaison - external	3	Public, NGOs
Community Action		
f) Landscape and recreation (SNH & CCW only)		
Landscape Designations		National Parks, AONBs, NSAs, Regional Parks, Heritage Coasts
Recreation Designations		Country Parks, National Trails, Long Distance Routes

For explanation of acronyms, please see the table of acronyms at the start of the document.

'Key business activities' of the Joint Nature Conservation Committee

The JNCC has an overall co-ordination role and tries to integrate and represent the nature conservation policy of Great Britain in Europe and internationally. This role is reflected in the following goals taken from the JNCC's mission statement (JNCC 1994b):

i. To devise and maintain common standards and protocols for nature conservation designations, for monitoring the overall effectiveness of this work and for environmental audit.

ii. To promote, through the establishment of common standards, the free interchange of data between the country agencies and with external partners.

iii. To advise on nature conservation and related issues affecting Great Britain as a whole.

iv. To pursue wider international goals for nature conservation (encouraging environmental sustainability, biological diversity and earth science conservation), including the provision of relevant advice to government.

v. To commission new research and collate existing knowledge internally, or by contract, in support of any of the activities listed above, and to disseminate the results.

JNCC seeks to deliver this mission through a collaborative network of staff from the country agencies. This embraces those staff assigned to the JNCC Support Unit and those working on special functions within their own agencies such as the Lead Co-ordination Networks and specialist inter-agency working groups. Responses to the workshop questionnaire from JNCC staff reflected the differences between their agencies' role and that of the country agencies. Although in many cases they are co-ordinating, supporting and developing the same initiatives to the country agencies, their functions are often very different, and thus their information requirements can also be different.

Discussion

There was general agreement on the KBAs of the three country agencies, although there were significant differences in emphasis and the manner in which they are achieved in each country. Some country agency staff did not identify strongly with the goals and KBAs of the JNCC and saw it as an unnecessary tier of organisation and administration.

Conclusions

i. Common KBAs can be identified for the three country agencies.

ii. The JNCC had some KBAs in common with the country agencies plus additional strategic activities specific to the Support Unit.

iii. The special role, function and value of the JNCC was not fully appreciated by staff from the country agencies.

iv. There were inevitable overlaps in several KBAs relating to the execution of different functions for the same initiative or programme (e.g. SSSIs) and this caused problems in relating use of, and requirements for, habitat-related data to specific KBAs.

v. An assumption behind the methodology employed for this study was that KBAs were clearly identified by staff, and that staff would be able to link their use of habitat-related data to these KBAs. It became clear that this was often not the case.

vi. Either staff found it difficult to identify the functions they had to undertake in relation to the KBAs of their organisations or the KBAs of the organisations did not reflect the activities actually undertaken by staff. Either or both scenarios suggest organisational and functional management problems.

3.1.6 The use of habitat-related data in relation to 'key business activities'

Individuals use information in two main ways. They can use habitat-related data or interpreted information to fulfil their key tasks, or they can supply data or interpreted information to others. For either of these stages, the data may be originated from a third party or by the individual. Information may be passed on through physical media or verbally. The questionnaire devised for this study examined the sources of habitat-related data that were used for different KBAs, the scale they were used at and the value that was given to each source in its use for particular KBAs.

Appendix 3 lists all the data sources and types used for the KBAs listed by all respondents. Data from this table are summarised within this section.

Data sources used for key business activities

A variety of data sources were recorded by workshop respondents. For the purposes of analysis and brevity, these have been grouped into the seven main categories outlined in Table 3.5.

Table 3.5 Data sources recorded by respondents during workshops.

Source category	Components
Surveys	NVC, Phase 2, MNCR, River Habitat Survey, 5 yearly water quality surveys, Phase 1, Marine Phase 1
Inventories and resource surveys (Including databases & information systems)	Habitat resource surveys & inventories (e.g. Ancient woodland inventory, Limestone Pavement, peatland survey), Forestry Commission Census, BRC atlases, Red Data Books, BSBI atlases, Countryside Information System, Invertebrate Site Register, Rivers Database, Loch Survey Database, IFE (RIVPACS) Database, NVC Survey database, COREDATA, Features Database, Databases for primary data sets, JNCC database & GIS for statutory sites, JNCC sites repository system, MNCR database, EC Habitats Directive databases, WCMC & related international data sets
Remote sensing, aerial photography and land cover maps	CORINE Biotopes, ITE Land Classification, Remote sensing data, LCS 88, Ordnance Survey data, Old maps, British Geological Survey, Air photos
Site files, surveys and assessments	SSSI Habitat surveys, Second tier site inventories, Site files, Environmental Assessments, Environmental Impact Statements
Literature	Published literature, Taxonomic literature, Lead Co-ordination Networks literature, Published international sources, EC literature on directives
Organisations	BRC, BSBI, Entomological societies, Local Record Centres. LPBR (JNCC), NRA/EA, Universities, Institutes, EU/International, NGOs (e.g. BTO, National Trust, Wildfowl Trust)
Staff/people	Local & national experts (internal and external) Self/Colleagues

For explanation of acronyms, please see the table of acronyms at the start of the document.

In functional terms, many of the data source categories are similar. The majority of sources listed in the 'Surveys', 'Inventories/Resource Surveys' and 'Site files' categories are inherently field survey-based in their origins but they contain a mixture of both raw and interpreted information. For example, a site survey may simply contain recorded quantitative observations relating to a feature; these results may then be analysed and interpreted within the relevant environmental or habitat context of that location in order to explain significance. The majority of these survey-related data are within direct agency control. Data sources from the 'Remote sensing', 'Literature', 'Organisations' and 'Staff/People' (external) categories are generally outside agency control.

Scale

Table 3.6 shows the scale at which different data sources were used. The numbers in each column refer to the frequency that data sources in each category were referred to for use at different scales. The absolute numbers should be treated with caution, as the sample of respondents was small, and not necessarily representative, but they do indicate trends in data source use. Most respondents used data sources at more than one scale; therefore, frequencies should not be summed across rows.

Table 3.6 The scales at which data sources were used.

Source	Local	Regional	National	GB	UK	Intl.
Surveys	22	16	14	6	0	0
Inventories and resource surveys (including databases & information systems)	13	18	19	14	1	2
Remote sensing, aerial photography and land cover maps	4	5	4	1	0	2
Site files, surveys and assessments	8	3	2	1	0	0
Literature	3	1	1	3	0	2
Organisations	12	5	11	0	0	2
Staff/people	6	3	2	1	1	1

The first two categories, 'Surveys and Inventories'/'Resource Surveys' were the most heavily used sources at local to national levels. This is as might be expected of organisations that have historically been survey-based. The slightly greater use of collated survey information at the national to international scales (in the 'Inventories' category) may indicate the greater value of interpreted data at these levels. The third category 'Remote sensing' includes air photos and the underlying data show these were cited only twice and were only used at the local level. In view of its great potential, the lack of easily accessible interpreted air photo data at wider geographic scales may preclude its wider use. The level of use of other organisations as data sources is relatively high. The conservation agencies will have less control, if any, over the quality and management of these data sources.

Table 3.7 is categorised by the different data sources used for KBAs and shows which KBAs each data source was most frequently used for. The categories 'Surveys' and 'Site files, surveys and assessments' from Table 3.6 were aggregated in Table 3.7 because functionally they are inherently similar. The distinction is basically one of scale and was therefore retained in Table 3.6. Table 3.7 shows a list of 'functional KBAs' derived from the master data table in Appendix 3. These data were effectively normalised and the KBAs summarised in terms of the operational function they were serving. As with the other tables, the frequencies quoted relate to KBAs conducted by staff who responded to the questionnaires in the workshops and do not necessarily represent a true cross-section of each conservation agency. Nevertheless, the data indicate a number of trends. Only two data source categories, 'Surveys and sites' and 'Landcover', represent any first degree or raw data. The remaining five columns are based upon second-degree or interpreted data. This raises two important points: firstly that primary data may not be accessible (either physically, or in an appropriate form) or that the demand is for secondary or interpreted data.

Table 3.7 Frequency of functional KBAs using different data sources.

Functional KBA	Surveys & sites	Inventories	Land cover	Literature	Organisations	People	Total
Advice	19	16	5	5	11	3	59
Audit	4	1	1	0	4	0	10
Contextual data	3	1	1	0	1	1	7
Designation	7	3	1	1	5	9	26
Education	4	1	1	1	1	0	8
Evaluation	11	14	1	3	10	6	45
Identification	1	3	0	0	0	2	6
Information systems	5	13	6	0	9	1	34
Management	10	12	4	2	8	7	43
Monitoring	5	7	0	0	6	4	22
Partnership	1	1	0	0	0	0	2
Planning casework	3	0	1	0	1	2	7
Reporting	6	18	0	2	1	2	29
Research	6	2	1	1	1	0	11
Selection	11	10	2	3	3	3	32
Standards	11	5	0	3	7	2	28
Survey	1	0	0	0	6	0	7

Table 3.8 The mean value of data sources for different functional KBAs; scale 1 (good) - 5 (poor). The values in grey-shaded cells indicate a mean that masks a bipolar set of rankings.

Functional KBA	Surveys & sites	Inventories	Land cover	Literature	Organisations	People
Advice	2	2	2	2	2	1
Audit	3	1	1		2	
Contextual data	1	2	2		5	1
Designation	2	2	3	4	3	1
Education	2	3	2	5	1	
Evaluation	2	2	1	4	2	2
Identification	1	2				1
Information systems	2	2	3		3	1
Management	2	2	3	4	3	2
Monitoring	3	1			3	1
Partnership	4	1				
Planning casework	4		3		3	2
Reporting	3	2		3	3	2
Research	2	2	1	2	1	
Selection	2	2	3	2	3	1
Standards	2	2	5	2	2	2
Survey	3				2	

The 'Surveys' and 'Inventories' sections were used significantly across most KBAs, forming the backbone of information supply. The 'Land Cover' and remotely-sensed data sources mostly consisted of primary data sources such as air photos and cartographic data. These were evenly used at a low frequency for most KBAs. Surprisingly they were not used, or little used, for providing contextual data, monitoring or survey activities. 'Organisations' and 'People' were the most used information sources after 'Surveys' and 'Inventories'.

The value of habitat-related data sources

Respondents' ratings of the value of different data sources were sought in the data sources evaluation section of the workshop questionnaire. The values given were specific to individual KBAs, and represent a qualitative judgement made by the respondent in rating the value of that source in conducting a particular KBA. Table 3.8 shows the average value score given by respondents for each category of data source for use for a particular KBA. In some cases, the average (mean) values show a score that masks a bipolar set of rankings. These cells are shaded grey and represent a situation where respondents differed widely in their assessment of the data source. The average values in each cell were derived from a greatly variable number of responses ranging from 1 to 19 and should therefore be treated as indicative (for frequency of responses see Table 3.7).

In general, data sources have been graded as above average or average value. The most often cited sources – 'Surveys' and 'Inventories' received higher than average ratings. 'Inventories' gained higher scores than 'Surveys'. In most cases inventories are collated and or summarised survey results. The results here suggest that either staff were less critical of the data they use that are interpreted, or that staff find interpreted data more useful or, perhaps, a combination of both factors. The 'Organisations' category receives a variable set of ratings but the 'People' category receives very good ratings. This is somewhat contradictory, as this category might include people from external organisations, though it may be partly explained by the fact that communication with individuals can be more easily suited to specific enquiries than can the official line of an organisation taking a corporate position through published information. Although 'Landcover' data sources are generally neglected (see Table 3.7) they were generally well rated when used which suggests that further use may be made of such sources.

Strengths and weaknesses of data sources

In the questionnaire, respondents were asked to list the strengths and weaknesses of the data sources they employed for KBAs, using and adding to a standard set of data evaluation criteria. The tables below (Table 3.9) are presented in terms of an expanded version of the data source groupings used above. They show the strengths and weaknesses respectively of each data source type, in terms of the frequency with which each criterion was mentioned by respondents. These tables use the standard criteria from the questionnaire (Appendix 2) and those added by individuals. Comments in *italics* refer to specific comments made by questionnaire respondents and are unrelated to the frequency score when this is greater than one.

Table 3.9 Strengths and weaknesses of data sources.

NVC

Strengths/Frequency		**Weaknesses/Frequency**	
Date	3	Date	2
Original aim and purpose	1	Original aim and purpose	1
Original customer/user	1	Original customer/user	0
Geographical coverage	4	Geographical coverage	7
Scale	1	Scale	1
Resolution	2	Resolution	1
Spatial unit of description	5	Spatial unit of description	1
Sampling methodology	3	Sampling methodology	0
Repeatability	4	Repeatability	2
Use of meta data	0	Use of meta data	1
Availability of area measurements	5	Availability of area measurements	2
Availability of mapped information	5	Availability of mapped information	0
Availability of associated variables or environmental data	2	Availability of associated variables or environmental data	1
Quality	3	Quality	2
History/processing chain	0	History/processing chain	0
Baseline data	5	Baseline data	0
Change data	2	Change data	2
Valuable for context & management	1	Does not work well in NW Scotland *(classification unrepresentative)*	1
Depends on scale of use	1	Expensive	1
Widely used	1	Not digitised	1

Phase 1

Strengths/Frequency		**Weaknesses/Frequency**	
Date	1	Date *(Some 9 years old)*	5
Original aim and purpose	3	Original aim and purpose	0
Original customer/user	0	Original customer/user	0
Geographical coverage	6	Geographical coverage	1
Scale	7	Scale	0
Resolution	6	Resolution	0
Spatial unit of description	3	Spatial unit of description	0
Sampling methodology	0	Sampling methodology *(not standard)*	2
Repeatability	5	Repeatability *(variable)*	1
Use of meta data	0	Use of meta data	0
Availability of area measurements	6	Availability of area measurements	0
Availability of mapped information	6	Availability of mapped information	0
Availability of associated variables or environmental data	1	Availability of associated variables or environmental data	0
Quality	5	Quality	2
History/processing chain	1	History/processing chain	0
Baseline data	0	Baseline data	0
Change data	0	Change data	0
		Not available digitally	1

Table 3.9 Strengths and weaknesses of data sources (Contd.).

Phase 2

Strengths/Frequency		**Weaknesses/Frequency**	
Date	1	Date	4
Original aim and purpose	1	Original aim and purpose	1
Original customer/user	0	Original customer/user	0
Geographical coverage	1	Geographical coverage *(Only a few give good coverage)*	1
Scale	2	Scale	0
Resolution	2	Resolution	0
Spatial unit of description	0	Spatial unit of description	0
Sampling methodology	2	Sampling methodology	1
Repeatability	3	Repeatability	1
Use of meta data	0	Use of meta data	0
Availability of area measurements	2	Availability of area measurements	0
Availability of mapped information	3	Availability of mapped information	1
Availability of associated variables or environmental data	1	Availability of associated variables or environmental data	0
Quality	1	Quality	2
History/processing chain	0	History/processing chain	0
Baseline data	1	Baseline data	0
Change data	0	Change data	0
Cost effective	1	Limited supply	1

Site files and Surveys

Strengths/Frequency		**Weaknesses/Frequency**	
Date	3	Date	5
Original aim and purpose	4	Original aim and purpose	2
Original customer/user	0	Original customer/user	0
Geographical coverage	5	Geographical coverage	1
Scale	1	Scale	0
Resolution	1	Resolution	1
Spatial unit of description	3	Spatial unit of description	0
Sampling methodology	0	Sampling methodology	5
Repeatability	1	Repeatability	3
Use of meta data	0	Use of meta data	1
Availability of area measurements	0	Availability of area measurements	3
Availability of mapped information	1	Availability of mapped information	5
Availability of associated variables or environmental data	1	Availability of associated variables or environmental data	0
Quality	4	Quality	8
History/processing chain	1	History/processing chain *(unstructured)*	1
Baseline data	1	Baseline data	0
Change data	0	Change data	1
Immediate low-cost access	1	Not documented	1
Only source	1	Accessibility poor	1

Table 3.9 Strengths and weaknesses of data sources (Contd.).

Inventories

Strengths/Frequency		**Weaknesses/Frequency**	
Date	8	Date	6
Original aim and purpose	7	Original aim and purpose	2
Original customer/user	2	Original customer/user	0
Geographical coverage	21	Geographical coverage	5
Scale	5	Scale *(Imbalanced classification - CORINE)*	3
Resolution	3	Resolution	0
Spatial unit of description	3	Spatial unit of description	3
Sampling methodology	8	Sampling methodology	4
Repeatability	5	Repeatability	3
Use of meta data	4	Use of meta data	0
Availability of area measurements	9	Availability of area measurements	2
Availability of mapped information	7	Availability of mapped information	5
Availability of associated variables or environmental data	3	Availability of associated variables or environmental data	4
Quality	8	Quality	5
History/processing chain	1	History/processing chain	4
Baseline data	9	Baseline data	0
Change data	2	Change data	5
Accessibility	1	Accessibility	1
Documentation	1	Documentation	3
Widely used by others	1	Outside Agency control	1
Selective	1	Poor linkage	1
Cost-effective	1	No original data	1

Landcover

Strengths/Frequency		**Weaknesses/Frequency**	
Date	3	Date	1
Original aim and purpose	1	Original aim and purpose *(Limits often exceeded - LCS88)*	2
Original customer/user	1	Original customer/user	1
Geographical coverage	6	Geographical coverage	1
Scale	2	Scale	1
Resolution	2	Resolution	2
Spatial unit of description	3	Spatial unit of description	1
Sampling methodology	2	Sampling methodology	1
Repeatability	2	Repeatability	2
Use of meta data	1	Use of meta data	0
Availability of area measurements	3	Availability of area measurements	1
Availability of mapped information	4	Availability of mapped information	0
Availability of associated variables or environmental data	1	Availability of associated variables or environmental data	0
Quality	2	Quality	2
History/processing chain	0	History/processing chain	0
Baseline data	4	Baseline data	1
Change data	1	Change data	1
		Prohibitive cost *(OS)*	1
		Requires interpretation *(Air photos)*	1

Table 3.9 Strengths and weaknesses of data sources (Contd.).

People

Strengths/Frequency		Weaknesses/Frequency	
Date	3	Date	2
Original aim and purpose	3	Original aim and purpose	0
Original customer/user	3	Original customer/user	0
Geographical coverage	1	Geographical coverage	3
Scale	0	Scale	1
Resolution	1	Resolution	0
Spatial unit of description	0	Spatial unit of description	0
Sampling methodology	0	Sampling methodology	2
Repeatability	0	Repeatability	5
Use of meta data	0	Use of meta data	0
Availability of area measurements	0	Availability of area measurements	2
Availability of mapped information	0	Availability of mapped information	2
Availability of associated variables or environmental data	0	Availability of associated variables or environmental data	1
Quality	3	Quality	5
History/processing chain	0	History/processing chain	2
Baseline data	0	Baseline data	0
Change data	0	Change data	1
Accessibility	1	Accessibility	2
Immediate and low-cost access	2	Subjective	1
Only source	2	Inconsistent effort	1
Basis of most routine decision making	1	Weak supporting information	1
		Information/skill lost with staff	2

NVC, Phase 1 and Phase 2 surveys have been chosen from the 'Surveys' category used above because these were the most frequently used data sources and it was considered important to examine any differences between these sources. The Literature and Organisations categories have not been further analysed here, due to patchy responses and because they referred to widely different types of information, which meant that further generalisation would have been inappropriate. The raw data can be examined in Appendix 3.

NVC sources were the most commonly referred to. Nearly twice as many strengths were listed compared to weaknesses, but there is also much contradiction. This contradiction is likely to be due to the use of different specific sources and the use of different data types for different functions or KBAs. This analysis shows that the greatest strength of NVC was the availability of maps and baseline data. Perhaps surprisingly, the fact that it was widely used and has become a *de facto* standard for survey in the country agencies was cited only once. On balance, there appeared to be a consensus that quality and repeatability were strengths but there are clearly significant exceptions. The most common weakness was geographical coverage which was possibly a function of the cost and effort required to produce NVC survey data, but may also have been due to lack of knowledge of available sources.

Phase 1 survey sources were less frequently used than NVC sources. The strengths and weaknesses for Phase 1 were less contradictory with geographic elements being cited as strengths, as might be expected. The significant weaknesses were that many sources were several years old and that there was some variation in methodology in early surveys. These results reflect Wyatt's (1991) findings from a review of Phase 1 survey. CCW had just completed an extensive Phase 1 mapping exercise which explained the fact that Welsh Phase 1 was more highly rated than English Phase 1. The high cost of Phase 1 survey means its further widespread use in England or Scotland may be limited.

Phase 2 sources appeared, on average, to have more strengths than weaknesses but again the picture was contradictory in relation to specific sources or applications. The main strengths appear to lie in repeatability and the provision of mapped information, with the main weaknesses being age and possibly quality. There will be some overlap with NVC sources as NVC is now used for Phase 2 survey.

In terms of frequency, the 'Site files' and 'Surveys' category had around a third more weaknesses than strengths. The most frequently quoted strengths were in all but one case offset by higher weakness counts. Although this means some staff found the source useful for some applications, there were more who found the source unsatisfactory. Quality, age and sampling methodology were the major weaknesses.

In contrast to 'Site files', 'Inventories' were generally well received, rating very highly for geographical coverage, and satisfactorily on the provision of mapped information, area measurements and baseline data. The different ages of sources and their variation in quality, combined with a poor rating for change data, indicated that inventories were rarely repeated. Quality may be rated as poor because inventories may now be used for applications outside their original purpose - in the absence of more relevant data. The lack of repeats may be because they involve significant resource investment.

Despite their low level of use, landcover-based data sources had, on balance, more strengths than weaknesses but again there was a contradictory picture related to the variation of source types, and original applications, and the purpose for which they were used. As may be expected, geographical coverage was a commonly cited strength. Additional comments gave some insight to the limited use of landcover-based data. Ordnance Survey data were regarded as being prohibitively expensive and air photos were not used in some cases when they required expert interpretation.

The data above suggest that people were used extensively as informal data sources. The ratio of cited strengths to weaknesses of using people as data sources was 2:3. The ratings show that information was strong on relevance and purpose but potentially weak on repeatability and quality. An important point made by two respondents was that information and skills are lost when staff leave.

Discussion

A variety of data sources were used. Most sources were derived from survey material which was then analysed, summarised and collated in different forms for different uses in survey reports, inventories, databases and personal experience. At each step, there may be some form of generalisation or loss of detail or specific meaning. If common standards are not rigorously enforced at all stages, information will become less reliable and less comparable at each stage. Much information is often old or of variable date. This problem will be compounded in inventories and databases unless very careful updating and version control is implemented. Most sources used related to describing the stock (extent), composition or distribution of habitats and or species. Most data sources presented interpreted data. This suggested that either primary data were unavailable or that there was no demand for them. If we consider the information used with respect to the ISPEMA model (Table 3.1) it is evident that relatively little information on influences (pressures and threats) impact or change (effects) was used. Most effort appeared to be concentrated on the 'State' category. While this is clearly important, the relative neglect of other categories may limit the capability of the statutory nature conservation agencies to examine environmental pressures and management responses or to develop more active preventive strategies.

Although landcover-based sources were well rated, they appeared to be used very little, possibly because they often require specialist expertise in interpretation. Sources such as air photos were not often used for survey-related applications (like targeting areas of search). However, the results from this project indicate that little survey work is now carried out by the country agencies, largely because most such work is now contracted out. The results may also suggest that less survey work in total was being conducted than in previous years. Survey-based data sources are currently the most heavily used for a range of KBAs. If less survey work is being conducted, the survey-based information base will progressively become more dated and less relevant. There is little indication that this potential gap is being met with any other new data sources. Thus the need specifically to target survey work is great, to support information on the state of habitats, but also to consider influences, effects and management actions. Comments made by staff through the study suggested that new survey work was limited due to lack of financial resources.

In general, data sources have been graded as above average or average value. This would suggest that staff were generally happy with the utility of the data that are available for them to conduct their KBAs. However, feedback during the workshops revealed that this was not always the case. Additionally, the evaluation of strengths and weaknesses showed a number of contradictions. In consequence, it is difficult to reach firm conclusions regarding the values attributed to data sources for different uses. It is probable that, when asked to provide an overall grading, staff were less critical than when asked to consider specific strengths or weaknesses. Particular data sources also vary in their applicability and utility according to specific applications or functions. In some cases, a stated weakness can be more limiting than in others. For example, the lack of quadrat data from a NVC survey may not preclude reporting in NVC communities but would limit the further interpretation of the data into new classifications or the identification of the presence of a key species.

If currently used data sources were meeting KBAs satisfactorily, a more frequent response to several of the strengths categories would have been expected. The responses for quality as a strength were not consistently high. Very few sources were described as cost-effective and compliance to common standards was not often cited as a strength. Either this was not regarded by staff as an important issue, or data sources were not generally regarded as being cost effective or conforming to common standards. If quality was not highly rated in many sources, this suggests there could be a problem for the provision of high quality data to support KBAs.

Organisations and people are relied upon to a large extent as data sources. The characteristics of such sources cause potential problems in terms of maintaining common standards and relevance to need. If staff move within an organisation or outside an organisation, then their knowledge could be lost if it was not documented.

Contextual data and information use was poor. Land use and agricultural-based data were little used. This must limit the ability to describe pressures upon the environment and causes for decline in conservation features. Because the conservation agencies concentrate often upon protected sites, not on the wider countryside, information is geared around semi-natural habitats and distributions that represent a fraction of the resources they are charged to protect. Effort is concentrated around measures of state and, to a lesser extent, process. Little information was available on pressures, magnitude of threat or effectiveness of management responses.

Conclusions

i. Survey-based data sources were the most often used but it is unclear whether these are being updated consistently. There is a continuing need to specifically target survey effort.

ii. There was a heavy reliance upon secondary or interpreted information - this suggests either inaccessibility to primary data, no demand for them or is a reflection of time pressures.

iii. Interpreted information was generally used more widely at the country or GB level and was more highly rated than primary data.

iv. Many of the information sources used, such as some surveys, organisations and literature, were outside the agencies' control.

v. Landcover-based data sources were little used.

vi. It was difficult definitively to assess data sources in terms of their strengths and weaknesses because answers were contradictory and the value of sources was always contingent upon a specific application.

vii. In general, staff in the study did not highlight quality and standards issues as major strengths of data sets they were using for KBAs.

viii. Data and information on environmental influences, processes, effects and management responses were generally neglected.

3.1.7 Potential sources of habitat-related data

Certain sources of habitat-related data exist but were not used or were inaccessible for various reasons. Six such reasons were identified in the questionnaire, and added to during the subsequent workshops:

i. Where there is imperfect knowledge about the data source(s), and potentially interested users are unaware of the available data or information.

ii. Situations of confidentiality, or commercial/military interest.

iii. Expense, where data are too expensive to purchase, or if accessible are in the wrong form and are consequently too expensive in terms of the time and human resources necessary to summarise them or supply them in a usable format.

iv. Where data quality is poor, data are not testable or representative, or there is incomplete coverage, either spatially or temporally.

v. Data quantity - there may be too little or too much data to provide the necessary information for the relevant KBA.

vi. Data ownership - the data may exist but are not made available because of sensitivity over ownership. Staff identified a perhaps quite natural tendency for collectors of hard-won data sometimes to be reluctant to release it to personnel in other organisations or even to other staff within their own organisations.

Potential data sources and reasons for non-availability

During the workshops, staff were asked to list potential sources of habitat-related data and to give the reasons why they were not used or were unavailable. These are summarised in Table 3.10; the figure in brackets after each source title represents the number of times a source was mentioned by the questionnaire respondents.

Table 3.10 Potential sources of habitat-related data listed by workshop participants and reasons for non-or limited availability. (The numbers in brackets refer to the number of responses).

Data/Information Source	Reason
Countryside Information System (2)	Too expensive Lack of expertise Potential utility needs evaluation in a test/pilot situation Resolution too coarse
Remote sensing - LANDSAT/Airborne Data (3)	Too expensive - both for data gathering and technology for interpretation Lack of expertise Technology unproven - at present resolution is too coarse other than for general habitat categories Research is required to link NVC classification to remote sensing categories Discriminatory power for habitats/communities as yet uncertain
MAFF Agricultural Statistics (2)	MAFF reluctant to make data available in some situations - confidentiality? Problems of linkage to other data both spatially and temporally
MAFF Data on Environmentally Sensitive Areas and ESA Monitoring (1)	MAFF reluctant to make data available in some situations - confidentiality? Variations in data quality and spatial and temporal coverage
Research in University Departments (3)	Problems of finding out what data and research projects exist Restricted access and confidentiality May require resources to work into useable format Problems of consistency Future problems over commercial charging?
Student Projects in University Departments (e.g. English Nature LINK Scheme) (1)	Staff unaware of projects Difficulty of access to reports Quality of data and reports variable
Invertebrate Site Register (ISR) (2)	ISR is used but could be more widely accessed if the necessary entomological expertise was available (e.g. for woodland survey/indicator species)
County Wildlife Trusts (1)	Data often available but not in easily accessible form Resources required to retrieve data Problems of data quality and consistency
National Vegetation Classification (1)	Original site records not available No mapping of spatial extent of communities
Institute of Freshwater Ecology Invertebrate Data (RIVPACS) (1)	Only available as a commercial source and thus expensive Problems of data ownership and potential infringement of 'intellectual property rights'
Institute of Hydrology (IOH) Digital Terrain Model (1)	Expensive
National Rivers Authority/River Purification Board, Invertebrate Data (1)	Data quality poor
Phase II Habitat Surveys (2)	Incomplete coverage Variable quality Stored on paper in files and thus not easily accessible Require resources for data retrieval and synthesis
Hydrographic Survey Seabed Data (1)	Restricted access, confidentiality
JNCC and country agency Specialists (1)	Significant amounts of data and information on data sources stored 'in-heads' Data ownership problems - access sometimes limited because individuals lose power and status by giving away data/information
FenBase - fen and bog database (1)	Held outside JNCC/Country agencies Incomplete spatial coverage Problems of consistency of habitat categories
Peatlands Database (SNH) (1)	Limited availability to JNCC and other country agencies outside SNH Technical problems (computer programming)
National Rivers Authority River Corridor Survey (1)	Confidentiality

Discussion

Most potential but inaccessible habitat-related data sources represent specific situations where data that have been collected by other agencies are required by one individual, probably to meet one specific KBA. There was not a high degree of commonality amongst the limited sample of staff over those potential sources that might be of value. Access to data was the most often cited reason for not utilising these data sources. Often, organisational factors prevented their use, although in a number of cases, the expense and resources required to render the data accessible were too high. The issue of data confidentiality was a problem, particularly in connection with agricultural information. As MAFF[5] and ADAS (and now FRCA) data relate to the management of land under the single largest land use in the UK, greater access to these data would clearly be valuable for the conservation agencies. In some cases, lack of appropriate expertise was a limiting factor (e.g. with CIS, the ISR and university research) to the further use of data sources and in these situations further training or access to interpreted data could prove useful.

A commonly cited problem was that of actually knowing which data exist and where. However, even more critical was the difficulty of evaluating the potential of a known data source on the basis of minimal information. The quality of information on which staff base their perception of the potential value of data sources appeared to be seriously limited in many situations. Thus it was difficult to assess the true degree of objectivity about the real potential value of any given data source.

Conclusions

i. A number of sources were not used to their greatest utility.

ii. The most common reasons for non-or under-use were lack of knowledge or supporting information, and the (perceived) poor quality of some data sets for country agency use.

iii. There appeared to be significant problems of confidentiality and data ownership rights.

iv. There was a need for data access protocols with data-holding organisations (particularly other government departments), and if those already existed, educating staff about their function.

v. In the case of data held by universities and potential problems of commercial interest, it would be worth exploring the trade of country agency data in exchange for college data and the exchange of research interests - 'data banking'.

vi. Decisions on the potential value of a system or data source need to be made objectively and on an informed basis. To achieve this it would be useful to develop criteria for evaluation so that suitability for different conservation functions could be rationally assessed.

---

[5] From June 2001 MAFF became part of the new Department for Environment, Food and Rural Affairs (DEFRA). FRCA became the Rural Development Service (RDS) in England and CAP Management Division in Wales.

3.2 REGIONAL DATA SOURCES AND USE – NATIONAL PARKS

3.2.1 National Parks as a case study

Management of the countryside is not systematically structured in Great Britain. Regional-level management focussed first in areas of great landscape beauty with the creation of National Parks and Areas of Outstanding Natural Beauty. More recently, recognition of the importance of regional level management has seen creation of other bodies such as the Sussex Downs Conservation Board and the New Forest Conservation Board. Regional level management is important for a number of reasons. It provides a strategic basis for the conservation and management of nationally or internationally important features that are present within a restricted geographical context. They reinforce local and regional identity and distinctiveness and provide realistic management units that can be dealt with at a sub-national level. In England and Wales over the last forty years, regional-level management has been focussed within National Parks. There are no directly comparable entities in Scotland, but the closest are National Scenic Areas. In many ways National Parks can be considered as the flagship organisations for countryside management in Great Britain encompassing environments of high landscape and nature conservation value and managed by functionally dedicated organisations. This case study seeks to explore the need for information at policy and operational levels and to explore the linkages between the two processes in a National Park.

3.2.2 Policy background

Ten National Parks were established in England and Wales during the 1950s after the 1949 National Parks and Access to the Countryside Act (HMSO 1949). In 1989, the Broads area was added (HMSO 1988) and although not a National Park by name, it has the same status. The 1949 Act defined the purposes of National Parks as:

- preserving and enhancing the natural beauty of the areas;
- promoting their enjoyment by the public.

While the accent was on recreation and access, National Parks became foci for conservation management. National Parks in England and Wales are different to those in other countries, such as the United States of America, where parks are usually state owned and often comprise large tracts of land relatively untouched by human intervention. In England and Wales, National Parks are relatively small, most land is privately owned and the countryside is often highly managed, reflecting the impact of centuries of human occupation. For example, around 40% of the total area of National Parks is agricultural land (Countryside Commission 1993).

Since their establishment in 1949, National Parks have been subject to a number of reviews. The Sandford Committee (HMSO 1974) conducted the first major review and one of its main recommendations was that the enjoyment of National Parks by the public:

"shall be in such a manner and by such means as will leave their natural beauty unimpaired for the enjoyment of this and future generations".

The Committee concluded that, although most conflicts could be resolved by means of planning and management, where the two National Park purposes were incompatible priority should be given to the conservation purpose. This concept became known as the 'Sandford Principle'. In 1991, the National Parks Review Panel, under the chairmanship of Professor Ron Edwards, reviewed the position of National Parks at the request of the Countryside Commission (Countryside Commission 1991a). The panel concluded that National Parks required a new strategic direction in order to meet their original purposes. The Government responded (DOE and Welsh Office 1992) the following year with the promise of legislation which eventually became enshrined in the Environment Act 1995, Part III National Parks (HMSO 1995c). This Act revised the two purposes of the National Parks to:

- conserve and enhance the natural beauty, wildlife and cultural heritage of the National Parks;
- promote the understanding and enjoyment of the special qualities of the Parks by the public.

The Environment Act 1995 also required that National Park Authorities should 'seek to foster' the economic and social well being of local communities. Additionally, the Act made a number of statutory changes:

- it provided for the establishment of National Park Authorities as unitary planning authorities;
- it placed a duty on public bodies whose decisions or activities affect land in the Parks to 'have regard to' the purposes of the Parks when conducting their work;
- where there is a conflict between the two Park purposes that the conservation case should take precedence;
- under Section 66, that each National Park Authority must prepare a National Park Management Plan (Countryside Commission and Countryside Council for Wales 1996).

Thus, the organisational and political environments that National Parks worked within had changed significantly in both character and emphasis. Now other 'public' bodies such as utilities had to have regard to National Park Purposes and National Park Authorities as unitary authorities had more control over their internal management and over planning control. Conservation considerations had a higher profile within the new National Park Authorities but these had to be tempered with a new requirement to foster economic well being. This concept of linking economic development to environmental stewardship closely echoed the tenets of

Agenda 21 and the concepts of environmental sustainability (World Commission on Environment and Development 1987).

The National Park Management Plan is the strategic planning tool for the activities of the National Park Authority. It fulfils many roles including:

- providing strategic objectives and policies for achieving National Park purposes;
- developing a co-ordinated approach to management in the Park;
- integrating the action of different organisations;
- forming the basis from which to bid for funding (Countryside Commission and Countryside Council for Wales 1996).

National Park Management Plans are thereby pivotal to the operation of National Park Authorities. They are considered to be important as much as a process as an end product. It is the production, maintenance and review of National Park Management Plans that requires appropriate information about the nature and characteristics of the Parks, how these are changing over time and what effects management actions have.

The passing of the 1995 Act thus led to a number of changes in National Park Authorities. All National Parks were to become independent bodies, like the Lake District and Peak District National Parks were prior to the 1995 act. As future funding structures changed to yearly bids made through Functional Strategies, Parks would increasingly find themselves accountable. Greater accountability was required by the (then) Department of the Environment and the general public. This change of political environment, in conjunction with perceived needs to justify their existence and effectiveness, led National Parks to consider the development of environmental indicators. Indicators provide not only a means for summarising information about the state or characteristics of a Park but, additionally, were seen as means for providing a monitoring and reporting structure for Park planning and management. In this way, indicators could potentially provide a link between data collection, policy, management and reporting.

3.2.3 Methods

The purpose of this case study was to examine the use of information in a National Park, the information sources available to the Park, and how these were used to inform policy and management. Because National Park Management Plans will become the main process for the creation and use of information a Park that was beginning to develop its Plan was sought. In consultation with the Countryside Commission, the North York Moors National Park (NYMNP) was chosen as a Park that was just beginning to review its National Park Plan and begin development of its National Park Management Plan. This provided an ideal opportunity to study the management planning process and to examine the use of environmental information. The

Countryside Commission and the National Park had the additional objective of wanting to test the use of environmental indicators for policy review as part of the planning process.

The study was initiated with a desk-based examination of the North York Moors National Park Plan and reports supplied by the National Park. This was followed by a series of workshops including staff from the National Park and also from external agencies with related interests. The research proceeded through a number of stages:

1. Review of the objectives from the existing National Park Plan.
2. Objective setting for the new National Park Management Plan.
3. Development of potential indicators for the new National Park Management Plan, including an assessment of the information requirements and data availability.

For each stage a series of workshops was conducted. These were held at Helmsley, in the National Park's headquarters. A timetable for the workshops was established with the National Park's Management Group and staff were allocated to the exercise by the National Park according to their areas of work. The majority of staff participating in each workshop were senior staff concerned with policy and management. The details of the workshops held at each stage varied and are described in further detail below.

1. Review of objectives from the National Park Plan.

At the time of the study (1996) the current Plan for the North York Moors National Park (Statham 1991) was due for review. Due to the change in legislation (1995 Environment Act) the review was to become the starting point for the National Park Management Plan. After discussion with staff from the NYMNP it was decided that the best starting point would be an evaluation of the objectives published in the 1991 plan. These could then be amended and developed to provide a basis for the forthcoming Management Plan. For the purposes of this study, the evaluation of the new objectives would provide a basis from which to formulate possible indicators and assess information requirements for the NYMNP.

The 1991 National Park Plan comprised eighteen chapters, of which fourteen covered features or activities within the Park and others included a number of general objectives and numbered 'priorities' (Appendix 4). These were taken as the starting point for the study. Workshop sessions with groups of staff responsible for each chapter subject area were held with the original intention of producing a working list of possible indicators for each priority in the National Park Management Plan and thence defining information needs. Workshops were held with eight groups of specialist staff to cover the fourteen subject groups. With the possible exception of the *Moorlands* group, it became clear that the exercise would not generate a basis from which to develop indicators. Progress in the workshops was hampered by the nature of the old National Park Plan:

i. The phrasing of the existing plan made it difficult to evaluate and formulate indicators. Each subject chapter in the Plan began with a several objectives. In most cases, these objectives were framed in general terms that made them unsuitable for assessment or measurement.

ii. Each chapter contained a series of numbered statements that are referred to here as 'policies'. In the Plan, these policies were actually a mixture of statements of intent, advocacy, programmes and objectives.

iii. The 'policies' were not clearly related to the objectives stated at the start of the chapter and did not necessarily have clear targets or endpoints. Additionally, few of the statements were time-bound so making it impossible to determine rates of progress.

iv. There was no differentiation between proactive and reactive work.

v. The 1991 Plan contained many priorities or policies that no longer appeared to be supported by current staff within the National Park.

vi. Some areas of work were inherently more easily measurable than others. This was often due to a history of data collection rather than differences in the subject matter.

vii. The picture was further clouded when measures of quality were required. For example, there was good information on the area and extent of woodland across the Park. However, the main issue in woodland management was quality, which is largely determined by 'appropriate management'. Appropriate management is highly context specific. It was not possible to specify what appropriate management was across the whole park.

viii. Other situations which have reasonably tight working definitions were easier to work with. For example:

Loss of Section 3 moorland to inappropriate development or use.

Nevertheless, even an indicator such as this still relies upon agreed definitions and suitable data. This indicator requires that:

Section 3 moorland is well defined and mapped, and that it is a meaningful classification in terms of National Park Objectives,

adequate definitions of inappropriate development and use and data to describe and quantify them are available.

Much more difficult still to measure were subjective evaluations such as a possible indicator for information provision:

Increase in visitor understanding.

ix. In most cases, no forms of baseline or reference data were available.

x. Staff were unfamiliar with the concept of indicators and, perhaps more fundamentally, the Park had no process for the review of objectives or policies.

As a result of this situation an alternative approach was agreed with the National Park Authority. The study was to progress by going through an objective setting exercise before indicators and information needs were considered. This benefited the National Park as the objectives set could form the basis for starting the new National Park Management Plan and was useful for this study as the objectives would provide a more feasible starting point from which to develop indicators and to identify their concomitant information requirements. It also helped to emphasise that an information-based approach to planning and management may involve far-reaching changes in the way in which organisations operate, think, communicate and interact.

2. Objective setting

The purpose of this exercise was to establish a set of objectives for each subject theme to be covered by the National Park Management Plan. In this context objectives fulfil a number of important roles in providing:

i. a focus on priorities for action;

ii. the link between policy and action;

iii. both a conceptual and operational system for managing the Park;

iv. a basis for the development of indicators.

The previous set of workshops had proved to be very time-consuming and the National Park was unable to commit staff across the board to the study. Because of this a further set of workshop sessions were established with three groups of staff from the following subject areas: moorlands, rivers and fresh waters. These three themes were chosen in conjunction with the National Park to outline the special characteristics of the North York Moors National Park and a range of issues representative of park management. As discussed earlier, the new National Park Management Plan diverged from the pattern established by National Park Plans in that it is directed to include other bodies and organisations that have interests in the National Park area, not simply the interests of the National Park Authority as had been the case previously. In addition to the presence of the appropriate National Park staff in each of the three topic groups, staff from the appropriate lead external agencies were invited to the workshop sessions. From the topic areas chosen these were English Nature for moorlands, the National Rivers Authority (now the Environment Agency) for rivers and fresh waters and the Yorkshire and Humberside Tourist Board for visitor attractions, facilities and services.

The review of the National Park Plan established that the Plan had a considerable range of institutional, organisational and political associations. To try to promote fresh consideration of the issues for this phase of the study the Park Plan was conceptually put to one side.

Each workshop session was structured as follows:

i. Scoping:
- definition of key management issues which need to be addressed;
- selection of priorities.

ii. Objective setting:
- specification of management aims and objectives.

For the scoping component each group was prompted and encouraged to list and discuss the key features, characteristics and qualities for their topic within the North York Moors. The products of these sessions are presented in tables 3.11-3.13 in terms of matrices between special qualities or characteristics and themes. The visitor attractions, facilities and services results are presented simply in terms of characteristics and not subdivided as the others were due to the different nature of the subject matter.

Using these matrices as a starting point, the second stage of each session was setting possible objectives for the National Park Management Plan. The draft objectives generated with each group are presented in Appendix 5. These aims and objectives were, by necessity of time constraints, indicative rather than comprehensive. They were further revised and developed by the North York Moors staff for the National Park Management Plan. For the purposes of this study, they provided a basis for the next stage of the work, the development of indicators and their information requirements.

Table 3.11 Rivers and Freshwaters Topic Group: key features.

Special Qualities/ issues	Economy/Society	Landscape	Ecology	Tourism/ Recreation	Culture/ Heritage
Groundwaters	Abstraction – bottling - water supply - irrigation		Springs		
Rivers	Fish farms Salmon Abstraction	River banks Linear woods Naturalness Secret places	River carr Acidity	Angling Paddling Picnics Walking	Water mills
Standing water	Storage reservoirs	Gormire	Ponds	Gormire Reservoirs - sailing, sport	Fish ponds
Marsh/mire		Reedbeds Wet grassland	Mires Bog Marsh Reedbed Wet grassland		
Wetland species			Overwintering birds Otters Crayfish	Duck Salmon	

Table 3.12 Moorland Topic Group: key features.

Special Qualities/ Issues	Economy/Society	Landscape	Ecology	Tourism/ Recreation	Culture/ Heritage
Moorland	Heather Grouse-shooting Sheep	Heather Burning patterns, colour and variation Open (size), wilderness Atmospheric Contrast of high moors and valleys Weather, light Scale (perceived) Bracken Topology	Heather Size Grouse Merlin Golden Plover Invertebrate diversity Structural diversity Large extent of dry heath Mire Grassland Mosaics Unique combination of dry and wet heath Geomorphology	Heather Grouse-shooting Sheep Bird/wildlife watching RoW Open access Walking Mountain biking 4x4 off-roading Open 'wilderness'	Archaeology Legend/myth Common rights Peat-cutting Tradition
Moorland fringe		Contrast with unenclosed moorland	Edge habitats Breeding/feeding areas Woodland		

Table 3.13 Visitor attractions, facilities and services Topic Group: key features.

Feature	Characteristic (or desired characteristic)
People	Warm friendly welcome Local
Services	Quality Customer care
Accommodation	Range Quality Availability
Attractions	Heritage Railway Villages - Coastal - Historic
Commercial services	Arts/crafts Eating Shops
Themes/Associations	Captain Cook
Information	Access/availability Messages Booking etc. services
Events (good for visitor management)	Cultural fairs/shows Markets

3. Indicator development

A further series of workshops was conducted with the three topic groups used in the objective setting exercise. The purpose was to test and demonstrate the use of indicators in the National Park Management Plan process. Where possible indicators were to be produced for each objective derived during the previous exercise together with an assessment of information requirements and lists of possible data sources. In practice, this proved to be a less than a straightforward exercise. Staff had difficulties in developing useful indicators based upon the objectives they had developed. The initial stage of each session therefore included some reconsideration of the objectives. There was detailed discussion in all groups about the relationships between aims and objectives and in particular whether and how these related to different timescales for implementation and monitoring. Another interesting issue was raised relating to remit or jurisdiction. As mentioned earlier, National Park Management Plans differed from the preceding National Park Plans in that they now include other relevant agencies in addition to National Park Authorities. The logical extension of this is that aims and objectives within the new National Park Management Plan should include items that the National Park Authority would like to influence, but over which they have no direct control. If this was the case it posed the question of how they should set targets or measures against objectives which they cannot then directly influence or control. In this case, it would clearly be possible to set desired targets for the National Park but conformance to these would be assessed outside the management structure of the National Park Authority.

Staff participating in the exercise also had problems conceptualising the link between objectives and indicators. This was largely addressed within the workshop sessions by the addition of a further stage - that of listing or defining desirable outcomes for each of the objectives. This process helped staff to make the conceptual link from an objective to a concrete outcome. It was then much easier to begin defining methods or characteristics for appropriate measurement. Another problem relating to indicator selection was one of a perceived lack of baseline or reference data. In many cases Park staff felt they had insufficient knowledge about the state of the Park and the range and magnitude of issues facing it. This is a fundamental issue: it suggests that in the past decision making was not information-led. There is also the further implication that the effectiveness of any indicators produced may be limited by a lack of baseline or reference data with which to evaluate progress or change.

The workshops resulted in the drafting of 47 indicators against objectives. These are presented in Appendix 6 in a table also showing the outcomes, information needs and possible data sources.

3.2.4 Information requirements

The table in Appendix 6 shows the wide range of information requirements for the draft indicators produced for the North York Moors National Park. For example, Table 3.14 shows an indicator produced for each topic group and its related information requirements.

Table 3.14 shows that each indicator has a number of information requirements including in most cases measures of type, location, quantity and quality (as discussed previously in Chapter 2). In practice, the framing of most indicators starts with issues of typology, Table 3.15 illustrates how information needs fit into these data groupings. Thus, the table represents a cross section of information requirements for indicators related to the three chosen topic groups. As Table 3.14 illustrated, typology issues are the most frequent, particularly when they refer to a qualitative aspect of the feature under consideration. For some areas, these can be relatively easily defined by the adoption of agreed standards. For example, vegetation can be described using the NVC (Rodwell 1991a *et seq.*) which provides a working typology for British

Table 3.14 Indicators and associated information requirements.

Topic Group	Indicator	Information requirements
Moorland	Proportion of different NVC vegetation types within semi-natural moorland habitat	Definition of 'semi-natural moorland' Area of semi-natural moorland habitat NVC communities and their extent Defined locations or geographic areas
Visitor attractions, facilities and services	Level of visitors/users satisfaction	Definition of visitor/user Visitor/user satisfaction Number of repeat visits Defined locations or geographic areas
Rivers and freshwaters	% of total length of river system meeting each target quality class 1) chemical 2) biological	Definition of river system Length of total river system within a pre-determined area Target quality classes (RQOs) Chemical and biological river quality scores

plant communities. Taking the first example from Table 3.14, 'semi-natural moorland' is not a universally recognised type. The features that constitute moorland in the North York Moors are in some cases unique to the area. It would be possible to list and describe the NVC vegetation communities present within the moorland but these taken in isolation cannot necessarily describe what is regarded as moorland in the North York Moors. It would therefore not necessarily be sufficient to locate and map all those vegetation communities that are present in moorland landscape types. Issues of typology were discussed in detail in Chapter 2, but examples from two other variables relevant to National Park indicators are shown below in Figure 3.1.

Because indicators are related to purpose, it clearly makes sense to adopt or construct typologies relevant to the field of use of the indicator. However, one of the main functions of indicators is for reporting. To fulfil this role they must employ readily recognisable terms and adopt established classifications and typologies. When established standards are not available, robust and repeatable typologies need to be developed in terms recognisable within the relevant area of expertise. These issues do not appear to have been tackled in a National Park context. However, work in this area has been conducted by the Country Conservation Agencies as part of the National Biodiversity Network (see Chapter 2) and the development and adoption of standards will increase the utility and range of applications for data.

Table 3.15 Grouping North York Moors indicator information requirements.

Data grouping	Moorlands	Rivers and fresh waters	Visitor attractions facilities and services
Type	moorland/open moorland/wild moorland semi-natural moorland suitable restoration natural succession management scheme semi-natural woodland industrial and archaeological features cultural heritage and traditions open access linear access	freshwater wetland associated habitats river system landscape quality key landscape elements suitable habitat (for crayfish) constraints criteria for expansion of native fish populations	visitor/user satisfaction repeat visit accommodation availability of facilities or services range of attractions special events information sources information categories
Location	of all features in 'type' (above) in terms of point, place or mapped boundaries semi-natural moorland suitable restoration areas	of all features in 'type' (above) in terms of point, place or mapped boundaries rivers boreholes sample flow points suitable habitat (for crayfish)	of all features in 'type' (above) in terms of point, place or mapped boundaries accommodation attractions information sources
Quantity	of all features in 'type' (above) e.g. area of open moorland number of management schemes and the area they cover length of linear access area of open access	of all features in 'type' (above) e.g. nitrate, phosphate and pesticide concentrations seasonal flow rates of NVC vegetation classes in habitats numbers and population profiles of native fish species	of all features in 'type' (above) e.g. number of facilities in quality classes number of repeat visits number of information sources number of people receiving information
Quality	desired habitat diversity level of public awareness level of public understanding of rights of way	target quality classes (RQOs) chemical and biological river quality scores	accommodation quality classes booking criteria quality criteria for user information

Figure 3.1 Issues for defining typologies.

1. Woodland

To identify woodlands consistently definitions are required for:

What trees are - e.g. the species, the difference between a shrub and a tree, size.

How many trees make a woodland? When does a group of trees become woodland? What density is required i.e. when do parkland trees become wood-pasture?

The difference between scrub and woodland – classification by size (area), density or species differences or combinations of all these factors?

How to deal with change - i.e. when does scrub become woodland? If a woodland is clear-felled is it still a woodland?

2 Visitors

To identify a National Park visitor definitions are needed for:

Residence and geography - where does the individual live? Can a resident from one part of the NP be classed as a visitor to another part of the Park?

Type of activity - are visitors defined by their activity rather than their geographical origin? Should only leisure visits be included (and what are leisure activities) or should business travel be included as well?

Time - how long does a visitor need to be in a geographical location or conducting a certain type of activity to be included? How should mixed situations be dealt with i.e. a quick sight-seeing trip squeezed into the end of a business trip? How should day visitors be dealt with compared with a holiday visitor who may stay for several days

How does one deal with change - e.g. when people move between destinations and activities or start from different locations.

Repeat visits - should repeat visits by the same individual be counted, and if so, over what time period?

In the examples considered here, many of the definitions required could be met by the adoption of established standards. However, in many cases these would have to be refined by specific criteria of relevance in the local context of the North York Moors. For example, suitable restoration of heather moorland in the North York Moors can differ from suitable restoration of heather moorland on other soil types or at different altitudes. In the same way, landscape quality is tied to local physical and aesthetic character. If local variation is required, it poses the question of how this can be reconciled with the need to conform to international, national or widely adopted standards. These are required to allow comparison between different National Parks and to allow reporting to other standards such as Natura 2000 and the Convention on Biological Diversity. An apparent compromise would be to collect and collate data to 'core' standards which comprise a minimum or standard reporting requirement to meet all 'upward' obligations and to gather regionally or locally specific data to meet the needs of local variation and distinctiveness. One example of this is provided by the EU CORINE land cover data, which uses a standard set of land cover classes for the uppermost three levels of its classification hierarchy, but allows additional, locally defined classes to be used at the fourth level in the hierarchy (Moss and Wyatt 1994).

Information requirements related to location also raise the issue of scale and timing. Considering the requirements in Table 3.15, locations of features such as semi-natural moorland, river systems, landscape quality and visitor facilities are required. The manner by which these are located, whether by position within a geographical unit, central point or mapped boundary, will be related to units and precision required for management and reporting. Reporting may only require the knowledge of how many, or what area of a feature/resource is present within a relatively broad geographic unit, but management may well necessitate knowledge of the precise position. An example of this may taken from Table 3.15. Suitable habitat for freshwater crayfish (*Astacus pallipes*) may be reported in terms of x km of streams within the National Park, but if the information is needed for species re-location or re-introduction programmes precise locations of suitable habitat would be necessary. The practicalities of this example may in fact mirror information required for typology. Coarser scale locational information may be required for reporting obligations than are needed for management activities. In both cases information requirements must therefore be assessed in terms of the greatest precision required for any of the intended applications. The issues discussed above in relation to type and location also apply to measures of quantity that, for generic considerations, have the same characteristics as location. Quality issues are less easily addressed in terms of quantitative criteria because they relate to objectives, and objectives inherently contain associations of value. Drawing from the examples in the North York Moors these are related to criteria such as desired habitat diversity, levels of public awareness, target water quality classes, and the qualities of accommodation and visitor information. Target water quality classes, qualities of accommodation and visitor information can be related to nationally accepted standards. Desired habitat diversity and levels of public awareness do not have any 'ready made' standards of classifications that can be picked up. They are more closely linked to management objectives specific to the North York Moors and to what the National Park Authority is attempting to achieve.

Thus information requirements are wide-ranging and in many cases context- or application-specific. For many features or criteria, established standards can be adopted. For specific management applications, these may need to be added to for use within the Park Authority. Data precision is determined by both management and reporting requirements, but these may not be the same.

3.2.5 Data availability

The examination of information requirements in the previous section shows that carefully targeted typologies are often needed. While these can often be met by adopting established standards, in some Park-specific cases they may have to be developed according to need. Data are required to support that process and to provide information on the other categories of location, quantity and quality. For this study, an attempt was made to examine the data available for the three topic groups. In practice, this proved problematic because only the moorlands topic group were able to provide a list of data sources they used, and an assessment of their quality. Thus, only data relevant to the moorlands topic group are examined in detail.

The fact that the other groups were unable to provide metadata for their subject areas can lead to several conclusions: firstly that they were unwilling to provide the data, secondly that providing metadata would be time-consuming and thirdly that they be unable to assemble the information. Comments from the moorlands group suggested that they found even the compilation of basic meta-information time-consuming, supporting the suggestion that lack of capability, rather than lack of willingness, is the main constraint. If true, this would represent a situation for concern for the National Park.

3.2.6 Data assessment

Data evaluation was only conducted for the moorland topic group because they were the only group to provide data assessments. The data sources used by this group are shown in Table 3.16. The data were based on a variety of sources but relied heavily upon field survey and aerial photography. Staff experienced a variety of problems with the data they used and these are summarised below.

Temporal variation

Park staff found it difficult to characterise and quantify habitat features at different points in time because many of the data sources they relied upon related to different periods. The update and survey repeat periods for aerial photography, the sites and monuments record and the NYMNP Upland Vegetation Survey were all variable, causing problems in estimating and detecting change over time. The Phase 1 survey was last added to in 1989 and is not clear when or whether it was to be updated.

Table 3.16. Possible 'moorlands' data sources available to North York Moors Park staff.

Data Sources	Strengths	Weaknesses
Air photos	Whole Park coverage Large scale (1:10,000)	Update variable Some sets have only partial cover Problems due to seasonality
Section 3 map (1995)	Regular update On-going Whole Park coverage Being digitised Verification procedures Most error known	Difficult to interpret Habitat areas accurate to +/- 10%
Sites and monuments record	Comprehensive On-going	Update frequency variable Computerised database is incomplete and requires extensive revision
North York Moors/EN Phase 1 survey	Virtually whole Park coverage Mapped information at 1:10,000 Some analysis and area measurements Limitation specified	Somewhat out of date (last field season 1989) Paper-based system No quadrat data
Phase 2 survey	Ongoing survey Mapped information at large scales (1:10,000 & 1:2,500) Some data in database Raw data available to check NVC Records sites of high botanical diversity 1993 survey included open moorland sites which contribute to interpretation of 'quality' of moorland	Started in 1991, still going Database not linked to GIS 35% of 1991/92 sites have incomplete records Favours sites of high botanical diversity
NYMNP Upland vegetation survey	Up-to-date Covers all moorland areas in Park Common standards adopted Database available Quadrat data available Vegetation boundaries partly mapped from 1995 air photos	Planned repeat is only partial Exact locations of quadrats not available
Monitoring landscape Change in National Parks	Available in SPANS & MapInfo formats Data moderately good at whole Park level It may be possible to integrate with Phase 1 data to create a base level digitised habitat map	Air photo interpretation for improved pasture, cultivated land and rough grazing is very inaccurate

While sources such as the MLCNP data relate clearly to a discrete period (Countryside Commission 1991b), others such as the Phase 2 survey of the North York Moors stretch over periods of up to six years. Thus, across the Park temporal coverage of specific areas or different data sources was variable.

Format

Some data were available in digital formats and this increased their utility and potential for integration with other sources. However, many of the key data sources, both spatial and attribute, were in paper formats or incomplete digital data sets.

Coverage

Geographic coverage was variable across the Park. The MLCNP data are Park-wide data based upon air photo interpretation and are also comparable with those from all the other National Parks, where the same survey was conducted. The data users within the NYMNP, however, felt that while the data were relatively robust at the whole park level, they displayed unacceptable levels of accuracy for use at more local scales.

Data sources such as aerial photography and the NYMNP Upland Vegetation Survey have only partial geographic coverage.

The coverage in terms of data content was also variable, the Phase 2 survey had only partial thematic coverage in that it focussed on sites of high conservation interest and was also incomplete. Therefore, certain vegetation categories are not well catered for as they appear to 'fall out' of the system. If improved pasture and cultivated land are considered, they were mapped at Phase 1 level at one point in time. They were inaccurately recorded by the MLCNP data and were excluded from the Upland Vegetation Survey and the Phase 2 survey. To some extent, they could be picked up by examination of the available air photos, but these have variable geographical coverage and are not complete for all periods. While these categories are relatively unimportant for nature conservation, it is important that their location, quality and change over time are known for farm management plans and stewardship schemes. Changes in improved status and the area of cultivated land result from changes in the agricultural economy and affect the viability and physical fragmentation of semi-natural habitats.

Classification

Different sources use different methodology both in recording and in reporting. This makes it difficult to relate information from different sources because of the prevalence of non-linear or many-to-many relationships between categories. This is an important issue because the inability to link data from different surveys limited the ability to monitor changes in environmental features and characteristics and compounded limitations in overall data utility introduced by deficiencies of geographic and temporal coverage. Therefore, to draw upon the example used above, if data from the MLCNP and Phase 1 surveys were comparable it may be possible to combine both to more accurately categorise and map improved pasture and cultivated land.

3.3 SUMMARY AND CONCLUSIONS

- There were clear parallels and differences between the SNCOs and the National Parks. Many aspects of information use were common to the two organisations. There were few real differences and these lay in the scale of operation and their different remits.
- The identification of information requirements is dependent upon clear objectives and these were often lacking in the organisations studied.
- In both organisations, staff did not tend to conceptualise their work in terms of objectives and outcomes and had few tangible targets.

- There was evidence that few staff used an information-led approach. In the NYM National Park, staff asked to rationalise management objectives thought their information/knowledge base was insufficient to make informed decisions.
- The previous NYMNP Plan was, by its nature, non-informational. Taking an information-led approach meant that the whole way in which objectives were set, defined, communicated and monitored then changed. This had implications for the way in which staff then were involved in the process, and in the longer term, for their training needs.
- Focusing on the National Park scale highlighted the importance of local/regional distinctiveness in countryside management. In some cases, nationally accepted classifications/definitions did not discriminate characteristics required at the local level.
- There are clearly different information requirements for management and for reporting. In many cases, more detailed information is required for management.
- The concept of habitat 'quality' is needed at different scales by both the SNCOs and the NPs. This is both highly contextual and locally variable and thus has implications for the scalability of local, regional and national data.
- In both organisations, there appeared to be significant deficiencies in information for supporting their activities. These surfaced in terms of thematic content, geographical and temporal coverage.
- Staff did not appear to value data quality and standards highly as positive criteria but often cited the lack of them as deficiencies.
- Land cover based data sources were not often used. Despite their ability to offer accurate and comprehensive area estimates most ecologists tend to ignore them because of (perceived) deficiencies in composition-related data.
- The most common reasons for non-or under-use of data sources were the lack of knowledge or supporting information, and the (perceived) poor quality of data sets.
- Decisions by staff on the potential value of a data source need to be made objectively and on an informed basis. This requires criteria for evaluation against different conservation functions.

3.3.1 Providing habitat information

At a fundamental level, the information needs of the organisations studied displayed a high degree of commonality. At a basic level information is needed on habitat stock and distribution, while other data are needed on habitat change and environmental impacts. Although there is a great and growing demand for information to inform activities there appears to be little effort spent on identifying information needs and planning to meet them. Staff experience a variety of problems with data sources, but in many cases feel ill equipped to critically review their choice and use of available data. Coupled with a lack of perfect knowledge of the data 'market', optimal data may not always be used. On the supply side, there are a variety of data sources, as described in the review in Chapter 2 and as referred to by respondents in the case studies. There are also clearly problems encountered in making optimal use of the available data, whatever its deficiencies. In many cases responses from staff in the different agencies indicated that they had poor knowledge of the range of available data sources and lacked expertise in their critical evaluation.

However, amongst all these issues one factor merits overriding consideration. At a strategic level, basic information on the quantity and distribution of habitats is still scarce, despite the fact that this basic information is needed to support nearly all the core functions of the organisations consulted. The protection of habitats is one of the key business activities of conservation and yet in many cases little is known about the quantity and distribution of this resource – let alone how and where it is changing.

The SNCOs and NPs require both national and regional habitat estimates for policy and reporting and these estimates should be comparable and interoperable. In addition, time-series data are required for the assessment of change. The obvious way to provide this information is by using remote sensing and land cover-based information sources, as these can provide comprehensive wide-area geographic data. However, they are little used. At the GB level, few national data sets are available that provide wide or complete geographic coverage and fewer still are repeated to provide either up-to-date information or time-series data that are required for monitoring change. Of the major data sets (see Table 2.4, Chapter 2) few are current, offer the possibility of imminent or planned repeats or are based on a hierarchical structure that allows ready scalability. Phase 1 survey is unlikely to be repeated again on a wide-scale basis. Phase 2 and NVC survey is very labour intensive and hence expensive. It is used to provide detailed information on specific sites or for specific projects but cannot be generated routinely on a wide-area basis. Woodland is perhaps unique amongst habitats/land cover types. The Census of Woodland and Trees/Woodland Inventory has been undertaken on a periodic basis and provides comprehensive data with complete geographic coverage. The situation for other habitats is very different. Estimates of the occurrence of semi-natural habitats can be provided by habitat inventories. However, these have traditionally been derived from field survey are very costly and to date none have been repeated. Only a few habitats have been covered, meaning that data are not available for many others. Of those habitats that have been covered, many were based upon survey which are up to thirty years old. In the current climate, with the SNCOs reducing the number of field survey contracts, it appears unlikely that new habitat inventories will be produced or existing ones repeated.

This process of elimination leaves one major information source – the Countryside Survey. This appears to offer all the requisite characteristics:

i. GB-wide geographical coverage;
ii. broad thematic coverage – all rural cover types;
iii. a repeat survey programme;
iv. repeatable methodology;
v. a combination of field survey and remote sensing methods;
vi. a hierarchical structure that offers potential scalability;
vii. digital outputs.

Given all these characteristics we might therefore expect Countryside Survey to be widely used in conservation agencies as a basis for habitat stock estimation. However, the review of data sources and both case studies revealed that the Countryside Survey was little used and not highly valued. The comments received indicated that staff felt that Countryside Survey data were not suitable for their needs and did not fulfil their requirements. Largely, because of the apparent scale limitations of these data: they were seen as unable to provide reliable regional or local-level data, at a resolution appropriate to the organisation's work activities. However, there are clearly fundamental gaps in the information required by conservation agencies and the characteristics of Countryside Survey, combined with the signal lack of alternatives, suggest that its utility for habitat stock estimation should be considered further. Therefore, it was decided to investigate in detail the scope for improving habitat estimates from Countryside Survey data on the basis that:

i. there are strategic deficiencies in countryside information at regional and national scales;
ii. geographic inventory-type habitat data are scarce and rarely updated;
iii. Countryside Survey data may be able to provide national habitat estimates as required by the conservation agencies. Their ability to provide regional estimates is unknown;
iv. Countryside Survey data include both field-survey and land cover methods and offer the opportunity for comparing their relative performance for habitat estimation;
v. it would be useful and novel to explore the utility and scope for improvement for Countryside Survey data.

"The study of physiognomic divisions is as much a mathematical and scientific as an aesthetic exercise." Humboldt, 1807.

4 METHODS

4.1 INTRODUCTION

The Countryside Information System (CIS) provides a means of access and display to CS90 and LCMGB data amongst others. These two major data sets represent one of the very few ways users can access GB-wide countryside data. The CIS was intended to provide country and regional-level statistics. In order to assess the contribution that CS90 and LCMGB can make to the estimation of habitat extent and distribution three different habitat types were chosen for this study. To some extent the choice was limited by the availability of alternative or 'reference' data sets needed to test the effectiveness of CS90 and LCMGB estimates. Reference data sets are only available in 'inventory' forms for a restricted number of habitat types - hence one of the reasons behind conducting this research. Additionally, habitat types that represented different types of landscape and environmental variables were required in order to test the ability of CS90 and LCMGB to resolve habitats under different circumstances. The three selected habitats therefore differ in terms of their strength of landscape association (and thus the likely ease of defining their distribution on the basis of extraneous physical geographic characteristics). The three habitat types chosen were:

- saltmarsh – a habitat with grouped and restricted geographical distribution and strong landscape associations;
- lowland heathland – a habitat with dispersed geographical distribution and moderate landscape associations;
- woodland – a relatively ubiquitous habitat with weak landscape associations.

These target habitats were used to test the validity of habitat estimates from CS90 and LCMGB and to assess the ability of CS90 and LCMGB to estimate their spatial distribution, each within different-sized spatial reporting units. Each target habitat is represented by a 'reference' data set that is used as a standard against which to test CS90, derived CS90 and LCMGB estimates. Table 4.1 shows the reference data sets used for each target habitat. These are described in detail in the following sections and these also describe the rationale behind reference data set selection and treatment of data. The assessment and development of CS90 and LCMGB comparison data are described below.

In each case, a range of different estimation methods were used, including simple interpolation (areal weighting) of CS90 and LCMGB data (Deichmann 1996; Fisher and Langford 1995), control zone estimation (Langford and Unwin 1992, Moxley and Allanson 1994), weighted control zone estimation (Flowerdew and Green 1992) and – for some habitats – a combined approach, using both CS90 and LCMGB data, referred to here as 'intelligent weighting'. GIS methods were employed for this purpose, using MapInfo and MapBasic software (MapInfo 1998a,b).

Table 4.1 Target habitats and reference data

TARGET HABITAT	REFERENCE DATA SET	DIGITAL SPATIAL FORMAT
Saltmarsh	NCC Saltmarsh Survey of Great Britain supplemented with OS data	Point – grid references/digitised polygons
Lowland heathland	EN/RSPB Lowland Heathland Inventory	Point – grid references
Woodlands	FC Census of Woodlands and Trees/Ordnance Survey 1:250,000 topographic data, woodland area	Grid - area within km square

4.2 BASE DATA PREPARATION

This section describes the data preparation common to all geographic and habitat data used. Due to the particular characteristics and circumstances presented by the three chosen habitats, there were differences in the detail of the data preparation in each case and these are described in the specific habitat sections.

4.2.1 Data export

The CIS is not a GIS system. It provides a database of CS data and a means for displaying generalised maps and conducting simple overlay analyses. It also offers a means of exporting spatially referenced (but not georeferenced[6]) tabular data. For the analyses required in this work analytical spatial tools were required and CS data were to be combined with external geographic data. Neither capability was present within CIS. For this reason all CS90 and LCMGB data used in these analyses were exported from CIS v. 5.40 as space-separated ASCII files for use in an external GIS. The export output from CIS is in the form of a header and three data columns: easting, northing (SW corner, 1km squares, British National Grid) and data value. Each row represents a 1km square. The data value is a figure in ha representing the value for the chosen data and therefore represents ha per 1km square. CIS export files were opened in Microsoft WordPad™ and the headers were removed; the files were then re-saved and imported into MapInfo Professional™ (V. 4.5, 5 and 5.5) as text files to create attribute data files. The size of the files (for GB 240,222 rows) precluded the use of standard office software such as Microsoft Excel™ which has a limit of c.63,000 rows. Microsoft Access™ database software was in practice found to be too slow with these quantities of data and MapInfo was therefore used for its speed and accuracy. The eastings and northings columns were concatenated in MapInfo to create a unique 1km grid square identifier named 'CIS_Ref' and an

[6] Spatially referenced data – data with (a) spatial attribute(s), georeferenced data is linked to a geographic object.

index was created for this field. The data column was converted from a character field to a floating point decimal and re-named appropriately using the database management tools within MapInfo.

4.2.2 Georeferencing data

The data exported from CIS and processed in MapInfo provided attribute data but no means for geographic analysis. For analysis, vector-based data were required to facilitate combination and querying with other vector-based data sets. To attach the data to geographic vector objects an additional stage was required. The CIS contains a definition of land squares, that is 1 km squares, referenced to the British National Grid, that are predominantly land – as opposed to sea. A representation of these data was required as a basis for the CS data in the project GIS. This was achieved by exporting a Region File for GB from the CIS. The data were imported into MapInfo and converted into a single column named 'CIS_Ref2' which consisted of a space separated concatenation of the British National Grid (BNG) easting and northing values for each row, each row representing 1 square kilometre. This 'CIS_Land' data set comprised 240,222 1 km rows representing 240,222 km^2.

A 1 km-vector grid for Great Britain was generated in MapInfo using a custom-written MapBasic™ program that also generated a unique geographic reference key to match those generated by CIS. The resulting 1 km grid contained 807,081 1 km squares aligned with the BNG and within a bounding rectangle around Great Britain and produced a MapInfo table of c.110MB in size. Such a file is very slow in use and it contained many redundant (c. 550,000 sea only) squares. Only the squares defined as predominantly land by CIS were required, thus a query was written to extract those squares that matched those included in the CIS_Land data set. MapInfo employs Structured Query Language (SQL, an industry-standard data sub-language) for querying data. For geographic analysis, MapInfo adds a number of specific geographic operators to standard SQL. All queries referred to in this study were executed using MapInfo SQL unless otherwise specified. The query resulted in a daughter table containing 240,222 1 km vector squares, aligned with the BNG, and attributed with a unique row identifier – the 'CIS_Ref' as described above.

4.2.3 Spatial querying/aggregation

The purpose of geographic analyses was to test the ability of CS90 data to quantify and locate different target habitats and to investigate whether these estimates could be improved by the use of extraneous geographic data. In assessing the performance of these estimates we are interested in testing the ability of data sets to quantify habitat resources within different sized reporting units and additionally to show us where target habitats may be found. All habitat estimates were therefore tested at three levels of spatial aggregation: 1 km^2, 100km^2 and county level. 1 km^2 was chosen as the smallest geographic reporting unit for published CS data. This scale is also the basic unit employed by the CIS. A 10x10 km square (100km^2) was chosen as the next level of spatial aggregation as it offers an even-sized reporting unit that provides some smoothing of 1km^2 level data with a resolution that is mappable at the GB scale whilst distinguishing geographic patterns of distribution. County-level aggregation is a variable sized unit but was used because many organisations report at these scales and it is therefore important to quantify the reliability of CS and LCMGB data at this level. All three of these resolutions are smaller than those normally considered reliable for CS data (use of county-level aggregates for CS90 data in CIS v.5.40 generate 'unreliable mean' and 'unreliable totals' for many counties).

To perform the necessary queries and aggregations at different scales, additional geographic data were required. No additional framework was required for 1km square analyses as this is the basic unit employed in CS data sets. For reference data different treatments were required to produce 1km square-level data and these are described in the specific habitat sections below. For aggregation within 10x10km squares (100km^2) a regular 10x10km grid, fitting the BNG was required. This was produced by adapting the MapBasic program used to produce the 1km base grid. This produced a 10km-based square vector grid covering Great Britain. Each square was labelled with a unique identifier based upon BNG easting and northing values and a unique integer ID which was indexed and used as the unique identifier in further queries. For county-based analyses it was necessary to identify which county each of the 240,222 1km squares (that represented GB land area) were within. This was achieved by overlaying the 1km-level vector grid with a vector county boundary data set and querying out the membership of the squares in counties using a geographic operator that was mutually exclusive in its operation. The county boundary data set was digitised at a scale of 1:625,000 and based upon local authority county and metropolitan districts as of re-organisation in 1987. The SQL query used selected 1km squares where their centroids were within the area of a county. This (centroid within) operator was used to avoid duplication of values (e.g. using an 'intersects' operator can place a border square in two counties). This query assigned the majority of 1km squares to counties but left several hundred 'orphaned' squares in coastal and estuary areas. These resulted where squares were defined as predominantly land by CIS but not by the county boundary data set - either by virtue of definition or as an artefact of scale. These 'orphaned' squares were manually allocated to counties by zooming in to a high level of magnification in the GIS and assigning 1km squares based on their proximity to county boundaries on a nearest neighbour basis. The end result of this operation was a county field in the 1km square data set that identified the county membership of each of the 240,222 1km squares. This master table was used as the basis for all subsequent county-based queries.

A number of other CIS-derived data sets were used in the analyses. Like all CIS data these represent a data value for the variable in question expressed in ha for each 1km square in GB. These were exported from CIS as ASCII text files and processed in MapInfo in the same way as the CIS land squares, as described above. The data used were OS95GREF.ccf and OS95TOPO.ccf, geographic reference and topographic reference data respectively. Their use is described below and Appendix 7 describes their definitions. In addition to these data major soil group data were purchased from Cranfield University. These were again in a CIS type format and were georeferenced in the same way as the other data.

4.3 SALTMARSH

Saltmarsh habitat is widely distributed throughout the UK and is present along nearly 1700km of the coastline (HMSO 1995b). Saltmarsh communities develop on marine substrates between mean low water spring tides (MLWS) and mean high water spring tides (MHWS). They require a net accumulation of sediment and consequently occur in situations where there is some shelter from wave action. Saltmarsh habitat represents a transition from mudflat and sand dominated lower marsh areas to upper marsh areas that are characterised by pans and creeks. Lower marsh areas are subject to frequent tidal inundation, upper marsh zones are subject to more infrequent and shorter periods of inundation. The composition of saltmarsh communities is greatly determined by the relative tolerance of different plant species to inundation by saline water and tend to follow gradients with respect to this factor. They are dominated by halophytic species, i.e. those able to withstand high salinity and inundation by seawater. British saltmarsh communities have been described by Burd (1989) and Rodwell (2000) who identify five vegetation zones:

i. pioneer marsh;
ii. low-mid marsh;
iii. mid-upper marsh;
iv. drift-line;
v. swamp.

Rodwell (2000) further categorises vegetation communities along phytosociological principles. Saltmarsh was identified in the biodiversity action planning process as a Broad Habitat type and Coastal Saltmarsh was identified as a Key Habitat type (a subset of a Broad Habitat). Saltmarsh is important in conservation terms because it supports both unique and important assemblages of flora and fauna (Davidson *et al.* 1991). Saltmarsh habitat is threatened by a number of factors. 'Coastal squeeze', is the loss of habitat area through development or agricultural reclamation on the landward side and the erosion of the lower saltmarsh edge through sea-level rise. Saltmarshes are also affected by changes to nutrient budgets, pollution incidents, interference with sediment budgets, grazing management and invasion by the non-native cordgrass (*Spartina anglica*), (HMSO 1995b).

4.3.1 Reference data

Burd (1989) produced an inventory of saltmarsh in Great Britain for the Nature Conservancy Council (NCC) based upon new and existing field survey (see Section 2.3.2, Chapter 2). Despite being completed over ten years ago this remains the only resource inventory with national coverage. Burd estimated a national saltmarsh total of c. 44,000 ha. For the purposes of the inventory saltmarsh habitat was defined to include the plant communities described by the (then) draft NVC. These occurred within tidal limits but excluded *Zostera* and *Ruppia* communities. The definition used also excluded inland saltmarshes, splash zones on cliffs, and all cliff, dune and shingle communities, but did include transitions to some of these habitats (Burd 1989). Table 4.2 shows the relationship between the Saltmarsh Survey and NVC communities.

As with all classification systems it is relatively easy to define core habitat types but transitional areas present further problems. Table 4.2 shows that the Saltmarsh Survey of Great Britain included some plant communities that by NVC standards were not strictly saltmarsh, notably the swamp communities and transitional communities. The detail of this nomenclature is important when making comparisons with other data and is revisited below when discussing equivalent CS90 data sets.

The published inventory was a printed report containing data tables and dot distribution maps. Two copies of county-based inventories were also published. Contact with the current (1998) coastal ecologists for English Nature, Scottish Natural Heritage and Countryside Council for Wales respectively, confirmed that the saltmarsh site maps had never been digitised and that in fact no-one was sure that a complete national set was still in existence; it was thought that some maps may be held in regional offices (Radley, G, Duncan, K and Hill, M. Pers Comm., 1998). It was beyond the scope of this project to attempt to locate paper copies of the original maps in all the regional offices of the country agencies and to digitise them. The value of doing so was also open to question because of the lack of a topographic base on field sketch maps. Therefore, an alternative approach was devised. Tabular data were entered from Saltmarsh Survey of Great Britain (Burd 1989) into a spreadsheet in Microsoft Excel. The site code, centroid alpha-numeric grid reference, site name, administrative unit (county), and site area were recorded. The input data were thoroughly checked against the original for errors and site area totals were checked with those in Burd's report to check for input errors. The alpha-numeric grid references were converted into separate x and y columns for eastings and northings respectively using 'text to columns' and 'search-replace' functions in Microsoft Excel. The data were imported into MapInfo and points were generated from the x/y grid reference columns by use of the 'Table>Create points' routine. This data set therefore captured the tabular data created by Burd representing 557 saltmarsh sites and is henceforth referred to as the NCC Saltmarsh Inventory.

Table 4.2 Relationship between the Saltmarsh Survey and NVC communities.

Community type	Saltmarsh Survey community		NVC communities
Pioneer	1.	*Spartina*	SM 4 *Spartina maritima* community SM 5 *Spartina alterniflora* community SM 6 *Spartina anglica* saltmarsh
	2a.	*Salicornia/Suaeda*	SM 7 *Arthrocnemum perenne* community SM 8 Annual *Salicornia* saltmarsh SM 9 *Suaeda maritima* saltmarsh
	2b.	*Aster*	SM11 *Aster tripolium* var. *discoideus* saltmarsh SM12 *Aster tripolium* (rayed) stands
Low-mid marsh	3a.	*Puccinellia*	SM10 Transitional low marsh vegetation SM13 *Puccinellia maritima* saltmarsh SM13a *Puccinellia maritima* subcommunity
	3b.	*Halimione*	SM14 *Halimione portulacoides* saltmarsh SM14a *Halimione portulacoides* subcommunity SM14b *Juncus maritimus* subcommunity SM14c *Puccinellia maritima* subcommunity
Mid-upper marsh	4a.	*Limonium /Armeria*	SM13 *Puccinellia maritima* saltmarsh
	4b.	*Puccinellia/Festuca*	SM13c *Limonium vulgare-Armeria maritima* subcommunity SM13 *Puccinellia maritima* saltmarsh SM13b *Glaux maritima* subcommunity SM13c *Limonium vulgare-Armeria maritima* subcommunity SM13d *Plantago maritima-Armeria maritima* subcommunity SM13e turf fucoid subcommunity
	4c.	*Juncus gerardii*	SM16 *Festuca rubra* saltmarsh SM16d *Festuca rubra* subcommunity SM17 *Artemisia maritima* saltmarsh SM16 *Festuca rubra* saltmarsh SM16a *Puccinellia maritima* subcommunity SM16b *Juncus gerardi* subcommunity SM16c *Festuca rubra-Glaux maritima* subcommunity SM16e *Leontodon autumnalis* subcommunity SM16f *Carex flacca* subcommunity
	4d.	*Juncus maritimus*	SM15 *Juncus maritimus-Triglochin maritima* saltmarsh SM18 *Juncus maritimus* saltmarsh SM18b *Oenanthe lachenalii* subcommunity SM18c *Festuca arundinacea* subcommunity
Drift line communities	5a.	*Agropyron* (*Elymus*)	SM24 *Elymus pycnanthus* saltmarsh SM28 *Elymus repens* saltmarsh
	5b.	*Suaeda fruticosa*	SM25 *Suaeda vera* saltmarsh SM25a *Elymus pycnanthus* subcommunity SM25b *Halimione portulacoides* subcommunity
Upper marsh swamps	6.	Upper marsh swamps	S4 *Phragmites australis* S19 *Eleocharis palustris* S20 *Scirpus lacustris* ssp. *tabernaemontani* S21 *Scirpus maritimus*
Transition communities	7i.	Shingle/dune transition	SM21 *Suaeda vera-Limonium binervosum* saltmarsh SM21a typical subcommunity SM21b *Frankenia laevis* subcommunity SM22 *Halimione portulacoides-Frankenia laevis* saltmarsh
	7ii.	Freshwater transition	MG11 *Festuca rubra-Agrostis stolonifera-Potentilla answerina* grassland MG11a *Lolium perenne* subcommunity MG11b *Atriplex hastata* subcommunity MG11c *Honkenya peploides* subcommunity
	7iii.	Grassland transition	MG12 *Festuca arundinacea* grassland MG12a *Lolium perenne-Holcus lanatus* subcommunity MG12b *Oenanthe lachenalii* subcommunity

Converting the NCC Saltmarsh Inventory into a 1km square-based data set still presented problems. The available data comprised only an estimate of the surface area and a point location, representing the centroid. For small areas of habitat, that were unlikely extend beyond the 1km square in which the centroid fell, the whole area of the habitat patch could be assigned with some confidence to the 1km square. For larger areas, however, a methodology was required to estimate the extent of the habitat patch. For a linear habitat, such as saltmarsh, approximating the extent by means of a circle, centred on the centroid, was clearly inappropriate. Consideration was given to assigning the area of the habitat patch to the coastal squares only – i.e. assuming that the saltmarsh extended no further than 1 km inland, and ran parallel to the coast, in each direction from the centroid. After examination of Ordnance Survey maps, however, this was thought to be an unreliable generalisation, particularly in estuarine situations and for areas behind shingle spits. Therefore there appeared to be no clear way of assigning habitat areas through a simple rule-based approach. Consequently a two-stage approach to assigning saltmarsh area values to individual 1km squares was developed. This was based upon assigning values below a selected threshold to a single square and digitising external geographic data to use for assignment across multiple squares.

1. Below a certain area it should be relatively safe to assume that if a site centroid falls within a km square all that sites saltmarsh area could be attributed to that square. It was decided to set this threshold at 50ha, which represents half the area of a 1km square. This threshold was based upon the following factors:

    i. given the limitations of the available map data, digitising units less than 50ha in area would be likely to introduce larger errors than those produced by assuming a <50ha site was within a single 1km square;

    ii. for small unit areas, translation of mapping symbolisation would be likely to introduce larger errors than those produced by assuming a <50ha site was within a single 1km square;

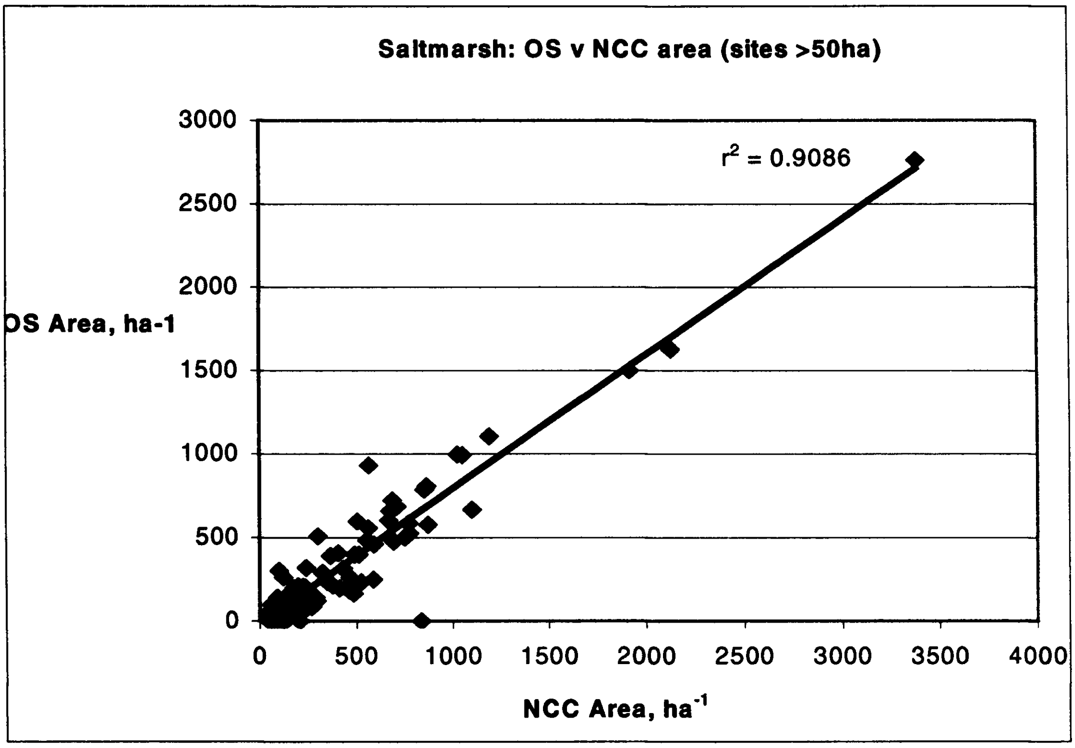
    iii. further analyses of the output data were to be conducted using 1km squares as the smallest unit - therefore any errors introduced by assigning up to 50ha saltmarsh to a particular 1km square would be small in relative terms.

2. Habitat areas for sites over 50ha in area were assigned to adjacent km squares (from given site centroid) by referral to external geographic information digitised from Ordnance Survey maps.

The NCC Saltmarsh Inventory was queried in MapInfo to divide it into two sets, set one containing sites of less than 50ha in area (n=434) and set two containing sites greater or equal to 50ha in area (n=123). Both these data sets were mapped in MapInfo and overlaid with the GB coastline to provide context. With reference to these data OS 1:50,000 Landranger maps were obtained for the entire GB coastline. For each site in set two the centroid grid reference was used to locate the site on the appropriate Ordnance Survey map. Saltmarsh areas as mapped by the Ordnance survey were identified and traced onto drafting film along with a minimum of eight widely spaced control points. Polygons were drawn freehand to produce a minimum bounding area around 'saltings' and 'marsh' symbol areas in the target locations suggested by reference to the NCC Saltmarsh Inventory and the potential site area suggested by this. This resulted in a set of tracings for Ordnance Survey polygons coded by presumed site name and Ordnance Survey mapped type (saltings or marsh). These sheets were digitised on a tablet digitiser using MapInfo and coded with the presumed site name and mapped type attributes. In all it took over twenty-five person days to capture 840 polygons with this method. The area of each Ordnance Survey site was calculated in MapInfo and each site was then examined in MapInfo in conjunction with the NCC Saltmarsh Inventory to equate NCC Saltmarsh Inventory sites over 50ha to Ordnance Survey digitised areas.

Before these data could be used for further analysis they were checked against the original NCC Saltmarsh Inventory data. Figure 4.1 shows a graph of NCC Saltmarsh Inventory area plotted against Ordnance Survey area. This produced an r^2 value of 0.909 showing a very strong relationship between the two data sets. However, the total area of saltmarsh within the Ordnance Survey-based data set (31,850.74ha) was lower than that of the NCC data (Saltmarsh Survey of Great Britain) derived data (40,285.98ha). The difference in area (8,435.24ha) was due to the fact that the Ordnance Survey areas were generally smaller than those identified by the Saltmarsh Survey of Great Britain. Although these figures represent a net loss of saltmarsh area for the reference data set they also represent the only possible way of constructing a spatial reference data set. Examination of the relationship between the two estimates (Figure 4.1) shows that no systematic error is being introduced. Moreover, in many respects the absolute totals of different habitat estimates to be investigated by this research are less important than accurate estimation of presence or absence or relative area. Greater emphasis should be also be placed upon success of location and geographic patterns rather than absolute totals because differences in nomenclature mean that in many cases slightly different 'habitats' are being compared. However, in interpreting the results, it should not be forgotten that the combined NCC/OS reference data set represents an underestimate compared to the raw NCC data.

Figure 4.1 Relationship between Ordnance Survey derived saltmarsh areas and NCC Saltmarsh Inventory derived areas.



4.3.2 CS90 data

The CS90 data were prepared by export from the CIS and import into MapInfo as described in the general methods above. Selection of which data to query from CIS was relatively straightforward but further interpretation was clouded by nomenclature issues. The CS 1990 field survey reporting classes include one category for saltmarsh:

> **8.2.1 Saltmarsh**
> Intertidal sand, silt or mud based habitats colonised by halophytic grasses (e.g. *Puccinellia* spp. and *Spartina* spp.), rushes (e.g. *Juncus gerardii* and *Blysmus rufus*) and herbs (e.g. *Aster tripolium* and *Salicornia dolichostachya*). Applies to all flowering plant communities which are submerged by high tides at some part of the annual cycle.
>
> Source: CIS v. 5.40, CS1990a.csf, Countryside Survey 1990 Wednesday July 19th 1995, 16.23

By using the Land Use Classifications facility within CIS (LUCID) it was possible to derive a list of NVC communities that are equivalent to the CS 8.2.1. These are presented in Table 4.3 against the NVC communities represented in the NCC Saltmarsh Survey of Great Britain saltmarsh definitions (Burd 1989).

Table 4.3 Relationship between the NCC Saltmarsh Survey, CS90 and LCMGB saltmarsh by NVC community.

NVC Communities	**NCC SMS**	**CS90 8.2.1**	**LCMGB D Saltmarsh**
SM03 *Eleocharis parvula* community		✓	✓
SM04 *Spartina maritima* community	✓	✓	✓
SM05 *Spartina alterniflora* community	✓	✓	✓
SM06 *Spartina anglica* saltmarsh	✓	✓	✓
SM07 *Arthrocnemum perenne* community	✓	✓	✓
SM08 Annual *Salicornia* saltmarsh	✓	✓	✓
SM09 *Suaeda maritima* saltmarsh	✓	✓	✓
SM10 Transitional low marsh vegetation	✓	✓	✓
SM11 *Aster tripolium* var. *discoideus* saltmarsh	✓	✓	✓
SM12 *Aster tripolium* (rayed) stands	✓		
SM13 *Puccinellia maritima* saltmarsh	✓	✓	✓
SM14 *Halimione portulacoides* saltmarsh	✓	✓	✓
SM15 *Juncus maritimus-Triglochin maritima* saltmarsh	✓	✓	✓
SM16 *Festuca rubra* saltmarsh	✓	✓	✓
SM17 *Artemisia maritima* saltmarsh	✓	✓	✓
SM18 *Juncus maritimus* saltmarsh	✓	✓	✓
SM19 *Blysmus rufus* saltmarsh		✓	✓
SM20 *Eleocharis uniglumis* saltmarsh		✓	✓
SM21 *Suaeda vera-Limonium binervosum* saltmarsh	✓	✓	✓
SM22 *Halimione portulacoides-Frankenia laevis* saltmarsh	✓	✓	✓
SM23 *Spergularia marina-Puccinellia distans* saltmarsh		✓	✓
SM24 *Elymus pycnanthus* saltmarsh	✓	✓	✓
SM25 *Suaeda vera* saltmarsh	✓	✓	✓
SM26 *Inula crithmoides* saltmarsh		✓	✓
SM28 *Elymus repens* saltmarsh	✓	✓	✓
Non-Saltmarsh communities (transition areas)			
MG11 *Festuca rubra-Agrostis stolonifera-Potentilla answerina* grassland	✓		
MG12 *Festuca arundinacea* grassland	✓		
S19 *Eleocharis palustris*	✓		
S20 *Scirpus lacustris* sp. *tabernaemontani*	✓		
S21 *Scirpus maritimus*	✓		
S4 *Phragmites australis*	✓		

It can be seen that the CS saltmarsh definition excludes some communities included by the NCC Saltmarsh Survey (SM12, MG11 and 12, S4, 20 and 21) and includes others not included by the NCC Saltmarsh Survey (SM3, 19, 20, 23 and 26). Therefore the definitions do not exactly correspond but have a large area of overlap. SM12 rayed *Aster tripolium,* occurs on marshes with a strong freshwater influence (Rodwell 2000) and is very restricted in its abundance and geographical distribution. The swamp and mesotrophic grassland categories included in the NCC Saltmarsh Survey represent transitional communities. After careful examination of the CS90 reporting categories, it was clear that it would not be possible to add elements from the CS90 data to represent these transitional areas because mesotrophic grassland communities are represented by a large and ubiquitous category (permanent non-intensive grassland) the inclusion of which would introduce errors far greater than those caused by the possible loss of small transitional areas from the reference data. Swamp community S4 is not represented within CS90 reporting categories and S20 and S21 are not differentiated, but are simply represented within the 'wetland' category.

Five saltmarsh communities were not included by the NCC Saltmarsh Survey but were included by the CS90 reporting category. SM3 *Eleocharis parvula* community is very rare, with only four records in Britain. SM19 *Blysmus rufus* saltmarsh has a restricted north-western distribution from north Wales to north-west Scotland. It generally occurs in small stands. SM20 *Eleocharis uniglumis* saltmarsh has a similar geographic distribution to SM19 and is described by Rodwell (2000) as:

> *"a widespread minor constituent of a variety of damp transitional communities along the upper marsh fringes".*

SM23 *Spergularia marina-Puccinellia distans* saltmarsh is generally found in the upper marsh areas of coastal marshes, but also in inland saline areas. The coastal distribution is widespread but scarce. SM26 *Inula crithmoides* has a restricted distribution in saltmarsh communities, occurring from Essex to Hampshire where it tends to occur in high marsh areas.

Without access to detailed spatial vegetation data it is not possible to quantify the effect that these nomenclature differences may have upon results. The saltmarsh community descriptions suggest that the communities omitted from the respective definitions are generally peripheral to the major and abundant saltmarsh communities. The possible error introduced by the differences in nomenclature are likely to be very small when conducting analyses at the 1km square level and negligible at larger scales such as the 10km square level.

4.3.3 LCMGB data

The LCMGB data were prepared by export from the CIS and import into MapInfo as described in the general methods above. Selection of which data to query from CIS was relatively straightforward but further interpretation was again clouded by nomenclature issues. The LCMGB data included one category for saltmarsh:

> **ITE Landcover Saltmarsh**
>
> Saltmarshes are intertidal sand-, silt- or mud-based habitats, colonised by halophytic grasses such as *Puccinelia* spp., and herbs such as *Limonium* spp., *Aster tripolium* and *Triglochin maritima*. They remain mostly green in winter. For the purposes of this classmap, only those marshes up to normal high water spring tides (i.e. those flooded monthly) are included. The upper saltmarsh, inundated only on extreme high-water spring tides, is dominated by coarse grasses such as *Agropyron* spp. These are classified accordingly as marsh / rough grass.
>
> Areas of seaweeds are sometimes sufficiently extensive to show as vegetated intertidal plant communities. The may comprise the green alga *Enteromorpha intestinalis* or the brown wracks (*Pelvetia caniliculata*, *Fucus* spp. and *Ascophyllum nodosum*) growing on rocks, boulders and sometimes gravels, sands and muds.
>
> Distinction of this cover type is dependent on the level of the tide on the days of imaging (the lower tide being used to define the lower limit of the seaweed beds or saltmarshes). Thus discrepancies can arise where high tides prevailed on imaging.
>
> Source: CIS v. 5.40, LCMGB2.csf, Land Cover Map (Revised) version 2 Wednesday July 17th 1996, 13.37

The correspondence between LCMGB and the NCC Saltmarsh Survey by NVC category is shown in Table 4.3. It can be seen that the LCGB Saltmarsh definition corresponds exactly (in NVC categories) with the CS90 8.2.1 Saltmarsh reporting category: therefore the correspondence with the NCC Saltmarsh Survey is the same as that provided by the CS90 data. The S4 community is not represented in LCMGB reporting categories while S20 and S21 are present within a much larger category of 'Fen and Marsh'. SM12 is not represented and MG11 and 12 are present, as with CS90, in a large grassland category – 'Lowland grass, non-sown grass >25%'. Therefore the same principles apply to LCMGB as to CS90 and no further attempt was made to add to the LCMBG saltmarsh category to make it more inclusive of transitional communities. Referring to the definition above it can be seen that the LCMGB saltmarsh category data may include some areas of seaweed not distinguished from other intertidal plant communities. These would not normally be included in any saltmarsh definitions as saltmarsh but are present within the original satellite imagery due to probable similarity in signature and discrepancies in tide states between different scenes.

4.4 LOWLAND HEATHLAND

Lowland heathland habitat is defined both in terms of its constituent plant communities and also by landscape-related characteristics. It is dominated by ericoid sub-shrubs such as heather (*Calluna vulgaris*), bell heather (*Erica cinerea*), cross-leaved heath (*Erica tetralix*) and gorse (*Ulex* spp.) but typically includes a range of habitats and habitat transitions to, and including, acid grassland mosaics, scrub, woodland, valley mires, open water and wet heath. These habitats tend to be associated with neutral to acid mineral soils, often characterised by nutrient leaching and podsol development. In England lowland heathland tends to occur at altitudes of less than 300m (Gimingham 1972).

Lowland heathland supports a range of characteristic plants and animals, some of which are rare and many of which are in decline (Michael 1994). The rare species present on heathlands include marsh gentian (*Gentiana pneumonanthe*), marsh clubmoss (*Lycopodiella holub*), Dorset heath (*Erica ciliaris*), sand lizard (*Lacerta agilis*), smooth snake (*Coronella austriaca*), silver-studded blue butterfly (*Plebejus argus*), raft spider (*Dolmedes fimbriatus*), small red damselfly (*Ceriagrion tenellum*), dartford warbler (*Sylvia undata*), woodlark (*Lullula arborea*) and nightjar (*Caprimulgus europaeus*). Overall, heathlands support an abundance of Red Data Book plants and animals and numerous species afforded the highest priority for protection with the UK Biodiversity Action Plan (HMSO 1995b).

Lowland heathland is now a restricted and threatened habitat; it was once more extensive in England than it is today. Indeed, only one sixth of the lowland heathland area present in 1800 AD now remains (Farrell 1989, HMSO 1995b). Around 32,000ha of lowland heathland remains in England (Michael 1994) with the major concentrations in the south-west (Cornwall, Dorset and Devon) and the south (Hampshire and Surrey). Threats to lowland heathland include the decline of traditional management, habitat fragmentation and disturbance and, to a lesser extent, agricultural improvement. Elevated nitrogen levels are also thought to be causing progressive replacement of *Calluna* with grassland species (Marrs 1993). In the recent past heathland losses were mainly due to urban and industrial development, agriculture and forestry. The UK and England in particular (with c.55% of the UK's resource [Michael 1994]) is internationally important for lowland heathland. This fact has been recognised by designation of heathland communities under the Bern Convention and European Community's Species and Habitats Directive.

4.4.1 Reference data

The development of a lowland heathland reference data set again illustrated the heterogeneous nature of environmental data. The English Nature/RSPB Lowland Heathland Inventory (LHI), (EN/RSPB 1994/5/6) was the only available data set that focussed specifically on lowland heathland. This data set covers England only; consequently the analyses conducted here were restricted to England, rather than GB as with the other target habitats. The LHI was essentially a desk-based inventory taking information from existing field survey and best available knowledge. It was produced with the aim of providing lowland heathland data to help target conservation management schemes including those under the Countryside Stewardship and Environmentally Sensitive Areas schemes. The LHI attempts to map all sites greater than 0.5ha in area that are known to contain areas of dry or wet heath. The LHI defined heathland vegetation loosely as:

> " *vegetation dominated by ericoids, dwarf gorse species or achidophilous lichens growing on generally nutrient-poor mineral soils. This includes dry, wet, chalk and lichen heaths.* " (EN/RSPB 1994/5/6).

The LHI did not specifically include areas of former heathland suitable for restoration and generally did not include sites of less than 0.5ha except where they were of local importance (N. Michael, pers. comm. 1998). Heathland areas above 250m altitude were not included except where larger sites below this level extended above 250m altitude. The LHI is referred to as a 'provisional' data resource as it is subject to periodic (but non-systematic) updating. Table 4.4 shows the Phase 1 Habitat codes included in the LHI.

Table 4.4 Phase 1 habitat coding and inventory coding in the EN/RSPB LHI (EN/RSPB 1994/5/6).

Phase 1 code	Inventory code
D1.1 (Dry heath)	DH (Dry heath)
D5 (Dry heath/acid grassland mosaic)	DH (Dry heath)
H6.6 (Dune heath)	DH (Dry heath)
H8.6 (Maritime/seacliff heath)	DH (Dry heath)
D2 (Wet heath)	WH (Wet heath)
D6 (Wet heath/ acid grassland mosaic)	WH (Wet heath)
D1.2 (Chalk heath)	CH (Chalk heath)
D3 (Lichen heath)	LH (Lichen heath)

The LHI data were supplied by Dr Nick Michael, English Nature's Heathland Ecologist in May 1998; there was no version coding available for the data set. The data were tabular in form and included attributes for site code number, site name, centroid grid reference, total site area, heathland type, status and date of survey. The column containing data on heathland type and area was a free-text field containing abbreviations for heath type (i.e. dry heath, wet heath) and followed by a numeric figure representing the area of that type in ha. The database had not been created using standard terms or in a relational structure. The result of this was that area-based terms occurred at different (and random) points in the text string. Spelling was inconsistent.

This 'format' precluded the use of standard data cleaning techniques and meant that the data were not amenable to querying. In order to create the reference data set these free-text fields were cleaned, parsed and aggregated using predominantly manual techniques and mostly on a row by row basis. Interpretation and cleaning of unstructured data in this manner is liable to produce some errors but is necessary to change the data into a format amenable to analysis. In a number of cases no area figures for heathland vegetation were given and the presented areas in the LHI therefore related to 'site' areas (mostly mixed habitat SSSIs) containing an undisclosed area of heathland vegetation. Because this matrix may contain other habitat types such as woodland, fen and acid grassland it was decided to use only those sites with specified areas of heathland vegetation. It should be noted that the derived totals would therefore represent an underestimate which it was not possible to quanitify. The methods employed for deriving heathland type areas were not specified in the LHI (EN/RSPB 1994/5/6) so it was not possible to quantify the effect of measurement methods on possible totals. However, known heathland areas were described for most sites and these were added to create a total heathland area column to include all recorded heathland types.

The grid references provided were for site centroids. For the purposes of analysis a 1 km square lowland heathland data set was required. The heathland area could not simply be attributed to the grid reference point for the site centroid because, in many cases, the heathland area would be greater than 1 km^2. In the absence of digital vegetation data it was impossible to assign appropriate heathland areas to adjacent 1 km squares. In the absence of any external reference data it was decided to assume that sites were circular in shape and to buffer the points with a variably sized buffer that was equal in area to the area of heathland indicated by the data. These were linked to the site centroids to produce a vector-based data set that could be intersected with the GB land area 1 km square vector data to produce figures of lowland heathland vegetation per km square. This data set was then overlaid with the 1 km square data set, in the project GIS, and the area of overlap with each buffer area was calculated. In reality heathland sites will not be circular, but the errors introduced at a site level should even out statistically for the whole population. It should also be noted that the centroids refer to a 'site' – normally a SSSI that will contain, to a greater or lesser extent, some heathland. Therefore the location of the site centroid will refer to the centroid of the heathland vegetation with differing (and unknown) levels of accuracy. The only way to improve this situation would be by use of digital vector site boundaries and these do not exist. The county-based LHI includes paper maps indicating site boundaries, but these are SSSI designated boundaries for SSSIs containing heathland habitat, not maps of heathland vegetation. No other national data set was available that could indicate lowland heathland vegetation boundaries. In the light of these factors it was thought that the methods employed were likely to produce the most accurate estimate possible in the circumstances.

4.4.2 CS90 data

The CS90 data were prepared by export from the CIS and import into MapInfo as described in the general methods above. Selection of which data to query from CIS was not straightforward because CS90 does not provide a single lowland heathland category. The box below shows the landcover/use categories pertaining to lowland heathland. Assessment of these categories (Table 4.5) against the Phase 1 categories included in the LHI (EN/RSPB 1994/5/6) suggested that categories 4.1.1, 4.1.2 and 4.1.3 should be aggregated to produce a 'heathland' data set for comparison and analysis.

4.0 HEATHLAND AND BOG

4.1 HEATHLAND

Land dominated by (>25% cover) dwarf shrubs. Dominant shrub species are invariably *Calluna* or *Vaccinium*. Heathland is traditionally divided by context into lowland types, usually characterised by dry soils, and moorland, often on peat substrates.

4.1.1 Dense Heath

Heath with >75% cover of *Calluna* and/or *Erica* in upland and lowland settings. Includes dune heath which occurs on consolidated and flattened dunes.

4.1.2 Open-Canopy Heath

Heath with 25-75% cover of *Calluna* and/or *Erica*, in a mosaic with grassy herbaceous vegetation, whether in upland or lowland settings. Includes lowland wet heath, where the ericoid element is high.

4.1.3 Berry-Bush Heath

Heath with >25% cover of *Vaccinium* + *Empetrum* + *Arctostaphylos* and <25% cover of *Calluna* + *Erica*. Includes non-alpine berry-bush heath, alpine and sub-alpine heaths, comprising *Arctostaphylos alpinus* heath, *Loiseluria* heath and other sub-alpine heaths.

Source: Countryside Survey 1990 Field Survey Reporting Classes – Land cover/use categories and definitions. CIS v. 5.40 Reference information, CS1990a.csf

Table 4.5 Phase 1 lowland heathland habitat categories in the EN/RSPB LHI and their equivalents in CS90 and LCMGB (EN/RSPB 1994/5/6 and CIS v. 5.40 LUCID 1998).

Phase 1 code	CS90	LCMGB
D1.1 (Dry heath)	3.6.2 Non-Molinia moorland and mountain grass	E Rough pasture/dune grass/grass moor
	4.1.1 Dense heath (*Calluna/Erica* >75%)	
	4.1.2 Open-canopy heath (*Calluna/Erica* 25-75%)	H Grass/shrub heath
	4.1.3 Berry-bush heath	I Shrub heath
D5 (Dry heath/acid grassland mosaic)	4.1.2 Open-canopy heath (*Calluna/Erica* 25-75%)	H Grass/shrub heath
H6.6 (Dune heath)	4.1.1 Dense heath (*Calluna/Erica* >75%)	H Grass/shrub heath
	4.1.2 Open-canopy heath (*Calluna/Erica* 25-75%)	I Shrub heath
H8.6 (Maritime/seacliff heath)	4.1.1 Dense heath (Calluna/Erica >75%)	H Grass/shrub heath
	4.1.2 Open-canopy heath (*Calluna/Erica* 25-75%)	I Shrub heath
D2 (Wet heath)	4.1.1 Dense heath (*Calluna/Erica* >75%)	H Grass/shrub heath
	4.1.2 Open-canopy heath (*Calluna/Erica* 25-75%)	I Shrub heath
	4.1.3 Berry-bush heath	
D6 (Wet heath/acid grassland mosaic)	4.1.2 Open-canopy heath (*Calluna/Erica* 25-75%)	H Grass/shrub heath
D1.2 (Chalk heath)	4.1.1 Dense heath (*Calluna/Erica* >75%)	H Grass/shrub heath
	4.1.2 Open-canopy heath (*Calluna/Erica* 25-75%)	I Shrub heath
D3 (Lichen heath)	3.6.2 Non-*Molinia* moorland and mountain grass	E Rough pasture/dune grass/grass moor

The CS90 category '3.6.2 Non-*Molinia* moorland and mountain grass' is described (CIS 5.40, CS90A.csf 1995) as:

> *"Coarse unimproved upland grass in a moorland setting, usually on soil with peaty top. Dominated by* Nardus, Deschampsia flexuosa *and* Juncus squarrosus. *Includes alpine and sub-alpine grasslands and* Racomitrium *heath."*

This definition describes upland blanket-bog type rough grassland dominant by grazing resistant coarse grass and rush species and was therefore not included in the combined CS90 data set because it is only relevant to upland areas. The combined CS90 'heathland' data set was thought to represent an over-estimate of lowland heathland because the constituent categories include both lowland and upland heath categories. Because the categories were so broad there was no clear way of separating the lowland and upland components. This issue was addressed during the next stage of data treatment where additional data were used to filter the vegetation data.

4.4.3 LCMGB data

Similarly to CS90, the LCMGB did not possess a single land cover category equivalent to lowland heathland. Additionally, the LCMGB categories do not correspond with the CS90 reporting categories as the LCMGB used a different (17 class) reporting structure. Suitable LCMGB land cover categories were therefore chosen by referring to online data set details provided within CIS (CIS 5.40 1998) and by reference to LUCID (CIS 5.40 1998). These categories are shown in Table 4.5 in terms of their correspondence with Phase 1 habitat categories. CS90 and LCMGB categories (indicated to be relevant by reference to Phase 1 categories) were compared using LUCID (within CIS v. 5.40) and were found to be consistent.

It was decided to combine LCMGB categories H and I (Grass/shrub heath and Shrub heath respectively) to form a compound heathland data set. LCMGB Category E (Rough pasture/dune grass/grass moor) was omitted because it appears to have been introduced mainly as a correspondence to lichen heath and this is a very scarce habitat. It did not appear prudent to introduce a large, and largely non-applicable, land cover category (E Rough pasture/dune grass/grass moor, 438,700ha) to try and include components of lichen heath habitat when this category represented only 161.65 ha (<0.006% of the LHI heathland total). Despite the exclusion of possible lichen heath the resulting heathland data set was thought to represent a large over-estimate of lowland heathland vegetation. This was because the LCMGB categories include upland heathland types in addition to lowland types. The compound 'heathland' data set was created within CIS and exported for processing within MapInfo according to the standard procedures described above.

4.5 WOODLAND

Woodlands are one of the more widely known and studied vegetation types in GB. Notwithstanding this, GB is one of the least wooded countries in Europe (HMSO 1995b). Woodlands were chosen for analysis because they represented a relatively ubiquitous habitat with weak landscape associations. Although woodland is widespread across the UK, distribution and size distribution varies greatly (Peterken 1981). The vast majority of native woodland is composed of broadleaf species with the notable exception of important yew (*Taxus baccata*) woodlands, native pinewoods (*Pinus sylvestris*) and juniper scrub (*Juniperus communis*). Woodland data are routinely collected under the census of woodland and trees conducted by the Forestry Commission (Locke 1987). For nature conservation purposes woodlands are classified into ancient and recent categories and a provisional Ancient Woodland Inventory is maintained by English Nature (Spencer and Kirby 1992). Ancient Woodlands are those where the land is believed to have been continuously wooded since 1600 AD (Rackham 1990). Because Ancient Woodlands are effectively a land use category (based upon continuity of land use) they are not studied explicitly in this work.

Woodlands are mentioned in various BAPs and Habitat Statements (HSs). These include: Upland Oakwood HAP, Native Pine Woodlands HAP, Broadleaved and Yew Woodland HS, Planted Coniferous Woodland HS, Native Pine Woodland HS, and the Lowland Wood Pastures and Parkland HS. Woodlands are highly represented because of the high number of species associated with them (HMSO 1995b).

4.5.1 Reference data

Of all possible habitats, woodland had the greatest choice for reference data sets; however, as was the case for the other habitats, the assembly of an accurate spatial data set for Great Britain was problematic. The Forestry Commission (FC) has recorded data through the Census of Woodlands and Trees (Locke 1987) and since 1994 through the National Inventory of Woodland and Trees (NIWT, Wright 1998). The NIWT will not be completed until 2001 (Wright 1998) and GB-wide data were not available when required for this study (Justin Gilbert pers. comm., 1997). For this reason, data from the Census of Woodlands and Trees (CWT) were used in this study (Locke 1987). These data, however, are only published in a collated form, aggregated to County level, and no detailed spatial data were available. It was therefore decided to use the FC data as a benchmark against which to test other candidate spatial data sets. Two candidate data sets were considered potentially suitable. Bartholomew's publish digital spatial data through the academic CHEST agreement; this includes a 1:250,000 topographic data layer for woodland based upon irregular polygons and with a minimum mapped unit of 2ha. The Ordnance Survey publishes geographical reference data within CIS (OS95GREF.ccf), and this includes a woodland data set giving area of woodland within 1km squares derived from map-based data (see Figure 4.2).

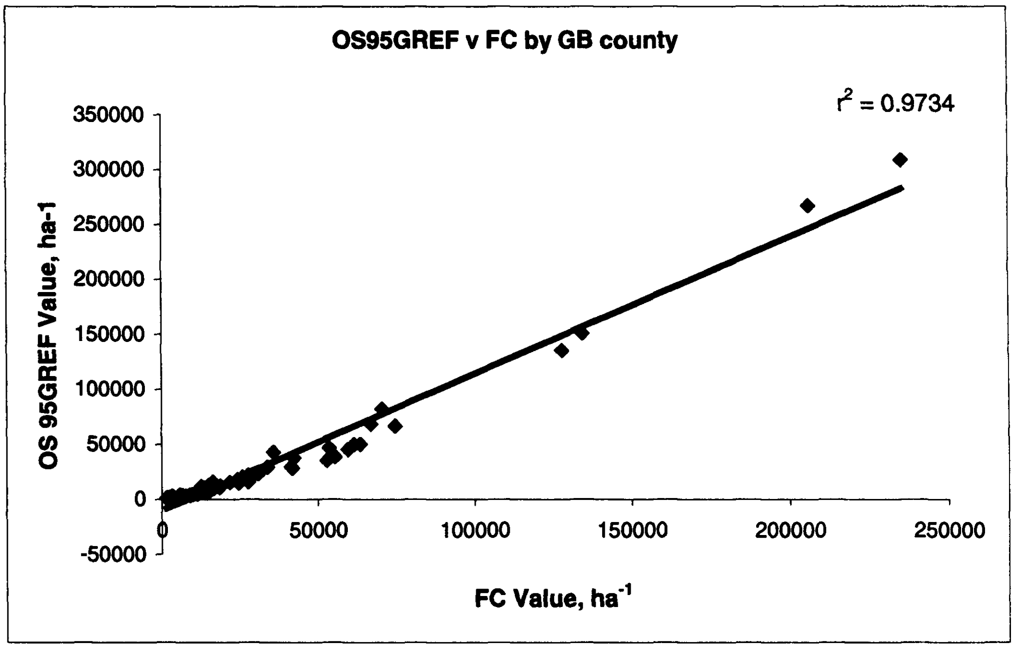
To assess their relative suitability both data sets were imported into the project GIS. The Bartholomew's data were converted from Arc E00 format into MapInfo using MapInfo's ArcLink software and an area overlay analysis was conducted to produce a figure for woodland area by county.

Figure 4.2 OSGREF95.ccf data set details.

> Geographic reference dataset created from the Ordnance Survey's 1995 digital 1:250,000 *Strategi* dataset by classifying the data each one kilometre square into 13 categories (one for each dataset) as a percentage, to a resolution of 0.01.
>
> Each kilometre square totals 100% with any residual areas being added to the Open Countryside category. The thirteen feature categories were extracted from vector data, each category allocated a colour and for linear features a width value. Each layer was then rasterised and the total number of each colour pixel outputted as a percentage of each kilometre square.
>
> Source: Data set details CIS v. 5.40 Reference information, OS95GREF.ccf

The OS data were exported from CIS and imported into MapInfo via the standard method described earlier. These data were joined to the CIS Land-based data set (which included a county attribute for each 1 km square) and county totals were calculated using a sum clause within an SQL query. The resulting county totals from each data set were exported to MS Excel and correlated with FC county-based data. The FC county-based data were based upon the figures published by Locke (1987). These included six woodland categories: Mainly coniferous high forest, Mainly broadleaved high forest, Coppice with standards, Coppice, Scrub and Cleared. These data were entered into MapInfo and figures were double-checked visually and by comparing county and category totals to remove any input errors. The definition of cleared woodland is highly temporal and subject to much variation in its definition (Sinclair 1992, Locke 1987). Additionally, recently felled woodland can be recorded in terms of the remaining vegetation during field survey or incorrectly classified from satellite data. Reference to the LUCID land use classifications program revealed that the LCMGB includes the FC felled woodland category within its 'G Marsh/Rough grassland' category. Therefore a decision was made to exclude the cleared category from the combined FC data set because this was likely to introduce error when comparisons were made with other data sets. The combined FC woodland data set was therefore made up from the summation of the five FC categories described above, excluding the 'Cleared' category. In both cases the correlations between the candidate spatial data sets and the FC data were very high (FC vs. Bartholomew's - 0.979, FC vs. OS – 0.973, see Figure 4.3). The decision was taken to use the OS-based woodland data set for further analysis because more information (metadata) was available about the provenance of the data set and because these data are included within the CIS and are therefore more widely available to a range of users.

Figure 4.3 Relationship between OS95GREF woodland data and FC data by GB counties (Locke 1987).



4.5.2 CS90 data

The preparation of a CS90 woodland data set was relatively straightforward because CS90 identifies clear woodland categories. However, as in the other cases, the nomenclature differs between that employed by the FC and the reporting categories used for CS90 field survey.

Table 4.6 Woodland type categories used by the Forestry Commission and their CS90 and LCMGB equivalents. CIS v. 5.40 LUCID 1998).

FC woodland type	CS90	LCMGB
Mainly coniferous high forest	5.1.1 Coniferous woodland 5.1.2 Mixed woodland >20% of each	K Deciduous/mixed wood L Coniferous/evergreen wood
Mainly broadleaved high forest	5.1.2 Mixed woodland >20% of each 5.1.3 Broadleaved woodland	K Deciduous/mixed wood L Coniferous/evergreen wood
Coppice with standards	5.1.2 Mixed woodland >20% of each 5.1.3 Broadleaved woodland	K Deciduous/mixed wood
Coppice	5.1.2 Mixed woodland >20% of each 5.1.3 Broadleaved woodland	K Deciduous/mixed wood
Scrub	5.1.1 Coniferous woodland 5.1.2 Mixed woodland >20% of each 5.1.3 Broadleaved woodland 5.3 Shrub	K Deciduous/mixed wood L Coniferous/evergreen wood
Cleared	5.4 Felled woodland (regrowth <1m high)	G Marsh/rough grass

As discussed in the reference data section above, it was decided to omit the FC 'Cleared' category from the reference data set to avoid the introduction of non-relevant data in the test data sets. Table 4.6 shows that developing a combined CS90 woodland data set was relatively straightforward categorically. The combined CS90 woodland data set was therefore produced by combining three woodland reporting categories (5.1.1, 5.1.2, 5.1.3) and shrub (5.3). The

Felled woodland category (5.4) was omitted. The data were aggregated within CIS and then prepared for export to MapInfo as described in the general methodology above.

4.5.3 LCMGB data

As with the CS90 data, woodland types raised no immediate nomenclature issues. Table 4.6 shows the two relevant LCMGB categories used to build a combined LCMGB woodland data set. Category G Marsh grass was omitted as discussed above. The data were aggregated within CIS and then prepared for export to MapInfo as described in the general methodology above.

4.6 DATA TREATMENT

The steps so far described in this chapter led to the development of spatial reference data sets for each target habitat and as far as practicable equivalent data sets based upon CS90 and LCMGB data. At this point CS90 and LCMGB data had only been filtered by categorical relevance. This appeared to be relatively close for saltmarsh and woodland but relatively coarse for lowland heathland, the test data sets including upland heathland categories.

As discussed earlier, one of the main problems users of CS90 (and to a lesser extent, LCMGB data) have is the production of county-level or smaller regional resource estimates. One of the objectives of this work was to look for means to improve the reporting of CIS data at these scales. Therefore it was decided to make use of additional data in an attempt to improve the habitat location and quantity estimates provided by CS90 data. This was done through three stages by the use of control zones and weighted control zones to adjust the geographical distribution of data estimates within ITE land classes. Control zones were to work at a relatively coarse level by introducing a 'filter' where no habitat would normally be expected. Weighted control zones were to introduce external data for use as weighting factors within the previously restricted control zone target area. A third technique, henceforth referred to as 'intelligent' weighting, was developed to re-distribute CS90 data based upon LCMGB mapped quantities and locations.

4.6.1 Control zones

Areal interpolation using control zones, also known as dasymetric mapping (Wright 1936), is a well-established technique, although as Langford *et al.* (1990) note, it has not been widely used. Areal weighting uses the known data from a source zone to construct estimates for another zone – the target zone. The inherent assumption in areal weighting is intrinsically simplistic: namely that data within the zones are uniformly distributed. Where the zones can be assumed not to be homogenous, interpolation can be aided by the use of control zones (Flowerdew and Green 1989). Control zones can be based upon a variety of sources including: remote sensing data (Langford and Unwin 1992) digital land use maps (Moxey and Allanson 1994) or *a priori* knowledge (Goodchild 1993).

The advantages of areal interpolation are that it is relatively easy to compute and the data requirements are low. However, the disadvantage of the technique lies in the rather simplistic assumption of uniform distribution of data in the target zone. Areal interpolation using control zones has the advantages of recognising non-uniformity in the data, allowing the input of prior knowledge and can provide greatly improved accuracy. However, areal interpolation using control zones still relies upon a relatively simplistic model and it is often difficult to find suitable data for their use. Because of the large data volumes in this study and the relative lack of suitable spatial data it was felt that areal interpolation with control zones offered a suitable method for the first stage of modelling.

The development of control zones as a basis for attempting to improve CS90 habitat estimates is based upon the rationale underlying CS90 data interpolation. As discussed in Chapter 5, CS90 field survey data are based upon fieldwork conducted in 508 km squares in GB. These squares were located according to a stratification derived from ITE's land class classification. For each vegetation reporting category, the mean area of vegetation recorded within a land class was calculated and the assumption made that this figure is representative of the land class in its entirety. There are two related problems with this approach: the total vegetation area figures generated and the accuracy of vegetation distribution. Area totals are thought to work well at a national level, but for smaller areas the standard error becomes very large and means are often unreliable. Distribution mapping is also problematic; while the area totals generated by multiplying the sample-based means appear to be relatively accurate at the land class or national level, presence/absence/quantity for a particular km square is not likely to be accurate because every square in a land class receives the same (low) total.

The rationale behind the development of control zones was to re-allocate land-class based totals within the same land class but differentially across member km squares. Control zones therefore limit the candidate locations (km squares) where habitat may be present. By using control zones, some km squares would be allocated no habitat total and others would receive a greater than land class average total. However, the sum total for an entire land class would be the same as that reported by field survey.

To be effective, the variables used as controls in control zone estimation need to be good covariates of the target variable. Ideally, they should relate to conditions that explicitly constrain or determine the distribution of the target habitat. Because of the variable nature of the target habitats, different control zone variables were used in this study. By its very definition, saltmarsh is overwhelmingly coastal in distribution. An artefact of the ITE land class interpolation model is that some land classes cover coastal and inland areas. Where saltmarsh was recorded by field survey within these land classes, each square in that land class was attributed with a saltmarsh area value. Thus, when CS90 saltmarsh data are mapped, many inland squares are shown to possess saltmarsh when clearly this is not the case. To address this issue, the saltmarsh control zone was based upon buffering the coastline. A 5km wide buffer was chosen to define the control zone. The buffer radius was set at 5km from a coastal centreline to be wide enough to include large estuarial areas of saltmarsh, to allow for generalisation errors introduced by using a 1:625,000 scale coastal boundary and by using 1km square level data.

Lowland heathland presented a more complex problem. The nomenclature accounts provided by LUCID and CIS show that the combined heathland category developed for this work included upland heathland types. Because the nomenclature employed was not fine grained enough to filter out upland vegetation types, external data were required to develop the control zone to eliminate these upland areas. The first choice for this purpose would appear to be the use of altitude data: the reference data used a classification that defined lowland heathland as occurring up to 250m altitude, although it also included sites that extended above this limit. The rationale behind this research was to develop test criteria from first principles rather than from analysing the available reference data and using these results to optimise the fit to the test data. Consequently, the altitude limit needed to be derived not from the reference data *per se*, but from independent evidence or logic. In this context, it is important to recognise that altitude itself does not provide a direct constraint on the distribution of lowland heath; rather, it represents a proxy for the more complex effects of climate and land use. In GB, there is clear east/west biogeographic variation (Barr 1997) which only loosely relates to altitude: vegetation types of a distinctly upland character can be found at low altitudes, or even at sea level, in western areas (Rodwell 1991 *et seq.*). In this study, therefore, a control zone variable was sought which better reflected this biogeographic pattern.

As discussed earlier, the ITE land class system stratifies the country by physical environmental characteristics. The 32 land classes are additionally categorised into Arable, Pastural, Marginal upland and Upland landscapes and thereby offer a more sophisticated means for categorising upland and lowland landscapes. The control zone was therefore based upon ITE lowland land classes that comprised land classes 1-16 and 25-26. This approach was also employed by the ITE for Key Habitats work conducted for DETR (Barr 1997) for the development of a lowland heath 'mask'.

The third control zone to be developed was for woodland. Unlike the other target habitats there were no obvious control zone variables in this case, for woodland habitats are almost ubiquitous across different landscape types and environmental conditions. The only consistent constraints are provided by standing water and, perhaps, the altitudinal treeline. The area of standing water does not provide an especially useful control variable at the scale of analysis used here, for it affects a relatively small area of GB. The altitudinal treeline is potentially more effective. Evans (1984) describes the altitudinal distribution of upland woods: oakwoods (*Quercus sp.*) occur up to an altitude of 400m, and birchwoods (*Betula sp.*) up to altitudes of 600m. Scots Pine (*Pinus sylvestris* occurs at altitudes of up to 670m (Clapham *et al.* 1989) but contiguous wooded areas rarely reach this limit. On the basis of these figures it was decided to set the altitude limit at 600m. Because the geographic data used (CIS V. 5.40, OS95GREF.ccf) were based upon a mean for a km square there was scope for local variation about the 600m limit and this was felt to represent a relatively conservative limit.

The modelling of control zone variables was conducted within the project GIS. As previously described a master database was established for each target habitat, including the km square reference, reference and test data area values and ITE land class. For each target habitat, geographic reference information was introduced as appropriate by importing data into MapInfo and adding them to the master database using an SQL join. For each target habitat master database, an integer column was added and labelled 'CZone' to record data indicating membership of the control zone category. In the case of saltmarsh a 5km-radius buffer was created around the GB coastline (based upon vector 1:625,000 data). This buffer (representing the control zone) was then intersected with the saltmarsh spatial database (based upon 1km squares) by means of an SQL query employing 'centroid within' criteria. The master database was then updated using the MapInfo Table>Update column function to modify the CZone column to equal a value of 1 where the centroid of the squares were within the buffer. By this means the membership of every GB km square (240,222) was attribute coded according to inclusion or exclusion of the coastal control zone.

The lowland heathland database was coded in a similar fashion. The lowland land classes were selected by SQL query and the CZone column was updated using the MapInfo Table>Update column function to attribute membership of the lowland control zone. The same method was employed for woodland, based upon altitude data for each square. The control zone comprised all squares with a mean altitude of <600m.

Once the membership of control zones had been coded for each database the CS90 data were re-assigned within the control zones according to land class. The same methods were employed for saltmarsh and woodland, but a slightly different technique was used for lowland heathland. For saltmarsh and woodland the rationale was to re-assign the same land class total, because this was based upon means derived from field survey and therefore could not be changed without knowledge of the precise locations and habitat areas measured for each sample square. Therefore, the same land class total was to be re-assigned across a smaller population of km squares, based upon membership of the control zone within a given land class. To achieve this the total area of CS90 reported habitat by each land class (all 32) was calculated in MapInfo by use of an SQL query. The total number of squares present (count) within each land class but also within the control zone was also calculated with a SQL query. The mean control zone value, i.e. the new CS90 control value was then calculated according to the following formula:

Control zone value calculation:

$$CZV = \frac{a}{n}$$

Where: CZV = Control Zone Value
a = Land class total (sum of habitat area)
n = Number of cases in each land class that are within the control zone

The resulting CZV values were added to new columns labelled 'CS90CZ' in the master databases using MapInfo's Table>Update column function. The CS90CZ data therefore had the same total sum as the original CS90 data and (by definition) the same land class totals. However, there were no values (zero) for non-control zone squares and consequently the data values for the member squares within the control zones were higher than for the same square in the original CS90 data.

The approach described above was modified for lowland heathland where the CS90 and LCMGB test data were known to include unsuitable vegetation types (upland heath categories) and therefore represented a large overestimate for lowland heathland habitat. In recognition of this the control zone (lowland land classes) was used to filter the test data and to eliminate upland areas from the test data. In the lowland heathland database the control zone value column was updated with a SQL query where control zone squares (lowland land classes) returned the raw CS90 or LCGB value and non-control zone squares returned 0 for their CS90 or LCGB values.

4.6.2 Weighted control zones

The control zones described in the previous section represented a relatively crude filter that divided the GB land area (England only for lowland heathland) into two type areas: those that were expected to contain habitat and those that were not. However, logic suggests that it should be possible to refine the control zone according to the probability of finding habitat in a particular km square. If the CS90 data could be further refined in this way they should represent more accurate regional totals and a more reliable distribution.

Various techniques have been developed to model the distribution of a spatial target variables in this way (see Flowerdew and Green 1994). One approach is to subdivide the study area into subzones, using one or more control variables, and then use maximum likelihood methods to compute weights for each subzone, reflecting the probability of occurrence of the target variable. Flowerdew and Green (1994) describe the use of this methodology to model populations of ethnic minorities and house prices. This approach has several advantages, not least that it can be conducted reiteratively in order to derive best estimates of the weights. A major disadvantage in this context, however, is the heavy processing demands of maximum likelihood methods, especially for large and complex data sets such as those considered here.

Alternatively, weights can be deduced for predictor variables (covariates) using regression methods – for example, by analysing the association between the predictor variables and the target variable using a 'training' data set, such as data from a sample area. This approach was used by Briggs and Gulliver (2000) to model distribution of Gross Domestic Product (GDP) at a small-area (Nomenclature des Unites Territoriales Statistiques 5) (NUTS) level across the Belgium-French border. In this case, weights were derived by modelling associations between GDP and the control variable (population) at a more aggregated level (NUTS 3). This approach has the advantage that it uses available data to derive the weights; a disadvantage, however, is that it assumes that the weights derived from the training area are valid for the whole study area. A third approach is to define weights for the predictor variable(s) deterministically – e.g. on the basis of prior experience or from 'first principles'. This was the approach adopted for testing here.

The variables employed for this purpose again need to represent good covariates of the target habitat. They also needed to be based upon available digital spatial data with GB coverage. It was decided to use CIS-based data where possible, as these are the data widely available to other users. Soils, climatic and land use data were all possible candidates for use as weighting variables. Soils were chosen as weighting variables in each treatment because of the potentially strong associations with habitat distribution and because GB-level digital data were available. Climatic variation affects habitat distribution mainly at a coarse scale. Habitats such as lowland heathland vary in composition and floristics across GB, but because the available reference and test data were relatively coarse, and did not differentiate such subtleties, it was not felt that climatic data would provide a strong discriminatory variable. Land use data would have undoubtedly been a useful determinant. However, no suitable GB-wide digital data were available. Therefore three variables were used here: soils data (for all three habitats), altitude data (for saltmarsh and lowland heathland) and urban area and slope angle (for woodland) (Table 4.9).

Reference to the Environmental Catalogue V.2 (W.S. Atkins 1996) showed the availability of soils data originated by Cranfield University and available in a CIS format. Contact with Cranfield (I. Bradley 1997 pers. comm.) revealed the data were available and a licence for the dominant soils data was purchased. The data could not be imported into CIS because they were not correctly formatted, but were imported into MapInfo after cleaning in MS Wordpad. These data related to 10 major soil groups (Table 4.7) based upon a generalisation of Soil Survey 1:250,000 data (Hudson *et al.* 1982, Jarvis *et al.* 1983, 1984, MLURI 1982, 1984).

Table 4.7 Major soil groups in GB (after Jarvis *et al.* 1984).

Name	Description
1. Terrestrial raw soils	Mineral soils in very recently formed material with no soil horizons other than superficial organic or organo-mineral layer less than 5 cm thick. The unit is used to describe bare rock, cliffs and scree slopes in Scotland.
2. Raw gley soils	Mineral soils in material that has remained waterlogged since deposition. Prominently mottled or greyish above 40 cm depth. Mainly confined to intertidal flats and saltings that represent stages in the development of mature salt marshes. They are found in fresh salt marsh and in fresh river deposits.
3. Lithomorphic soils	Shallow soils with a distinct, humose or peaty topsoil but no subsurface horizons more than 5 cm thick. Normally over bedrock, very stony rock rubble or little altered soft unconsolidated deposits within 30cm depth. They are most extensive in the chalk downs and on hill crests or mountains in the uplands.
4. Pelosols	Non-alluvial clayey soils that crack deeply in dry seasons, but are slowly permeable when wet. They have a coarse blocky or prismatic structure and no prominently mottled non-calcareous subsurface horizons within 40 cm depth. They are most extensive in eastern England in chalky boulder clay and in the south Midlands in pre-Quaternary clays.
5. Brown soils	Soils with dominantly brownish or reddish subsoils and no prominent mottling or greyish colours (gleying) above 40 cm depth. They are developed mainly on permeable materials at elevations below about 300 m O.D. Most are in agricultural use. These loamy or sandy well-drained soils are found extensively across most of lowland Great Britain.
6. Podzolic soils	Soils with black, dark brown or ochreous humus and iron-enriched subsoils formed as a result of acid weathering conditions. Under natural or semi-natural vegetation, they have an unincorporated acid organic layer at the surface. Such soils typify upland areas of south west England, Wales, the Lake District, the Southern Uplands and Highlands. They are found in smaller area on existing or lowland heaths in England.
7. Surface-water gley soils	Non-alluvial, seasonally waterlogged, permeable soils formed above 3 m O.D. and prominently mottled above 40 cm depth. They have no relatively permeable material starting within and extending below 1 m of the surface. These soils are typical of the wetter areas in the lowland and upland fringes of England, especially in the northern. They dominate the Clyde and Forth valleys.
8. Ground-water gley soils	Seasonally waterlogged soils affected by a shallow fluctuating groundwater-table. They are developed mainly within or over permeable material and have prominently mottled or greyish coloured horizons within 40cm depth. Most occupy low-lying or depressional sites. The most extensive area is the Wash in eastern England. The remaining areas are much smaller in extent and very often linear in shape along river valleys.
9. Man-made soils	Soils with man-made topsoil or a disturbed subsurface layer (containing disturbed fragments of soil horizons) to at least 40cm depth. They result from the addition of earth containing manures, or the restoration of soil material after mining or quarrying. The main areas covered by this map unit are in midland and northern England associated with former opencast and ironstone workings.
10. Peat soils	Soils with more than 40 cm of organic material in the upper 80 cm or with more than 30 cm of organic material over bedrock or very stony rock rubble. These soils are primarily found in upland Scotland particularly in the west. They are also extensive in the Pennines. Lowland peat soils are found in the Fens, Somerset, Lancashire and Yorkshire.
14. Lakes and inland water	(No other description)
15. Bare rock	(No other description)
16. Pixels dominated by sea	(No other description)

By reference to the soil descriptions shown in Table 4.7 and by reference to habitat accounts (Rodwell 1991a *et seq.*) each major soil group was coded for each target habitat. The coding was based upon attributing a soil score value between 1 and 0. Soils were scored as 1 where the soil class concerned was thought to be highly favourable for the habitat in question and 0 where it was thought that the soil class was likely to exclude the habitat. Intermediate cases were scored 0.2, 0.5 or 0.6 respectively according to the strength of association anticipated (Table 4.8): no attempt was made to apply a continuous scaling of the soil variable, because of the imprecise nature of the process, and the heterogeneous and generalised nature of the mapped soil units. Heterogeneity derives from the way in which the soils data were compiled. The major soil group data were originally based upon 1:250,000 mapped data, themselves based upon sample data. The sampling intensity for these original data varied considerably: some observations were based upon sample farms, others on transects related to relief or known underlying geology. The original minimum mapping unit was 0.5km^2 in England and Wales and 0.75-1.0km^2 in Scotland, from which 296 soil associations in England and Wales and 109 associations in Scotland were built. The digital data were derived by scanning the resulting 1:250,000 maps at the equivalent of 100m resolution and the Scottish data were then re-classified to match the English and Welsh data. The 100m pixels were aggregated to a 1km resolution. At each stage of the data observation, capture and processing, errors or logical inconsistencies may have been introduced. The resulting dominant soil groups data were given a single soil code for each 1km square. This means that the code is likely to represent the major soil group in that square but not the exclusive soil type. Associations between soil type and habitat are therefore likely to be confounded by local variations in soil conditions, within the 1 km squares. These effects are likely to be especially important for habitats influenced by local soil conditions (e.g. chalk heath).

Table 4.8 Target habitats and major soil group scoring for weighted control zones.

Description	Saltmarsh Score	Lowland heathland Score	Woodland Score
1. Terrestrial raw soils	0	0	1
2. Raw gley soils	1	0	1
3. Lithomorphic soils	0	0.2	1
4. Pelosols	0	0.2	1
5. Brown soils	0	0.6	1
6. Podzolic soils	0	1	1
7. Surface-water gley soils	0	0.2	1
8. Ground-water gley soils	0.6	0.6	1
9. Man-made soils	0	0	1
10. Peat soils	0.6	1	1
14. Lakes and inland water	0.2	0	0.5
15. Bare rock	0	0	0
16. Dominated by sea	1	0.2	0

The major soil group scoring values were added to the master habitat databases in MapInfo by querying major soil group types from each database and updating a new soil score column with the values indicated in Table 4.8.

Major soil group constituted one of the constituents for the establishment of weighted control zones. In addition other data were sought. The geographic reference data (OS95GREF) supplied with CIS included data for average slope and average altitude and these were employed for the target habitats as indicated previously and summarised in Table 4.9.

Table 4.9 Target habitats and additional weighted control zone variables.

Target habitat	Control Zone	Weighting variables
Saltmarsh	5km buffer	Control Zone, soil score, altitude score
Lowland heathland	Lowland land classes	Control Zone, soil score, altitude score
Woodland	Average altitude below 600m AOD	Control Zone, soil score, slope score, urban area score

Saltmarsh

The saltmarsh control zone was based upon a 5km-radius buffer from the coastline. The weighting variables employed added a soil score as described above and an altitude score. The altitude score was based upon likelihood of saltmarsh habitat occurring within chosen altitude ranges. The altitude data were based upon mean altitude within 1km squares. The threshold for the occurrence of saltmarsh was set at 100m AOD. It was thought unlikely that saltmarsh would occur at higher average altitudes than 100m but that it was not possible to differentiate classes for altitudes less than 100m because saltmarsh occurs in tidal estuarial situations and high marsh areas are often significantly higher than low marsh areas (Burd 1989).

Table 4.10 Weighted control zones: altitude scoring for saltmarsh.

Altitude/m	Score
<100 m AOD	1
>= 100 m AOD	0

The saltmarsh control zone squares were selected by SQL query within MapInfo from the saltmarsh master database. A new column for altitude score was created and updated using the Table>Update column function. Altitudes over 100m were coded as 1 and altitudes below 100m were coded as 0. A new column 'CellWeight' was created. This column was calculated by multiplying the row values for the soil score and altitude score to produce the 'control zone weighting factor'. This cell-weighting factor was used to calculate a novel CS90 cell mean value by land class using the same general methods for all target habitats, as shown in Figure 4.4.

Lowland heathland

Lowland heathland weighted control zones were developed in the same way as the saltmarsh zones, using the same weighting variables (Table 4.9) but different scoring factors (Table 4.11) to represent the different habitat associations. The upper altitude cut-off point was set at 500m AOD to include sites starting below 300m in the EN/RSPB definition. The 300-500m AOD class was given a low probability value (0.2) because few sites were likely to be present in this class.

Figure 4.4 Calculation of weighted control zone values.

1. Calculate cell weight (CW) where CW = $CZS.WS_1.WS_2.(WS_3$ for lowland heathland and woodland).
2. Calculate the sum of cell weights for each land class - $\sum CW_{L1..L2..Ln}$
3. Calculate the sum of 'raw' CS90 values for each land class - $\sum CS90_{L1..L2..Ln}$
4. For each land class calculate cell mean (CM) where $CM = \frac{\sum CS90_{L1..L2..Ln}}{\sum CW_{L1..L2..Ln}}$
5. Calculate Weighted control zone value (WCZ) where WCZ = CM.CW

Where:
CW = Cell Weight
$L_{1..2..n}$ = Land class 1,2,3...32
CZS = Control Zone Score (Presence/absence)
$WS_{1..2..3}$ = Weighting score 1, 2 or 3 (soil, altitude, slope, urban)
CS90 = CS90 target habitat value areas ha^{-1}
CM = Cell Mean (Intermediate step in SQL calculation)
WCZ = Weighted control Zone Value – the re-assigned CS90 derived value for target habitat area ha^{-1} per land class.

Because the reference data for lowland heathland included a cut-off of 300m in its definition, land squares occurring within the <300m interval were given a probability score of 1. These weights were used to calculate WCZ values for lowland heathland in the master database using a SQL query, according to the general procedure described above and in Figure 4.4.

Table 4.11 Weighted control zones: altitude scoring for lowland heathland.

Altitude/m	Score
>=500m AOD	0
>=300m to <500m AOD	0.2
<300m	1

Woodland

Woodland weighted control zones were developed in the same way as the other target habitats but using different weighting variables (Table 4.9). Because altitude had been used to develop the control zone for woodland it could not be used as a weighting variable. The OS95GREF topographic data available from CIS also included average slope data and these were used as a positive weighting variable, as woodland tends to occur more on steep areas and escarpments than on areas with less variable relief (Rackham 1990) due to the lower suitability for agriculture. The other potential variable available for use as a negative weighting factor was urban development area. This was not used for other habitats as it represents a rather weak weighting factor due to the low land area that is urban (c.0.04%, based upon OS95GREF Built-up, towns). However, because woodland is such a ubiquitous and catholic habitat urban area offered a means of attempting to refine CS90 area and locational estimates. The scoring employed for these weighting factors is shown in Table 4.12.

Table 4.12 Weighted control zones: slope and urban area scoring for woodland.

Slope % (per 1km square)	**Score**
>=0 – <20	0.8
>=20	1
Urban area % (per 1km square)	**Score**
0	1
>0 – <33	0.75
>=33 – <66	0.5
>=66	0.25

These weighting variables were used to calculate slope and urban area scores respectively within the woodland master database in MapInfo according to the general methodology described above.

4.6.3 'Intelligent' weighting

The final data treatment stage was to weight the allocation of CS90 'raw' data by LCMGB data. The aim of this methodology was to refine the weighting process by incorporating the additional data, and the increased spatial resolution, offered by the LCMGB data. 'Intelligent' weighting used the LCMGB data to weight and redistribute CS90 data, thereby employing LCMGB mapped resource distributions as a stratification method. Because the highest level of CS90 data available is effectively the land class total (the locations of the original 508 sample squares are confidential) it was not possible to test this method outside the land class system.

Because lowland heathland presented a particular problem due to classification and nomenclature issues for CS90 and LCMGB data, as previously discussed, the intelligent weighting treatment for this habitat was based upon the control zone data. For the other two habitats the treatment was based upon the 'raw' data.

The 'intelligent' weighting (IW) values were calculated according to the method described in Figure 4.5. For each target habitat the calculations described above were carried out using SQL in MapInfo and by using the Table>Update column function to return data values from expressions.

Figure 4.5 Calculation of 'intelligent' weighted values.

1. Calculate the sum of 'raw' CS90 values for each land class - $\sum CS90_{L1..L2..Ln}$
2. Calculate the sum of LCMGB values for each land class - $\sum LCMGB_{L1..L2..Ln}$
3. Calculate Intelligent Weighting (IW) value where IW $= \sum CS90_{L1..L2..Ln} \cdot \dfrac{LCMGB}{\sum LCMGB_{L1..L2..Ln}}$

Where:
CS90 = Individual CS90 data value ha^{-1}
$L_{1..2..n}$ = Land class 1,2,3...32
LCMGB = Individual LCMGB data value ha^{-1}
IW = Intelligent weighting value ha^{-1}

4.7 EVALUATION METHODS

Each of the methods described above added new data columns to the master databases. For the GB-level analyses these contained over 5 million values for each habitat. The master databases were based upon 1km^2 level data and this was to form the basic unit for analysis and interpretation of the results. One of the main aims of the research was to test the viability of data use at different spatial scales. To achieve this the data were aggregated at 10km square (100km^2) and county levels using the relevant attribute codes created in the 1km-level databases. This resulted in nine databases, representing each target habitat at each spatial scale. As described above, each database contained values for the reference data set, CS90 data, CS90 Control Zone data, CS90 Weighted Control Zone data, LCMGB data, and 'Intelligent' Weighted data. These databases therefore represented the collated test data required for testing.

The performance of each estimation method needed to be assessed both in terms of their ability to define *where* habitats/landscapes occur and the *area* of habitat/landscape at each location at each scale. In order to assess each of these factors three main measures were used to assess performance:

- Measures of the overall deviations between predicted and measured areas of habitat, computed in two ways:

  NMSE the Normal Mean Square Error.

  RMSE Root Mean Square Error.

- Spatial Similarity Index (SSI) – the proportion of areas (e.g. grid squares) in which the test data correctly predicts the presence or absence of the target habitat. The SSI was calculated in three ways (after Stehman 1999):

  SSI-1 for all squares (i.e. total where presence or absence was correctly estimated);

  SSI-2 for absence (i.e. proportion of squares not containing habitat where absence was correctly predicted);

  SSI-3 for presence (i.e. proportion of squares containing habitat where presence was correctly predicted).

- Correlation – the test data were regressed against the reference data to produce a measure of the coefficient of determination (r^2).

- All these methods were applied at three levels of spatial aggregation:

  i. 1 km grid square (1km^2);

  ii. 10 km grid square (100km^2);

  iii. County (Variable area).

4.7.1 Normal Mean Square Error

The Normal Mean Square Error (NMSE) was calculated for each target habitat, at each scale according to the formula given below (after Kukkonen *et al.* 2000):

NMSE calculation:

$$NMSE = \frac{1}{n}\left(\frac{\sum (x-y)^2}{(\bar{x}\cdot\bar{y})}\right)$$

Where: x = Reference (observed) value
y = Estimated (predicted) value
n = Number of cases

The calculations were made in MapInfo using separate SQL expressions for each NMSE to calculate the numerator and denominator components of the expressions:

Numerator - Select avg ((*tablename.RefData* – *tablename.TestData*) ^ 2) from *tablename* into selection01

Denominator - Select avg (*tablename.RefData*) * avg (*tablename.TestData*) from *tablename* into selection02

Selection01 was then divided by selection02.

4.7.2 Root Mean Square Error

The Root Mean Square Error (RMSE) was calculated at each scale according to the formula given below (after Kashyap 2001):

RMSE calculation:

$$RMSE = \sqrt{\left(\frac{\sum (x-y)^2}{n}\right)}$$

Where: x = Reference (observed) value
y = Estimated (predicted) value
n = Number of cases

The calculations were made in MapInfo using an SQL expression:

Select (avg ((*tablename.RefData* – *tablename.TestData*) ^ 2)) ^ 0.5 from *tablename* into selection01

4.7.3 Spatial Similarity Index

The Spatial Similarity Index (SSI) was calculated in three ways according to the index being calculated (1, 2 or 3).

To calculate SSIs, each habitat database, at each scale had additional columns added to represent presence or absence (a 'PA' column). In each database, PA columns were added for the reference, CS90, CS90CZ, CS90WCZ, LCMGB and IW data columns. By use of SQL expressions the PA columns were updated to equal 1 where the row data indicated a habitat area value (present) and to 0 where the row data was null (absent). These coded PA columns were then used to calculate the SSI values according to the rules described in Table 4.13.

Table 4.13 Calculation of SSI, classification of presence and absence.

	SSI-1	**SSI-2**	**SSI-3**
Case where statement is true	RefPA – TestPA = 0	RefPA + TestPA = 0	RefPA + TestPA = 2
n used in calculation	n = population	n = No. squares where estimate = 0	n = No. squares where estimate = 1

Note: 'Squares' relate to geographic areas, thus they refer to counties for county scale calculations.

SSI-1 is the percentage of all squares where **presence or absence** was correctly estimated. This could be calculated in MapInfo using an SQL expression, the PA columns and by referring to Table 4.13. Thus for SSI-1 squares where the presence or absence of habitat was correctly estimated are where RefPA – TestPA = 0. The occurrence of these cases were calculated:

Select count (*) from *tablename* where (RefPA – TestPA) = 0 into selection01

Selection01 was then divided by n (in this case n = the population) and multiplied by 100 to express the figure as a percentage.

SSI-2 is the percentage of squares correctly estimated for **absence** of habitat. Therefore this could be calculated according to:

Select count (*) from *tablename* where (RefPA + TestPA) = 0 into selection02

Selection02 was then divided by n (in this case n = No. cases where TestPA = 0) and multiplied by 100 to express the figure as a percentage.

SSI-3 is the percentage of cases correctly estimated as **present**. Therefore this could be calculated according to:

Select count (*) from *tablename* where (RefPA + TestPA) = 2 into selection03

Selection03 was then divided by n (in this case n = No. cases where TestPA = 1) and multiplied by 100 to express the figure as a percentage.

Each of these steps was conducted for each target habitat, test data estimate and geographic scale. The results were tabulated by habitat and scale.

4.7.4 Calculation of r^2

The final testing stage was to test the relationship between the reference data sets and the estimated values. The sheer size of the data sets precluded the use of standard office software (i.e. MS Excel™ has a row limit of c. 64,000), so the data were exported from MapInfo in a dbf format and imported into SPSS for Windows™ V.9.0.

The regression analyses were carried out using SPSS' 'Analyze>Regression>Linear' function. In each case the reference data was set as the dependent variable. Regression statistics were computed for each target habitat, test data estimate and geographic scale. Examination of the data, using scatterplots, showed that marked heteroscedasticity existed in many cases, and emphasised the non-normality of the data (due to a large number of squares with zero values, where the habitat did not exist). Both these characteristics violate the strict assumptions of regression analysis, although the extent to which such violations invalidate exploratory investigations, such as those applied here, as opposed to statistical inference is a matter of some debate. Overall, the effect is likely to be to overestimate the significance of the regression model. In this study, such effects are of only limited concern because the aim is to *compare* results from different methods, tested under broadly the same statistical conditions, rather than to use the regression results as a basis for hypothesis testing.

Various methods were nevertheless examined to reduce the effects of these model violations. One widely recommended approach is to log the data. This, however, failed in this case to normalise the data, or greatly to reduce heteroscedasticity; log transformation also makes interpretation and reporting of the results more difficult (and potentially misleading), since it means that slope coefficients and other parameters of the regression model are not constant across the full range of data. On advice from Prof. Sylvia Richardson (pers. comm.), therefore, two separate analyses were performed: one on the entire data set using untransformed data; the other using data only for those cases where the habitat was present (i.e. where the area of the reference data >0). The first of these analyses thus assesses the extent to which (within the admitted limitations of the data) estimates of habitat area are correlated with actual habitat area (as defined by the reference data), across all cases. The second analysis assesses the more limited condition of how well the estimates of habitat area correlate with actual habitat area within those areas in which the habitat actually exists. Differences between the regression parameters (especially the r^2 value and standard error of the estimate) give a general indication of the extent to which statistical violations in the full model (analysis 1) are affecting the regression results.

The results were tabulated by habitat and scale and are presented in the following chapter.

4.7.5 Geographical patterns of deviances

To assess the geographical distribution of deviance between data treatment methods and the reference data further calculations were made. In each target habitat database, and at each scale, additional columns were created to calculate the difference between the reference data estimate and the test data estimate in each case. For the 1km^2 and 100km^2 levels these were simply expressed as the arithmetic difference in ha. At the county level the area differences were expressed as a percentage of the total county land area to allow for the different geographic unit size. The calculations were executed in MapInfo using SQL e.g.:

```
Update SM_compCountyGeog Set CS90~OSNCC = ((CS90-OSNCC)/Area[obj, "hectare"])*100
```

The results tables were then linked with vector geographic data using an SQL join, to enable thematic mapping of the results. The thematic maps were developed for each habitat, at each scale, and were based upon data ranges derived from the maximum and minimum ranges present within each habitat across all treatments at a given scale. Using this method the graphical values (colour shading) depicted by each map were directly comparable with each other.

4.7.6 Land Class testing

The ITE land class system is integral to the CS90 estimation model. Because of this, testing was conducted to investigate its performance with the three target habitats.

In order to investigate the effect of ITE land class upon the results, data for each target habitat were queried out in MapInfo using SQL. Land class totals for each target habitat were calculated for reference, CS90 and LCMGB estimates. In the case of lowland heathland, the CS90 and LCMGB totals were based upon control zone values because of the nomenclature issues previously discussed. The land class totals were adjusted for land class area and used for linear regression analyses at the whole land class level.

To further investigate land class performance additional testing was conducted at the 1km square level. For each target habitat and each land class, the respective values for the LCMGB data and reference data were queried and entered into linear regression analyses of LCMGB data against the respective reference data. In the case of lowland heathland, the LCMGB control zone data were used. It was not applicable to use CS90 data for this test, as only an arithmetic mean total is available for each land class. In contrast, the LCMGB provides a (measured) data value for each km square of the GB and therefore was amenable to analysis in this fashion.

5 RESULTS

5.1 INTRODUCTION

The data treatments and evaluative processes described in the previous chapter produced a huge quantity of results. For each habitat tested large tables were generated in MapInfo. For the two GB wide habitats at a 1km square level, these tables constituted nearly a quarter of a million records and thirty-two data fields, thereby constituting nearly eight million values. These tables were repeated for each additional geographic scale (10km square and county) and for different stages in the analysis and together resulted in over 1GB of data. It is therefore impossible to present all the results here: summaries are given in this chapter and detailed results are available in MapInfo formats from the author.

This chapter summarises the results obtained for the three test habitats, using the various data sources and modelling methods described earlier. Performance of the various approaches is evaluated and compared using measures of bias (differences in total area, where relevant), spatial location (SSI values), correlation (r^2) and mean error (NMSE and RMSE). The results related here are grouped by target habitat and are mainly presented as tables and thematic maps. The maps are presented for county level and 10km square data. Maps for 1km square data are not presented, as at this scale it is not possible visually to interpret the printed thematic maps. The data were held and processed to a very high level of precision (15 decimal places). However, the data are unlikely to be accurate to more than three decimal places at best and are therefore only expressed to two or three decimal places as appropriate.

5.2 SALTMARSH

5.2.1 Quantity/area

The first comparison measure for saltmarsh habitat was the total area habitat estimated by each data source. These figures are presented in Table 5.1. It should be noted that the reference data set, based upon a combination of the NCC Saltmarsh Inventory and Ordnance Survey mapped data, represents an underestimate of 8,435 ha (19%) compared to the original (NCC Saltmarsh Inventory) total figure of 44,370 ha. In the light of this, relatively little significance should be attached to differences in the absolute totals between the reference data set and the test data sets.

Table 5.1 Total saltmarsh habitat areas for GB by data set.

Data set	Sum of saltmarsh areas/ha
OS/NCC (Reference)	35,934.7
CS90	39,667.9
CS90 Control Zone	39,667.9
CS90 Weighted Control Zone	39,667.9
LCMGB	38,938.5
CS90LCMGB Intelligent Weighting	39,667.9

Because the nomenclature for saltmarsh classifications in the reference and test data sets were broadly equivalent, the original CS90 landclass totals were used in each test scenario, resulting in the same overall totals. The total reported by LCMGB is very close to that reported by CS90, differing by only c.730ha.

5.2.2 County level

The results outlined in Table 5.2 show that all treatments produced reasonable estimates for saltmarsh at the county level; Normal Mean Square Error (NMSE), Root Mean Square Error (RMSE), Linear regression and Spatial Similarity Index (SSI) analyses are presented. A lower NMSE/RMSE score indicates a more accurate estimate. As explained in Chapter Four, the SSI was calculated in three ways: SSI-1 is the percentage of all squares where **presence or absence** was correctly estimated, SSI-2 is the percentage of squares correctly estimated for **absence** of habitat, SSI-3 is the percentage of cases correctly estimated as **present**.

There was relatively good correspondence between the nomenclature used in all the data sets so we would expect broadly similar estimates from each treatment. The CS90 based estimates all produce good estimates for saltmarsh. The 'raw' CS90 estimate produces a relatively low NMSE score, which increases with the addition of the control zone. However, adding weighting factors improves the score compared to both the control zone and 'raw' cases. LCMGB has a specific saltmarsh category and should be expected to perform strongly. The LCMGB estimate produces the joint lowest NMSE score, equal to that of the CS90 weighted control zone method. The intelligent weighting technique does not improve upon this or the 'raw' CS90 score. The RMSE scores confirm the NMSE results and this pattern of performance is also reflected by the coefficients of determination (r^2). Correlation coefficients are statistically significant for all cases; but in all cases r^2 values for the presence-only data are somewhat lower than those for the full data set. This suggests that, in this case, violations of the regression model for the full data set might be causing some over-estimation of the strength of the correlation with the observed habitat area.

Table 5.2 Saltmarsh County validation: Mean Predicted, Mean Measured, NMSE, RMSE, r^2 and SSI.

DATA	$\bar{P}$ (ha)	$\bar{M}$ (ha)	NMSE (ha)	RMSE (ha)	r^2	r^2 (Ind>0)	SSI-1	SSI-2	SSI-3
CS90 NCC/OS v CS90	592.06	536.34	1.31	643.99	0.538	0.526	64.18% 43/67	60.00% 3/5	64.52% 40/62
CS90 Control Zone NCC/OS v CS90CZ	592.06	536.34	1.84	767.38	0.416	0.320	89.55% 60/67	90.91% 20/22	88.89% 40/45
CS90 Weighted Control Zone NCC/OS v CS90WCZ	592.06	536.34	0.92	541.93	0.697	0.602	85.07% 57/67	74.19% 23/31	94.44% 34/36
LCMGB NCC/OS v LCMGB	581.16	536.34	0.92	534.31	0.778	0.750	92.54% 62/67	100% 20/20	89.36% 42/47
Intelligent Weighting NCC/OS v CS90-LCMGBW	592.06	536.34	1.59	710.56	0.733	0.703	88.06% 59/67	86.96% 20/23	88.64% 39/44

The relative performance of each measure changes slightly when the location (SSI) measures are considered. In principle, we may expect the 'raw' CS90 data to perform least well in terms of these measures, because CS90 applies a standard land class mean to all squares in land classes containing any saltmarsh. This results in saltmarsh habitat being mapped in inland areas, as well as along entire lengths of coast. The use of the control zone restricts the modelled distribution to the coastal buffer. As a consequence, SSI values improve relative to the 'raw' CS90 estimate. The weighted control zone method further improves the locational measure of presence (SSI-3), but performs less well in terms of the other two SSI measures. Overall, however, the LCMGB performs best, slightly ahead of the intelligent weighting method. These results thus suggest that the LCMGB is intrinsically more reliable in predicting habitat distribution at this scale than the CS90 data, perhaps because the CS90-based sample squares are not adequately representative for this habitat or, more likely, that the underlying land class system is not well attuned to the habitat distribution.

Figure 5.1 shows the reference data set mapped at the county-level. The reference data are presented at this scale here to aid interpretation of the following error maps depicting discrepancies between habitat estimates at the county-level. When interpreting the results, it should be remembered that counties are variable-sized units. However, saltmarsh area is not directly related to county area; if anything it is perhaps more closely associated with length of coastline. Consequently, the errors (differences between estimates) in this case are simply expressed as the (un-normalised) difference in ha. Thus, positive values (blue) represent an overestimate by the test data with respect to the reference data and negative values (red) represent an underestimate. The magnitude of these differences is, however, at least partly determined by the length of coastline in each county.

According to the reference data, the resource distribution shows saltmarsh occurring in all coastal counties with the exception of North Yorkshire, County Durham and Borders Region. The highest concentrations of saltmarsh are present in the Wash, Upper Thames estuary, Hampshire, Dyfed, Lancashire, Cumbria, Dumfries and Galloway and Highland Region. Saltmarsh distribution is therefore widespread and at this scale there is no obvious geographic bias to the distribution. By county, totals the saltmarsh resource ranges from 0 to nearly 4000ha.

The CS90 maps show a markedly different distribution to the reference data. As discussed earlier, this is mainly due to an artefact in the CS90/CIS interpolation model. Sample data are interpolated by land class; hence land classes containing saltmarsh in the field samples but not restricted to purely coastal distributions produce estimates for coastal saltmarsh in inland counties. This is shown clearly on the map by an overestimate for saltmarsh in land-bound counties. Only three counties are correctly estimated as having no saltmarsh. There is widespread overestimation of saltmarsh area across the entire country, but mainly restricted to areas where the reference data suggests low levels of the resource. Thus, there appears to be consistent overestimation in areas with low levels of saltmarsh, with the exception of Dyfed and Dumfries and Galloway, where the CS90 data overestimates for counties with large quantities of habitat. The largest overestimates are in the south east, south west, Gwynedd, Strathclyde region, Humberside and Cambridgeshire. The maximum overestimate is up to 2000ha; the maximum underestimate is 3000ha. In relation to the maximum resource value of 4000ha in any one county, these error values are very high. Underestimation occurs largely in those counties with large quantities of saltmarsh, such as Essex, Highland Region, Lincolnshire, Lancashire and Cumbria. This is probably because CS90 spreads estimates for saltmarsh across the country, thereby smoothing the geographical distribution and overestimating in counties with little or no resource and underestimating in those with larger quantities of saltmarsh.

The use of the control zone method, based on the CS90 data, changes the picture significantly (Figure 5.3) and corrects the estimates for inland counties. It also reduces the range of underestimation to -2000ha. However, the range for overestimation increases to a maximum of 4,400ha, over 100% of the maximum possible resource according to the reference data (though it should be remembered that the reference data represent an underestimate compared to the original NCC data). In most cases underestimation still occurs in areas where more saltmarsh occurs. The weighted control zone method changes the geographic distribution of error little (Figure 5.4) and the picture produced by the LCMGB estimate (Figure 5.5) is also very similar.

Figure 5.1 Saltmarsh, county level OS/NCC reference data, area ha^{-1}.

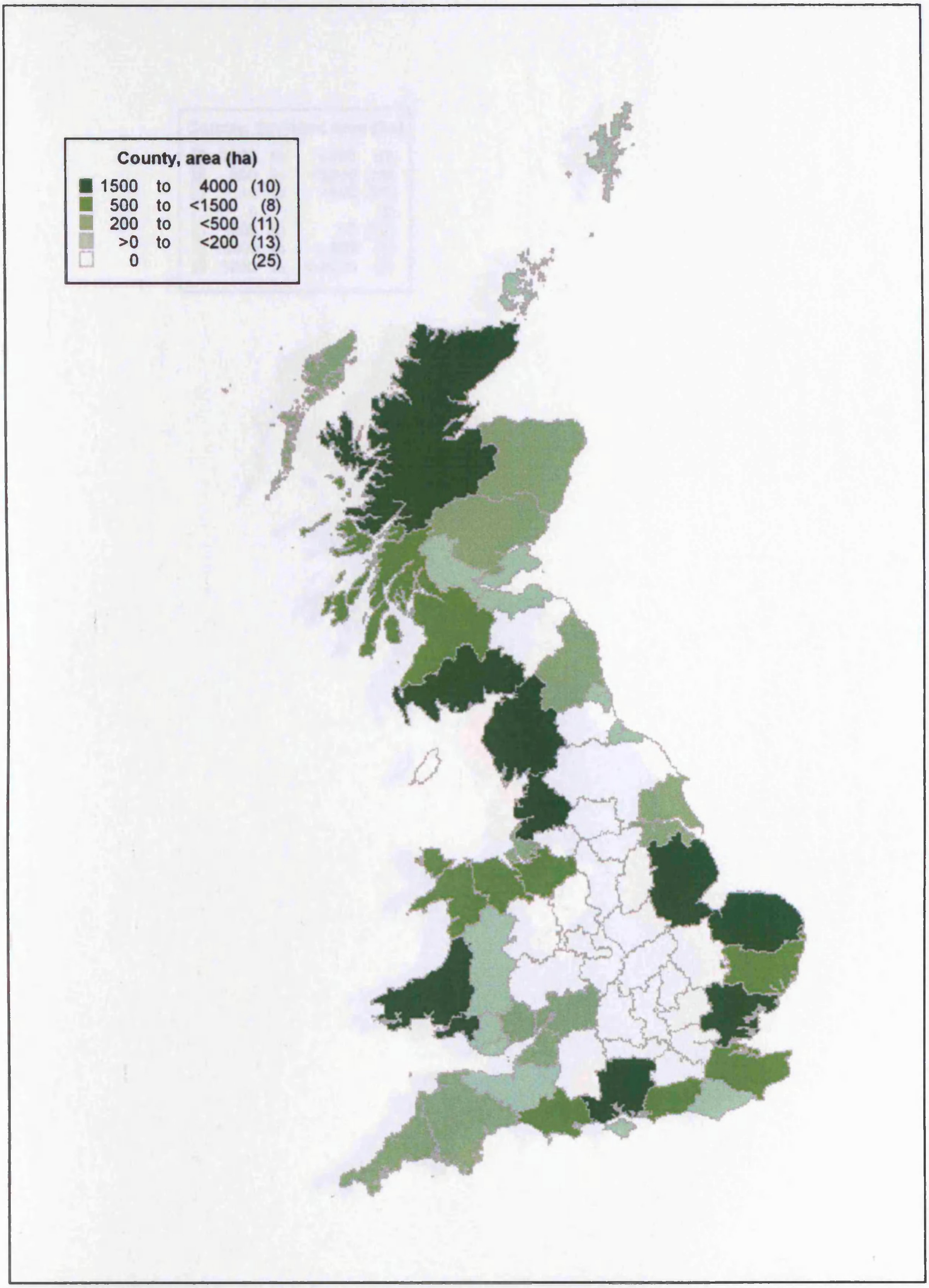


Figure 5.2 Saltmarsh County level, CS90-OS/NCC error map, area ha^{-1}.

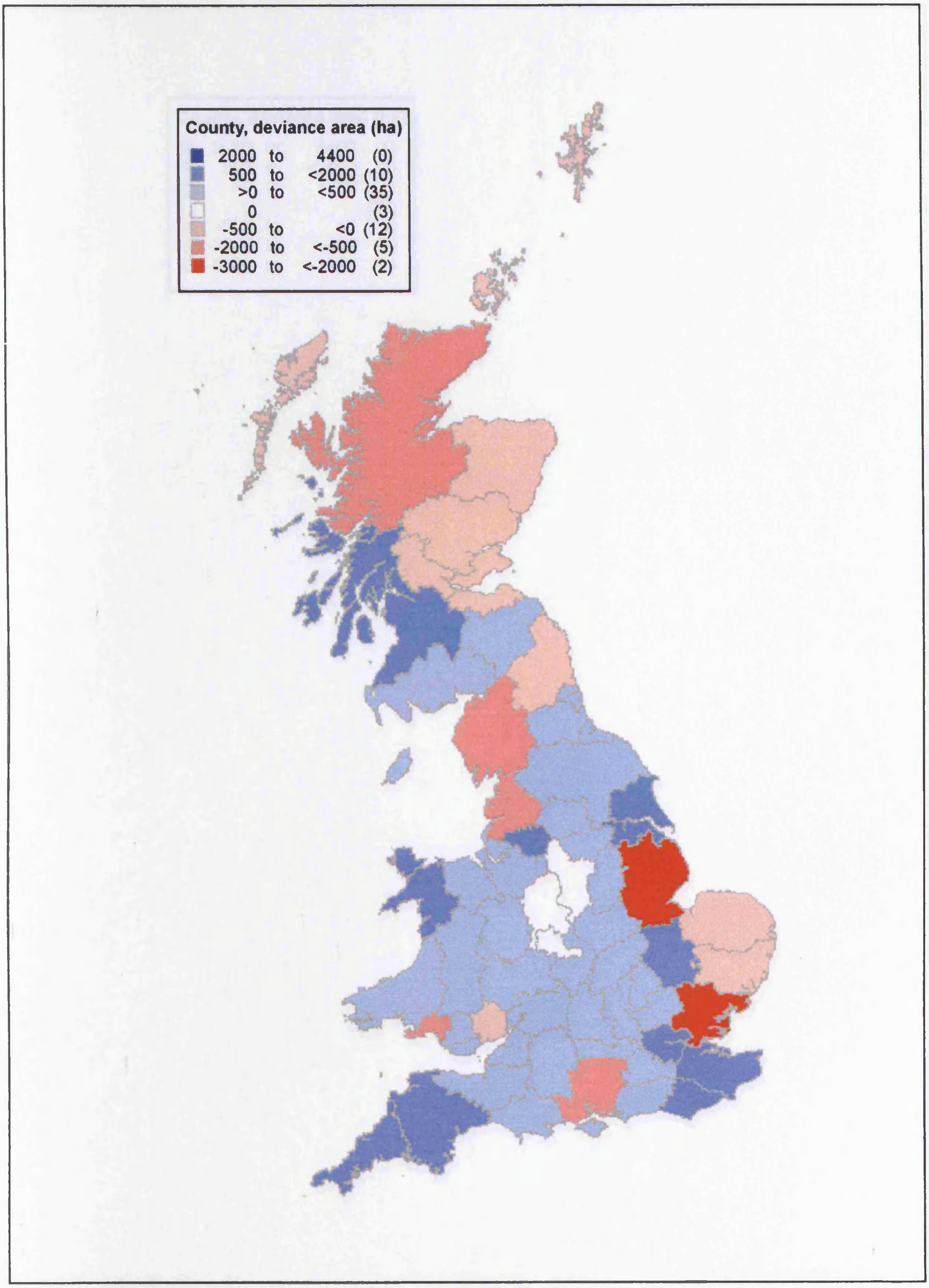


Figure 5.3 Saltmarsh County level, CS90CZ-OS/NCC error map, area ha^{-1}.

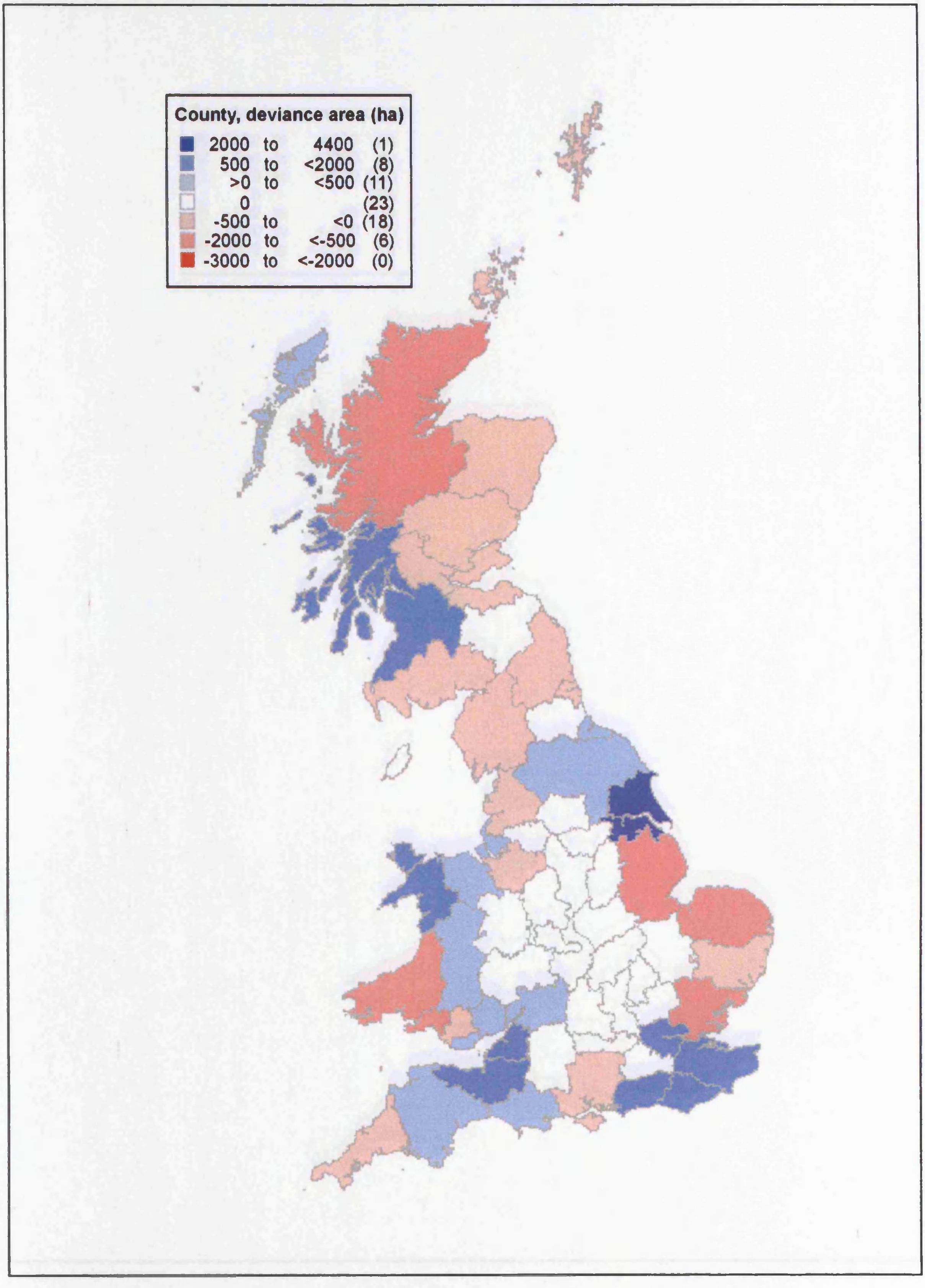


Figure 5.4 Saltmarsh County level, CS90WCZ-OS/NCC error map, area ha^{-1}.

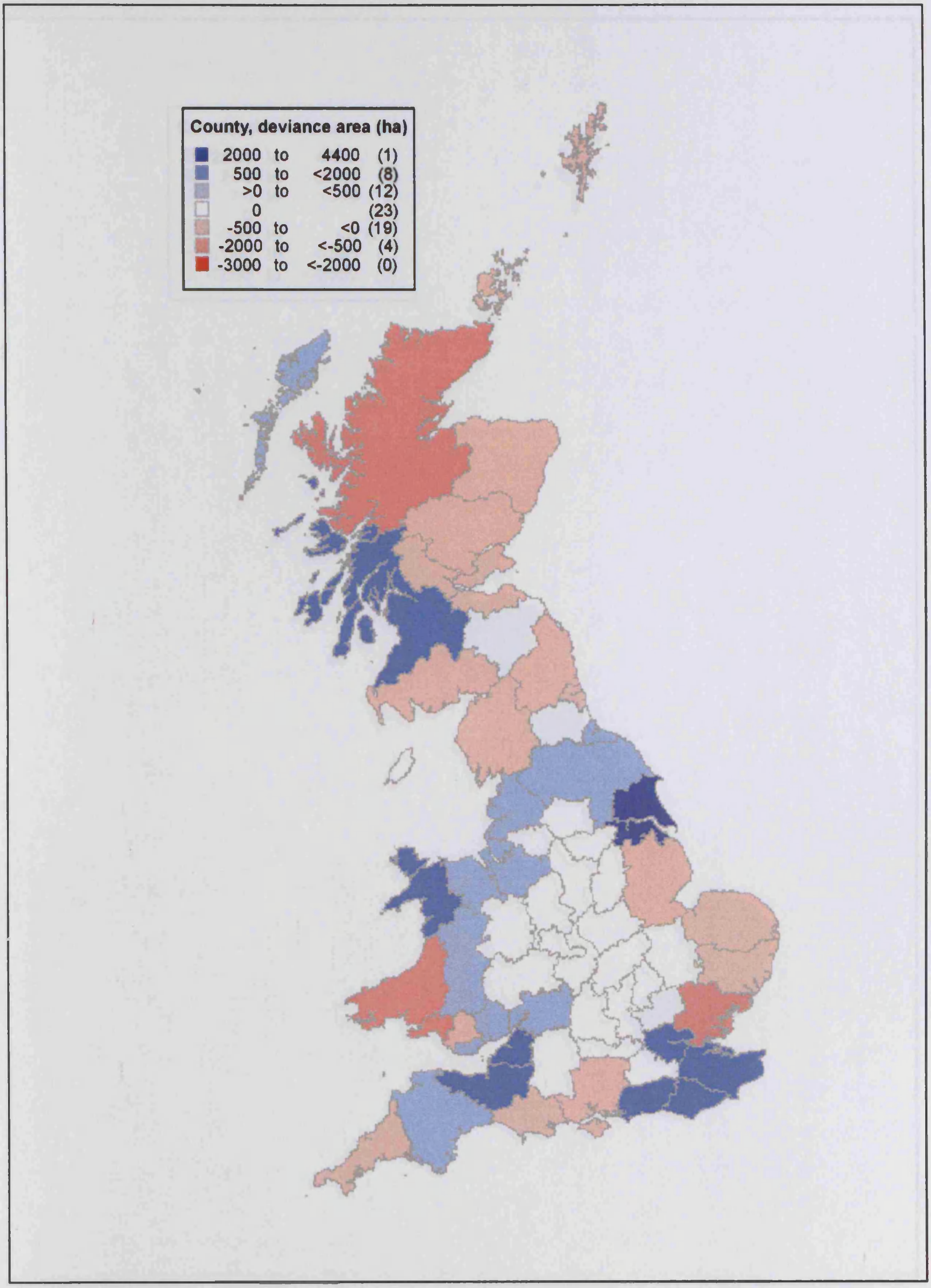


Figure 5.5 Saltmarsh County level, LCMGB-OS/NCC error map, area ha^{-1}.

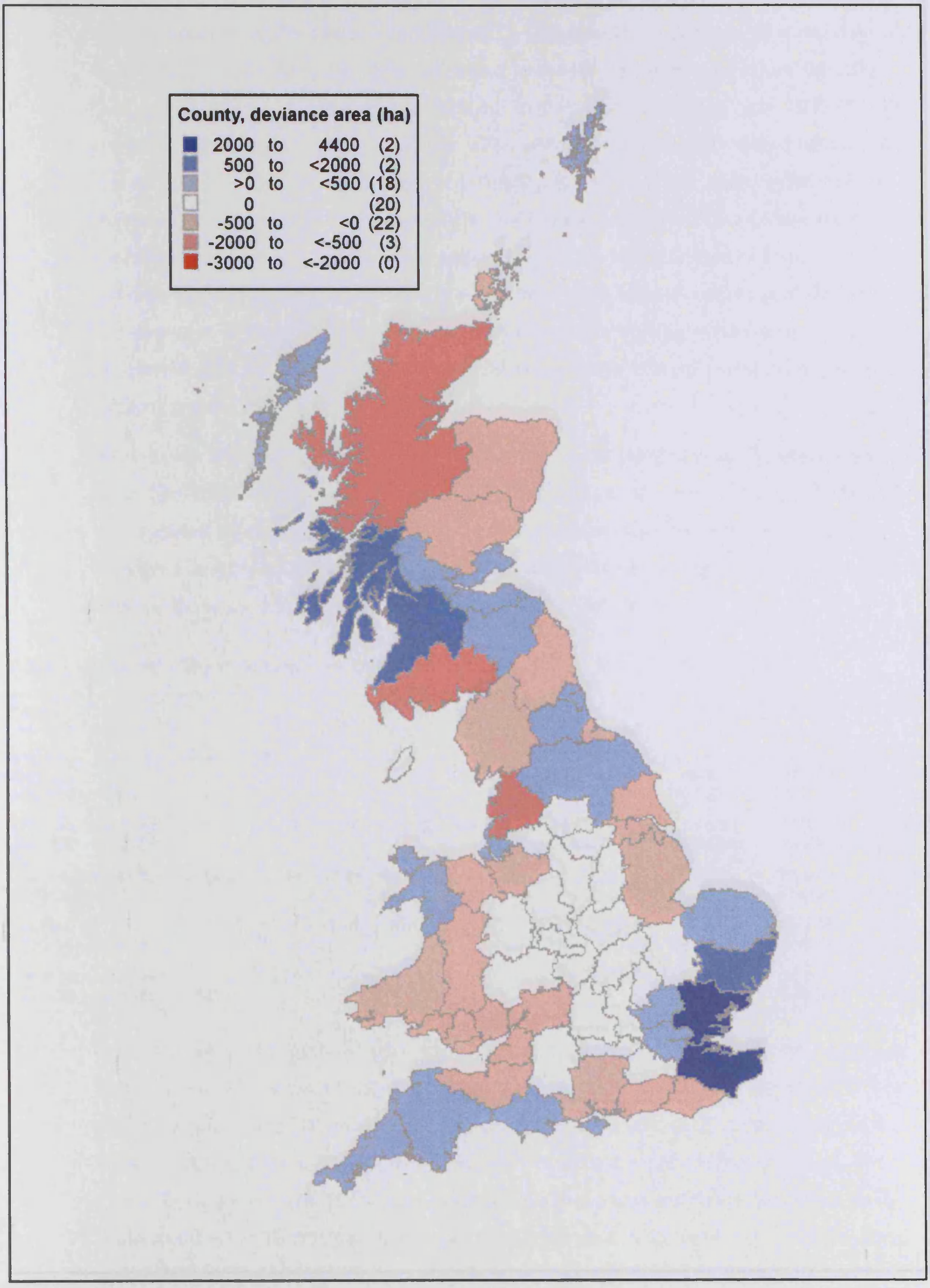
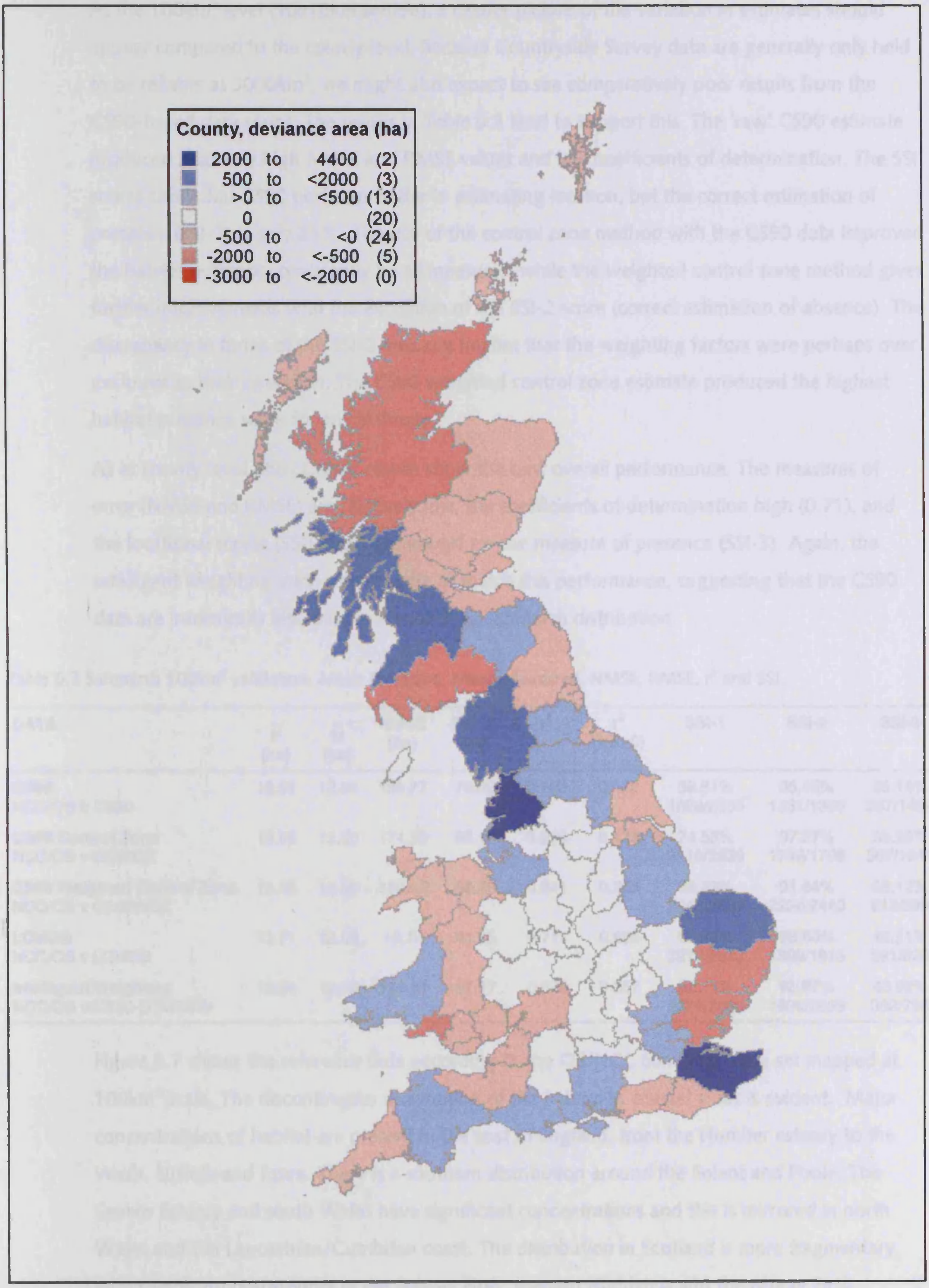


Figure 5.6 Saltmarsh County level, IW-OS/NCC error map, area ha^{-1}.



5.2.3 100 km² level

At the 100km² level (10x10km square), a clearer picture of the variation in estimates should appear compared to the county level. Because Countryside Survey data are generally only held to be reliable at 3000km², we might also expect to see comparatively poor results from the CS90-based data series. The results in Table 5.3 tend to support this. The 'raw' CS90 estimate produced relatively high NMSE and RMSE values and low coefficients of determination. The SSI scores show that CS90 performs better in estimating location, but the correct estimation of presence (SSI-3) is only 25%. The use of the control zone method with the CS90 data improved the habitat estimate consistently by all measures, while the weighted control zone method gives further improvements with the exception of the SSI-2 score (correct estimation of absence). This discrepancy in terms of the SSI-2 measure implies that the weighting factors were perhaps over exclusive in their operation. The CS90 weighted control zone estimate produced the highest habitat presence score for any estimate.

As at county level, the LCMGB results show the best overall performance. The measures of error (NMSE and RMSE) are relatively low, the coefficients of determination high (0.71), and the locational scores (SSI) also high, except for the measure of presence (SSI-3). Again, the intelligent weighting method fails quite to match this performance, suggesting that the CS90 data are intrinsically less reliable in modelling saltmarsh distribution.

Table 5.3 Saltmarsh 100km² validation: Mean Predicted, Mean Measured, NMSE, RMSE, r^2 and SSI.

DATA	$\bar{P}$ (ha)	$\bar{M}$ (ha)	NMSE (ha)	RMSE (ha)	r^2	r^2 (Ind>0)	SSI-1	SSI-2	SSI-3
CS90 NCC/OS v CS90	13.96	12.66	185.77	70.32	0.153	0.132	59.81% 1698/2839	96.45% 1331/1380	25.14% 367/1459
CS90 Control Zone NCC/OS v CS90CZ	13.96	12.66	174.20	68.10	0.209	0.173	74.53% 2116/2839	97.27% 1749/1798	35.25% 367/1041
CS90 Weighted Control Zone NCC/OS v CS90WCZ	13.96	12.66	160.16	65.29	0.341	0.255	86.23% 2448/2839	91.64% 2236/2440	53.12% 212/399
LCMGB NCC/OS v LCMGB	13.71	12.66	10.14	41.95	0.711	0.695	80.26% 2279/2839	98.68% 1888/1913	42.21% 391/926
Intelligent Weighting NCC/OS v CS90-LCMGBW	13.96	12.66	172.51	67.77	0.573	0.557	82.11% 2331/2839	95.97% 1999/2083	43.92% 332/756

Figure 5.7 shows the reference data according to the OS/NCC combined data set mapped at 100km² scale. The discontinuous distribution of the habitat in coastal areas is evident. Major concentrations of habitat are present in the east of England, from the Humber estuary to the Wash, Suffolk and Essex. There is a southern distribution around the Solent and Poole. The Severn Estuary and south Wales have significant concentrations and this is mirrored in north Wales and the Lancastrian/Cumbrian coast. The distribution in Scotland is more fragmentary, with significant occurrences in the Solway Firth, western Highlands and the Moray Firth.

Figure 5.7 Saltmarsh 100km^2 square, OS/NCC reference data, area ha^{-1}.

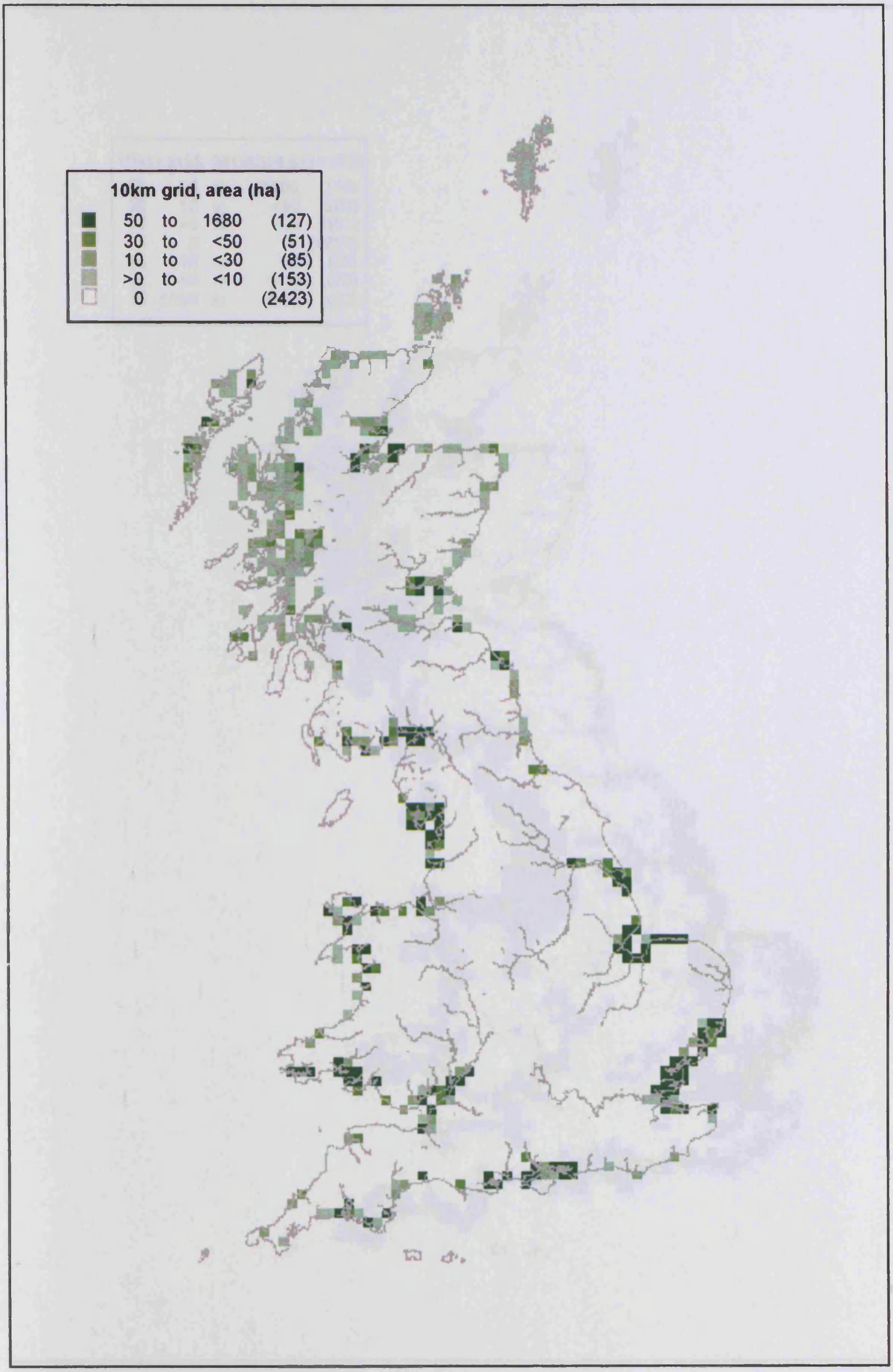


Figure 5.8 Saltmarsh 100km^2 square, CS90-OS/NCC error map, area ha^{-1}.

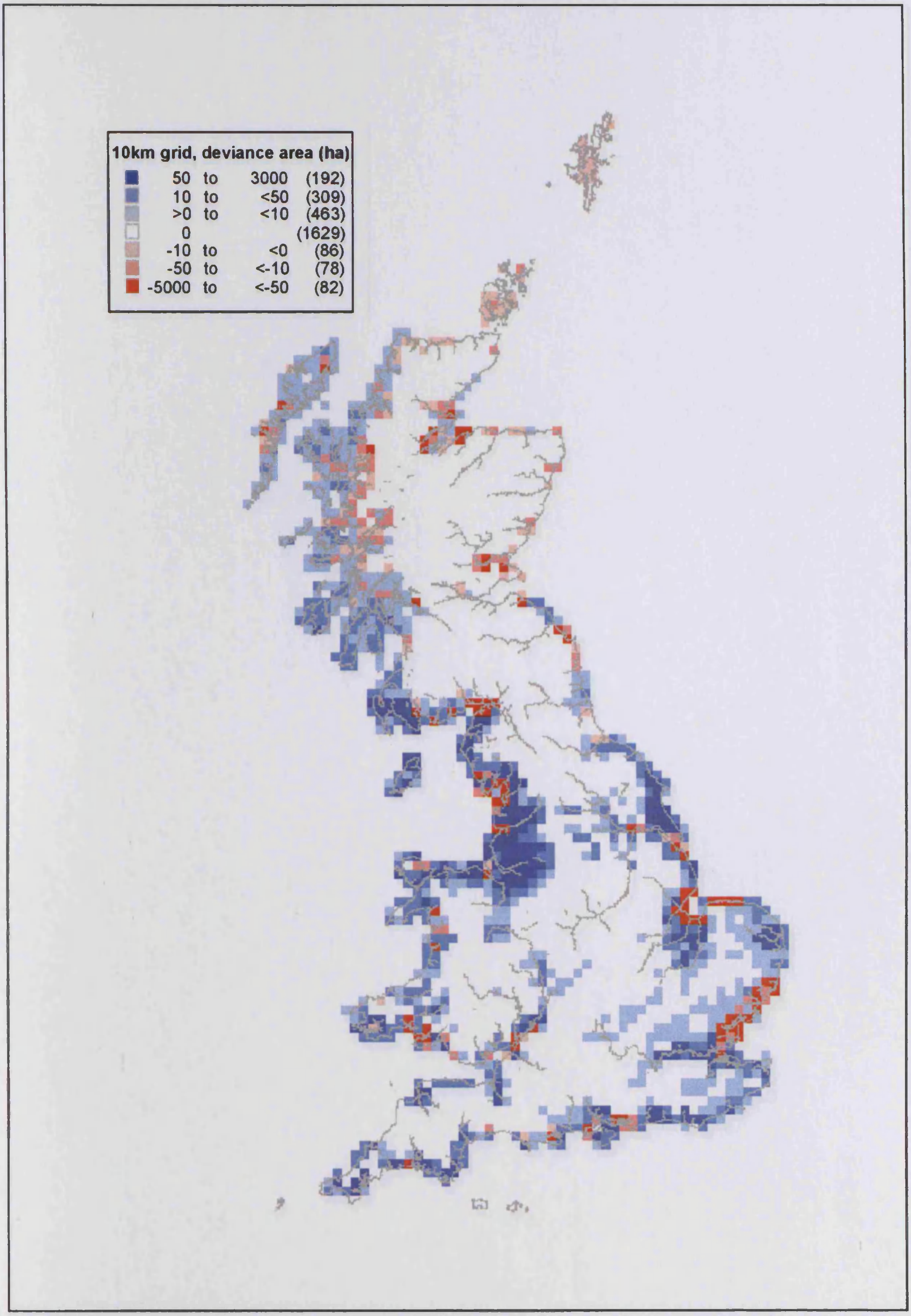


Figure 5.9 Saltmarsh 100km^2 square, CS90CZ-OS/NCC error map, area ha^{-1}.

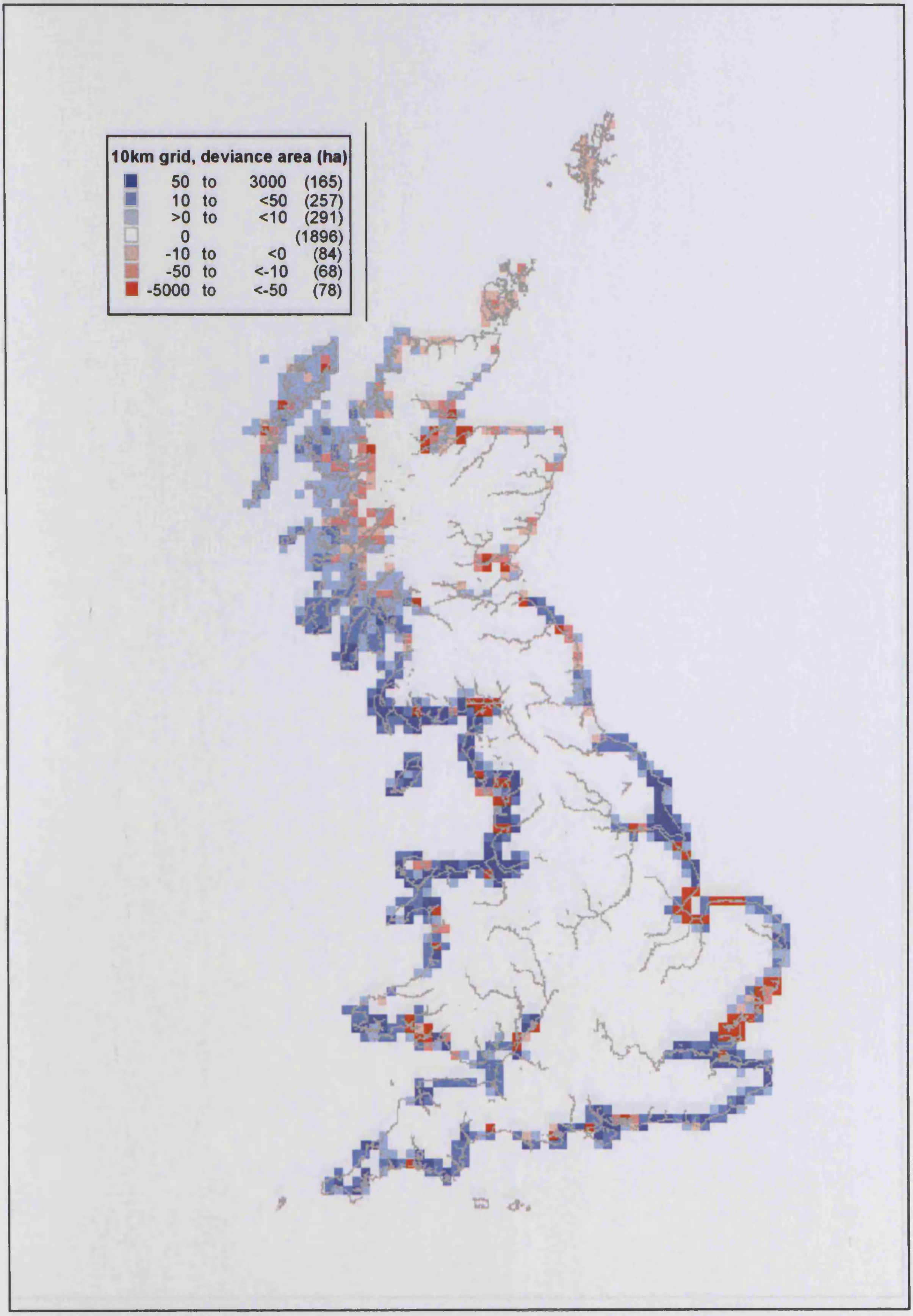


Figure 5.10 Saltmarsh 100km² square, CS90WCZ-OS/NCC error map, area ha⁻¹.

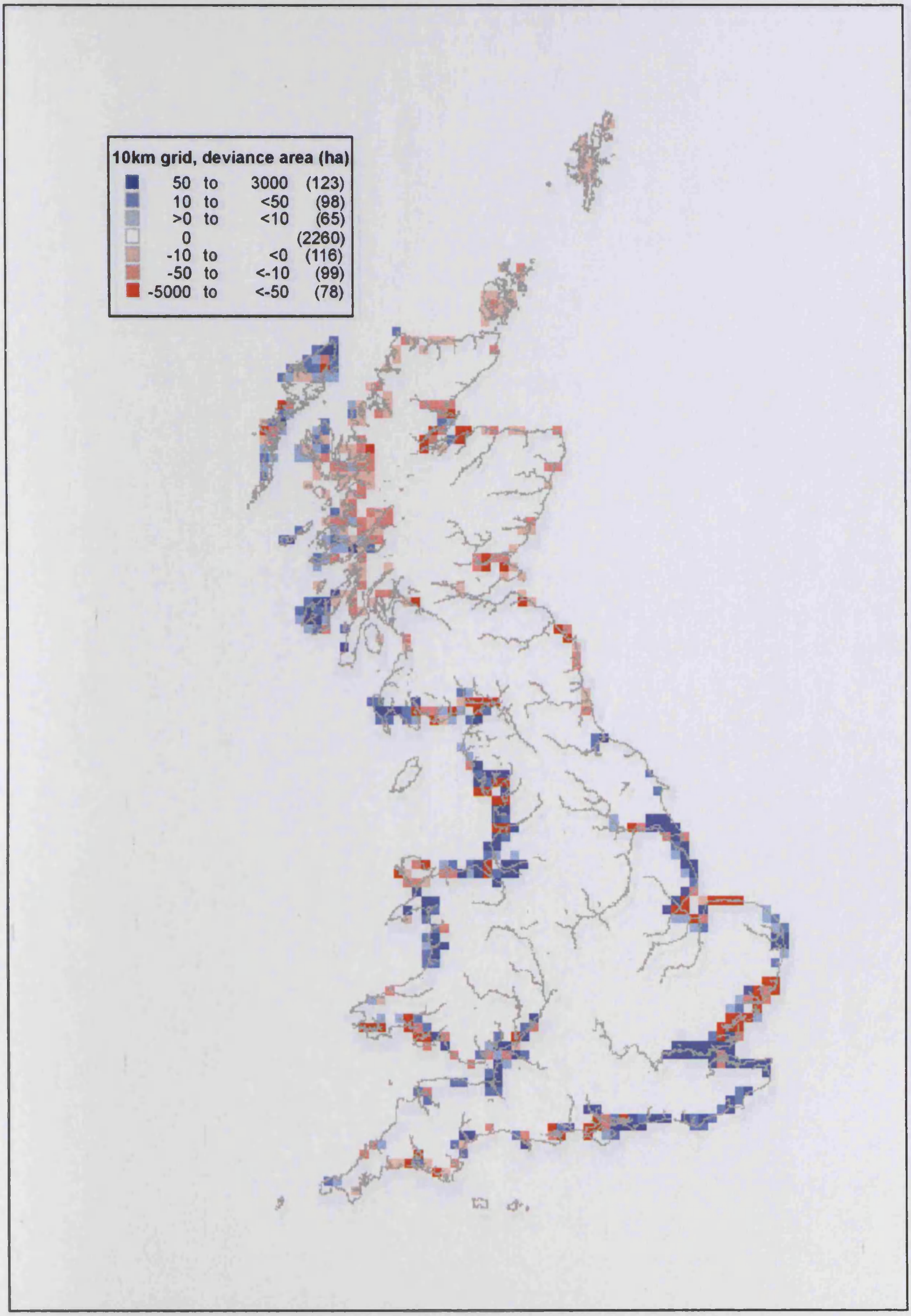


Figure 5.11 Saltmarsh 100km^2 square, LCMGB-OS/NCC error map, area ha^{-1}.

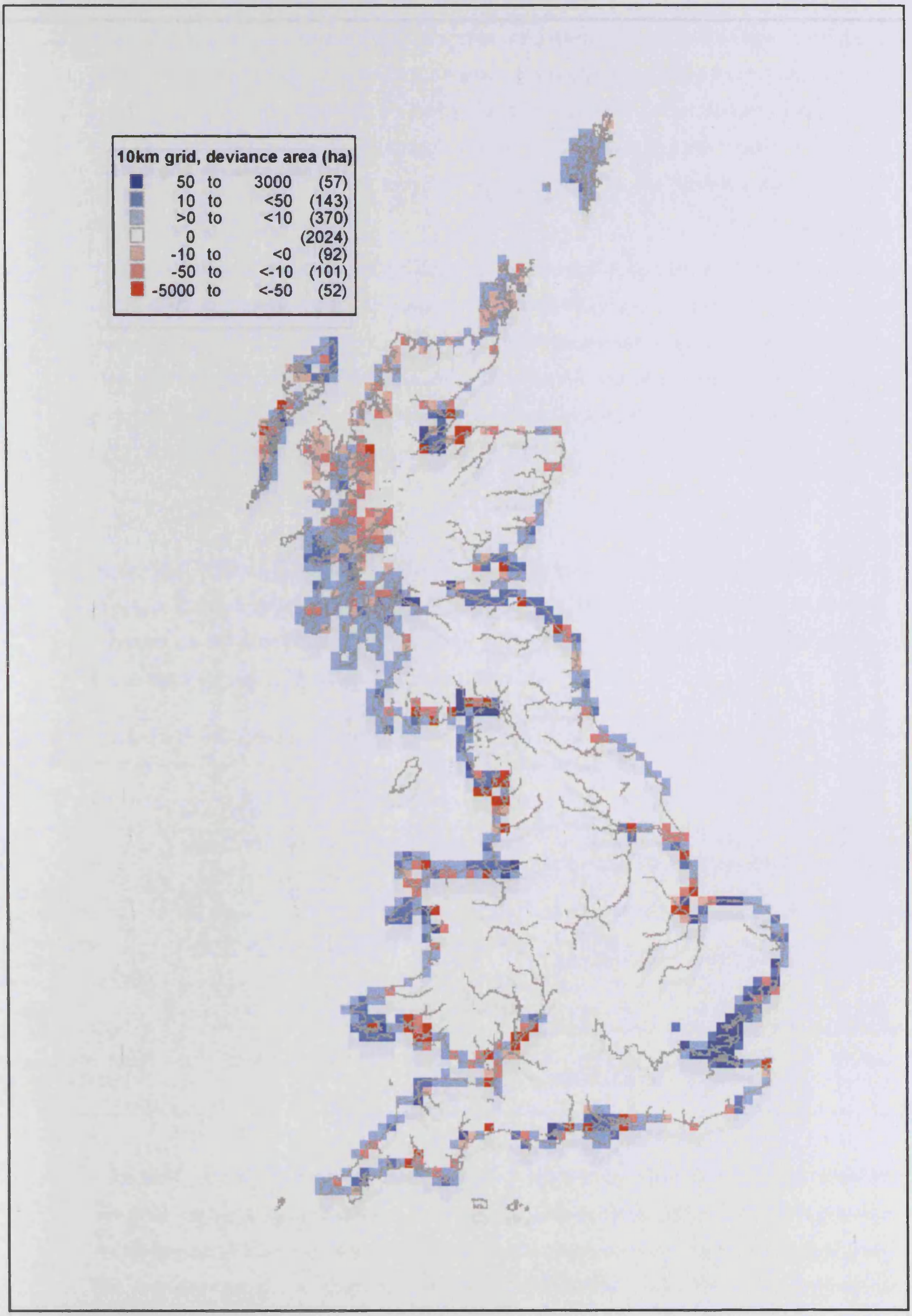
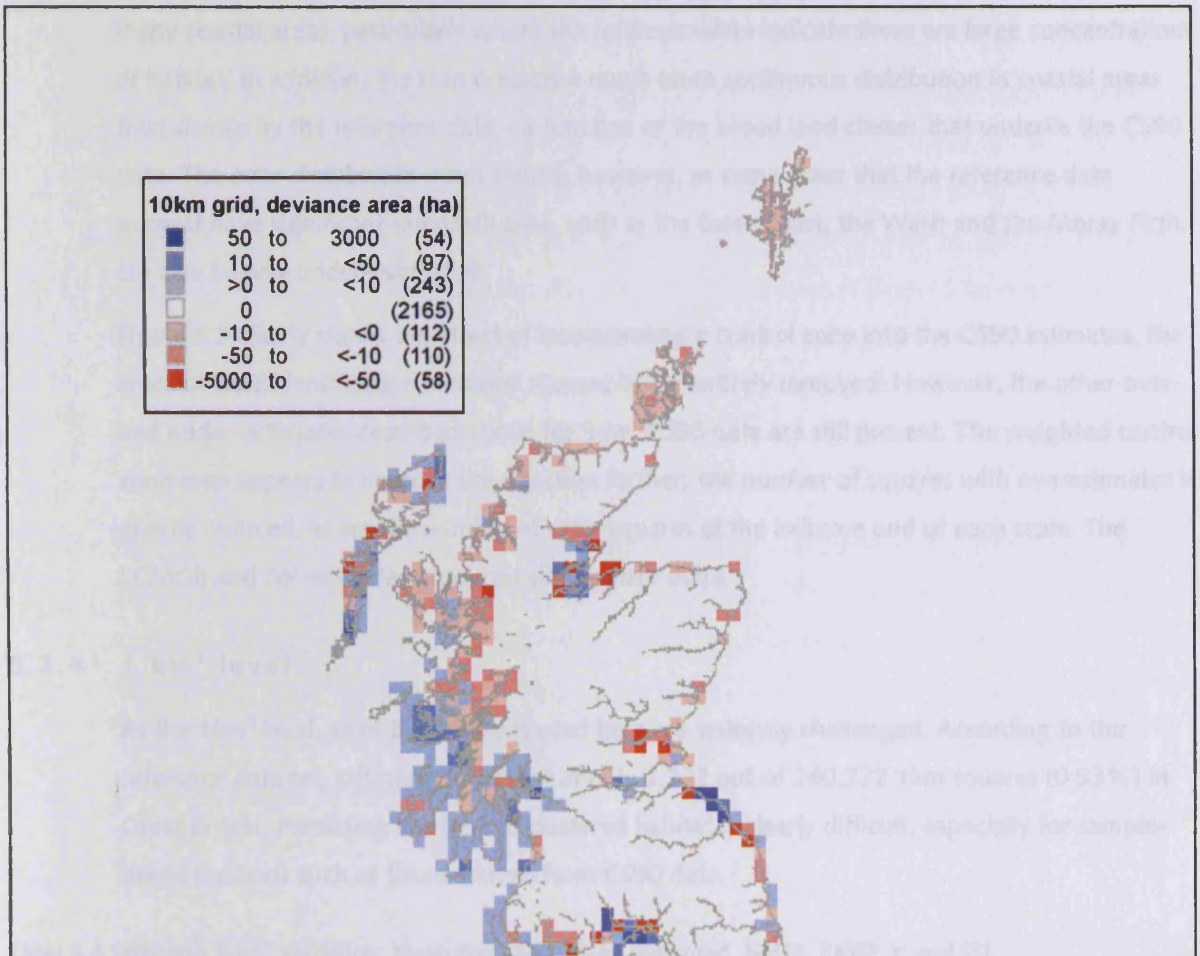


Figure 5.12 Saltmarsh 100km^2 square, IW-OS/NCC error map, area ha^{-1}.



Figures 5.8 to 5.12 show the geographic error maps. As at the county level, CS90 is seen to overestimate in inland areas where there is no coastal saltmarsh. Overestimation also occurs in many coastal areas, particularly where the reference data indicate there are large concentrations of habitat. In addition, the map predicts a much more continuous distribution in coastal areas than shown by the reference data – a function of the broad land classes that underlie the CS90 data. The error distribution is not simple, however, as some areas that the reference data suggest have significant saltmarsh area, such as the Essex coast, the Wash and the Moray Firth, are also heavily under-estimated.

Figure 5.9 clearly shows the effect of incorporating a control zone into the CS90 estimates, the areas of false identification in inland squares being entirely removed. However, the other over- and under-estimates described above for 'raw' CS90 data are still present. The weighted control zone map appears to improve the situation further; the number of squares with overestimates is greatly reduced, as are the number of error squares at the extreme end of each scale. The LCMGB and IW estimates produced similar error maps.

5.2.4 1 km² level

At the 1km² level, all of the methods used here are seriously challenged. According to the reference data set, saltmarsh is present in only 2,222 out of 240,222 1km squares (0.93%) in Great Britain. Predicting this type of clustered habitat is clearly difficult, especially for sample-based methods such as those derived from CS90 data.

Table 5.4 Saltmarsh 1 km² validation: Mean Predicted, Mean Measured, NMSE, RMSE, r^2 and SSI.

DATA	$\bar{P}$ (ha)	$\bar{M}$ (ha)	NMSE (ha)	RMSE (ha)	r^2	r^2 (Ind>0)	SSI-1	SSI-2	SSI-3
CS90 NCC/OS v CS90	0.16	0.15	236.75	2.38	0.042	0.032	85.21% 204684/240222	99.80% 202868/203274	4.91% 1816/36948
CS90 Control Zone NCC/OS v CS90CZ	0.16	0.15	233.83	2.37	0.050	0.032	89.23% 214363/240222	99.81% 214363/214769	7.13% 1816/25453
CS90 Weighted Control Zone NCC/OS v CS90WCZ	0.16	0.15	280.07	2.59	0.111	0.130	97.98% 235406/240222	99.34% 234744/236304	16.89% 662/3918
LCMGB NCC/OS v LCMGB	0.15	0.15	125.91	1.68	0.536	0.530	96.80% 232534/240222	99.79% 230802/231292	19.38% 1732/8930
Intelligent Weighting NCC/OS v CS90-LCMGBW	0.16	0.15	424.71	3.19	0.283	0.293	97.36% 233906/240222	99.71% 232369/233054	21.43% 1537/7168

Accordingly, we would expect the SSI-1 and SSI-2 scores to be relatively high when modelled at the 1km² scale, and this was borne out by the results. By the same token, the habitat presence detection rates (SSI-3) were poor. On its own, CS90 achieves a habitat detection rate of only 5%. Incorporation of the control zone increased the detection rate slightly to 7%. The addition of the weighting methods provided greater success with SSI-3 reaching 17%. LCMGB and the intelligent weighting method provided the best detection rates, at 19% and 21% respectively.

A similar pattern is seen in relation to the coefficients of determination: r^2 values increase with incorporation of the control zone and weighting variables for CS90 data, and are highest (ca. 0.53) for LCMGB (again, the intelligent weighting method performs less well).

In interpreting the results at this scale, it is important to bear in mind the limitations of the reference data. As explained in Chapter 4, these were derived from an inventory that contained only point locations and area totals for the mapped saltmarsh. To convert these to a 1 km^2 grid structure, larger sites were matched to OS data, while smaller sites were assigned to the 1 km square in which the centroid occurred. Both of these procedures is likely to have led to localised errors: saltmarsh will, at times, have been attributed to the wrong grid squares. At the broader scales discussed above, these effects are unlikely to have been significant; at the 1 km^2 level, they may have a substantial effect on model performance. At least part of the error identified at this scale may thus be attributed to errors in the reference data.

Against this background, the performance of the LCMGB in relation to this habitat is especially notable. While presence is accurately predicted only about 20% of the time, more than half the variation in saltmarsh area, across the whole of GB, is being detected by the LCMGB data. The RMSE value is also relatively low (1.68 ha), indicating a small mean error in the estimated habitat area. The implication is that the LCMGB classification provides relatively sensitive discrimination of saltmarsh at the 1 km^2 scale.

5.2.5 Summary

For a very uncommon habitat, the results for saltmarsh were encouraging, producing relatively good levels of prediction for some data treatments at more than one scale. The overall quantity estimates were relatively close, even allowing for differences in nomenclature and the lower total used in the reference data. At the county level, prediction rates were good, although, as may be expected, they decreased at the other levels. With the CS90 data the coastal control zone worked reasonably well, increasing the accuracy and success rate of CS90-based estimates at the county and 100 km^2 scales; the effect at the 1 km^2 was less pronounced. Scarce habitats are difficult to sample and map using stratified random sampling and interpolation, but in conjunction with the control zone and weighting factors the CS90 results came close to those achieved by the land cover map. At the 100 km^2 level the CS90 weighted control zone estimate produced the best habitat presence (SSI-3) estimation of 53%, exceeding the performance of LCMGB. However, for smaller unit areas, the detection of habitat presence was still rather poor by all methods.

One possible contributing factor relates to the c.8000 ha of 'lost' habitat in the reference data. This, however, is unlikely to be the main cause of error for the lost area is known to relate mainly to larger sites. In order for this to have biased the results for the CS90 estimates in the observed direction, the missing data would have to be systematically related to large saltmarsh patches located in a highly distributed fashion – i.e. not within the main areas of habitat occurrence. There is no evidence that this was the case. A more likely explanation is that the underlying land class system fails accurately to reflect the appropriate deterministic environmental factors (few land classes are completely coastal), or that the original sampling intensity for saltmarsh was not sufficient. The former would appear to be the more likely, and to relate to patch size distribution, because the overall national habitat total produced by CS90 is close to that suggested by the reference data.

Overall, therefore, the LCMGB performed the most successfully, and indeed appears to provide a relatively reliable basis for saltmarsh mapping. Further improvement on this performance might also be possible, for example using contextual data to help distinguish between saltmarsh and other coastal grassland habitats. In contrast, use of the CS90 data led to widespread overestimation of saltmarsh habitat in areas where low-levels actually occur and, in general, underestimation where there are greater quantities of habitat. Use of control zones and weighting variables enhanced the performance of the CS90 data, by helping to eliminate areas of non-presence. Further improvements may also be possible, for example by using more powerful and discriminatory control zone variables, or by deriving more sensitive weighting variables.

5.3 LOWLAND HEATHLAND

5.3.1 Quantity/area

Using Countryside Survey-based data for the estimation of lowland heathland habitat presented far greater problems than its use for predicting saltmarsh. This was because CS reporting categories did not include lowland heathland as a discrete vegetation class. This problem is highlighted by the results, shown in Table 5.5. The reference data set estimated the total area of lowland heathland in England at less than 30,000 ha. By contrast, the CS90 derived estimate is over ten times this quantity. This is because the nomenclature for CS90 includes upland heath classes; not just those restricted to lowland heathland. This issue was anticipated (see Methods chapter) but the magnitude of the discrepancy was unknown. A geographic control zone of 'lowland' ITE land classes was employed to restrict CS90 estimates to lowland areas, and this reduced the total to c.75,000 ha, still over double the reference data set estimate. The 'raw' LCMGB estimate was subject to similar nomenclature issues and was of the same order as the CS90 estimate. Because of the nomenclature issue, the control zone was also applied to the LCMGB data for lowland heathland. However, the control zone only reduced the total to c. 130,000ha, over four times that estimated by the reference data. A possible cause of this situation is that LCMGB data were estimating higher levels of heath vegetation in lowland land classes than CS90, but that these areas were not recorded as lowland heathland by the reference data. This discrepancy might have arisen due to misclassification of the satellite data, or more probably by 'contamination' from upland heath types in marginal areas still included within lowland land classes.

Table 5.5 Total lowland heathland habitat areas for England by data set.

Data set	Sum of heathland areas/ha
EN LHI (Reference)	28,313.07
CS90	342,445.03
CS90 Control Zone	74,926.16
CS90 Weighted Control Zone	74,926.16
LCMGB	364,446.32
LCMGB Control Zone	128,441.26
CS90LCMGB Intelligent Weighting	74,926.16

5.3.2 County level

The county level results are summarised in Table 5.6 and maps of resource distribution and mean error are shown in Figure 5.13 to Figure 5.19. Because of the dispersed and ubiquitous distribution of lowland heath at this scale, none of the presence/absence measures are very informative: only one county (Cleveland) contains no habitat, so SSI-2 is zero in all cases, and the other two measures thus report the same high values (98%) for all methods.

The NMSE values are very similar and fail to reflect the large differences in the predicted and measured means. The RMSE figures are more revealing. High RMSE values (indicating high levels of error) are produced by both the raw CS90 and raw LCMGB methods at this scale (18082 ha and 16307 ha respectively). This reflects the broad definition of heathland in these two data sets, neither of which discriminate between upland and lowland heathland types. However, the use of the control zone, based on lowland land classes, markedly improves the estimates from these methods, with RMSE values falling over six-fold for the CS90 control zone method and nearly five-fold for the LCMGB control zone method. Notably, use of weighting methods fails to improve on the CS90 estimate, suggesting that the weighting variables used (or the weights attached to them) are not sufficiently discriminatory. The addition of CS90 data to the LCMGB-based distribution in the intelligent weighting data set slightly decreased the RMSE compared to the LCMGB control zone estimate, suggesting that the land class-based sample data provides additional information to that contained within LCMGB.

With the exception of the LCMGB control zone method, coefficients of determination (r^2) are very low, and non-significant. The LCMGB control zone estimate, however, achieved a moderate and significant level (0.466), suggesting that it models the geographic distribution of the habitat reasonably accurately at this scale.

Table 5.6 Lowland heathland county validation: Mean Predicted, Mean Measured, NMSE, RMSE, r^2 and SSI.

DATA	$\bar{P}$	$\bar{M}$	NMSE (ha)	RMSE (ha)	r^2	r^2 (Ind.>0)	SSI-1	SSI-2	SSI-3
CS90 EN v CS90	7444.46	615.50	7.14	18082.99	0.005	0.005	97.83% 45/46	N/A% 0/0	97.83% 45/46
CS90 Control Zone EN v CS90CZ	1628.83	615.50	7.81	2798.84	0.022	0.022	97.83% 45/46	N/A% 0/0	97.83% 45/46
CS90 Weighted Control Zone EN v CS90WCZ	1628.83	615.50	9.33	3058.53	0.010	0.010	97.83% 45/46	N/A% 0/0	97.83% 45/46
LCMGB EN v LCMGB	7922.75	615.50	5.45	16307.67	0.003	0.003	97.83% 45/46	N/A% 0/0	97.83% 45/46
LCMGB Control Zone EN v LCMGBCZ	2792.20	615.50	6.47	3335.60	0.466	0.466	97.83% 45/46	N/A% 0/0	97.83% 45/46
Intelligent Weighting EN v CS90-LCMGB/CZW	1628.83	615.50	8.42	2905.61	0.049	0.049	97.83% 45/46	N/A% 0/0	97.83% 45/46

The distributions of the estimation errors are shown in Figures 5.14-5.19. In order to standardise by county area, the estimation errors were computed in percentage terms: i.e. the difference between the modelled percentage of the county area and the reference percentage for the county area. The poor performance of the 'uncontrolled' methods is evident: both the raw CS90 and raw LCMGB maps show large over-estimations in the north, and greatly underestimate the extent of the habitat in Dorset. Incorporation of the control zone reduces the errors in the north, but in both cases leaves an arc of counties, running from North Yorkshire, through the Welsh Marches, to Hampshire and Dorset, in which relatively large errors occur. Through much of this zone, the methods over-estimate the heathland area, in part, perhaps,

Figure 5.13 Lowland heathland county level, EN LHI reference data, percentage of county area.

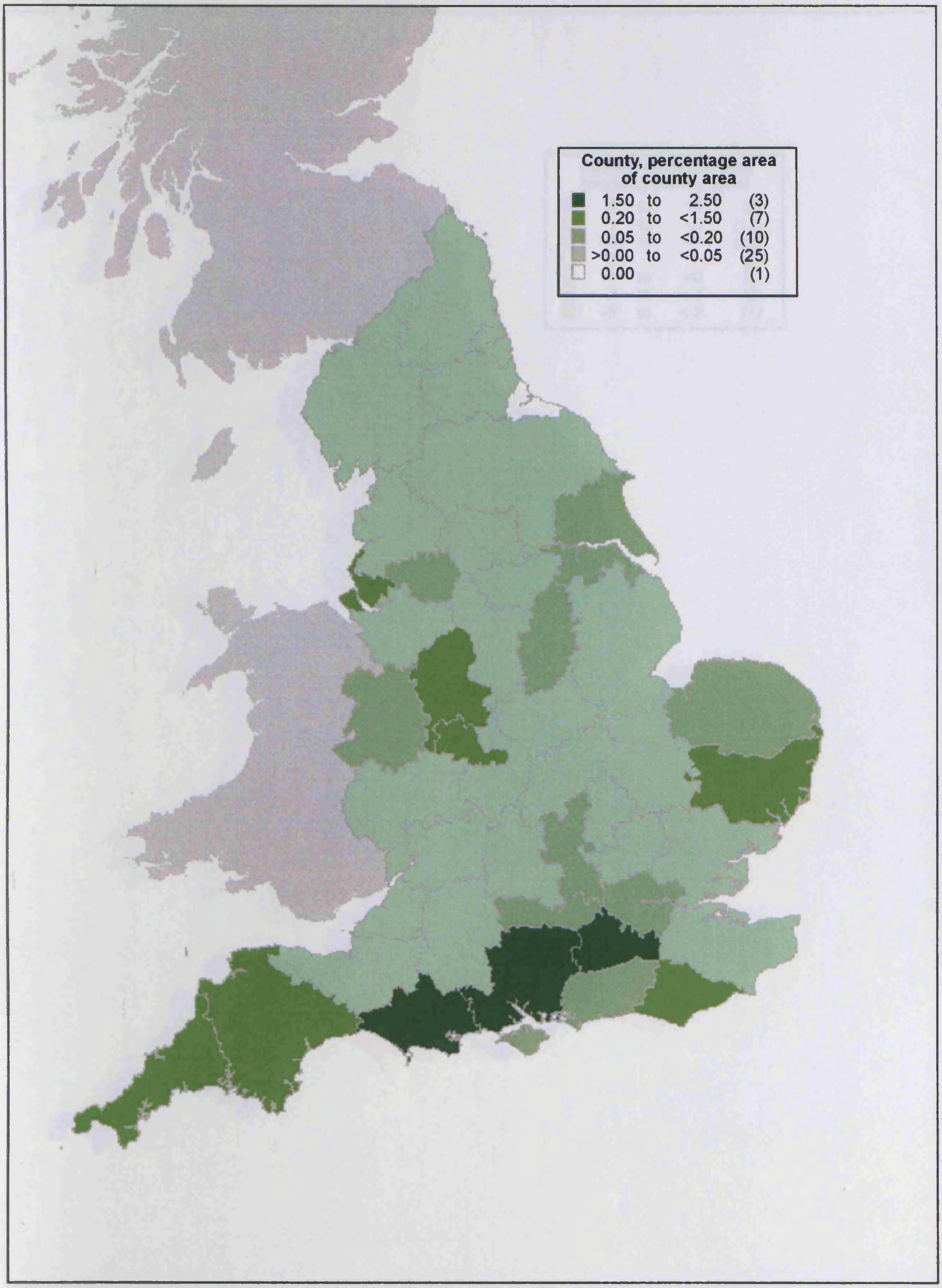


Figure 5.14 Lowland heathland county level, CS90 - EN error map, percentage of county area.

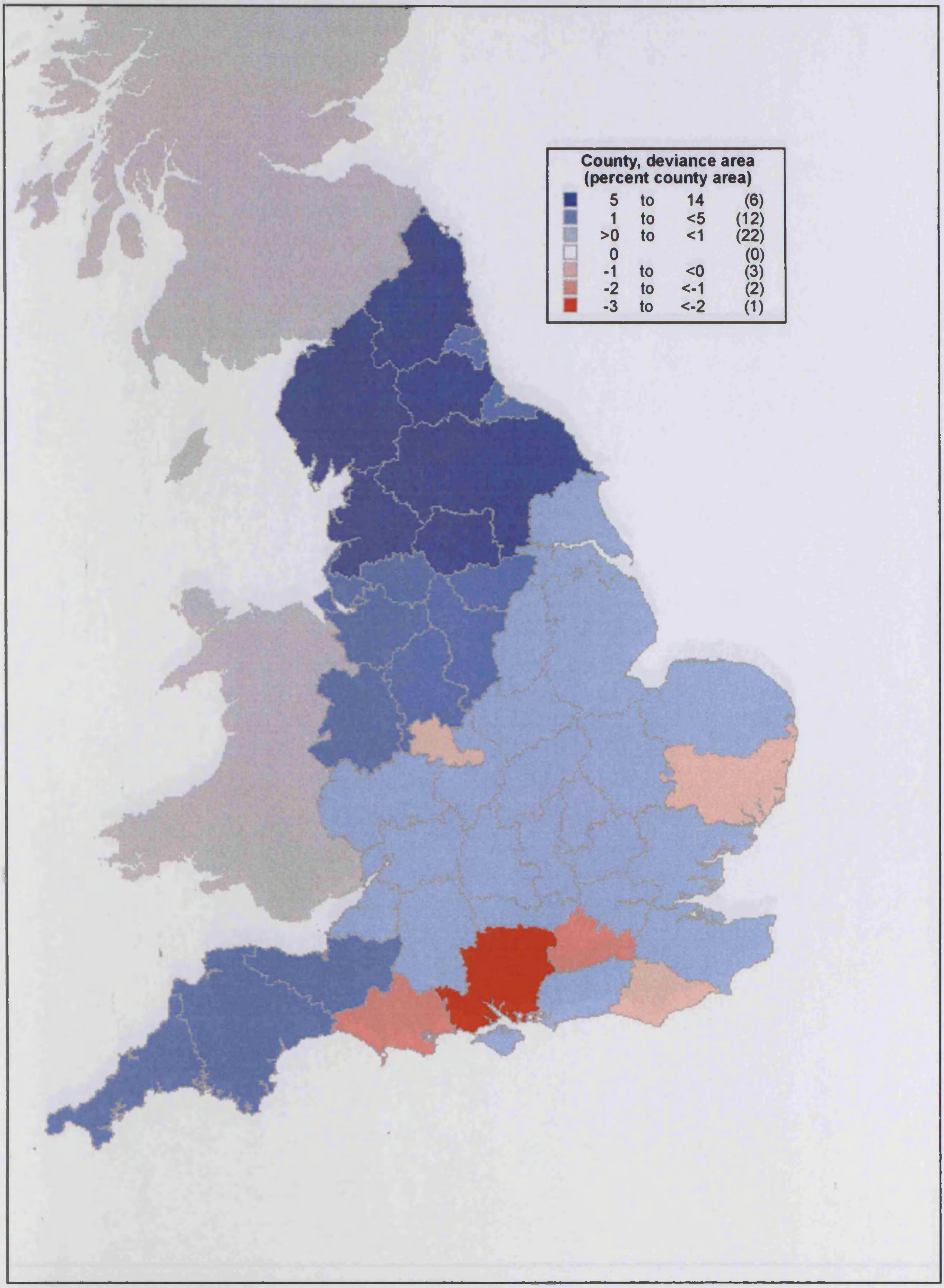


Figure 5.15 Lowland heathland county level, CS90CZ- EN error map, percentage of county area.

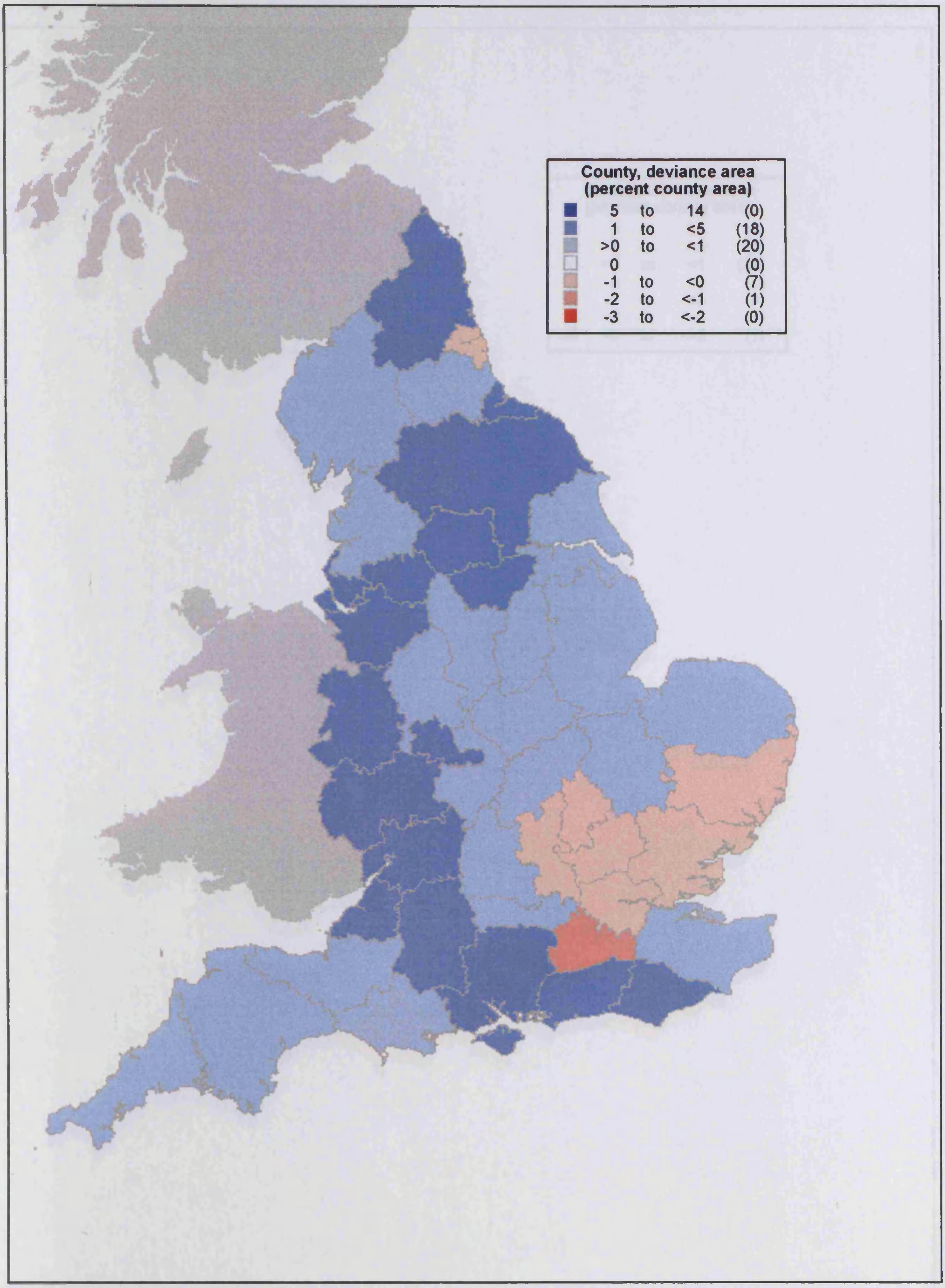


Figure 5.16 Lowland heathland county level, CS90WCZ - EN error map, percentage of county area.

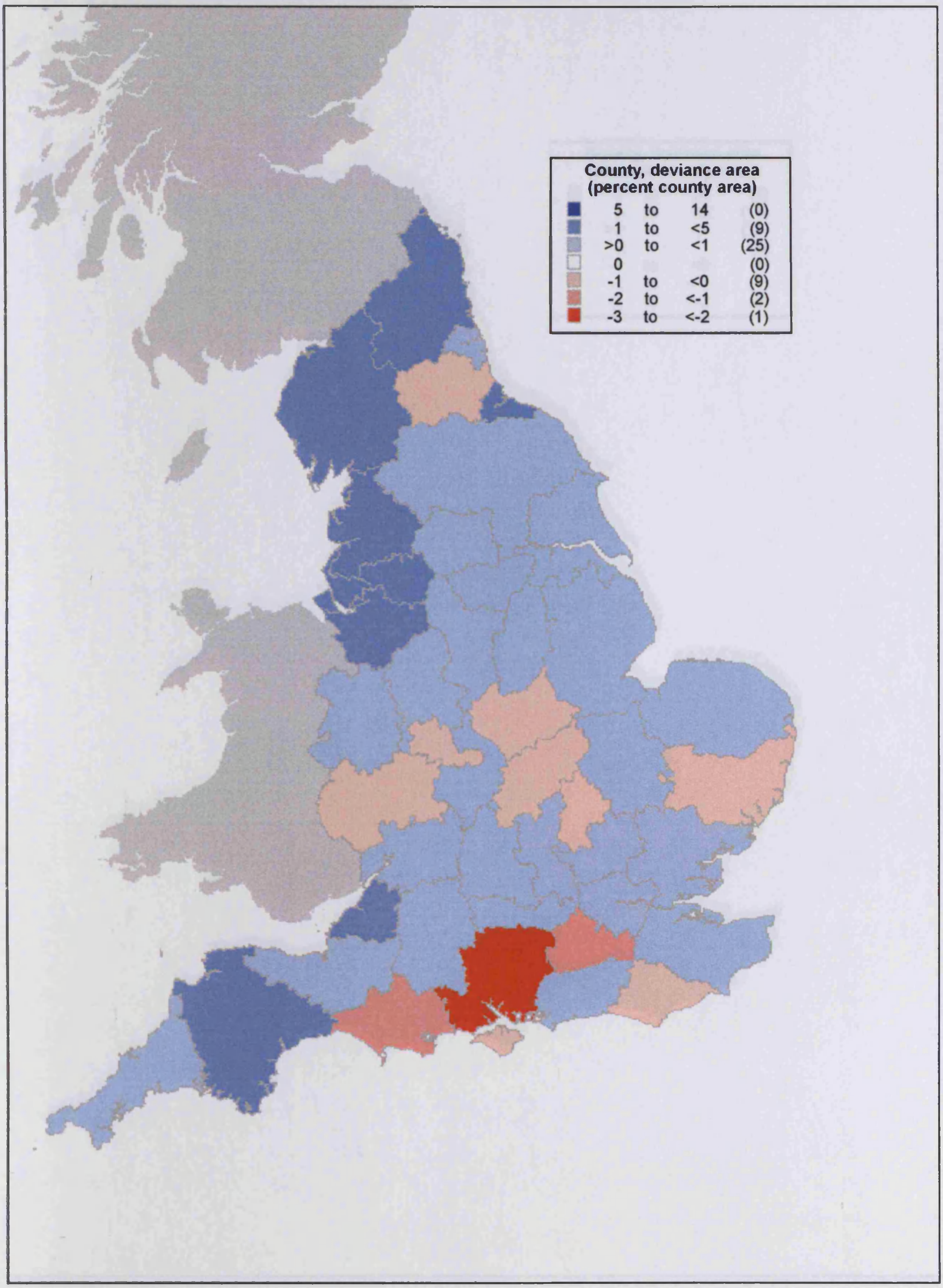


Figure 5.17 Lowland heathland county level, LCMGB - EN error map, percentage of county area.

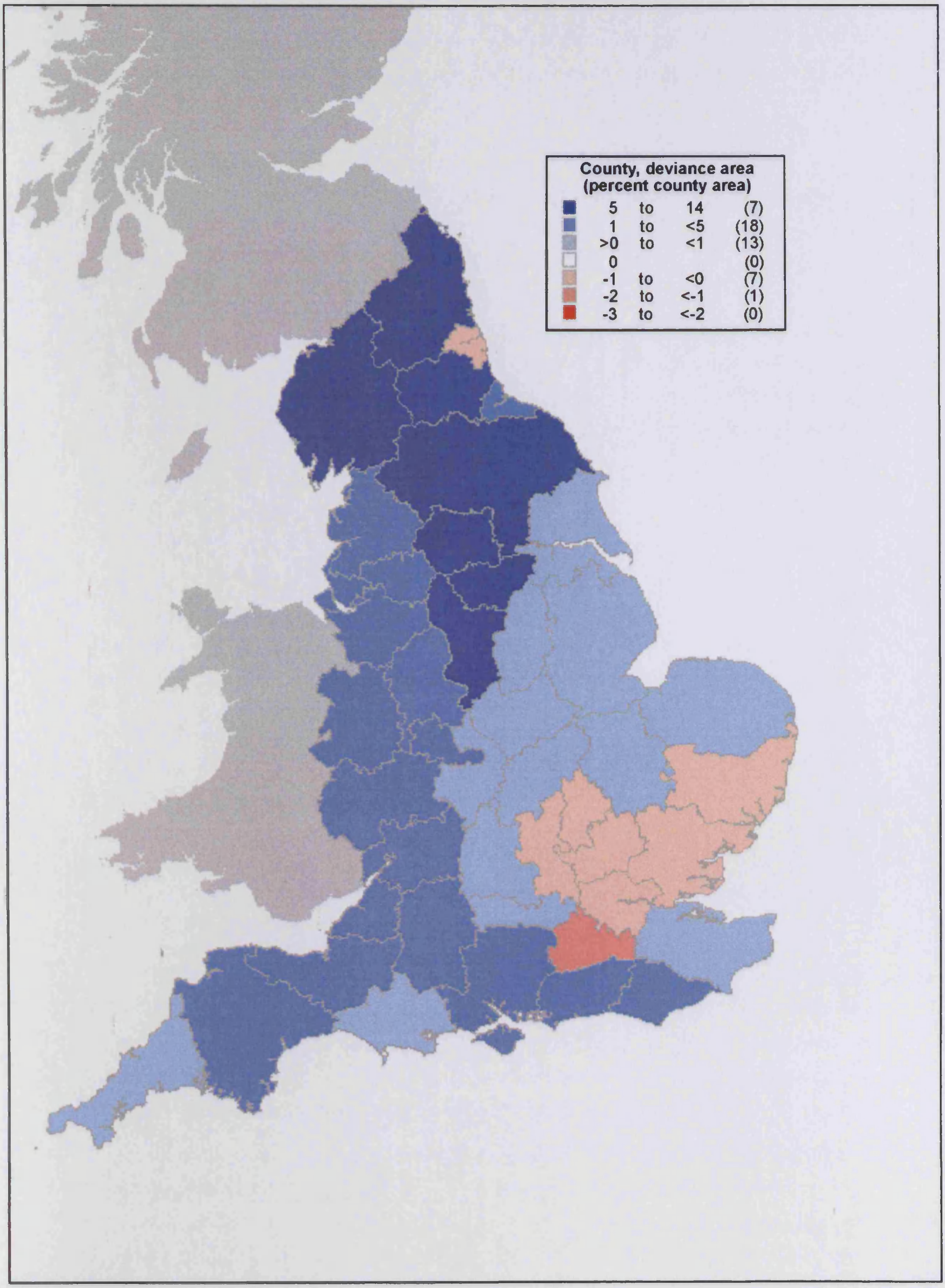


Figure 5.18 Lowland heathland county level, LCMGBCZ - EN error map, percentage of county area.

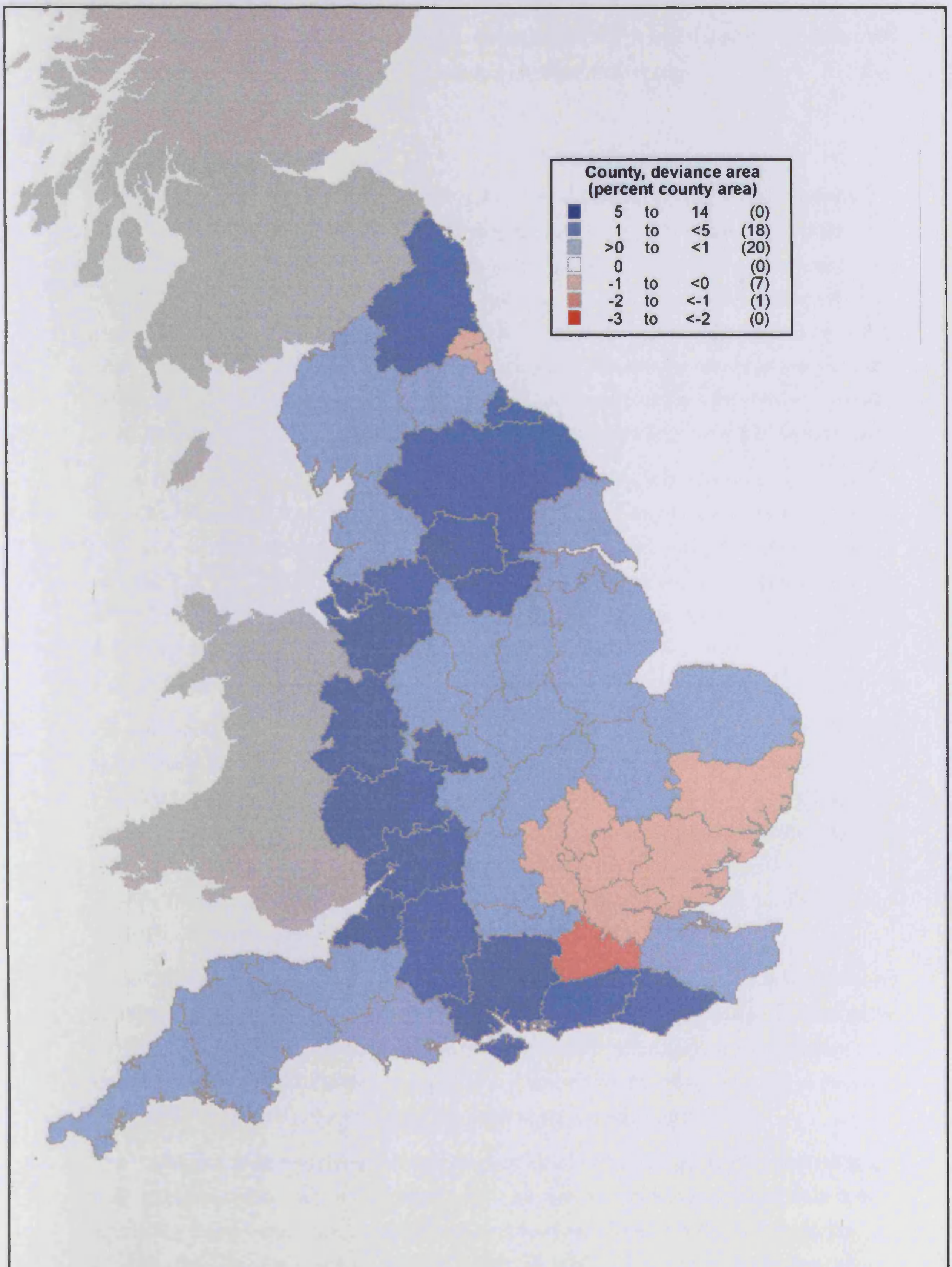
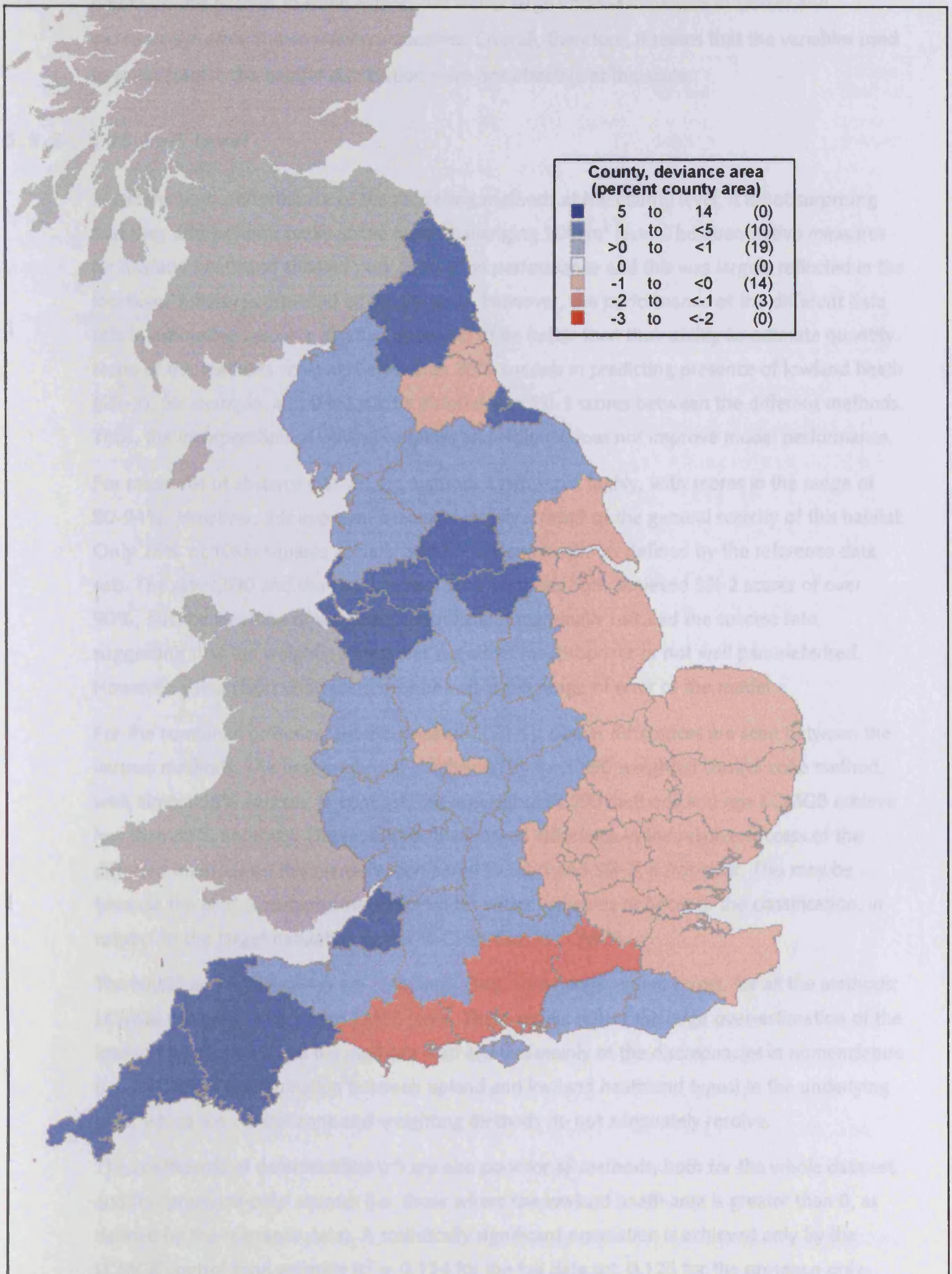


Figure 5.19 Lowland heathland county level, IW - EN error map, percentage of county area.



because the control variable used is insufficiently rigorous. The use of the weighting methods breaks up this pattern to some extent, but leaves large under-estimations in Dorset and increases the error in many eastern counties. Overall, therefore, it seems that the variables used to try to resolve the habitat distribution were not effective at this scale.

5.3.3 100 km^2 level

Given the poor performance of the modelling methods at the county level, it is not surprising that they also perform badly at the more challenging 100km^2 level. The quantitative measures for lowland heathland showed poor estimation performance and this was largely reflected in the locational measures provided by the SSI tests. However, the performance of the different data sets in estimating resource location appeared to be better than their ability to estimate quantity. None of the methods achieves better than 30% success in predicting presence of lowland heath (SSI-3), for example, and there is little difference in SSI-3 scores between the different methods. Thus, the incorporation of control variables and weights does not improve model performance.

For measures of absence (SSI-2), the methods again score highly, with scores in the range of 80-94%. However, this apparent success is largely a result of the general scarcity of this habitat. Only 25% of 10km squares actually contain lowland heath (as defined by the reference data set). The raw CS90 and the CS90 control zone methods both achieved SSI-2 scores of over 90%, but the inclusion of weighted control zones marginally reduced the success rate, suggesting that the weighting measures are either inappropriate or not well parameterised. However, these slight differences may be within the range of error of the model.

For the combined presence/absence measure (SSI-1), clearer differences are seen between the various methods. The best prediction is achieved by the CS90 weighted control zone method, with almost 58% success. In contrast, the unweighted CS90 methods and raw LCMGB achieve less than 33% accuracy. The reason for the marked difference in the relative success of the different methods on this measure, compared to SSI-3 and SSI-2, is not clear. This may be because the land class-based system provides better estimates or because the classification, in respect to the target habitat, is tighter in CS90 than in LCMGB.

The NMSE and RMSE scores are high, indicating large measurement errors, for all the methods; LCMGB produced a very high RMSE score. These results reflect the large over-estimation of the lowland heath area by all the methods – an artefact mainly of the discrepancies in nomenclature (i.e. the lack of discrimination between upland and lowland heathland types) in the underlying data, which the control zone and weighting methods do not adequately resolve.

The coefficients of determination (r^2) are also poor for all methods, both for the whole data set, and for 'presence-only' squares (i.e. those where the lowland heath area is greater than 0, as defined by the reference data). A statistically significant association is achieved only by the LCMGB control zone estimate (r^2 = 0.124 for the full data set, 0.125 for the presence-only squares).

Table 5.7 Lowland heathland 100km^2 validation: Mean Predicted, Mean Measured, NMSE, RMSE, r^2 and SSI.

DATA	$\bar{P}$ (ha)	$\bar{M}$ (ha)	NMSE (ha)	RMSE (ha)	r^2	r^2 (Ind>0)	SSI-1	SSI-2	SSI-3
CS90 EN v CS90	230.28	19.03	80.85	595.27	0.001	0.001	28.65% 426/1487	92.42% 61/66	25.69% 365/1421
CS90 Control Zone EN v CS90CZ	50.39	19.03	53.69	226.90	0.000	0.000	31.54% 469/1487	93.81% 106/113	26.42% 363/1374
CS90 Weighted Control Zone EN v CS90WCZ	50.39	19.03	106.52	319.62	0.000	0.000	57.70% 858/1487	79.61% 656/824	30.46% 202/663
LCMGB EN v LCMGB	245.09	19.03	95.02	665.71	0.005	0.004	32.08% 477/1487	86.90% 126/145	26.15% 351/1342
LCMGB Control Zone EN v LCMGBCZ	86.38	19.03	30.79	225.00	0.124	0.125	36.52% 543/1487	89.50% 196/219	27.37% 196/219
Intelligent Weighting EN v CS90-LCMGB/CZW	50.39	19.03	58.13	236.09	0.010	0.009	42.97% 639/1487	89.44% 305/341	29.14% 334/1146

The Lowland Heathland Inventory map (Figure 5.20) shows the distribution of lowland heathland according to the reference data set. The distribution this reveals is widespread but localised and patchy. Lowland heathland occurs in 370 of the 1487 10km grid squares (25%). The maximum quantity of lowland heathland within any 10km grid (100km^2) is 7,700ha. This represents a maximum possible cover of 77%. The highest concentrations of heathland are in Surrey and Hampshire, Dorset and Cornwall. There are also high concentrations in the West Midlands Plateau, Staffordshire, and Shropshire and in East Anglia in the Brecks, North Norfolk and the Suffolk coast, and south of the Humber Estuary. Surrounding these areas are other low-level dispersed occurrences of lowland heathland. There is a NE/SW running belt from the Wash to the Severn Estuary that is almost completely devoid of lowland heathland. In the north of England, the Pennines area including much of North Yorkshire and Cumbria also has very little habitat.

Figure 5.21 shows the error map for CS90 data: for each 10km grid square the reference data value has been subtracted from the estimated value. Positive (blue-shaded) squares therefore show an overestimate and negative values (red-shaded) show an underestimate. The most striking feature of the CS90-based map is the large number and wide distribution of grid squares in which the data predict habitat presence. As with saltmarsh, this stems from the structure of CS90 data that produces land class-based estimates from intra-land class samples. This issue is compounded by nomenclature issues: we know that the CS90 estimate included upland heath categories. This is clearly shown by the map as areas of overestimation in relatively high altitude areas such as the Pennines, Cotswolds, Wiltshire and Hampshire Downs, Exmoor, Dartmoor and the Quantocks. The majority of northern England is categorised as an overestimate for lowland heathland. CS90 manages correctly to estimate part of the Wash/Severn belt area as containing little lowland heathland, but produces overestimates at the western side of the area. The map is seen to be skewed towards overestimation, reflecting the overall overestimation of the national stock of lowland heathland.

Figure 5.20 Lowland heathland 100km^2 square, EN LHI reference data, area ha^{-1}.

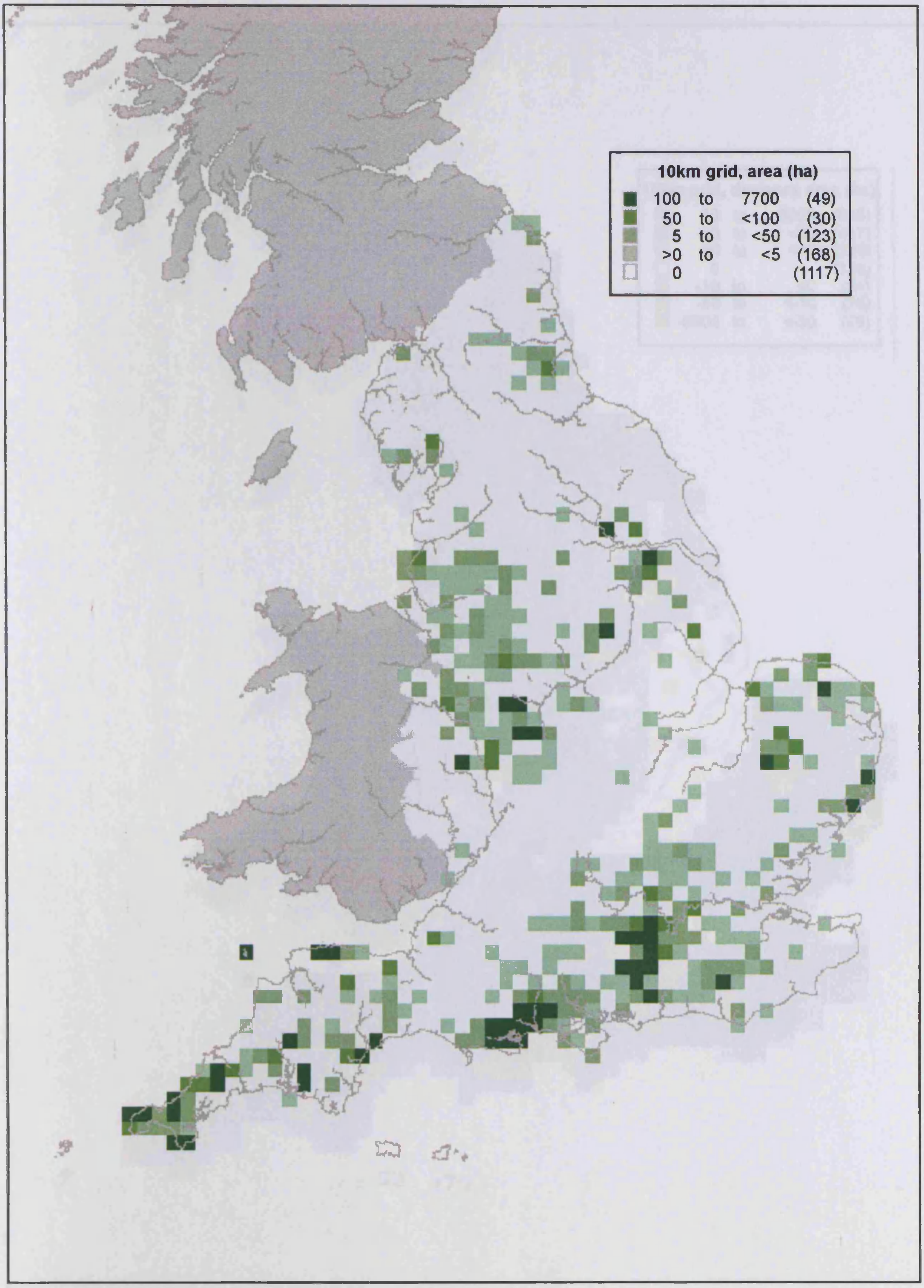


Figure 5.21 Lowland heathland 100km² square, CS90- EN error map, area ha⁻¹.

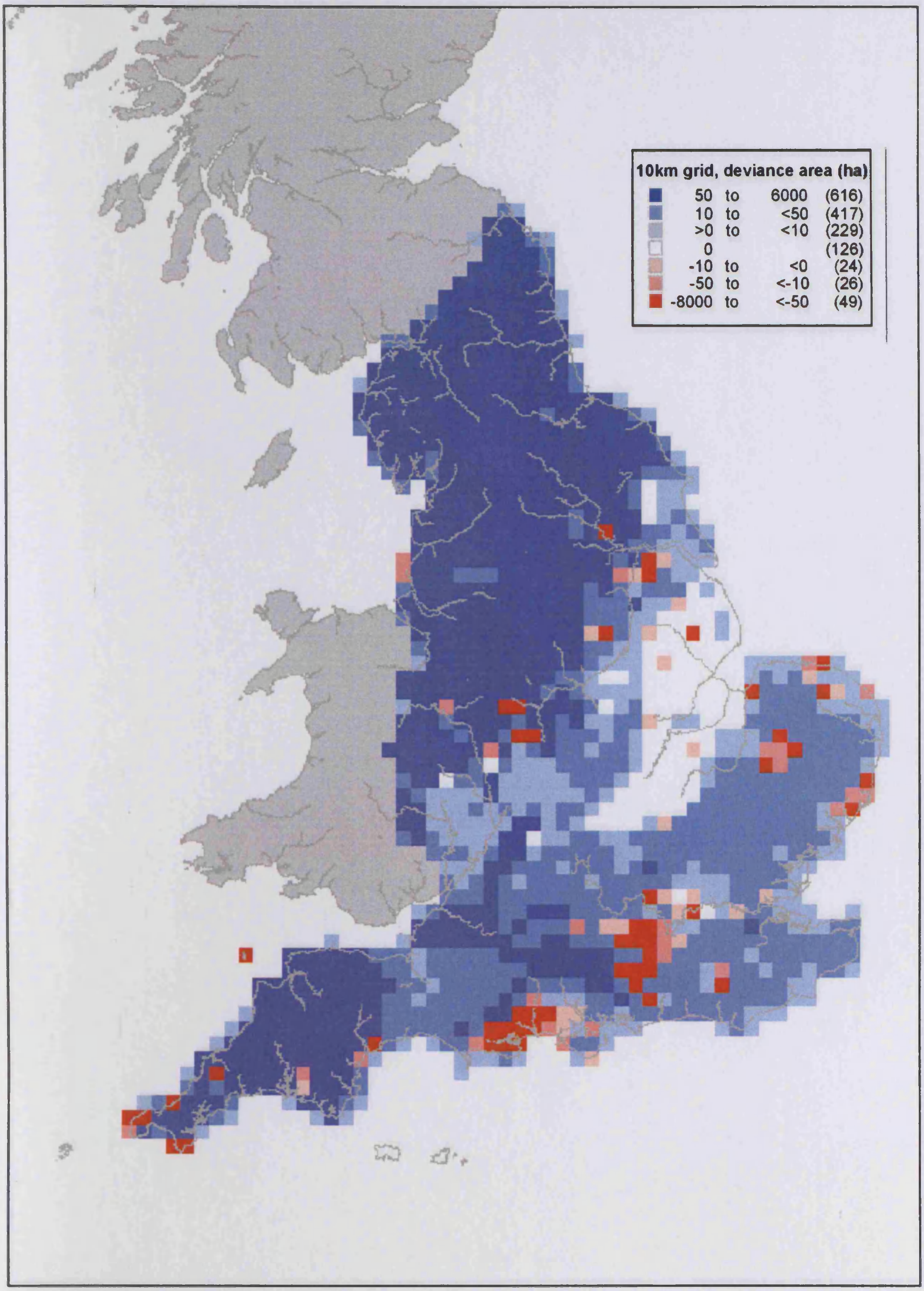


Figure 5.22 Lowland heathland 100km² square, CS90CZ-EN error map, area ha⁻¹.

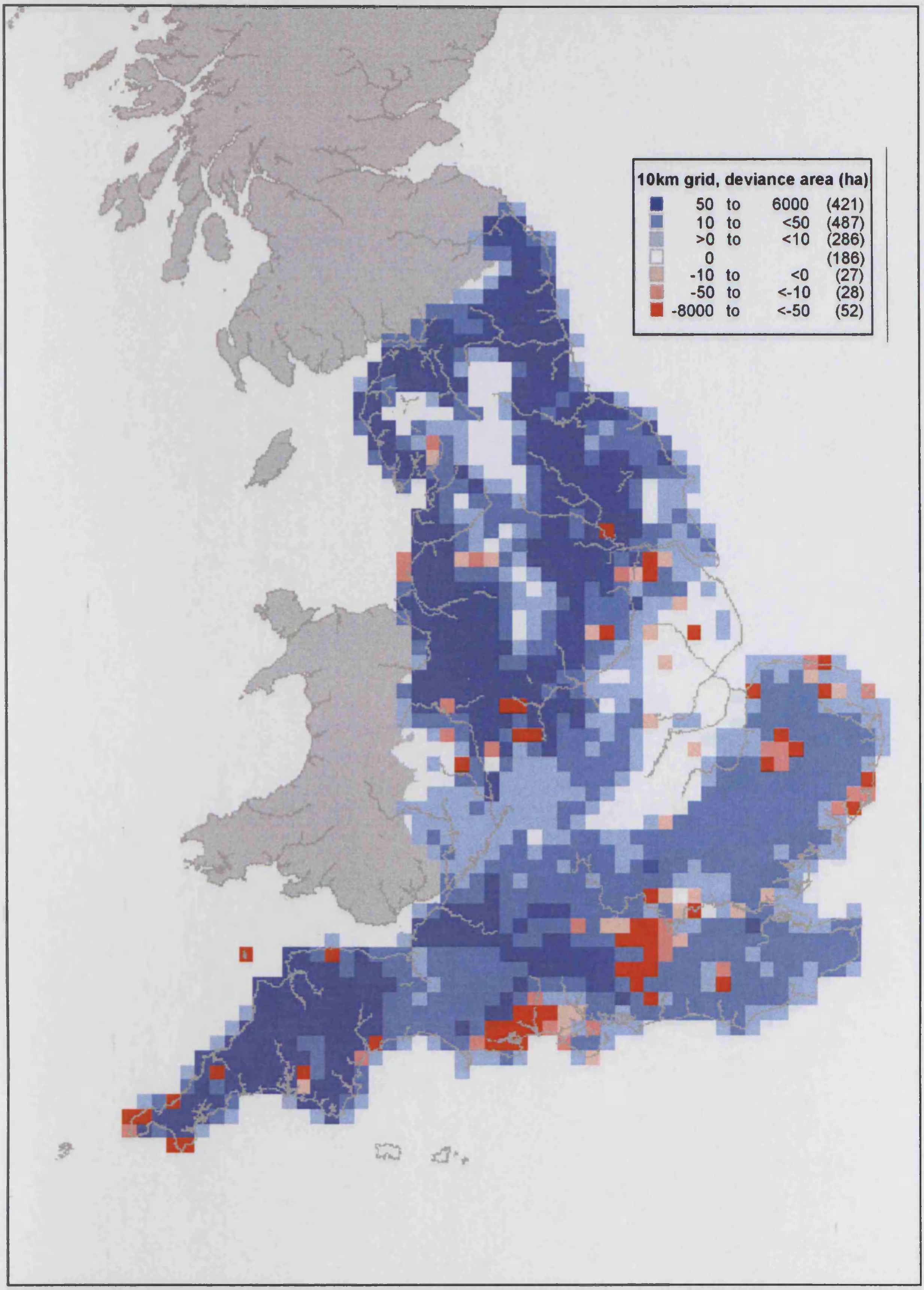


Figure 5.23 Lowland heathland 100km^2 square, CS90WCZ-EN error map, area ha^{-1}.

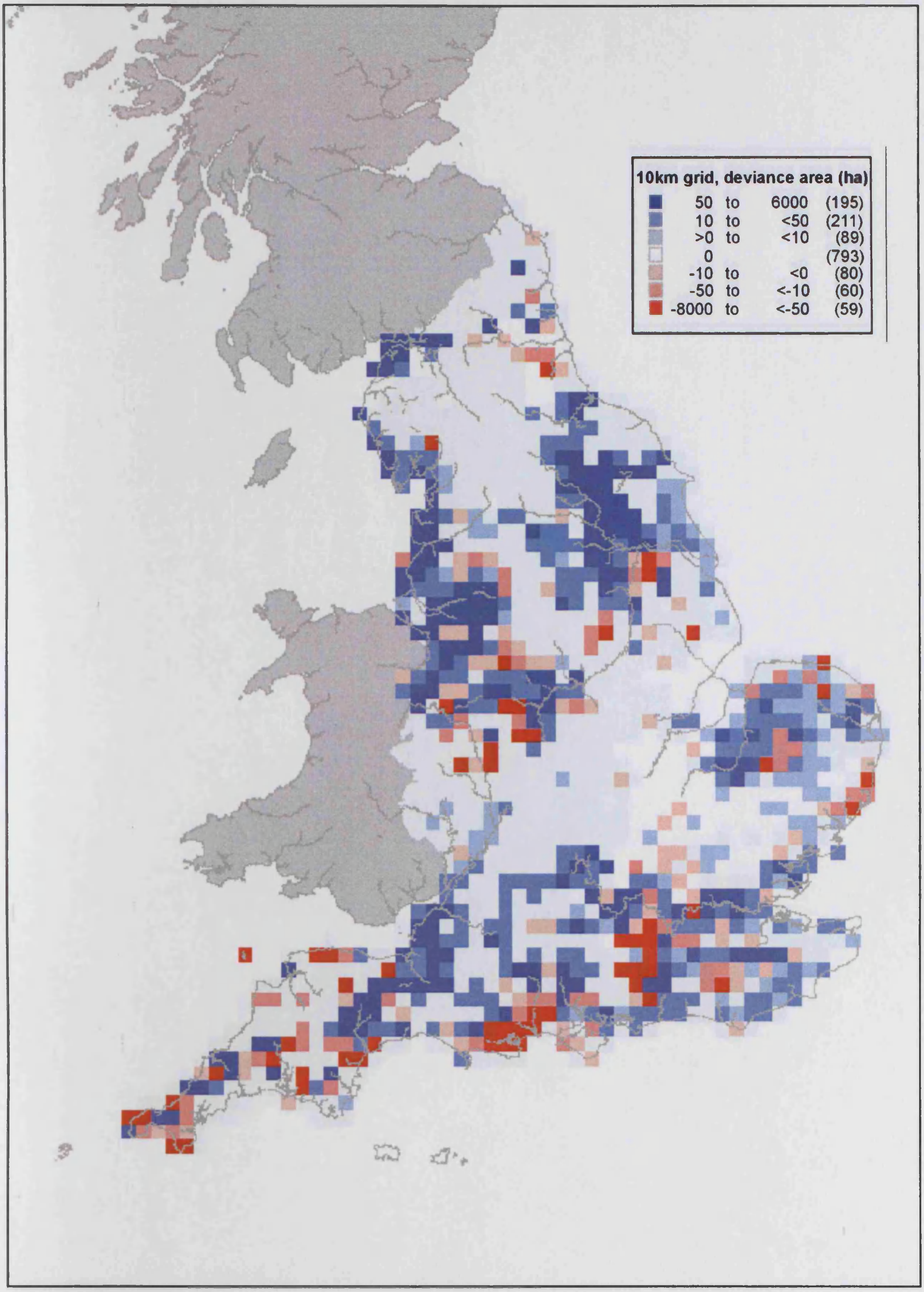


Figure 5.24 Lowland heathland 100km² square, LCMGB-EN error map, area ha⁻¹.

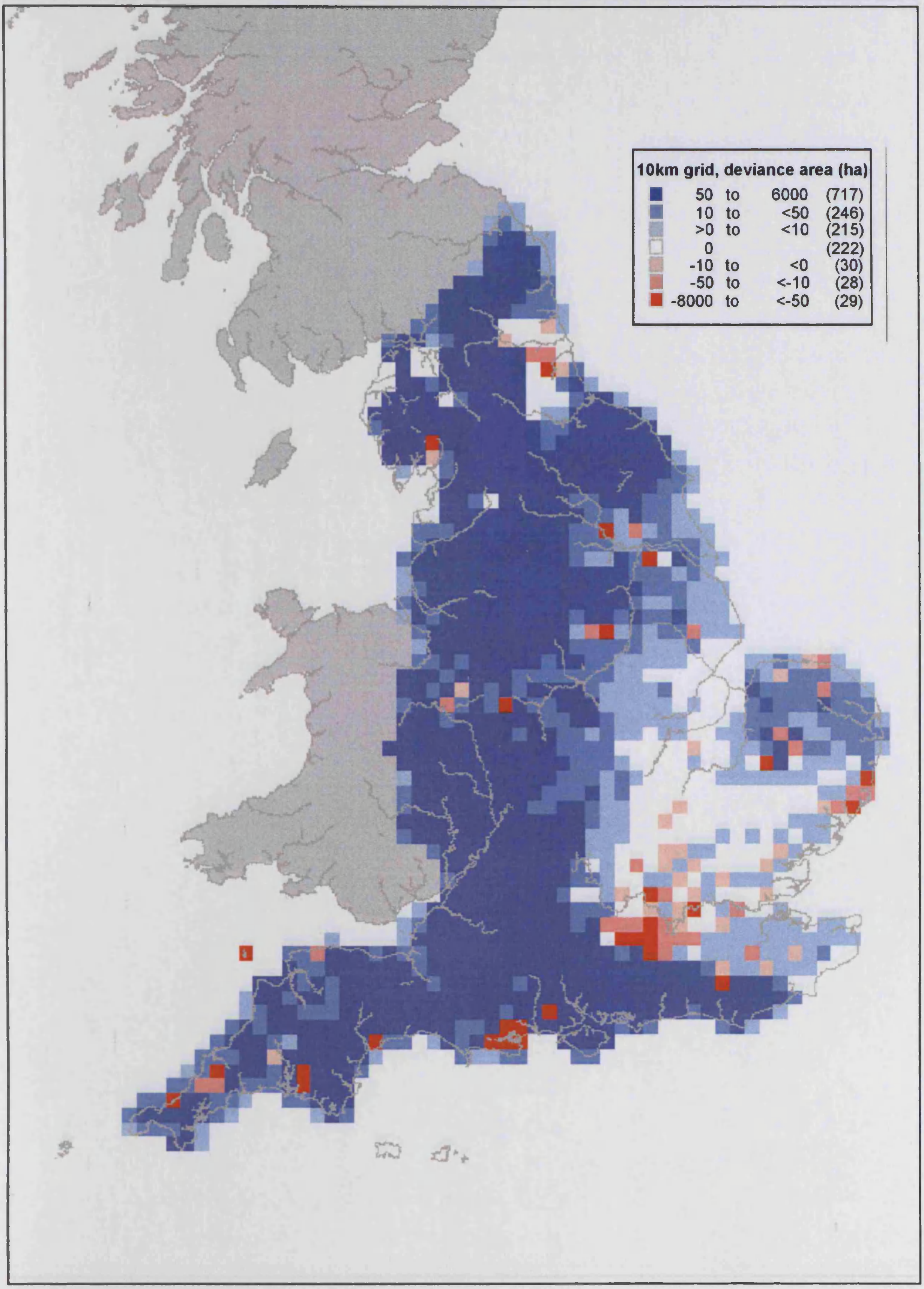


Figure 5.25 Lowland heathland 100km² square, LCMGBCZ-EN error map, area ha⁻¹.

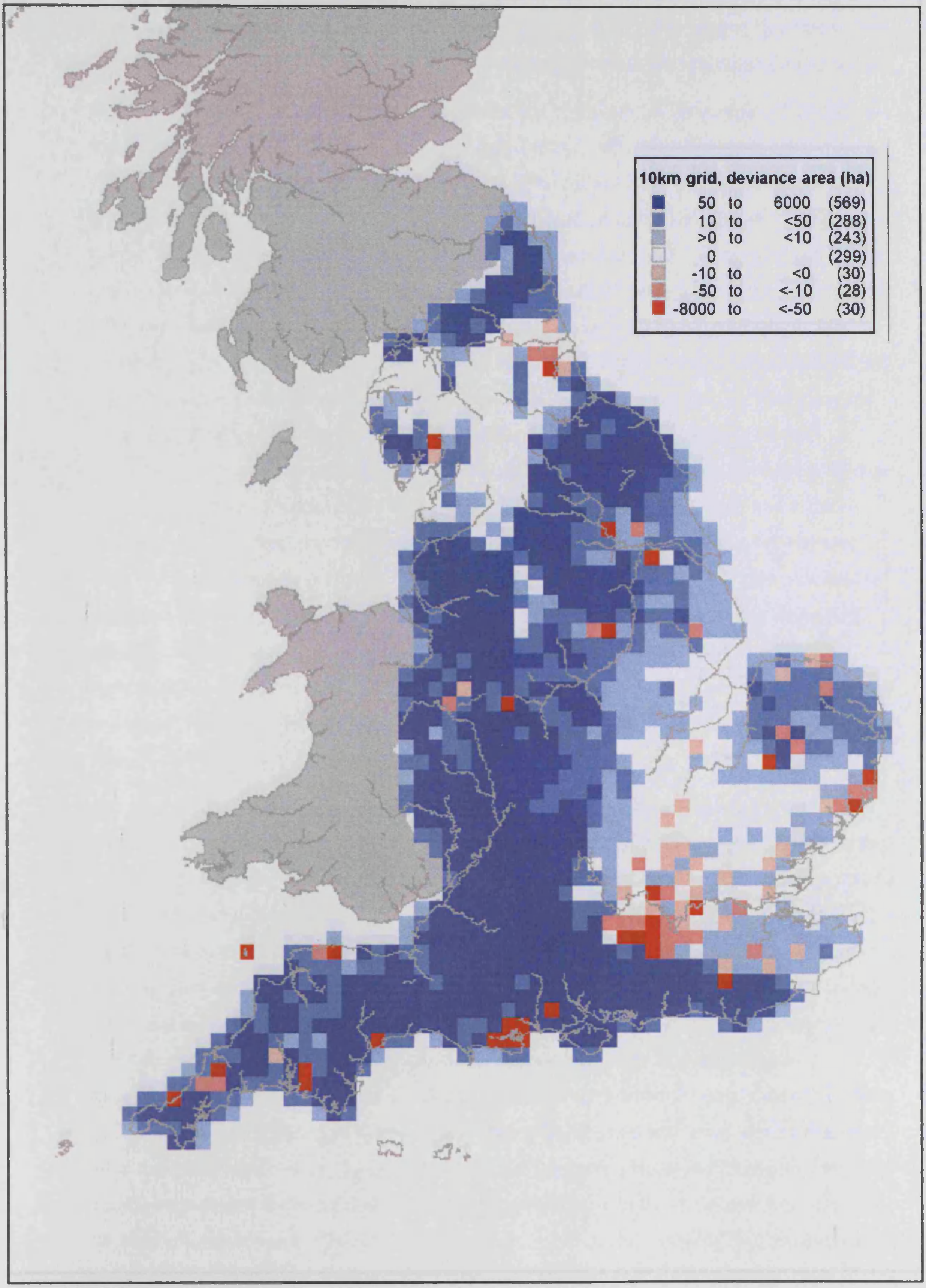
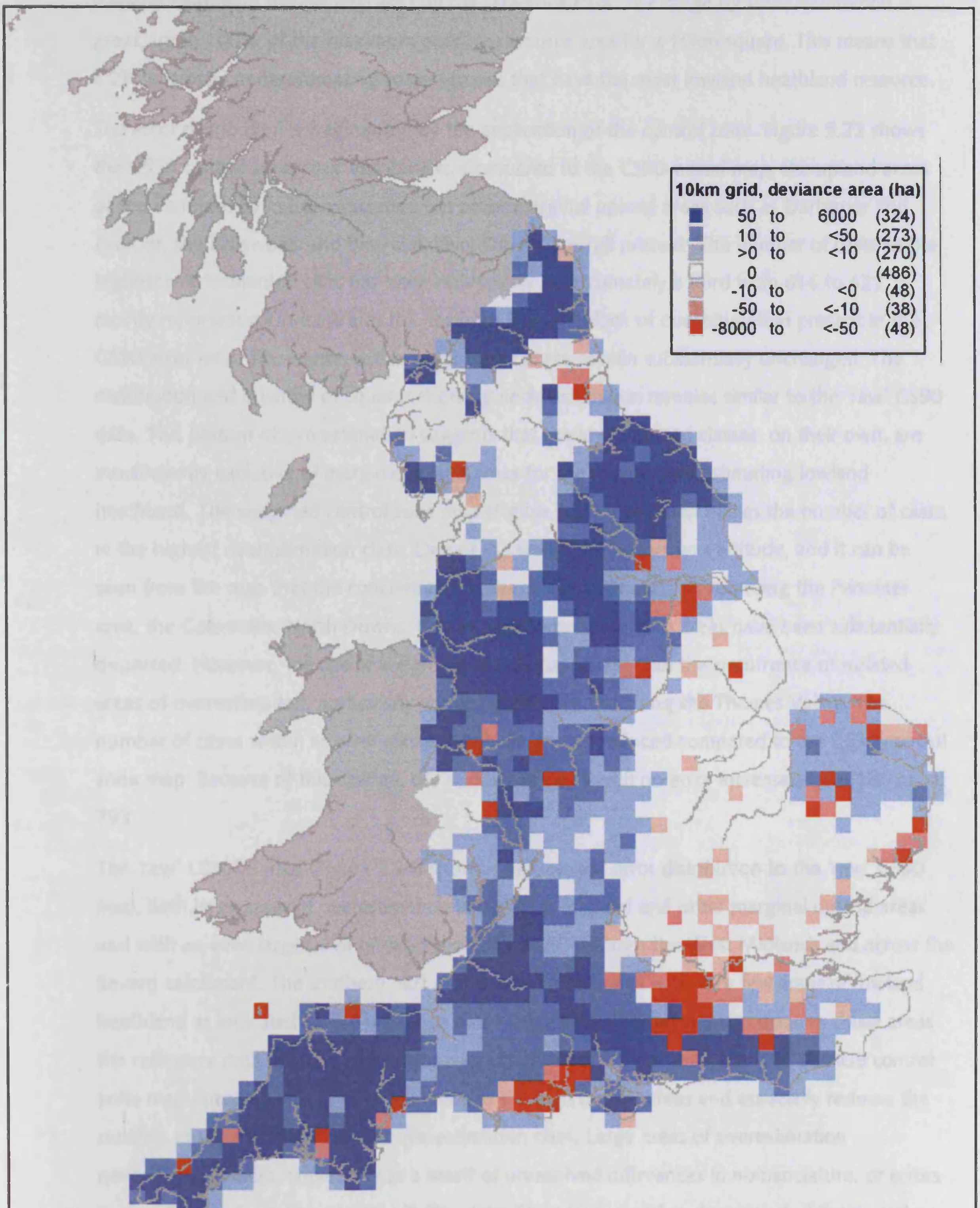


Figure 5.26 Lowland heathland 100km² square, IW-EN error map, area ha⁻¹.



The few areas of notable underestimation appear to coincide with the major concentrations of lowland heathland habitat indicated by the reference data. The range for underestimation is great, up to 100% of the maximum possible resource area for a 10km square. This means that CS90 is grossly underestimating some squares that have the most lowland heathland resource.

The error distribution is fragmented by the application of the control zone. Figure 5.22 shows the CS90 control zone error distribution. Compared to the CS90-based map, the upland areas of the Pennines have been excluded but other marginal upland areas such as Dartmoor and Exmoor, the Cotswolds and the Hampshire Downs are still present. The number of cases in the highest overestimation class has been reduced by approximately a third from 616 to 421, mainly representing inroads into the northern England block of overestimation present in the CS90 error map. The counts within other class ranges remain substantially unchanged. The distribution and quantity of squares showing underestimation remains similar to the 'raw' CS90 data. This pattern of overestimation suggests that the lowland land classes, on their own, are insufficiently exclusive of marginal upland areas for the purposes of estimating lowland heathland. The weighted control zone map (Figure 5.23) more than halves the number of cases in the highest overestimation class. One of the weighting factors was altitude, and it can be seen from the map that the concentrated areas of overestimation surrounding the Pennines area, the Cotswolds, North Downs, Devon, and Somerset upland areas have been substantially dispersed. However, the use of weighting factors has also spread the occurrence of isolated areas of overestimation, particularly within East Anglia and along the Thames Valley. The number of cases within all error classes has been greatly reduced compared to the CS90 control zone map. Because of this change, the number of cases with no error increased from 186 to 793.

The 'raw' LCMGB map (Figure 5.24) shows a very similar error distribution to the 'raw' CS90 map, with large areas of overestimation in northern England and other marginal upland areas and with an even larger area of overestimation south through the West Midlands and across the Severn catchment. The southern part of this area corresponds with very few areas of lowland heathland as indicated by the reference data. Large underestimations also occur in those areas the reference data define as the major concentrations of habitat resource. The LCMGB control zone map reduces areas of overestimation in the high upland areas and especially reduces the number of squares in the highest overestimation class. Large areas of overestimation nevertheless remain, apparently as a result of unresolved differences in nomenclature, or errors in classification from the original satellite data. The areas of underestimation and their number of occurrences appear to be changed little by use of the control zone. The intelligent weighting treatment increases the number of squares with no error by a factor of 1.6 and decreases the number of cases in the highest overestimate range by 245 (44%). Nevertheless, the map also shows that the number of underestimates increases. The geographical distribution suggests that the weighting by CS90 data improves performance in predicting resource absence but not presence - a hypothesis which is supported, albeit only weakly, by the SSI results.

5.3.4 1 km² level

In view of the preceding results, poor levels of prediction can be expected at the 1km² level for all methods, and as Table 5.8 shows, this is indeed the case. SSI-3 values are consistently less than 3%, indicating extremely poor capability to detect the presence of lowland heath habitats. The consistently high scores for SSI-2 merely show that – for a scarce habitat such as this – detecting absence is not difficult. The SSI-1 scores, however, do show interesting variation. Values range from 28% for the raw CS90 method to 94% for the CS90 weighted control zone method; the methods based on LCMGB range from 48% for the raw data to 66% for the intelligent weighting approach. For both the CS90-based measures and LCMGB, the control zone improved the SSI-1 scores.

Table 5.8 Lowland heathland 1km² validation: Mean Predicted, Mean Measured, NMSE, RMSE, r^2 and SSI.

DATA	$\bar{P}$ (ha)	$\bar{M}$ (ha)	NMSE (ha)	RMSE (ha)	r^2	r^2 (Ind>0)	SSI-1	SSI-2	SSI-3
CS90 EN v CS90	2.56	0.21	129.20	8.33	0.000	0.000	27.5% 36591/132993	99.08% 35134/35455	1.48% 1457/97538
CS90 Control Zone EN v CS90CZ	0.55	0.21	126.66	3.82	0.000	0.000	38.33% 50979/132992	99.24% 49578/49955	1.69% 1401/83038
CS90 Weighted Control Zone EN v CS90WCZ	0.55	0.21	1258.80	12.06	0.000	0.000	94.38% 125523/132992	98.67% 125439/127132	0.18% 84/45815
LCMGB EN v LCMGB	2.73	0.21	189.96	10.44	0.003	0.002	48.16% 64048/132992	49.51% 62582/62893	2.09% 1466/70099
LCMGB Control Zone EN v LCMGBCZ	0.96	0.21	95.64	4.39	0.043	0.041	57.53% 76518/132993	99.51% 75112/75484	2.43% 1406/57509
Intelligent Weighting EN v CS90-LCMGB/CZW	0.55	0.21	214.04	4.97	0.004	0.003	65.95% 87706/132992	99.28% 86553/87177	2.52% 1153/45815

NMSE and RMSE scores are also very variable between the different methods. The lowest NMSE score is provided by the LCMGB control zone estimate, which also returns a low RMSE score. The CS90 weighted control zone estimate produces a very high NMSE and also the highest RMSE value. The RMSE scores show that the control zone reduces mean errors for CS90 and LCMGB estimates and that these two treatments perform best by this measure.

Coefficient of determination (r^2) values are consistently very low and non-significant, both for the data set as a whole and the presence-only squares, and show that none of the methods provides a reliable measure of the area of the habitat at this scale. There is little difference between the r^2 values generated on all data and the second column r^2 values generated for cases where the independent variable > 0, as is to be expected given the inherently low level of correlation.

5.3.5 Summary

Overall it is evident that the distribution of lowland heathland is poorly predicted and estimated by all the methods used here. To a large degree this reflects the much more inclusive nomenclature employed by the CS90 and LCMGB data sets, neither of which discriminates between lowland and upland heathland types. Predictions from all the methods are thus 'contaminated' by the inclusion of upland heath vegetation types.

Inclusion of a control zone (based on land class) provides some improvement to the estimates at all three scales of analysis, though at finer scales especially the performance of the models is still poor, particularly for the correct estimation of resource presence. The use of weighting methods (incorporating major soil type and altitude into the CS90 models), fails to improve predictions, however, and in practice slightly worsens model performance. Combination of the CS90 and LCMGB data (in the intelligent weighting method) also fails to improve on the control zone methods.

Several factors might be adduced to explain this pattern of results. One is that the lowland land classes used as a control in this analysis are insufficiently discriminatory for lowland heathland vegetation types and were still contaminated by 'leakage' from upland heath vegetation types. This would account for the over-estimation of the total lowland heath area. Lowland land classes were chosen, in preference to using an altitudinal limit, because they are also designed to reflect north/south and east/west climatic variation. The error maps at the 100km^2 scale, however, show a tendency for overestimation in the west and north of the country. These results would tend to suggest that the target vegetation types still included vegetation with upland/western characteristics. The relatively low presence (SSI-3) scores at both 100km^2 and 1km^2 scales is more likely to be due to failure to detect lowland heathland vegetation types. In addition, it should not be forgotten that there is also likely to be error in the reference data and some lowland heathland ecologists believe that the Lowland Heathland Inventory is an underestimate of the resource (Dr N. Michael, pers. comm. 1998). Because the inventory did not provide boundaries for lowland heathland areas, it was also necessary in applying these data here to model habitat extent by simple buffering around the centroid. This, also, will have contributed to errors, especially at the 1km^2 scale.

Another likely source of error relates to modelled versus actual distributions. If the LCMGB contained a more precise lowland heathland category, we might expect to be able to map the habitat successfully. However, lowland heathland represents a complex and 'moving' target. The vegetation composition varies east-west and north-south. Lowland heathland in Cornwall is very different to the Breckland heathland in Norfolk and Suffolk. Additionally, lowland heathland comprises an intimate mosaic of other habitat types, including scrub, acid grassland mosaics and valley mire. The identification of these, either in their own right, or as members of a biotope complex, is problematic and highly scale-dependent. In environmental terms, the occurrence of lowland heathland is dependent upon highly sensitive local variation in soil and drainage characteristics. Distribution is therefore related to localised drift geology and edaphic factors. This level of detail is not resolvable with the relatively coarse data employed here. Lowland heathland is also a remnant land use. It remains not so much where there is environmental potential for it, but where it has not been removed and where there has been sufficient historical continuity of land use. The current distribution of lowland heathland is therefore to a large extent 'incidental' compared to its possible environmental range. Modelling this distribution on the basis of simple environmental and deterministic concepts, using coarse-scale data, is therefore extremely challenging and unlikely to succeed. For this approach to be successful it is likely to require more sophisticated techniques, perhaps employing more locally detailed data and incorporating land use-related variables.

5.4 WOODLAND

5.4.1 Quantity/area

Woodland habitat is far more abundant and widespread than either of the previous target habitats. The totals produced by each data set are shown in Table 5.9. The Forestry Commission (FC) data were not employed in the analysis (see Methods chapter) but are included here as an additional point of comparison. The Ordnance Survey (OS95Gref) data, used as the reference, produced a total very close to the Forestry Commission total and to that derived from the LCMGB data. The CS90 data, however, produced a total over 660,000ha (25%) higher. The Forestry Commission and Ordnance Survey totals could be considered to represent an underestimate due to their use of a threshold woodland size for inclusion, but no other data were available to quantify this effect.

In contrast to the other habitats, nomenclature issues were less apparent, partly because woodland is more easily and consistently defined, but perhaps also because the differences are subtle and have simply become hidden by the use of generalised descriptions. With the exception of their broad categorisation into broadleaf or coniferous, the Forestry Commission's categories are structurally rather than compositionally based. For CS90 and LCMGB, woodland types were condensed into four and two categories respectively. Because these represented 'many-to-many' relationships with the Forestry Commission and the reference data, there was little scope for assessing nomenclature-based effects.

The control zone was based upon extreme altitude and developed areas, but in contrast to the other target habitats was relatively small in area.

Table 5.9 Total woodland habitat areas for GB by data set.

Data set	Sum of woodland areas/ha
FC (Excluding 'cleared' total)	2,063,340.0
OSGRef (Reference)	1,930,528.3
CS90	2,593,714.2
CS90 Control Zone	2,593,714.2
CS90 Weighted Control Zone	2,593,714.2
LCMGB	1,979,844.5
CS90LCMGB Intelligent Weighting	2,593,714.2

5.4.2 County level

In contrast to the two previous habitats, woodland is both widespread and relatively common and therefore it should be relatively easy to detect.

Reflecting this, the NMSE score for CS90 is low but is not influenced by the application of the control zone or weighting factors. The LCMGB data provides a higher score and the intelligent weighting returns a value similar to the other CS90-based estimates. The RMSE values show a similar pattern to the NMSE results. Overall, the weighting factors had a negligible effect on any of the quantitative or locational scores. This is because they apply to relatively few of the original 1 km squares and will not affect membership of land classes at the county scale.

The two sets of r^2 values were virtually identical (reflecting the circumstance that only a few counties are predicted as not containing woodland, so the full and presence-only data sets are essentially the same); both showed strong associations in all cases, suggesting that CS90 works well at this scale for woodland. LCMGB, alone, performs least well in relation to both the measures of mean error and the coefficient of determination. The intelligent weighting method, in contrast, has the best scores in relation to these measures, suggesting that the CS90 data adds valuable information to that contained in LCMGB.

Table 5.10 Woodland County-level validation: Mean Predicted, Mean Measured, NMSE, RMSE, r^2 and SSI.

DATA	$\bar{P}$ (ha)	$\bar{M}$ (ha)	NMSE (ha)	RMSE (ha)	r^2	r^2 (Ind>0)	SSI-1	SSI-2	SSI-3
CS90 OS v CS90	38712.15	28813.85	0.374	20423.90	0.891	0.891	98.51% 66/67	N/A	98.48% 65/66
CS90 Control Zone OS v CS90CZ	38712.15	28813.85	0.374	20429.64	0.891	0.891	98.51% 66/67	N/A	98.48% 65/66
CS90 Weighted Control Zone OS v CS90WCZ	38712.15	28813.85	0.372	20378.17	0.891	0.891	98.51% 66/67	N/A	98.48% 65/66
LCMGB OS v LCMGB	29549.92	28813.85	0.975	28814.28	0.792	0.793	97.01% 65/67	0% 0/1	98.48% 65/66
Intelligent Weighting OS v CS90-LCMGBW	38712.15	28813.85	0.310	18606.05	0.915	0.916	97.01% 65/67	0% 0/1	98.48% 65/66

The SSI-1 scores are uniformly high for all treatments. The calculation of SSI-2 values, a measure of the correct estimation of absence, was not possible for the CS90-based data and yielded scores of zero for the other two treatments. This was because CS90 estimated habitat presence in all counties. For the LCMGB-derived estimates, these scores were zero because no counties were correctly estimated as having no woodland when, according to the reference data, one county (Shetland Isles) had no woodland present.

The county-based distribution of woodland is expressed as a percentage of county area to adjust for differences in county sizes. The reference data (Figure 5.38) evince major woodland concentrations in south-east England in East and West Sussex, Surrey, Hampshire, in northern England in Northumbria, in Wales in Powys, Mid- and West Glamorgan and in Scotland in Borders, Dumfries and Galloway, Strathclyde, Central, Tayside, Highland and Grampian Regions. Only the Shetland Islands possess no woodland and the SSI-3 results show that all treatments correctly estimated woodland presence in all but one case. There are likely to be some small patches of woodland in Shetland but these probably fall below the threshold size for the reference data set.

The CS90 error map (Figure 5.27) shows considerable underestimation in south east England and Wales and south western Scotland, all of which have high concentrations of woodland. On the other hand, the relatively highly wooded counties of northern and eastern Scotland and Northumberland are overestimated by CS90. Generally, the map overestimates woodland area in less wooded counties and underestimates the area in more wooded counties. This reflects the tendency of CS90 to produce relatively 'flat' maps, due to its distribution of habitat widely across complete land classes.

The application of the control zone at the county scale made no difference to category memberships and consequently the error map was identical to that for 'raw' CS90 data. Similarly the weighting factors had little effect. There were no transfers either side of the zero point and consequently the distribution map is substantially unchanged to the other CS90-based estimates. The LCMGB error map (Figure 5.30) shows a tendency for more error in the west. In contrast to the CS90-based data, the highly wooded counties in the south east of England are slightly overestimated rather than underestimated. This pattern is repeated in Wales for Powys and West Glamorgan. The LCMGB map produces underestimates for most regions in Scotland, compared to the reference data; these areas are those with large areas of woodland. This cannot easily be explained in terms of a broadleaf/coniferous split because the Welsh counties do not appear to display a pattern that would support this hypothesis. The intelligent weighting error map (Figure 5.31) slightly compresses the error range for underestimates with no cases in the lowest class. However, membership of the overestimate classes increased. The east/west bias apparent in the 'raw' LCMGB data is no longer evident and systematic error related to areas of high woodland occurrence has also been largely dispersed. However, there are still large amounts of overestimation, in eight cases of up to 12%, which is equivalent to half of the total possible maximum woodland amount in any one county.

Figure 5.27 Woodland county level, OS95GRef reference data, percentage of county area.

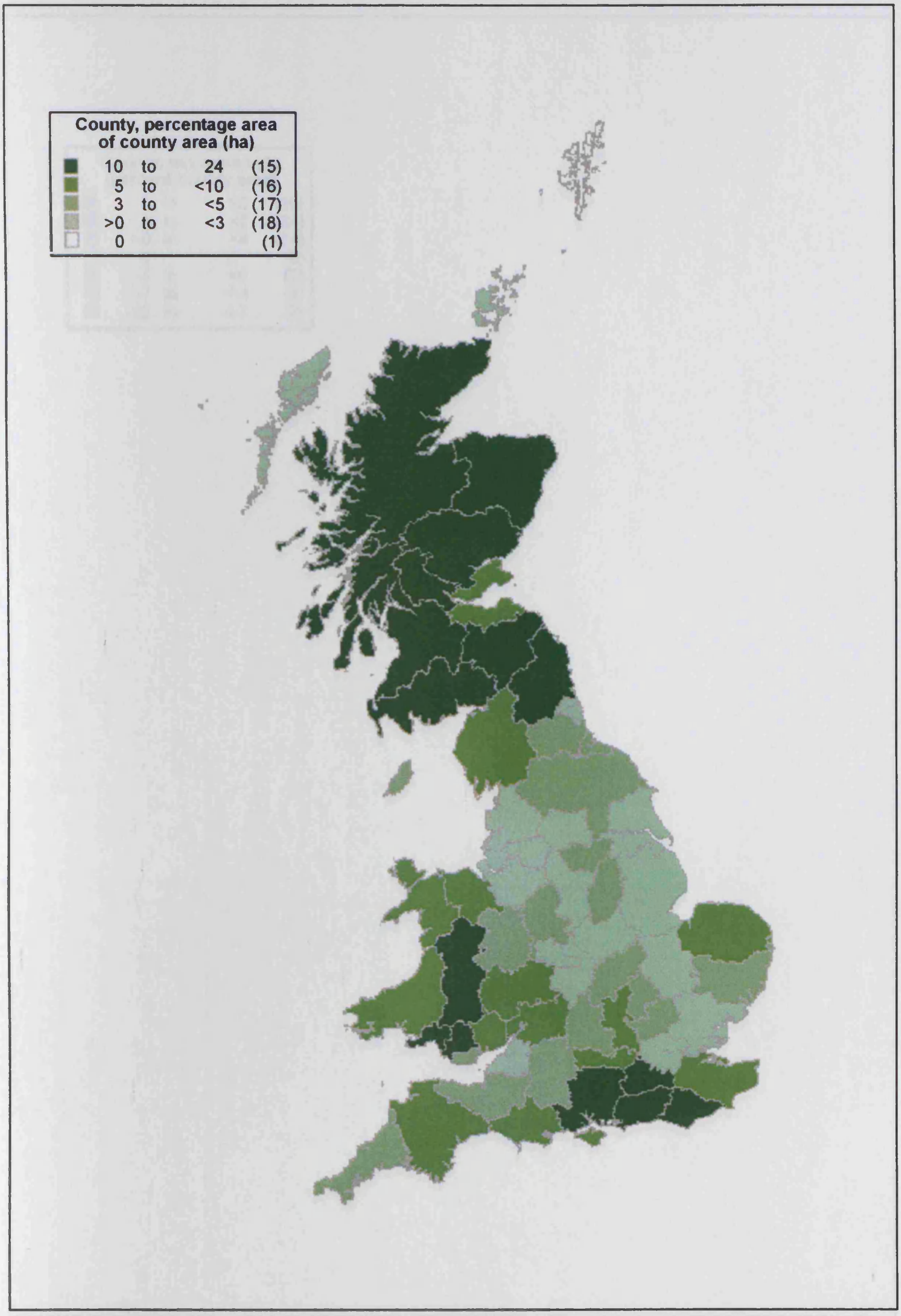


Figure 5.28 Woodland county level, CS90 – OS95GRef error map, percentage of county area.

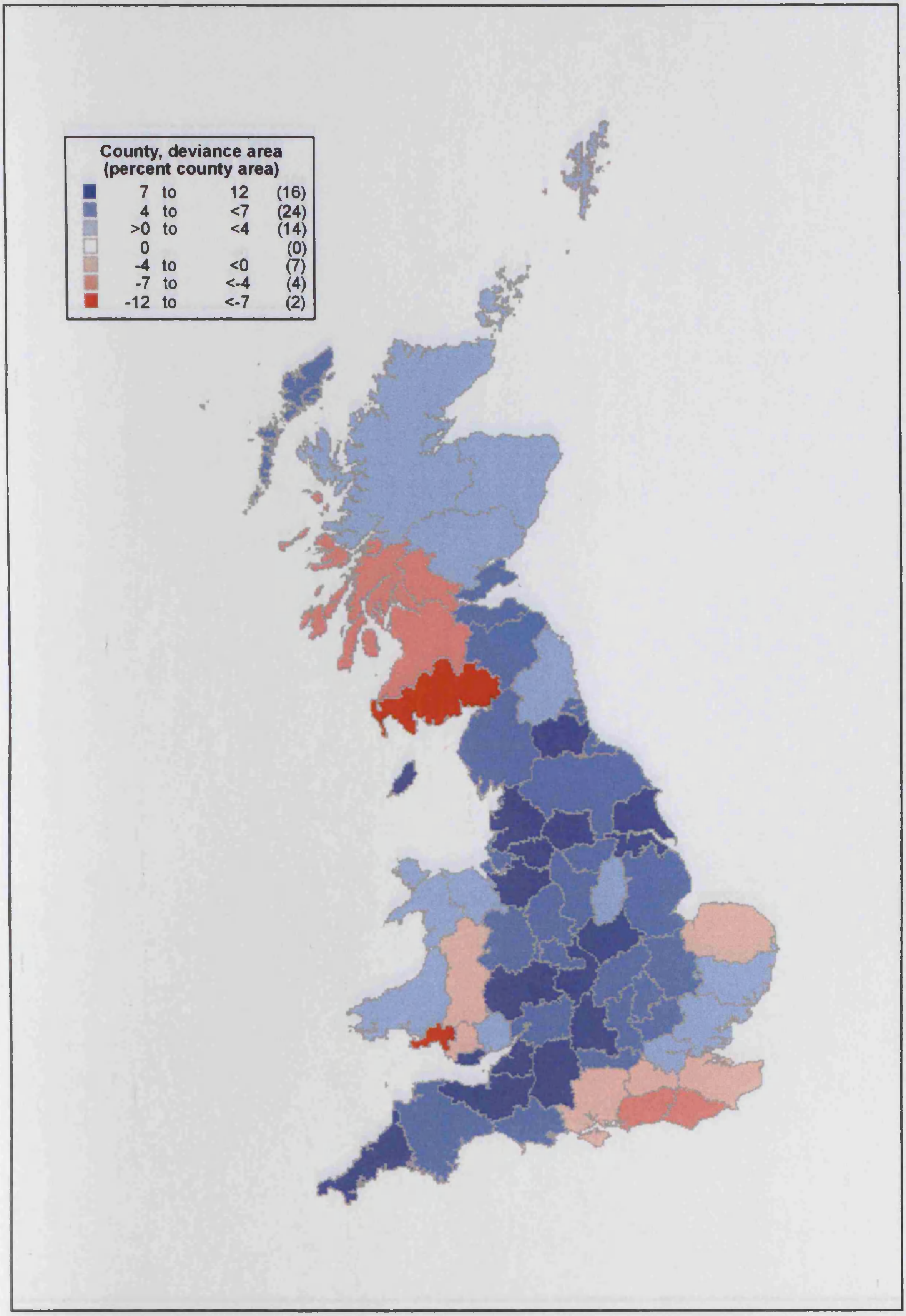


Figure 5.29 Woodland county level, CS90CZ – OS95GRef error map, percentage of county area.

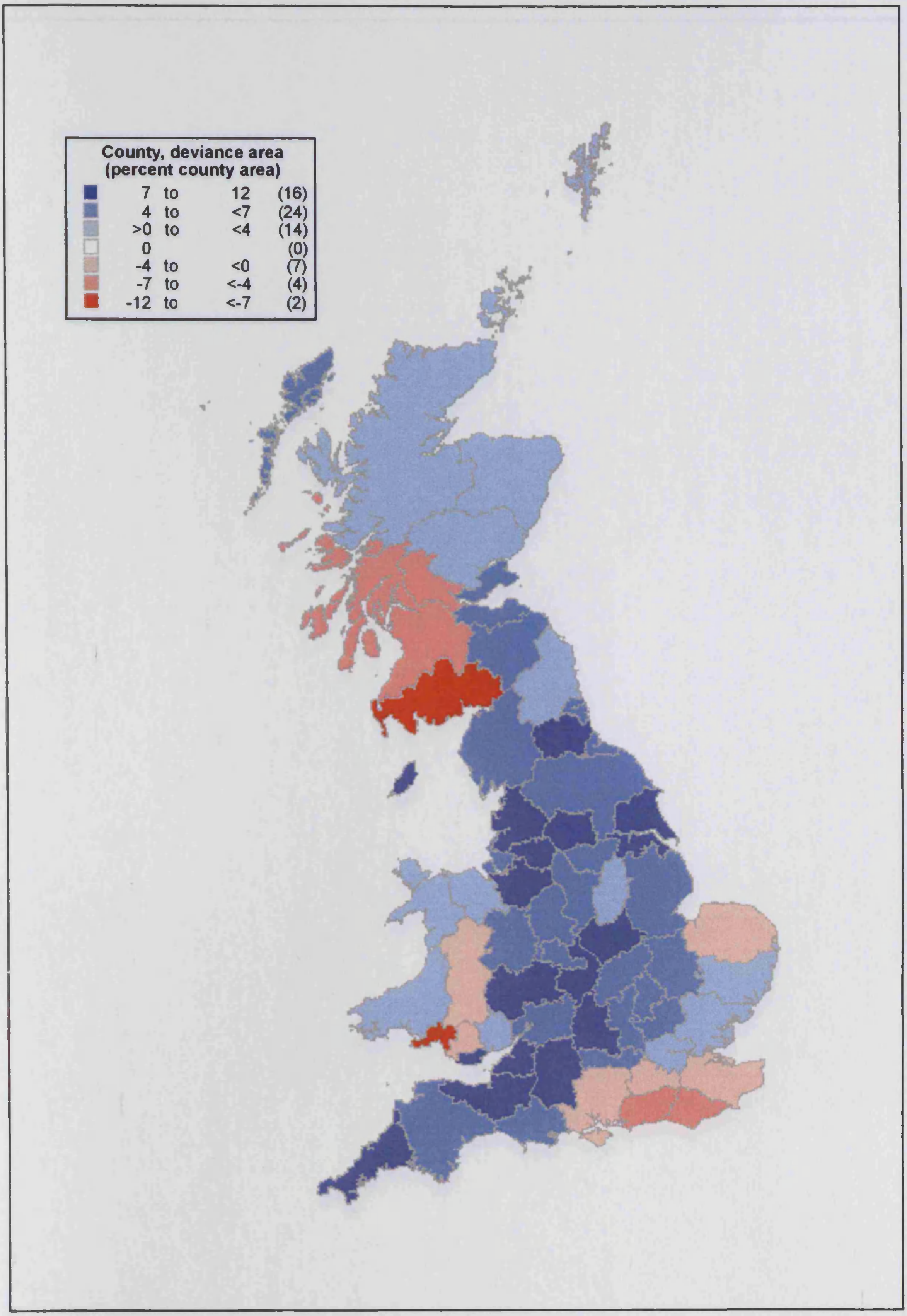


Figure 5.30 Woodland county level, CS90WCZ – OS95GRef error map, percentage of county area.

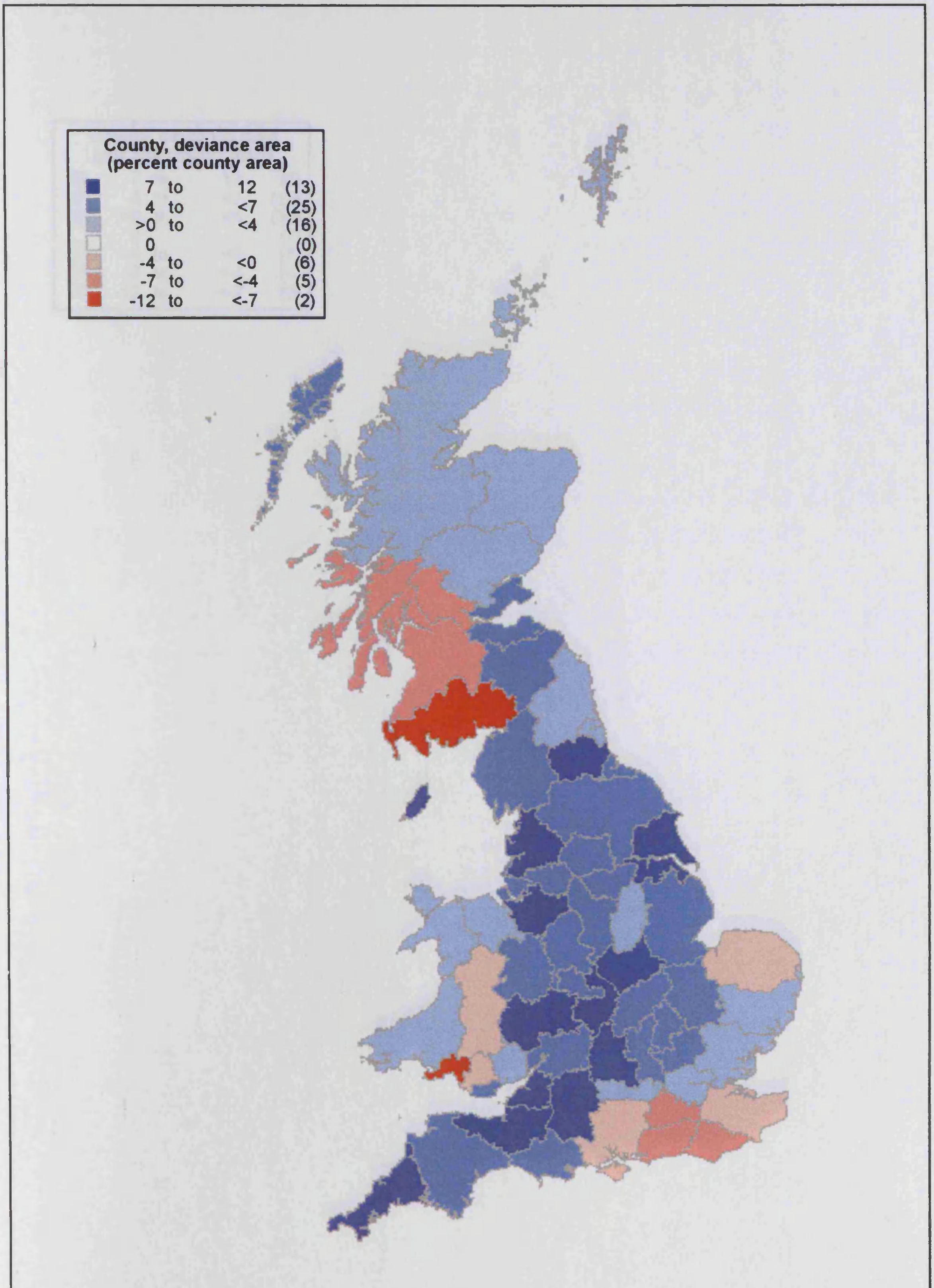


Figure 5.31 Woodland county level, LCMGB – OS95GRef error map, percentage of county area.

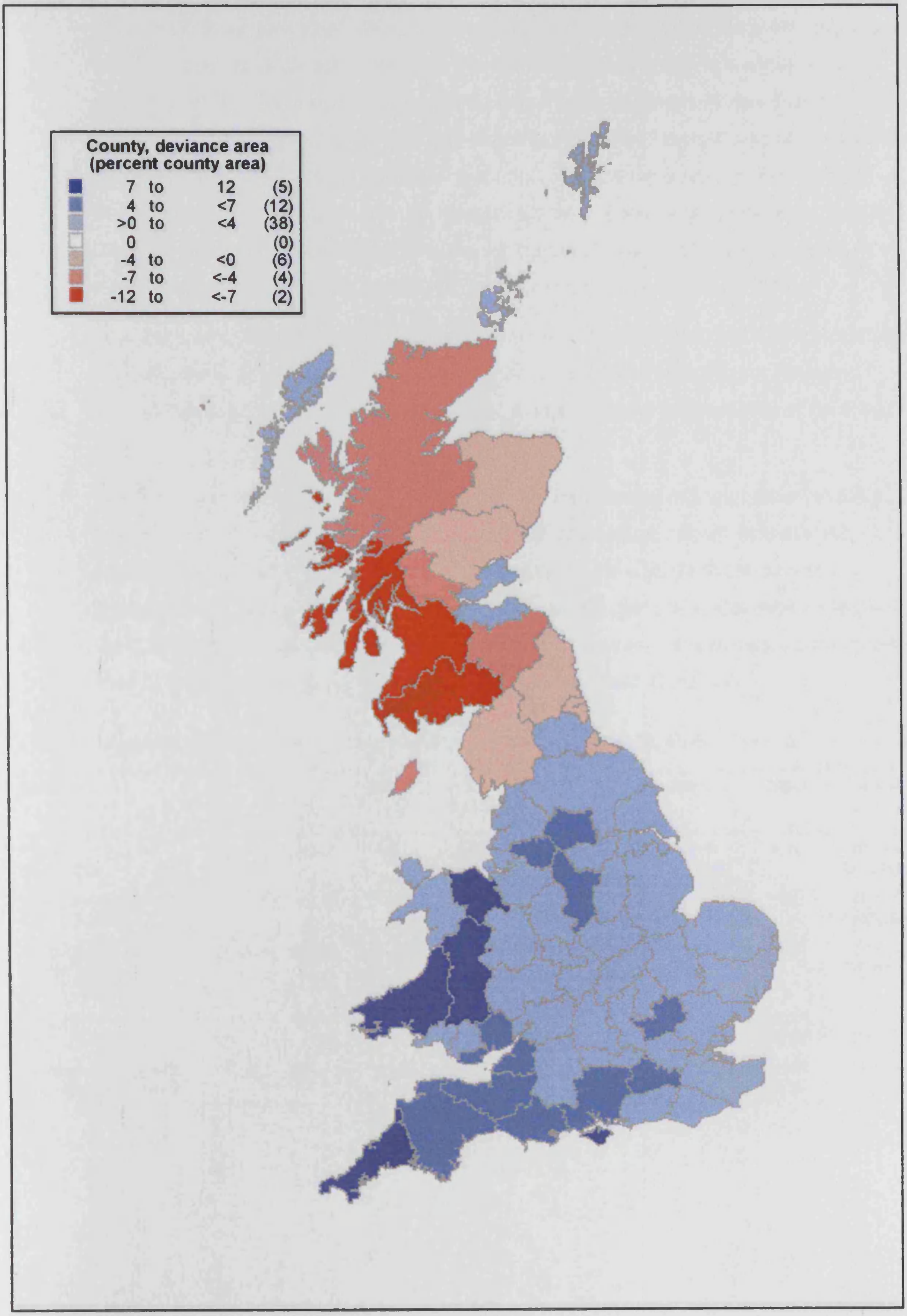
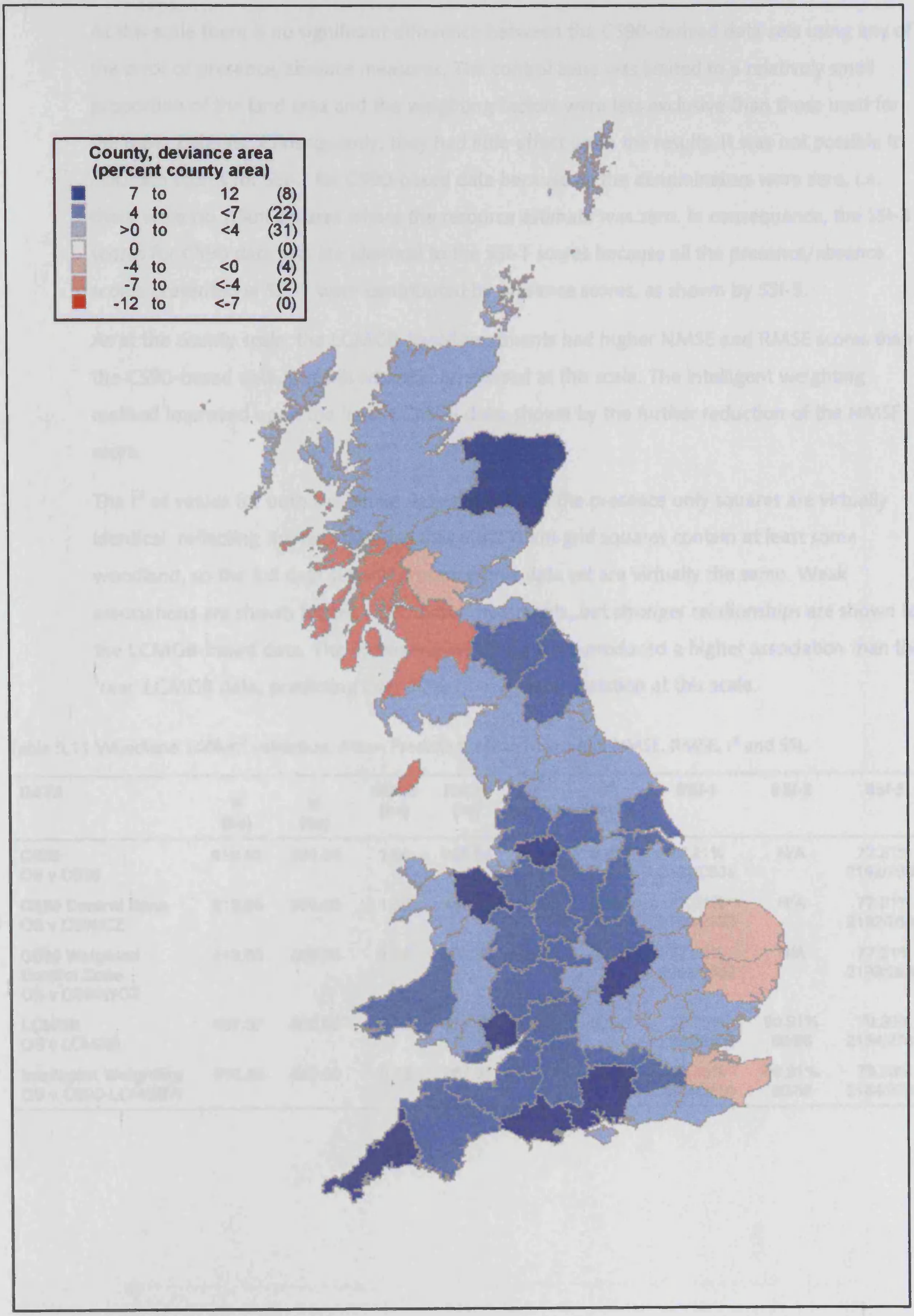


Figure 5.32 Woodland county level, IW – OS95GRef error map, percentage of county area.



5.4.3 100 km² level

At this scale there is no significant difference between the CS90-derived data sets using any of the error or presence/absence measures. The control zone was limited to a relatively small proportion of the land area and the weighting factors were less exclusive than those used for the other habitats. Consequently, they had little effect upon the results. It was not possible to calculate scores for SSI-2 for CS90-based data because all the denominators were zero, i.e. there were no 10km squares where the resource estimate was zero. In consequence, the SSI-3 scores for CS90 data sets are identical to the SSI-1 scores because all the presence/absence scores presented in SSI-1 were contributed by presence scores, as shown by SSI-3.

As at the county scale, the LCMGB-based treatments had higher NMSE and RMSE scores than the CS90-based data, but this situation is reversed at this scale. The intelligent weighting method improved upon the 'raw' LCMGB data, shown by the further reduction of the NMSE score.

The r^2 of values for both the whole data and set and the presence only squares are virtually identical, reflecting the circumstance that most 10km grid squares contain at least some woodland, so the full data set and presence-only data set are virtually the same. Weak associations are shown for the CS90-based treatments, but stronger relationships are shown for the LCMGB-based data. The intelligent weighting score produced a higher association than the 'raw' LCMGB data, predicting over 70% of woodland variation at this scale.

Table 5.11 Woodland 100km² validation: Mean Predicted, Mean Measured, NMSE, RMSE, r^2 and SSI.

DATA	$\bar{P}$ (ha)	$\bar{M}$ (ha)	NMSE (ha)	RMSE (ha)	r^2	r^2 (Ind>0)	SSI-1	SSI-2	SSI-3
CS90 OS v CS90	913.60	680.00	1.56	983.54	0.232	0.232	77.21% 2192/2839	N/A	77.21% 2192/2839
CS90 Control Zone OS v CS90CZ	913.60	680.00	1.55	983.16	0.233	0.233	77.21% 2192/2839	N/A	77.21% 2192/2839
CS90 Weighted Control Zone OS v CS90WCZ	913.60	680.00	1.56	985.23	0.230	0.230	77.21% 2192/2839	N/A	77.21% 2192/2839
LCMGB OS v LCMGB	697.37	680.00	0.96	674.28	0.619	0.615	79.75% 2264/2839	90.91% 80/88	79.39% 2184/2751
Intelligent Weighting OS v CS90-LCMGBW	913.60	680.00	0.72	667.65	0.719	0.716	79.75% 2264/2839	90.91% 80/88	79.39% 2184/2751

Figure 5.33 Woodland 100km² square, OSGRef95 reference data, area ha^{-1}.

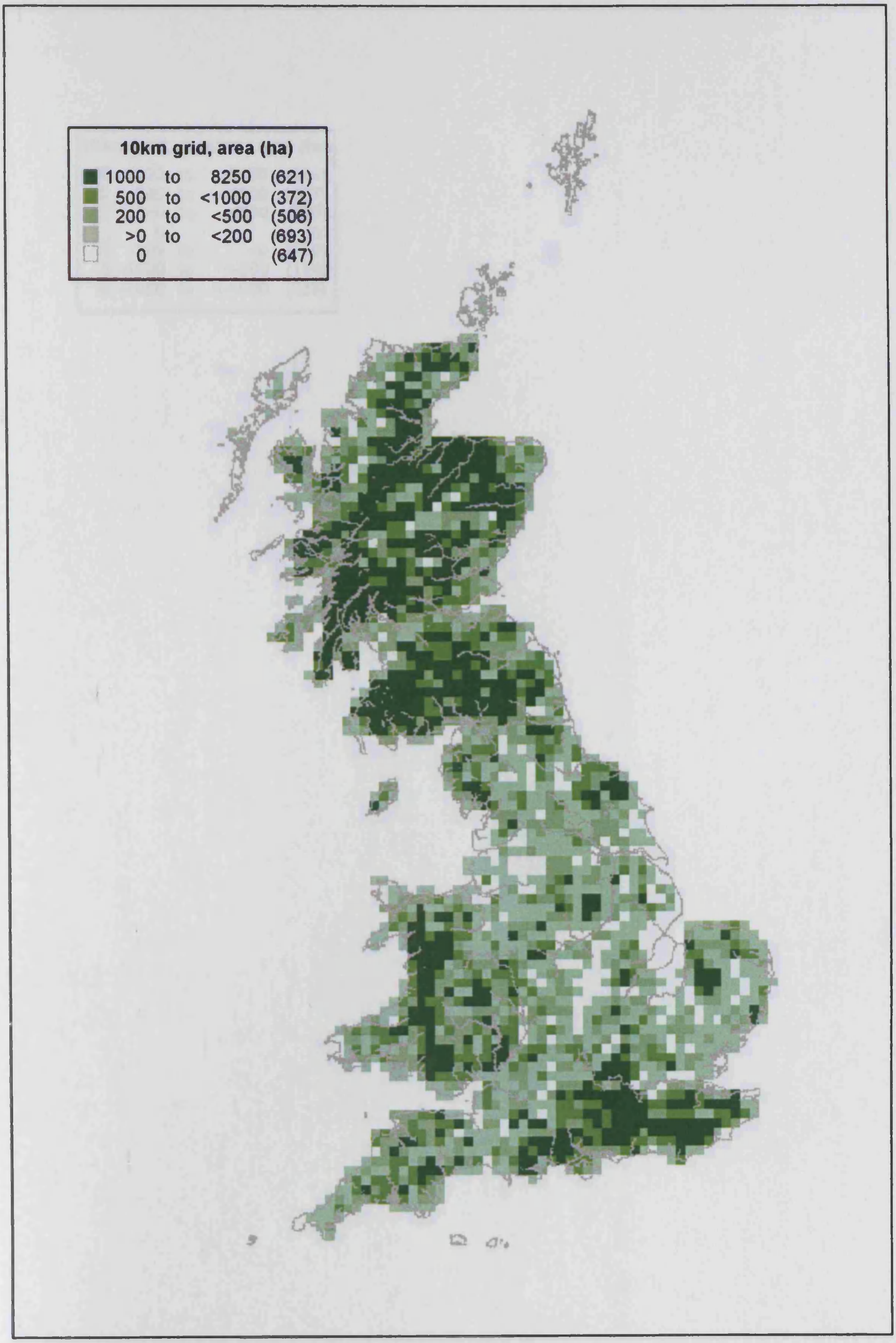


Figure 5.34 Woodland 100km^2 square, CS90- OS95GRef error map, area ha^{-1}.

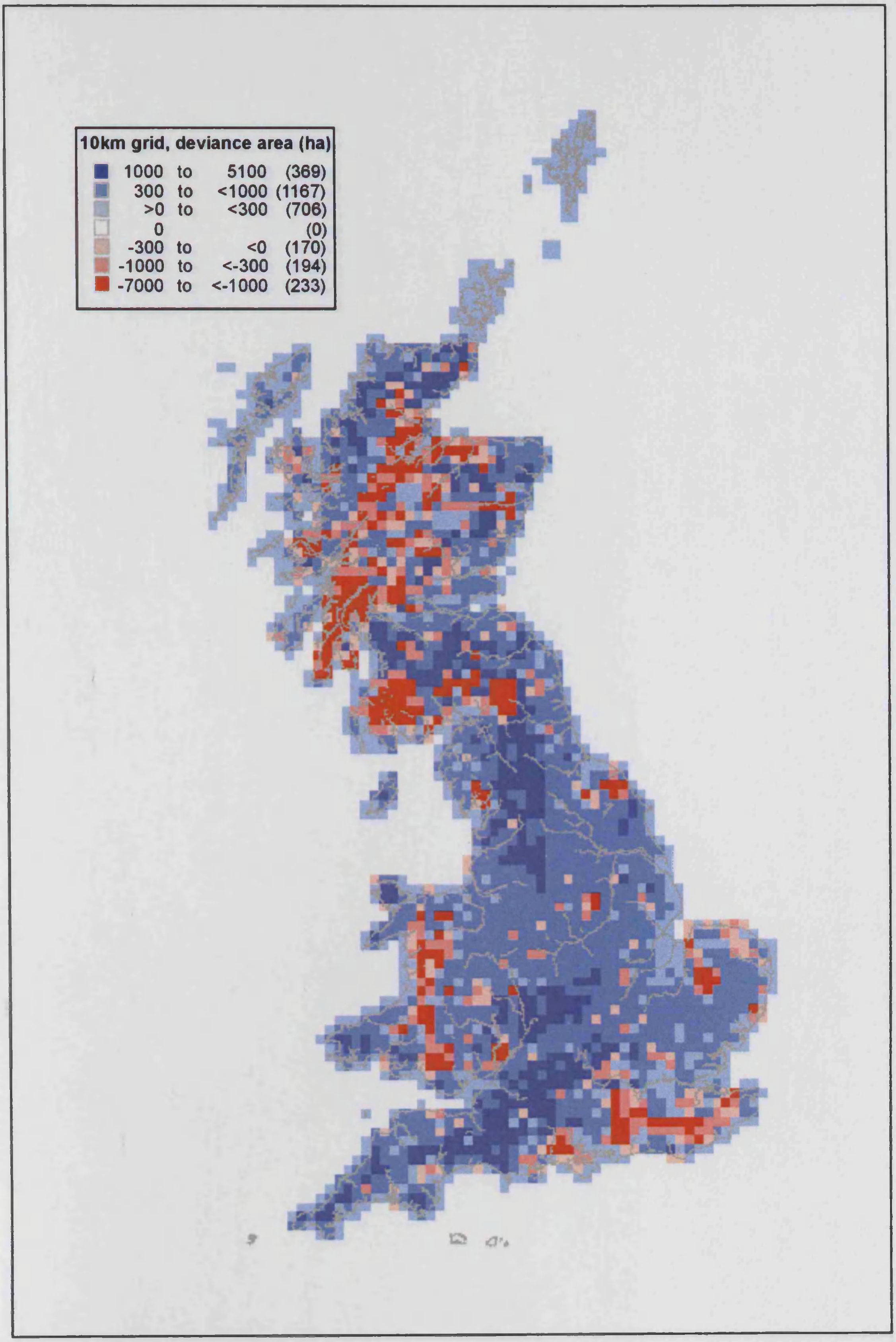


Figure 5.35 Woodland 100km² square, CS90CZ - OS95GRef error map, area ha^{-1}.

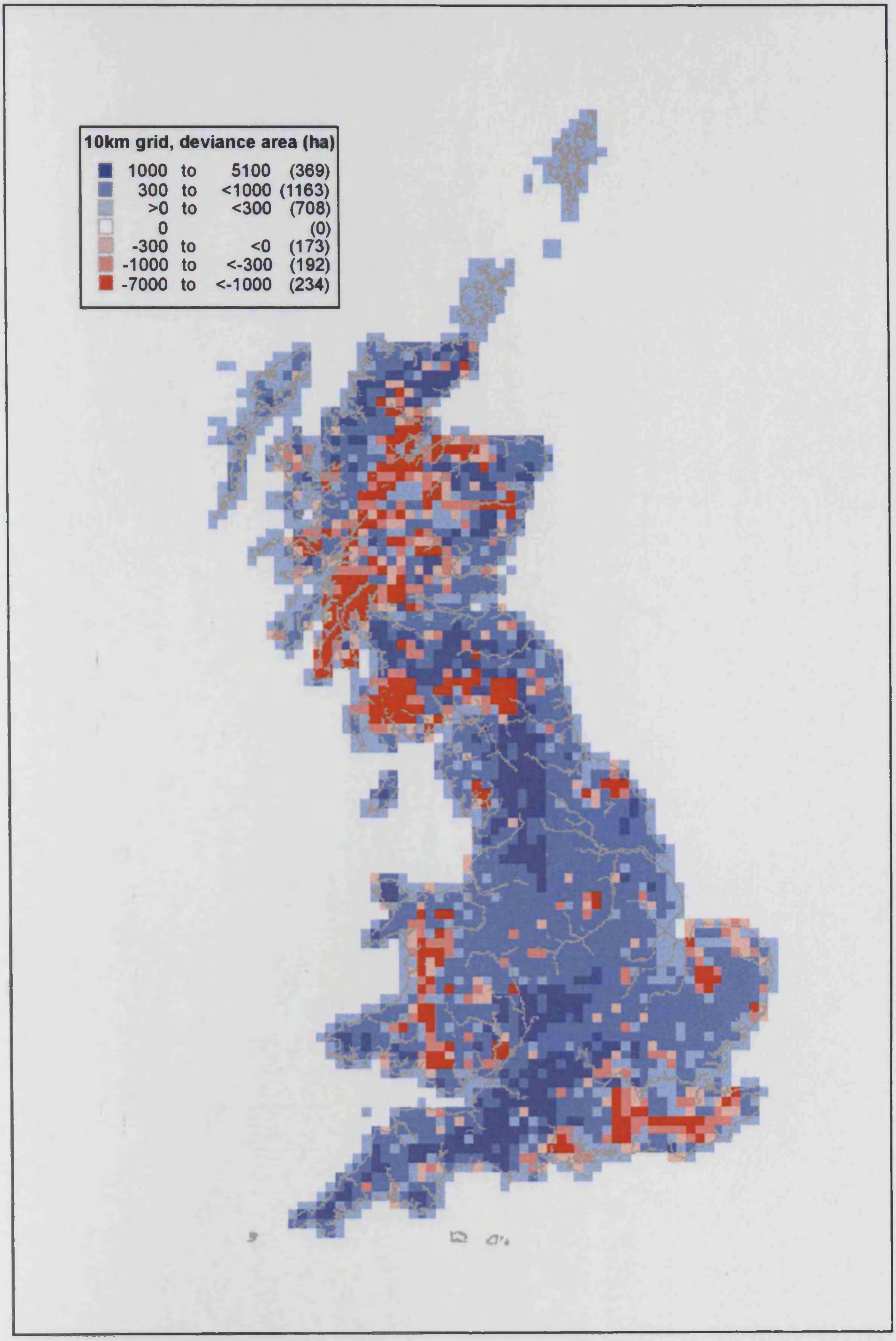


Figure 5.36 Woodland 100km² square, CS90CZW - OS95GRef error map, area ha^{-1}.

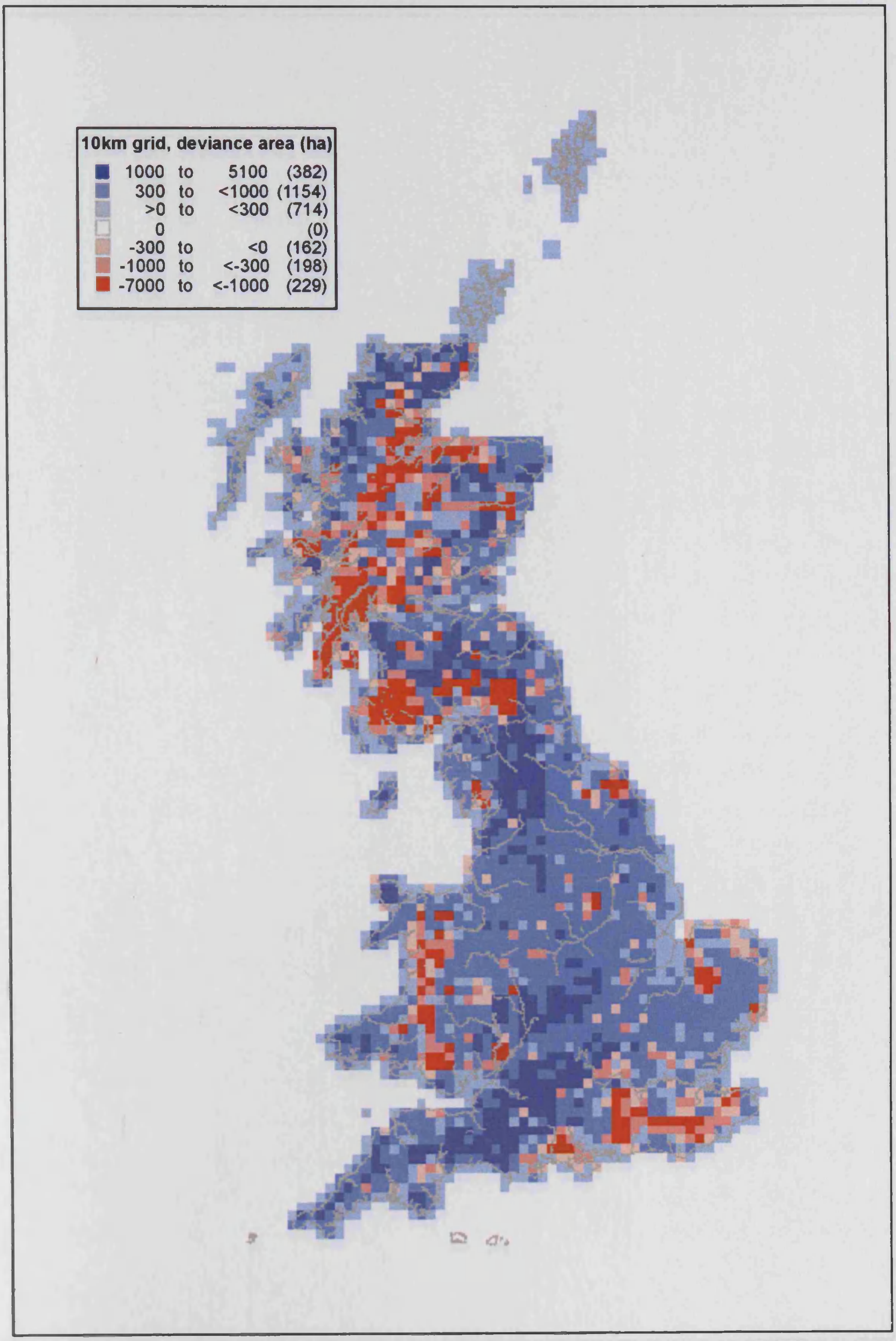


Figure 5.37 Woodland 100km^2 square, LCMGB - OS95GRef error map, area ha^{-1}.

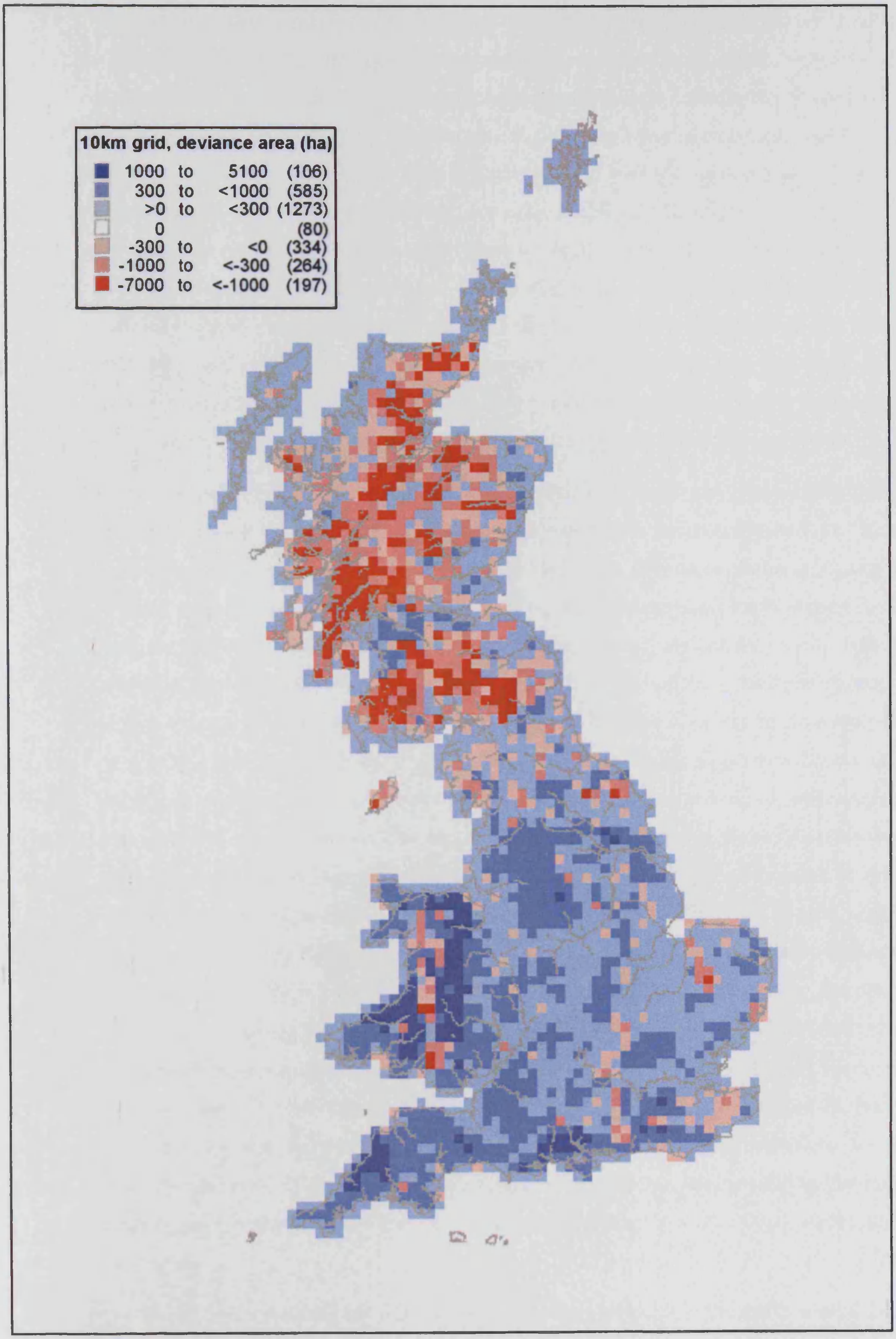
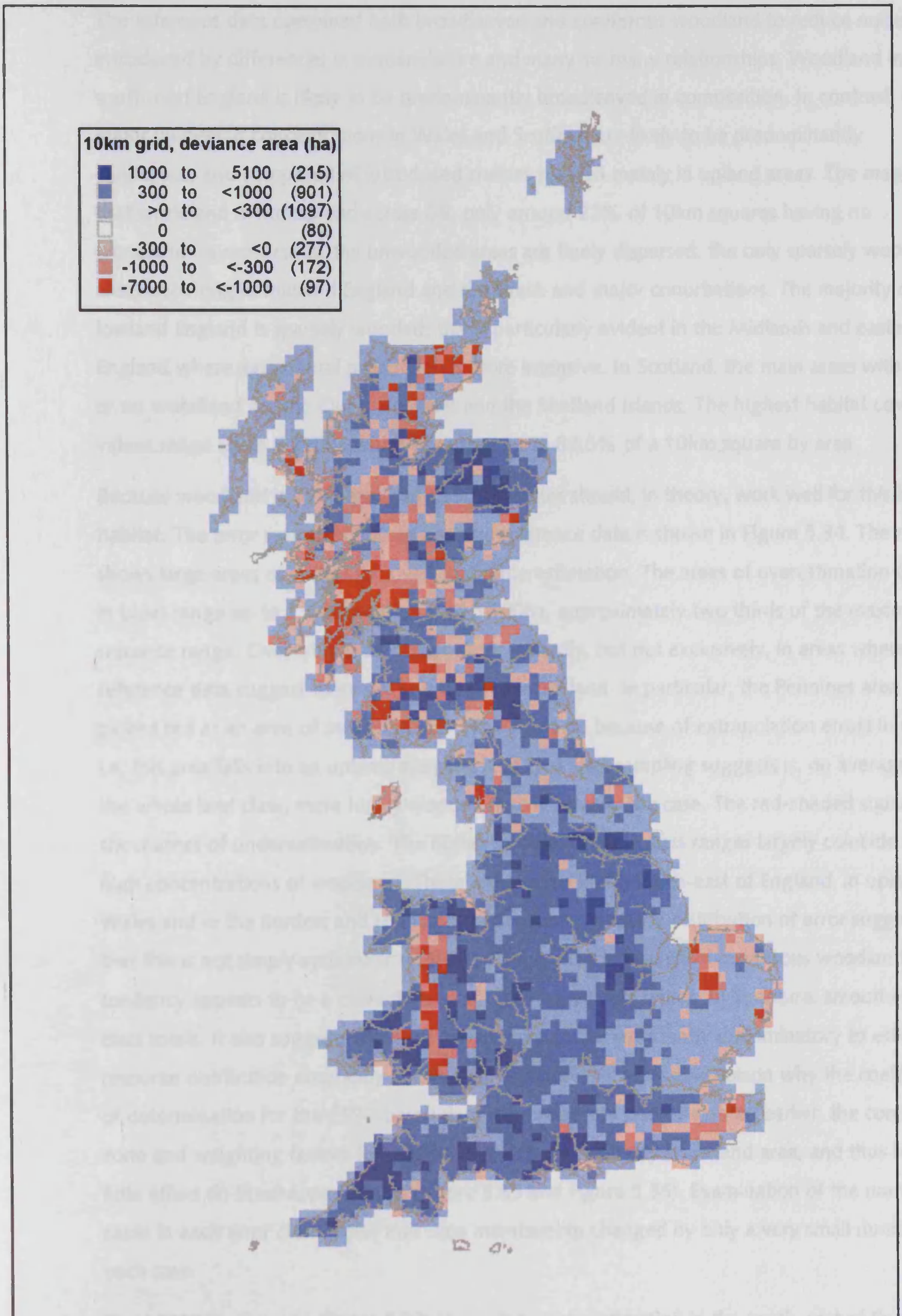


Figure 5.38 Woodland 100km^2 square, IW - OS95GRef error map, area ha^{-1}.



Mapping the woodland reference data (Figure 5.33) shows major concentrations in the south east of England, in south, mid and north Wales, the Scottish Borders and the Scottish Highlands. The reference data combined both broadleaved and coniferous woodland to reduce noise introduced by differences in nomenclature and many-to-many relationships. Woodland in south-east England is likely to be predominantly broadleaved in composition. In contrast, the major woodland concentrations in Wales and Scotland are likely to be predominantly coniferous, and composed of introduced species planted mainly in upland areas. The map shows that woodland is widespread across GB, only around 22% of 10km squares having no woodland. In most cases, the unwooded areas are finely dispersed, the only sparsely wooded areas occurring in midland England and the Wash and major conurbations. The majority of lowland England is sparsely wooded; this is particularly evident in the Midlands and eastern England where agricultural production is more intensive. In Scotland, the main areas with little or no woodland are the Outer Hebrides and the Shetland Islands. The highest habitat cover values range up to 8,250ha which is equivalent to 82.5% of a 10km square by area.

Because woodland is so widespread CS90 estimates should, in theory, work well for this kind of habitat. The error map for CS90 against the reference data is shown in Figure 5.34. The map shows large areas of systematic over- and underestimation. The areas of overestimation (shaded in blue) range up to a maximum error of 5,100ha, approximately two thirds of the maximum resource range. Overestimation occurs predominantly, but not exclusively, in areas where the reference data suggest there are low-levels of woodland. In particular, the Pennines area is picked out as an area of overestimation. This could be because of extrapolation errors in CS90: i.e. this area falls into an upland/marginal land class that sampling suggests is, on average across the whole land class, more highly wooded than is actually the case. The red-shaded squares show areas of underestimation. The higher underestimation class ranges largely coincide with high concentrations of woodland. These areas occur in the south-east of England, in upland Wales and in the Borders and Highlands of Scotland. The wide distribution of error suggests that this is not simply systematic underestimation of broadleaved or coniferous woodland. This tendency appears to be a characteristic of CS90-based data which, by it nature, smoothes land-class totals. It also suggests that land classes may not be sufficiently discriminatory to estimate resource distribution accurately at this scale and provides a possible reason why the coefficients of determination for the CS90-based data were relatively poor. As noted earlier, the control zone and weighting factors for woodland affected a relatively small land area, and thus have little effect on the mapped results (Figure 5.35 and Figure 5.36). Examination of the number of cases in each error class shows that class membership changed by only a very small number in each case.

The LCMGB error map (Figure 5.37) shows less underestimation in the south-east of England and in Northern Wales. However, there are still nearly 200 cases in the greatest underestimation error class and this is reflected by a greater concentration of error in the Highlands of Scotland.

It is possible that LCMGB is better at estimating broadleaved woodland, but produces underestimates for coniferous woodland. LCMGB produced fewer cases of overestimation than the CS90-based series and these are concentrated in Hampshire, north-eastern Cornwall and Wales. Combining LCMGB-based locations with CS90-derived data in the intelligent weighting estimate resulted in far fewer cases of underestimation and dispersed the concentrated areas of these in the Highlands. In the south-east of England the highly wooded areas are subject to both over- and underestimation. In England and Wales more error, in both directions, is evident in the west than in the east. The intelligent weighting results also show a tendency towards general overestimation compared to the LCMGB-based results.

5.4.4 1 km² level

The results for woodland at the 1km² level are shown in Table 5.12. The CS90-based treatments all produced very similar results. The NMSE and RMSE scores varied little, as did the coefficients of determination and the presence/absence measures. The CS90 derived data performed well for estimating resource absence, but by the other two SSI measures performed relatively poorly.

The LCMGB and intelligent weighting measures performed very well at this scale, producing lower NMSE and RMSE scores than the CS90-based methods and showing relatively strong coefficients of determination. The SSI scores were also relatively high, especially for the overall distribution (SSI-1) while there was a slight improvement in performance for correctly estimating resource presence (SSI-3) compared to the CS90-based treatments. Overall, for all methods, the performance in correctly locating habitat (SSI-3) was relatively poor.

Table 5.12 Woodland 1km² validation: Mean Predicted, Mean Measured, NMSE, RMSE, r^2 and SSI.

DATA	$\bar{P}$ (ha)	$\bar{M}$ (ha)	NMSE (ha)	RMSE (ha)	r^2	r^2 (Ind>0)	SSI-1	SSI-2	SSI-3
CS90 OS v CS90	10.80	8.04	4.85	20.52	0.051	0.051	23.51% 56468/240222	100% 20/20	23.50% 56448/240202
CS90 Control Zone OS v CS90CZ	10.80	8.04	4.85	20.51	0.052	0.049	22.50% 61255/240222	97.46% 4935/5063	23.95% 56320/235159
CS90 Weighted Control Zone OS v CS90WCZ	10.80	8.04	4.86	20.53	0.051	0.057	25.52% 56511/240222	100.00% 63/63	23.50% 56448/240159
LCMGB OS v LCMGB	8.24	8.04	2.96	14.00	0.546	0.566	41.32% 99252/240222	94.45% 45477/48150	28.00% 53775/192072
Intelligent Weighting OS v CS90-LCMGBW	10.80	8.04	3.25	16.79	0.504	0.517	41.32% 99252/240222	94.45% 45477/48150	28.00% 53775/192072

5.4.5 Summary

As a widespread and ubiquitous habitat, woodland should be inherently easier to detect than the previous target habitats. CS90 estimates are not thought to be accurate below the 3000km² level, just below the average county area in Great Britain. However, at the county level the

CS90-based data produced good estimates of woodland area despite representing a probable overestimate in total and producing relatively large errors in some counties.

The tendency for CS90 to 'smooth' data across land classes was evident. CS90-based estimates produced overestimates where the reference data suggested there were low levels of woodland cover and underestimates where there were higher concentrations of woodland. These results suggest either that the original sample was not representative enough or that the land classes are insufficiently discriminatory of the appropriate environmental variables. The widespread nature of woodland meant that it was not appropriate to develop a very prescriptive control zone or very exclusive weighting factors and for these reasons they had minimal effects upon the CS90 resource estimates.

Overall, the LCMGB-based data provided the most accurate estimates by all measures used. The intelligent weighting methods improved the LCMGB data at the county and 10km grid square scales but had little effect at the 1km square level. These results suggest that the use of CS90 data can improve resource estimates for woodland when combined with the satellite-based distribution model.

5.5 LAND CLASS TESTING

The ITE land class system is integral to the CS90 estimation model. Because of this testing was conducted to investigate its performance with the three target habitats. The first set of tests were conducted at the land class level; the results are shown in Tables 5.13-5.15. The tables confirm some of the earlier results but also reveal interesting differences in land class performance. Supporting the earlier results, there were strong associations for saltmarsh and woodland, but no significant relationship was found for lowland heathland. Despite the significant r^2 values for saltmarsh and woodland, there are still large differences between the habitat land class totals for the reference data, CS90 and LCMGB estimates. For saltmarsh, LCMGB displays a stronger relationship with the reference data than CS90, but the positions are reversed for woodland. The significant values shown by CS90 in both cases suggest that, at a whole land class level, CS90 totals are strongly related to the habitat distribution. However, the (often) large differences between the reference and estimated land class totals suggest that the land class system is less good at stratifying for or modelling habitat quantity.

The second tier of testing was conducted at the 1km^2 level to assess the relative performance of different ITE land classes in estimating each of the target habitats. In the case of lowland heathland, this test used the control zone based data and therefore no regression results were available for 'upland' land classes. The results are presented in Table 5.16 and show that the r^2 values vary greatly by land class and by habitat. For saltmarsh, for example, r^2 values for the LCMGB estimates vary from 0.000 in land classes 21, 22 and 24 to over 0.8 in land classes 6 and 12. For woodland, the range is from about 0.35 in land classes 7, 15, 18 and 23 to over 0.7

in land classes 1 and 3. For heathland, the coefficients of determination are universally low, never exceeding far above 0.1.

One possible explanation of these patterns might be that the models perform better in land classes which have larger areas, and greater within-class variation, of the target habitat, since these should be easier to detect. In order to test this, the relationships between the land class mean, land class standard deviation and the r^2 values were plotted for each target habitat.

Table 5.13 Saltmarsh estimates; land class totals and r^2.

Land class	LC Area Km^2	OS/NCC ha^{-1}	CS90 ha^{-1}	LCMGB ha^{-1}	OS/NCC % LC	CS90 % LC	LCMGB % LC
1	14159	45.66	0.00	5.32	0.00	0.00	0.00
2	14463	20.68	0.00	123.76	0.00	0.00	0.01
3	15452	1.79	0.00	42.80	0.00	0.00	0.00
4	9012	4543.99	901.20	7420.28	0.50	0.10	0.82
5	3877	83.63	0.00	17.11	0.02	0.00	0.00
6	10340	248.40	0.00	94.51	0.02	0.00	0.01
7	2532	1729.51	66.38	2095.04	0.68	0.03	0.83
8	4412	18040.23	24532.54	18784.73	4.09	5.56	4.26
9	11781	648.67	0.00	391.90	0.06	0.00	0.03
10	13905	1120.33	0.00	981.17	0.08	0.00	0.07
11	8895	0.00	0.00	0.00	0.00	0.00	0.00
12	3543	31.36	0.00	17.50	0.01	0.00	0.00
13	7263	3379.79	4975.44	3358.12	0.47	0.69	0.46
14	933	1173.14	8.51	1041.00	1.26	0.01	1.12
15	4195	187.33	0.00	105.97	0.04	0.00	0.03
16	3089	1215.21	7478.79	809.23	0.39	2.42	0.26
17	12999	0.00	0.00	4.75	0.00	0.00	0.00
18	6732	0.00	0.00	0.32	0.00	0.00	0.00
19	5421	0.00	0.00	0.00	0.00	0.00	0.00
20	2508	0.00	0.00	0.12	0.00	0.00	0.00
21	9717	60.65	0.00	13.90	0.01	0.00	0.00
22	12549	19.88	0.00	8.12	0.00	0.00	0.00
23	6951	0.00	0.00	0.00	0.00	0.00	0.00
24	7207	25.63	0.00	0.62	0.00	0.00	0.00
25	10552	87.23	0.00	83.42	0.01	0.00	0.01
26	6876	648.84	0.00	573.13	0.09	0.00	0.08
27	6881	83.01	0.00	68.69	0.01	0.00	0.01
28	7464	403.05	0.00	161.36	0.05	0.00	0.02
29	5465	1065.15	951.69	1977.70	0.19	0.17	0.36
30	4254	463.54	753.36	267.73	0.11	0.18	0.06
31	3016	314.91	0.00	396.75	0.10	0.00	0.13
32	3779	293.15	0.00	93.48	0.08	0.00	0.02
r^2 (based on % land class values, OS/NCC % LC dependent variable)						0.89	0.99

Table 5.14 Lowland heathland estimates; land class totals and r^2.

Land class	LC Area Km^2	EN LHI ha^{-1}	CS90CZ ha^{-1}	LCMGBCZ ha^{-1}	EN LHI % LC	CS90CZ % LC	LCMGBCZ % LC
1	13103	4609.48	94.68	18355.36	0.35	0.01	1.40
2	14459	8840.40	7436.53	22090.10	0.61	0.51	1.53
3	15360	3125.96	2801.03	7990.45	0.20	0.18	0.52
4	8954	3927.21	0.00	6261.53	0.44	0.00	0.70
5	2473	333.62	3230.68	3049.25	0.13	1.31	1.23
6	7276	2965.63	16953.93	9420.48	0.41	2.33	1.29
7	1374	931.71	903.13	1666.09	0.68	0.66	1.21
8	3214	289.30	6.79	2216.51	0.09	0.00	0.69
9	11027	868.35	0.00	12100.75	0.08	0.00	1.10
10	13641	1595.32	14849.55	20596.90	0.12	1.09	1.51
11	8895	14.71	0.00	2601.18	0.00	0.00	0.29
12	3542	86.87	0.00	1432.63	0.02	0.00	0.40
13	4794	216.85	16386.83	6506.04	0.05	3.42	1.36
14	603	6.69	0.00	386.15	0.01	0.00	0.64
15	1390	42.09	4.92	3463.17	0.03	0.00	2.49
16	2451	21.51	248.88	1818.17	0.01	0.10	0.74
25	2011	124.09	10652.69	3416.32	0.06	5.30	1.70
26	1192	31.87	0.00	1076.87	0.03	0.00	0.90
r^2 (based on % land class values, EN LHI % LC dependent variable)						-0.04	0.17

Table 5.15 Woodland estimates; land class totals and r^2.

Land class	LC Area Km^2	OS ha^{-1}	CS90 ha^{-1}	LCMGB ha^{-1}	OS % LC	CS90 % LC	LCMGB % LC
1	14159	86060.43	197646.25	138021.21	6.08	13.96	9.75
2	14463	131365.26	202959.37	174008.23	9.08	14.03	12.03
3	15452	97574.11	94631.77	127910.78	6.31	6.12	8.28
4	9012	23119.13	19811.48	43382.75	2.57	2.20	4.81
5	3877	17418.01	47955.89	41221.48	4.49	12.37	10.63
6	10340	44548.09	129305.52	110831.50	4.31	12.51	10.72
7	2532	4196.08	9632.35	9997.52	1.66	3.80	3.95
8	4412	7641.12	4534.09	15930.66	1.73	1.03	3.61
9	11781	40017.46	125119.34	74163.54	3.40	10.62	6.30
10	13905	38794.64	111062.44	76214.96	2.79	7.99	5.48
11	8895	21748.16	73596.23	39171.85	2.44	8.27	4.40
12	3543	5564.57	15538.40	9018.29	1.57	4.39	2.55
13	7263	26034.42	70787.15	35358.66	3.58	9.75	4.87
14	933	2785.55	838.48	3475.90	2.99	0.90	3.73
15	4195	23298.70	33107.17	59833.04	5.55	7.89	14.26
16	3089	14758.93	20827.64	25699.98	4.78	6.74	8.32
17	12999	155194.49	111932.44	216112.96	11.94	8.61	16.63
18	6732	99927.94	128553.02	61320.28	14.84	19.10	9.11
19	5421	84542.87	45660.60	60166.54	15.60	8.42	11.10
20	2508	29263.94	26195.73	27204.07	11.67	10.44	10.85
21	9717	154823.36	132421.63	73542.66	15.93	13.63	7.57
22	12549	280076.01	397210.03	165436.09	22.32	31.65	13.18
23	6951	10126.88	9135.35	10382.36	1.46	1.31	1.49
24	7207	72663.25	85776.56	33720.94	10.08	11.90	4.68
25	10552	108559.67	133463.89	94926.97	10.29	12.65	9.00
26	6876	72686.39	79643.49	64478.89	10.57	11.58	9.38
27	6881	69657.44	48656.11	58253.79	10.12	7.07	8.47
28	7464	132994.83	185807.02	79398.66	17.82	24.89	10.64
29	5465	28153.67	20800.88	23307.91	5.15	3.81	4.26
30	4254	39959.86	29143.85	19252.29	9.39	6.85	4.53
31	3016	509.99	1560.96	1417.04	0.17	0.52	0.47
32	3779	6463.03	399.10	6682.67	1.71	0.11	1.77
r^2 (based on % land class values, OS % LC dependent variable)						0.80	0.65

Table 5.16 Landclass testing 1km^2, r^2 values for saltmarsh, lowland heathland and woodland against respective reference data sets.

Land Class	Saltmarsh r^2 (LCMGB)	Lowland heathland r^2 (LCMGBCZ)	Woodland r^2 (LCMGB)
1	0.001	0.079	0.714
2	0.221	0.104	0.682
3	0.035	0.060	0.707
4	0.585	0.111	0.562
5	0.084	0.055	0.528
6	0.838	0.050	0.405
7	0.467	0.045	0.357
8	0.575	0.022	0.484
9	0.461	0.007	0.569
10	0.687	0.014	0.563
11	N/A	0.000	0.627
12	0.833	0.002	0.626
13	0.529	0.006	0.440
14	0.277	0.000	0.456
15	0.453	0.004	0.354
16	0.666	0.007	0.489
17	N/A	N/A	0.531
18	N/A	N/A	0.352
19	N/A	N/A	0.632
20	N/A	N/A	0.584
21	0.000	N/A	0.525
22	0.000	N/A	0.662
23	N/A	N/A	0.332
24	0.000	N/A	0.510
25	0.004	0.000	0.647
26	0.484	0.000	0.615
27	0.030	0.000	0.602
28	0.001	N/A	0.510
29	0.025	N/A	0.526
30	0.002	N/A	0.469
31	0.040	N/A	0.475
32	0.001	N/A	0.584

Figure 5.39 shows the results for saltmarsh. With the exception of an outlier, there is little relationship between habitat area and the coefficient of determination. The r^2 value for the outlier is not exceptional.

In contrast, the results for lowland heathland (Figure 5.40) show a general trend for the r^2 value to increase with both increasing mean habitat area and increasing variation in the habitat area. This suggests that predictions are poorest in land classes with small amounts of habitat and/or limited spatial variation in habitat area. These results nevertheless need to be interpreted with caution: as Table 5.16 shows, the coefficients of determination are, in every case, low and correlations are generally non-significant.

Figure 5.39 Saltmarsh r^2, standard deviation and mean total by land class.

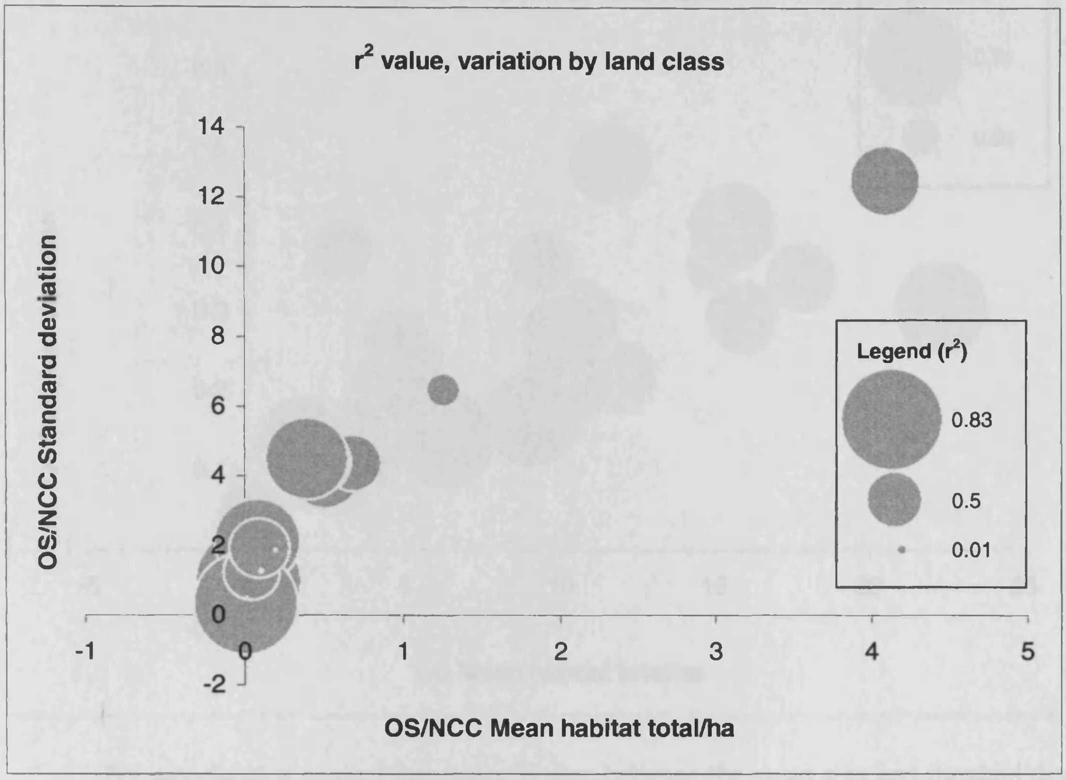


Figure 5.40 Lowland heathland r^2, standard deviation and mean total by land class.

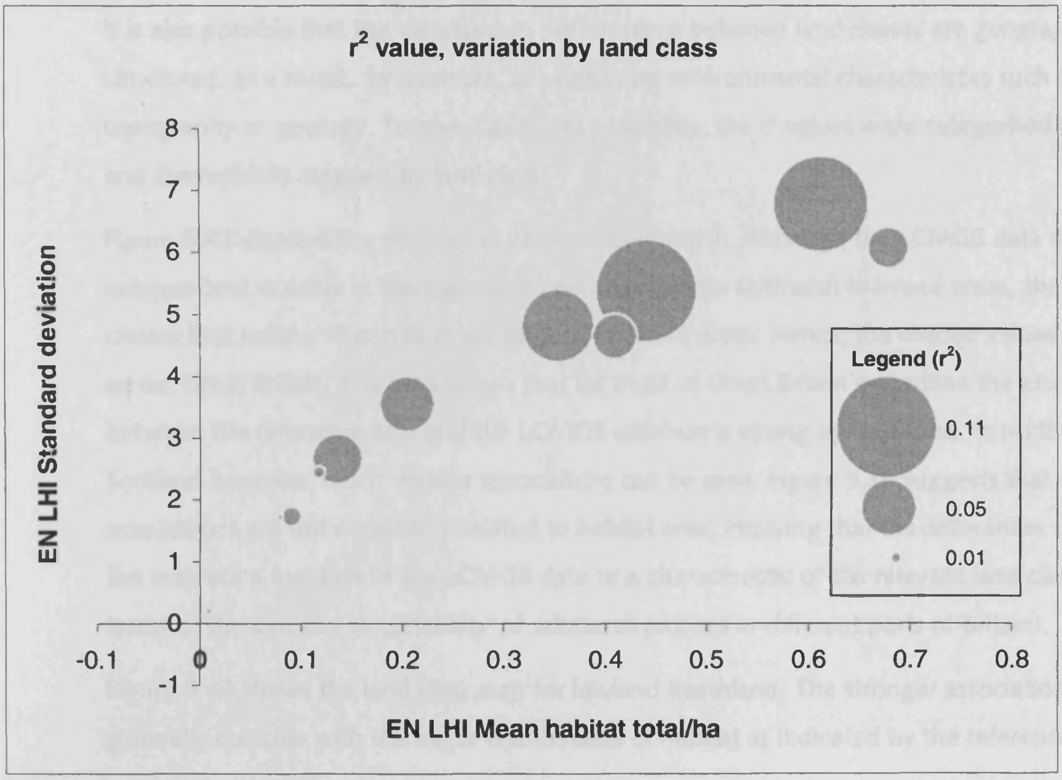
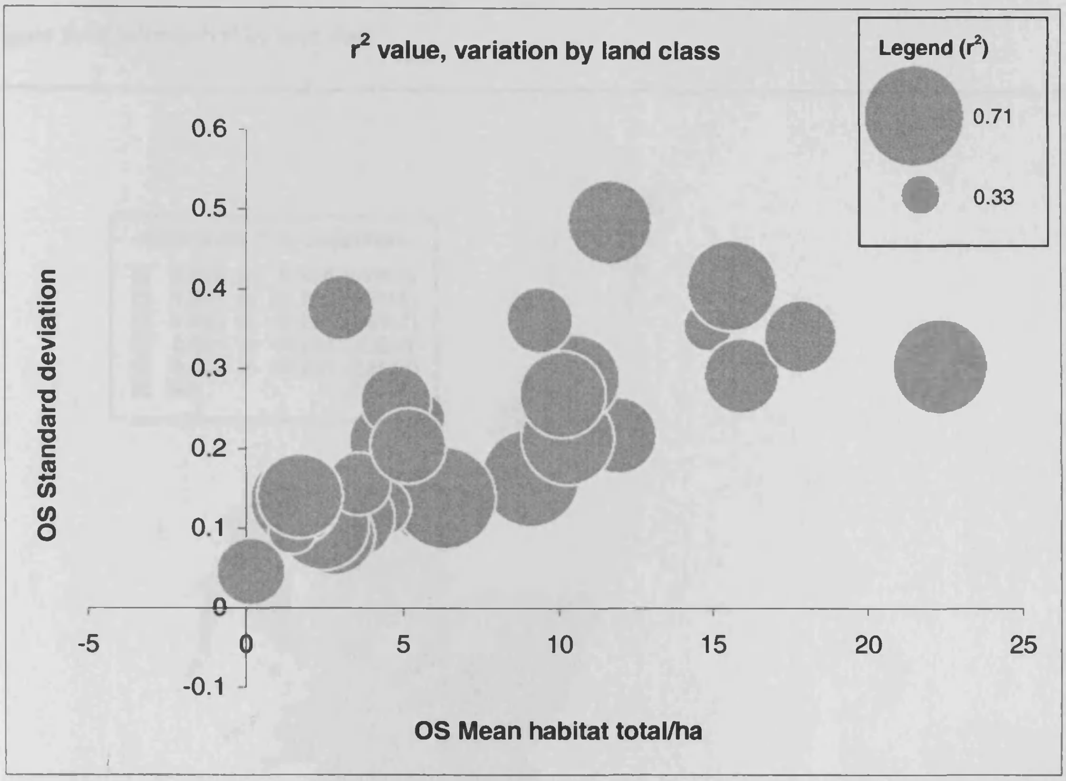


Figure 5.41 Woodland r^2, standard deviation and mean total by land class.



For woodland, a general association is seen between the mean area and standard deviation (indicating that woodland is more variable in extent in the more heavily wooded land classes), but there is no association with the r^2 value.

It is also possible that the variations in performance between land classes are geographically structured, as a result, for example, of underlying environmental characteristics such as climate, topography or geology. To investigate this possibility, the r^2 values were categorised into ranges and thematically mapped by land class.

Figure 5.42 displays the mapped r^2 values for saltmarsh. Although the LCMGB data (the independent variable in the regression) do not estimate saltmarsh in inland areas, the land classes that saltmarsh occurs in are present in inland areas. Hence, the shaded values extend across Great Britain. The map shows that for most of Great Britain's coastline the association between the reference data and the LCMGB estimate is strong by land class. In northern Scotland however, much weaker associations can be seen. Figure 5.39 suggests that weaker associations are not necessarily related to habitat area, implying that the differences shown by the map are a function of the LCMGB data or a characteristic of the relevant land classes (e.g. in terms of the size and 'detectability' of saltmarsh patches in different parts of Britain).

Figure 5.43 shows the land class map for lowland heathland. The stronger associations shown generally coincide with the major distributions of habitat as indicated by the reference data.

Because the association (r^2) is so weak it is difficult to draw any firm conclusions but the map suggests that the land class system may be reflecting variables at least partly related to the habitat distribution.

Figure 5.42 Saltmarsh r^2 by land class.

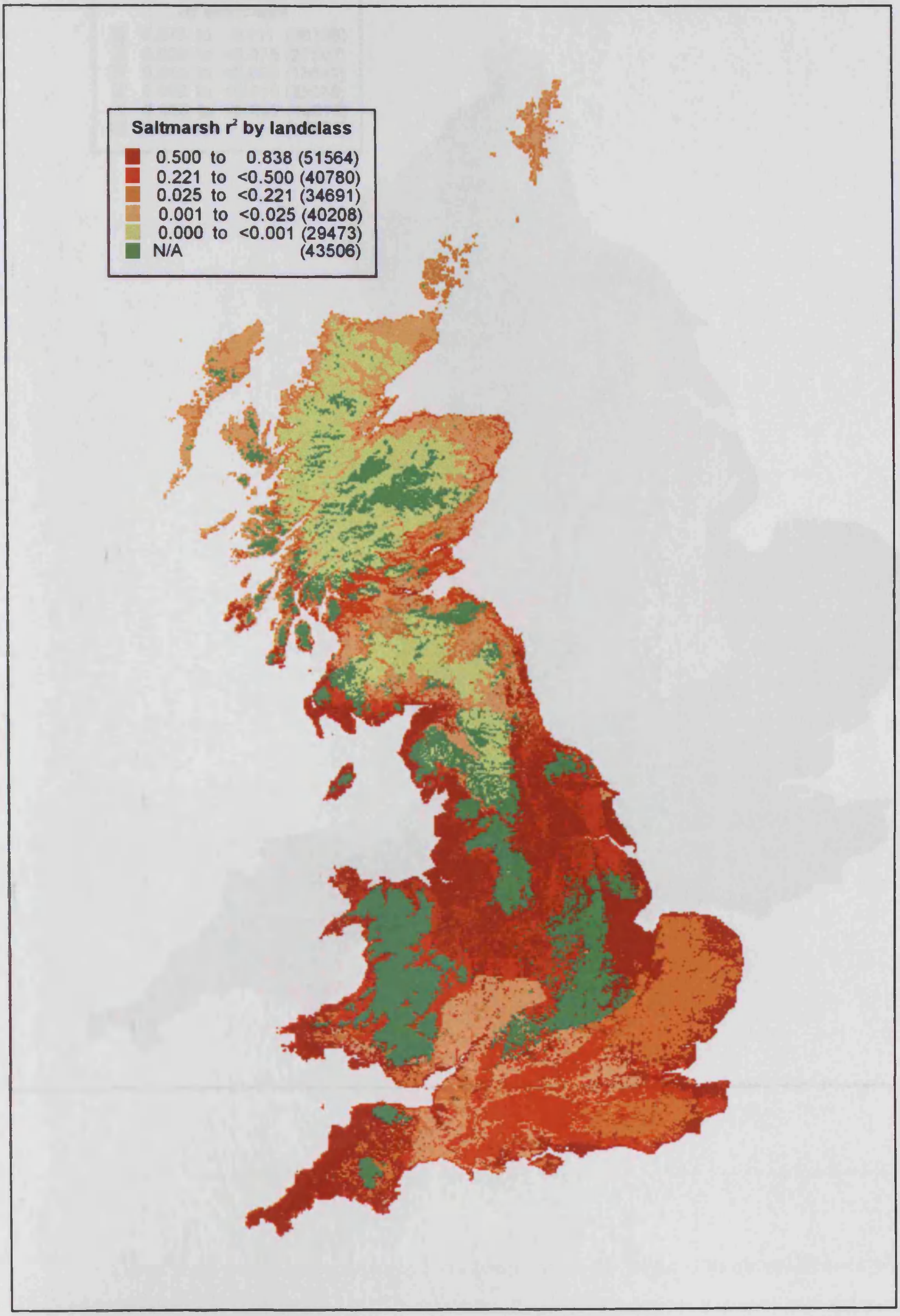


Figure 5.43 Lowland heathland r^2 by land class.

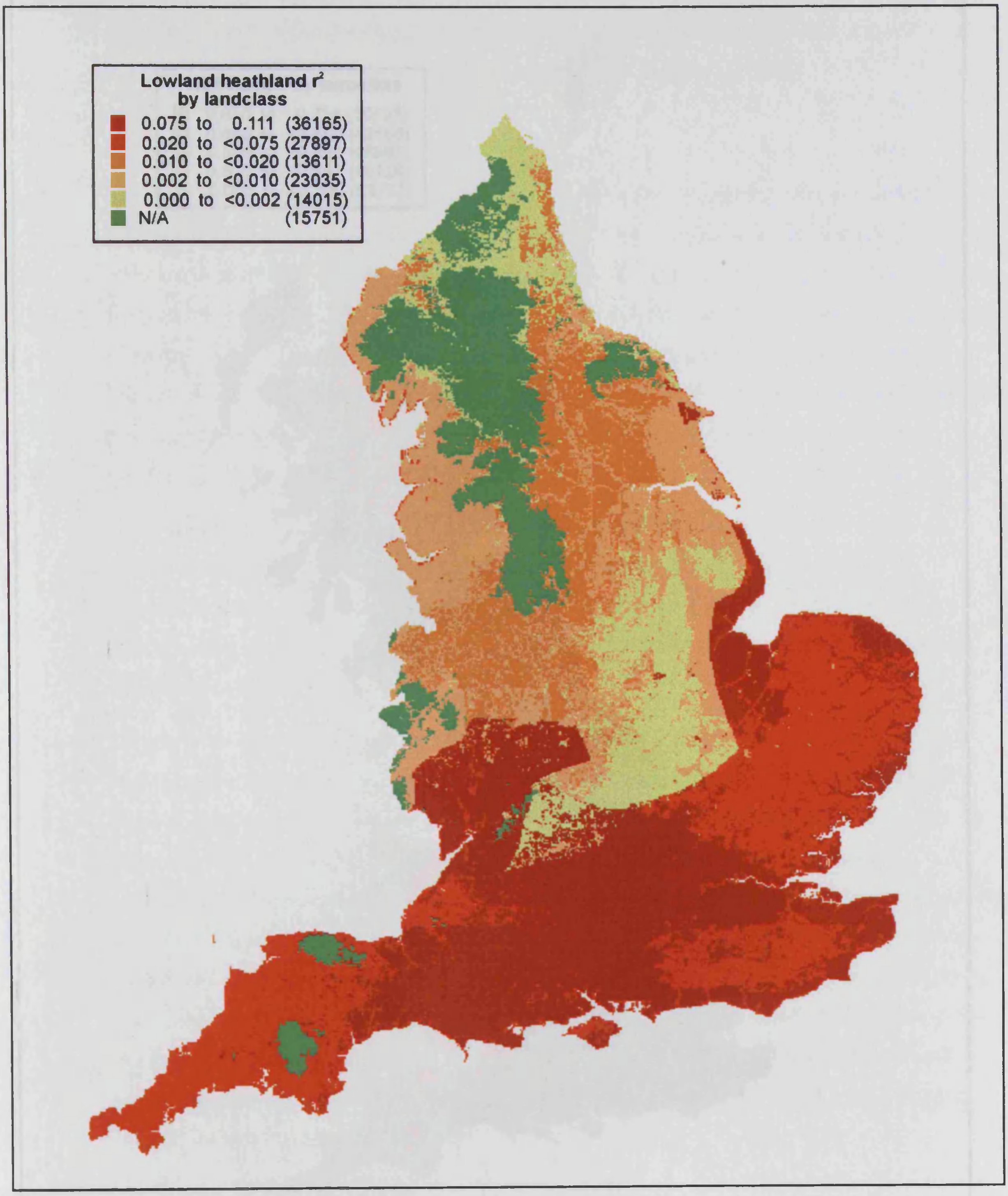


Figure 5.44 Woodland r^2 by land class.

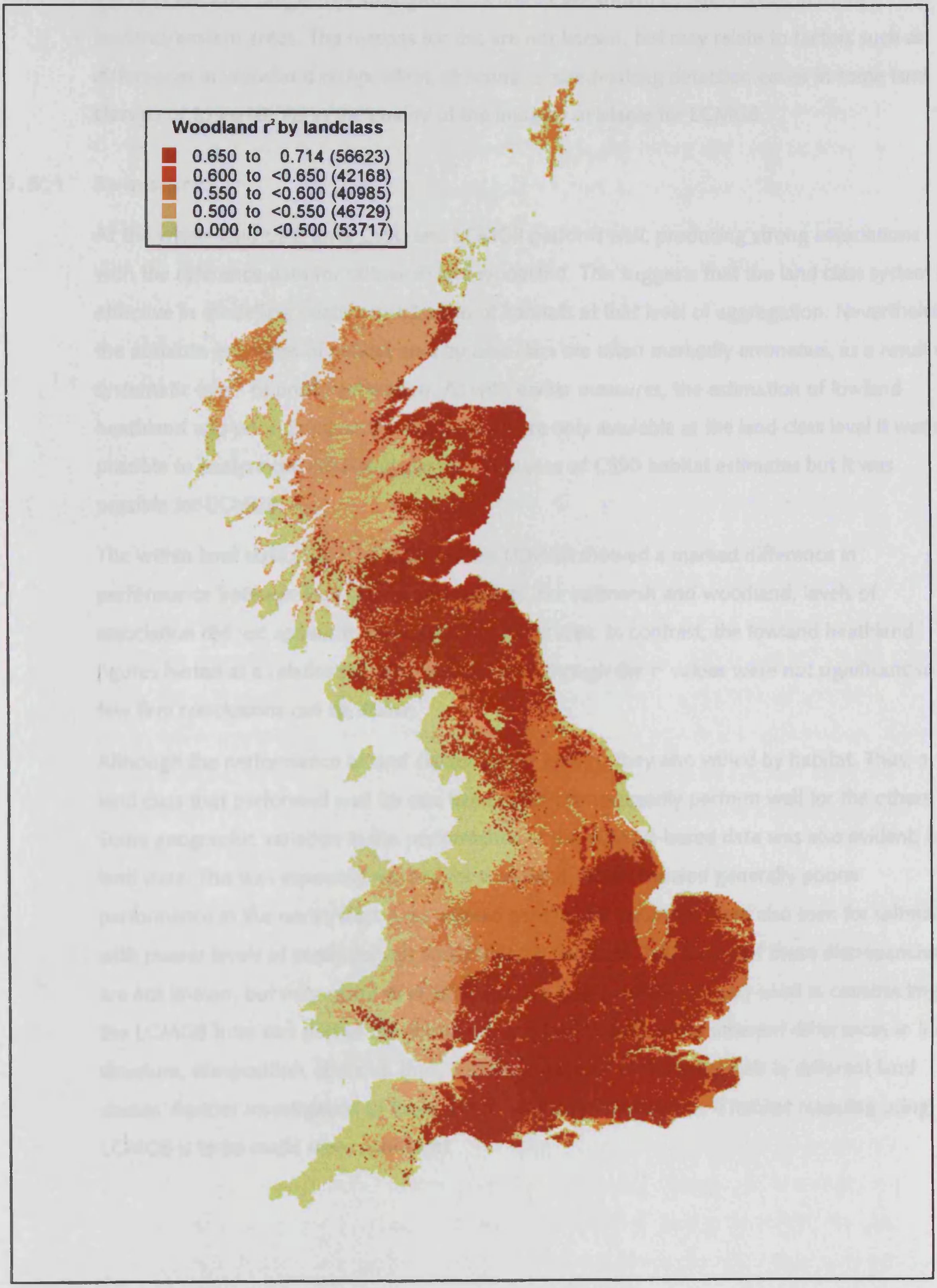


Figure 5.44 shows the map for woodland by land class. Some clear geographic variation is evident in the performance of the different land classes. The land classes with the weakest

associations have a distinct western bias for the lowest range and an upland/western bias for the next weakest range. The strongest associations are shown by land classes in more lowland/eastern areas. The reasons for this are not known, but may relate to factors such as differences in woodland composition, structure or size (making detection easier in some land classes) or to variations in the quality of the imagery available for LCMGB.

5.5.1 Summary

At the whole land class level CS90 and LCMGB perform well, producing strong associations with the reference data for saltmarsh and woodland. This suggests that the land class system is effective in modelling relative distribution of habitats at that level of aggregation. Nevertheless, the absolute estimates of habitat area by land class are often markedly erroneous, as a result of systematic over- or under-estimation. As with earlier measures, the estimation of lowland heathland was poor. Because the CS90 data were only available at the land class level it was not possible to analyse the intra-land class performance of CS90 habitat estimates but it was possible for LCMGB.

The within land class (1 km^2) associations for LCMGB showed a marked difference in performance between land classes and habitats. For saltmarsh and woodland, levels of association did not appear to be related to habitat area. In contrast, the lowland heathland figures hinted at a relationship with habitat area, though the r^2 values were not significant so few firm conclusions can be drawn.

Although the performance of land classes varied greatly, they also varied by habitat. Thus, a land class that performed well for one habitat did not necessarily perform well for the others. Some geographic variation in the performance of the LCMGB-based data was also evident, by land class. This was especially marked for woodland, which showed generally poorer performance in the north/west. Less marked geographic variations were also seen for saltmarsh, with poorer levels of prediction for land classes in Scotland. The causes of these discrepancies are not known, but may relate to variations in the quality of the imagery used in constructing the LCMGB from one part of the country to another, or perhaps to inherent differences in the structure, composition, size and, thus, 'satellite signature' of these habitats in different land classes. Further investigation of these effects would be worthwhile, if habitat mapping using LCMGB is to be made more consistent.

6 DISCUSSION

This study has examined the provision and use of habitat data. Through a combination of surveys of user and data providing organisations, and exploratory analysis of available data sets, it has attempted to assess the extent to which the data needed by users can be provided in an appropriate form and at an appropriate geographic scale and resolution. These issues are inter-related and it can be difficult to isolate different processes inherent in the conceptualisation, creation, supply and use of information. Whilst the inherent links between the creation of use of habitat-related data should not be ignored, it can be useful to further consider these processes in terms of the selection and use of information - the 'demand' side; and the creation, processing and supply of data – the 'supply' side.

6.1 DEMAND SIDE ISSUES

Fundamental to any analysis of data utilisation is the assumption that decision-makers concerned with the countryside actually need and – in so far as they are available – are prepared to make use of habitat data. Based on the findings of this research (Chapter 3), even this basic assumption needs examination. It is apparent that the mere creation of data does not guarantee appropriate use. In many cases, decisions on policy and management do not appear to be information-driven. Results from the two case studies showed that both sets of organisations appeared to make some decisions on the basis of very little information. In many cases, objective setting and policy formulation were neither a response to, nor based on, information, but were apparently a product largely of historical precedent intermixed with 'professional' judgement.

There are a number of reasons why this may be so. Firstly, those making decisions may in fact be making them from an informed position, founded on a summation of their personal professional knowledge. Secondly, decision-makers may not feel that suitable information is available to support the development of objectives or decision making, and that an information-based approach is not worthwhile. Each of these interpretations merits consideration.

In the first case, it may be true that personal professional knowledge provides a suitable platform from which to make decisions. This is especially valid, perhaps, in situations where scientific knowledge is weak or undeveloped, where general precepts and principles cannot easily be formulated (e.g. because the types of problem that need to be tackled are almost always new so that knowledge is not easily shared), and where practice is essentially localised and case-specific. Whether this scenario accurately describes countryside management and policy in the western world is highly debatable. In recent decades, there has been a considerable growth in, and strengthening of, ecological and landscape sciences, supported in many cases by extensive monitoring, advances in modelling and major developments in theory (e.g, Griffiths *et* *al.* 2000, Haines-Young and Chopping 1996, Jones *et al.* 1996, Servigne 2000, Steele 2000, Wielemaker *et al.* 2001). It is therefore difficult to argue that science is inadequate, and professional expertise alone is sufficient. Moreover, organisational reliance on professional judgement and experience is problematic because they are not amenable to audit and are highly dependent upon the ability and knowledge of specific staff (Burnett *et al.* 1995). In the long term, knowledge can be lost to the organisation when staff leave. Conflict resolution is also difficult and liability (e.g. as a result of mismanagement or poor decision-making) difficult to defend. The lack of hard data can also make it more difficult to win resources in a world which, increasingly, depends on systems of competitive tendering and accountability.

In the second case, staff may consider that no suitable information exists and base decision-making on little or inadequate information. Again, this argument is difficult to sustain in general, given the major improvements that have occurred in countryside survey and monitoring in recent years. Nevertheless, lack of relevant data – or lack of knowledge about the data that do exist – is evident, especially at local to regional scales. This situation can be alleviated by the more widespread provision of data catalogues and metadata implementation, but this is not a simple panacea. In many cases, suitable data may not exist in a readily usable form or may, erroneously, be considered as unsuitable by users and potential users. Often, it appears that potential users are also reluctant to use data because they feel that they lack the appropriate skills or expertise needed to process and interpret them, or perhaps because they see the data as a threat to the position that their experience otherwise gives them within their organisation.

Users' awareness of data sources certainly appears to be influenced by the formats in which they are made available. Data sources are often perceived as being obscure or overly complex, a situation which is compounded by the lack of user or application oriented data 'products' (Burnett *et al.* 1995). In many ways, this is a marketing problem, but is also partly an educative issue. Land cover and land use based data sources, especially, appear to be under-used in habitat mapping and estimation. This is in part due to a lack of understanding of how they are created, what they show and how this is reported. It is also, perhaps, a product of over-expectation (or possibly double-standards) about the accuracy and resolution of land cover data.

Many land cover-based data sets, for example, quote accuracy rates; these are often cited by (non) users as examples of how inaccurate such data are. However, it is extremely unusual to find any quantitative error estimations for ecological field survey or mapping. Cherrill and McClean (1999) quantified observer error for Phase 1 survey within a very limited area and found great variation in categorical attribution. Unpublished work by the (then) Nature Conservancy Council's England Field Unit found large differences in the ability of different surveyors to record plant species within the same quadrat. Angold *et al.* (1996) describe and quantify locational errors in field survey maps, concluding that the level of accuracy must be quantified and made explicit if they are used in databases and GIS. Work conducted for the

BSBI monitoring scheme found bias in botanical surveys associated with the recording behaviour of botanists and systematic error related to recording methods. These biases did not necessarily affect the overall validity of survey results, but were relevant in their interoperation (Rich and Woodruff 1992). Clearly, errors exist in traditional ecological field survey, but the description and quantification of these errors is not common and is rarely explicit. This may lead to a spurious concept of survey accuracy amongst users. In reality, land cover-based data sources may compare favourably with other methods in terms of accuracy.

Another reason cited for the non-use of remotely-sensed data in the case studies was that such data do not adequately resolve the detail required to identify semi-natural habitats. This is partly a question of the inherent resolution of the data and partly related to application. Any classification of the real world inevitably involves generalisation; within broad classifications, such as LCMGB (as expressed in the CIS) this generalisation may be considerable and may indeed present constraints on data use at the local level (Polley 1993). This is potentially a particular problem in the UK countryside, which comprises a complex mosaic of often small and remnant habitat patches, set within a relatively intensively used agricultural landscape. As analysis of the reference data for the study habitats considered in this study show, the majority of habitat sites in the country are small: only 80 ha ($\pm$ 247 ha), (mode 3ha) for saltmarsh, only 29 ha ($\pm$ 257 ha), mode (0.1 ha) for lowland heathland (site sizes were not available from the woodland data but are clearly highly variable). These habitats are also likely to vary considerably in terms of structure, composition and state depending on local environmental conditions (e.g. soil type, drainage, slope angle, aspect) and land use (both current and historical). In addition, the pressures acting on any habitat patch are likely to vary depending on local circumstances, such as the type or intensity of land use in the surrounding area. The resilience of most habitat patches to these pressures is also dependent to a large extent upon their specific characteristics (e.g. size, shape) and their connectivity to other habitats that might act as sources for replenishment – factors which are rarely detectable in generalised data. Furthermore, many management and policy decisions are focused on questions of habitat change rather than stock or quality at a single point in time. Data on change are especially vulnerable to problems of scale and resolution, since change is often small (relative to the initial stock or state of the habitat) and because measurement of change usually depends on the ability to compare data sets from different survey periods: the error in the change data are thus a product of the errors in the two source data sets. Management therefore requires local, site-specific data; interpolation from more generalised data can be potentially dangerous.

Nevertheless, generalised data should still have relevance for the local user, both as a context for policy-making and management, and as a means of communication and learning – a way, for example, of comparing local conditions with those elsewhere. The resolution of many of the available data are also far better than is often implied; if remotely sensed data are trained on a specific target habitat, for a particular application, for example, then fine levels of detail can often be resolved (e.g. Reid and Quarmby 1997). Opportunities also exist to improve on the available land cover data by combining their specific strengths (e.g. identification of spatial extent, feature mapping) with the strengths of field survey-based techniques (detail resolution, description of composition). Combining data from different sources, however, poses its own problems. It multiplies the constraints of knowledge about data availability; it adds to the technical demands on the user; it may add to the cost of data acquisition; and it can add to – as well as reduce – problems of interpretation, not least because errors and uncertainties propagated during the process of data integration and analysis may be even less apparent than with a single data set. Accurate matching of different data sets is also hampered in many cases by the inherent differences between the data sets concerned – for example, in terms of classification system, scale, georeferencing system and survey date. Problems of relating the field-based data from CS90 to the satellite-based data from LCMGB were thus cited as a major constraint on the use of CS data by many users – and was one of the primary motivations for this research. The apparently very low uptake of such techniques may thus be due as much to a lack of 'mapping' awareness in ecological surveyors or the inaccessibility, unavailability or apparent complexity of suitable, complementary data sources as it is to inherent problems with the data themselves. Another reason is undoubtedly the lack of suitable training or tools for the manipulation and reporting of such data. The recent development of desktop GIS and associated tools may help to alleviate this problem, but only if users' perceptions change.

As this implies, the use, or non-use, of information is not solely a question of the availability and quality or technical capability of the data. It depends equally on the experience, attitudes, perceptions and expectations of potential users. These factors operate at both a personal and an organisational, as well as a scientific, level. There is, in particular, a strong link between organisational culture and information use, as was highlighted by the National Park case study. The previous National Park Plan was, by its nature, non-informational. The requirement for a new Park Management Plan, and the development of environmental indicators, required taking an information-led approach. At both a staff and organisational level this meant that the whole way in which objectives were set, defined, communicated and monitored then changed. This type of change also has ramifications for the structure and management of an organisation and the relationships between staff and the management team. In this case, adopting an information-led approach changed the way in which staff were involved in the objective-setting process, and therefore their data and training needs also changed. One of the most significant comments from staff in the objective setting exercises was that they had insufficient information available to them to set realistic objectives for the National Park.

Scalability of countryside data is also an important issue. Scalability is required both to enhance information use within individual organisations, and to improve flow of information and knowledge between the different tiers of decision-making and government (e.g. from local authorities up to national government). Results from the case study on the SNCOs showed that agency staff used different data sources for different tasks at different scales. In itself, this is not surprising. One would expect detailed information about a SSSI to reside in a local team's site file. However, the regional and national-level information, which may include the same SSSI, is often not derived from or linked to that original site-based information and few standards are maintained to ensure comparability. As a result, incompatibilities may exist between the information and knowledge held and used at different levels, and it may be difficult either to translate national guidance or policy to regional or local level or, vice versa, to derive national policy from the sum of local information.

Differences in information requirements can also give rise to a range of problems when data developed for one purpose are used for another. A data source may, for example, be based upon inappropriate measurement thresholds, have been conducted at the wrong time of year, or may use a classification that is insufficiently precise for a different application. Nevertheless, the case studies carried out showed how users often relied on data which they knew were not adequate. Again, there are a range of reasons why this may be so. There are, for a start, several strong motivations to reuse existing data. Data collection is expensive, so collecting new data, to meet a specific need, is often seen as non cost-effective; using existing data, on the other hand, helps to reduce the unit cost of their acquisition. Data collection is also time-consuming; given the often reactive nature of many information needs/enquiries, the use of readily available existing data is therefore essential. Additionally, use of the same data, for different purposes, can help to ensure compatibility. Other, more practical reasons, also exist. The most common appears to be that there is actually no or very little choice available to the data user. In these cases, the data user may be making a rational choice in the face of limited information supply. Such problems were faced by the author in this study in the selection of reference data for testing the target habitat estimates. In such situations, the data user has to make the best of the available data sources. If these data sources are comprehensively documented that task is easier and less prone to error. Where documentation or metadata are lacking, however, there are potential pitfalls, not least the introduction of unknown amounts of uncertainty.

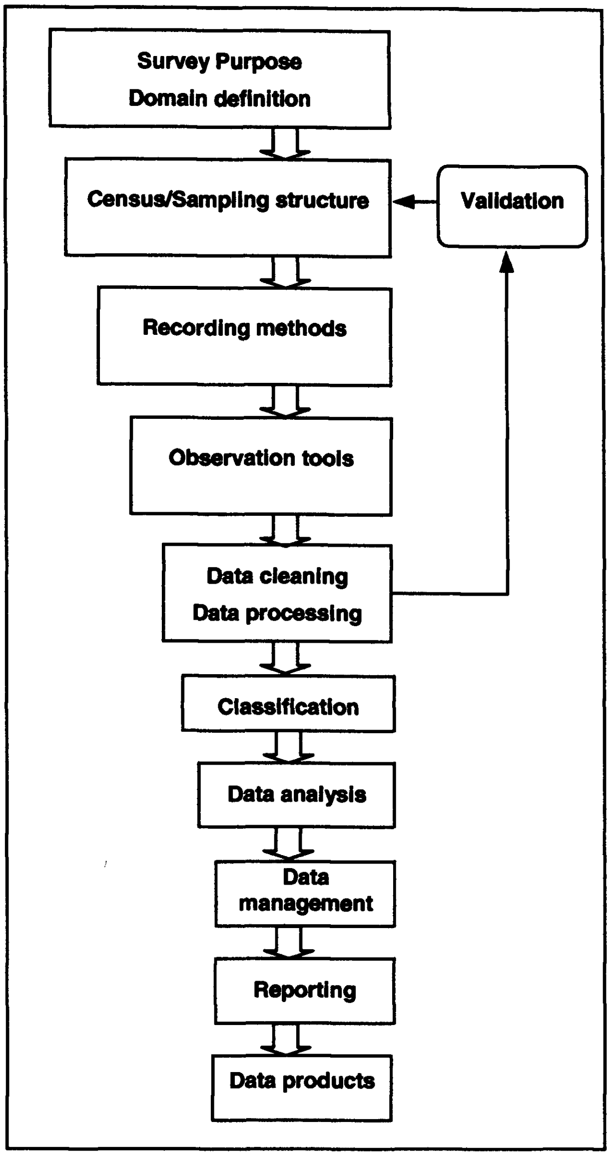
Thus, there are problems in data comparability when using data sources for different applications. However, requirements also change within specific policy areas as policy needs develop of change direction. Just as habitats change, so information requirements also change. Although there is a core of ecological information that is probably always needed for surveillance, management aims are to a large degree linked to the state of knowledge and current environmental priorities and objectives. This is evidenced by the change in emphasis from protected areas to area-wide conservation and management and to the recognition that the wider countryside is important outside small and often isolated protected areas. The advent of biodiversity and sustainability concepts has strengthened the idea of conservation within the wider countryside, as has growing recognition of the need for more inclusive and holistic approaches. Thus, those attempting to provide comprehensive and timely habitat information are faced with a 'moving target'. Firstly, the habitats themselves change and, secondly, the information required changes according to the current policy 'paradigm' (McCollin 2000). The policy paradigm has shifted from the intense study of small protected areas to landscape-level considerations. Increasingly, also, policy requirements have been driven by Europe and the need to report against varied and wide-ranging objectives and targets. In addition, the adoption by many bodies of environmental indicators has created a regular demand for information on habitat loss and/or change, often at a new and higher level of abstraction than previous countryside data (e.g. Briggs *et al.* 1995, HMSO 1996, MacGillivray 1994). Because new indicators are constantly being demanded, to address new problems, data needs are also in a perpetual state of flux. Because habitats are difficult to define and their definition is usually linked to a particular application or conservation objective, habitat data sources are inextricably linked to habitat classifications. Currently the main reporting demands require information to be NVC or Broad Habitat based and links have been established between these two classifications (Jackson 2000). If policy requirements alter, as they did for the Convention on Biological Diversity and Natura 2000, then information demands change. This is most recently illustrated by the need to map open land for the Countryside and Rights of Way Act 2000. This required the identification of landscape types (e.g. mountain, moorland and downland) that could not easily be met through existing land cover or habitat-based data sources (Clark *et al.* 2000). The reasons why existing sources were not suitable were varied and related to specific definitions, but as with other conservation information needs were often related to issues of classification.

6.2 SUPPLY SIDE ISSUES

The provision of information is influenced by a range of factors. Information products are the outputs of a data creation process. This process, or 'data creation chain', as illustrated in Figure 6.1, includes a number of stages that influence the characteristics and utility of both intermediate and final data products. As data pass through each of these filters they are subject to a range of definitions, restrictions, generalisations or assumptions, each of which influences the nature of the output data and its final content and utility. For the conceptual purposes of this diagram there is an implied order to each stage; however, in practice this order may be different and is likely to be iterative.

Most surveys are designed to collect and record the information needed for a specific task or application in the most cost-effective manner whilst providing adequate precision and accuracy. The survey purpose determines what should be measured, the thematic coverage, and to what levels of precision and accuracy. Definitions at this stage are often highly pragmatic, particularly where resources or necessity dictate the use of existing data. For example, the Lowland Heathland Inventory targeted only lowland heath and excluded upland vegetation types. Because it was based upon existing sources, there was little scope for defining the levels of accuracy and precision.

Figure 6.1 The data creation chain process.



In contrast, the Saltmarsh Survey of Great Britain was essentially based upon primary field survey. In this case, resource constraints meant that a full Phase 2 (NVC) survey could not be followed and a hybrid approach between Phase 1 and Phase 2 was adopted. NVC communities were grouped together and, although sketch maps were produced in the field, these were generally not drafted on a topographic base. In both these examples, the surveys met their primary requirements, but the further and future utility of the data sources was severely limited by the methods employed. Ideally, therefore, the purposes of a data source should be defined as a result of an assessment of users' needs and as the result of a data modelling exercise. In many cases these are not conducted and data modelling is often limited. In many cases, also, such assessments are difficult to undertake or, by their very nature, incomplete since data needs often emerge only once potential users can see and handle the data, and because – as noted earlier – data needs are perpetually changing in response to external forces. In some situations, also, the implied order in Figure 6.1 is not followed and surveys are designed on the basis of what can be held in data management systems. These are often based upon flat tables and are unable to manage multiple entity relationships. In such situations, where there is negative feedback in the model, limitations in the data can be created. In such cases there is a need to consider more sophisticated data models and physical implementations.

Amongst the main constraints on the reuse and reanalysis of habitat data is the use of classification nomenclature and its potential for scalability. Data-driven classifications offer the possibility of returning to the original data for re-analysis and re-classification for an alternative application, but only if the sampling technique and measurement methods produced suitable data in the first place. Using NVC survey as an example, the benefits of collecting floristic data during field survey are apparent, for if these are accessible, they provide the opportunity for re-analysis or re-use. However, because NVC samples (quadrats) are placed within stands of homogenous vegetation, the data may be of little use for applications that need to include or evaluate transitional vegetation types. Similarly, the level at which data are made available is important. In the case of Countryside Survey, the field data are only made available on a general basis in an aggregated form and using the Countryside Survey classification's reporting categories. This classification is different from those used by the SNCOs and National Parks. Because the raw data are not easily accessible, comparisons have to be made between the Countryside Survey reporting categories and other data sources. Thus, the use of a classification where the raw data are not accessible imposes a set of limitations related to the smallest descriptive units available within the classification. When comparisons are made between two or more classifications, the use of differently defined categories can lead to the forced comparison of one-to-many or many-to-many relationships and the consequent generation of uncertainty, as illustrated in this research.

The choice of measurement methods and observation tools also has implications for the relative discrimination of data values and for the resolution and precision possible within a data set. Raw data are seldom made available to users; therefore the nomenclature employed, characteristics of reporting and the nature of data products all influence the final utility of a data set. In many cases, data are only published at the very end the 'chain' shown in Figure 6.1; data from intermediate stages may not be available, often for reasons of confidentiality or cost. In these situations, users are tied by the sets of assumptions and treatments applied. Only if access is possible to earlier versions of data is it likely to be possible to re-analyse, recompile and reclassify the data to meet specific needs.

In the majority of cases, access into the data creation chain is limited or non-existent and users are therefore left with no choice other than to use the supplied data products. In many applications users will often require further specificity and will want to re-analyse or re-use part or all of the data. In this situation, each of the factors outlined in the data creation model is relevant, but the classification stage provides the greatest barrier to further adaptability.

6.2.1 Classification

The lack of a single standard method for classifying habitats means that many of the main uses of habitat-related data commonly involve the need to translate data from one classification system into another (Gibson 1996). For example, BAPs are derived from habitats defined by Annex 1 of the EC Habitats Directive and require translation between the Annex 1 and BAP (Broad Habitat) systems (HMSO 1995a, Jackson 2000) and NVC field survey data need to be related to Broad Habitats (Hill *et al.* 2001). Making comparisons between different classifications is therefore necessary, but is often a convoluted and indefinite exercise.

The identification and mapping of habitats is a complex process. Habitats represent a difficult and at times rather ambiguous target. In many ways this is because our concepts of habitats are based upon a mixture of ecological and historical/cultural references (Southwood 1977). Habitats are both a product of the landscape and an integral part of its structure and process.

Thus, the description of habitats involves a variety of criteria such as floristic composition, vegetation structure, edaphic characteristics and environmental gradients. The identification and mapping of a particular habitat requires the specific integration and application of all these criteria, either explicitly as part of the habitat classification or implicitly as unrecognised covariates of habitat distribution. Additionally, habitats vary geographically, in response to local environmental characteristics and to wider biogeographical trends.

At the same time, the very concept of 'habitat' is nebulous and highly contingent; it is effectively a scale-dependent conceptual 'snapshot' of what in reality are actually continua, in both time and space. Spatially, habitats comprise complex mosaics and transitional entities, in many cases lacking clear boundaries: the 'habitat' is thus a human artefact, imposed on reality. Temporally, habitats evolve and change, over different time scales, and in response to both natural succession and anthropogenic influences. Successional change is, for the most part, relatively slow and linear, and is thus predictable. In contrast, effects of human impacts (e.g. habitat loss or modification through development, changes in management or more subtle influences such as pollution or nutrient enrichment) are often abrupt, even catastrophic, and difficult to predict. Thus, in Great Britain, semi-natural habitat distribution is perhaps characterised as much by habitat survival as it is by habitat potential.

By their very nature, therefore, classification systems are generalised models of reality devised to help us conceptualise and describe the environment. As such, they are inevitably impositions upon reality rather than a direct representation of reality. This is one of the characteristics that cause complications in the comparison of different classifications, because classification divisions are created/drawn in different conceptual and spatial frameworks (Lees 1998, Mackey *et al.* 1988).

As already noted, converting data from the classification system in the source data to that needed by the user commonly introduces many-to-many relationships. Even with 'clean' classifications, where data are correctly classified and categories are mutually exclusive and correctly attributed as such, degrees of overlap and underlap will occur. The direction of categorical comparison is therefore relevant – it can be easier to convert classes by aggregating them than it is disaggregating them (Van Beurden and Douven 1999). To enable conversion between land cover definitions Wyatt *et al.* (1993) developed another 'baseline' classification which was used as a key to which all other classifications were compared. This can be a useful approach but can also create an additional filter for categorical association, and thus another layer of potential uncertainty.

There are also problems in matching habitat-based definitions with land-cover ones. LCMGB, for example, employed 18 reporting categories (as presented in CIS) compared to CS90's 58. Inevitably, any given LCMGB category is likely to relate to a number of CS90 categories. Taking the example of woodland, the CS90 category 'broadleaved woodland' relates to both 'deciduous' and 'coniferous' categories in LCMGB and the CS90 category 'conifer woodland' includes elements of both 'coniferous' and 'marsh/rough grass' categories in LCMGB. The marsh/rough grass category is included because the satellite signature for young coniferous plantations is influenced by the land cover type the plantation is on, before canopy closure, and therefore cannot be distinguished within a generalised classification. By contrast, a field-based survey can distinguish such types because further structural, compositional and land use characteristics can be used to define class membership. Therefore, even if the signatures produced from the original Thematic Mapper data are consistently assigned according to their spectral signature, there can still be mismatches between field- and land cover-based data. In practice, this detection-based leakage into other reporting categories will also be dependent, in some cases, upon context. The adjacent land cover may affect the signature of detection and other scattered features may contaminate the overall spectral signature, giving rise to mismatches in classification.

Thus, there are practical and logistical problems in identifying and classifying habitats. Additionally, there are also conceptual problems that help increase uncertainty. One of the problems with accurately comparing classifications is the concept of habitat units and their description. It is possible to describe specific habitats, based upon their core characteristics and features, but it is less straightforward to map habitat boundaries and transitional zones. This issue is exacerbated where classification systems involve defining homogenous areas, like the NVC. In turn, unit identification and description are scale-related; most classifications employ size/area thresholds for feature inclusion. Others may use relative composition (e.g. percentage of coniferous species in a 'mixed' woodland) as a discriminating characteristic. The use of a size threshold can be useful in streamlining surveys and helping to eliminate noise or misclassification; for example defining woodland as groups of trees over 2 ha in area (e.g. Spencer and Kirby 1992). However, to be representative of the resource distribution, size thresholds must be set at levels appropriate to the parcel size and frequency distribution of the habitat in question.

Scale is also important in terms of habitat classification and data scalability. Scale of detection or reporting is related to that of habitat definition and identification. Lowland heathland is a really a biotope complex comprising a range of different habitat types. At one (general) scale, elements of scrub, woodland or valley mire are components of a larger lowland heathland complex. At a finer, sub-landscape scale, they can be considered as habitats in their own right. Thus, definition is both scale and context dependent. If relevant scale and contextual rules are not enforced within a classification, then there are also problems of mutual exclusivity. Thus, there are a number of sources of potential misclassification in habitat estimates:

i. the target is not 'real' – habitats are, to a large extent, conceptual constructs and can reflect divisions imposed upon a continuum. The continuum itself has within class variation and geographical trends. However, if these are systematic they should be predictable;

ii. there may be random variation within classes;

iii. surveys/data may be unable to discriminate the target, due to the insufficient resolution of data and the availability of detectable indicators, both in original data and in reporting;

iv. the mixing of conceptual boundaries with real boundaries (e.g. in mapped data where field boundaries are utilised) can cause confounding.

In addition, when different habitat classifications/data sources are compared, the level of uncertainty is increased because:

i. classifications employ different threshold characteristics, definitions and hierarchies and thus superficially similar categories do not necessarily describe the same thing;

ii. the use of different observation tools/methods introduces further ambiguity;

iii. by definition different classifications/data sources have different (and often unknown) levels of error.

This range of issues raises very real problems in comparing habitat estimates: in short, it is almost impossible to determine to what degree one is comparing estimates of the same thing. For this reason, measures of both quantity and location were used in this research to help identify any systematic discrepancies and statistical analysis was only taken to a level considered consistent with the comparative robustness of the estimates.

Little research appears to have been conducted into the quantification of discrepancies introduced by the comparison of different classifications, perhaps due to lack of data. In this study, it was not possible to further investigate the impact of different classification measures because of the lack of high-resolution comparative data for the same geographical area. By using GIS techniques with detailed and explicit spatial data it should, however, be possible to investigate the implications of using different classifications on the same land area and to quantify the differences between two or more classifications.

6.3 MODELLING HABITAT DISTRIBUTION

The varied, demanding and evolving uses for information place heavy demands upon the available data. Current data sources have a range of limitations, in meeting current needs and in adapting to developing and forthcoming information requirements. The costs of data acquisition are often very high (Burnett *et al.* 1995) and the potential for the creation of new data sets is thus restricted. Because of the large (and changing) range of data uses for different applications there is a requirement to make better use of the available data. Enhancement of the available data is therefore crucial. Fortunately, with the advent of GIS and associated analytical procedures, this may be possible through linkage with other data/variables and by the use of modelling techniques.

For habitat comparison or estimation, modelling essentially involves the interpolation and/or aggregation of available data into different classification system and/or different spatial systems. The results from this study showed that modelling could improve data outputs and facilitate conversion to other classification/geographic systems. GIS provide powerful tools for the integration of different data and methods for spatial transformation, but, in most cases, the success of habitat modelling or data conversion is dependent upon the nature of the habitat and of the data available to describe it. The key determinants are:

i. the characteristics of the source data:
   - spatial referencing;
   - geographical structure and level of aggregation;
   - classification system.

ii. the characteristics of the habitat:
   - geographical structure and distribution;
   - how close it is to the classes in the original or target classification;
   - relationship with covariates.

The nature of the source data determines much of its capability and utility for modelling. The geographical structure of the data and the level of aggregation is crucial and affects their potential for conversion. In this study, variability in spatial referencing caused problems for the development of reference data sets. In the cases of saltmarsh and lowland heathland, data were only initially available in point forms. Thus interpolation methods were needed for creation of a geographical data set. The saltmarsh and lowland heathland inventories were based upon the concept of 'sites'. For saltmarsh, these comprised sets of habitat units based upon a combination of contiguous or semi-contiguous habitat patches grouped by their co-location and sub-divided by membership of an administrative unit (county) (Burd 1989). In the case of lowland heathland, however, a site was usually based upon an existing SSSI unit. Inclusion in the Inventory was based upon the occurrence of lowland heathland vegetation within the SSSI unit which may have included large areas of different vegetation (EN/RSPB 1994/5/6). Thus, these sources presented different kinds of spatial ambiguity.

The CS90 data were interpolated from sample data aggregated at the land class level. Aggregation for large areas such as this limits the utility of the data for use in different geographies/units. At a national level (e.g. GB or country) the output scale is significantly larger than a land class unit, but if the use or conversion of the data are required at a sub-land class level or in different geographies, the user is forced to disaggregate the data and therefore to make assumptions about the habitat distribution. The ITE land classes range in size from 933km^2

to over 15,000km^2 and were differentially representative of the different habitat types tested. If the nature of this heterogeneous relationship is not known modelling is problematic. The LCMGB presented fewer problems in this study because the data were available at a fine-grained level (1km^2) and hence organisation into different (larger) geographies was straightforward.

The classification nomenclature employed can also limit the ability for data comparison. The classification units determine the specificity and relevance of the data. In this study, the inclusivity of CS90 and LCMGB heathland classes caused problems in identifying lowland heathland because this was a subset of the classes available. Saltmarsh proved far less problematic, with clear relevant categories in both CS90 and LCMGB. However, there were differences between the classifications used in both these data sources and that of the reference data. Linkage between the classifications had to be made via a fourth (NVC) classification and this, inevitably, will introduce ambiguity into the results. In some situations even subtle differences in classifications can limit the capability for comparison. For example, the definitions of coniferous/broadleaf woodland overlapped between the sources tested in this study, despite the fact that all three classifications used identified these as separate categories. In this case, it was necessary to combine these classes to avoid creating further classification ambiguity/error and, hence, differentiation between the two types was lost. Thus, although different classifications occupy different conceptual spaces, these can result in concrete effects when converting data or comparing habitat estimates.

The physical characteristics of a habitat also have implications for modelling or data comparisons. The geographical structure and distribution of different habitats are relevant in any attempts to model them. Coastal saltmarsh is a relatively rare habitat occurring in restricted geographical locations in highly variable quantities. This kind of distribution presents problems for sampling/interpolation methods such as those used by CS90. With a highly dispersed and variable resource it is difficult to obtain a representative sample; therefore within a given land class sampling may overlook many small sites or a few very large sites. The generalised CS90 model means that, while a good land class total may be obtained, sub-land class estimates will underestimate where there are large areas of habitat and overestimate where there is little or none.

Variability in the composition of a habitat will also affect the closeness of fit with the original classification used to describe it. If there is much variation, error will be included in the original data and this could cause ambiguity in further comparisons. Biogeographic variability is present in saltmarsh, lowland heathland and woodland in terms of floristic variation (Rodwell 1991a *et seq.*). This was most pronounced in lowland heathland where the habitat definition, at best, includes much variability. Additionally, there is much biogeographic variation: the Norfolk/Suffolk Breckland heathlands have a low ericoid scrub component and a high acid grassland content, and thus they are very different to the Dorset heathlands with high levels of *Calluna* and other ericoids (Farrell 1989).

In circumstances where modelling or interpolation of a habitat is required or desirable, the ability to do so will largely be dependent upon the relationship between the habitat and covariates. If the chosen habitat has few or weak associations with environmental variables it will provide a difficult target. In this study, exogenous data were used in three main ways: to develop control zones, to weight the data within control zones and (using LCMGB) to provide an alternative stratification for CS90 data. The use of exogenous data helped improve some CS90 estimates and in other cases weighting CS90 estimates by the land cover map improved upon 'raw' CS90 estimates and these results suggest that greater intra-land class discrimination is possible. The choice of exogenous data is nevertheless critical. A good exogenous variable should display the following characteristics:

i. a strong association with the target habitat distribution;

ii. known specificity/integrity of this association;

iii. an appropriate range of sensitivity (related to strength of association);

iv. availability across a suitable geographic area;

v. availability at a scale/resolution sufficient to provide appropriate spatial discrimination.

The choice of data for this study was severely limited by availability and cost. Few nationwide (GB) environmental data sets are available in accessible digital formats or at reasonable cost. The external data used here were chosen on an *a priori* basis according to available knowledge of habitat association. The data used were based upon the 1 km resolution employed by CIS. These data were chosen because:

i. an objective of the research was to explore the wider utility of CIS-based data sets;

ii. the CS90 and LCMGB test data were available at the 1 km square level and were not divisible below this resolution;

iii. the spatial referencing of the reference data was not sufficiently robust to go below the 1 km square level;

iv. higher resolution data were only available for some variables and the costs were too high for use in this study.

These exogenous data were admittedly limited, both in the strength and specificity of their association with the target habitats and their resolution/scale. These limitations undoubtedly affects their capability to enhance habitat discrimination, especially of habitats such as saltmarsh and lowland heathland which are related to often subtle and local environmental variation. Saltmarsh, for example, occurs within a specific topographic niche and saltmarsh communities follow salinity gradients. The distribution of lowland heathland is closely linked to local drift geology and soil variations. Such subtle variations are unlikely to be fully discriminated by the exogenous data used in this study. More detailed variables may provide a stronger basis for accurate modelling. In the case of saltmarsh, for example, more detailed topographic data would allow closer resolution of tidal zones, within which saltmarsh be found. Use of CASI data might be helpful in this context (Hill *et al.* 2001). Equally, the particular soil requirements for lowland heathland may be lost in the generalisation when creating the dominant soil type for an entire 1 km square. The same problem was encountered by ITE during the development of a lowland heathland potential map (Barr 1997). When comparing their 1 km soil and land class-derived map with English Nature's data they achieved a correspondence with 55% of sites. The 45% of sites not covered by the map were scattered throughout England, but concentrated in Cornwall and Hampshire, and were related with 1 km squares containing dominant or sub-dominant soil types not associated with lowland heathland. The field sampling exercise associated with the heathland potential map found that just 5% of the habitat mask was estimated to be lowland heathland habitat. The ITE study used more detailed soil data (including sub-dominant soils) than were used here. Their results are not directly comparable with those from this study due to the different methods used but the results found here (e.g. habitat detection of 30% and 2.5% at 100km^2 and 1 km^2 levels respectively) compare favourably. More detailed soils data are also unlikely to improve modelling of woodland distribution because woodland occurs on nearly all major soil types. It has also been argued here that habitat distribution in GB is as much a function of (past and present) land use as it is of environmental potential. If some target habitats have only weak or moderate associations with environmental variables the use of different exogenous data to those used here should therefore be considered. Land use data would have been seriously considered if any suitable sources were available and further research should investigate its potential utility. Developed land was used here as a negative covariate for woodland, but because it only represents a small proportion of the land area, and CS90/LCMGB excluded urban areas, it had only a very marginal affect in the model.

Another alternative to the methods used here would be the use of different modelling/interpolation techniques. One approach may be to use regression mapping, a development of modified areal weighting using statistical and/or regression methods. This approach involves using one or more covariates as a weighting variable to redistribute the data. The relationship between the weighting variables and the target variable may be obtained using regression or other statistical methods (e.g. Dempster *et al.* 1977, Flowerdew and Green 1992) based on a sample of locations. Such methods work well if there is a strong association with available variables, but if regression analysis is carried out in an uncontrolled environment, spurious and counter-intuitive associations may arise. Because of the levels of uncertainty in the reference, test and environmental data available for this study, this was felt to be a major concern here, so the method was not used. Future research might, however, evaluate this approach. Notably, this approach has been found to work well in other areas of environmental science. Briggs *et al.* (2000), for example, showed that regression-based models of air pollution, locally calibrated against monitored data, performed as well as more sophisticated dispersion models. Similarly, following a nationwide comparison of alternative methods, Newton and Herrin (1983) recommended the use of regression-based models over deterministic watershed models for estimating flood flows.

Another possibility could have been the use of some form of spatial interpolation such as kriging or trend surface analysis (Bailey and Gatrell 1995). Kriging is a weighted moving average technique based upon regionalised variable theory. It assumes that the spatial variation in the variable comprises: a systematic spatial trend, a random but spatially dependent variation, and a random, spatially independent component. Kriging is a well-used technique that provides a robust interpolation of point data and an estimate of the error. However, it is sensitive to clustering and assumes spatial dependency, neither of which are necessarily the case with habitats. Trend surface analysis is a technique for spatial interpolation and surface fitting from sample point data that uses an extension of regression analysis to three dimensions, in which the two independent variables are spatial co-ordinates. However, despite offering a reliable modelled surface it is only suited to continuously variable data, relies upon a clear spatial dependence and is sensitive to clustering, hence, it would be an unsuitable method for modelling habitats.

A development of the approach used in this study could be, for example, to obtain higher resolution soils data and combine these with more spatially explicit digital maps for lowland heathland. Such maps do not exist at the moment, although they should be available for the south-west of England in the near future as a result of English Nature's South West Pilot. By using a deductive approach, the relationship between heathland areas and soil type data could be investigated and used to develop more accurate and precise weighting factors that might help achieve better parameterisation of the habitat model.

One of the main problems encountered with lowland heathland estimates was that the use of lowland land classes as a control zone proved insufficiently discriminatory to remove upland heathland types. Employing average altitude data in combination with the lowland land classes improved estimates in some cases but still did not adequately remove the unwanted vegetation types. Using climatic data may provide a means for further discriminating these vegetation types. As with detailed soils data, the use of climatic data would be better suited to a deductive approach rather than the *a priori* method used. Such a method would involve sampling a stratified sub-set of the reference data and using techniques such as multiple regression analysis to establish associations with the environmental variables. This technique was tested for this study at an early stage, but yielded ambiguous results, probably due to the uncertainties in the spatial location of the habitats in the reference data and to the generalised nature of the soils and altitude data used. However, further research, aided by higher resolution data, would undoubtedly prove useful. A different and more novel approach would be to employ basic environmental variables to further subdivide or stratify the ITE land classes. Currently little is known about intra-land class variation, but GIS techniques could be used to locally refine the land class system on an application specific basis. Additionally, further research into habitat associations with different biogeographical regions, such as the English Nature/Countryside Agency Natural Areas (Brooke 1994) or biogeographical zones in Scotland (Carey *et al.* 1994), could provide an alternative spatial stratification to ITE land classes and would help provide an assessment of relative performance between the different geographies.

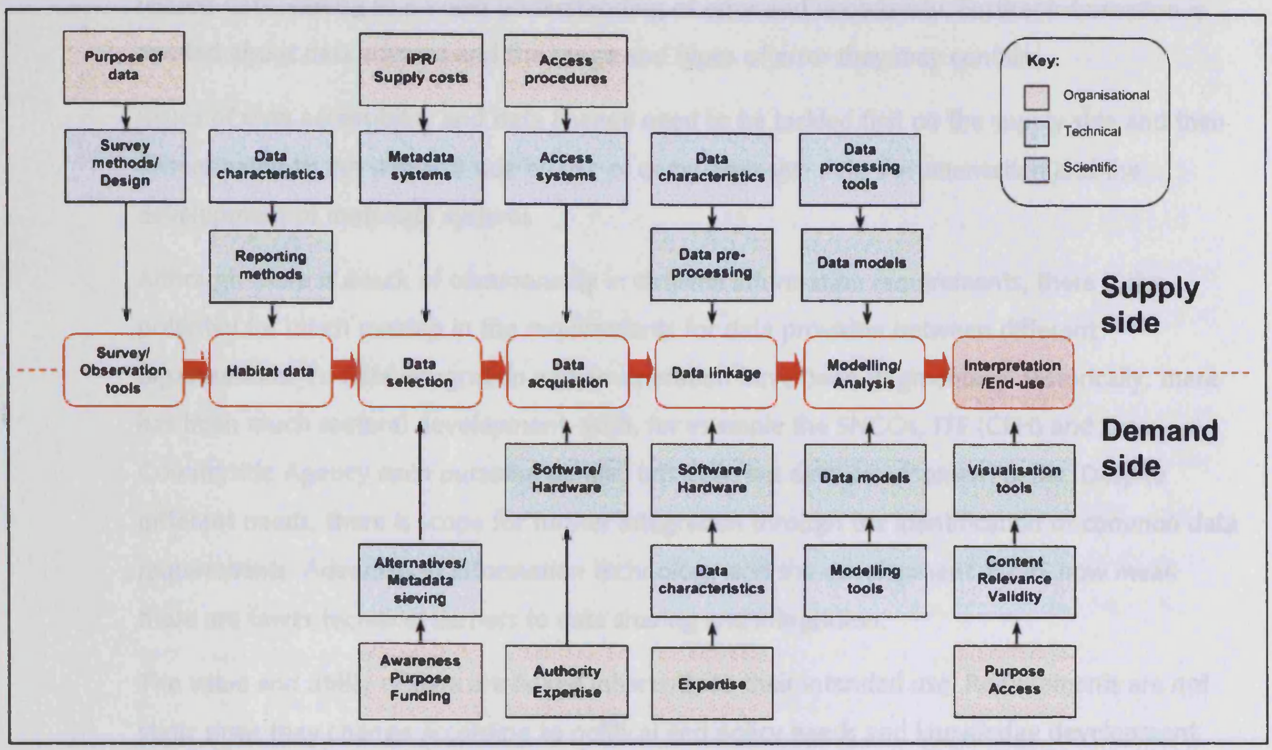
A further avenue for research would be to combine the essentially deterministic approaches used here with more probabilistic approaches. The distribution of habitat *potential*, it can be argued, is largely environmentally determined: it relates to environmental factors such as climate, soils, altitude and the other variables either inherent in the ITE land classification scheme, or incorporated here as additional control zone and weighting variables. The actual occurrence of any habitat within this area of potential, however, is largely a product of contingency: it depends on the many, often random, events and decisions that have affected its survival over time – events such as fires, storms, pollution episodes, changes in land use or changes in management. As a result of these, very little of the area of habitat potential now contains the habitat; as Barr (1997) found, only 5% of the sample squares within a heathland mask contained heathland. These contingent effects cannot be modelled deterministically. They may, however, be amenable to more probabilistic analysis, perhaps using measures such as proximity of urban areas, local land use, slope angle or other factors likely to be correlated with habitat disturbance as indicators of survival. Thus, it would be possible to model, for each land class, the probability distribution of finding the habitat, taking account of local land use and other conditions.

6.4 ADDRESSING THE INFORMATION 'PROBLEM'

Addressing the information 'problem' is a complex task. There are few static policy or management objectives and the complexity of habitats, in terms of variety, composition, structure, geographic distribution/variation and change over time present a difficult target. This research has considered the requirements for habitat-related data, their use and the nature of data content. Essentially, there are two aspects to the problems facing the provision of information: the 'cultural' use of information by individuals and organisations and deficiencies in the actual available data. There are limitations and uncertainties within the data sources currently used, but the use of the available data is in itself sub-optimal.

The organisational and technical issues surrounding the creation and utilisation of data have been considered in terms of supply and demand. To aid this process a model of data utilisation is proposed (Figure 6.2) that describes the processes inherent in information use. The model shows the organisational, technical and scientific processes relevant to the creation, processing and supply of data on the supply side and to the selection and use of information on the demand side.

Figure 6.2 A supply/demand model for data provision and utilisation.



The model shows the influence of organisational, technical and scientific criteria on the availability and utilisation of information. Each of these criteria are related and organisations must work efficiently in each of these spheres in order to make optimal use of their information. The central 'spine' of the model is a functional generalisation of the processes described in Figure 6.1 – the data creation process. At each stage there are a set of organisational, technical or scientific 'filters' that affect the choice or use of data within an organisation.

In order to make better use of available information and to influence the capture and collection of new data, users must follow a more information-based approach. Inherent in doing so is the need to recognise the use and true value of information in supporting work activities. As organisations are increasingly required to work to targets and to report progress and effectiveness, these requirements will grow. Such a change in culture is likely to presage a change in the skill base of many staff and management-related support. There is a need to build technical capacity (the technical level in Figure 6.2) and for management to fully support organisational activity. The work conducted for this study in the North York Moors National Park and within the SNCOs suggested that management recognition of the role and importance of information to operational support and long-term effectiveness was poorly developed. Thus, a more comprehensive understanding of the value of information is needed at a staff and at a corporate level.

In addition to recognition of the value of information to their own work activities, staff also require a fuller understanding of the value of information within their organisation and beyond. Implicit in this is the development of common criteria for the testing and evaluation of habitat-related data leading to a wider understanding of error and uncertainty. Further information is needed about data sources and the range and types of error they may contain.

Issues of data accessibility and data lineage need to be tackled first on the supply side and then disseminated to the demand side by use of comprehensive data documentation and the development of metadata systems.

Although there is a lack of commonality in detailed information requirements, there is the potential for much overlap in the requirements for data provision between different organisations. To date integration and co-operation have been fragmentary. Historically, there has been much sectoral development, with, for example the SNCOs, ITE (CEH) and the Countryside Agency each pursuing parallel but different data development paths. Despite different needs, there is scope for further integration through the identification of common data requirements. Advances in information technology and the development of GIS now mean there are fewer technical barriers to data sharing and integration.

The value and utility of data are linked inherently to their intended use. Requirements are not static since they change according to political and policy needs and knowledge development. Information requirements will tend to evolve, in most cases ahead of data creation/availability. Thus, there will generally be a need for the adaptability of old data to meet new requirements, at least in the short term. The question therefore arises, is it possible to 'future proof' habitat information either by ensuring it has a wide potential utility, or through building in comparability and scalability? Because different applications require different methods, classifications and data products, it is difficult for a data source adequately to meet different information needs. The problems are introduced by the need for different levels of discrimination, precision, accuracy and geographical coverage. The extent to which a data source can be remodelled for alternative use depends largely upon where it is possible to intervene into, or gain access to, the data lineage chain (Figure 6.1). The closer to the beginning of this chain that it is possible to gain access to 'raw' data, the wider the potential utility of the data. In some cases, it is possible to decompose data below measured or published levels (e.g. De Genst *et al.* 1999) but the ability to do so is highly application specific and dependent upon the aggregation methods already employed (Van Beurden and Douven 1999).

The results from this study have shown limited success in decomposing CS90 data from the whole land class level. It is clear that aggregation techniques are important and influence the future utility of the data. Particularly important are the relationships between classifications and areal units (e.g. Openshaw 1981). Any conceptual or spatial framework will limit the potential utility of a data source but some structures may provide a more generic basis for use in a variety of applications. There is also scope for re-assessment of conventional data models. Aspinall and Lees (1994), Lees (1998) and Lees and Ritman (1991) argued that the continuum of change represented by categorical maps in effect imposes a set of overlapping gaussian distributions upon reality, building-in errors of commission and omission to habitat models and imposing a ceiling upon accuracy. Such problems may be overcome by representing species distribution by fuzzy membership methods.

Another possible solution may be to adopt Dunn and Harrison's (1994) 'atomistic' approach whereby features are described by attributes and can be grouped according to user-defined classifications. Using these methods, data-driven classifications could be developed, or existing classifications refined, on a specific application basis. In developing such an approach, perhaps the most difficult element would be deciding how to identify and describe the scale and resolution of the core framework for spatial parameterisation and description, as this will have upstream and downstream implications for the further utility of the nomenclature. Should the framework be related to policy needs or to environmental characteristics? The problem with prioritising policy needs is that they are constantly changing and it is difficult or impossible to predict what they will be in the future. The main alternative is to base any framework upon relatively static environmental variables as, for example, the ITE land class system does in providing both a sampling and modelling framework for Countryside Survey. Difficulties with this approach arise when the chosen or diagnostic variables do not have strong associations with the habitats in question. As has been argued earlier, habitats in GB are as much a function of land use and historical accident as of ecological potential. Thus, the way forward may be to work with a relatively static framework based upon environmental variables, and on an application basis, use other exogenous data to refine models or classifications.

Against this, there is also clearly a need to strive for integration and comparability between data sources and to provide users with more 'standard' information products. However, it is suggested that these objectives are not necessarily in conflict. The key to the problem is the identification of the appropriate scale/resolution to set for the minimum identifiable/mappable unit for the atomistic approach. Ideally this should meet national reporting requirements but also adequately meet many local/regional needs. A balance needs to be struck between scientific rigour and practical needs. Thus, a national standard for habitat classification would be used to describe general habitat units at all scales. A data-driven approach based upon a spatially explicit framework would then be used to provide further detail at a local level. Such a model would therefore service both upstream and downstream information requirements whilst providing room for specialisation within a spatially and thematically consistent framework.

The scale and resolution for a minimum mapping unit should be assessed after further research and based upon a balance of utility to cost effectiveness. It is suggested that the basic descriptive unit for habitat description should be one tied to core reporting requirements and the obvious candidate is the Broad Habitat, which has now been adopted by Countryside

Survey 2000 (Haines-Young *et al.* 2000). Despite some inconsistencies in the classification, the relationships with other widely used systems such as NVC are well understood. If Broad Habitats provide the basic classification framework, spatial descriptors are also required. Conformance with the OS's National Topographic Database/Digital National Framework would provide comparability with other geographic information users and equivalence with a universal spatial standard.

7 SUMMARY AND CONCLUSIONS

7.1 SUMMARY

7.1.1 Information needs and supply

There are a wide range of conservation organisations using countryside information. These share many common characteristics in the information needed for support of policy and management requirements, but they also show many subtle differences.

The link between information needs and information supply is tentative and usually organisationally based. Some organisations act as data suppliers and/or users, though in many cases these roles are poorly defined and rarely fully integrated. There is also often little linkage between different organisations or sectors, despite the fact that they may have overlapping responsibilities and many common information requirements. Since information need tends to drive information supply, the relative lack of homogeneity between the purposes and objectives of different information suppliers and users is likely to lead to important (though perhaps often subtle) differences in the character and ultimate utility of the available data.

Identifying data needs within organisations is not always straightforward. Evidence from the case studies revealed organisational cultures in the SNCOs and North York Moors NP that were not based upon an information-led approach. This has a number of implications. Firstly, at both a staff and organisational level, the lack of perceived value for information has fundamental ramifications for the acquisition and management of suitable data. Secondly, any moves towards an information-based approach would necessitate a change in working practice and a different focus to organisational management. In turn this would change the skills profile required from staff and have implications for training.

The lack of perceived value of information by staff is a question of both attitude and emphasis. When asked to review their work activities and to consider their information use and requirements most staff readily identified shortcomings in information supply or in data sources used. Implicitly, therefore, they recognised the need for information to support their roles. Unprovoked by such questioning, however, they would have rarely considered taking a critical view of their information use or that of their organisation.

Information requirements are generated both internally and externally to an organisation. Many of the external requirements are policy driven and in turn are related to legislative requirements. Neither legislative or policy requirements are static; they change with developing knowledge of ecosystem process and function and according to political paradigms. In the UK, major recent forces for change have included the UN Convention on Biodiversity and the EC Habitats Directive.

In producing habitat information, and in particular in estimating habitat occurrence and distribution, it is necessary to identify and measure various habitat characteristics and to categorise and quantify those of interest. The nature of habitat-related data and their wider utility are governed by the techniques used to collect, process and report them. A variety of methods have been developed to measure and record habitat data, each of which imposes a particular view of 'reality' upon the outputs. Each stage within the collection, processing and reporting of data has implications for the wider utility of the data, but, in particular, the use of classification methodologies can cause major comparability problems between different data sources. They can also constrain fundamentally the ability of users to access the lower level data, in order to customise them to specific needs.

There are a wide range of data sources for habitat and habitat-related data, created by different organisations, using different methods and for different purposes. This heterogeneity creates a lack of comparability in survey methods, classification and reporting frameworks that make it difficult to combine or link data from different sources. From the (potentially) large choice of data available to users, relatively few data sources are easily available or actively used.

The systematic recording and reporting of data quality is important for the value assessment of data sources. However, there are few absolutes in this assessment: data quality relates to both data content and data format and data quality is intimately related to fitness for purpose. Adherence to data standards within the organisations and data sources studied in this research was highly variable. There appeared to be few widely adopted data standards and their routine use, in data creation or evaluation, was not widespread. Conservation staff do not appear to value data quality and standards highly as positive criteria but often cite the lack of them as deficiencies in the data sources they use. The main reasons for the non- or under-use of data sources were related to lack of knowledge of data availability and to the paucity of supporting information. In many cases, data sets were perceived as being poor quality, but with little evidence to support this assertion.

Despite the common requirements for comprehensive and accurate habitat area estimates, land cover-based data sources were not often used by those consulted in this study. It was apparent that many ecologists tend to ignore them because of perceived deficiencies in composition-related data. The hybrid use of land cover data in conjunction with field survey does not appear to have been fully considered or tested.

In considering the requirements for information, the activities and uses of the conservation agencies and evaluating the available data sources, a number of deficiencies are evident in the available data. These are generally contingent to specific applications but include factors such as: information content, timeliness, geographic coverage, temporal coverage, resolution, precision, suitability of data formats and accessibility. Perhaps the first filter on use of any countryside data, however, is knowledge about their availability: users need to be aware of the existence of suitable data sources and this requires access to metadata, especially 'discovery metadata' that can tell users about the availability of potential information sources. The case studies suggested that there is no routine use of discovery metadata and potential users are often unaware of the range of data available.

One of the fundamental requirements in conservation agencies is to obtain information on the occurrence and distribution of semi-natural habitats. Such information is needed at all scales, from local to national, although the precision required changes with scale. Many users have traditionally relied upon field survey to produce habitat inventories to meet these needs. However, these are seldom repeated and, in the past, have not always been spatially explicit. Detailed, or even sometimes general, georeferencing is often lacking. In many ways, the Countryside Survey would appear to provide the solution to many of the information 'gaps' present. It has followed a substantially standard methodology and is regularly repeated. It also provides national coverage and includes a wide range of relevant data on habitat extent and quality. Despite these characteristics, Countryside Survey was little used by the organisations contacted in this study. In part, this is because Countryside Survey was mainly designed to provide national policy support, and the extent to which the data could be used at a regional or local level was unclear. The development of the CIS and parallel developments in GIS nevertheless offer the opportunity to use Countryside Survey/CIS data at different scales and to make use of other environmental data to help to refine habitat estimates. The potential of this approach was therefore investigated in this research.

7.1.2 Countryside Survey/Countryside Information System

Three different habitats were selected for investigation: saltmarsh, lowland heathland and woodland. These were chosen because they are each of great importance in conservation terms, they provide contrasting types of habitat with different spatial distributions, and reference data were available for each of them from independent field-based inventories.

Various methods of modelling habitat distribution were explored. The 'benchmark' approach was to use CS90 data, as available from CIS, in their raw form. Sub-national estimates of habitat area were thus estimated by intersecting the land classes with the relevant 'test' geographies. Two methods for enhancing CS90 data were also investigated: one using control zones to reassign the land class totals to areas considered likely to contain the habitat; the other incorporating also weighting variables to model distributions within these control zones. The capability of the Land Cover Map of Great Britain to estimate habitat distribution was also assessed, using the 1 km^2 data available through CIS. In addition, a hybrid approach ('intelligent weighting') was used, in which LCMGB data were employed to redistribute CS90 totals within land classes.

All these approaches were applied at three geographic levels for each habitat: for counties, for a 100km^2 grid, and for a 1km^2 grid. Performance of the various methods was assessed using measures of mean error (NMSE, RMSE), spatial location (SSI scores) and correlation (r^2).

The best overall estimates for saltmarsh were provided by the LCMGB. However, CS90-based data produced good results for larger areas when combined with a coastal control zone and exogenous weighting factors. Even using the control zone and weighting factors, CS90 overestimated areas with little habitat and underestimated those with larger quantities. It may be possible to improve these results by using more refined weighting factors, but even this might be adversely affected by the unrepresentativeness of the original CS90 sampling.

Lowland heathland estimates were less successful than those for saltmarsh or woodland. The distribution of lowland heathland was poorly predicted and estimated by all the methods used. Despite the use of a control zone and weighting factors, the estimates from all the treatments used were 'contaminated' by the inclusion of upland heath vegetation types originating from the over-inclusive CS90 and LCMGB reporting classes. In most cases the weighting factors degraded model performance, but the use of the control zone provided some improvement to poor estimates at all three scales tested. Inclusion of a control zone (based on land class) provided some improvement to the estimates at all three scales of analysis, though at finer scales especially the performance of the models is still rather poor, particularly for the correct estimation of resource presence. The use of weighting methods (incorporating major soil type and altitude into the CS90 models), fails to improve predictions, however, and in practice slightly worsens model performance. Combination of the CS90 and LCMGB data (in the intelligent weighting method) also failed to improve on the control zone methods.

The results achieved for woodland improved upon the performance of the other two target habitats. CS90 estimates worked well at the county level, although the error levels were relatively high in some cases. Because of the widespread distribution of woodland and its wide environmental tolerances, it was not possible to develop powerful control or weighting variables. Like saltmarsh, the LCMGB provided the best overall estimates. However, the intelligent weighting methods improved the LCMGB data at the county and 10km grid square scales, indicating that CS90 data add useful information to that provided by the satellite-based distribution model.

The CS90 model relies heavily upon the ITE land class system. At the whole land class level, comparisons of CS90 and LCMGB habitat estimates with the reference data for saltmarsh and woodland produced strong associations but relatively poor estimations of habitat area within individual land classes. In the case of saltmarsh, CS90 based measures tended to cause under-estimation in those land classes with small amounts of habitat, and over-estimation in land classes with large areas of habitat. This may be because the field sample squares, on which

CS90 estimates are based, tend to be biased towards areas with larger areas of saltmarsh habitat, while smaller patches of saltmarsh in other habitats are under-reported.

At the within-land class level, associations for LCMGB against reference data showed a marked difference in performance between land classes and between habitats. The performance of land classes was mixed, thus a land class that produced a low association for one habitat did not necessarily do so for another. For woodland, and to a lesser extent for saltmarsh, levels of association showed a geographical pattern across Britain, suggesting that the ability of LCMGB to model habitat distribution varied systematically, either because of differences in the quality of the available imagery, or because of geographical variations in the size, composition, structure (and hence 'detectability') of the target habitats.

Overall, results of these analyses suggested that regional and local (county to 1 km^2) scale estimates of habitat extent and distribution are possible, either by combining CS90 data with other, predictor information, or through use of LCMGB. The value of GIS-based methods, such as control zone estimation and weighted control zone estimation, has also been demonstrated. Nevertheless, scope for considerable improvement in these estimation procedures exists, either by enhancing the source data (CS or LCMGB), or by using different modelling techniques. Improvements to CS data are likely to come mainly from more intensive sampling of the land classes. However, improvements may also be possible by developing a finer level of landscape subdivision within the landclass system – for example, by recognising (and then sampling) subclasses based on some of the variables investigated here (e.g. altitude, soils). LCMGB data might be improved by better discrimination of some habitats, for example using contextual analysis. Such an approach has already been adopted in LCM2000, with the use of image segmentation techniques.

The study suggested a number of ways by which the utility of countryside survey could be improved. Exporting the data from the CIS into a GIS enables the user to recombine and report the data more flexibly and to combine the results with other geographic data. The study showed that sub-national estimates from field and satellite data could yield comparable and relatively accurate results. CS data can be further improved, in specific circumstances, through the use of control zones and areal weighting factors and users should investigate their use for particular applications. The LCM-based data proved the most reliable for mapping habitat distribution and also produced good stock estimates.

Users of CS data should pay careful attention to the nomenclature used by CS and by any other data they use for comparison. For regional and national resource estimates, the results showed that CS90 field survey data yielded good stock estimates. However, if the user requires estimates of habitat distribution for smaller areas, the LCMGB data provides the best solution.

The major barriers to achieving more accurate resource estimation of quantity and distribution relate to sampling and reporting models. Greater specificity and accuracy could be achieved by gaining access to CS field survey data before they are aggregated into the CS reporting classes. This would help diminish ambiguity introduced by the use of different classification schemes. Access to the original sample squares data would allow users to examine in detail the relationship between target habitats and the ITE land classes and enable the investigation of sub-land class divisions. However, confidentiality issues mean that this is not feasible. Despite this, access to field code-based statistics at land class levels would prove useful, but for the average user it would be more helpful to re-engineer the CS90 data into more commonly used classifications such as Broad Habitats, as has now been tested in the development of CS2000 (Barr 1999).

The GIS modelling used here was relatively simple, mainly in consideration of the limited spatial resolution and accuracy of the available data. Various opportunities nevertheless exist to improve on these methods. Improvements in the input data used to define the control and weighting variables would, for example, enhance the analysis – for example by using higher resolution soils data. For some habitats, improvements might also be possible by incorporating additional, or alternative, control and weighting variables, including meteorological variables (e.g. precipitation or temperature) and land use variables (e.g. agricultural land use). Maximum likelihood or regression methods might also be used to help select the most appropriate weighting variables and to derive weights. In addition, more complex models, including interaction terms between the various predictor variables, might be applied. All these possibilities highlight the need for new research.

7.2 CONCLUSIONS AND IMPLICATIONS

7.2.1 Organisational information use

- In the conservation agencies studied, there was little organisational culture of information use. With ever-growing demands for information for a variety of applications, staff training and organisational management must evolve to include a more critical focus upon the use and value of information. Failure to do so will impact upon organisational efficiency and effectiveness.
- Conservation staff had imperfect knowledge of the range of potential data sources available. This can lead to the use of sub-optimal data sources for particular applications.
- Conservation organisations require a more critical and strategic approach to the gathering and assessment of data so that can adequately support their requirements. Addressing this issue will require further staff training in issues related to information gathering, review and management.

- Changing requirements mean that organisations must adopt an iterative approach to securing and developing data sources to meet their requirements. Failure to do so will result in poor quality and outdated information sources.

7.2.2 Data sources

- Although there appear to be a wide range of available data sources, the choice for specific applications can be limited. This can mean that work is not always based upon a sound information base and is likely to have an impact upon the quality of the results.
- To-date, SNCO derived habitat data sources have lacked a geographical basis, and require a more spatially explicit framework to provide comparability and to enable further improvement and update. The lack of a spatial approach can mean that the quantity and distribution of habitats are often unknown in any detail.
- Differences between data sources in terms of data collection methods, scale, data processing, classification and outputs lead to comparability problems that contribute to poor data quality within inventory data sets.
- Comparability issues create uncertainty when attempting to link or integrate different data sources.
- Local and regional applications can work within national standards, but in some cases these require further refinement/detail to encompass local distinctiveness.
- Poor data documentation, and the widespread lack of metadata, compound the above problems. The lack of metadata means that organisations have little idea of the type, quantity and quality of their data holdings. Widespread and comprehensive metadata systems are needed to address the lack of knowledge about data availability and content.
- Data accessibility and use relates to both data content and to data format. The provision of data in unsuitable formats appears to severely limit use. Unsuitable formats include: unavailability in electronic formats, non-standard formats or inappropriate levels of detail or generalisation. Unless these factors are addressed, users will continue to experience problems that limit their ability to make best use of available data.
- Information requirements are likely to continue changing. More widespread, documentation combined with greater transparency in data source creation would enable better use of different data sources in GIS and for novel applications. This would help increase the utility of current data sources.

- Standards provide the key to data comparability; current and future data sources may be improved by greater adherence to common data standards where these exist, and development of standards where they are lacking.
- Greater access to raw data could help the development of application specific data-driven classifications for specific uses. However, these should, where possible, work within established standards.
- Linkage of different data sets would, in many cases, help to add value to the available data, and help to mitigate some of the spatial and temporal limitations of individual data sets. This would require a great deal of work and would be only be appropriate for specific applications.
- GIS technology offers the ability to test and develop new models for habitat distribution. The use of GIS in conservation organisations is varied and follows different development models. Further co-operation and exchange of expertise should be encouraged to aid future co-ordination.
- Research should be undertaken into the development of application specific approaches based upon novel data models and making use of explicit object-based frameworks such as the OS DNF/NTD. The lead should be taken by the SNCOs and could provide the basis for more sophisticated methods for the estimation of habitat stock and change.

7.2.3 Countryside Survey/Countryside Information System

- For an initiative of the scale and cost of the Countryside Survey, its data are used infrequently by conservation agencies. This is a function of perceived data value but is also related to product perception and the non-strategic use of information in conservation organisations. CS2000 has sought to address many of these issues through lengthy policy review. This may mean that CS2000 will be seen in a more favourable light and achieve greater uptake by users than CS90.
- This study found that regional habitat estimates from CS90 data could be improved, in some circumstances, through the use of GIS and additional environmental variables. This suggests that the provision of CS90 data in other forms would be useful to users.
- In most cases, LCMGB provided better estimates of habitat location and quantity than CS90. This was impressive because the LCMGB was originally intended as a research and test exercise and suggests that the more sophisticated LCM2000 will prove a valuable data resource when launched in October 2001.
- In a few limited cases, the use of LCMGB to weight the distribution of CS90 land class estimates improved upon the standard interpolation based upon land classes. This

suggests that the linkage between the two sources, for particular habitats, is variable. Further work should investigate the nature of these relationships.

- The study results suggested that regional habitat estimates may be improved further through:
  - greater within-land class differentiation based upon weighting by environmental variables on an application specific basis;
  - an increase in within-land class sampling intensity or stratification;
  - further refinement of the LCMGB;
  - further integration of CS field survey and LCMGB nomenclature and greater comparability between these classifications and those used by other agencies.
- The wider utility and use of Countryside Survey 1990 field survey and land cover map was hampered by its failure to adopt more commonly used classification systems and thereby provide comparability with the NVC and Broad Habitat systems in use by the majority of practitioners. The Broad Habitat approach has now been adopted by CS2000 and this should greatly enhance its utility and uptake for conservation organisations. This will require the re-engineering of CS90 and previous Countryside Surveys to provide time-series data and may have implications for (historic) data quality.

8 REFERENCES

Adams, W. M. 1996. *Future Nature*, Earthscan, London.

Adinarayana, J., Flach, J. D. and Collins, W. G. 1994. Mapping land use patterns in a river catchment using geographical information systems. *Journal of Environmental Management*, **42**, 55-61.

Angold, P. G., Gurnell, A. M. and Edwards. P. J. 1996. Locational errors in maps from environmental surveys and their implications for information extraction. *Journal of Environmental Management*, **47**, 341-354.

Aspinall, R. J. 1995. Geographical Information Systems: their use for environmental management and nature conservation. *Parks*, **5**, 20-31.

Aspinall, R. and Lees, B. G. 1994. Sampling and analysis of spatial environmental data. in T. C. Waugh and R. G. Healey (Eds) *Advances In GIS Research*, 1086-1099. Taylor and Francis, Southampton.

Austin, G. E., Thomas, C. J., Houston, D. C. and Thompson, D. B. A. 1996. Predicting the environment distribution of buzzard *Buteo buteo* nesting areas using a geographical information system and remote sensing. *Journal of Applied Ecology*, **33**, 1541-1550.

Bailey, T. C. and Gatrell, A. C. 1995. *Interactive spatial data analysis*. Longman Scientific and Technical, London.

Barr, C. 1997. (Ed). *Current status and prospects for key habitats in England. Part 1 Lowland heath landscapes*. Department of the Environment Transport and the Regions, London.

Barr, C. 1999. Broad Habitats – how have they been translated into CS2000? *Countryside Survey 2000 News*. December 1999, **4**, 4-5. Institute of Terrestrial Ecology, Grange over Sands.

Barr, C. .J., Bunce, R. G. H., Clarke, R. T, Fuller, R. M., Furse, M. T, Gillespie, M. K., Groom, G. B., Hallam, C. .J., Hornung, M., Howard, D. C., & Ness, M. G. 1993. *Countryside Survey 1990: Main report*. Countryside 1990 series, Volume 2. Department of the Environment, London.

Bird, A. C., Peccol, E., Taylor, J., Brewer, T. R. and Keech, M. 1994. Monitoring landscape change-the role for GIS. *Landscape Research*, **19**, 120-127.

Birnie, R. V., Morgan, R. J., Bateman, D, McGregor, M. J., Potter, C, Shucksmith, D. M., Thompson,T. R. E., & Webster, J. P. G. 1995. *Review of land use research in the UK*. Report to SOAFD, Edinburgh.

Brabyn, L. 1996. Landscape classification using GIS and national digital databases. *Landscape research*, **21**, 277-300.

Braun-Blanquet, J. 1932. *Plant sociology*. McGraw-Hill, New York.

Briggs, D. J., de Hoogh, C., Gulliver, J., Wills, J., Elliott, P., Kingham, S. and Smallbone, K. 2000. A regression-based method for mapping traffic-related air pollution: application and testing in four contrasting urban environments. *The Science of the Total Environment*, **253**, 151-167.

Briggs, D. J. and Gulliver, J. 2000. *Spatial transformation methods for the analysis of geographic data*. Paper presented to the UN/ECE Work Session on methodological issues involving the integration of statistics and geography, Neuchâtel, Switzerland, 10-12 April 2000.

Briggs, D. J., Kerrell, E., Stansfield, R. and Tantram, D. A. S. 1995. *State of the countryside environment indicators*. Unpublished report to the Countryside Commission. University College Northampton, Northampton.

Briggs, D. J., and Tantram, D. A. S. 1997. *Using GIS for countryside management: the experience of the National Parks*. University College Northampton, Northampton.

Briggs, D. J., Tantram, D. A. S., & Scott, P. 1996. *Improving information for management and decision making in National Parks. The report of the PIMS Development Project*. Unpublished report to the Countryside Commission, Nene Centre for Research, Northampton.

Brooke, D. 1994. A countryside character programme. *Landscape Research*, **19**, 128-132.

Bunce, R. G. H., Barr, C. J., Gillespie, M. K & Howard, D. C. 1996a. The ITE Land Classification: providing an environmental stratification of Great Britain. *Environmental Monitoring and Assessment*. **39**, 39-46.

Bunce, R. G. H., Barr, C. J., Clark, R. T., Howard, D. C. and Lane, A. M. J. 1996b. Land classification for strategic ecological survey. *Journal of Environmental Management*, **47**, 37-60.

Bunce, R. G. H., Barr, C. J., Clark, R. T., Howard, D. C. and Lane, A. M. J. 1996c. ITE Merlewood land classification of Great Britain. *Journal Biogeography*, **23**, 625-634.

Bunce, R. G. H., Morrell, S. K. and Stel, H. E. 1975. The application of multivariate analysis to regional survey. *Journal of Environmental Management*, **3**, 151-165.

Burd, F. 1989. *The saltmarsh survey of Great Britain. An inventory of British saltmarshes*. Research and survey in nature conservation No. 17. Nature Conservancy Council, Peterborough.

Burnett, J., Copp, C. and Harding, P. 1995. *Biological recording in the United Kingdom. Present practice and future development*. Vols. 1 & 2. Department of the Environment, London.

Burrough, P. A. 1986. *Principles of Geographic Information Systems for land resources assessment*. Clarendon Press, Oxford.

Carey, P. D., Dring, J. C. M, Hill, M. O., Preston, C. D. and Wright, S. M. 1994. *Biogeographical zones in Scotland*. Scottish Natural Heritage Research Survey and Monitoring Report No. 26, Scottish Natural Heritage, Perth.

Causton, D. R. 1988. *Introduction to vegetation analysis*. Unwin Hyman, London.

Chapin, F. S., Zavaleta, E. S., Eviner, V. T., Naylor, R. L., Vitousek, P. M., Reynoulds, H. L., Hooper, D. U., Lavorel, S., Sala, O. E., Hobbie, S. E., Mack, M. C., Diaz, S. 2000. Consequences of changing biodiversity. *Nature*, **405**, 234-242.

Cherrill, A. and Lane, M. 1995. The survey and prediction of land cover using an environmental land classification. *Applied Geography*, **15**, 69-85.

Cherrill, A. and McClean, C. 1999. Between-observer variation in the application of a standard method of habitat mapping by environmental consultants in the UK. *Journal of Applied Ecology*, **36**, 989-1008.

Clapham, A. R, Tutin, T. G and Moore, D. M. 1989. *Flora of the British Isles*. Third Edition, Cambridge University Press, Cambridge.

Clark, M., Hill, C. and Priest, S. 2000. *Mapping Open Access Land: assessing the scope and utility of existing datasets.* Unpublished report to the Countryside Agency, GeoData Institute, Southampton.

Clements, F. E. 1928. *Plant succession and indicators.* H. W. Wilson, New York.

Cochran, W. G. 1977. *Sampling techniques.* Third Edition, Wiley, New York.

Commission of the European Communities. 1991. *CORINE Biotopes.* Report and manual, 4 volumes. Office for official publications of the European Community, Luxembourg.

Cooper, A., McCann, T. and Power, J. 1995. Grassland diversity in relation to field parcel size and management. In *Fragmentation in agricultural landscapes. Proceedings of the third annual IALE conference, Preston, 13-14th September 1994.* J. W. Dover (Ed) pp. 62-70. IALE, Aberdeen.

Countryside Agency 1999. *The state of the countryside 1999.* CA 3. Countryside Agency, Cheltenham.

Countryside Commission 1991a. *Fit for the future.* CC P334. Countryside Commission, Cheltenham.

Countryside Commission 1991b. *Landscape change in the National Parks.* Countryside Commission, Manchester.

Countryside Commission 1993. *The National Park Authority. Purposes, powers and administration: a guide for members of National Park Authorities.* CCP 230. Countryside Commission, Cheltenham.

Countryside Commission and Countryside Council for Wales 1996. *Draft guidance on producing National Park Management Plans.* Unpublished report by Countryside Commission and Countryside Council for Wales.

CCW - Countryside Council for Wales 1994. *Annual Report 1994-95.* Countryside Council for Wales, Bangor.

CCW - Countryside Council for Wales 1995. *Corporate Plan 1996-97 - 1998-99.* Countryside Council for Wales, Bangor.

Crawley, M. J. 1987. What makes a community invasible? In *Colonisation, succession and stability.* A. J. Gray, M. J. Crawley & P. J. Edwards (Eds) pp. 429-454. Blackwell, Oxford.

Dargie, T. C. D, Dargie, M. M. & Tantram, D. A. S. 1995 *Lowland wet grassland resource survey. Module 2: Biota reference inventory and quality profiling report.* English Nature Research Report No. 149. English Nature, Peterborough.

Davidson, N. C., d'A. Laffoley, D., Doody, J. P., Way, L. S. Gordon, J., Key, R., Pienkowski, M. W., Mitchell, R. and Duff, K. L. 1991. *Nature conservation and estuaries in Great Britain.* Nature Conservancy Council, Peterborough.

De Genst, W., Canters, F., Jacquet, W. and Vermeersch, S. 1999. Development of a technique to decompose heterogeneous mapping units in a categorical map of the biotic environment using correlated categorical data. *International Journal of Geographical Information Science,* 13, 591-614.

Deichmann, U. 1996. *Smart interpolation. A review of spatial interpolation, database design and modelling.* Use of GIS in Agricultural Research. UNEP GRID Report.

Dempster, A. P., Laird, N. M. and Rubin D. B. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, **39**, 1-38.

DETR 1999a *A better quality of life: a strategy for sustainable development for the UK.* The Stationery Office, London.

DETR 1999b. *Access to the countryside in England and Wales: the Government's framework for action.* Department for the Environment Transport and the Regions, London.

DETR 2000 *Our countryside: the future.* Cm 4909. HMSO, London.

DOE 1993. *Land use change in England, No. 8.* DOE Statistical Bulletin (93)1, Department of the Environment, London.

DOE 1994. *Biodiversity: the UK action plan.* HMSO, London.

DOE 1995 *The Habitats Directive: how it will apply in Great Britain.* DOE, London.

DOE and Welsh Office. 1992. *Fit for the future: a statement by the Government on policies for the National Parks.* DOE/Welsh Office, London.

Dunn, R. and Harrison, A. 1994. *Feasibility study for deriving information about land use stock.* Final report to the Department of the Environment, PECD 7/1/411. University of Bristol, Bristol.

EN - English Nature 1994. *Third Report.* English Nature, Peterborough.

EN - English Nature 1995. *Corporate Plan.* Unpublished document, English Nature, Peterborough.

EN/RSPB. 1994/5/6. *English Nature/Royal Society for the Protection of Birds Lowland Heathland Inventory.* English Nature, Peterborough.

European Commission 1992. *The state of the environment in the European Community: overview,* Volume 3, Com(92) 23, Commission of the European Communities, Brussels.

European Commission 2001. *Sixth Environment Action Programme*: http://europa.eu.int/comm/environment/newprg/index.htm European Commission [accessed 20/06/2001].

Evans, J. 1984. *Silviculture of broadleaved woodland* Forestry Commission Bulletin 62. HMSO, London.

Farrell. L. 1989. The types and importance of British heaths. *Botanical Journal of the Linnean Society*, **101**, 291-299.

FGDC – Federal Geographic Data Committee 1997. *Content standards for digital geospatial metadata.* Revised draft for peer review, April 1997. [online] Federal Geographic Data Committee. Available from: http://www.mews.org/nsdi/revis497.pdf [accessed 18/05/2001].

Fisher, P. and Langford, M. 1995. Modelling the errors in areal interpolation between zonal systems by Monte Carlo simulation. *Environment and Planning A*, **27**, 211-224.

Fishwick, A. and Clayson, J. 1995. *GIS development project.* Final report to the Countryside Commission. Lake District National Park Authority, Kendall.

Flowerdew R, and Green M, 1989. Statistical Methods for Inference between incompatible zonal systems. In *Accuracy of Spatial Databases,* M. Goodchild and S. Gopal (Eds) pp. 239-247. Taylor and Francis, London.

Flowerdew, R. and Green, M. 1992. Developments in areal interpolation methods in GIS. *Annals of Regional Science,* **26**, 79-95.

Flowerdew, R. and Green, M. 1994. Areal interpolation and types of data. In *Spatial analysis and GIS.* S. Fotheringham and P. Rogerson (Eds) pp. 121-145. Taylor and Francis, London.

Forestry Commission 2001 *About the Forestry Commission.* [online] Forestry Commission. Available from: http://www.forestry.gov.uk/website/OldSite.nsf/ByUnique/HCOU-4U4HZT [accessed 20/06/2001].

Fuller, R. and Brown, N. 1996. A CORINE map of Great Britain by automated means. Techniques for automatic generalisation of the Land Cover Map of Great Britain. *International Journal of Geographical Information Systems,* **10**, 937-953.

Fuller, R. M., Groom, G. B. and Jones, A. R. 1994. The landcover map of Great Britain: an automated classification of Landsat Thematic Mapper data. *Photogrammetric Engineering and Remote Sensing,* **60**, 553-562.

Gibson, C. W. D. 1996. *Harmonisation of habitat classifications.* Bioscan (UK), Oxford.

Gilbert, G. & Gibbons, D. 1996 *A review of habitat, land cover and land-use survey and monitoring in the United Kingdom.* Royal Society for the Protection of Birds, Sandy, Beds.

Gimingham, C. H. 1972. *The ecology of heathlands.* Chapman & Hall, London.

Gleason, H. A. 1927. Further views on the succession concept. *Ecology,* **8**, 299-326.

Goodchild, M. F. 1993. Data models and data quality: problems and prospects. In *Visualisation in Geographical Information Systems.* M. F. Goodchild, B. O. Parks and L. T. Steyaert (Eds) pp. 141-149, Wiley, New York.

Green, B. H. 1985. *Countryside conservation: the protection and management of amenity ecosystems.* Allen & Unwin, London.

Greig-Smith, P. 1983. *Quantitative plant ecology.* Third edition. Blackwell, Oxford.

Griffiths, G. H., Lee, J. and Eversham, B. C. 2000. Landscape pattern and species richness; regional scale analysis from remote sensing. *International Journal of Remote Sensing,* **21**, 2685-2704.

Haigh, N. 1991. The European Community and international environmental policy. *International Environmental Affairs,* **3**, 163-180.

Haines-Young, R. H. and Chopping, M. 1996. Quantifying landscape structure: a review of landscape indices and their application to forested landscapes. *Progress in Physical Geography,* **20**, 418-445.

Haines-Young, R. H., Barr, C. J., Black, H. I. J., Briggs, D. J., Bunce, R. G. H., Clarke, R. T., Cooper, A., Dawson, F. H., Firbank, L. G., Fuller, R. M., Furse, M. T., Gillespie, M. K., Hill, R., Hornung, M., Howard, D. C., McCann. T., Morecroft, M. D., Petit, S., Sier, A. R. J., Smart, S. M., Smith, G. M., Stott, A. P., Stuart, R. C. and Watkins, J. W. 2000. *Accounting for nature: assessing habitats in the UK countryside.* Department of the Environment, Transport and the Regions, London.

Hall, J. 1996. *An NVC database for woodlands.* English Nature Research Report No. 181, English Nature, Peterborough.

Harrison, A. R., Dunn, R., Brown, L., and Turton, P. J. 1991. *The use of Geographical Information Systems in the analysis of countryside data.* Final report to the Department of the Environment, PECD/7/7/77. HMSO, London.

Hellawell, J. M. 1992. Development of a rationale for monitoring. In *Monitoring for conservation and ecology.* F. B. Goldsmith (Ed) pp. 1-14, Chapman and Hall, London.

Hill, M. O. 1979. *TWINSPAN a FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes.* Ithaca, New York.

Hill, C., Ball, J., Dargie, T. C. D., Tantram, D. A. S. and Boobyer, M. G. 2001. *Maritime cliffs and slope inventory.* English Nature Research Report MAR 02-03-02, English Nature, Peterborough.

Hill, M. O., Bunce, R. G. H., and Shaw, M. W. 1975. Indicator species analysis: a divisive polythetic method of classification and its application to a survey of native pinewoods in Scotland. *Journal of Ecology*, **63**, 597-613.

Hill, R. A., Fuller, R. M., Smith, G. M. and Veitch, N. 2001. Classification of airborne CASI and LIDAR data of selected CS2000 sample squares. Countryside Survey 2000. Module 8: Airborne Scanner Applications. Contract Report to DETR, CR0210. Centre for Ecology and Hydrology, Huntingdon.

HM Government 1990. *This common inheritance: Britain's environmental strategy*, HMSO, London.

HMSO 1949. *1949 National Parks and Access to the Countryside Act.* HMSO, London.

HMSO 1974. *Report of the National Park Policies Review Committee 1974.* HMSO, London.

HMSO 1988. *Norfolk and Suffolk Broads Act 1988.* HMSO, London.

HMSO 1990. *Environmental Protection Act 1990 (c.43).* HMSO, London.

HMSO 1994a. *Biodiversity: The UK action plan.* HMSO, London.

HMSO 1994b. *The Conservation (Natural Habitats & c.) Regulations 1994.* Statutory Instrument No. 2716. HMSO, London.

HMSO 1995a *Biodiversity: The UK Steering Group report. Volume 1 meeting the Rio challenge.* HMSO, London.

HMSO. 1995b *Biodiversity: The UK Steering Group report. Volume 2 action plans.* HMSO, London.

HMSO 1995c. *Environment Act 1995.* HMSO, London.

HMSO 1996. *Indicators of sustainable development for the United Kingdom.* HMSO, London.

HMSO 2001. *The Countryside and Rights of Way Act 2000.* [online] HMSO. Available from: http://www.legislation.hmso.gov.uk/acts/acts2000/20000037.htm [accessed 20/06/2001].

Howard, D. C. and Bunce, R. G. H. 1996. The Countryside Information System. a strategic-level decision support system. *Environmental Monitoring and Assessment*, **39**, 373-384.

Howard, D. C., Jones, M., and Haines-Young, R. H. 1994. *Development of the Countryside Information System.* Countryside 1990 Series, Volume 4. Department of the Environment, London.

Hudson, G., Towers, W., Bibby, J. S. and Henderson, D. J. 1982. *Soil and land capability for agriculture: The Outer Hebrides 1:250,000 Sheet 2.* Soil Survey of Scotland. The Macaulay Institute for Soil Research, Aberdeen.

Hunting Technical Services 1986. *Monitoring landscape change*: Final report to Department of the Environment. Hunting Technical Services, Borehamwood.

Ireland, P. A. R., and Maycock, P. J. 1996. *Countryside Information System data management contract: final report.* Department of the Environment, Bristol.

IUCN. 1994. *Parks for life: action for protected areas in Europe.* IUCN, Cambridge.

Jackson, D. L. 2000. *Guidance on the interpretation of the Biodiversity Broad Habitat Classification (terrestrial and freshwater types): Definitions and the relationship with other habitat classifications.* JNCC Report, No. 30. Joint Nature Conservation Committee, Peterborough.

Jarvis, R. A., Bendelow, V. C., Bradley, R. I., Carroll, D. M., Furness, R. R., Kilgour, I. N. L. and King, S. J. 1983. *National Soil Map Sheet 1 - Northern England.* 1:250,000 scale, Ordnance Survey, Southampton.

Jarvis, R. A., Bendelow, V. C., Bradley, R. I., Carroll, D. M., Furness, R. R., Kilgour, I. N. L. and King, S. J. 1984. *Soils and their use in Northern England.* Soil Survey; Bulletin No 10.

JNCC - Joint Nature Conservation Committee 1994a. *Guidelines for selection of biological SSSIs: bogs.* Peterborough, Joint Nature Conservation Committee.

JNCC – Joint Nature Conservation Committee 1994b. *Annual Report 1994.* Joint Nature Conservation Committee, Peterborough.

JNCC – Joint Nature Conservation Committee 1995. *Work Plan 1995-96.* Unpublished document, Joint Nature Conservation Committee, Peterborough.

JNCC - Joint Nature Conservation Committee 1996. *Guidelines for selection of biological SSSIs: inter-tidal habitats.* Joint Nature Conservation Committee, Peterborough.

JNCC - Joint Nature Conservation Committee 1997. *Corporate Plan 1997-982000.* Unpublished document, Joint Nature Conservation Committee, Peterborough.

JNCC - Joint Nature Conservation Committee 2001a. *Other ways that sites can be protected.* {online] JNCC. Available from: http://www.jncc.gov.uk/idt/ramsar_spa/others.htm. [accessed 19/06/2001].

JNCC - Joint Nature Conservation Committee 2001b. *National report of the United Kingdom for COP7.* [online] JNCC. Available from: http://www.ramsar.org/cop7_nr_uk.htm [accessed 19/06/2001].

Jones, C. B., Kidner, D. B., Luo, L. Q., Bundy, G. L1. and Ware, J. M. 1996. Database design for a multi-scale information system. *International Journal of Information Systems,* **10**, 901-920.

Kaiser, J. 2001. Climate change – 17 national academies endorse Kyoto. *Science,* **292**, 1275-1275.

Kashyap, P. S. and Panda, R. K. 2001. Evaluation of evapotranspiration estimation methods and development of crop-coefficients for potato crop in a sub-humid region. *Agricultural Water Management*, **50**, 9-25.

Kerr, R. A. 2001. Climate change – it's official: humans are behind most of global warming. *Science*, **292**, 566-566.

Kirby, K. 1995. *Rebuilding the English countryside: habitat fragmentation and wildlife corridors as issues in practical conservation.* English Nature Science No. 10. English Nature, Peterborough.

Krebs, C. J. 1978. *Ecology: the experimental analysis of distribution and abundance.* Second Edition. Harper and Row, New York.

Kukkonen, J., Hakrkoknen, J., Walden, J., Karppinen, A. and Lusa, K. 2000. Evaluation of the CAR-FMI model against measurements near a major road. *Atmospheric Environment*, **35**, 949-960.

Langford, M. and Unwin, D., J. 1992. Generating and mapping population density surfaces within a GIS. *The Cartographic Journal*, **31**, 21-26.

Langford, M., Unwin, D., J. and Maguire, D., J. 1990. Generating improved population density maps in an integrated GIS. In: J. Harts, H.F.L. Ottens and H.J. Scholten (Eds) pp. 651-660. *EGIS '90: Proceedings of the First European Conference on Geographical Information Systems.* EGIS Foundation, Utrecht, Netherlands.

Lanter, D. P. 1990. *Lineage in GIS: the problem and solution.* NCGIA Technical Paper 90-6, Santa Barbara, California, USA.

Lavers, C., P. and Haines-Young, R. H. 1996. Using models of bird abundance to predict the impact of current land-use and conservation policies in the flow country of Caithness and Sutherland, Northern Scotland. *Biological Conservation*, **75**, 71-77.

Lees, B. G. 1998. *Improving the spatial extension of point data by changing the data model.* http://geography.anu.edu.au/people/lees/santafe2.html [accessed 19/06/2001].

Lees, B. G. and Ritman K. 1991. Decision-tree and rule-induction approach to integration of remotely sensed and GIS data in mapping vegetation in disturbed or hilly environment. *Environmental Management.* **15**, 823-831.

Lescourret, F and Genard, M. 1994. Habitat, landscape and bird composition in mountain forest fragments. *Journal of Environmental Management.* **40**, 317-328.

Locke, G. M. L. 1987. *Census of Woodlands and Trees 1979–82.* Forestry Commission Bulletin, 63, HMSO, London.

Long, B. G. and Skewes, T. D. 1996. A technique for mapping Mangroves with Landsat TM satellite data and geographic information system. *Estuarine, Coastal and Shelf Science*, **43**, 373-381.

MacGillivray, A. (Ed) 1994. *Environmental measures: indicators for the UK environment.* New Economics Foundation, London.

Mackey, B. G., Nix, H. A., Hutchinson, M. F., McMahon, J. P. and Fleming, P. M. 1988. Assessing representativeness of places for conservation reservation and heritage listing. *Environmental Management*, **12**, 502-514.

MapInfo 1998a. *MapInfo Professional user's guide.* MapInfo Corporation, Troy, New York.

MapInfo 1998b. *MapBasic user's guide.* MapInfo Corporation, Troy, New York.

Marrs, R. H. 1993. An assessment of change in *Calluna* heathlands in Breckland, eastern England, between 1983 and 1991. *Biological Conservation,* **65**, 133-139.

McCollin, D. 2000. Editorial: Hedgerow policy and protection – changing paradigms and the conservation ethic. *Journal of Environmental Management,* **60**, 3-6.

Michael, N. 1994. Lowland heathland. In *Habitat conservation in England.* English Nature Research Report No.96. *A. M.* Moffat (Ed) pp. 103-113. English Nature, Peterborough.

Milton, N. 1994. Nature conservation and arable farming. *British Wildlife,* **5**, 229-335.

MLURI 1982. *The Outer Hebrides 1:250,000 Sheet 2. Soil Survey of Scotland.* The Macaulay Institute for Soil Research, Aberdeen.

MLURI 1984. *Organisation and methods of the 1:250,000 soil survey of Scotland.* Soil Survey of Scotland. The Macaulay Institute for Soil Research, Aberdeen.

Morgan, C. 1999. Journals metadata: information about content. *Learned Publishing,* **12**, 191-195.

Moss, D. and Wyatt, B. K. 1994. The CORINE Biotopes Project: a database for conservation of nature and wildlife in the European Community. *Applied Geography,* **14**, 327-349.

Moxley, A. and Allanson, P. 1994. Areal interpolation of spatially extensive variables: a comparison of alternatives. *International Journal of Geographical Information Systems,* **8**, 479-487.

NBN – National Biodiversity Network 2001. *NBN Index.* [online] NBN. Available from: http://www.nbn.org.uk (Framed site/Index menu). [accessed 20/06/2001].

NCC - Nature Conservancy Council 1984. *Nature conservation in Great Britain.* Nature Conservancy Council, Shrewsbury.

NCC - Nature Conservancy Council 1989. *Guidelines for selection of biological SSSIs.* Nature Conservancy Council, Peterborough.

NCC - Nature Conservancy Council 1990. *Handbook for Phase 1 habitat survey: a technique for environmental audit.* Nature Conservancy Council, Peterborough.

NERC – Natural Environment Research Council. 1996. CIS - *Countryside Information System V5.40.* 1996. Computer software for MS Windows. NERC.

Newton D. W. and Herrin J. C. 1983. Estimating flood frequencies at unguaged locations. In *Proceedings of the Conference in Hydraulic Engineering,* pp. 528-533, American Society of Civil Engineers, New York.

NGDF – National Geospatial Data Framework 2000. *Discovery metadata guidelines Version 1.2.* National Geospatial Data Framework Management Board, Southampton.

OECD – Organisation for Economic Co-operation and Development. 1991. *Environmental indicators.* OECD, Paris.

Openshaw, S. 1981. *The modifiable areal unit problem.* Geo Books, Norwich.

OS – Ordnance Survey 2001. *The Digital National Framework.* [online] OS. Available from: http://www.ordnancesurvey.gov.uk [accessed 20/06/2001].

Peccol, E., Bird, A. C. and Brewer, T. R. 1996. GIS as a tool for assessing the influence of countryside designations and planning policies on landscape change. *Journal of Environmental Management,* **47**, 355-367.

Peck, D. 1996. *The Ramsar Convention on Wetlands.* [online] The Ramsar Bureau. Available from: http://www.ramsar.org [accessed 22/06/2001].

Pepper, D. 1989. *The roots of modern environmentalism.* Routledge, London.

Peterken, G. F. 1981. *Woodland conservation and management.* Chapman and Hall, London.

Phillips, M. and Huggett, D. 2001. From passive to positive – the Countryside Act 2000 and British Wildlife. *British Wildlife,* **12**, 237-243.

Polley, M. D. 1993. *The use of remotely sensed data in English Nature. Phase 1 – a comparison of the ITE Land Cover Map with other data sets.* Geographic Information Unit, English Nature, Peterborough.

Porter, R. 1985. Holes in Part 1 of the Wildlife and Countryside Act. *Ecos,* **6**, 5-8.

Rackham, O. 1990. *Trees and woodlands in the British Landscape.* Revised Edition. Dent, London.

Ratcliffe, D. A. 1977. (Ed) *A nature conservation review.* 2 vols. Cambridge University Press, Cambridge.

Reid, E. and Quarmby, N. 1997. Blanket coverage: mapping and assessing Scotland's bogs. *Mapping Awareness,* **11**, 5, 34-37. Also available [online] from: http://www.geoplace.com/ma/1997/0597/ma970500feat1.asp [accessed 21/06/2001].

Rhind, D., and Hudson, R. 1980. *Land use.* Methuen, London.

Rice, T. 1994. *Losing interest: a survey of threats to Sites of Special Scientific Interest in England and Wales.* Friends of the Earth, London.

Rich, T. C. G. and Woodruff, E. R. 1992. Recording bias in botanical surveys. *Watsonia,* **19**, 75-95.

Robinson, G. 1999. New business model, pricing and licensing announced for Ordnance Survey. *Mapping Awareness,* **13**, 9, 17-19.

Rodwell, J. S. (Ed). 1991a. *British plant communities, Volume 1: woodlands and scrub.* Cambridge University Press, Cambridge.

Rodwell, J. S. (Ed). 1991b. *British plant communities, Volume 2: mires and heaths.* Cambridge University Press, Cambridge.

Rodwell, J. S. (Ed). 1992. *British plant communities, Volume 3: grassland and montane vegetation.* Cambridge University Press, Cambridge.

Rodwell, J. S. (Ed). 1995. *British plant communities, Volume 4: aquatic communities, swamp and tall herb fens.* Cambridge University Press, Cambridge.

Rodwell, J. S. (Ed). 2000. *British plant communities, Volume 5: maritime communities and vegetation of open habitats.* Cambridge University Press, Cambridge.

Rowell, T. 1991. *SSSIs: a health check.* A *review of the statutory protection afforded to Sites of Special Scientific Interest in Great Britain.* Wildlife Link, London.

Rowell, T. A., and Reed, T. M. 1994. Monitoring in the United Kingdom: an integrated perspective. In *Ecosystem monitoring in protected areas: 2nd international conference on science and management of protected areas (SAMPA),* n. k.

Royal Commission on Environmental Pollution. 1994. *Transport and the environment.* HMSO, London.

Seabra, C. 1996. *CORINE Land cover directory.* Centro Nacional de Informação Geográfica, Portugal.

Servigne, S., Ubeda, T., Puticelli, A. and Laurini, R. 2000. A methodology for spatial consistency improvement of geographic databases. *GeoInformatica,* **4**, 7-34.

Sinclair, G. 1992. *The lost land: land use change in England 1945-1990.* Blackmore Press, Dorset.

Smith, R. 1995. *The Association of National Parks: annual review 1995.* ANPA, Newton Abbott.

SNH - Scottish Natural Heritage 1994. *Annual Review 1993-94.* Scottish Natural Heritage, Edinburgh.

SNH - Scottish Natural Heritage 1995. *Plans and Progress 1995-96.* Scottish Natural Heritage, Edinburgh.

Southwood, T. R. E. 1977. Habitat, the templet for ecological strategies? *Journal of Animal Ecology,* **46**, 337-365.

Spencer, J. W. and Kirby, K. J. 1992. An inventory of ancient woodland for England and Wales. *Biological Conservation,* **62**, 77-93.

Statham, D. 1991. *North York Moors National Park Plan: second review.* North York Moors National Park, Helmsley.

Steele, B. M. 2000. Combining multiple classifiers: an application using spatial and remotely sensed information for land cover type mapping. *Remote sensing of environment,* **74**, 545-556.

Stehman, S. V. 1999. Comparing thematic maps based upon value. *International Journal of Remote Sensing,* **20**, 2347-2366.

Stoate, C. 1996. The changing face of lowland farming and wildlife Part 2: 1945-1995. *British Wildlife,* **7**, 162-172.

Stott, A. 1994. Monitoring land-use change-the context for Countryside Survey 1990. *Ecos,* **15**, 3-8.

Tantram, D. A. S. 1994. *Woodlands: Colne Valley countryside management strategy.* Research document No. 6. Terra Environmental Consultancy, Huddersfield.

Tantram, D. A. S., Dargie, T. C. & Kent, M. K. 1997. *Habitat information scoping study,* Report to the Joint Nature Conservation Committee, Terra Environmental Consultancy, Northampton.

Thomas, R. C. & Phillips, P. M. 1994. *Ancient Woodland Inventory (provisional) England.* English Nature Research Report No. 72. English Nature, Peterborough.

Tomkins, S. 1989. *Forestry in crisis.* Helm, London.

United Nations 2001. *Convention on Biological Diversity.* United Nations [online] Available from: http://www.biodiv.org United Nations [accessed 22/06/2001].

Van Beurden, A. U. C. J. and Douven, W. J. A. M. 1999. Aggregation issues of spatial information in environmental research. *International Journal of Geographical Information Science,* **13**, 513-527.

Warren, A. and Goldsmith, F. B. 1974. *Conservation in practice.* John Wiley & Sons, London.

Watt, A. S. 1947. Pattern and process in the plant community. *Journal of Ecology,* **35**, 1-22.

Webb, N. R. and Hopkins, P. J. 1984. Invertebrate diversity on fragmented *Calluna* heathlands. *Journal of Applied Ecology,* **21**, 921-933.

Wielemaker, W. G., de Briun, S., Epema, G. F. and Veldkamp, A. 2001. Significance and application of the multi-hierarchical land system in soil mapping. *Catena,* **43**, 15-34.

Williamson, M. 1981. *Island populations.* Oxford University Press, Oxford.

World Commission on Environment and Development. 1987. *Our common future.* Oxford University Press, Oxford.

Wright, D. 1998. *The national inventory of woodland and trees. Information Note FCIN8.* Forestry Commission, Edinburgh.

Wright, J. K. 1936. A method of mapping densities of population with Cape Cod as an example. *Geographical Review,* **26**, 103-110.

W. S. Atkins. 1996. *Environmental Catalogue V.2 1996.* Software for IBM PC.

Wyatt, B. K., Greatorex-Davies, J. N., Hill, M. O., Parr, T. W., Bunce, R. G. H. and Fuller, R. M. 1993. *Comparison of land cover definitions.* Countryside 1990 series, Volume 3. DOE, London.

Wyatt, G. 1991. A *review of Phase 1 habitat survey in England; volume 1 General report.* Great Britain nature conservation resource survey project No. 4. Nature Conservancy Council, Peterborough.

Wynne, G., Avery, M., Campbell, L., Gubbay, S., Hawkswell, S., Juniper, T., King, M., Newberry, P., Smart, J., Steel, C., Stones, T., Stubbs, A., Taylor, J. Tydeman, C., and Wynde, R. 1995. *Biodiversity Challenge.* Second Edition. Royal Society for the Protection of Birds, Sandy.

Wynne, G., Avery, M., Campbell, L., Juniper, T., King, M., Smart, J., Steel, C., Stones, T., Stubbs, A., Taylor, J. Tydeman, C., and Wynde, R. 1993. *Biodiversity Challenge.* Royal Society for the Protection of Birds, Sandy.

Young, S. 1993. *The politics of the environment.* Baseline, Manchester.

APPENDIX 1

CLAUDE PROJECT DATA SOURCE PROFORMA

Source Name	
Purpose and countries	
Originating organisation & contact	
Implementing organisation/contractor	
Availability/Confidentiality/Cost	
Copyright	
Geographical coverage	
Classes/classification used	
Start Date	
Last/end Date	
Updates/Frequency **Ongoing?**	
Timescale (frequency of observations, duration of sampling/averaging times)	
Resolution/scale	
Accuracy	
Precision, level of aggregation	
Georeferencing (point/boundary/polygon)	
Representativeness	
Format (forms, report, maps) **Digital availability**	
Documentation - Lineage/Audit trail/version control	
References	

CLAUDE PROJECT DATA PROFORMA

1	**Name**	
2	**Organisation**	
3	**Type of organisation**	
4	**Address**	
5	**Phone**	
6	**Fax**	
7	**Email**	
8	**Data products supplied**	Complete data supplier form
9	**Data used**	Complete data user form
10	**Work conducted in relation to LU/LC**	
11	**Scale of working**	
12	**Difference between LU & LC**	
13	**Work with partners**	
14	**Do you monitor LU/LC *change***	
15	**Main problems encountered in work regarding LU/LC data**	
16	**Needs/wishes for the future**	
17	**Is CLAUDE useful**	
18	**Further comments**	
19	**Would like further information**	

CLAUDE PROJECT DATA SUPPLIER PROFORMA

DATA SOURCE	
Supplier	
Do you supply other data in combination with this source	
What type of organisations have you supplied this data source to	
What applications are these data used for, and are all of these appropriate uses of the data	
What are the major limitations of these data	
Do you have plans to change any aspects of data supply or specification - please detail	

CLAUDE PROJECT DATA USER PROFORMA

DATA SOURCE	
User	
What applications do you use these data for. **Do you think these are appropriate uses**	
Do you use other data in combination with this source	
How far do you find this source compatible with others	
What are the major limitations of this data source	
If you have problems what do you think are the possible solutions	

APPENDIX 2

SNCO WORKSHOP QUESTIONNAIRE

All answers will be treated in the utmost confidence. We are only asking for names so we can clarify answers if necessary.

Questions are in *italics.*

Name: ***Agency:*** ***Position & contact No.:***

1. HABITAT INFORMATION

What do you think habitat related data are ?

1.1 Use the list below as a guide, and add to it if you can...

habitat descriptions/classifications (quantitative or qualitative)

species distribution and abundance

soils, geomorphology, geology

associated environmental data e.g. nitrogen deposition, air quality, water quality

data on conservation designations, land use, land use restrictions

data on nature and magnitude of threats and impacts

Please give your definition of habitat-related data...

1.2 Please give an example of your use of habitat-related data from your own area of work...

2. *KEY BUSINESS ACTIVITIES*

The table below outlines a possible general classification for 'key business activities' (KBA's) of the country agencies. It is not intended to be rigid or necessarily comprehensive.

2.1 *Please add any categories you think have been missed to columns 1&2 (and number them in sequence), and indicate in columns 3&4 (using a five point scale, 1-often to 5-not at all) how often you have used or supplied information for any of these activities that you have a working knowledge of.* In column 5 please indicate (5 point scale, 1-important to 5-not important) how important habitat-related data is to each KBA. Only fill in those that are **important to you.**

KEY BUSINESS ACTIVITY		**How often do you:**		
1	**2**	**3** **Use habitat related information**	**4** **Supply habitat related information**	**5** **Importance of habitat related information**
A. Legal Requirements				
1 W&C act				
2 SSSIs	Site Integrity Monitoring (SIM)			
	Site Quality Monitoring (SQM)			
3 NNRs				
4 Habitats Directive				
5 Planning casework				
B. Advice				
1 To govnt.- internal	1 General policy			
	2 Specific cases			
2 To govnt.- external	1 Europe - EC/EU			
	2 International - IUCN, UN			
	3 Public bodies/QUANGOS			
	4 NGOs			
	5 Individuals			
C. Internal liaison/use				
1 Within agency advice				
2 Within agency links				
D. Knowledge provision				
1 Research				
2 Education				

2.2 *Do you think that there are certain 'key business activities' that are specific to each of the country agencies CCW, SNH, & EN ?...*

3. *USE AND DEMAND FOR HABITAT RELATED DATA*

3.1 Using a code from the 'Key Business Activity' table please fill in the first two columns with KBAs your work is involved with. Then using the last four columns tick the scale where you provide or use information.

Fill in a table for information provision/supply

Fill in the next table for information use

PROVISION/SUPPLY

KEY BUSINESS ACTIVITY OR FUNCTION		**SCALE**			
		Local	**Regional**	**National**	**International**

USE

KEY BUSINESS ACTIVITY OR FUNCTION		**SCALE**			
		Local	**Regional**	**National**	**International**

3.2 Can you provide examples of the kinds of common requests for habitat related data that you have received over the past five years, and specify who requested them...

4. *SOURCES OF HABITAT RELATED DATA*

Below is a provisional list of habitat-related data sources to use as an aid to memory.

4.1 Please add any sources we have missed.

Phase 1 survey	Air Photos	Land use change in England (OS/DOE)	BSBI
Phase 2 survey	Remote sensing data	Monitoring Landscape Change	BRC
NVC surveys	ITE landcover	National Countryside Monitoring Scheme (NCMS)	National Trust
SSSI habitat (site files)	CORINE Biotopes	Countryside Survey	BTO
River corridor surveys	Land Cover Scotland	ITE land classification	NGOs e.g. RSPB, Plantlife, etc.
Environmental Assessments	Soil Survey	Invertebrate Site Register (ISR)	Wildlife Trusts
Environmental Impact Statements	Forestry Commission Census	Countryside Information System (CIS)	Local and county naturalists societies
	Ancient Woodland Inventory	COREDATA	Wildfowl Trust
		Habitat resource surveys	Agricultural Parish Returns
			IACS

*4.2 Please add sources that you USE in your area of work to the **Data Sources Evaluation Table** below.*

We would like you to distinguish between data sources and data types. For example if you refer to a **Phase 1 survey** report (the *source*) for data on habitat extent within a region, then in this case the data *type* is **area.** Therefore data types will include information on: area, volume, number, distribution, range, quality.....

Please fill in each column for each data source and type. If you refer to one data *source* for different distinguishable *types* of data please fill in a new row for each *type*. Please rate the strengths and weaknesses of each data *source* for each data *type*. To help you with your judgement refer to the ***Data Assessment Criteria*** below. After you have done this please rate the value of the data sources/types in the penultimate column using a 5 point scale (1-high, 5-low), and add a KBA code to relate what function you commonly use this data source for.

DATA ASSESSMENT CRITERIA

1	Date (current?)	13	Availability of associated variables or environmental data
2	Original aim and purpose (is this relevant?)	14	Quality
3	Original customer/user	15	History/processing chain
4	Geographical coverage	16	Baseline data
5	Scale	17	Change data
6	Resolution	18	
7	Spatial unit of description	19	
8	Sampling methodology	20	
9	Repeatability	21	
10	Use of meta data	22	
11	Availability of area measurements	23	
12	Availability of mapped information	24	

DATA SOURCES EVALUATION TABLE

SOURCE	*TYPE*	*STRENGTHS*	*WEAKNESSES*	*VALUE*	*SCALE*	*KBA*
e.g. *Phase I*	*Habitat quality*	*4, 9, 12*	*1- not current* *14-not consistent* *insufficient detail on habitat quality*	*3*	*Regiona l*	*A5*

4.2 Although you have listed strengths and weaknesses can you elaborate on the particular merits of the three data source(s) you have rated most highly...

4.3 Similarly, can you list the strengths (if any) and weaknesses of the three data sources you have rated as the poorest...

5. POTENTIAL SOURCES OF HABITAT RELATED DATA

The next question is about why data may not be used.

There are normally four main reasons why data are not used:

a) imperfect knowledge, where potentially interested users are unaware of available data or information

b) data restrictions, where data is restricted for reasons of confidentiality, commercial or military interest. This category can include situations where data may simply be too expensive for some potential users.

c) data quality, where data are not of sufficient quality, or are not testable or representative.

d) data form, where data are not in a suitable form for the required function or use.

5.2 Please list any potential habitat related data sources you are aware of that are not currently used by you or your organisation, please add the reasons why you think they are not utilised...

APPENDIX 3

SNCO DATA SOURCES

Source	Type	Scale	Key Business Activity	Strengths	Weaknesses	Value
SURVEYS						
NVC	Habitat quality	Local	Site selection Audit Influence research Partnership Develop quality standards Advise colleagues Advise government	Sampling methodology Availability of area measurements Baseline data Change data	Date not consistent Does not work well in NW Scotland Now used differently to original purpose Does not have GB geographical coverage Difficult to use at large scale Repeatability unknown Quality surveyor dependant	4
NVC	Habitat quality	Local Regional National	Management of SSSIs SSSI designation NNR review and management Species recovery Natural Areas Advise other agencies	Date Relevance of original purpose Original customer/user Geographical coverage Resolution Spatial unit of description Repeatability Availability of area measurements Availability of mapped information Availability of associated variables or environmental data Quality	Use of meta data unless databased Expensive	1
NVC	Habitat distribution	National	Habitats Directive Biodiversity convention	Date Geographical coverage - best available Sampling methodology Availability of mapped information Quality	Geographical coverage not complete No area data available	2
NVC Surveys	Habitat distribution	Local Regional National GB	State of resource Designated sites Provide contextual information Advise externally Maintain NNR series for research, conservation, demonstration	Valuable information for context and management	Date, many records old Incomplete coverage of country Repeatability unknown Often lacks associated data Quality variable	1
NVC	Habitat Distribution	Local Regional National GB	Designated sites Habitat evaluation-wider countryside Inventory/baselining Monitoring sites	Date, recent system Repeatability Baseline data Change data	Resolution	3
NVC	Habitat extent	Local	Habitats Directive	Baseline data	Geographical coverage, not complete Availability of area measurements	3
NVC	Biodiversity	Local Regional National	Designated site selection Audit Education Advise colleagues Advise government	Depends upon scale of use	Spatial unit of description, as it does not always take into account the area of description	4

Source	Type	Scale	Key Business Activity	Strengths	Weaknesses	Value
NVC	Classification	National	Habitat definitions Relating habitats	Original aim and purpose Geographical coverage Spatial unit of description Repeatability Widely used	Geographical coverage, some areas incomplete	1
NVC/ Phase 2	Community composition	Local Regional National	Designated sites Advise government Advise colleagues Research Education Setting common standards	Relevance of original purpose Geographical coverage Scale Resolution Spatial unit of description Sampling methodology Repeatability Availability of area measurements Availability of mapped information Availability of associated variables or environmental data Quality	Not digitised	1
Phase 2 NVC Survey	Habitat classification	Local	Inventory Developing Information systems	Spatial unit of description Availability of area measurements Availability of mapped information Baseline data	Geographical coverage Change data	2
Phase 2 NVC Survey	Habitat area	Local	Site designation, Natura 2000 Site management SSSIs & SACS	Spatial unit of description Availability of area measurements Availability of mapped information Baseline data	Geographical coverage Change data	2
Phase 2	Habitat assessment	Regional	SSSI resource audit Natural Areas	Repeatability Availability of mapped information	Date Sampling methodology Availability of mapped information Repeatability Quality	2
Phase 2 (Non NVC)	Habitat quality	Local Regional	SSSI designation Natural Areas	Sampling methodology Repeatability Availability of mapped information	Date Original aim and purpose	2
Phase 2	Habitat area, distribution, quality	Wales	SSSI selection pSAC selection Advise colleagues Research Education, lectures, training	Date Original aim and purpose Geographical coverage Scale Resolution Sampling methodology Repeatability Availability of area measurements Availability of mapped information Availability of associated variables or environmental data Quality Baseline data	-	1

Source	Type	Scale	Key Business Activity	Strengths	Weaknesses	Value
Phase 2 Habitat resource	Habitat quality, extent	GB	Developing advocacy Setting conservation priorities Providing contextual information Improving conservation monitoring ? Site selection	Scale Resolution Availability of area measurements	Date, now very old Geographical coverage, only minority give full GB cover	2
Phase 2	Quality	Local	Advice - All categories	-	Date Quality, not collected for this purpose	3
Phase 2	Distribution	Local	Advice - All categories Establish standards - habitat classification Inventory	Cost Effective	-	1
Phase 2	Area	Local	Advice - All categories	-	This type of information is typically limited (in supply?)	4
MNCR Phase 2	Habitat descriptions	GB	Site selection- Statutory & non-statutory Site management Site monitoring W&C Act Quinquennial Review Advise Country Agencies Advice -All categories Establish standards Inventory Setting conservation priorities	Date Spatial unit of description Sampling methodology Use of meta data Availability of associated variables or environmental data Baseline data	Geographical coverage, gaps may limit ability to define/classify habitats	1
River Habitat Survey	Quality	Local Regional National GB	Site management Advise other agencies Setting conservation priorities	Spatial unit of description Repeatability Baseline data Change data	Geographical coverage, at present patchy, will improve as the technique becomes more widely adopted	1
River corridor surveys	Quality	Local Regional	Site management Advise other agencies Setting conservation priorities	Scale Spatial unit of description Repeatability Availability of mapped information	Quality, large inter-surveyor differences	4
5 yearly water quality surveys	Quality	Regional National GB	Site management Site monitoring Advise other agencies Advise externally Setting conservation priorities	Geographical coverage Repeatability Baseline data Change data	Sampling methodology, ignores several important quality variables	4
Phase 1	Habitat extent	Regional	SSSI site survey SSSI resource audit Species recovery, site & species selection Natural Area profiling	Original aim and purpose Resolution Spatial unit of description	Date, not current Sampling methodology, pre-dates standard system Quality, not consistent, too hurried	3
Phase 1	Habitat type	Local Regional	Habitats Directive, SAC selection	Geographical coverage Scale Resolution Availability of area measurements Availability of mapped information	-	2

Source	Type	Scale	Key Business Activity	Strengths	Weaknesses	Value
Phase 1	Habitat area	Local Regional National	Parliamentary questions	Geographical coverage Scale Resolution Repeatability Availability of area measurements Availability of mapped information Quality	Date, some now 9 years old	1
Phase 1	Habitat quality	Local Regional	Advise externally, NGOs Site identification SNCI	Geographical coverage Scale Resolution Repeatability Availability of area measurements Availability of mapped information Quality	Date, some now 9 years old	1
Phase 1	Habitat distribution	Local Regional National	Advise externally, individuals, consultants Casework, road schemes	Geographical coverage Scale Resolution Repeatability Availability of area measurements Availability of mapped information Quality	Date, some now 9 years old	1
Phase 1	Habitat composition	Local Regional National	Designated sites Advise government Advise colleagues Research Education Setting common standards	Original aim and purpose Geographical coverage Scale Resolution Spatial unit of description Repeatability Availability of area measurements Availability of mapped information Availability of associated variables or environmental data Quality History/processing chain	Not available in digitised form	1
Phase 1	Habitat area, distribution and quality	National, Wales	Site selection, SSSI, pSAC	Date Original aim and purpose Geographical coverage Scale Repeatability Availability of area measurements Availability of mapped information Quality	-	1
Phase 1	Habitat location	Local	Evaluating habitat fragmentation	Scale Spatial unit of description, enables examination of range of habitats in an area	Date, often old Geographical coverage, some gaps Sampling methodology, somewhat variable Quality variable	2
Marine Phase 1	Habitat area	Local	Inventory	Availability of mapped information Baseline data	Geographical coverage very limited Scale not consistent Spatial unit of description, inconsistent classification	3

Source	Type	Scale	Key Business Activity	Strengths	Weaknesses	Value
INVENTORIES/RESOURCE SURVEYS						
Habitat resource surveys	Habitat area	GB	Habitats Directive Biodiversity convention Site designation, SACs Advise government Liaise with NGOs & public Provide information NGOs & public Provide information Country Agencies, government	Date Geographical coverage Availability of area measurements Availability of mapped information Quality Accessibility	Accessibility, not always within Country Agencies Geographical coverage, depends upon resource	1
Ancient woodlan d inventory	Habitat quality	National	Inventory Restoration	Original aim and purpose Geographical coverage Scale Repeatability Baseline data	Date History/processing chain	1
Ancient woodlan d inventory	Habitat area	National	Inventory Restoration	Original aim and purpose Geographical coverage Scale Availability of area measurements Availability of mapped information	Date History/processing chain	1
Ancient Woodlan d Inventory	Habitat area, state	Local Regional National	National state reporting Advise externally Advocacy	Date, on-going revision Original aim and purpose, remains relevant Geographical coverage, complete for England Use of meta data, brings together data from photos etc. Availability of area measurements Availability of mapped information Change data	Availability of associated variables or environmental data, only limited data for each site Quality, variable but improving Change data, only at large time intervals Documentation , satisfactory but could be improved	1
Forestry Commiss ion Census	Woodland area & composition	Regional Country GB	National state reporting Advise externally Provide context	Geographical coverage, complete GB cover Sampling methodology, sound methodology Availability of area measurements, County to national areas Baseline data Change data, change can be estimated from previous censuses	Date, Use of meta data years old but re-survey starting Original aim and purpose, composition data do not quite match EN categories but still useful Availability of mapped information, not site based but next one will be	2
Forestry Commiss ion Census	Area, type	National	Advise colleagues Advise externally Research Education	Geographical coverage Availability of area measurements Baseline data	Sampling methodology, sample rather than inventory	3
Habitat inventori es	Site type, distribution	Local Regional	Natural Areas Provide advice externally Planning casework	Date Sampling methodology Repeatability	Original aim and purpose, means limited uses Repeatability, not consistent	2

Source	Type	Scale	Key Business Activity	Strengths	Weaknesses	Value
Habitat inventories	Habitat status	National GB	Biodiversity, information evaluation and investment Biodiversity Action Plan reporting Habitats Directive reporting Assessment of favourable status for BAP and HD	Original aim and purpose Geographical coverage	Spatial unit of description Change data Documentation	2
Limestone Pavement survey	Habitat status	Regional National GB	Biodiversity, information evaluation and investment Biodiversity Action Plan reporting Habitats Directive reporting Assessment of favourable status for BAP and HD	Documented	-	2
Habitat resource surveys	Habitat extent	National	Habitats Directive	-	Geographical coverage Outside agency control	3
Habitat Inventory (peatlands)	Quality, distribution, area	Local Regional National	Advise landowners Site selection Inventory Audit Influence research Partnership Advise colleagues Advise government	Scale, different scales Use of meta data Availability of area measurements Availability of mapped information Quality History/processing chain Baseline data	Date, inconsistent	1
BRC atlases	Distribution	Local National	Biodiversity Action Plans (mostly species)	Geographical coverage Quality Baseline data	Scale Availability of mapped information	2
Red Data Books	Status	Local National	Biodiversity Action Plans (mostly species)	Quality	-	3
BSBI atlases	Distribution	Local National	Biodiversity Action Plans (mostly species)	Geographical coverage Quality Baseline data	Scale Availability of mapped information	2
CORINE Biotopes	Habitat extent	EU	Developing advocacy Setting conservation priorities Site selection, SAC	Geographical coverage		5
CORINE Biotopes	Classification	International	Habitat definition Relating habitat classifications Habitats Directive	Geographical coverage Widely used by others	Spatial unit of description, imbalanced classification No original data for comparison Repeatability, classification constantly changing	5
REMOTE SENSING/AIR PHOTOS/LAND COVER/MAPS						
ITE Land Classification	Distribution	Regional National GB	Developing evaluation techniques	Geographical coverage Scale Spatial unit of description Sampling methodology Repeatability Availability of area measurements	-	1
Remote sensing data	Land cover Area	National (Scotland) Regional	Inventory Habitat restoration Developing information systems	Geographical coverage Availability of area measurements Availability of mapped information	Resolution Sampling methodology Repeatability Quality	3

Source	Type	Scale	Key Business Activity	Strengths	Weaknesses	Value
LCS 88	Landcover type, area	National (Scotland)	Habitat restoration Developing information systems Advise externally Advise internally Research Education	Geographical coverage Availability of area measurements Availability of mapped information Baseline data	Date Original aim and purpose, limits often exceeded Spatial unit of description, mosaics are problematic Repeatability	1
Ordnance Survey data	Spatial orientation	Local Regional National	Inventory Audit Advise colleagues Advise government	Date Original aim and purpose Original customer/user Geographical coverage Scale Resolution Spatial unit of description Sampling methodology Use of meta data Availability of mapped information Availability of associated variables or environmental data Quality Baseline data	Prohibitive cost	1
Old maps	Historical woodland cover	Local Regional	Designated site selection Disseminate knowledge Contextual information Advise externally	Baseline data Change data	Geographical coverage variable Resolution variable, often uncertain Quality variable	2
British Geological Survey	Area Habitat type	Regional	Inventory	Geographical coverage Availability of mapped information	Original aim and purpose often not relevant to biology Scale too broad	4
Air photos	Habitat cover within and outside sites	Local	Site management, SSSIs Site designation, SSSIs	Date Resolution Spatial unit of description Quality Baseline data	Original customer/user Requires interpretation	3
Air photos	Resource audit	Local	Management of designated sites, SSSIs, NNRs Advise externally Planning casework	Date Geographical coverage Repeatability	Availability of area measurements Baseline data Change data	3
SITE FILES, SURVEYS & ASSESSMENTS						
SSSI Habitat surveys	Habitat representation	Local Regional	Management of designated sites Site designation, SSSIs Natural Areas	Date Original aim and purpose Geographical coverage Scale Repeatability Availability of mapped information Quality Quick low-cost access	Date Resolution, often too coarse Quality sometimes poor	1
Second tier site registers	Status of land unit	Local	Planning casework	Date, variable Original aim and purpose Spatial unit of description Quality	Date, variable Sampling methodology, maps & lists collated over a period of time	3
Second tier site inventories	Habitat assessment	-	-	Geographical coverage Spatial unit of description	Date Sampling methodology Availability of area measurements Availability of mapped information Quality Change data	-

Source	Type	Scale	Key Business Activity	Strengths	Weaknesses	Value
Site files	Site-related	Local	Management of SSSIs Designation and management, NNRs Natural Areas Planning casework	-	Date Sampling methodology Repeatability Availability of mapped information Quality	3
Site files	Range	Local Regional National GB	Statutory sites Site monitoring	Geographical coverage History/processing chain Baseline data	Availability of mapped information, absent or insufficient detail	5
Site briefs, citations	Reasons for site selection	Local	Biodiversity Action Plan reporting Habitats Directive reporting	-	Original aim and purpose Quality History/processing chain, don't contain information in a structured way	4
SSSI site files	Habitat characteristics	-	-	Availability of associated variables or environmental data	Date, inconsistent across files Original aim and purpose, often misused Sampling methodology inconsistent Repeatability, often not possible to repeat Use of meta data, not used Availability of mapped information, often no maps Quality, inconsistent Not documented	-
SSSI Schedules	Site quality	Local Regional National	State of habitat reporting Designated site support Advise externally Maintain NNR series for research, conservation, demonstration	Original aim and purpose, relevant data Geographical coverage, complete cover Quality, high	Availability of area measurements Availability of mapped information, not available in HQ (EN)	3
Files	Quality Extent	Local	Habitats Directive products	Only source	Repeatability Quality Accessibility	1
Environmental Assessments, Environmental Impact Statements	Habitat quality	-	Planning casework	Original aim and purpose Geographical coverage Spatial unit of description Quality	Quality	-
Environmental Assessments	Quality	Local	Planning casework	Date Resolution	Geographical coverage Sampling methodology Availability of area measurements Quality	4
DATABASES/INFORMATION SYSTEMS						
Countryside Information System	Woodland area and general composition	Regional National GB	Advise externally Evaluating habitat fragmentation	Date more recent than FC Census Geographical coverage, whole of GB	Sampling methodology Quality, uncertainty on reliability of methods and results	3
Invertebrate Site Register	Quality	Local Regional National GB	SSSI selection SSSI notification Advise other agencies	Geographical coverage	Sampling methodology, variable methodology Geographical coverage, patchy, incomplete	3

Source	Type	Scale	Key Business Activity	Strengths	Weaknesses	Value
Invertebrate Site Register	Species data	Local Regional National	Review protected species schedules Species status listing International conventions and directives Advise on SSSI selection Data collection Data collation & computerisation Setting/maintaining standards Advise government agencies Consultation with DOE External liaison Internal liaison	Original aim and purpose Original customer/user Scale Resolution Spatial unit of description Availability of mapped information Baseline data Good quality data provided free by volunteers	Date, variable Geographical coverage, gappy cover Sampling methodology variable Repeatability variable Availability of associated variables or environmental data, variable Quality variable	2
Rivers Database	Distribution Quality	Local Regional National GB	Statutory sites Site management Site monitoring UK Biodiversity Action Plan Habitat evaluation Inventory Developing evaluation techniques	Geographical coverage Baseline data	Geographical coverage throughout GB, but still patchy	1
Loch Survey Database	Distribution	Local Regional National GB	Statutory sites Site management Site monitoring UK Biodiversity Action Plan	Geographical coverage Sampling methodology Repeatability Availability of mapped information	Availability of associated variables or environmental data, more environmental data are needed	1
Loch Survey Database	Quality	Local Regional National GB	Statutory sites Site management Site monitoring UK Biodiversity Action Plan	Geographical coverage Scale	Availability of associated variables or environmental data, more environmental data are needed	1
IFE RIVPACS Database	Quality	Local Regional National GB	Statutory sites Site monitoring UK Biodiversity Action Plan Habitat evaluation Advise other agencies Advise externally Developing evaluation techniques	Geographical coverage Resolution Sampling methodology Repeatability Availability of associated variables or environmental data	-	2
NVC Survey database	Habitat distribution, quality	Local Regional	SSSI management SSSI designation Natural Areas	Geographical coverage Spatial unit of description Sampling methodology Repeatability Availability of area measurements Availability of mapped information Availability of associated variables or environmental data Quality	Date Change data	2
COREDATA	Woodland area in SSSIs	Regional National	State of habitat reporting Designated site support Dissemination of ideas and knowledge	Geographical coverage, most wooded sites covered Availability of area measurements	Availability of mapped information, not available in HQ (EN) Quality uncertain in places	3

Source	Type	Scale	Key Business Activity	Strengths	Weaknesses	Value
Features Database	Habitat area Conservation status	-	Advise government	Date Original aim and purpose Geographical coverage Availability of area measurements	-	1
Databases for primary data sets	Status	National GB	Biodiversity, information evaluation and investment Biodiversity Action Plan reporting Habitats Directive reporting Assessment of favourable status for BAP and HD	Original aim and purpose Resolution	Spatial unit of description Change data Documented Expertise required to interpret	2
JNCC database & GIS for statutory sites	Status and conservation action	National	Biodiversity Action Plan reporting Habitats Directive reporting Setting conservation priorities	Original customer/user Geographical coverage	Availability of mapped information Quality History/processing chain Change data	2
JNCC sites repository system	Status and conservation action	GB UK	Biodiversity Action Plan reporting Habitats Directive reporting	Selective	Not yet possible to make linkage between habitat distribution and sites	-
JNCC sites repository system	Working versions of classifications	-	Background information for most functions	-	-	3
MNCR database	Habitat quality	GB	Site selection- Statutory & non-statutory Site management Site monitoring W&C Act Quinquennial Review Advise Country Agencies Advice -All categories Establish standards Inventory Setting conservation priorities	Date Spatial unit of description Sampling methodology Use of meta data Availability of associated variables or environmental data Quality Baseline data	Geographical coverage, sometimes gaps make GB perspective limited	1
MNCR database	Habitat/species distribution	GB	Site selection- Statutory & non-statutory Site management Site monitoring W&C Act Quinquennial Review Advise Country Agencies Advice -All categories Establish standards Inventory Setting conservation priorities	Date Sampling methodology Use of meta data	Geographical coverage, some significant gaps in GB coverage	1

Source	Type	Scale	Key Business Activity	Strengths	Weaknesses	Value
MNCR database	Area (extent)	GB	Site selection- Statutory & non-statutory Site management Site monitoring W&C Act Quinquennial Review Advise Country Agencies Advice -All categories Establish standards Inventory Setting conservation priorities	Date Sampling methodology Use of meta data	Availability of area measurements Availability of mapped information, generally poor/not prime aim of survey	3
EC Habitats Directive databases	Legal position GB context	Internatio nal	Biodiversity Action Plan reporting Habitats Directive reporting Assess conservation status	Yet to be developed, but will become important	History/processing chain, processing chain not evident	4
WCMC & related internatio nal datasets	International context	Internatio nal	Biodiversity Action Plan reporting Habitats Directive reporting Assess conservation status	-	-	5
LITERATURE						
Publishe d literature	Distribution	Local Regional National GB	Habitat evaluation Inventory External education	Availability of associated variables or environmental data	Sampling methodology, patchy, incomplete Geographical coverage patchy	5
Literatur e	Quality	Local	Management of SSSIs Designation and management, NNRs Species recovery Advise internally Advise externally	Geographical coverage Quality	Date Availability of area measurements Availability of mapped information Change data	4
Literatur e	Site descriptions and processes	Local	Designated site selection and defence Disseminate ideas and knowledge Advise externally Maintain NNR series for research, conservation, demonstration	Quality, often very high level of detail	Date, often old Geographical coverage, very limited	2
Literatur e	Master versions of classifications	?	Background information for most functions	-	Quality control	4
Taxonom ic literature	Distribution	GB	Advise internally Advise externally Establish standards Inventory	Geographical coverage Use of meta data Availability of associated variables or environmental data	Date, often outdated unless recently published Poor for most taxa	3
Lead Networks literature	Issues affecting conservation status	GB	Biodiversity Action Plan reporting Habitats Directive reporting Assess conservation status	-	-	2
Publishe d internatio nal sources	Status beyond GB Context	Internatio nal	Biodiversity Action Plan reporting Habitats Directive reporting Assess conservation status	-	-	4

Source	Type	Scale	Key Business Activity	Strengths	Weaknesses	Value
EC literature on directives	Classifications Reporting requirements	International	Background information for most functions	-	Quality, poor version control Poor relationship to GB	-
Literature & Universities	Species data	Local Regional National	Review protected species schedules Species status listing International conventions and directives Advise on SSSI selection Advise government agencies Consultation with DOE Species translocation policy	All criteria - depends upon query	All criteria - depends upon query	1-3
ORGANISATIONS						
Wildfowl Trust	Site quality	Local	Management of SSSIs Designation of SSSIs	Date Repeatability Quality	Original customer/user	3
BRC	Species data	National	Review protected species schedules Species status listing International conventions and directives Data collection Data collation & computerisation Setting/maintaining standards Advise government agencies Consultation with DOE Species recovery Time series monitoring External liaison Internal liaison	Geographical coverage (recent data) Scale (recent data) Resolution (recent data) Spatial unit of description (recent data) Use of meta data Quality History/processing chain Baseline data Change data Good quality data provided free by volunteers	Date, variable Original aim and purpose (early data) Original customer/user (early data) Geographical coverage (early data) Scale (early data) Resolution (early data) Spatial unit of description (early data)	2
Entomological societies	Species data	National	Review protected species schedules Species status listing International conventions and directives Data collection Data collation & computerisation Setting/maintaining standards Advise government agencies Consultation with DOE Species recovery Time series monitoring External liaison Internal liaison	Geographical coverage (recent data) Scale (recent data) Resolution (recent data) Spatial unit of description (recent data) Use of meta data Quality History/processing chain Baseline data Change data Good quality data provided free by volunteers	Date, variable Original aim and purpose (early data) Original customer/user (early data) Geographical coverage (early data) Scale (early data) Resolution (early data) Spatial unit of description (early data)	2

Source	Type	Scale	Key Business Activity	Strengths	Weaknesses	Value
National Trust	Species data	Local National	Review protected species schedules Species status listing International conventions and directives Data collection Species recovery Time series monitoring External liaison Internal liaison	Date Original aim and purpose Original customer/user Scale Resolution Spatial unit of description Sampling methodology Repeatability Availability of area measurements Availability of mapped information Availability of associated variables or environmental data Quality History/processing chain Baseline data Change data Good quality data provided free by volunteers	Geographical coverage, only on NT sites	3
Universities Institutes Literature	Habitat description	Local	Advice - All categories Establish standards - habitat classification Inventory	Spatial unit of description Availability of associated variables or environmental data (but varies) Baseline data Cost-effective	Date, may be old Geographical coverage, often has different aims to us	2
NGOs	Quality Extent	Local National	Habitats Directive, Annex 1 development	-	Geographical coverage Outside agency control	4
NGOs	Habitat representation	Local	Management of SSSIs Designation of SSSIs Review & management of NNRs Species recovery Natural Areas	-	Quality, subjective in some areas	3
NGOs	Habitat quality	National	Habitats Directive Biodiversity Convention	Date Availability of area measurements Availability of mapped information	Geographical coverage not always complete	3
Local Record Centres	Site quality Species	Local	Management of SSSIs Designation of SSSIs Review & management of NNRs Species recovery Natural Areas	-	Date Spatial unit of description Sampling methodology Availability of associated variables or environmental data Quality, in coverage and consistency	3
Local Record Centres	Databases	Local	Natural Areas Planning casework	Date Geographical coverage Spatial unit of description	Resolution Sampling methodology Availability of area measurements Availability of mapped information Quality Baseline data Change data	3

Source	Type	Scale	Key Business Activity	Strengths	Weaknesses	Value
Local Record Centres	Species data	Local Regional	Review protected species schedules Species status listing International conventions and directives Advise on SSSI selection Data collection Data collation & computerisation Setting/maintaining standards Advise government agencies Consultation with DOE External liaison Internal liaison	Original aim and purpose Original customer/user Good quality data provided free by volunteers	Date, variable Geographical coverage, only a few LRC's used Availability of associated variables or environmental data, little data available Quality, variable	3 (when used)
BSBI	Species data	National	Review protected species schedules Species status listing International conventions and directives Data collection Species recovery Time series monitoring External liaison Internal liaison	Geographical coverage Resolution Repeatability Availability of mapped information Quality Change data Good quality data provided free by volunteers	Date, variable Availability of associated variables or environmental data, variable	2
BSBI	Plant species recording	Local Regional National	Designated sites Advise government Advise colleagues Research Education Setting common standards	Geographical coverage Availability of mapped information Availability of associated variables or environmental data	Repeatability not planned in all cases	1
LPBR (JNCC)	Plant species recording	Local Regional National	Review protected species schedules Species status listing International conventions and directives Advise on SSSI selection Data collection Data collation & computerisation Setting/maintaining standards Advise government agencies Consultation with DOE External liaison Internal liaison	Geographical coverage Resolution Repeatability Availability of mapped information Quality Change data Good quality data provided free by volunteers	Date, variable Availability of associated variables or environmental data, variable	2
EU/International e.g. EUCC	Habitat extent	EU	Developing advocacy Setting conservation priorities Providing contextual information Improving conservation monitoring ? Site selection	Geographical coverage, at best	Very difficult....all areas.....????	5 (some 3)
International (various)	Description	International	Inventory	-	Geographical coverage very poor Quality Habitat coverage limited	4

Source	Type	Scale	Key Business Activity	Strengths	Weaknesses	Value
British Trust for Ornithology	Species numbers/distribution	National	Advise government	Geographical coverage	Resolution, satisfactory for general data	3
NRA	Species data	Regional National	Review protected species schedules Species status listing International conventions and directives Advise on SSSI selection Data collection Data collation & computerisation Advise government agencies Consultation with DOE Species translocation policy Species recovery Time series monitoring	Geographical coverage Scale Sampling methodology Repeatability Availability of associated variables or environmental data Baseline data Change data	Date, variable Original aim and purpose Original customer/user	4 1 (NRA crayfish database)

APPENDIX 4

NORTH YORK MOORS NATIONAL PARK PLAN OBJECTIVES

1. MOORLAND

AIMS	*POLICIES*	*INDICATOR*	*DATA NEEDS*	*DATA SOURCES*
Protection and management of moorland	M1 Encourage co-operation between moorland owners/occupiers to develop comprehensive/positive moorland management	Level and effectiveness of co-operation	Perceptions of NYM staff and of other public agencies on levels/results of co-operation	Questionnaire survey
	M2 Resist development or conversion of Section 3 (W&C Act) land	Proportion of Section 3 heather moorland affected by development or conversion.	Moorland extent and quality (area Section 3). Area affected by development or conversion	*Section 3 map; air photos when on GIS*
	M3 Resist development on other open moorland	Proportion of other open areas affected by development	Moorland extent and quality. Area affected by development	*Section 3 map; air photos when on GIS*
	M4 In event of conversion take mitigation measures	Proportion of moorland affected by conversion or change, and proportion covered by satisfactory mitigation measures	Moorland extent and quality. Area affected by conversion and type of change	*Section 3 map; air photos when on GIS*
	M5 Take steps to prevent moorland fires, co-ordinate training of vol. fire-fighters	*% of moorland covered by agreed fire plans (by year)*		
	M6 Encourage moorland owners to continue long-term heather management	Area and % of heather cover in moorland areas under effective management	Moorland extent and quality. Area of heather cover. Area of heather under management.	*Heather condition survey*
	M7 Disseminate information on moorland management/ecology	Level of awareness by moorland owners/occupiers	Perceptions of moorland owners/occupiers	Questionnaire survey *No. of publications*
Bracken encroachment and control	M8 Eradicate/reduce bracken to 10% of the moorland area by 2000	Proportion of moorland under bracken	Moorland extent; bracken extent	Remote sensing; field-work *Air photos -doesn't include effectiveness*
	M9 Create actively managed buffer zones around cleared areas	Proportion of cleared areas with actively managed buffer zones - *linked to overall control*	Extent of cleared bracken (by year); proportion managed as buffer zone	Remote sensing; field-work; management records
Moorland fencing	M10 Resist permanent new fencing (except category 1 roads)	Length of permanent new fencing	Length and location of permanent new fencing	Highways Department; MAFF; fieldwork
	M11 Consider requests for cattle grids; provide grants if appropriate *now changed*			
Moorland research	M12 Continue moorland research programme	Amount of funding to moorland research programme	Expenditure on research programme	Financial records
	M13 Develop and use remote sensing and air-photo techniques to monitor land use management and change	Proportion of moorland effectively monitored *Don't actively participate in*	Area of moorland; extent and accuracy of monitoring	In-house records; field validation of survey data
	M14 Continue co-operation with other agencies on moorland ecology research programme	*Produce research register*		

2. FARMING

AIMS	*POLICIES*	*INDICATOR*	*DATA NEEDS*	*DATA SOURCES*
Environmentally sensitive farming	F1 Support conservation oriented farming through grant schemes and management objectives	Area or proportion of farmland covered by a grant scheme	Area of farmland. Area of farmland covered by separate schemes	*Already do for Functional Strategy*
Farm diversification	F2 Support farm diversification where compatible with NP objectives	*No longer a relevant issue*		
	F3 Work jointly on specific projects to promote greater awareness of diversification projects which support NP objectives	*No longer a relevant issue*		
Alternative land use *(Creative side)*	F4 Investigate the feasibility of alternative land management techniques through the Alternative Land Use Project	*No. of new shames a year* *Specifically don't want to use expenditure.*		
Impact of farming on the landscape	F5 Resist changes development and land use changes arising from new small farms where they affect landscape quality	*No. of new farms visited*		
	F6 Monitor the condition of traditional field boundaries and take action to maintain and improve through grants and advice	*Hedges are in Farm Schemes, walls are not*		*Hedge survey done in 1991. Also in farm scheme surveys*
	F7 Give guidance on new farm building and road design to minimise landscape impacts			

APPENDIX 5

NORTH YORK MOORS NATIONAL PARK – OUTPUTS FROM AIMS AND OBJECTIVES WORKSHOPS

Aims and objectives produced during the objective setting workshops.

These aims and objectives were later amended for the indicator development workshops, but are included here to illustrate the development process.

RIVERS AND FRESH WATERS

Preface

The following aims and objectives were developed during discussions in February. Discussion participants are invited to suggest amendments to these statements to correct any mis-recording by the Consultants and to help to progress the aims and objectives development process. The following statements have been developed for the purposes of the PIMS project and are illustrative. They are not intended to be comprehensive or definitive and Park staff will require to further develop aims and objectives for these topics as part of the Park Plan preparation process. The aims and objectives are not in any order of priority.

[Comments in italics are notes by the Consultants intended to provoke further thought or discussion by relevant staff. They should be considered and acted on or dismissed as appropriate!]

Aims

i. to conserve and enhance the quality and diversity of fresh water, wetland and associated wildlife.

ii. to protect and maintain the supply and quality of surface and ground waters.

iii. to maintain and enhance the essential characteristics of the river corridor landscapes within the Park.

[These characteristics will need to be defined by the staff concerned. What does this aim imply about the other (non-essential) characteristics?]

iv. to maintain and promote opportunities for recreation and access associated with rivers and other waters, provided that these do not conflict with conservation objectives.

[There may be a need to develop an 'economic' aim related to rivers and fresh waters]

Objectives

a. to maintain and, where appropriate, improve the quality of surface and ground water resources within river catchments in the Park, with the intention that these meet the chemical and biological quality targets set out within the NRA's Catchment Management Plans for the Rivers Esk and Derwent.

[The phrase 'with the intention that' may cause problems for monitoring and measuring this objective at later dates. Can it be avoided? This objective also implies that criteria will be established at a later date, to define where improvements might be appropriate.]

b. to protect, re-establish and, where appropriate, extend woodland and other semi-natural habitats associated with river valleys, in recognition of the contributions of these features to the ecological diversity and characteristic landscapes of the Park.

[Park staff should define which river valleys are of priority within the Plan period, in respect of this objective].

c. to maintain, enhance and, insofar as practical, extend wetland habitats in the Park, in recognition of the particular values of wetland habitats and species.

[Supporting text in the Park Plan may refer to opportunities to redress the loss of wetland habitats and increase biodiversity]

d. to identify, extend and improve opportunities for access and informal recreation associated with river valleys and inland waters, insofar as these do not conflict with conservation objectives.

[Park staff should define which river valleys and inland waters are of priority within the Plan period, in respect of this objective].

e. to conserve and enhance native fish stocks within the rivers of the Park and, in particular, the salmon fisheries of the River Esk and its tributaries, in recognition of their conservation, recreation and economic values.

MOORLANDS

Preface

The following aims and objectives were developed during discussions in February. Discussion participants are invited to suggest amendments to these statements to correct any mis-recording by the Consultants and to help to progress the aims and objectives development process. The following statements have been developed for the purposes of the PIMS project and are illustrative. They are not intended to be comprehensive or definitive and Park staff will require to further develop aims and objectives for this topic as part of the Park Plan preparation process. The aims and objectives are not in any order of priority.

Aims

i. to conserve and enhance the characteristic landscapes of the Park's moorlands[1] .

ii. to protect and, where possible, extend and enhance the biodiversity of moorland habitats.

iii. to promote the restoration and/or regeneration of moorland habitats and landscapes.

iv. to protect and enhance the cultural heritage associated with the moorlands.

v. to maintain, extend and enhance opportunities for moorland access and informal recreation.

vi. to maintain and promote economic activities which are traditionally associated with the moorlands and/or contribute to the essential characteristics of the moorlands.

Objectives

a. to promote the adoption of moorland management regimes which maintain and enhance the special landscape characteristics and biodiversity of the heather moorlands.

[Will it be possible to define indicators which allow this objective - i.e. the 'promotion' of moorland management practices - to be monitored and measured?]

b. to encourage management practices which extend and restore the extent of, and linkages between, semi-natural moorland habitats and identify opportunities to encourage natural succession in specific locations.

[As above: can we measure 'encouragement' and 'identification of opportunities']

c. to safeguard, manage and enhance moor-edge woodlands in recognition of their valuable landscape and nature conservation contributions.

d. to ensure that development and other activities do not diminish the characteristic open and wild landscapes of the moors.

e. to safeguard the traditional moorland economies based on grouse and sheep.

f. to conserve and promote the appropriate management of archaeological and industrial archaeological features associated with the moorlands.

---

1 The term 'moorland' includes moorland fringe areas and associated areas.

g. to protect the cultural heritage and promote awareness, interest and understanding in the living culture and traditions associated with the moorlands.

h. to safeguard and enhance the availability and positive management of, and provision of information on, opportunities for access to moorlands for walking, cycling and riding and other recreational activities, which do not conflict with the special characteristics and values of the moorlands.

[This is a very lengthy and cumbersome objective. Can it be better expressed - e.g. as two separate statements?]

[Many of these objectives will require that Park staff define the key areas, habitats, sites or features of interest, in order to target action effectively

VISITOR[1] ATTRACTIONS, FACILITIES AND SERVICES

Preface

The following aims and objectives were developed during discussions in February. Discussion participants are invited to suggest amendments to these statements to correct any mis-recording by the Consultants and to help to progress the aims and objectives development process. The following statements have been developed for the purposes of the PIMS project and are illustrative. They are not intended to be comprehensive or definitive and Park staff will require to further develop aims and objectives for these topic areas as part of the Park Plan preparation process. The aims and objectives are not in any order of priority.

Aim

To maintain and enhance the range, quality and availability of attractions, facilities and services, which contribute to Park users' enjoyment and are appropriate to the special characteristics of the Park.

[There will be a need within the Park Plan to identify what the special characteristics of the Park are - if that or a similar term is used.]

Objectives

a. to encourage mutual understanding and respect between Park users and those who live and work in the Park and promote a friendly and genuine welcome to visitors.

b. to maintain and encourage the provision, and promote enhancement, of the range, quality and availability of accommodation and related commercial services, for the benefit of Park visitors.

c. to maintain and promote the enhancement of the range, quality and availability of attractions which reflect the special characteristics of the Park.

d. to encourage the development and promotion of events which relate to the special characteristics and economic and community activities of the Park for the mutual benefits of Park visitors and those who live and work in the Park.

e. to improve the integration, delivery, range and quality of Park user information and related services, including the 'book-ability' of attractions, facilities and services.

[The majority of these are qualitative objectives. This may be appropriate in this area, but those concerned will need to find appropriate indicators].

---

1 For the purposes of this topic, 'visitor' includes tourists, day visitors and those who live and work in the Park and use many of its attractions, facilities and services for leisure and recreation in a similar way to visitors. The terms 'visitor' and 'user' are used inter-changeably.

APPENDIX 6

NORTH YORK MOORS NATIONAL PARK – OUTPUTS FROM INDICATORS AND DATA EVALUATION WORKSHOPS

DATA SOURCES EVALUATION

This Annex represents results from the North York Moors indicator development workshops and provided the basis for the data sources evaluation exercise. Indicators were produced for each objective resulting from the objective setting exercise. The objectives shown here were developed from the aims and objectives shown in Appendix 5. At this stage North York Moors staff decided to drop topic based aims. It was decided that, in the future, Park aims would be produced at a broad level for the whole Park area. These would then be used as a basis for the development of specific objectives within topic areas.

The 'Indicator' column largely represents the results of the indicator development workshop, in some cases additional indicators were added by the consultants at a later date. At this point the consultants also added the 'outcomes' column. These outcomes are interpreted from each objective and are intended to help clarify what needs to be measured for that objective. In most cases an indicator measures one or more of these outcomes (but not necessarily all). This additional step helps show which component of the objective that the indicator is intended to measure.

For each indicator a summary of information needs required by that indicator is listed. The data sources column indicates how these information needs may be met. For the data sources evaluation exercise data holders were asked to evaluate the suitability of each data source for each indicator.

MOORLANDS

a. to sustain and enhance the open and wild moorland landscapes of the Park.

OUTCOME(S)	*INDICATOR*	*INFORMATION NEEDS*	*DATA SOURCE(S)*
No adverse change in, or an expansion of, the extent of open and wild moorland landscapes (e.g. areas unaffected by intrusion by fences, conifers, pylons, development. etc.)	Extent of open and wild moorland	Definition and identification of open moorlands Definition and identification of 'detractor elements' - e.g. Area of moorland > 1km from major developments (e.g. pylons, roads) and > 500m from minor developments (fences, buildings)	OS data Section 3 Map Air-photo interpretation
No adverse change in, or an improvement in, the quality of open and wild moorland landscapes	Quality of open and wild moorland	Definition of quality criteria for open moorlands	Air-photo interpretation Field surveys

b. to safeguard and restore the extent and biodiversity of semi-natural moorland habitats and encourage natural succession in specified locations.

OUTCOME(S)	*INDICATOR*	*INFORMATION NEEDS*	*DATA SOURCE(S)*
Maintenance of extent of semi-natural moorland habitats	Area of semi-natural moorland habitat	Definition of 'semi-natural moorland' Area of semi-natural moorland habitat	Section 3 Map EN NVC survey
Maintenance of the quality (biodiversity) of semi-natural moorland habitats	Proportion of different NVC vegetation types within semi-natural moorland habitat	above + NVC communities and their extent Defined locations or geographic areas	EN NVC survey
Restore the extent of semi-natural moorland habitats	Area of semi-natural moorland restored	location and number of suitable restoration area sites Area of restored sites	Air photo interpretation Field logs Management records
Increase in (suitable) moorland area subject to natural succession	Proportion of suitable moorland area undergoing natural succession	Definition and identification of 'suitable' moorland areas Area undergoing natural succession	EN NVC survey Air-photo interpretation

c. to safeguard, extend and enhance moor-edge woodlands, in recognition of their valuable landscape and nature conservation contributions.

OUTCOME(S)	*INDICATOR*	*INFORMATION NEEDS*	*DATA SOURCE(S)*
Maintenance of, or increase in, extent of moor-edge woodlands	Total area of semi-natural woodland habitat	Definition of 'semi-natural woodland' Area of semi-natural moorland habitat	Air photos EN NVC survey ?
Maintenance of, or increase in, landscape quality of moor-edge woodlands	Proportion of semi-natural woodland in management schemes	as above + area of semi-natural woodland in management schemes	NYM/FA records
Maintenance of, or increase in, nature conservation value of moor-edge woodlands	as above + Habitat diversity of woodlands	as above + Measure of desired habitat diversity Measure of habitat diversity	Field surveys Air photo interpretation

d. to protect and enhance the archaeological and industrial archaeological features associated with the moorlands.

OUTCOME(S)	*INDICATOR*	*INFORMATION NEEDS*	*DATA SOURCE(S)*
Maintenance (no deterioration), or repair/restoration of, industrial and archaeological features	Number and condition of features % of sites in need of restoration/ repair % of sites restored/repaired	Definition of 'industrial and archaeological features' Number and location of features Condition of features Number of sites needing restoration/ repair Number of sites restored/repaired	NYM data ? Archaeological site database Field logs Management records Site audits NB - requires liaison with archaeology section

e. ***to protect and promote awareness, interest and understanding of the cultural heritage, living culture and traditional management activities of the moorlands.***

OUTCOME(S)	*INDICATOR*	*INFORMATION NEEDS*	*DATA SOURCE(S)*
Increase in awareness of cultural heritage and traditions	Level of documentation [what does this mean? recording of knowledge?] Level of awareness of cultural heritage and traditions amongst Park users	Definition of 'cultural heritage and traditions' Level of public awareness	NB - requires liaison with education and interpretation section Visitor surveys
Increase in understanding of cultural heritage and traditions	Level of understanding of cultural heritage and traditions amongst Park users	Level of public understanding	as above

f. ***to safeguard and enhance opportunities for access to the moorlands for walking, cycling, riding and other recreational activities, insofar as these do not conflict with the special characteristics of the moorlands.***

OUTCOME(S)	*INDICATOR*	*INFORMATION NEEDS*	*DATA SOURCE(S)*
Maintenance of opportunities for open access	Extent of open access in moorland areas	Definition of moorland area for access purposes Definition of 'open access' Area of moorland under open access agreements	Section 3 map Standard CoCo criteria NYMNP management agreements CoCo Local authorities
Maintenance of opportunities for linear access	Extent of linear access in moorland areas Quality of rights-of-way	Definition of moorland area for access purposes Definitions of 'linear' access Length of rights of way in moorlands Quality of rights of way	Section 3 map Standard CoCo criteria Local authorities NYMNP records Field surveys
Development of opportunities for access	Area under new access agreements Length of new rights of way in moorlands	as above	NYMNP management records ?MAFF Local authorities

VISITOR[2] ATTRACTIONS, FACILITIES AND SERVICES

a. to encourage mutual understanding and respect between Park visitors and those who live and work in the Park and promote a friendly and genuine welcome to visitors.

OUTCOME(S)	*INDICATOR*	*INFORMATION NEEDS*	*DATA SOURCE(S)*
Development of mutual understanding and respect	Level of visitors/users satisfaction	Definition of visitor/user Visitor/user satisfaction Number of repeat visits Defined locations or geographic areas	Visitor surveys
Provision of a friendly and genuine welcome	Perceived level of welcome	Visitor satisfaction	as above

b. to maintain and promote the enhancement of the range, quality and availability of accommodation and related commercial services for the benefit Park users.

OUTCOME(S)	*INDICATOR*	*INFORMATION NEEDS*	*DATA SOURCE(S)*
Maintenance of, or increase in, the range of accommodation and related commercial services	Number of accommodation units available (by type) Quality of accommodation available Availability of accommodation	Number of accommodation types available in agreed classes Definition of quality classes Number of facilities in each quality class Definition of availability of facilities (e.g. bookability) Average availability of facilities (by type) based on this definition	National Accommodation Database All Parks Visitor Survey NYMNP/Tourist Board Visitor surveys Surveys of facilities as above

2 'Park users' include tourists, day visitors and those who live and work in the Park and use its attractions, facilities and services for leisure and recreation.

c. ***to maintain and promote the enhancement of the range, quality and availability of attractions which reflect the special characteristics of the Park.***

OUTCOME(S)	*INDICATOR*	*INFORMATION NEEDS*	*DATA SOURCE(S)*
Maintenance of, or increase in, the range of attractions	Number of attractions available by type	Range of attractions available in agreed type classes	National Attractions Database All Parks Visitor Survey
Maintenance of, or increase in, the quality of attractions	Quality of attractions available	Definition of quality classes (quality is linked to expectation) Number of facilities in each class	 Visitor surveys
Maintenance of, or increase in, the availability of attractions	Geographic distribution by type 'Bookability'	Location of attractions, by type Ease of booking	National Attractions Database Bookability surveys

d. ***to sustain and encourage the development and promotion of events, which relate to the special characteristics and economic and community activities of the Park for the mutual benefits of Park visitors and those who live and work in the Park.***

OUTCOME(S)	*INDICATOR*	*INFORMATION NEEDS*	*DATA SOURCE(S)*
Maintenance of, or increase in, the range of special events	Number of events by type	Definition of 'special events' Number of events based on this definition	Events lists and publications
Maintenance of, or increase in, the number of people attending special events	Number of people at events by type	Numbers of people at events	Surveys of attendees
Maintenance of, or increase in, the level of satisfaction with, and understanding derived from, special events	Level of satisfaction with events by type Level of understanding gained about the Park	Definition of satisfaction Levels of satisfaction of attendees Definition of understanding Levels of understanding gained by attendees	Surveys of attendees Surveys of attendees

e. ***to improve the integration, delivery, range and quality of Park user information and related services, including the 'book-ability' of attractions, facilities and services.***

OUTCOME(S)	*INDICATOR*	*INFORMATION NEEDS*	*DATA SOURCE(S)*
Improvement in delivery of user information		Definition and identification of information sources	
	Levels of dissemination of information	Numbers of information 'packs' issued/sold	NYMNP/Tourist Board records
	Visitor satisfaction with information and information services	Visitor satisfaction with information provision	Twice yearly TIC assessment (National?) Attractions survey
Improvement in range of user information	Extent of information provided about the Park, by type and source	Definition and identification of information categories	Surveys of documentation and information sources
Improvement in quality of user information		Quality criteria for user information	as above Visitor surveys
Improvement in the 'bookability' of attractions, facilities and services	Ease of booking of attractions, services and facilities	Definition of ease of booking criteria Average bookability of attractions based on this definition	Surveys of facilities

RIVERS AND FRESH WATERS

a. ***to maintain and, where possible, improve the quality of surface and ground water resources within the Park, with the intention that these meet the chemical and biological quality targets set by the Environment Agency.***

OUTCOME(S)	*INDICATOR*	*INFORMATION NEEDS*	*DATA SOURCE(S)*
Maintenance of, or improvement in, surface water quality (where possible)	% of total length of river system meeting each target quality class 1) chemical 2) biological	Definition of river system Length of total river system within a pre-determined area Target quality classes (RQOs) Chemical and biological river quality scores	OS data Environment Agency Environment Agency, water quality data
Maintenance of, or improvement in, groundwater quality (where possible)	Levels of nitrate, phosphate and pesticides in groundwaters	Nitrate, phosphate and pesticide concentrations in borehole waters	Environment Agency

b. to protect and maintain the supply of surface and ground waters within the Park.

OUTCOME(S)	*INDICATOR*	*INFORMATION NEEDS*	*DATA SOURCE(S)*
Maintenance of minimum stream/river flow	% of sample points below 'minimum acceptable flow' Change in minimum (summer) baseflow	Number and location of sample flow points. Definition of 'minimum acceptable flow' Seasonal flow records	Environment Agency NYMNP Institute of Hydrology Environment Agency Institute of Hydrology

c. to maintain and enhance the characteristic river corridor landscapes within the Park.

OUTCOME(S)	*INDICATOR*	*INFORMATION NEEDS*	*DATA SOURCE(S)*
Maintenance of, or improvement in, landscape quality	Change in quality of 'key landscape character elements'	Identification of 'key landscape character elements' Landscape quality classification	New landscape assessment MLC data Landscape Character Assessments

d. to conserve and enhance the wildlife values and diversity of fresh water, wetland and associated habitats.

OUTCOME(S)	*INDICATOR*	*INFORMATION NEEDS*	*DATA SOURCE(S)*
Maintenance of or improvement in habitat quality	Native crayfish numbers in limestone streams	Location of suitable habitat Crayfish numbers	OS and water quality data. Stream survey En/NYM survey ?
Maintenance of or improvement in habitat diversity	(Change in?) extent of habitat in different NVC classes	Definition of freshwater, wetland and associated habitats Extent of freshwater, wetland and associated habitats Extent and distribution of vegetation in NVC classes of above	Phase 2 survey New NVC mapping survey in sample areas Biological records

e. to conserve and enhance native fish stocks within the rivers of the Park, insofar as this objective does not conflict with other conservation objectives.

OUTCOME(S)	*INDICATOR*	*INFORMATION NEEDS*	*DATA SOURCE(S)*
Maintenance of, or increase in, native fish populations	Condition of stock	Constraints criteria and map to indicate areas where stock enhancement and expansion is acceptable Numbers and population profiles of native fish species	Environment Agency Angling groups
Geographic expansion of native fish populations	Condition and distribution of stock	as above	as above

f. to sustain and improve opportunities for access, sport and informal recreation associated with river valleys and inland waters, insofar as these do not conflict with conservation objectives.

The group felt unable to tackle this objective without consultation with recreation specialists. This issue highlights the need for an integrative approach to objective setting and indicator development.

APPENDIX 7

CIS GEOGRAPHIC DATA DEFINITIONS

Source: CIS 1996. Countryside Information System V5.40 – Supplementary data details. Computer software for MS Windows. NERC.

WOODLAND	
OS95GREF.ccf Geographic Reference Data.	Friday April 28 13:16:36 95
Woodland : hectares or %, per sq. km: OS95	
Description of data set Geographic reference data set created from the Ordnance Survey's 1995 digital 1:250,000 Strategi data set by classifying the data each one kilometre square into 13 categories (one for each data set) as a percentage, to a resolution of 0.01. Each kilometre square totals 100% with any residual areas being added to the Open Countryside category. The thirteen feature categories were extracted from vector data, each category allocated a colour and for linear features a width value. Each layer was then rasterised and the total number of each colour pixel outputted as a percentage of each kilometre square.	
Accuracy information A nominal resolution of 10m was used throughout (i.e. the smallest measurable unit will be a 10m square and the widths of linear features are in multiples of 10metres). Although data set covers GB approximately 200 coastal squares identified by CIS as "land" are not recognised as such by this data set.	

BUILT-UP AREAS	
OS95GREF.ccf Geographic Reference Data.	Friday April 28 13:16:36 95
Built up - Towns : hectares or %, per sq. km: OS95	
Description of data set Geographic reference data set created from the Ordnance Survey's 1995 digital 1:250,000 Strategi data set by classifying the data each one kilometre square into 13 categories (one for each data set) as a percentage, to a resolution of 0.01. Each kilometre square totals 100% with any residual areas being added to the Open Countryside category. The thirteen feature categories were extracted from vector data, each category allocated a colour and for linear features a width value. Each layer was then rasterised and the total number of each colour pixel outputted as a percentage of each kilometre square.	
Accuracy information A nominal resolution of 10m was used throughout (i.e. the smallest measurable unit will be a 10m square and the widths of linear features are in multiples of 10metres). Although data set covers GB approximately 200 coastal squares identified by CIS as "land" are not recognised as such by this data set.	

MEAN ALTITUDE	
OS95TOPO.ccf Altitude Data.	Monday November 6 13:56:33 95
Mean altitude : m per sq. km: OS95	
Description of data set Topographic data created from the Ordnance Survey's 1995 digital 1:50,000 Panorama data set, in turn derived from the Digital Terrain Model data consisting of height values at each intersection of a 50m horizontal grid, the values of which have been mathematically interpolated from contours on the 1:50,000 Landranger maps. Height values are rounded to the nearest metre with coastline - MHW being given a value of 1m. For each kilometre square the mean altitude is determined from the average of 400 (20 x 20 matrix) interpolated values; the 10%iles and 90%iles being calculated from the same 400 values.	
Accuracy information Mean altitude, together with the 10 percentile and 90 percentile are to the nearest metre. Although the data set covers GB, approximately 500 coastal squares identified by CIS as "land" are not recognised as such by this data set.	

MEAN SLOPE	
OS95TOPO.ccf Altitude Data.	Monday November 6 13:56:33 95
Mean Slope (Percent) : per sq. km : OS95	
Description of data set Topographic data created from the Ordnance Survey's digital 1:50,000 Panorama data set in turn derived from the Digital Terrain Model data consisting of height values at each intersection of a 50m horizontal grid, the values of which have been mathematically interpolated from contours on the 1:50,000 Landranger maps. The mean gradient or slope of each kilometre square is calculated using the default slope algorithm employed by the OS. This uses a 3x3 operator that is passed over the 20x20 matrix column in each kilometre square column by column and row by row. The operator determines the slope at the centre point of the matrix by calculating the change in slope in the x and y direction for the surrounding matrix points, and combining the x and y components to provide an average measure of slope.	
Accuracy information Values for slope (both degrees and percent) are to the nearest one degree. Values for slope in degrees were calculated from the tan-1 percent slope values. Although data set covers GB approximately 500 coastal squares identified by CIS as "land" are not recognised as such by this data set.	