# Evaluation and Quality Control for the Copernicus Seasonal Forecasts and Products

**Poster EGU2019-18891-1**

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Photography encouraged

**Pierre-Antoine Bretonnière[1], Stephan Hemri[2], Nicolau Manubens[1], Joaquín Bedia[3], Jonas Bhend[2], Alasdair Hunter[1], Daniel San-Martín[3], Marta Burno Soares[4], and Francisco-Javier Doblas Reyes[1,5]**

[1]Earth Sciences Department, Barcelona Supercomputing Center, Barcelona, Spain , [2]MeteoSwiss, Zurich-Airport, Switzerland, [3]Predictia Intelligent Data Solutions, Santander, Spain,
[4]University of Leeds, Leeds, United Kingdom, [5]Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain

## 1. Introduction

This poster presents the work realized in the C3S 51 Lot 3 contract, devoted to the development of a strategy for the **evaluation and quality control (EQC) of the multi-model seasonal forecasts provided by the Copernicus Climate Change Service (C3S)**. The strategy aims at assessing product suitability and building trust on the service, while responding to the needs identified among a wide range of stakeholders. To achieve the objective, the consortium has evaluated the multi-faceted aspects of the EQC, performed a gap analysis of current information available to users, and developed a comprehensive framework and a proof-of-concept software prototype.

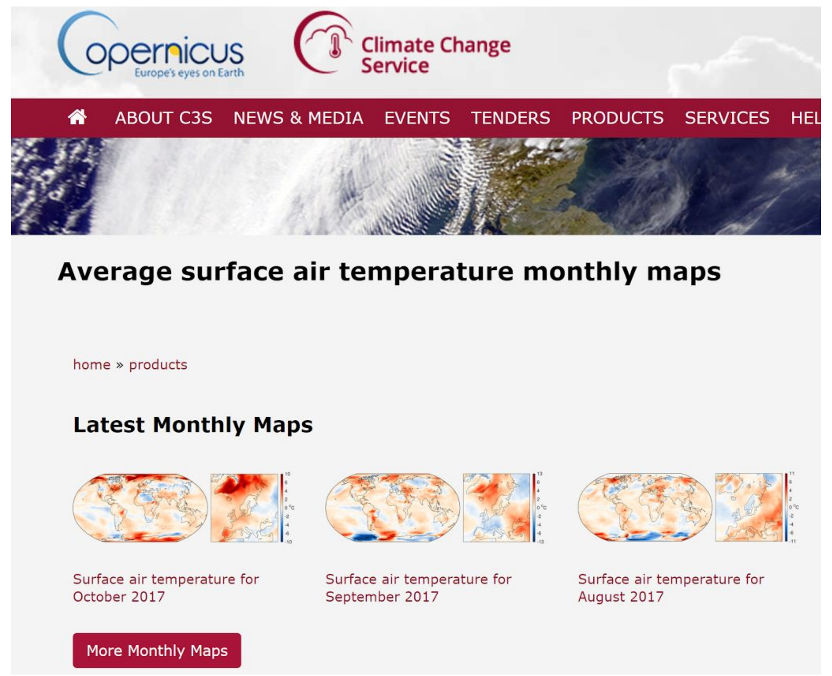



**Fig 1**: Screenshot of a section of the C3S web portal (still under development) showing records of monthly averages of air temperature at surface level.c

The C3S will provide authoritative information about past, current and future climate for a wide range of users, from climate scientists to stakeholders from a wide range of sectors including insurance, energy or transport.

Two examples of applications of this service are policy making in the insurance sector and decision making in the energy sector. Hence, the publication of erroneous products, or having a user choose an unsuitable product, could lead to huge third-party financial losses.

Establishing an EQC framework that assesses the suitability of the products is fundamental in order to minimize issues and build trust in the offered service.

## 2. Evaluation and Quality Control framework

The EQC strategy developed in the contract is multifaceted and comprehensive, including activities from assessment of user needs to the development of a software prototype.

These activities build upon three main pillars: first, providing **user guidance** for the user to select the correct product; second, estimates of the quality of the products, or **QA measures**, for the user to know how trustworthy a product is; and finally **provenance information** which, by one side, contributes to user guidance and, by the other side, allows for product inspection and comparison.

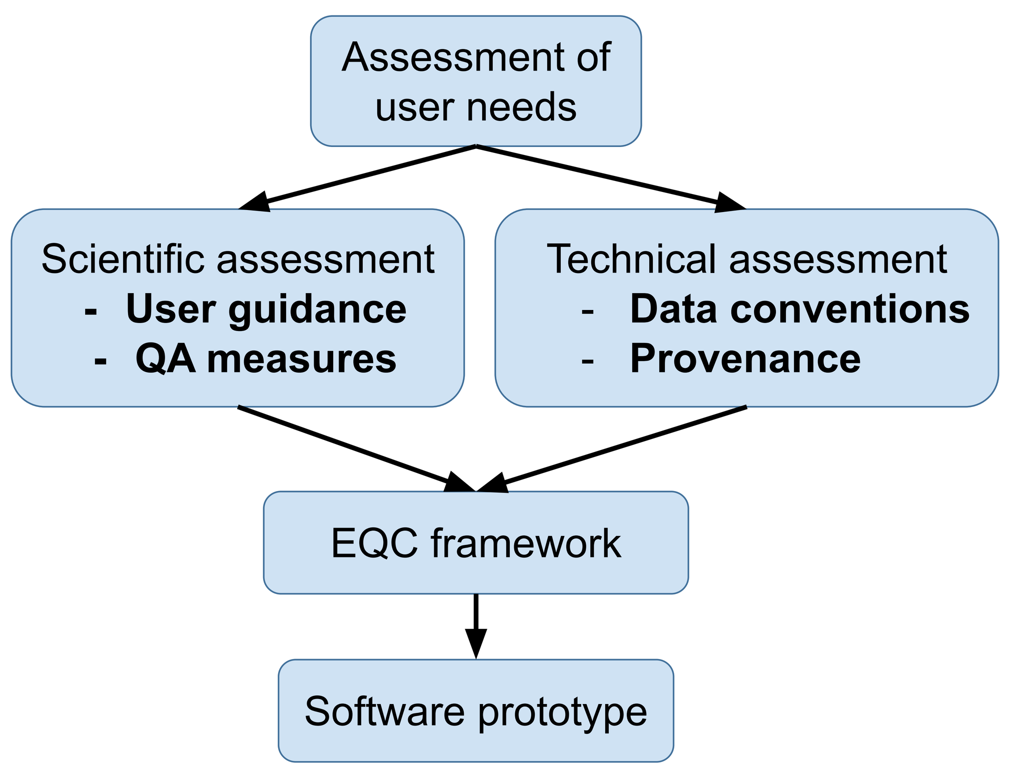



**Fig 2**: Diagram with the activities conducted in the C3S 51 Lot3 contract for the elaboration of the EQC strategy.

As part of the scientific assessment, a climate forecast verification framework has been developed. It defines a set of statistical verification scores to be calculated and provided together with each kind of forecast product provided in the C3S, as well as a graphical representation to convey such information.

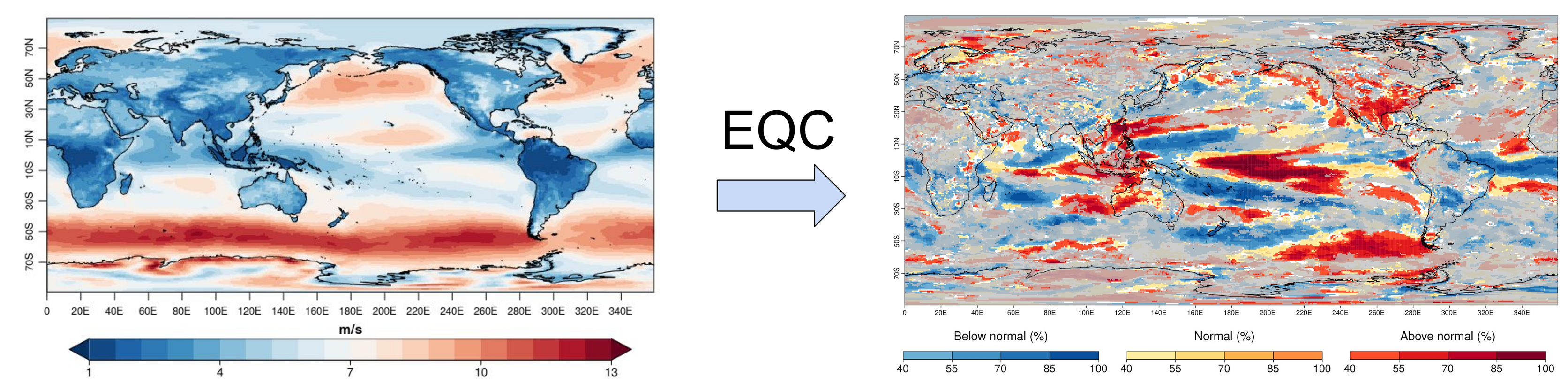



**Fig 3**: On the left, an example of a seasonal forecast product provided without EQC information: a map of the average wind speed forecast for the upcoming season, with all ensemble members averaged. On the right, the same forecast product with EQC information: the forecast is shown in terms of how likely it is to belong to the lower/medium/upper tercile (according to an observational climatology), and the areas where the forecast RPSS skill score is below 0 (no improvement compared to a simply issuing the climatology as forecast) are masked in gray.

## 3. Data conventions

Establishing a well-defined data convention for the seasonal forecasts to be stored in a homogeneous way and with all the required information, in addition to making the work viable and efficient, is the first indispensable step towards the development of a provenance mechanism. The convention defined by C3S and used in the EQC framework covers the following aspects:

File content | File format | Data structure | Metadata

It results from combining and extending other existing conventions:

CF | SPECS | CMIP | ACDD

The resulting convention establishes NetCDF3/4 as the standard file format, and provides indications on how to name and distribute the files, which attributes must be defined, how to define the space and time coordinates and axes, and includes also, for each climate variable, an extensive list of standard names, units, dimension names, cell methods, and others.

## 4. The METACLIP provenance framework

Having the data conventions established, a **data provenance framework**, called **METACLIP**, has been developed. It allows to encode and propagate the metadata involved in climate forecasting analyses in the form of a **directed graph**. In Figure 4, the blue balls represent a graph containing the information on the involved data sources and the essential metadata.

Once the basic set of metadata on the data sources involved in a use case is encoded, the resulting graph can easily be **extended with information on the steps applied** to the data (coloured in green in Figure 4). After all the processing steps have been encoded, the provenance graph can be **serialized** as an XML string, which can be **attached** to NetCDF or image files.

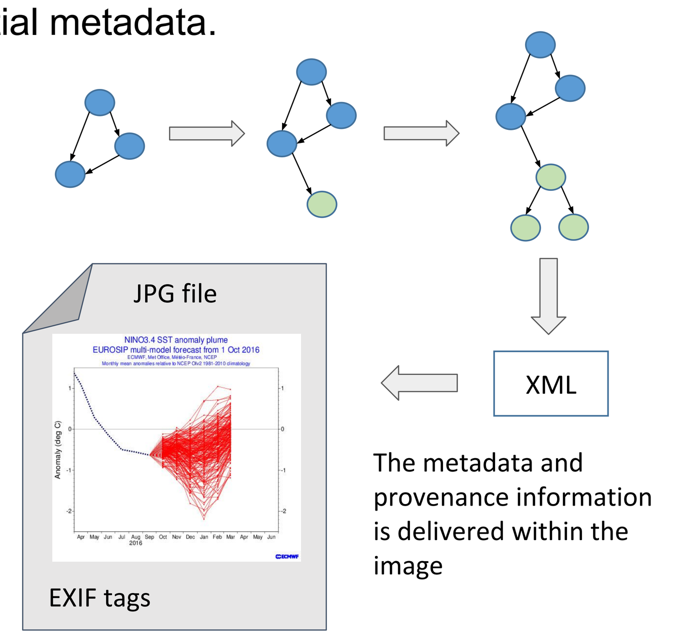
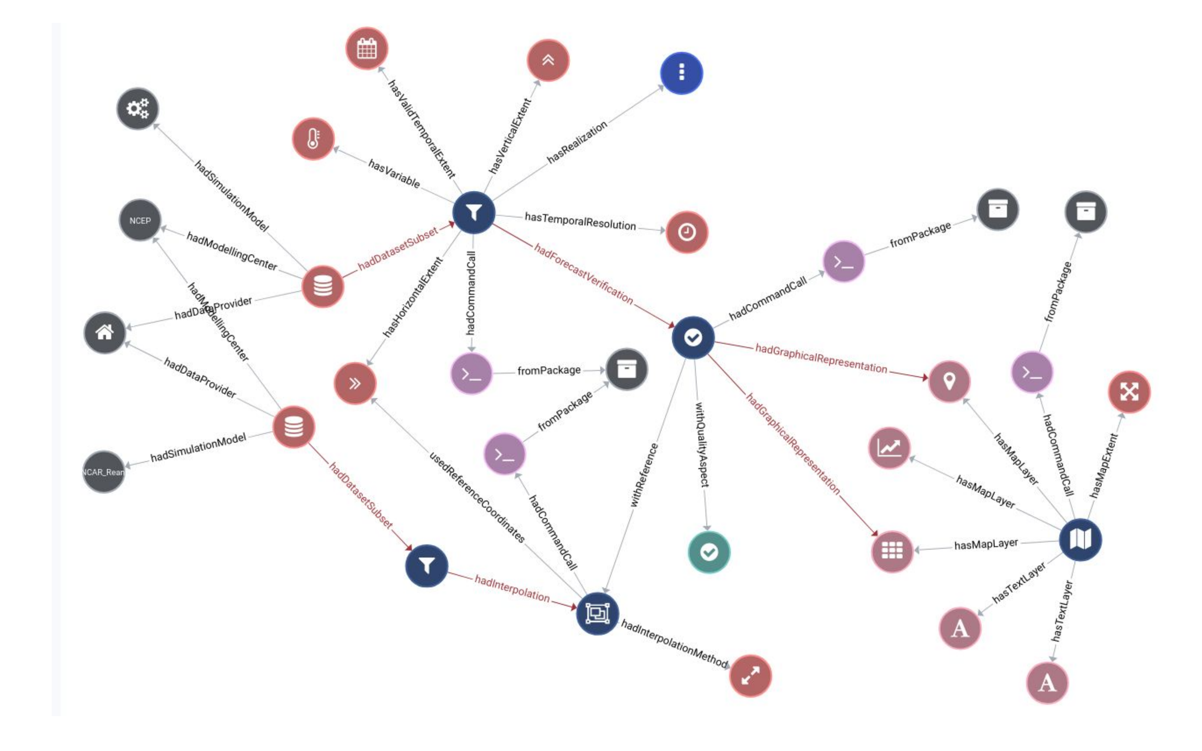



**Fig 4**: In the METACLIP data provenance framework, metadata is encoded in a directed graph, serialized and attached to final products..

Finally, the produced file with attached metadata can be ingested by a **provenance visualization engine** also developed in this contract (see Figure 5), which interprets and graphically displays the provenance and metadata graph in an interactive way. See live examples at **http://metaclip.org**.




For example, the forecast product with EQC measures shown in Figure 3, after being generated with METACLIP-compatible software, can be dragged and dropped onto the provenance visualizer, and the information on the applied steps would be displayed.

**Fig 5**: Screenshot of the web-based METACLIP provenance visualization engine. Forecast products with attached provenance information can be dragged and dropped, and the provenance graph is be displayed in an interactive way. The data sources and workflow steps are shown with different symbols and colours.

METACLIP is based on the **Resource Description Framework** (RDF), a popular W3C standard widely used for the web semantics, and benefits from existing community efforts by importing some already defined RDF vocabularies such as the PROV, the GEO or the SEQ. Thanks to its modularity, METACLIP can be extended to other research fields.

## 5. Software prototype

The last piece of the EQC strategy has consisted in the development of a proof-of-concept software prototype that demonstrates the viability of the recommended methods and technologies.

In Figure 6, a diagram is shown with the modules of the developed software prototype. Newly developed software modules are coloured in green, and other existing software modules used in the prototype are coloured in white, or yellow if adaptations have been required. Interactions between modules are represented with labelled arrows.

From perspective of the EQC prototype user, the prototype is started via a Python script, which automatically generates an extensive EC-Flow suite (workflow of EQC steps) according to the parameters defined in a configuration YAML file. The suite will dispatch EQC tasks to retrieve data, perform the required computations efficiently (multinode, multicore), and generate forecast products with EQC and provenance information.

Finally, the provenance of a forecast product can be inspected by uploading it onto the METACLIP provenance visualization engine.

The dashed red box highlights those software modules that are publicly available. They contain the essential EQC functionality considered in the developed suite, including the forecast verification scores and graphical products.

All of them are R packages published on the CRAN archive or on GitHub repositories.

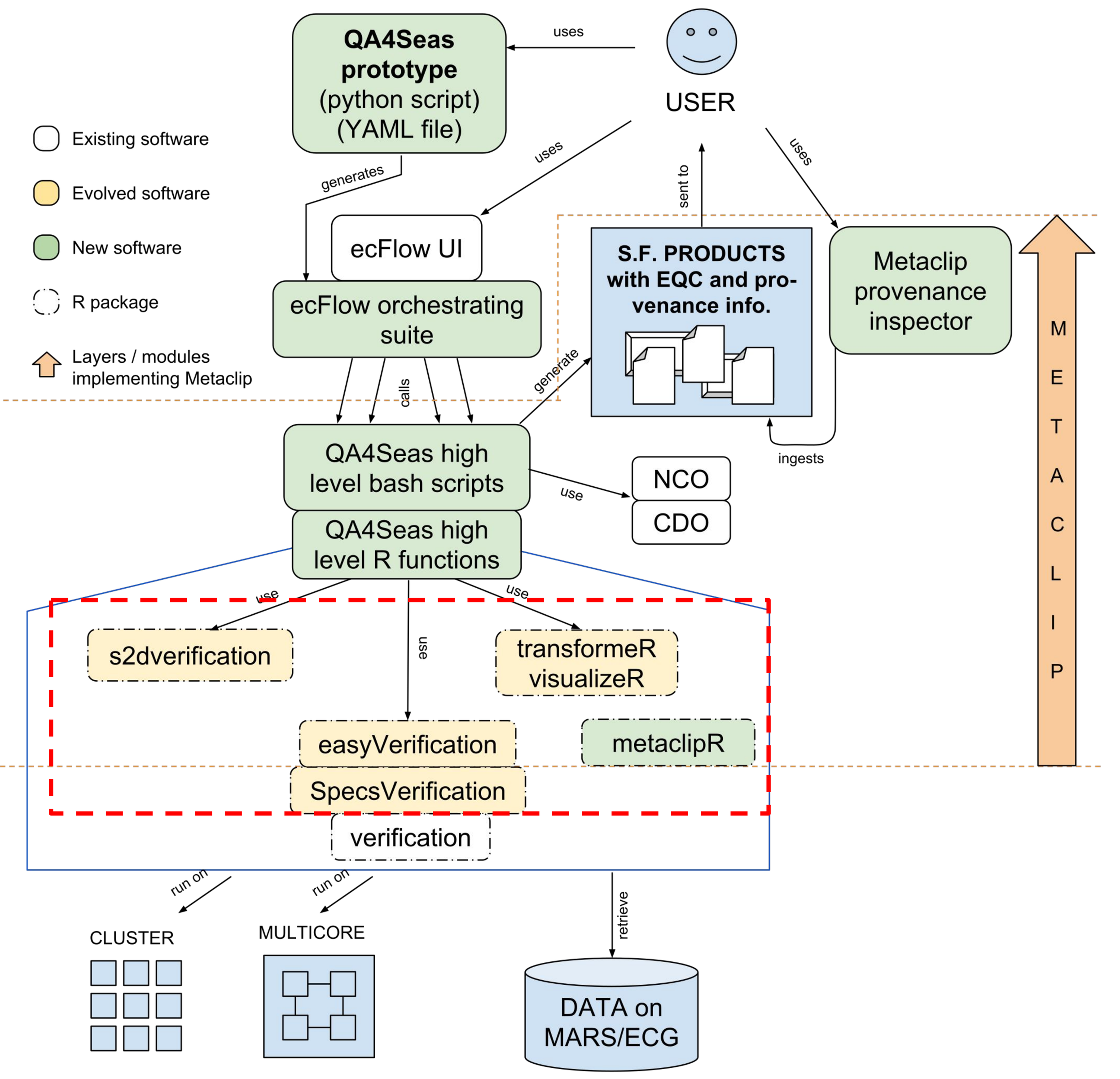



**Fig 6**: Module diagram of the software modules of the C3S 51 Lot 3 EQC prototype.

## 6. Achievements and future work

- An **evaluation and quality control strategy for climate seasonal forecasts** has been developed. It consists essentially in defining **quality measures for the different offered products** in the C3S, and in establishing a **provenance mechanism**.
- **User guidance** and **data conventions** are key.
- An extensible **provenance framework** called **Metaclip** has been developed.
- Open-source software for the EQC of seasonal forecasts has been developed.
- The presented EQC framework will be used and evolved for the **EQC of the overall C3S** data store in the C3S 512 contract.

## 7. Acknowledgments

The research leading to these results is part of the Copernicus Climate Change Service (C3S) (Framework Agreement number: C3S_51_Lot3), a programme being implemented by the European Centre for Medium-Range Weather Forecasts (ECMWF) on behalf of the European Commission.

Corresponding authors:
**pierre-antoine.bretonniere@bsc.es**
**nicolau.manubens@bsc.es**