SUPPLEMENT: Developing the Informationscape Approach to Environmental Change Detection

The main question that this research will pursue is: **How does spatial propagation of information (i.e., signals passed from one state variable to another) in the environment reflect past trajectories and indicate future change?** The work brings emerging statistical tools from information and network theory into the environmental sciences through a novel *informationscape* approach, which enables inductive determination of critical feedbacks, impending shifts in the underlying dynamics, and identification of key spatial and temporal scales of interaction from environmental observatory and remote sensing data. These powerful analysis tools have potentially transformative applications in environmental science (see pre-proposal and Table 1). I will continue advancing a diverse range of these applications, through mentorship, teaching, and evangelization to the broader scientific community. However, there are two key focus areas that I believe will have the most significant impact:

- **Emphasis 1.** Developing a spatially explicit, information-based early warning method of detecting impending critical transitions (i.e., between two system states with markedly different behavior or structure), testing its performance on systems that undergo well-documented critical transitions, and applying it to address management concerns in a high-stakes ecosystem restoration project.
- **Emphasis 2.** Establishing links between information-based process networks (i.e., how environmental systems are "wired") and spatiotemporal indicators. Information transfer happens over a wide range of spatial and temporal scales, but environmental data are generally either spatially *or* temporally dense. This emphasis addresses the need to identify comprehensive mechanistic indicators from commonly available datasets. It will also provide guidance on ideal locations for multi-sensor deployment.

Research need and novelty

One of the defining challenges of modern environmental science is that of making predictions in an era of nonstationarity. Namely, correlative relationships derived from past observations do not necessarily apply to future conditions, due to shifts in underlying drivers (e.g., climate). Predictions instead need to be based on a *mechanistic* understanding of the interactions and feedbacks between system components. The proposed work produces this mechanistic understanding inductively, through data-driven statistics that resolve causative—rather than just correlative—relationships between variables. The result will be improved simulation models for prediction, new understanding of key drivers and sensitivities, and data-driven early warning metrics for critical transitions between different system states (e.g., vegetated vs. desert; oxygenated/biodiverse vs. oxygen depleted; retention phase vs. release phase).

Critical transitions are diverse and often catastrophic (e.g., stock market crashes, disease outbreaks, onset of sepsis, governmental coups). Hence, they have garnered much interdisciplinary research interest. Leading techniques for predicting their proximity are based on statistical analyses of time series from a single system state variable (e.g., dissolved oxygen) at a point. However, these approaches rely on the assumption that the system's potential function (describing its dynamical equilibrium) is smooth near the transition point, which limits their applicability. They have also exhibited false positives and negatives and have not been successful at predicting transitions from ordered to chaotic regimes or in heterogeneous systems[1,2]. Recent reviews[3,4] have highlighted the potential for spatially rich datasets to warn of impending transitions but decry a lack of quantitative methods.

I propose an alternate approach to the detection of impending transitions, grounded in direct delineation of the spatiotemporal scale and information exchange rates of the interactions and feedbacks that govern system behavior. Theory suggests that near catastrophic transitions, the total information transferred through feedback relationships may decline, while that transferred through forcing relationships may grow[5]. As large-scale negative feedbacks that tend to maintain resilience disappear near a transition[6], the spatial scale of information coupling may decrease. My work will test and refine these hypotheses against existing datasets from systems undergoing well-understood transitions. It will also apply the refined methods to near real-time sensor data from the lower Owens River, a major water supply for Los Angeles and focus of a large-scale restoration effort. Recent fish kills and encroachment of vegetation on the main channel lead restoration managers to suspect that the river is near a critical

[1]Batt et al. 2013. PNAS 110, 17398-17403.
[2]Hastings and Wysham 2010. Ecol. Lett. 13, 464-472.
[3]Scheffer et al. 2012. Science 338, 344-348.
[4]Kéfi et al. 2013. Oikos 122, 641-648.
[5]Larsen and Harvey, 2010. Am. Nat. 176, E66-E79.
[6]Pascual and Guichard, 2005. Trends Ecol. Evol. 20, 88-95.

transition to a state characterized by excessive organic sediment accumulation and hypoxic conditions. The availability of sensor data, spatial nature of the feedbacks controlling organic carbon accumulation and water quality (i.e., linked to flow-vegetation interactions and channel-riparian connectivity), and urgent need for restoration decision support make this site an ideal application.

The tools: Information entropy and network statistics

Transfer entropy statistics (*T*) resolve directional information transfer (i.e., causal interactions) between pairs of variables that may interact nonlinearly. Specifically, transfer entropy measures the reduction in uncertainty in variable *y* that results from knowledge of variable *x* at a particular time lag *L* relative to the uncertainty reduction that results from knowing *y* at a previous time step (*t*-1; Table 1). Statistical significance can be assigned based on distributions of transfer entropy constructed through Monte Carlo shuffling of the input time series, and critical timescales of information exchange can then be defined as either the *L* that provides the earliest significant *T* or as the peak in significance[1]. A related quantity is mutual information (*I*), which assesses the extent to which *x* and *y* change synchronously. The relative values of *I* and *T* classify pairwise relationships as synchronous, feedback-dominated, or forcing-dominated. By examining pairwise relationships between all sets of variables within a system, subsystems can be identified as components of the system that share similar types of interactions among themselves and to other variables and interact on similar timescales. In this way, a mechanistic "wiring" diagram—or process network—for complex environmental systems can be inductively produced from data[7]. Other quantities, such as total system transport (the balance of significant information exchange across nodes), directional information flows, or node centralities can then be computed[8].

Though the type of analysis described above has been applied in the environmental sciences to resolve soil-plant-atmosphere feedbacks at a flux tower[7], many environmental processes involve spatial interactions between entities heterogeneously distributed in space. My work will build on at-a-point resolution of process networks by incorporating space, as elaborated in Table 1. Open-source code for all computations below will be freely available for other users, through the Environmental Systems Dynamics webpage (www.esdlberkeley.com) and through the Berkeley Institute for Data Sciences.

Table 1. Proposed transfer entropy metrics, based on generic form $T = \sum_t p(A,B,C)\log\frac{p(A|B,C)}{p(A|B)}$

A	*B*	*C*	Description	Use in environmental science (*italic = novel use*)
$y_{t,z}$	$y_{t-1,z}$	$x_{t-L,z}$	Transfer entropy at a point	Resolving process networks; *in flow-through systems: determining limiting factors for biotic growth*
$y_{t,z}$	$y_{t-1,z}$	$x_{t-L,z-Z}$	Spatially explicit transfer entropy	*Resolving fine-scale spatial process networks, determining critical interaction length scales*
y_{t,z^*}	y_{t-1,z^*}	x_{t-L,z^*}	Spatially averaged transfer entropy	*Resolving dominant system-scale interactions in spatially heterogeneous systems*
$y_{t,z}$	$y_{t-1,z}$	$y_{t-D/v,z-D}$	Self transport entropy	*Determining conservatively transported constituents*
$y_{t,z}$	$y_{t-D/v,z-D}$	$y_{t-L,z-D}$	Self delayed transport entropy	*Determining timescales of dispersion/tailing behavior of conservative constituents (L>D/v)*
$y_{t,z}$	$y_{t-D/v,z-D}$	$x_{t-D/v,z-D}$	Along-path transfer entropy	*Determining dominant interactions along a flow path*

t = current time, *L* = time lag, *z* = spatial position; *Z* = spatial lag; *z** = compiled over all positions, *D* = distance between upstream and downstream sensors, *v* = mean flow velocity (i.e., of water)

The spatially explicit transfer entropy, a prime focus here, produces a 4D adjacency matrix of transfer entropies, in which the dimensions represent the information inflow and outflow nodes, time lag *L*, and spatial lag Z. Network visualization and statistics will be key for interpreting these matrices, and a major output products of this project will be open-source software for this purpose. The adjacency matrix will first be culled based on a user-selected significance threshold or a maximum number of links. Users will be able to plot significant transfer entropies between nodes as directional links, weighting link thickness by *L* and coloring links by type of pairwise interactions. Nodes may be depicted according to their centrality or by the total amount of information transfer. Users can zoom in on nodes, restrict the range of variables/interaction types plotted, or page between different process networks generated for

[7]Ruddell and Kumar, 2009. Water Resour. Res. 45, doi:10.1029/2008WR007279
[8]Ruddell and Kumar, 2009. Water Resour. Res. 45, doi:10.1029/2008WR007280

sequential periods of time. The software will also compute the difference between networks and compare physical networks (e.g., channel networks, vegetation corridors) or transition-based networks computed from remote sensing observables (e.g., greenness, turbidity) to process networks.

Research steps and assessment of progress:
Progress will be measured in terms of successful completion of the following phases of research:

1. Development of code and software resources. Spatial entropy analysis codes and software for network visualization and statistics will be produced early, with the goal of making products available to the public as quickly as possible. In coordination with the Berkeley Institute for Data Science, project personnel, collaborators, and I will organize an annual summer "boot camp" to train users on the use and application of information and network theory statistics and the specific coding resources.
2. Testing/refining spatial early warning signals: Process networks will be computed from existing simulation-generated data from several systems undergoing known transitions. Initial tests will be a wetland undergoing a shift between a patterned, heterogeneous state to a homogeneously vegetated state[5] and a shear flow undergoing a spatial transition from a laminar (ordered) to turbulent (chaotic)[9] regime. Datasets consist of all state variables at all model time steps and at all cells in the gridded domain. The latter system[9] will provide a test of whether the spatial early warning signals apply to a type of transition that at-a-point early warning signals have been unsuccessful at predicting.
3. Linking process network statistics with mechanistic knowledge and spatial analysis: The focus of this step will be on identifying spatial locations and metrics indicative of critical landscape dynamics. Indicator locations from the process network analysis will be compared to node centrality statistics for physical networks generated from skeleton and edge features, as well as to changes in integrative spatial landscape characteristics, such as directional connectivity[10]. The process network structure relevant to early-warning signals will also be compared to synchronization networks (e.g., Boers et al.[11]) generated from the timing of transitions in gridded observables from remote sensing imagery.
4. Expansion of scope: The methods developed above will be applied to a broader set of critical transitions in landscapes, using model data within the Community Surface Dynamics Modeling System database. Simulation data from models of coral reefs, deltas, and river networks are among the candidate systems available. Notably, the analyses proposed here may also point to measurements that can discriminate between different mechanisms that result in similar landscape structures.
5. Application to natural system observations: Spatial early warning signals will be tested in the real world using remote sensing and available sensor data from parts of the world undergoing rapid change, including Gulf Coast wetlands, rivers prone to avulsion (e.g., the Brahmaputra), and arctic ecosystems experiencing effects of climate change. Techniques will also be applied to sensor data from the lower Owens River in an ongoing collaboration with Inyo County resource managers.

Summary

The informationscape approach is novel in 1) its expansion of applications of information- and network-based data analysis techniques in the environmental sciences and 2) its incorporation of space in information-based techniques. It addresses defining challenges in the fields of hydrology and ecology, including how to 1) make predictions in a nonstationary regime, 2) synthesize large environmental datasets to enhance our mechanistic understanding of environmental systems, and 3) make predictions of environmental change in poorly instrumented landscapes. Ironically, the test cases for method development are model simulation datasets that are extremely rich spatially and temporally. This rich information is needed, however, to identify critical scales for short-term sensor deployment or key statistics obtainable through remote sensing. An exciting future direction to emerge from successful completion of the above phases of research, together with the increasing availability of remote sensing products and near real-time data from environmental observatories, will be data-driven global mapping of regions most at risk of sudden change as a result of ongoing shifts in climatic and other drivers. Finally, advances in the detection of critical transitions, spatial entropy statistics, and network visualization methods are expected to be of broad utility in neurology, economics, epidemiology, biomedical engineering, and other fields that require complex-systems approaches to solving problems.

[9]Wu and Moin, 2009. J. Fluid Mech. 630, 5-41.
[10]Larsen et al., 2012. Ecol. Appl. 22, 2204-2220.
[11]Boers et al., 2013. Geophys. Res. Lett. 40, doi:10.1002/grl.50681.