Space in the Informationscape:
Resolution of critical spatial scales and functional connectivity in heterogeneous landscapes

Laurel Larsen[1], Mollie van Gordon[1], Christopher Tennant[1], Saalem Adera[1], Dino Bellugi[1], Hong-xu Ma[1], Theresa Oehmke[2]
1.) Dept. of Geography, University of California, Berkeley; 2) Dept. of Civil and Environmental Engineering, University of California, Berkeley; Contact: laurel@berkeley.edu

Problem Statement

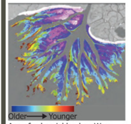
Transfer entropy is advancing our ability to quantitatively resolve causal mechanisms (e.g., feedbacks and forcings) through which environmental variables interact, based on data time series (Ruddell and Kumar, 2009). However, earth systems are inherently spatial. Here we showcase the Environmental System Dynamics Laboratory's early and ongoing efforts to unlock the potential of entropy-based methods to resolve spatially explicit earth system processes (i.e., **process connectivity**), the critical spatial scales and pathways along which information and mass travels (i.e., **functional connectivity**), and the implications of the spatial flow of information for system behavior. We show how entropy-based approaches may be used with different types of datasets, including data that are spatially dense but temporally sparse (e.g., remote sensing data), spatially and temporally dense but irregularly spaced (e.g., sensor network data), spatially and temporally sparse (e.g., environmental sampling data), and gridded model output data. Analysis strategies and limitations may vary depending on the existence of directional informaiton flow (e.g., water) and discrete spatial sources of perturbations to the system.

Data Type 1: Spatially Dense, Temporally Sparse (Remote sensing data)

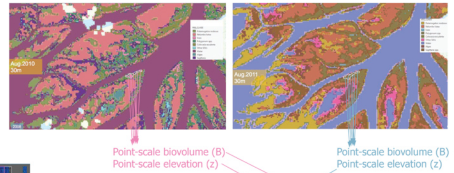
Many remote sensing datasets fall into this category. Depending on the motivating scientific question, researchers attempting to perform causal inference may take several approaches, which we detail here and in the next section.

Space-for-time substitution
If the dominant process of interest is expected to occur uniformly over the patch scale (i.e., in a way that is not spatially explicit at the sub-patch scale), information from patches of a particular type can be aggregated to develop the joint and marginal probability distributions from which transfer entropy is computed. Appropriate questions are restricted to those of causal inference between variables at the patch scale at the interval(s) between images.

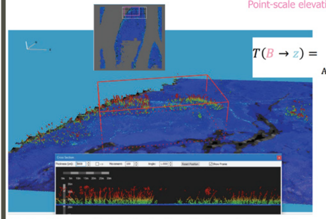
Example 1:
How does the geomorphic role of different vegetation communities vary on a growing river delta? In which communities does vegetation serve as a dominant influence on topography vs. vice-versa? In which communities can a bi-directional feedback be resolved, and how strong are the relationships?

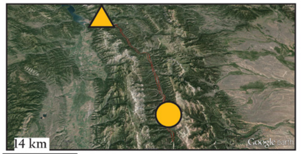


Age of subaerial land on Wax Lake Delta, Louisiana, USA (Turnipseed et al., 2014)



$$T(B \to z) = \sum_{\text{All tree pixels}} p(z_{2011}, z_{2010}, B_{2010}) \log \frac{p(z_{2011}|z_{2010}, B_{2010})}{p(z_{2011}|z_{2010})}$$


Extraction of biomass and bed elevation from lidar within tree vegetation community



Example 2: Spatially explicit information transfers with directionality

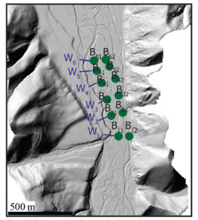
In the case of clear directionality (such as a river), spatial lags (ξ) in the parallel or perpendicular direction can substitute for temporal lags (τ) in the computation of transfer entropy. Two types of questions are readily addressable:

1) What are critical spatial scales of influence? (Example: Over what downstream distances do geomorphic perturbations--e.g., in river width--propagate?)

$$T(W_{upstream} \to W) = \sum_x p(W_x, W_{x-1}, W_{x-\xi}) \log \frac{p(W_x|W_{x-1}, W_{x-\xi})}{p(W_x|W_{x-1})}$$

2) What are critical processes governing directional systems, and the spatial scales over which they occur? (Example: Does riparian biomass/lateral valley slope, etc. play a governing role in channel characteristics? What is the spatial scale of that lateral connectivity?

$$T(B_x \to W_x) = \sum_x p(W_x, W_{x-1}, B_x) \log \frac{p(W_x|W_{x-1}, B_x)}{p(W_x|, W_{x-1})}$$



Green lidar imagery showing high-resolution river bathymetry and floodplain elevation in the vicinity of a tributary confluence. Here we have superimposed schematic transects along which river width would be calculated and discretized adjacent portions of the floodplain for the computation of biomass (or slope).

Acknowledgements and References

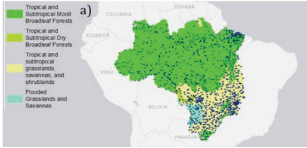
This work is supported by the Moore Foundation Data-Driven Discovery Investigator award and NSF CAREER award to LL, and a NASA Earth and Space Systems Fellowship to MVG. Vegetation classification maps were supplied by Melissa Carle (LSU and NOAA), and the Wax Lake Delta lidar imagery was supplied by Wayne Wagner. Special thanks to Ben Ruddell for transfer entropy process network toolbox and many useful conversations.

Ruddell, B. L., and P. Kumar. 2009. Ecohydrologic process networks: 1. Identification. Water Resources Research 45.
Ulanowicz,R .E. 1980. An hypothesis on the development of natural communities. J. Theor. Biol. 85, 233-245.

Data Type 2: Spatially and Temporally Dense, Irregularly Spaced (Sensor network data)

Spatially and temporally dense data offer researchers the ultimate flexibility. Transfer entropies may be computed between all pairs of variables, at all points in space, at all relevant time lags. Key challenges lie in the post-processing of the resulting highly dimensional process networks. The challenge may be addressed at the front end, by using knowledge of the system and driving questions to select which transfer entropy calculations are likely to yield the greatest insight, or at the back end, through complex network or machine learning approaches. Below we show how spatial networks of transfer entropy are likely to yield new insight into pressing questions in earth systems science, and how these challenges will be addressed.

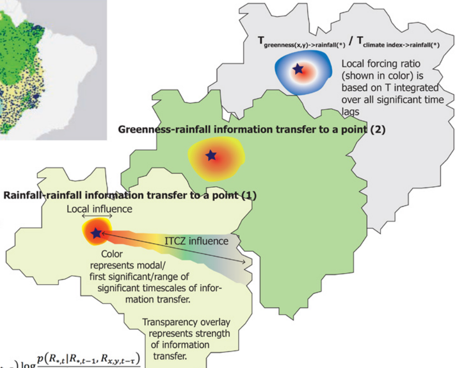
Example application: How is deforestation in the forest-savannah transition of Brazil impacting rainfall recycling and propagation?



Create these maps for all recipient points.

How do these scales of feedback and forcing change from forested to deforested areas?

Here, critical scales of information transfer are anchored to points in geographic space of interest.



Modified transfer entropy calculations:

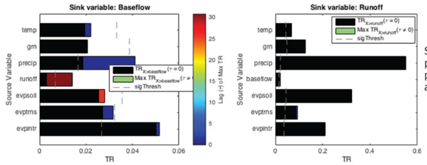
$$(1)\ T(R_{x,y} \to R_*) = \sum_t p(R_{*,t}, R_{*,t-1}, R_{x,y,t-\tau}) \log \frac{p(R_{*,t}|R_{*,t-1}, R_{x,y,t-\tau})}{p(R_{*,t}|R_{*,t-1})}$$

$$(2)\ T(NDVI_{x,y} \to R_*) = \sum_t p\left(R_{*,t}, NDVI_{*,t-\tau_{sig}}, NDVI_{x,y,t-\tau}\right) \log \frac{p\left(R_{*,t}|NDVI_{*,t-\tau_{sig}}, NDVI_{x,y,t-\tau}\right)}{p\left(R_{*,t}|NDVI_{*,t-\tau_{sig}}\right)}$$

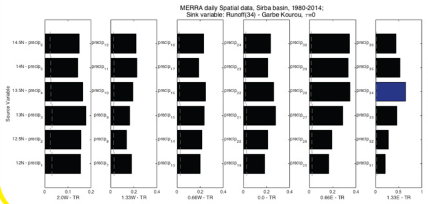
Example application 2: How have land use changes impacted hydrologic feedbacks in the Sahel?

Climate models perform notoriously poorly in representing land-atmosphere interactions over the African continent. In part due to a lack of data, scientists have a poor understanding of hydrologic feedbacks in the vicinity of the Sahel. Especially perplexing is the Sahelian Paradox, which describes decadal-scale increases in streamflow in some basins despite deepening drough conditions. In this study, we analyze for the first time a multi-source dataset, consisting of rain gauge network data, hand-classified land use data, and remote sensing data to resolve feedbacks, forcings, and their spatial scale. We test the hypothesis that changes in water balance partitioning due ultimately to land-use change are ultimately responsible for increased streamflow in the Sahelian Paradox. As shown in the figures below, process networks will be derived based on spatially averaged data (to delineate critical variables) and spatially explicit data (to delineate the spatial scales over which water recycling occurs and over which human influence changes the hydrologic cycle).

Sirba basin study area



Spatially-aggregated 0-lag and maximum transfer entropies to baseflow and runoff in the Sirba basin. One surprising finding was the long time lags over which runoff appears to influence baseflow.

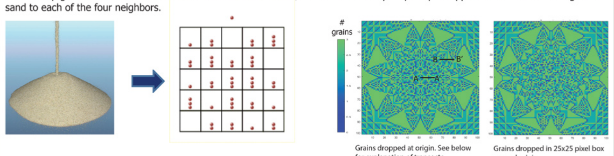


Transfer entropies between spatially distributed precipitation and runoff. Results likely illustrate the spatial coherence of rain events in this basin. Teasing apart critical spatial scales over which precipitation influences runoff will likely require formulating transfer entropy in a form similar to (2) above, conditioning probability distributions with respect to point-scale local precipitation at the most significant time lag (here, 0).

Data Type 3: Model Generated (Proof-of-concept)

In adapting transfer entropy concepts to spatially explicit systems, it is helpful to have proof-of-concept examples that show 1) That transfer entropy is able to resolve spatially explicit feedbacks, and 2) Transfer entropy can capture key properties of the system. Working with gridded spatially- and temporally-extensive datasets from models will also aid in the development of strategies for post processing highly dimensional information.

Example 1: Sand pile model
Rules: Drop grains of sand onto model domain. If there are four grains of sand on any cell, the pile topples and distributes one grain of sand to each of the four neighbors.



Grains dropped at origin. See below for explanation of transects.
Grains dropped in 25x25 pixel box around origin.

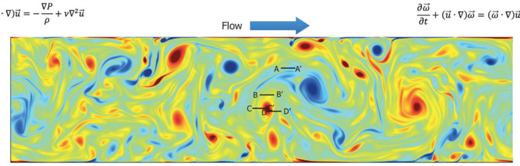
Example 2: Turbulent flow
Direct numerical simulations of the Navier-Stokes equations provide state-of-the-art resolution of fluival dynamics.

Incompressible Navier-Stokes:

$$\frac{\partial \vec{u}}{\partial t} + (\vec{u} \cdot \nabla)\vec{u} = -\frac{\nabla P}{\rho} + \nu \nabla^2 \vec{u}$$

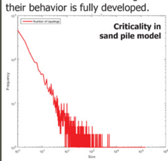
Written as vorticity equation:

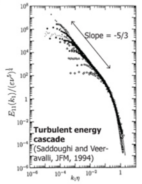
$$\frac{\partial \vec{\omega}}{\partial t} + (\vec{u} \cdot \nabla)\vec{\omega} = (\vec{\omega} \cdot \nabla)\vec{u} + \nu \nabla^2 \vec{\omega}$$



Vorticity field of a turbulent flow. Transects are paths with identical spatial lags, which in this data snapshot, would have distinctly different contributions to the joint and conditional probability distributions that comprise transfer entropy, due to the spatial nonuniformity of perturbations. Image source: http://www.win.tue.nl/smarter/index.php?page=werner

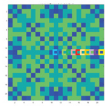
Key behavior: Both systems exhibit a power law distribution of critical length scales when their behavior is fully developed.

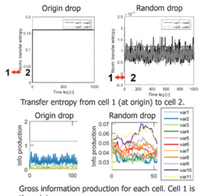




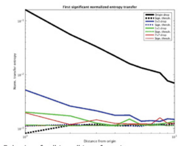
Hypotheses:
- Pixels on which sand grains are dropped should be producers of information.
- Maximum information transfer as a function of spatial lag will decline more steeply than the power-law function for event size (as whether a toppling at the "source" pixel influences a "sink" pixel depends on states of all intervening pixels.
- Critical spatial scales of information transfer should not be dependent on where the perturbation occurs.
- There should be an information analog for criticality and the turbulent energy cascade.
- In shear flow, information should propagate upstream for Fr<1.

Preliminary test: Drop 10,000 grains at origin vs. drop 10,000 grains randomly in 5 x 5 box around origin


Simplified computational domain, showing pathway along which information transfer is computed. Transfer entropy is based on the number of grains in a cell, conditioned by the number of grains in that cell at the previous time step.



Transfer entropy from cell 1 (at origin) to cell 2.
Gross information production for each cell. Cell 1 is the origin.



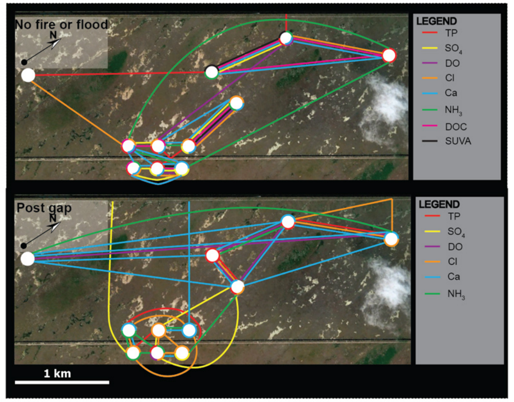
Behavior of cell-to-cell transfer entropy over length scale. For drops at the origin, the behavior is in accordance with a power law. Transfer entropy declines significantly when the source of information is spatially variable.

Methodological challenges unique to "information space":
- What variables to use in transfer entropy calculations to reveal critical spatial scales of information transfer? State change or sandpile height? Gross grains lost and gained? Vorticity or velocity? What are the implications of variable choice on information properties of the system?
- For each of the variables above, how to condition the probabilities in the computation of transfer entropy? By adjacent cell, previous time step, or both?
- How to compute significance level? By shuffling values in time, space, or both? How is the choice dependent upon the choice of variables?
- How do point-scale perturbations (e.g., vortices, sand grains), superimposed on a spatially heterogeneous system, impact the ability to inductively determine system properties? The challenge is that perturbations propagate directionally and in a spatially heterogenous manner. The probability distributions that comprise transfer entropy will be different depending on where they are anchored and the directions along which they are computed (see transects drawn within the figures). Is the solution spatial averaging? Identification of source nodes and key directions, and computation of probability distributions anchored to those sources and along those directions?

Data Type 4: Spatially and Temporally Sparse (Sampling data)

With spatially and temporally sparse data, transfer entropy techniques can not be used to resolve functional or process connectivity (Ruddell and Kumar, 2009). However, a conceptually analogous strategy based on tracking the propagation of perturbations through a system, can still be used.

Example: Tracing the impact of flow releases on solute connectivity in the Everglades
Part of Everglades restoration involves removal of manmade barriers to flow. As a pilot test of that management activity, flows and solute chemistry were monitored over an experimental landscape that experienced controlled flow releases in 2013 and 2014, shortly after gaps were breached in the surrounding levees. We modeled water quality at each sampled station as a linear combination of a site effect and time effect and computed residuals. The residuals (i.e., water quality signal perturbations) were then modeled with stepwise regression as a function of residuals and raw variable values at surrounding sites. Significant connections formed a link in the spatial functional connectivity network. By plotting the networks for all monitored solutes simultaneously, it is apparent that the flow releases substantially changed functional connectivity pathways in the Everglades landscape.



Interpreting Process and Functional Connectivity Networks

Spatial transfer entropy networks are highly dimensional and pose challenges for interpretation, highlighting a need for network characterization techniques. Information entropy provides several tools useful for characterizing and interpreting networks.

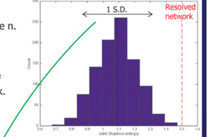
Example: How does a change in process or functional connectivity network structure impact the probable resilience of the system?
We asked this question for the water quality functional connectivity networks resolved above. Systems that feature a range of fundamental, discrete scales and diversity of functional entities within and across scales are typically viewed as having a high potential for resilience. This potential can be quantified by:
1) Determining the size of network components (connected portions of the network node connected to other nodes) and how components of each size are distributed among solutes, computed using Shannon entropy statistics.
2) Calculating the **capacity** and **redundancy** of the weighted network (Ulanowicz, 1979). Capacity refers to the evenness with which flows in a network visit the nodes. Redundancy refers to the multiplity and evenness of flow paths out of each node once a solute has reached the node.

$$C = \sum_{n=1}^{N} Q_n \log Q_n$$

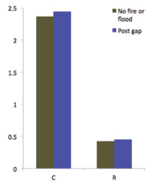
Capacity. Q_n represents the proportion of total network flows visiting node n.

$$R = \frac{1}{C}\left[C - \sum_{k=1}^{N}\sum_{j=1}^{N} g_{jk} Q_j \log\left(\frac{g_{jk}}{\sum_{n=1}^{N} g_{ik} Q_n}\right)\right]$$

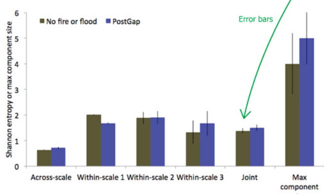
Redundancy. g_{jk} represents the proportion of flows passing through node j that go to node k.



PDF of joint Shannon entropy of shuffled surrogate networks with same number of connections as resolved network.



Capacity and redundancy of Everglades water quality networks, pre- and post-flow.



Shannon entropy within and across network components consisting of 1, 2, and 3 nodes Pre-flow-release conditions are in brown; post flow in blue.

By all metrics examined, water quality networks seemed to marginally increase in their capacity for resilience following restoration of flow in the Everglades.