WATER RESOURCES RESEARCH, VOL. 45, W03420, doi:10.1029/2008WR007280, 2009

Click Here for Full Article

# Ecohydrologic process networks: 2. Analysis and characterization

Benjamin L. Ruddell[1] and Praveen Kumar[1]

Received 11 July 2008; revised 25 November 2008; accepted 19 December 2008; published 25 March 2009.

[1] Ecohydrological systems are complex, open dissipative systems characterized by couplings and feedback between subsystems at many scales of space and time. The information flow process network approach is developed to analyze such systems, using time series data to delineate the feedback, time scales, and subsystems that define the complex system's organization. Network statistics are used to measure the statistical feedback, entropy, and net and gross information production of subsystems on the network to study monthly process networks for a Midwestern corn-soybean ecosystem for the years 1998–2006. Several distinct system states are identified and characterized. Particularly interesting is the midsummer state that is dominated by regional-scale information feedback and by information flow originating from the ecosystem's photosynthetic activity. In this state, information flows both "top-down" from synoptic weather systems and "bottom-up" from the plant photosynthetic activity. A threshold in air temperature separates this summer state where increased organization appears from other system states. The relationship between Shannon entropy and information flow is investigated. It is found that information generally flows from high-entropy variables to low-entropy variables, and moderate-entropy variables participate in feedback.

**Citation:** Ruddell, B. L., and P. Kumar (2009), Ecohydrologic process networks: 2. Analysis and characterization, *Water Resour. Res.*, *45*, W03420, doi:10.1029/2008WR007280.

## 1. Introduction

[2] Ecohydrologic processes comprise nonlinear couplings between climate, soils, water cycle and vegetation [*Reiners and Driese*, 2003]. In the first part of the paper [*Ruddell and Kumar*, 2009] it is argued that in this context, system state is best represented as a pattern of couplings between the various processes, rather than by measurements of individual variables. This arrangement may be described as a network of directional couplings and feedback cycles between a system's variables at a range of spatiotemporal scale. A process network was defined as "a network of feedback loops and the associated time scales that depicts the magnitude and direction of flow of matter, energy, and/or information between the different variables" [*Ruddell and Kumar*, 2009, paragraph 3]. Process networks were developed where the strength of each directional coupling is measured by the information flow between pairs of variables at a specified time scale. Information flow is the contribution of uncertainty-reducing or predictive information provided by the time lag history of one variable to the future value of another. The transfer entropy statistic [*Schreiber*, 2000] was used to measure the asymmetric information flow between two variables.

[3] Detailed analysis of process networks constructed for drought and healthy states of an agricultural ecosystem in the Midwestern United States showed that a healthy system is characterized by a predominance of feedbacks at a variety of time scales. On the other hand a drought system is characterized by a breakdown in the number of information feedback loops in the interaction between the various variables and the system seems to shift more to a source-sink type coupling rather than a feedback driven coupling [*Ruddell and Kumar*, 2009, Figures 7 and 8].

[4] Analysis of dynamics over a network poses significant challenges [*Strogatz*, 2001]. While a number of approaches exist for the analysis of structural properties of large network such as small world [*Watts and Strogatz*, 1998] and scale-free properties [*Albert and Barabasi*, 2002; *Newman*, 2003], the study of asymmetric information flow through a network remains an open and challenging problem. If a system's state is defined as a network of feedback couplings, then one should be able to learn about system states and dynamics by observing changes in statistics that characterize the properties of information flow in the process network.

[5] The goal of the present paper is to develop measures to characterize the organization of process networks, understand how the organization changes in time, and identify and characterize network-scale emergent properties. By summarizing a complicated process network using network statistics, many system states can be quickly and quantitatively compared. The result is a powerful and flexible statistical approach to the analysis of complex ecohydrologic systems, which can resolve key characteristics such as feedbacks, time scale, and subsystem organization. The organization of the process network is analyzed by identifying feedback, sources and sinks of information, and their seasonal variability. After characterizing distinct system states, specific patterns of organization are studied and the

[1]Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA.

Copyright 2009 by the American Geophysical Union.
0043-1397/09/2008WR007280$09.00

parameters controlling the emergence of these patterns are identified.

[6] This paper is organized as follows. In section 2, statistical measures for characterizing the organization of the asymmetric information flow in a process network are derived. In section 3, the seasonal and annual ecohydrological patterns that appear in the network statistics for the years 1998–2006 are analyzed. Conclusions and discussion are given in section 4.

## 2. Statistical Measures of Information Flow in a Network

### 2.1. Information Flow Process Networks

[7] Methods used to compute the entropy and information flow will now be briefly reviewed. They are presented in detail by *Ruddell and Kumar* [2009]. Let $X_t = \{x_t\}_{t=1,2,\ldots,n}$ be a discrete time series with marginal probability density function $p(x_t)$. Two time series are denoted as $X_t$ and $Y_t$, or equivalently as $X_t^{(i)}$ and $X_t^{(j)}$. The Shannon entropy [*Shannon*, 1948] of time series $X_t$ is $H(X_t) = -\sum p(x_t) \log p(x_t)$. $H$ is bounded as $0 \leq H(X_t) \leq \log(m)$, where $m$ is the number of discrete states that may be taken by $X_t$ or the number of bins used to discretize the probability density function. The normalized Shannon entropy is $H' = H/\log(m)$, and it takes values between zero and one. Larger $H'$ means a time series is less organized and less predictable, and therefore has greater variability.

[8] The time-lagged transfer entropy $T$ quantifies the reduction in the Shannon entropy of time series $Y_t$ provided by knowledge of time series $X_t$ at some time lag $\tau\Delta t$, which is additional to the reduction of entropy because of knowledge of the single point immediate history of time series $Y_t$, given by

$$T(X_t > Y_t, \tau) = \sum_{y_t, y_{t-\Delta t}, x_{t-\tau\Delta t}} p(y_t, y_{t-\Delta t}, x_{t-\tau\Delta t}) \cdot \log \frac{p(y_t | (y_{t-\Delta t}, x_{t-\tau\Delta t}))}{p(y_t | y_{t-\Delta t})}. \quad (1)$$

$T$ is bounded by $\log(m)$; the normalized quantity $T' = T/\log(m)$ is used. The transfer entropy effectively measures the directional flow of information from $X_t$ to $Y_t$ at the specified time scale.

[9] The computation of the entropy statistics $T'$ and $H'$ from discrete time series data depends on the estimation of a discrete marginal and joint probability densities. The state space for each variable is discretized by identifying the minimum and maximum observed values assumed by the variable, and then dividing the resulting range into $m$ discrete state partitions (bins) of equal size. The number of discrete states is fixed at $m = 11$ on the basis of arguments presented by *Ruddell and Kumar* [2009].

[10] *Ruddell and Kumar* [2009] compute entropy statistics for a single month at a time, so the minimum and maximum bounds of each variable's states are chosen to match the range of values observed during the specific month. This will be termed the "local" scheme for the discretization of state. However, in this paper entropy statistics are computed for 108 months, with the intention of comparing the results of one month with those of another. It is impossible to compare one month to another if the frame of reference is continually shifting [*Gershenson and Heylighen*, 2003]; when making these intermonthly comparisons, the minimum and maximum bounds are chosen to match the extreme range of values observed during all months compared. This will be termed the "global" scheme for the discretization of state, and will be used in this paper unless otherwise stated. A chronic problem facing discretization schemes is that of inadequate bins and sample sizes. Using a global discretization scheme increases the likelihood that inadequate bins will be used, since during each month a variable may only visit a few of the eleven discrete global states, effectively reducing the number of bins used. To control for negative impacts on the quality of results presented here, sensitivity analysis was performed with varying numbers of bins, and it was found that the methods are adequately robust, if not ideal.

[11] The basis of this analysis is the information flow process network consisting of pair wise couplings between a number $n_V$ of observed variables in an ecohydrologic system. For each month it is rendered as an adjacency matrix $\mathbf{A}(i, j, \tau) = T'(X_t^{(i)} > X_t^{(j)}, \tau)$ of dimension $[n_V \times n_V \times n_\tau]$, where $n_\tau$ is the number of time steps under consideration. The property $\mathbf{A}(i, j, \tau) \neq \mathbf{A}(j, i, \tau)$ captures the asymmetry of information flow between the two variables. Consider the entire network $(i, j) \in \mathrm{V}$ or a subset $(i, j) \in S \subseteq \mathrm{V}$ where the subset $S$ characterizes a subsystem consisting of $n_s$ variables or $n_s \times n_s$ nodes (or elements) in the matrix. For convenience this is denoted as $\mathbf{A}(S, \tau)$. When $n_s = 1$ the subsystem $S$ contains a single node $i$ in the network which corresponds to a time series variable $X_t^{(i)}$. The vector $\mathbf{H}$ is employed; $\mathbf{H}$ is of length $n_S$, and contains the normalized Shannon entropy for each variable in $S$ obtained as $H(i) = H'(X_t^{(i)})$ or $H'_X$.

[12] All analyses in this paper are performed using a weighted network adjacency matrix; unweighted or weighted-cut matrices are not used [see *Ruddell and Kumar*, 2009]. The weighted matrix provides clearer results than the cut matrices, in agreement with the experiments of *Wilhelm and Hollunder* [2007]. The authors believe that cut matrices may give higher-quality statistics for a much larger number of links in the matrix (i.e., thousands). However, in the present work the network has on the order of 100 links. For such a small number of links, the location of the cutting threshold tends to dominate the resulting network statistics, obscuring the meaning of the process network structure itself.

### 2.2. Derivation of Network Statistics

[13] A number of network statistics can be computed from the entropy vector $\mathbf{H}$ or the adjacency matrix $\mathbf{A}$. Network statistics can be computed for the entire network $V$ or for a subset $S \subseteq \mathrm{V}$ consisting of $n_s$ nodes.

[14] 1. The mean normalized Shannon entropy of a subsystem $S$ is computed from $\mathbf{H}$ as

$$H_S^m = \frac{1}{n_S} \sum_{i \in S} \mathbf{H}(i). \quad (2)$$

$H_S^m$ is bounded between zero and one. $H_S^m$ is the subset $S$'s equivalent to the basic Shannon entropy $H'$. It gives the average entropic uncertainty of the nodes in the subsystem, and allows comparison of one subsystem with another. It is important to note that $H$ is computed for the data set time

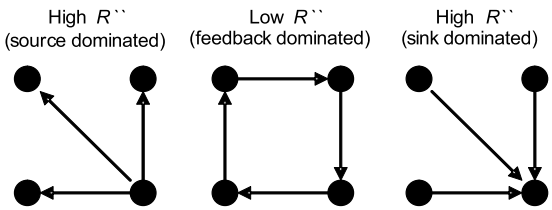


**Figure 1.** Network feedback statistic illustrated for three example cases on a four-node network: (middle) circular flows dominate the network, resulting in a minimum $R''$ which indicates feedback dominance; (left) flows from a single source dominate the network, resulting in a higher $R''$; and (right) flows into a single sink dominate the network, resulting in a higher $R''$.

scale $r$; $r$ is omitted from notation because in this study all analyses are for $r$ = 30 min.

[15] 2. The total system transport (TST) is the sum of all information flowing within a subsystem. It is computed for subsystem $S$ for a specific time scale $\tau$ from $\mathbf{A}$ as

$$TST_S(\tau) = \sum_{i,j \in S} \mathbf{A}(i,j,\tau). \tag{3}$$

The mean TST for subsystem $S$ is

$$TST_S^m(\tau) = \frac{1}{n_S^2} \sum_{i,j \in S} \mathbf{A}(i,j,\tau). \tag{4}$$

$TST_S$ is bounded as $H_S^m \times \mathrm{n}_S^2$ and $TST_S^m$ is bounded as $H_S^m$. $TST_S^m$ measures the total flow of information in the subsystem, as a fraction of the total possible flow $n_S^2$, which occurs if all weights of the couplings between nodes in the subsystem have their maximum value of 1.

[16] 3. The mean gross information production of a subsystem is computed for a specific time scale as

$$T_S^{[+]}(\tau) = \frac{1}{n_V \cdot n_S} \sum_{\substack{i \in S \\ z \in V}} \mathbf{A}(i,z,\tau). \tag{5}$$

The information production measures the total predictive information provided by this subsystem to all nodes in the system. A subsystem which is strongly coupled to other subsystems will have a large gross information production. Note, however, that this does not necessarily mean such a subsystem is controlling the other subsystems, only that it can be used to predict other subsystems because it is strongly coupled to them on average.

[17] 4. The mean gross information consumption is computed similarly as

$$T_S^{[-]}(\tau) = \frac{1}{n_V \cdot n_S} \sum_{\substack{i \in S \\ z \in V}} \mathbf{A}(z,i,\tau). \tag{6}$$

The information consumption measures the total predictability that all nodes in the system provide to this subsystem $S$.

[18] 5. The mean net information production is the difference between the gross production and consumption,

$$T_S^{net}(\tau) = T_S^{[+]}(\tau) - T_S^{[-]}(\tau). \tag{7}$$

The net information production becomes meaningful in the case where subsystem $S$ is synchronized with another subsystem. Synchronization can be caused by one-way forcing or balanced two-way feedback, and the net information production measures the extent to which of these alternatives applies. A subsystem which is controlling the rest of the network has a large net positive information production, serving as a net source of information in the network. This subsystem controls other subsystems more than they control it; these tend to be the "original causes" of changes which occur in the system. The net and gross information production should be used together, such that the gross information production measures the strength of coupling of the subsystem to the rest of the network (on average), and the net information production measures the extent to which the subsystem is controlling the rest of the network, given its gross information production. A ratio of the two, $T^{net}/T^{[+]}$, is potentially useful as an index which quantifies how large the net export is, compared with the total export of information. All types of information production are bounded above by $H_S^m \times n_S$. Information production and consumption are conserved on the total network; the net production of the whole system $V$ sums to zero.

[19] 6. A final statistic, the source-sink redundancy $R$, is a measure of the topology of the network's couplings. Many interesting topology statistics may be computed, such as the radius, mean connection length, clustering coefficient [*Strogatz*, 2001], and the medium articulation [*Wilhelm and Hollunder*, 2007]. The redundancy is particularly interesting because it measures whether the network topology is dominated by circular connections, that is, feedback between nodes, or whether the topology is dominated by a few nodes which are the source or sink of most of the information (Figure 1). Because the redundancy serves as an index of feedback in the network, it can be used to identify subsystems and time scales where feedback dominates the flow of information.

[20] *Wilhelm and Hollunder* [2007] generalized existing methods that had been applied for the study of food webs and ecosystem structure, applying the statistics to directional, weighted binary networks. Their generalization applies a source-sink analogy to every coupling in the network, and asks the questions (1) given a source node $i$, what is the level of entropy (uncertainty) that exists with regard to the identity of the sink node $j$ and (2) given a sink node $j$, what is the level of entropy (uncertainty) that exists with regard to the identity of the source node $i$? To answer these questions the weighted information flow adjacency matrix $\mathbf{A}$ for subsystem $S$ must be converted into the form of a joint probability density matrix $\mathbf{Ap}$ by dividing it by the TST as

$$\mathbf{Ap}(S,\tau) = \begin{cases} \dfrac{\mathbf{A}(i,j,\tau)}{TST_S(\tau)} & \text{if} \quad (i,j) \in S \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

For each time scale $\tau$, the matrix $\mathbf{Ap}$ sums to one across the indices $i$ and $j$ of subsystem $S$. For convenience, $\mathbf{Ap}$ can take two forms of notation, $\mathbf{Ap}(i,j,\tau)$, $(i,j) \in S$, and $\mathbf{Ap}(S,\tau)$, in the same manner as $\mathbf{A}$.

[21] Using **Ap** as a two-dimensional joint probability density estimate for the identities of the source and sink nodes, compute the conditional entropies of subsystem $S$ as

$$H_S^{[i|j]}(\tau) = -\sum_{i,j\in S} \mathbf{Ap}(i,j,\tau) \log \frac{\mathbf{Ap}(i,j,\tau)}{\sum_{z\in S} \mathbf{Ap}(z,j,\tau)}, \tag{9}$$

which is the answer to question 1 above, and

$$H_S^{[j|i]}(\tau) = -\sum_{i,j\in S} \mathbf{Ap}(i,j,\tau) \log \frac{\mathbf{Ap}(i,j,\tau)}{\sum_{z\in S} \mathbf{Ap}(i,z,\tau)}, \tag{10}$$

which is the answer to question 2 above. The conditional entropies are bounded as $\log(n_S)$. The redundancy of subsystem $S$ is their sum,

$$R_S(\tau) = H_S^{[i|j]}(\tau) + H_S^{[j|i]}(\tau). \tag{11}$$

The source-sink redundancy is bounded as $0 \leq R_S \leq 2\log(n_S)$. Normalizing $R_S$ by this bound results in a quantity $R'$ that is bounded between zero and one. For convenience the subscript $S$ is dropped when the context of a subsystem is unambiguous.

[22] The source-sink redundancy may be intuitively understood as an index for feedback on the system; when $R$ is low, the network is characterized by equal participation of all nodes in cyclical flows of mass, energy, or information (in which case there is a lot of feedback). The lowest value of $R$ occurs when there is no uncertainty in the identity of sources and sinks, which can only occur when each node sends to and receives from exactly one other node. Logically, in such a case, the network graph must form a loop about which information flows in a single direction. Such networks are rare in the physical world; it is more common for a subset of nodes to dominate the pattern of connectivity [*Wilhelm and Hollunder*, 2007].

[23] The absolute value of $R'$ is not very meaningful, because this statistic is relative to the size of the network and the distribution of weights. It is more meaningful to compare $R'$ with the equivalent statistic computed from a shuffled surrogate version of **Ap**, where the coupling probability weights are maintained but the source and sink connectivity of each coupling is randomly "rewired" (rewiring is performed independently for each time scale). By computing the shuffled surrogate redundancy many times in Monte Carlo fashion (50 times in this paper), a mean may be obtained for the distribution of shuffled surrogate redundancies, $\mu(R'_{ss})$, using a procedure similar to *Ruddell and Kumar* [2009] and *Wilhelm and Hollunder* [2007]. The difference between $R'$ and the mean of the surrogates gives a more useful statistic, the surrogate-relative redundancy of subsystem $S$ at time scale $\tau$, $R''_S(\tau) = R'_S(\tau) - \mu(R'_{ss}(\tau))$. $R''_S$ is bounded as $-1 \leq R''_S \leq 1$.

[24] Values of $R''$ below zero indicate that subsystem $S$ at time scale $\tau$ is relatively feedback dominated (i.e., characterized by circular information flow) as compared with a randomly connected network of the same size and distribution of weights. Values of $R''$ above zero indicate that $S$ is relatively source-sink dominated (i.e., most information flows in or out of a few dominant nodes).

## 3. Results: Seasonal Patterns in a Corn-Soybean Ecosystem

[25] These results attempt to answer three key questions about the Bondville, Illinois corn-soybean ecohydrologic system (see section 3.1). First, how is the process network organized? Corollary questions include: What are the coupling time scales? When does feedback couple those to form self-organizing structures? What nodes are sources or sinks of information? What distinct states does the system occupy? Are there seasonal patterns in network organization? Second, when does the process network show "emergent" properties [*Corning*, 2002], where emergence is defined as a sudden increase in the production of information by a process network subsystem, or the strengthening of a key information feedback loop between two subsystems, during a specific system state? Finally, what are the system control parameters [*Haken*, 1988] which explain the emergence of organized structures in the process network?

[26] Section 3.1 provides a review of the experimental framework used for this study, including the data and preprocessing methods applied. In section 3.2 the derived network statistics are validated by demonstrating that the statistics provide qualitatively similar results when compared with a more detailed manual process of network analysis described by *Ruddell and Kumar* [2009]. The key network feedback time scales are also identified, and turn out to be 30 min and approximately 14 h. In section 3.3 the seasonal and annual patterns in the network statistics are analyzed, and it is found that two distinct system states (summer and winter) and three summer substates (A, B, C) are distinguishable. One substate, "summer B," is centered on the month of July and characterized by emergent properties including regional feedback and information production by the ecosystem. In section 3.4, emergent properties of the ecohydrological system are analyzed, and the parameters which control the emergent structures are identified.

### 3.1. Experimental Framework

[27] To answer the questions posed above, the network statistics derived in section 2 are applied to the Midwestern corn-soybean ecohydrological system for each of 108 months spanning the years 1998–2006. The FLUXNET [*Baldocchi et al.*, 2001b] eddy flux tower at Bondville, Illinois [*Hollinger et al.*, 2005, *Meyers*, 2008] is this study site. See *Ruddell and Kumar* [2009] for a detailed description of the data and quality issues. The only departure from *Ruddell and Kumar* [2009] is that the cloud cover and net ecosystem exchange variables are excluded. The net ecosystem exchange is highly redundant with the gross ecosystem production, and the meteorological data from which cloud cover is computed is not uniformly available across the full time range 1998–2006.

[28] During each month, 37 independent time scales of information flow coupling, ranging from 0 to 18 h, are analyzed using half-hour increments. The time series data set is preprocessed and transformed using a 5 day periodic anomaly, such that the data used for this study is the departure from the average 5 day diurnal pattern, rather

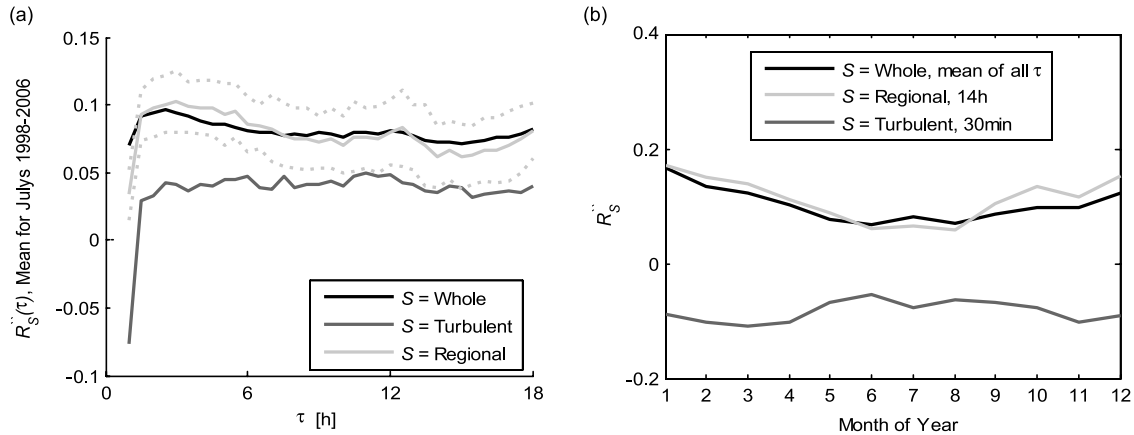


**Figure 2.** (a) Mean July values of $R''$, for the regional and turbulent subsystems and the whole system, for each $\tau$. Lower values of $R''$ mean that more feedback is occurring. $R'' < 0$ means that the subsystem is feedback dominated. The 30 min time scale has the lowest $R''$ for the turbulent subsystem, and the 14 h time scale has the lowest feedback for the regional subsystem. (The regional subsystem includes the turbulent and ABL subsystems.) Dashed lines show the standard deviations of $R''$ for the regional subsystem; $R''$ for the 14 h time scale is more than 1 standard deviation below $R''$ at the minimum regional feedback time scale of 2 h. (b) Mean annual pattern in $R''$ values for the whole system, regional subsystem, and turbulent subsystem for each month 1998–2006. $R''$ values for the whole and regional subsystems decrease (increased feedback) during the warmest months (June, July, August), but $R''$ values increase (decreased feedback) for the turbulent subsystem during the same months.

than observed values [*Ruddell and Kumar*, 2009]. A typical month of 30 days has 1440 data, which is a sufficient number for the robust computation of the transfer entropy using these methods. However, because of instrument malfunctions and the circumstances of weather, substantial gaps exist in the data record for certain months. Any results computed for months where fewer than 850 data points are available are dropped from the analysis.

### 3.2. Subsystems and Time Scales That Characterize the System

[29] The first challenge is to validate the network statistics discussed in section 2 by comparing them with the more detailed process network analysis presented by *Ruddell and Kumar* [2009, Figure 7]. Network statistics cannot provide the same level of detail as the detailed process network analysis, but they should be able to capture the pattern of feedback, time scale, and subsystem organization.

[30] *Ruddell and Kumar* [2009] identified three subsystems for the month of July 2003, which is the peak of the growing season and the warmest month of the year. The turbulent subsystem includes the sensible heat flux ($\gamma_H$), latent heat flux ($\gamma_{LE}$), and gross ecosystem carbon production (GEP) variables and is characterized by substantial feedback at the <30 min time scale associated with turbulent mixing processes on the land surface [see *Katul et al.*, 2001]. The atmospheric boundary layer (ABL) subsystem includes the global radiation ($R_g$) and precipitation ($P$) variables that are coupled to each other at time scales from 3 to 18 h, with coupling peaking at 12 h. This subsystem is associated with ABL formation processes on the subdaily time scale [*Juang et al.*, 2007]. The synoptic subsystem includes the air temperature ($\Theta_a$), vapor pressure deficit (VPD), soil temperature ($\Theta_s$), soil water content ($\theta$), and gross ecosystem respiration (GER) variables which force the rest of the network on very short >30 min time scales. These variables are associated with continental-scale weather patterns [*Baldocchi et al.*, 2001a]. An additional regional subsystem exists because of the subdaily time scale feedback of information between the ABL and turbulent subsystems and is the hierarchical aggregate of these two subsystems.

[31] To identify the relevant time scales, the feedback index $R''$ is computed for the turbulent subsystem, regional subsystem, and the whole system at time scales from 30 min to 18 h, for the month of July, for the years 1998–2006. Figure 2a illustrates the mean of the nine July values of $R''$ for each time scale. The turbulent subsystem is feedback dominated ($R'' < 0$) at the same 30 min time scale identified by *Ruddell and Kumar* [2009, Figure 7]. The regional subsystem, which is a coupled hierarchical aggregation of the turbulent subsystem and the ABL subsystem, shows a local minimum of $R''$ (where peak feedback occurs) in the vicinity of the 14 h time scale, which is similar to the 12 h time scale identified for the regional hierarchical subsystem by *Ruddell and Kumar* [2009, Figure 7]. This 14 h local minimum is less pronounced than the 30 min minimum of the turbulent subsystem, but it is significant because it lies more than one standard deviation below the global maximum (at 2 h) of $R''$ for the regional subsystem. These results demonstrate that the network statistic $R''$ is able to capture the same time scale dynamics that were identified using a more detailed process network analysis, but with an additional capability to do interseasonal comparisons of dynamics (see section 3.3). 30 min is used in the rest of the paper as the characteristic time scale of turbulent subsystem processes, and 14 h is used as the characteristic time scale of regional and ABL processes.

[32] The mean annual pattern in values of $R''$ is plotted in Figure 2b, which shows that the regional subsystem has $R'' > 0$ during all months, but by contrast the turbulent subsystem at the 30 min time scale is feedback dominated

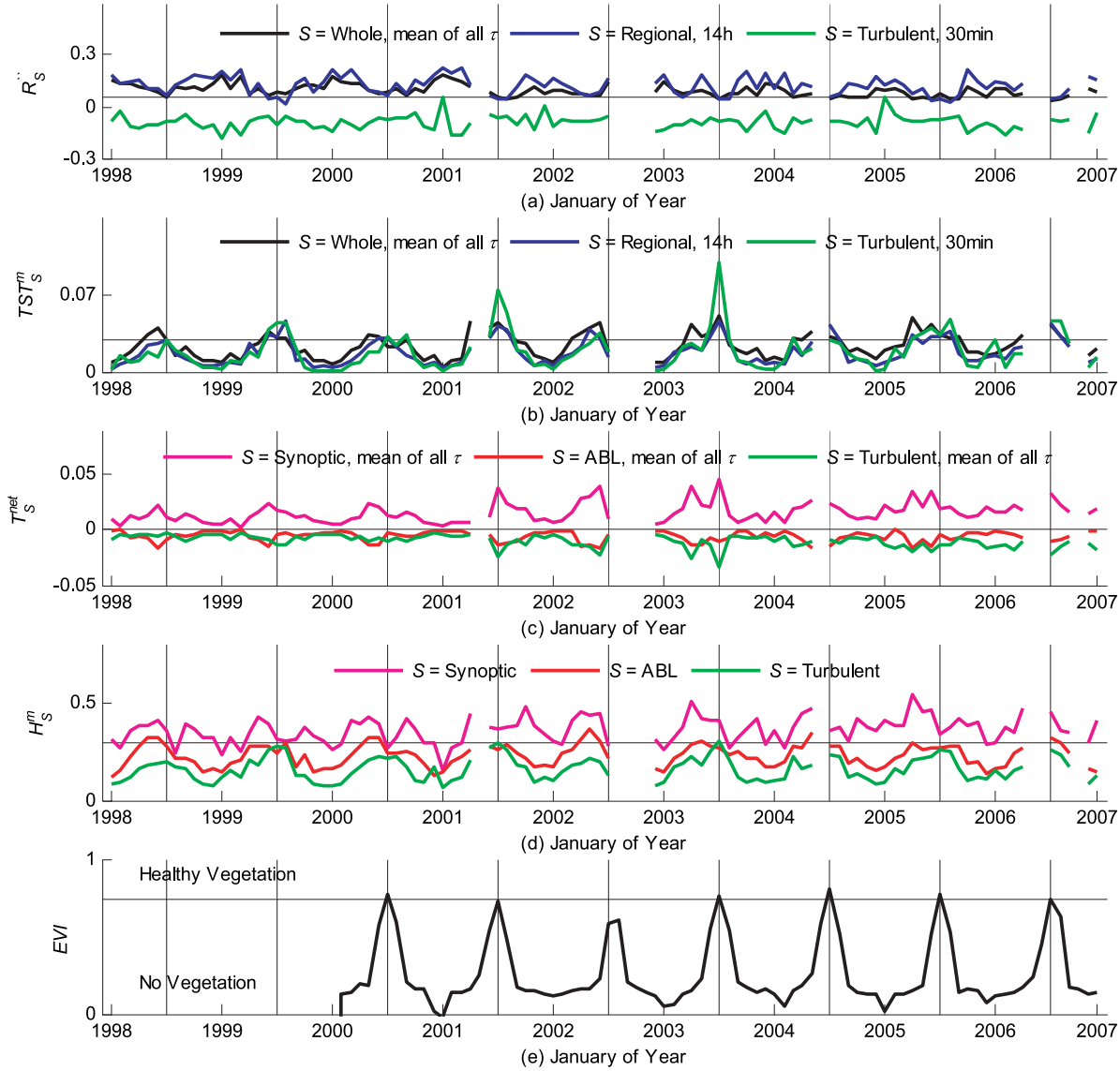


**Figure 3.** Process network statistics plotted for comparison over 9 years, 1998–2006. Vertical lines indicate July; ticks indicate January. Gaps indicate that fewer than 850 data are available in that month, so results are discarded. (a) Feedback index $R''$ for the regional and turbulent subsystem and the whole system; (b) mean total system transport $TST^m$ for the regional and turbulent subsystems and the whole system; (c) net information production $T^{net}$ for the synoptic, ABL, and turbulent subsystems; (d) mean Shannon entropy $H^m$ for the synoptic, ABL, and turbulent subsystems; and (e) enhanced vegetation index EVI, collection 4 MODIS subset for Bondville site (http://www.modis.ornl.gov/modis/index.cfm).

($R'' < 0$). The feedback index $R''$ shows an interesting seasonal trend, in that the regional and turbulent subsystems both move toward the neutral state ($R'' = 0$) during the summer. A physical interpretation of this pattern may be that more moisture and energy feedback occurs in the system as a whole during peak summer growing conditions, as interaction between the land surface and atmosphere increases. However, at the land surface, the strong dominance of the ecosystem processes over the network's dynamics causes a decrease in apparent feedback in the turbulent subsystem, during the peak summer growing season. These interpretations are evaluated in section 3.3 below, in which patterns explaining this seasonal variability of $R''$ are analyzed.

### 3.3. Seasonal Patterns in Network Statistics

[33] This section explores the seasonal and annual patterns in system organization, identifies distinct system states, and explains the occurrence of emergent structures during some system states. $R''$, $TST^m$, $T^{net}$, and $H^m$ are computed for all months; the results are plotted in Figure 3 along with the MODIS-derived enhanced vegetation index (EVI) for the Bondville site. Broad seasonal patterns are evident: $H^m$, TST, and the absolute value of $T^{net}$ are highest in the summer months and lowest in the winter months for

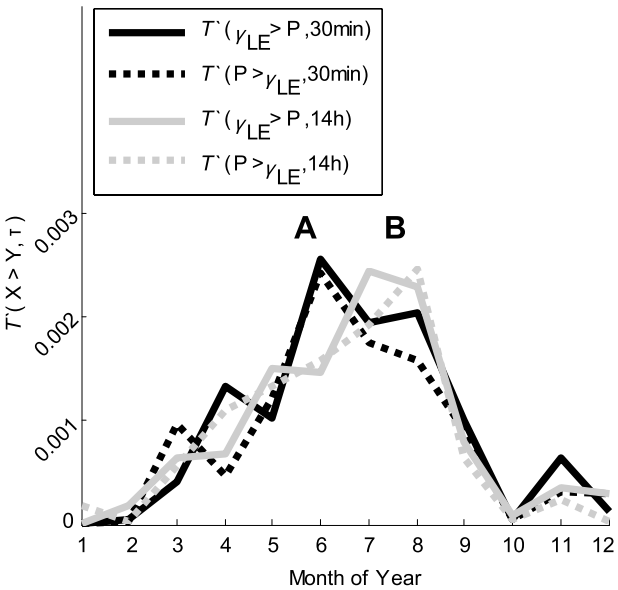


**Figure 4.** Mean annual pattern in the strength $T'$ of a key feedback coupling between the atmospheric boundary layer and turbulent subsystems, that of latent heat $\gamma_{LE}$ and precipitation $P$, plotted for 30 min and 14 h time scales. Winter and summer are distinguished because summer has far more information flow than winter. Two summer substates are apparent, with "A" peaking in June and "B" peaking in July and August.

all subsystems. This means that total Shannon entropy and information production in the system are higher during the warm months than during the cold months. Specifically, $H^m$ and $TST^m$ peak on average in the month of June, but $R''$ and $T^{net}$ peak on average in the month of July along with the EVI.

[34] The synoptic subsystem has the highest Shannon entropy $H^m$, and the turbulent and ABL subsystems have lower Shannon entropy. This means that there is more variability and uncertainty associated with the synoptic variables, as compared with the turbulent and ABL variables. From the plot of $T^{net}$, observe that the synoptic subsystem is a net producer of information, and the turbulent and ABL subsystems are net consumers of information. When two variables are synchronized but one drives the other, the driver is a net exporter of information. In this sense, the synoptic (weather-related) subsystem is the dominant controller of the system. This finding is consistent with the observation of *Ruddell and Kumar* [2009] that the synoptic subsystem is a source of information to the other two subsystems. Subsystems that are net exporters of information appear to be those with higher Shannon entropy; more on this is discussed in section 3.4.

[35] Compare Figure 2b with Figure 3a. Both show that the regional subsystem and the whole system are source-sink dominated ($R'' > 0$) during all months and time scales, but the turbulent subsystem at the <30 min time scale is feedback dominated ($R'' < 0$). The feedback index $R''$ shows an interesting seasonal trend, in that the regional and turbulent subsystems both move toward the neutral state ($R'' = 0$) during the summer (Figure 2b). This pattern may be interpreted to indicate that the feedback in the regional subsystem increases during the summer, while that in the turbulent subsystem decreases. What causes this increase in 14 h time scale feedback in the regional subsystem, and a decrease in <30 min time scale feedback in the turbulent subsystem, centered on the month of July?

[36] To answer this question, examine the feedback coupling between the turbulent and ABL subsystems which causes the emergence of a regional subsystem. *Ruddell and Kumar* [2009] demonstrated that the coupling between $\gamma_{LE}$ and $P$ at the subdaily time scales is a key feedback coupling that links the turbulent and ABL subsystems. Figure 4 plots the mean annual pattern in the coupling strength $T'$ between latent heat flux ($\gamma_{LE}$) and precipitation ($P$) for time scales of 30 min and 14 h. The 30 min feedback coupling peaks in June, while the 14 h feedback coupling associated with the emergence of the regional subsystem peaks during July and August. This plot demonstrates that one coupling can have different feedback organization at different time scales during different system states. As shown in Figure 4, multiple warm weather substates exist: an early summer April–June substate, which is termed "summer A," where short time scale localized processes are dominant, and a mid summer June–August substate, which is termed "summer B," where regional feedback processes mediated by the ABL at 14 h time scale are dominant.

[37] Attention is now shifted to the turbulent subsystem, to learn which variable in that subsystem is dominant during the "summer B" substate. Figure 5 plots the mean annual pattern in the gross information production and consumption at a 30 min time scale for sensible heat flux ($\gamma_H$), latent heat flux ($\gamma_{LE}$), and gross ecosystem carbon uptake (GEP) that comprise the turbulent subsystem. Heat fluxes $\gamma_H$ and $\gamma_{LE}$ are net information consumers since $T^{[-]}$ is greater than

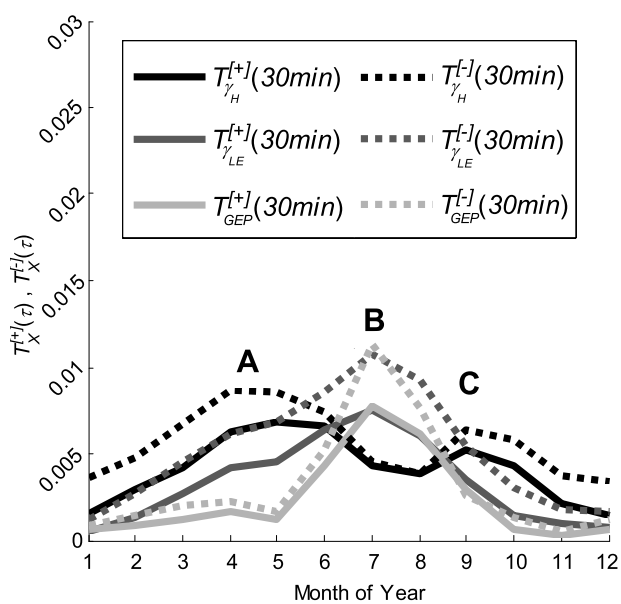


**Figure 5.** Mean annual pattern (for 9 years, 1998–2006) of gross information production $T^{[+]}$ and consumption $T^{[-]}$ for the turbulent subsystem variables sensible heat flux $\gamma_H$, latent heat flux $\gamma_{LE}$, and gross ecosystem carbon uptake GEP. Three summer substates, "A," "B," and "C" are indicated. GEP and $\gamma_{LE}$ become dominant during summer B.

$T^{[+]}$, meaning that they receive more predictive information from the network than they export to the network. There is also a complementary relationship between them such that when $T^{[+]}_{\gamma_{LE}}$ increases, $T^{[+]}_{\gamma_H}$ decreases and vice versa; $T^{[+]}_{\gamma_H}$ peaks in April through June, then gives way to $T^{[+]}_{\gamma_{LE}}$ in June through August. The July peak in $\gamma_{LE}$ information consumption corresponds to a dramatic increase in gross information export by *GEP* (Figure 5). An interpretation is that the turbulent subsystem is controlled by information export from GEP during summer B, and this explains the reduced system feedback observed in summer in Figure 3.

[38] The July peak in $\gamma_{LE}$ activity (Figure 5) corresponds to the increase in 14 h time scale feedback in the regional subsystem shown in Figure 4; *GEP* is the source of much of the information flowing to the turbulent subsystem (at 30 min time scales) and the ABL subsystem (14 h times scales, via $\gamma_{LE}$) during the peak growing months. A third warm weather state, resembling "summer A," appears in September after the peak of the growing season; this late summer substate is termed "summer C." "Summer C" is different from "summer A" in that little moisture recycling occurs between the latent heat flux and precipitation variables during "summer C," at any time scale (recall Figure 4). Figure 5 adds to the understanding of the "summer A" and "summer B" substates. "Summer A," peaking in April–June, is characterized by strong short time scale sensible heat flux processes on the land surface, driven by the synoptic weather patterns (especially air temperature). "Summer B," peaking in July, is characterized by strong latent heat fluxes and strong feedback between the land surface and ABL, and by strong information export and coupling strength from GEP and $\gamma_{LE}$. In April and May ("summer A") temperatures are no longer below the freezing point, and the fields are being cultivated, but no meaningful vegetation cover has emerged. The planting of the crops by farmers peaks the first week of May, but no significant biomass appears until June. June through August drought during this phase can delay the development of crops, weakening their later yield potentials. "Summer B" corresponds to the season of peak carbon and nutrient assimilation, biomass growth, and transpiration for both the corn and the soybean plant [*Hanway*, 1966; *Hanway and Thompson*, 1967]. The peak of the growth occurs in July and early August. Drought during one critical phase of growth, the "tasseling stage," can disrupt pollination and seed filling, crippling the crop. By the start of September most of the growth is complete and the crops are entering the drying stage, in preparation for harvest, which occurs from mid-September through October ("summer C").

[39] Much of the information flow which forms the emergent regional feedback structure during the "summer B" substate originates in the ecosystem photosynthetic-evapotranspirative processes. This finding is consistent with the findings of *Juang et al.* [2007] that plants have a substantial effect on local ABL conditions, including cloud cover and precipitation, via modification of $\gamma_H$ and $\gamma_{LE}$. Furthermore, "summer B" corresponds to the season of peak carbon and nutrient assimilation and biomass growth for both the corn and the soybean plant [*Hanway*, 1966; *Hanway and Thompson*, 1967].

[40] The gross information production of all ten variables for each month 1998–2006 is plotted in Figure 6a to allow evaluation of the seasonal patterns in total coupling strength and predictive information for each variable. It is apparent that all variables show greater information production in the summer when temperatures and energy fluxes are much higher, except for $P$ and $\theta$ which show little seasonal variability. The strongest gross information producers in Figure 6a are $R_g$, *VPD*, $\Theta_a$, and *GEP*, in that order. Information production by $\Theta_a$ and $R_g$ is strong during the entire year, but peaks during the "summer A" substate early in the growing season. During the "summer B" substate VPD and GEP are also strong producers of information. Even when compared with the synoptic subsystem's variables, the carbon uptake variable GEP emerges as a strong gross producer of information.

[41] In feedback-based complex systems, two variables may appear to be strongly coupled and synchronized with each other (showing a high gross information production), but the source of the apparent control may be an indirect "third-party" coupling [*Kurths et al.*, 2003; *Jorgensen et al.*, 2007, p. 85]. In other words, a variable may recycle information it has received from another variable on the network, or it may trade equal amounts of information with a tightly coupled variable with which it participates in feedback. However, when the net information production is computed, this recycled information is canceled on average, leaving only that information for which the variable is the original source. The net information production of each variable for each month 1998–2006 is plotted in Figure 6b. In general, synoptic variables are net producers (sources) and turbulent and ABL variables are net consumers (sinks) of information, in keeping with Figure 3 and earlier findings [*Ruddell and Kumar*, 2009]. GER, $P$, and $\theta$ are relatively weak participants in the system's information flows. The strongest net information exporters are $\Theta_a$ and GEP, with $\Theta_a$ net export peaking during "summer A" (April–June) and GEP net export peaking during "summer B" (especially July, when EVI also peaks). This result means that the "summer A" subsystem is controlled most strongly by information flow from the air temperature (which is controlled by synoptic weather patterns rather than localized feedback), but the "summer B" subsystem is controlled by information flow from the ecosystem photosynthetic processes (GEP). During "summer B," both the synoptic subsystem (largest scale) and GEP (smallest scale) are sources of information and the regional feedback subsystem (midscale) is receiving information from both higher and lower scales.

[42] Strong net information flow from GEP (a short time scale variable [*Baldocchi et al.*, 2001a]) to the regional subsystem (which operates on a 14 h feedback time scale), via modification of $\gamma_H$ and $\gamma_{LE}$, is a good example of "bottom-up" emergence [*Hubler*, 2005]. Bottom-up emergence is identified by a flow of order which originates at the smallest scale in a system, but which ends up defining order at larger scales. Bottom-up emergence is easy to identify because it stands out in contrast to the more common situation where order originates at the largest scale in the system hierarchy (the largest scale in this ecohydrologic system is the synoptic or weather-related subsystem). During the "summer B" substate (July), it appears that order is flowing both from the bottom up (from the ecosystem carbon uptake variable GEP) and from the

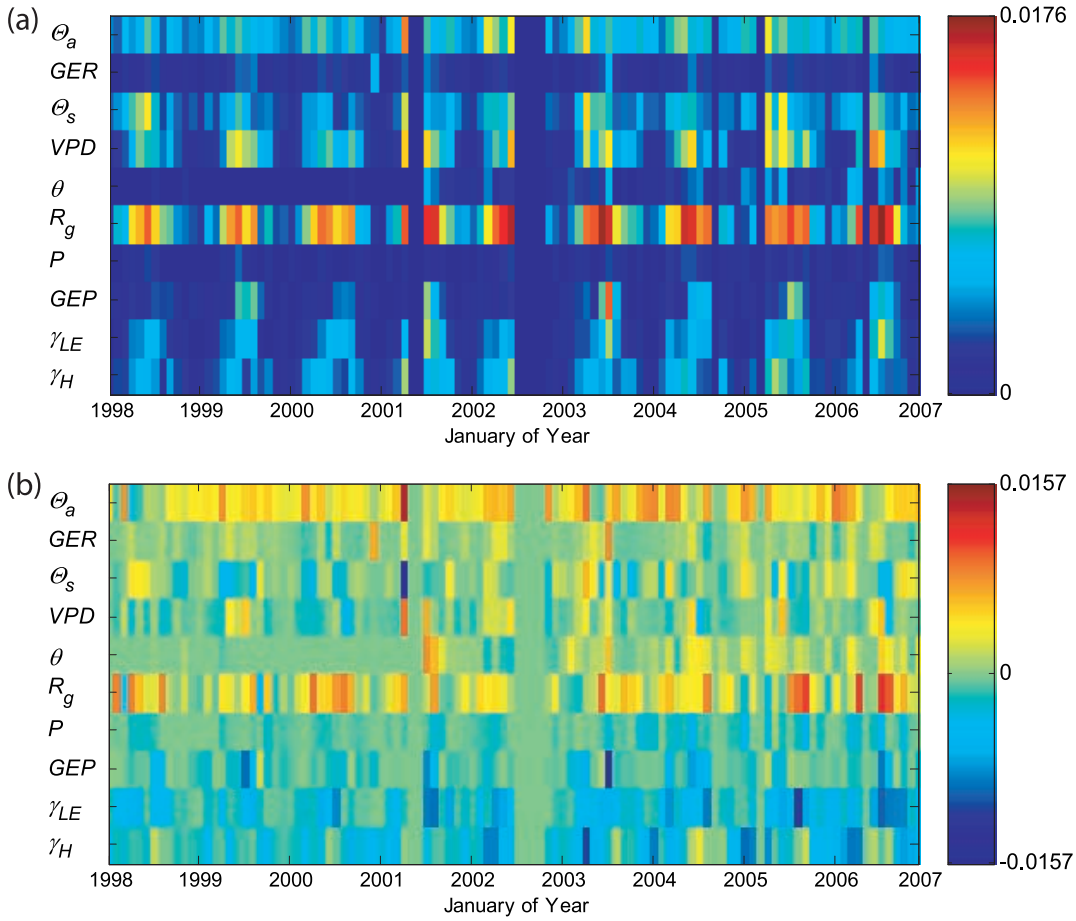


**Figure 6.** (a) Mean gross information production $T^{[+]}$ and (b) net information production $T^{net}$, for all 10 variables, every month for 1998–2006 at the Bondville site. Computations are performed at the characteristic time scale of each subsystem: 14 h for the atmospheric boundary layer subsystem (shortwave radiation $R_g$ and precipitation $P$) and 30 min for the synoptic and turbulent subsystems (all other variables).

top down (from the weather-related synoptic subsystem, especially $\Theta_a$).

### 3.4. Emergent Properties of the System: Control and Order Parameters

[43] It is evident from the above analysis that the Bondville ecohydrological system's process network feedback structure changes with the seasons. While the radiation is an important driver all the time, and the air temperature seems to be an important driver during the early summer, the photosynthetic process in the cultivated crop ecosystem becomes an important driver during the peak of the growing season in July.

[44] Ecohydrologic systems are complex systems, in that they comprise many parts linked by cycles of feedback [*Kumar*, 2007]. Feedback between multiple interacting parts can give rise to "emergent" organized structures (system structures which result from a feedback interaction of multiple system parts, rather than from the direct action of a single part [*Corning*, 2002]). A metric which directly measures the level of emergent organization resulting from feedback is called an "order parameter" [*Haken*, 1979]. Because the mechanism which enables self-organization and emergence is feedback, the feedback index $R''$ is a reasonable order parameter for this system (Figure 2). The information flow $T'$ between precipitation ($P$) and latent heat flux ($\gamma_{LE}$) measures the strength of the feedback which gives rise to the emergent regional subsystem that forms during "summer B" (Figure 4) and, therefore, may be considered as another order parameter. The gross production of information $T^{[+]}$ by the gross ecosystem carbon uptake ($GEP$) measures the strength of the emergent "bottom-up" flow of control during "summer B" (Figure 5) and is therefore a third-order parameter.

[45] If $R''(S, \tau)$, $T'(P > \gamma_{LE}$, 14 h), and $T^{[+]}_{GEP}$(0.5 h) are order parameters, what are the independent "control parameters" which drive the emergence of order on the process network? In complex systems, the relevant control parameter is the throughput [*Hubler*, 2005]. When the control parameter measuring throughput exceeds a certain threshold (or instability point), ordered structures spontaneously emerge in the system [*Haken*, 1988; *Nicolis and Prigogine*, 1989]. Examples of throughput include heat flow, fluid flow, information flow, or entropy production [*Hubler*, 2005]. What are the relevant throughputs in this ecohydrological system, and furthermore, are there two types of throughputs, separately controlling the physical and statistical aspects of the system?

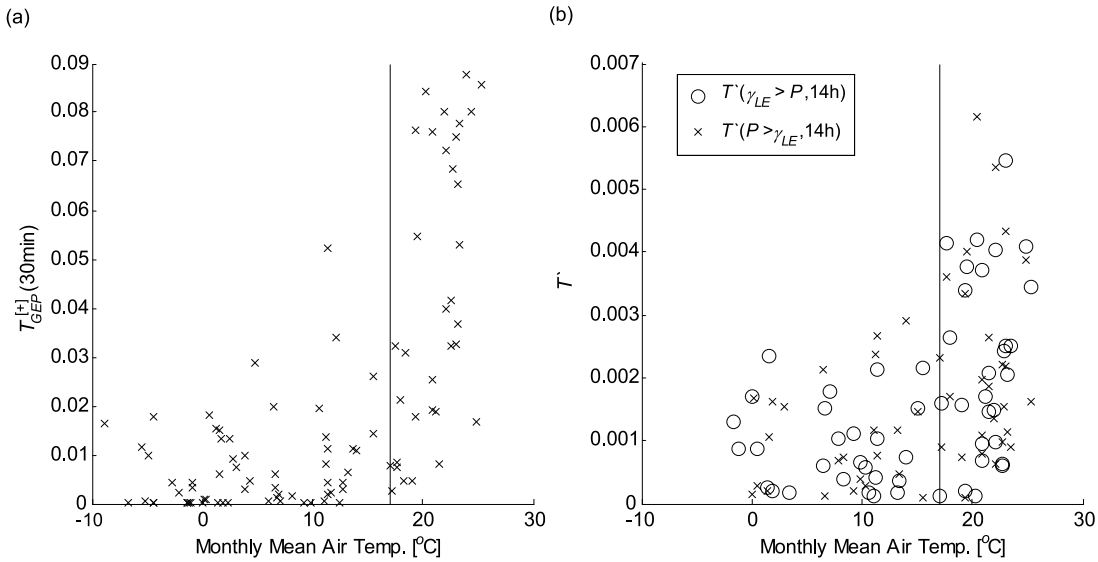


**Figure 7.** (a) Mean gross information production of the GEP variable at a 30 min time scale and (b) transfer entropy of the coupling between the $\gamma_{LE}$ and $P$ variables at a 14 h time scale (regional time scale) plotted against the monthly mean air temperature. A threshold exists near 17°C such that information flow can be much greater during the months of June, July, and August than during other months.

[46] *Ozawa et al.* [2003] found that the energy dissipation and thermodynamic entropy production in the atmospheric portion of the climate system is governed by the throughput of heat and momentum, because the amount of heat and momentum in the system, rather than the radiative input of energy, governs the turbulent fluid transport properties which dominate energy transport in the climate. The mean air temperature therefore provides a first-order measure of the heat in the atmosphere near the land surface, and is a reasonable choice for a control parameter which approximates the thermodynamic energy throughput of this ecohydrological system. Figure 7 plots two-order parameters, the gross information production of ecosystem carbon uptake GEP and the information feedback between the latent heat flux $\gamma_{LE}$ and precipitation $P$, against the control parameter, monthly mean air temperature $\Theta_a$. Figure 7 makes it clear that there is a strong and positive relationship between the energy throughput of the system, the gross information production of the ecosystem and the information flow and feedback in the regional-scale moisture feedback process.

[47] An apparent threshold exists in Figure 7a at approximately 17°C, which is the average air temperature exceeded during the "summer B" substate (section 3.3) months of June, July, and August. Strong gross and net production of information by the ecosystem, and increased feedback of information in the system at the regional scale, both emerge above this threshold in mean air temperature. Likewise, the $R''$ statistic for the regional subsystem shows increased feedback above the 17° threshold in Figure 2b. On the basis of the physical arguments of *Ozawa et al.* [2003] and this empirical evidence, the authors suggest that mean air temperature (which approximates atmospheric heat and momentum) is an important physical control parameter driving the emergence of information feedback-based self-organization in this ecohydrological system. Furthermore, there is clearly a link between the energy throughput of the system (approximated on average by the air temperature) and the production and feedback of statistical information in the system. More energy throughput occurs when more Shannon entropy and more statistical information are being produced.

[48] However, the mean air temperature is not the only control parameter. The Shannon entropy (measuring uncertainty in the variables' values) is, like the air temperature, an independent property of the system which varies from season to season (recall Figure 3). Is there a second layer of emergent behavior which lies latent in the information statistics themselves, independent from the physical processes which the statistics are used to measure? The answer is "yes," and the results may be important for a wide range of complex systems. The mean total information flow of the whole system, $TST^m_V$, is plotted against the mean Shannon entropy of the whole system, $H^m_V$, in Figure 8. It is clear from Figure 8 that there is a power law relationship (with a strong 89% $R^2$ fit) between the average production of information and the average production of Shannon entropy in this system. The Shannon entropy controls the production of information in the system.

[49] Interestingly, the total production of Shannon entropy, $H^m_V$, and the total flow of information, $TST^m_V$, do not peak in July during the summer B substate, but rather during the April–June summer A substate (recall Figure 3). Maximum temperature, ecosystem activity measured by GEP and EVI, and regional feedback measured by $R''$, all peak in July. These metrics are one month out of phase with the maximum Shannon entropy and the information flow (and also the input of solar radiation $R_g$; compare with *Jorgensen et al.* [2007, p. 132]. In other words, there is a physical control parameter (energy throughput approximated as air temperature) and a statistical control parameter (throughput of stochastic variability, measured as Shannon entropy), and the two are operating slightly out of phase in the Bondville system. This begs a question for future work, as to whether all ecohydrological systems behave this way.

[50] To investigate the relationship between $H$ and $T$ in greater detail, it is necessary to shift the point of view

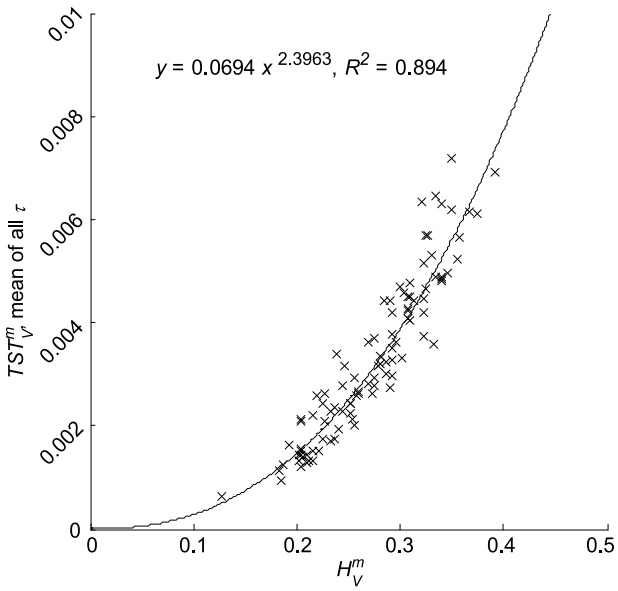


**Figure 8.** Normalized mean total system transport $TST_V^m$ of the whole system $V$, averaged across all time lags from 30 min to 18 h, plotted against the mean Shannon entropy $H_V{}^m$ of the whole system $V$, for each month in the years 1998–2006. $TST_V^m$ scales as a power law of $H_V^m$.

[*Gershenson and Heylighen*, 2003] from the "global" perspective that uses the same discrete state space to embed the values of all 108 months, to a "local" perspective, which embeds the observed values of each month into a discrete state space that is relative to that particular month (recall section 2.1). This shift in perspective allows consideration of the relationship between $H'$ and $T'$ independently from the confounding effects of the seasonal patterns in the physical system and the physical control parameter. Figure 9 shows a box plot of the range of Shannon entropy values assumed by each variable and scatterplots (Figure 9a), net information production versus Shannon entropy (Figure 9b), gross information production versus Shannon entropy (Figure 9c), and gross information consumption versus Shannon entropy (Figure 9d), for all ten variables in all months 1998–2006.

[51] In Figure 9a the turbulent subsystem variables $\gamma_H$, $\gamma_{LE}$, and GEP have a moderate Shannon entropy $0.4 < H'_X < 0.7$, synoptic subsystem variables $\Theta_a$, VPD, $\Theta_s$, $\theta$, and GER have large Shannon entropy $H'_X > 0.7$, and ABL subsystem variables $P$ and $R_g$ span moderate to low Shannon entropies. In Figure 9b, variables with Shannon entropy greater than 0.7 (synoptic subsystem variables) tend to be net producers (sources), those with less than 0.4 (ABL variables, especially $P$) tend to be net consumers (sinks), and those in between (turbulent variables) can be net producers or consumers depending on the month (*GEP* is usually a net producer, $\gamma_H$ and $\gamma_{LE}$ are usually net consumers). In Figure 9c, higher Shannon entropies are associated with higher gross information production. This simple relationship is not surprising, because it was demonstrated in section 2 that the transfer entropy is bounded as a function of the Shannon entropy of the source variable. However, Figure 9d shows the surprising result that variables with high Shannon entropies $H'_X > 0.7$ consume much less information than those where $0.4 < H'_X < 0.7$. This results in an imbalance in information production and consumption for variables above the 0.7 threshold.

[52] The variability measured by Shannon entropy is a control parameter that determines the organizational structure of the process network at the statistical level. Information flow, as measured by transfer entropy, can be considered as an order parameter. Information tends to flow from high Shannon entropy variables to low Shannon entropy variables, and only the moderate Shannon entropy variables participate in feedback. The control parameter $H'$ features two critical thresholds at approximately 0.4 and 0.7, and emergent behaviors are associated with variables in the moderate range between these thresholds. The hierarchy of subsystems observed in the ecohydrological system can be explained in relation to these thresholds: high Shannon entropy synoptic variables drive the moderate and low Shannon entropy variables, while moderate Shannon entropy variables (turbulent subsystem) form feedback-based subsystems. These approximate thresholds in the control parameter $H'$ hold for all subsystems, time scales, and months observed in this ecohydrological system, and as such appear to be independent of any specific physical process in the system.

[53] These results demonstrate that the transfer entropy $T'$, feedback index $R''$ and gross information production $T^{[+]}$ serve as order parameters where as the air temperature $\Theta_a$ and Shannon entropy $H'$ serve as control parameters for the ecohydrologic system studied.

## 4. Summary and Conclusions

[54] It has been demonstrated that the network statistics $H'$, $R''$, $T^{[+]}$, $T^{[-]}$, and $T^{net}$ may be used to robustly characterize the subsystems, time scales, and feedback structures that exist in a process network based on transfer-entropy derived information flow couplings. The network statistics are shown to reveal similar patterns of organization as compared with a more complete and detailed process network analysis performed by *Ruddell and Kumar* [2009]. They are used to study both the ecohydrological and the emergent properties of the system.

[55] The network statistics are used to identify two primary ecohydrological system states, summer and winter. In agreement with previous work on ecosystem networks [*Jorgensen et al.*, 2007], connectivity and information flow in the process network are found to increase dramatically during summer. A more detailed analysis of the turbulent land surface subsystem reveals three summer substates. The dynamics of the first summer substate ("summer A," April–June, peaking in June) are dominated by information export from weather-related synoptic subsystem variables (especially air temperature) at <30 min time scales. The dynamics of the second summer substate ("summer B," June–August, peaking in July) are dominated by the emergence of regional time scale (~14 h) feedback between the atmospheric boundary layer (ABL) and turbulent subsystems, which is indirectly controlled by the ecosystem photosynthetic process via the strong control of stomatal transpiration over latent heat during July. This substate is where emergent properties appear, such as self-organization in the regional subsystem and strong coupling between ecosystem and atmosphere controlled by the gross ecosystem production (GEP) variable via latent heat flux ($\gamma_{LE}$).

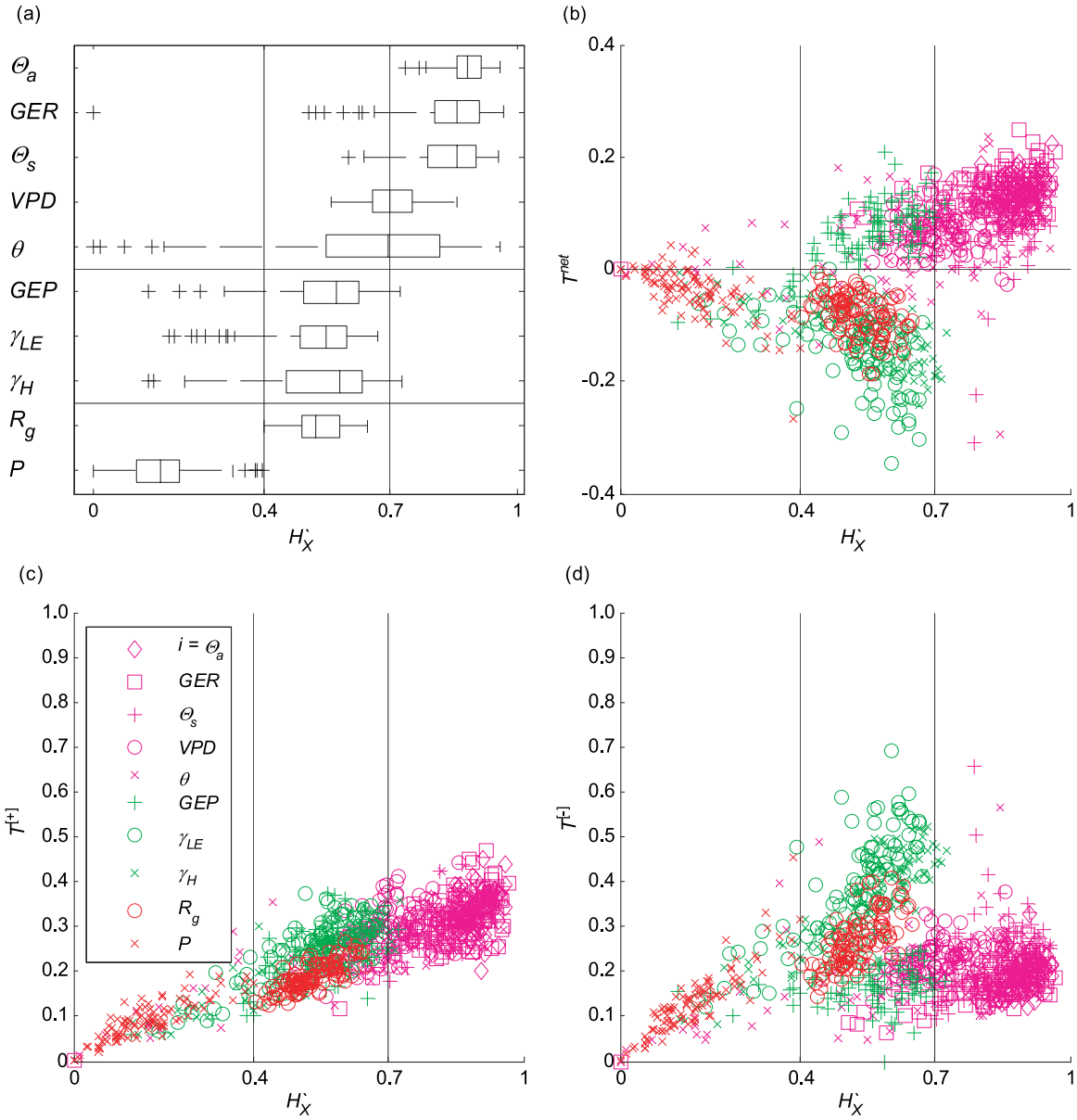


**Figure 9.** Distribution of information transport values compared with the normalized Shannon entropy $H'_X$ for all months in 1998 through 2006. (a) Box plot of 10 variables; (b) net information export $T^{net}$, plotted at the variable $X$'s characteristic time scale, 30 min or 14 h; (c) gross information export $T^{[+]}$; and (d) gross information import $T^{[-]}$.

The third summer subsystem ("summer C," September–October) is like "summer A," except that short time scale coupling between the turbulent and ABL subsystems is weaker.

[56] Air temperature ($\Theta_a$) and radiation ($R_g$) are the strongest net producers of information on the network; the net production of information by $\Theta_a$ peaks during summer A, but the net production of information by $R_g$ exceeds that of air temperature during July, at the peak of summer B, when latent heat flux ($\gamma_{LE}$) and gross ecosystem production (GEP) dominate the land-atmosphere coupling. This means that there are two primary sources of order and control of the Bondville ecohydrological system's states: weather-related synoptic variables (especially $\Theta_a$) during all seasons, and the radiative-photosynthetic-evapotranspirative [*Ball and Berry*, 1987] ecosystem activity during summer B. The ABL subsystem (incoming shortwave radiation ($R_g$) and precipitation ($P$)) and the surface energy fluxes (sensible heat flux ($\gamma_H$) and latent heat flux ($\gamma_{LE}$)) serve as a sink of and a medium for transmission of information from these two sources [*Kumar*, 2007]. These results show that information flows both from the "top down" and from the "bottom up," as is expected from emergent structures in complex systems [*Hubler*, 2005]. The most significant finding about this ecohydrological system is that plants are not passive recipients of information and order from their environment, but rather are themselves a large producer of information during July, acting collectively via feedback to control the land surface energy balance and ABL at the regional scale via latent heat flux modification.

[57] Following the arguments of *Ozawa et al.* [2003] that the throughput of heat and momentum in the atmosphere are

the relevant physical controls on the atmospheric system, evidence demonstrates that the mean air temperature (a first-order approximation of atmospheric heat near the land surface) is the control parameter which governs the emergence of regional-scale feedback and "bottom-up" information flow from the ecosystem. Above a threshold of approximately 17°C, the emergent behaviors associated with the "summer B" substate begin to appear in the Illinois corn-soybean ecosystem, as measured by marked increases in key information flow order parameters. This 17° threshold begs a question for future work, as to whether all ecohydrological systems are adapted to follow the same threshold relationship controlling the information production response.

[58] The information production and consumption of a variable is also related to the variable's Shannon entropy. Low Shannon entropy variables serve as net consumers of information, high Shannon entropy variables serve as net producers of information, and moderate Shannon entropy variables can both consume and produce information. It is predominantly moderate Shannon entropy variables ($R_g$, $\gamma_H$, $\gamma_{LE}$, GEP) that participate in feedback, and the information flow couplings of moderate Shannon entropy variables tend to be of the second type (type 2) identified by *Ruddell and Kumar* [2009]. Moderate Shannon entropy variables are those with a normalized Shannon entropy between approximately 0.4 and 0.7. A possible interpretation is that moderate Shannon entropy variables in ecohydrological systems may be operating at the proverbial "edge of chaos" [*Kaufmann*, 1993], organizing themselves via feedback into states which achieve maximum throughput (of energy, information, etc.), and allowing as much variability (Shannon entropy) as possible without lapsing into chaos (high Shannon entropy).

[59] The authors conclude that there are at least two control parameters in the Bondville ecohydrological system, which have a positive relationship to the system' information production (information measures organization, order, and predictability in the system). The first control parameter is physical energy throughput (approximated by mean air temperature), and the second is the stochastic variability (measured using mean Shannon entropy). Of the two control parameters, the mean Shannon entropy has a stronger and cleaner relationship to system information production. Does this mean that the process networks are not realistically measuring the structure of the complex ecohydrological system, but rather are predetermined by a simple Shannon entropy parameter? The authors think not, because previous work has validated the physical accuracy of the process networks [*Ruddell and Kumar*, 2009]. Rather, the authors believe that these results cast light on the general structure of complex systems. The authors hypothesize that, for complex open dissipative systems, the level of statistical Shannon entropy and information production of a subsystem is closely related to and inseparable from its physical function in the hierarchical structure of the system. This hypothesis applies in the context of a scale of reference, related to the data set time scale $r$, which defines which level of the system's hierarchy that is experimentally observed.

[60] Readers of part 1 of this paper, *Ruddell and Kumar* [2009], may at this point ask as to whether the network statistics of information production and feedback can clearly resolve the pattern of drought that impacted the Bondville, Illinois site during the spring and summer of 2005. The answer so far is negative. No clear pattern in Figure 3 appears to separate information production, feedback statistics, or Shannon entropy of the system during the drought-afflicted 2005 year from those during more well watered years such as 2003 and 2004. This is perhaps surprising, given that *Ruddell and Kumar* [2009] found a clear difference between 2003 and 2005 in terms of the system's regional-scale moisture recycling feedback on the process network. However, it appears that the process network's average structure, as measured by several network statistics, is not strongly affected by drought. The Shannon entropy and feedback index $R''$ of July 2005 are slightly lower than July 2003 or 2004, but one sample is not enough to draw conclusions. If many more drought-afflicted summer data sets could be analyzed, it might be possible to say more.

[61] In summary, this paper provides a conceptual approach and a working set of statistical tools for the analysis of process networks, identifies characteristic emergent behaviors of a Midwestern corn-soybean ecohydrological system, and demonstrates that the order parameters and control parameters which characterize this complex ecohydrological system can be quantified empirically by analyzing the process networks. The generality of the results across different ecosystems is being investigated through the analysis of other FLUXNET sites.

## Notation

| | |
|---|---|
| $\gamma_H$ | sensible heat flux [W m$^{-2}$]. |
| $\gamma_{LE}$ | latent heat flux [W m$^{-2}$]. |
| $\Delta t$, $dt$ | discrete interval of time, the units of time lags and steps [T]. |
| $\Theta$ | soil water content of the surface layer [m$^3$ m$^{-3}$]. |
| $\Theta_a$ | air temperature [°C or K]. |
| $\Theta_s$ | soil temperature (surface layer) [°C or K]. |
| $\tau$ | time lag between variables $X_t$ and $Y_t$ [$\Delta t$]. |
| $\tau'$ | characteristic time lag of the coupling between two variables [$\Delta t$]. |
| $\tau_{\max}$ | time lag of maximum information flow between two variables [$\Delta t$]. |
| $\omega$ | number of time lags skipped for variable $Y_t$'s own history [$\Delta t$]. |
| $\mathbf{A}(j, i, \tau)$ | network adjacency matrix where indices store $T'$ [arbitrary units]. |
| GEP | estimated gross ecosystem production [$\mu$mol $CO_2$ m$^{-2}$ s$^{-1}$]. |
| GER | estimated gross ecosystem respiration [$\mu$mol $CO_2$ m$^{-2}$ s$^{-1}$]. |
| $\mathbf{H}$ | vector storing values of $H'$ for each variable [fraction]. |
| $H'$ | normalized Shannon entropy [fraction]. |
| $H_S^{[i\|j]}(\tau)$ | source-conditional network Shannon entropy of $S$ [fraction]. |
| $H_S^{[j\|i]}(\tau)$ | sink-conditional network Shannon entropy of $S$ [fraction]. |
| $H_S^m$ | mean normalized Shannon entropy of subsystem $S$ [fraction]. |
| $H(X_t)$ or $H_{X_t}$ | Shannon entropy of variable $X_t$ [bits]. |

$i, j$ matrix indices for $X$ and $Y$ [positive integer].
$I'$ normalized mutual information [fraction].
$I(X_t, Y_t)$ mutual information of variables $X_t$ and $Y_t$ [bits].
$k, l$ length of time series history used for variables $X_t$ and $Y_t$ [$\Delta t$].
$m$ number of states used to classify the data [positive integer].
$N$ number of data points in the data set [positive integer].
$n_\tau$ number of time lags $\tau$ being considered [positive integer].
$n_S$ number of variables in the subsystem $S$ [positive integer].
$n_V$ number of variables in the system $V$ [positive integer].
$P$ precipitation [mm month$^{-1}$].
$p(x_t)$, $p(y_t)$ marginal probability distribution of variables $X_t$ and $Y_t$ [fraction].
$p(x_t, y_t)$ joint probability distribution of variables $X_t$ and $Y_t$ [fraction].
$r$ resolution of the time series data set [T].
$R_g$ total incoming shortwave radiation [W m$^{-2}$].
$R_S(\tau)$ redundancy, a measure of feedback in the process network [fraction].
$R'_S(\tau)$ normalized version of $R_S$ [fraction].
$R'_{ss}(\tau)$ $R'_S$ computed using a shuffled-surrogate network [fraction].
$R''_{ss}(\tau)$ quantity comparing $R'_S$ with the surrogate $R'_{ss}(\tau)$ [± fraction].
$T(X_t > Y_t, \tau)$ abbreviated version of $T^G$ [bits].
$T'$ normalized transfer entropy [fraction].
$T_S^{[+]}(\tau)$ mean normalized gross information production of $S$ at $\tau$ [fraction].
$T_S^{[-]}(\tau)$ mean normalized gross information consumption of $S$ at $\tau$ [fraction].
$T_S^{net}(\tau)$ mean normalized net information production of $S$ at $\tau$ [± fraction].
$TST_S(\tau)$ normalized total system transport of $S$ at $\tau$ [fraction].
$TST_S^m(\tau)$ mean normalized total system transport of $S$ at $\tau$ [fraction].
$VPD$ vapor pressure deficit [kPa].
$X, Y$ source and sink variables, respectively [arbitrary units].
$X_t^{(i)}$ and $X_t^{(j)}$ time series versions of $X$ and $Y$ [units of data].

[62] **Acknowledgments.** This research is funded by the 2006–2009 NASA Earth Systems Science (ESS) Fellowship Program grant NNX06AF71H and NSF grant ATM 06-28687. The authors would like to acknowledge Richard Robertson and Darren Drewry, who provided valuable feedback.

## References

Albert, R., and A.-L. Barabasi (2002), Statistical mechanics of complex networks, *Rev. Mod. Phys.*, *74*, 47–97, doi:10.1103/RevModPhys.74.47.
Baldocchi, D., E. Falge, and K. Wilson (2001a), A spectral analysis of biosphere-atmosphere trace gas flux densities and meteorological variables across hour to multi-year scales, *Agric. For. Meteorol.*, *107*, 1–27, doi:10.1016/S0168-1923(00)00228-8.
Baldocchi, D., et al. (2001b), FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, *Bull. Am. Meteorol. Soc.*, *82*, 2415–2434, doi:10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2.
Ball, J. T., and W. I. E. Berry (1987), A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions, in *Progress in Photosynthesis Research*, vol. 4, edited by I. Biggins, pp. 221–224, Martinus-Nijhoff, Dordrecht, Netherlands.
Corning, P. A. (2002), The re-emergence of "emergence": A venerable concept in search of a theory, *Complexity*, *7*(6), 18–30, doi:10.1002/cplx.10043.
Gershenson, C., and F. Heylighen (2003), When can we call a system self-organizing?, in *Advances in Artificial Life: European Conference, ECAL 2003*, *Lecture Notes Artificial Intell.*, vol. 2801, edited by W. Banzhaf et al., pp. 606–614, Springer, Berlin.
Haken, H. (1979), *Synergerics: An Introduction*, 2nd ed., Springer, Berlin.
Haken, H. (1988), *Information and Self-Organization*, Springer, Berlin.
Hanway, J. J. (1966), How a corn plant develops, *Spec. Rep. 48*, Coop. Ext. Serv., Iowa State Univ., Ames, Iowa.
Hanway, J. J., and H. E. Thompson (1967), How a soybean plant develops, *Spec. Rep. 53*, Coop. Ext. Serv., Iowa State Univ., Ames, Iowa.
Hollinger, S. E., C. J. Bernacchi, and T. P. Meyers (2005), Carbon budget of mature no-till ecosystem in north central region of the United States, *Agric. For. Meteorol.*, *130*, 59–69, doi:10.1016/j.agrformet.2005.01.005.
Hubler, A. W. (2005), Predicting complex systems with a holistic approach, *Complexity*, *10*(3), 11–16, doi:10.1002/cplx.20071.
Jorgensen, S. E., B. D. Fath, S. Bastianoni, J. C. Marques, F. Muller, S. N. Nielsen, B. C. Patten, E. Tiezzi, and R. E. Ulanowicz (2007), *A New Ecology: Systems Perspective*, Elsevier, Amsterdam.
Juang, J. Y., G. G. Katul, A. Porporato, P. C. Stoy, M. S. Siqueira, M. Detto, H. Kim, and R. Oren (2007), Eco-hydrological controls on summertime convective rainfall triggers, *Global Change Biol.*, *13*, 1–10, doi:10.1111/j.1365-2486.2006.01315.x.
Katul, G. G., C. Lai, K. Schafer, B. Vidakovic, J. Albertson, D. Ellsworth, and R. Oren (2001), Multiscale analysis of vegetation surface fluxes: From seconds to years, *Adv. Water Resour.*, *24*, 1119–1132, doi:10.1016/S0309-1708(01)00029-X.
Kaufmann, S. A. (1993), *The Origins of Order*, Oxford Univ. Press, Oxford, U. K.
Kumar, P. (2007), Variability, feedback, and cooperative process dynamics: Elements of a unifying hydrologic theory, *Geogr. Compass*, *1*(6), 1338–1360, doi:10.1111/j.1749-8198.2007.00068.x.
Kurths, J., S. Boccaletti, C. Grebogi, and Y.-C. Lai (2003), Introduction: Control and synchronization in chaotic dynamical systems, *Chaos*, *13*(1), 126–127, doi:10.1063/1.1554606.
Meyers, T. P. (2008), Bondville 30 minute L4 product, IL, U.S.A., 1998–2007, http://www.daac.ornl.gov, Oak Ridge Natl. Lab. Distrib. Active Arch. Cent., Oak Ridge, Tenn.
Newman, M. E. J. (2003), The structure and function of complex networks, *SIAM Rev.*, *45*(2), 167–256, doi:10.1137/S003614450342480.
Nicolis, G., and I. Prigogine (1989), *Exploring Complexity*, W. H. Freeman, New York.
Ozawa, H., A. Ohmura, R. D. Lorenz, and T. Pujol (2003), The second law of thermodynamics and the global climate system: A review of the maximum entropy production principle, *Rev. Geophys.*, *41*(4), 1018, doi:10.1029/2002RG000113.
Reiners, W. A., and K. L. Driese (2003), Transport of energy, information, and material through the biosphere, *Annu. Rev. Environ. Resour.*, *28*, 107–135, doi:10.1146/annurev.energy.28.050302.105452.
Ruddell, B. L., and P. Kumar (2009), Ecohydrologic process networks: 1. Identification, *Water Resour. Res.*, *45*, W03419, doi:10.1029/2008WR007279.
Schreiber, T. (2000), Measuring information transfer, *Phys. Rev. Lett.*, *85*, 461–464, doi:10.1103/PhysRevLett.85.461.
Shannon, C. E. (1948), A mathematical theory of communication, *Bell Syst. Tech. J.*, *27*, 379–423.
Strogatz, S. (2001), Exploring complex networks, *Nature*, *410*, 268–276, doi:10.1038/35065725.
Watts, D. J., and S. H. Strogatz (1998), Collective dynamics of 'small-world' networks, *Nature*, *393*, 440–442, doi:10.1038/30918.
Wilhelm, T., and J. Hollunder (2007), Information theoretic description of networks, *Physica A*, *385*, 385–396, doi:10.1016/j.physa.2007.06.029.

P. Kumar and B. L. Ruddell, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. (kumar1@illinois.edu)