Information Theory in Ecology

Robert E. Ulanowicz
University of Maryland
Chesapeake Biological Laboratory
Solomons, MD 20688-0038 USA

Tel: 410-326-7266
FAX: 410-326-7378
Email: ulan@cbl.umces.edu

Citation: Computers and Chemistry 25:393-399 (2002).

Abstract:

The application of information theory (IT) to ecology has occurred along two separate lines: (1) It has been used to quantify the distribution of stocks and numbers of organisms; and (2) it has been employed to quantify the pattern of interactions of trophic processes. By and large, the first endeavor has resulted in relatively few insights into ecosystem dynamics and has generated much ambiguity and disappointment, so that most ecologists remain highly skeptical about the advisability of applying IT to ecology. By contrast, the second, and less well-known application has shed light on the possibility that ecosystem behavior is the most palpable example of a purely natural “infodynamics” that transcends classical dynamics, but remains well within the realm of quantitative description.
Introduction:

Ecosystem theory, or “synecology” (Odum 1959) is mostly a relational endeavor. That is, its practitioners are usually more intent upon elucidating the relationships among the various populations that make up an ecosystem than they are with determining how much or what kind of organisms comprise those elements. Such emphasis on relationships allows the discipline to be interpreted as the study of the *communication* of material, energy and information among systems components. Whence, synecology is especially well-suited for quantification by the same methods that were developed for communications theory, i.e., *information theory* (IT).

By way of contrast, the ecology of individual organisms and populations, or *autecology*, is far more involved with the nature of systems elements themselves. It tends to be less process-oriented; and, therefore, the application of a mathematical formulation arising out of communications theory appears more *ad hoc* and strained.

The thesis I wish to propose here and discuss below is that the introduction of information theory into ecology has proceeded along two identifiably separate lines, each of which has met with markedly different degrees of success: The application of information indices to quasi-static stocks or numbers (autecology) has led to a number of difficulties with interpretations that, at one point in time resulted in an almost total repudiation of any role for information theory in ecology. Because the larger majority of ecologists traffic in stocks and sizes rather than processes, this failure was widely advertised and engendered negative attitudes towards information theory in ecology. The echoes of this debacle persist to this very day. On the other hand, the use of information theory to quantify patterns of *processes* has become a wellspring of new insights and theories for ecology. Unfortunately, because processes and flows are much more difficult to conceptualize and measure, and because far fewer ecologists concern themselves with systems processes, these contributions of information theory to ecosystems research remain today in relative obscurity.
A Propitious Beginning:

Although there exist earlier formalisms that some place under the rubric of information theory (e.g., Boltzmann 1872), the beginnings of IT are usually dated from Claude Shannon’s (1948) exegesis of the mathematics used to break codes during World War II. Because it derives from the field of communications, IT soon took on the rather rigid format of “sender, receiver, channel, alphabet,” etc. Such formalities did not, however, stop many from adapting this new calculus to a variety of fields. Thus it was in 1955 that Robert MacArthur interpreted how IT might be used to characterize networks of ecological processes or flows. His formulation drew heavily upon the earlier narratives of Raymond Lindeman and Eugene Odum. MacArthur followed Lindeman’s (1943) lead by describing the ecosystem as a configuration of flows.

He compared various configurations in the effort to identify which ones were likely to be most stable over time, and his criterion for stability he drew from Odum (1953), “The amount of choice which the energy has in following the paths up through the food web is a measure of the stability of the community.” He argued that if one species happened to be abnormally common in the system, it would exert but a small effect upon the rest of the community only if its energy were distributed among a large number of predators. Similarly, if a population were abnormally uncommon, it could persist only if its predators had a large number of alternative hosts from which to draw their sustenance.

MacArthur proposed the Shannon-Weaver index to gauge the uncertainty (choice) in the flows. If, for example, $f_i$ represents the $i^{th}$ flow, and $F$ represents the sum of all the flows ($F = \sum_{i} f_i$), then $p_i = f_i/F$ becomes the fraction that the $i^{th}$ flow comprises of the total system activity. MacArthur’s candidate for the measure of stability then became
\[ S = -\sum_i p_i \log(p_i). \]  

(1)

MacArthur made no effort whatsoever to force his formalisms to comply with the categories of communication theory.

MacArthur demonstrated his extraordinary insight when he discussed the limitations of the index he had just proposed. He noted that stability rests upon species that have diets with many degrees of freedom; a restricted diet tends to lower stability. But a restricted diet is essential for efficiency, and both efficiency and stability are necessary for survival under natural conditions. That is, efficiency and stability tend to be antagonistic, but simultaneously necessary attributes of persistent systems. It therefore seemed to MacArthur at the time that (1) did not encompass all the attributes necessary for survival. With those observations MacArthur quit the argument, and his promising lead remained virtually untouched for another 20 years.

A Costly Diversion:

Although many subsequent investigators acknowledged MacArthur’s 1955 paper as the origin for information theory in ecology, subsequent authors seemed to have abandoned MacArthur’s approach. In particular, it was perhaps understandable that investigators might want to attempt another tack towards measuring system diversity, given the relative difficulty of measuring ecosystem flows. Besides, there was precedent for measuring community diversity in terms of population sizes and stocks using more conventional mathematics (Simpson 1949.) So it was that investigators after MacArthur chose to apply the Shannon-Weaver, not to system flows, but to the contents of each population. Thus, Margalef (1957, 1961) chose as \( p_i \), not the fraction of total activity, but rather the fraction a population comprises of the total number of organisms, or their total biomass. In symbols, he set \( p_i = n_i/N \), where \( n_i \) is the number of organisms in population \( i \) and \( N = \sum_i n_i \), so that the biodiversity, \( D \), became
\[ D = -\sum_i p_i \log(p_i). \]

For any number of reasons, not unrelated to those cited by MacArthur, a higher biodiversity was thought to impart greater stability to the system.

In the argument, “diversity begets stability”, ecology had a seemingly testable hypothesis, and the decade of the 60’s saw widespread and intense efforts among ecologists to test and elaborate the hypothesis (Woodwell and Smith 1969.) From the outset, the Shannon index was prominent among the candidates for quantifying biodiversity. Methods for assessing stability, however, arose more slowly. Because the study of dynamical stability was a well-developed component of applied physics, developments there eventually were imported into ecology, most prominently by the erstwhile physicist, Robert M. May. May (1973) recapitulated the results of Gardner and Ashby (1970), who studied the stability properties of randomly-assembled collections of coupled linear differential equations. These theoreticians discovered that systems of low connectivity and moderate interaction were almost always inherently stable. As both the degree of connectivity and the strength of interaction increased, however, the systems became unstable. He used the stability criterion of Wigner (1959) to quantify the point at which the probability for stability suddenly becomes negligible. One may conclude that, for linear dynamics among randomly-connected systems at least, diversity implies decreased stability.

May did cautioned that, since many diverse stable systems do exist, one should make the effort to discover why such diversity persists in ecosystems, but his warning was largely ignored. In the mind of most ecologists, it was the refutation of “diversity begets stability” that persisted. In the end, because May had stated his argument so eloquently, both mathematically and rhetorically, the feverish activity that had been devoted to the diversity-stability hypothesis collapsed very abruptly among a widespread but tacit embarrassment that swept through ecology. Information theory, because it had figured so prominently in the debate, was painted with the stigma of failure in the minds of many ecologists, and even today an undercurrent of this negative attitude continues --
some 30 years after these events -- to ramify among ecologists, despite numerous opinions and evidence to the contrary (e.g., Van Voris et al. 1980, Hastings 1984, Tilman and Downing 1994.)

Notwithstanding the disdain for information theory as a theoretical device in ecology, nor the lack of any significant theoretical justification to support it, popular interest in biodiversity remains quite healthy today. The Shannon measure of biodiversity often is invoked and/or measured when assessing the integrity of an ecosystem. For purely technical reasons, therefore, the Shannon index has been studied in much detail – centering on issues such as the dual meaning of the term (number and evenness), or the statistics of how to estimate the index using finite collections (Pielou 1969.) In fact, biodiversity has become such a prominent theme in the realm of environmental protection, that one is left to wonder why the fundamental assumption to focus on the diversity of population contents rather than the configuration of ecosystem processes has received so little scrutiny?

MacArthur Revisited:

MacArthur had been frustrated by the ostensible inability of the Shannon measure of choice in energy flow to embody both the notion of stability and that of efficiency. Demonstration of the fact that both attributes contribute to the Shannon measure awaited further refinements in information theory itself. The solution to parsing the contributions of correlative constraints to complexity from those made by dissociative factors called for Bayesian considerations. Instead of regarding only a single distribution of probabilities, A, which yields a Shannon measure H(A), one must as well consider the distribution of a second character, B, and study how A distributes when it interacts with B. If we call p(a_i) the probability that a_i occurs, p(a_i,b_j) the joint probability that a_i and b_j occur together, and p(a_i|b_j) the conditional probability that a_i occurs, given that b_j has transpired, then we may define a term called the average mutual information, I(A;B), such that
\[ I(A;B) = \sum_{i,j} p(a_i, b_j) \log \left( \frac{p(a_i | b_j)}{p(b_j)} \right) \quad i,j = 1,2,3,\ldots n \]

One can prove analytically that \( H(A) \geq I(A;B) \geq 0 \), and the interpretation of \( I(A;B) \) is that it gauges how much of the complexity measured by \( H(A) \) is resolved by the action of \( B \). I.e., it measures the degree to which \( A \) is constrained by \( B \) (and vice-versa, because \( I(A;B) = I(B;A) \) [the information is \textit{mutual}.]) The complement of \( I(A;B) \) w.r.t \( H(A) \) is called the \textit{conditional entropy}, \( H(A|B) = H(A) - I(A;B) \). In effect, the Shannon index \( H(A) \) has been decomposed into two components according to its relationship to \( B \),

\[ H(A) = I(A;B) + H(A|B), \quad (2) \]

where \( I(A;B) \) describes how much \( A \) is correlated with (constrained by) \( B \), and \( H(A|B) \) quantifies how much freedom \( A \) exhibits in the presence of \( B \).

Rutledge et al. (1976) used this decomposition in a very clever way to resolve MacArthur’s predicament. They identified \( a_i \) with the set of events, “a quantum of material or energy enters compartment \( i \),” while \( b_j \) was denoted as “a quantum leaves compartment \( j \).” In this way the resulting measures all become reflexive (a key attribute of self-organizing systems.) The interpretation given to equation (2) thereby becomes, “The complexity of the behavior of feeding (inputs) in the ecosystem equals the sum of the constraints that exist between predators and prey and the degree of choice left to the typical predator.” Since predator-prey constraints are usually associated with feeding efficiencies, it became explicitly clear how Shannon’s index embodies \textit{both} elements of efficiency and freedom. Rutledge et al. suggested that \( H(A|B) \) was the appropriate measure of choice extant in the system, and it should be related to system stability.

Rutledge et al. also suggested how the three measures could be estimated consistent with MacArthur’s earlier conventions. If \( T_{ij} \) represents the magnitude of flow
from i to j, then the total system activity becomes $T = \sum_{i,j} T_{ij}$. The joint probability, $p(a_i, b_j)$ takes the form $T_{ij}/T$, $p(a_i)$ becomes $\sum_j T_{ij}/T$, $p(b_j)$ is denoted by $\sum_i T_{ij}/T$, and $p(a_i|b_j)$ is estimated by $T_{ij}/\sum_i T_{ij}$. Using these estimates, the indices

$$H(A) = -\sum_i \left( \frac{\sum_j T_{ij}}{T} \right) \log \left( \frac{\sum_j T_{ij}}{T} \right),$$

and

$$I(A; B) = \sum_{i,j} \frac{T_{ij}}{T} \log \left( \frac{T_{ij}T}{\sum_p T_{pj} \sum_q T_{iq}} \right)$$

ensue, and $H(A|B)$ can be calculated by difference using (2).

Rutledge et al. sought to pursue Odum’s idea that those systems with the most choice were the most stable and persisted longest. Thus, they hinted at the conditional entropy of flows as the prime measure of system status. Ulanowicz (1980), on the other hand, was not convinced that system development could be identified entirely with stability. Of course, when indeterminate perturbations are significant, stability will play the pivotal role in defining system status. In the absence of major perturbations, however, one could argue that systems develop more along the lines of increasing efficiency. He, therefore, focussed upon mutual information as the measure of system organization. Earlier, in a seminal paper Eugene Odum (1969) had enumerated 24 attributes that could be used to identify ecosystems in the later stages of succession. Ulanowicz noted significant overlap between many of these indices and the average mutual information. Howard Odum (and Pinkerton 1955), however, had shown that efficiency is a poor indicator of system development in its inchoate early stages. During the initial stages of

---

1 The term entropy, chosen by Shannon, is retained only because it has become common usage. No relationship to thermodynamical entropy is implied thereby.
development, when resources often abundant, system changes seem to follow more the
direction of increasing power throughput. In an effort to encapsulate the whole course of
system development in a single index, Ulanowicz borrowed a cue from Tribus (and
McIrvine 1971) and scaled the average mutual information of the network of trophic
processes (organization) by the total system throughput, $T$, (power) to create a new
variable he called the system *ascendency*, $A$, where

$$ A = T \times I(A; B) $$

$$ = \sum_{i,j} T_{ij} \log \left( \frac{T_{ij}T}{\sum_p T_{pj} \sum_q T_{iq}} \right) \quad (3) $$

Because so many of Odum’s criteria for development appeared to correlate with increases
in the ascendency index, Ulanowicz proposed the phenomenological principle: “*In the
absence of major perturbations ecosystems exhibit a propensity to increase in
ascendency.*

Perturbations always do intervene, however, so it was necessary to track these
impacts as well. Whereas Rutledge et al. had used the conditional entropy, $H(A) - I(A; B)$
to measure freedom or dissociation, Ulanowicz and Norden (1990) pointed out how this
difference is biased in favor of inputs over outputs, and suggested that the upper limit for
the ascendency be estimated by the *joint entropy* instead. That is, they scaled the
difference between the joint entropy and the mutual information by the total system
throughput (as done earlier with the ascendency) to yield a more symmetrical system
property, $\Phi$, called the system *overhead.*
Their inference was that the stability of the system is maintained by the disorganized, incoherent, inefficient system processes that contribute to the overhead. Under quiescent or benign conditions, these inefficiencies tend to encumber system performance. During times of novel or stochastic stress, however, the same collection of such processes acts as a reservoir from which the system can draw to reconfigure itself (adapt) to the perturbation.

As MacArthur had surmised, both efficiency and stability are necessary for system persistence over the long run. It follows then from MacArthur’s lead that system integrity or sustainability should be related to the sum of these agonistic measures. Accordingly, the system capacity, $C$, can be gauged by the product of the total system throughput and the joint entropy,

$$
C = T \times H(A, B)
$$

$$
= A + \Phi
$$

$$
= -\sum_{i,j} T_{ij} \log \left( \frac{T_{ij}}{T} \right).
$$

Because the capacity serves as an upper bound on both $A$ and $\Phi$, anything which would contribute to an increase in $C$ would also provide capacity for either ascendency or overhead to grow. Brooks and Wiley (1986), for example, argue as how there is a natural tendency for $C$ to grow that is analogous to the thermodynamic drive towards increasing the entropy of the universe.

Critics of the information theoretic approach are quick to point out that the suite of variables defined thus far can be applied only to static or quasi-static configurations of
flows. Real ecosystems are heterogeneous in space and dynamical in time. Although gathering sufficient data to quantify networks over time and space can pose significant logistical problems, there is no conceptual barrier to extending the indices into the temporal and spatial domains, for all the information indices treated thus far possess counterparts in 3 or more dimensions (Abramson 1963.) Claudia Pahl-Wostl (1992) cautions, however, that the proper extension for the average mutual information in 3 dimensions is not the 3-dimensional counterpart, I(A;B;C), but rather the cluster {I(A;B;C) + I(A;B|C) + I(B;C|A) + I(C;A|B)}. In quantitative terms, if $T_{ijk}$ denotes the amount of medium flowing from i to j during time interval k (or in spatial cell k), then the corresponding 3-dimensional ascendency becomes

$$A = \sum_{i,j,k} T_{ijk} \log \left( \frac{T_{ijk}^2 T}{\sum_p T_{pjk} \sum_q T_{iqk} \sum_r T_{ijr}} \right). \quad (4)$$

In similar manner, four and higher dimensional versions of ascendency can be constructed to treat more complicated systems. The principle of increasing ascendency is presumed to apply as well to spatial and temporal dynamics. That is, systems are assumed to adapt to perturbations that possess regularities in time and/or space by adjusting their flow distributions in (4) so as to achieve higher values of the system ascendency. In part, at least, the answer to one of the central questions in ecology, “Why are organisms distributed over time and space according to the patterns we observe?” should lie in the application of the principle of increasing (multi-dimensional) ascendency (Ulanowicz 2000.)

Yet another criticism of the theory as presented thus far might be that they are all estimated solely in terms of processes, and dynamics are not controlled by the rates of processes alone. The criticism made early in this essay of the awkwardness of applying information theory to population numbers and stocks notwithstanding, it remains inconceivable that the magnitudes of the contents of the compartments should play no role in guiding system dynamics. How then to introduce compartmental stocks into the calculation of the ascendency in a manner that is consonant with the formalisms of
information theory? One effective solution (Ulanowicz and Abarca 1997) appears to be to regard the apriori probability for an exchange of medium between i and j to be proportional to the product of the stocks contained in those respective nodes (the law of mass action.) If \( B_i \) represents the level of the contents of i and \( B_j \) that of j, then the joint probability that a flow will transpire from i to j could be estimated by the product \( B_i B_j / B^2 \), where \( B = \sum_i B_i \). The corresponding aposteriori probability may be taken simply as the fraction that the observed flow, \( T_{ij} \), constitutes of the total activity, i.e., \( T_{ij} / T \). The weighted difference between the aposteriori and apriori probabilities is what in information theory is called the Kullback-Leibler difference. In terms of the measured quantities already described, the biomass-inclusive form of the ascendency, \( A_b \), becomes

\[
A_b = \sum_{i,j} \left( \frac{T_{ij}}{T} \right) \log \left( \frac{T_{ij} B^2}{B_i B_j T} \right),
\]

which can be generalized for 3 or more dimensions as discussed in the previous paragraph.

The partial derivatives of the ascendency w.r.t. either flows, \( T_{ij} \) or biomasses, \( B_i \) represent the sensitivities of the whole system status w.r.t changes in the corresponding individual system elements. Typically, large values of these derivatives signify where resistance is controlling further system development. Ulanowicz and Baird (1999, SCOPE 1999) therefore employed a sensitivity analysis of the spatial networks of several nutrient elements in the Chesapeake Bay ecosystem to pinpoint which element should control system dynamics during each season.

**Infodynamics:**

MacArthur paid little heed to the conventions of communications theory when he framed his index of flow diversity, and subsequent developments discussed here have departed even further afield. The characteristics measured by Rutledge’s indices bear
little conceptual relation to the quantitative attributes of channels for communication of code (unlike the applications of IT to molecular biology, where the coding theory is explicit.) Rather, they represent the overall effect of numerous unidentified constraints at various scales that serve to guide physical medium through the ecosystem. Collier (1990) chose to refer to such constraints as “enformation” to distinguish their immanent nature from that of conventional information. In one sense, at least, information theory applied to ecosystem networks makes explicit and measurable the effects of “signs” and other agencies that one encounters today as somewhat more cryptic narrative under the rubric of “semiotics” (Hoffmeyer 1993.)

More than the nature of information is at issue when IT is applied to ecology, however. System “dynamics” can take on a radically new guise. In conventional science, the world is perceived as a uni-directional hierarchy, wherein causes arising at microscopic scales ramify upwards to generate macroscopic patterns. The goal of theory usually is to describe events at the lowest level possible in terms of deterministic mechanical dynamics. The effects of these mechanical agencies may be modified by chance interferences, but their effect is believed to be expressed at higher levels in some simple aggregated fashion. Under this scenario, any correspondence of conventional dynamics with the applications of information theory could be considered only accidental, at best.

There is another possibility, however – that the dynamics expressed in terms of information variables are legitimate in their own right. That is, the world is not structured simply by the machinations of energy and mass. Structures, such as those containing positive feedback or autocatalysis, can act as agencies in the sense of Aristotle’s formal cause (Ulanowicz 1999), and their effects can ramify downwards as well as up the hierarchy of scales. In this view, the increase of ascendency is more than a phenomenological quirk to be explained away by reductionistic exegesis. Increasing ascendency constitutes the controlling dynamic. Particular mechanisms and elements can come and go and are selected for (and against) according to the driving dynamic. A scenario wherein information plays a key causal role has been termed “infodynamics” by
Weber et al. (1989), and we leave the reader to ponder whether ecology might be governed as much (or more) by infodynamics as by the conventional sort.

At this point one might inquire exactly where in the information indices do system dynamics lie? One answer would be to note how increasing ascendency represents changing constraints, but that is only part of the story. A peculiar formal property of the ascendency is that, even when it is calculated on a static network, clues to the dynamics behind the network pattern are built into the index of status. To appreciate this remarkable circumstance, one need only calculate the derivatives of the ascendency w.r.t. the individual flows. Thus, taking the derivative of (3) w.r.t. $T_{wx}$, one arrives, after some algebra, at

$$\frac{\partial A}{\partial T_{wx}} = \log \left( \frac{T_{wx}T}{\sum_p T_{px} \sum_q T_{wj}} \right),$$

so that the equation for the ascendency can be rewritten as

$$A = \sum_{i,j} T_{ij} \frac{\partial A}{\partial T_{ij}}.$$ 

But this relation is homologous to the chain rule,

$$dA = \sum_{i,j} \frac{\partial A}{\partial T_{ij}} dT_{ij}.$$ 

Whence we conclude that, whenever information theory is applied to the system structure, one obtains implicit clues about how the system is likely to change. This
contrasts with the situation in conventional dynamics, where information on static configurations tells one nothing at all about the system’s dynamical behavior.

**Summary and Conclusions:**

It appears the study of ecology will always be pursued along two separate lines. The majority of ecologists will continue to pursue autecology, or the study of single populations in relation to an external environment. Emphasis in autecology is upon population size and contents, and relationships with other populations remain of secondary concern. As suggested above, information theory arose out of a relational context, so that any applications it may find in autecology are likely to appear awkward and not wholly natural. Past attempts to apply information theory to population distributions appeared ad-hoc, and the confusions that arose as a result have soured autecologists on the utility of IT. It seems unlikely that in the eyes of autecologists the reputation of information theory will ever recover from such setbacks, so that the tool probably will continue to be deemed ineffective by the great majority of ecologists.

Synecology, by contrast, places primary emphasis upon the relationships of populations with one another. As MacArthur and Rutledge have demonstrated, information theory can be applied in a very natural way to the investigations of interrelated processes. As a result, the development of theory cast in terms of IT and the rate at which this theory has generated hypotheses has outpaced the ability of community ecologists to produce systems-level data with which to test the propositions. Furthermore, the nature of the dynamics suggested by the use of information theory seem to be quite unconventional, albeit intriguing and full of promise. Because synecologists remain such a small minority, however, it remains questionable whether such promise will ever be appreciated by the community of ecologists at large.

This predicament is perhaps well-illustrated by how IT has been applied to the very visible and politically popular issue of biodiversity. There seems to exist an overwhelming consensus that global biodiversity is worth preserving. The use of IT to
quantify biodiversity in terms of the diversity of populations contributes to a theory that stubbornly refuses to justify such concern, predicated as such conventional theory is upon contents and mechanical dynamics. Information theory, used to quantify the diversity of trophic flows, on the other hand, opens a vista on a whole new relational dynamics – one which provides ready justification for the protection of the diversity of ecosystems processes. Perhaps it is time to reconsider the assumptions we use to view the living world?

**Acknowledgements:**

The author was supported in part by the US Geological Survey Program for Across Trophic Levels Systems Simulation (ATLSS, Contract 1445CA09950093) and by the US Environmental Protection Agency’s Multiscale Experimental Ecosystem Research Center (MEERC, Contract R819640). He also wishes to acknowledge encouragement from the Editor, Andrzej Konopka and an anonymous reviewer.
References:


