

Time Series Analysis of Ecosystem Variables with Complexity Measures

Holger Lange

Ecological Modeling Group

Bayreuth Institute of Terrestrial Ecosystem Research (BITÖK)

University of Bayreuth

D-95440 Bayreuth, Germany

holger.lange@bitoek.uni-bayreuth.de

<http://www.bitoeck.uni-bayreuth.de/Mitarbeiter/000125/EN.html>

1.1.1 Science and Management of Ecosystems

Natural forested watersheds are prototypes of ecosystems. They have a relatively well-defined boundary, closed cycles for fluxes of elements, contain a collection of species of all kinds of spatial sizes (microbes to trees), and are of important human concern as timber source and drinking water supply. They have been the focus of scientific investigation since many years now, and it seems to be a rather trivial statement that they exhibit rather complex behavior. Considering the amount of scientific knowledge gained, the intricate nature of food webs, transport mechanisms for matter and energy, growth and succession patterns, or responses to natural hazards, it almost occurs like a miracle that they are nevertheless manageable from a practitioner's point of view. Forestry has been working since centuries and has come up with simple empirical rules which allow timber yield predictions for a full rotation period (e.g. 80 years). Drinking water supply works properly in many cases. Science has not yet come up with an explanation for the existence or effectiveness of these empirical rules. Some of these rules are devaluated nowadays due to anthropogenic disturbances, e.g. in the context of forest decline symptoms. However, thus far science failed as well to give conclusive hints why these managing rules do not work properly. There is an obvious conflict between the managing and the scientific approach to such ecosystems.

The reason behind this conflict is in our opinion the importance of the unique historical development of each given system (Hauhs and Lange 1996). They are rendered manageable by an effective *biological reset*, e.g. by planting operations, successive thinning etc., and thus a reduction or "amputation" of historical influence.

On the other hand, a de novo (re)construction of such systems in the laboratory has not been possible; no scientific substitute for practical experiences has been found. It may well be that the type of knowledge residing in experienced practitioners is not only nonscientific in nature, but to a certain extent nonverbal also. An approach relying heavily on visualization tools to explicate the inferential behavior of foresters is currently under investigation and discussed in (Lange et al. 1998a).

1.1.2. Problems with process-oriented models

The convincing success of deterministic models in physics and other “hard” sciences has let ecosystem researchers to follow similar lines, especially in the context of transport simulation. Here, appropriate simplifications of the Navier-Stokes equation have been developed for the description of water movement in soils and other porous media, and the movement of solutes with convection-diffusion models (Bear and Verruijt 1990).

Direct modeling of observed transport patterns is virtually never possible. The available field data are rare and of limited quality; the experimental constraints on parameter functions describing soil hydraulic and hydrochemical properties are almost never very restrictive. Mismatches between model reconstructions and measurements are easily attributable to spatial heterogeneity. Therefore, inverse modeling is applied, fitting the transport properties of the system at an assumed level of heterogeneity and a given scale (Lange et al. 1996). In many cases, this procedure leads to successful data reproductions; however, the fitted properties are far from being unique or very near so; the danger of overparametrization is notorious. In addition, the information contained in output variables is too limited to justify an approach requiring such a high amount of assumptions on the interior. This could be demonstrated with a series of tracer experiments for a well-controlled small catchment (Lange et al. 1996).

We conclude that if the focus is on characterizing the relationship between input and output of an ecosystem, an explicit modeling assuming lots of internal and unobserved processes should be avoided. The mapping performed by the system should be characterized rather by direct data analysis.

1.2. Time Series Investigation Methods

The selection of appropriate measures to quantify ecosystem input and output is primarily dictated by the length and quality (w.r.t. gaps, outliers etc.) of the available data. Some of the monitored data, especially related to hydrology, are taken since decades on a daily basis. Others are irregularly and sparsely sampled, e.g. monthly runoff chemical composition. Secondary, the type of information sought for determines the choice as well. Some of the methods quantify short-term dynamics, others long-term correlations, and yet another group the relationship between two or more data sets. We therefore suggest to use a variety of methods, and to start with standard techniques like periodograms, or auto- and cross-correlation functions.

However, experience shows that in our case, some nonlinear methods are more powerful or give important additional information. We give three examples for them in the following.

1.2.1 Hurst analysis

The phenomenon of persistence (extension of periods with systematic deviations from the long-term mean), first considered in the context of Nile floodings (Hurst 1951) and often present in hydrological data sets, is quantified by calculating the Hurst exponent H using the rescaled range or R/S statistics (Mandelbrot and Wallis 1969). This is a prototypical example of a long-term nonlinear measure.

Given a time series X of length n and a time scale k , we plot the quantity

$$q(n, k) = R(n, k) / S(n, k) \quad (1)$$

versus k for various values of n (e.g. 20 different realizations), and the expected persistence behavior $q \propto k^H$ is fitted to the results. Here, R denotes the *total range statistics*: Building partial sums from the original series X ,

$$Y_n = \sum_{i=1}^n X(t_i) \quad (2)$$

the deviation from a linear increase (as for a constant or essentially stationary series) is calculated as

$$D(n, i, k) = Y_{n+i} - Y_n - \frac{i}{k} (Y_{n+k} - Y_n) \quad (3)$$

and its range is given by

$$R(n, k) = \max_{0 \leq i \leq k} D(n, i, k) - \min_{0 \leq i \leq k} D(n, i, k) \quad (4)$$

In order to obtain a dimensionless test quantity, R is divided by the standard deviation of the series at timescale k :

$$S(n, k) = \sqrt{\frac{1}{k} \sum_{i=n+1}^{n+k} (X(t_i) - \bar{X}(n, k))^2} \quad (5)$$

thus arriving at eq. (1). If Hurst scaling is found, the expected H values are $0 \leq H \leq 1$; $H \leq 0.5$ means antipersistent behavior, which is seldom observed in experimental data, $H=0.5$ is ordinary Brownian motion, and $H \geq 0.5$ is fractional Brownian motion with increasing persistence strength as H approaches 1.

A possible caveat is that the fixation of the range of time scales where Hurst scaling is assumed is a somehow subjective manner. In some cases, multiple scaling regimes are

observed, indicating a finite memory contained in the data, or broken scaling in general. Then it is more a matter of taste whether one attributes Hurst behavior to the data at all.

1.2.2 Complexity measures

Short-term fluctuations can be investigated through information-theoretic measures. They allow to quantify the information content, or the randomness, of data sources on one hand, and their complexity on the other. By the latter we intuitively mean a measure which vanishes or is very small for constant or periodic sequences, but also for completely random data, as both types are easy to describe. For time series which are not amenable to such an easy description involving only a few parameters, the complexity measures should show high values.

We will discuss here only one randomness and one complexity measure appropriate for our purposes. Their calculation is performed by first constructing a symbol string from the data X via partitioning:

$$\Sigma: X \rightarrow A, x_i \rightarrow a_j \quad A = \{a_j\}, j=0, \dots, A-1 \quad (2)$$

where A is the alphabet for the symbol sequence. In the examples presented, we have chosen a binary alphabet, and the partitioning was performed in a static manner: the values of the observed quantities were cut at the median of their distribution, and all values below (above) the median were assigned the symbol 0 (1).

Then, one defines a word length L to group symbols together (in this article, $L=4$). The relative word frequencies $p_{L,i}$ for each possible word i (i runs from 1 to A^L) and conditional (or transition) probabilities $p_{L,ij}$ (i.e. the probability to find word j immediately after word i) are calculated. These are the ingredients to calculate the *generalized Shannon entropy*

$$H_L = - \sum_{i=1}^{A^L} p_{L,i} \log_2 p_{L,i} \quad (3)$$

and our measure for randomness, the *mean information gain*

$$MIG_L = H_L - H_{L-1} \quad (4)$$

Finally, the fluctuation complexity (Bates and Shephard 1993) is given by (index L suppressed for simplicity)

$$\sigma_{fc}^2 = \sum_{i,j} p_{ij} \left(\log \frac{p_{ij}}{p_j} \right)^2 \quad (5)$$

These two quantities have the desired features, as can be demonstrated e.g. for binary Bernoulli sequences (Lange 1999). Whereas MIG is nonlinearly proportional to randomness, being more sensitive to structural changes in the region of low randomness, FC exhibits a maximum and vanishes for constant as well as completely random sequences. Thus, FC is close to our intuition of what a "true" complexity measure should be.

1.2.3. Recurrence quantification analysis

Possible instationarities, trends, extreme periods, periodicities and many other (nonlinear) features of the time series are visualized and quantified by this technique (Zbilut et al. 1998a). It essentially calculates distances between time series values at all possible different times and stores and displays them in a matrix. The properties of this recurrence matrix are the basis for several derived quantities like the local recurrence rate, the degree of determinism, the Shannon entropy of line segments, or an approximation to the highest Ljapunov exponent present in the system. The method does not presuppose stationarity or large amounts of data and is not very restrictive when data points are not equidistant.

After normalization, one also can compare two different time series for the same observation period, which is the subject of cross-recurrence quantification (Zbilut et al. 1998b). One application for this rather new technique will be given.

1.3 Data sets and results

In this short overview, examples from long-term monitoring data on hydrology are shown only. Precipitation and runoff from natural catchments have been measured daily for decades at a reasonable large number of places. The quality of the data is generally high, with the possible exception of data from places where a substantial amount of precipitation occurs as snowfall.

Fig. 1 shows a Hurst analysis for one of our investigation sites, located in the Fichtelgebirge, NE Bavaria, Germany. The typical result is that the output from the system is much more persistent than the input. This long-term correlational structure is imprinted by the system and not present in other driving variables. It is also a universal feature of runoff from catchments of drastically different sizes (Pelletier and Turcotte 1997). The analysis also shows that rainfall is definitely different from a pure noise process.

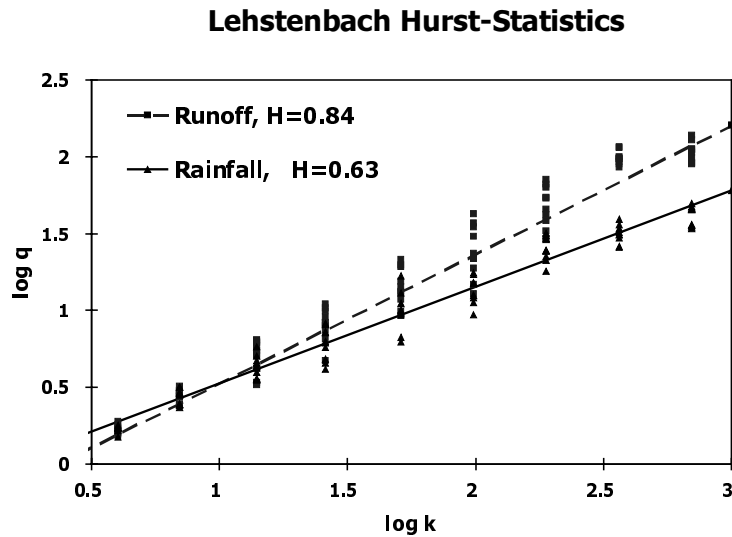


Figure 1: Hurst analysis for a small (4.2 km²) forested catchment in NE Bavaria. Runoff is more persistent than rainfall, which is different from pure noise. The scale k is in days.

An example of a complexity analysis is shown in Fig. 2. Here, 30 catchments from different climatic regions are compared, using the randomness and complexity measures discussed above as abscissas and ordinates, respectively. The Bernoulli process provides a limiting curve: apart from possible finite size effects, all experimental data sets should lead to points below it (Lange 1999). Our observation is that for most data from natural systems, the deviation from the maximum complexity at given randomness (mean information gain) is small. On the other hand, artificial or highly controlled systems usually have much larger distance from the limit. This is demonstrated here: whereas runoff from the natural catchments operates at high complexities and intermediate randomness, urban (channeled) systems have much more randomness, indicating that biological uptake (dominantly evapotranspiration) smoothes signals. It is thus feasible to think of ecosystems as information filters (Hauhs and Lange 1996).

Precipitation is characterized by extremely high randomness and consequently low complexity. The difference in randomness between input and output from the same system serves as measure for its efficiency, whereas the distance to the limiting curve is considered as an indicator for the « naturalness » of the system.

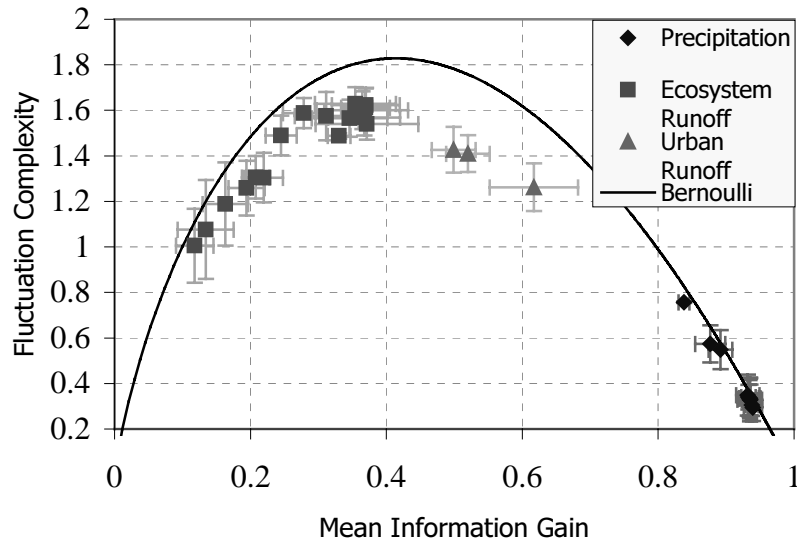


Figure 2: Complexity diagram for daily data from 30 different catchments. Natural runoff is significantly different from urban one; rainfall is highly random. The symbols and error bars denote the mean and standard deviation for windows of length 4 years each.

An example of recurrence analysis is given in Fig. 3. The data are runoff values from watershed 2 of the Hubbard Brook Experimental Forest (White Mountains, New Hampshire), the time span is 9 years (1962-1970). Both axes of this plot are time axes. A dark pixel at position (i, j) of this plot indicates that the time series values at position i and j are similar in magnitude (according to a threshold parameter which has to be fixed); otherwise the pixel is left white.

At the beginning, a clear periodic pattern can be recognized (chess-board like structure of black and white rectangles), which is to be expected since tree water uptake during the vegetation period induces seasonality. The dark rectangles correspond to dry periods in summer (or low flow conditions).

After four years, however, the periodic pattern is suddenly lost, obviously a dramatic change in the underlying dynamics. This is the consequence of a drastic experimental manipulation: a clear-cut of the vegetation, followed by pesticide treatment for the subsequent 3 years (Likens and Bormann 1995). Whereas hydrochemically, clear indications of the clear-cut were found, the hydrological response was not detected by linear methods. This demonstrates the power of recurrence quantification.

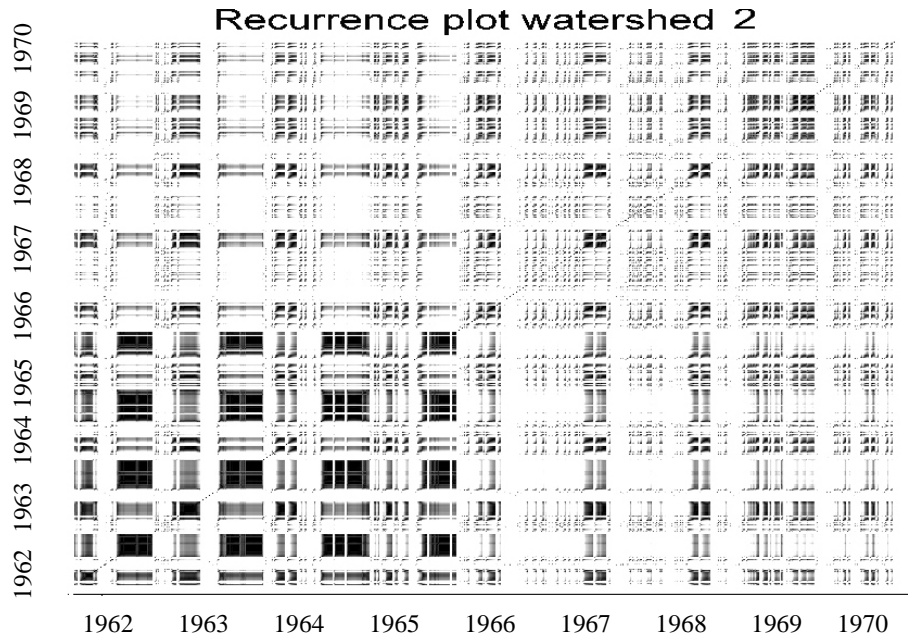


Figure 3: Recurrence plot for the Hubbard Brook watershed 2 for 9 years (1962-1970). The periodic pattern visible for the first four years vanishes suddenly at the end of 1965 : the watershed was clear-cut in that year and prevented from regrowth (by pesticides) for three more years.

1.4 Conclusions

It could be demonstrated that new (generally nonlinear) techniques of time series analysis can provide new information about the detailed structure of data from natural and anthropogenic ecosystems, which is not available by standard tools. The results indicate that extended memory effects are present in the systems and that biological activity induces them (cf. recurrence plots). As the history of system development seems to be crucial, an approach to model the system as a whole (biotic as well as abiotic parts) as a state-determined dynamical system is not expected to work.

Acknowledgements: This work was funded by the German Ministry of Education and Research (BMBF) under grant No. PT BEO 51-0339476B.

References

- Bates, J. E., & Shephard, H. K., 1993, Measuring complexity using information fluctuation. *Physics Letters A* **172**, 416-425.
- Bear, J., & Verruijt, A., 1990, *Modeling groundwater flow and pollution*. Reidel (Dordrecht).
- Hauhs, M., & Lange, H., 1996. Ecosystem dynamics viewed from an endoperspective. *The Science of the Total Environment* **183**, 125-136.

- Hurst, H.E., 1951, Long-Term Storage Capacity of Reservoirs. *Transactions of the American Society of Civil Engineering* **116**, 770-799.
- Lange, H., Lischeid, G., Hoch, R., & Hauhs, M., 1996, Water Flow paths and Residence Times in a Small Headwater Catchment at Gårdsjön (Sweden) during Steady State Stormflow Conditions. *Water Resources Research* **32**, 1689-1698.
- Lange, H., Thies, B., Kastner-Maresch, A., Dörwald, W., Kim, J.T., & Hauhs, M., 1998a, Investigating forest growth model results on evolutionary time scales. In: Adami, C., Belew, R.K., Kitano, H., & Taylor, C.E. (eds.), *Artificial Life VI*, pp. 418-422. MIT Press (Cambridge, Mass.).
- Lange, H., Newig, J., & Wolf, F., 1998b: Comparison of complexity measures for time series from ecosystem research. *Bayreuther Forum Ökologie* **52**, 99-116.
- Lange, H., 1999, Are Ecosystems Dynamical Systems? *International Journal of Computing Anticipatory Systems* **3**, 169-186.
- Likens, G.E., & Bormann, H., 1995, *Biogeochemistry of a Forested Ecosystem*, Springer-Verlag (New York).
- Mandelbrot, B.B., & Wallis, J.R., 1969, Robustness of the rescaled range R/S and the measurement of noncyclic long run statistical dependence. *Water Resources Research* **5**, 967-988.
- Pelletier, J.D., & Turcotte, D.L., 1997, Long-range persistence in climatological and hydrological time series: analysis, modeling and application to drought hazard assessment. *Journal of Hydrology* **203**, 198-208.
- Zbilut, J.P., Giuliani, A., & Webber Jr., C.L., 1998a, Recurrence quantification analysis and principal components in the detection of short complex signals. *Physics Letters A* **237**, 131-135.
- Zbilut, J.P., Giuliani, A., & Webber, C.L., 1998b, Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. *Physics Letters A* **246**, 122-128.