Journal of Hydrology 414–415 (2012) 99–107

Contents lists available at SciVerse ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Scale effects on information theory-based measures applied to streamflow patterns in two rural watersheds

Feng Pan [a,c,*], Yakov A. Pachepsky [b], Andrey K. Guber [d], Brian J. McPherson [a,c], Robert L. Hill [e]

[a] Energy and Geoscience Institute, University of Utah, 423 Wakara Way Suite 300, Salt Lake City, UT 84108, USA
[b] USDA-ARS Environmental Microbial and Food Safety Laboratory, 10300 Baltimore Ave., BARC-EAST Bldg. 173, Beltsville, MD 20705, USA
[c] Department of Civil and Environmental Engineering, University of Utah, Salt Lake City, UT 84112, USA
[d] Department of Crop and Soil Sciences, Michigan State University, East Lancing, MI 48824, USA
[e] Department of Environmental Science and Technology, University of Maryland, College Park, MD 20742, USA

ARTICLE INFO

Article history:
Received 26 April 2011
Received in revised form 22 September 2011
Accepted 14 October 2011
Available online 25 October 2011
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Erwin Zehe, Associate Editor

Keywords:
Information content measures
Complexity measures
Streamflow patterns
Temporal and spatial scales

SUMMARY

Understanding streamflow patterns in space and time is important for improving flood and drought forecasting, water resources management, and predictions of ecological changes. Objectives of this work include (a) to characterize the spatial and temporal patterns of streamflow using information theory-based measures at two thoroughly-monitored agricultural watersheds located in different hydroclimatic zones with similar land use, and (b) to elucidate and quantify temporal and spatial scale effects on those measures. We selected two USDA experimental watersheds to serve as case study examples, including the Little River experimental watershed (LREW) in Tifton, Georgia and the Sleepers River experimental watershed (SREW) in North Danville, Vermont. Both watersheds possess several nested sub-watersheds and more than 30 years of continuous data records of precipitation and streamflow. Information content measures (metric entropy and mean information gain) and complexity measures (effective measure complexity and fluctuation complexity) were computed based on the binary encoding of 5-year streamflow and precipitation time series data. We quantified patterns of streamflow using probabilities of joint or sequential appearances of the binary symbol sequences. Results of our analysis illustrate that information content measures of streamflow time series are much smaller than those for precipitation data, and the streamflow data also exhibit higher complexity, suggesting that the watersheds effectively act as filters of the precipitation information that leads to the observed additional complexity in streamflow measures. Correlation coefficients between the information-theory-based measures and time intervals are close to 0.9, demonstrating the significance of temporal scale effects on streamflow patterns. Moderate spatial scale effects on streamflow patterns are observed with absolute values of correlation coefficients between the measures and sub-watershed area varying from 0.2 to 0.6 in the two watersheds. We conclude that temporal effects must be evaluated and accounted for when the information theory-based methods are used for performance evaluation and comparison of hydrological models.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Quantification of streamflow patterns, especially observable and to some extent predictable recurrent changes in hydrographs, is of paramount importance for improvement of flood and drought forecasting and water resource management (Black, 1996; Simonovic, 2009). The structure and composition of lotic communities in fluvial ecosystems strongly depend on streamflow patterns (Poff and Allan, 1995), and these patterns may be observed and interpreted at various temporal and spatial scales (Poff, 1996; Matteau et al., 2009). Poff and Ward (1989) noted that detection of recovery from natural or anthropogenic disturbance in lotic ecosystems depends on the selection of appropriate spatial and temporal scales for the variables of interest. And, complementary information about hydrologic systems may be inferred from the analysis of streamflow patterns at different scales (Carey et al., 2010).

Various parameterization methods may be used to characterize streamflow patterns and detect changes in them. Historically, the first group of methods proposed rely on probability distribution functions of several flow characteristics. Poff and Ward (1989) developed a classification system of regional hydrologic regimes based on four flow characteristics: duration of intermittency, high flow frequency, high flow predictability, and overall flow variability. Subsequently, Richter et al. (1996) assessed flow alteration of regulated rivers in terms of five characteristics of flow: magnitude,

* Corresponding author at: Energy and Geoscience Institute, University of Utah, 423 Wakara Way Suite 300, Salt Lake City, UT 84108, USA. Tel.: +1 801 587 9788; fax: +1 801 587 9148.
E-mail address: fpan@egi.utah.edu (F. Pan).

0022-1694/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2011.10.018

duration, frequency, timing, and rate of change. With these methods, a relatively small number of parameters derived from the probability distribution functions may be used to characterize streamflow patterns. For example, Conrad and Booth (2005) discriminated between streamflow patterns using nine parameters that characterize the high flow frequency, the flow distribution, the daily variation, and the low flow magnitude.

A large group of methods to characterize streamflow patterns rely on multivariate statistical methods. Large numbers of streamflow metrics can be computed and then a principal component analysis (PCA) can be performed to create a small number of predictors that describe the most variance in the data. For example, Sanborn and Bledsoe (2006) computed 84 parameters to characterize streamflow patterns and applied a PCA to define aggregated metrics of flow patterns. Similarly, Clausen and Biggs (2000) used 35 parameters and defined four principal components as aggregated streamflow parameters. Canonical correlations or classification methods have been used together with PCA to relate flow properties to watershed properties (Detenbeck et al., 2005; Matteau et al., 2009), while cluster analysis with multiple streamflow pattern variables facilitates classification of streams into flow regime groupings (Baeza Sanz and García del Jalón, 2005; Moliere et al., 2009).

Various decompositions of streamflow time series have been proposed to characterize and classify streamflow hydrographs. Initially, Fourier series analysis was applied to evaluate climatic effects of streamflow at coarse temporal scales (Kahya and Dracup, 1993). Later, wavelet decompositions were used to characterize streamflow patterns at a progressively-increasing series of temporal scales (Smith et al., 1998; Anctil and Coulibaly, 2004).

Yet another group of methods of streamflow characterization concentrates on features in streamflow hydrograph shapes. For example, symbolic coding of hydrographs was proposed by Lange (1999). The range of flow variability is divided into a number of intervals, each interval is denoted with a symbol, and each flow observation is coded with the symbol of the interval within which this observation is found. The resulting string of symbols can be analyzed with information theory methods. As a result, streamflow patterns may be characterized, classified, and compared in terms of their information content and complexity (Engelhardt et al., 2009). This method of pattern classification has provided useful complementary information about the soil water model performance, especially comparison of results from different models and measured soil water time series (Pachepsky et al., 2006; Pan et al., 2011).

To our knowledge, scale effects on information content and complexity of streamflow patterns have not been studied in detail previously. The objectives of this work include (a) to characterize the spatial and temporal patterns of streamflow using information theory-based measures at thoroughly monitored agricultural watersheds located in different hydroclimatic zones with similar land use, and (b) to evaluate temporal and spatial scale effects on these measures.

2. Materials and methods

2.1. Study sites and data collection

The study sites are two USDA-ARS experimental watersheds, including the Little River experimental watershed (LREW) in Tifton, GA, and the Sleepers River experimental watershed (SREW) in North Danville, VT (Fig. 1). Both watersheds include several sub-watersheds with more than 30 years of continuous data records of precipitation and streamflow (USDA-ARS, 2009).

The LREW is located in the Western headwaters of the Suwannee River Basin (Bosch et al., 2007a). The watershed has an area of 334.2 km^2 with seven nested sub-watersheds, areas of which range from 2.6 km^2 to 114.8 km^2 (Table 1, Fig. 1) (Bosch et al., 2007a). The watershed is located in the Gulf-Atlantic Coastal Plain physiographic region in southeastern United States and has a humid climate with an annual average rainfall of 120.3 cm. The topography of the watershed includes broad flat floodplains, river terraces, and gently-sloping uplands. Land-surface elevation of the watershed ranges from 80 m to 150 m with the slopes less than 5%. Soil types within the watershed are mainly Tifton loamy sand, Alapaha loamy sand, and Kinston and Osier fine sandy loam (Sullivan et al., 2007). The land cover consists of 50% woodland, 31% row crops, 10% pasture, 2% water, and 7% others (Sullivan et al., 2007). The precipitation network has been designed to measure the rainfall within and in immediately surroundings of the LREW with 55 rain gauges installed and measuring streamflow data since 1967, with the number of gauges reduced to 31 since 1982 (Bosch et al., 2007b; Bosch and Sheridan, 2007).

The SREW, located in northeastern Vermont, has an area of 111 km^2 with a total of eleven sub-watersheds ranging from 0.5 km^2 to 43.5 km^2 (Table 1, Fig. 1). The watershed is located in a glaciated highland region of the northeastern U.S. and has a humid continental climate. Its annual average precipitation of 110 cm includes 25% snow (USGS, 2005). The elevation of the watershed ranges from 195 m to 780 m with slope varying from 3% to 35%. Bedrock within the watershed is the Waits River Formation, a quartz mica phyllite with layers of calcareous granulite (Shanley et al., 2002). The land cover consists of 67% forest and 33% agriculture, including pasture, hay and corn fields (USGS, 2005). Precipitation data have been observed at 13 sites and streamflow measured at 17 gauged watersheds since 1959 (USGS, 2005).

2.2. Information content and complexity measures

We use symbolic strings (Lange, 1999; Wolf, 1999) to approximate the time series of hydrologic variables in this study, including precipitation and streamflow, with systems that have a finite number of states.The median value of the measurements is found for each time series. The symbol "1" is assigned to the measurements that exceed the median value, and the symbol "0" is assigned to the measurements that fall below or at the median value. The word of length L is defined as a group of L consecutive symbols, and 2^L different words could be encountered in a string. For example, if the word length is three, then the eight possible words are 000, 001, 010, 100, 011, 101, 110, 111. Each word represents the state of the system. The transition from one state to another is defined as the change in the words starting from two consecutive observation times. For example, if the string is '100110' and the word length is three, then the shift from the first word '100' to the second word '001' represents the transition from the '100' state to '001' state, the shift from the second word '001' to the third word '011' represents the transition from the '001' state to '011' state, etc. Three sets of empirical probabilities are defined: (a) $p_{L,i}$: state probability of the occurrence of the ith in the symbolic string, $i = 1, 2, \ldots, 2^L$; (b) $p_{L,ij}$: probabilities of the transition from the ith to the jth L-word, $i = 1, 2, \ldots, 2^L$, $j = 1, 2, \ldots, 2^L$; and (c) $p_{L,i\to j}$ the conditional probability of the occurrence of the jth word provided that the ith word has occurred, $i = 1, 2, \ldots, 2^L$, $j = 1, 2, \ldots, 2^L$ (Wolf, 1999).The information content is quantified with two measures – metric entropy and mean information gain. The Shannon entropy (Shannon, 1948), $H(L)$, for words of length L is defined as a measure of information in the time series after it has been encoded with symbols:

$$H(L) = -\sum_{i=1}^{2^L} p_{L,i} \log_2 p_{L,i}, \tag{1}$$

Shannon's entropy measures the information contained in a message as opposed to the portion of the message that is determined

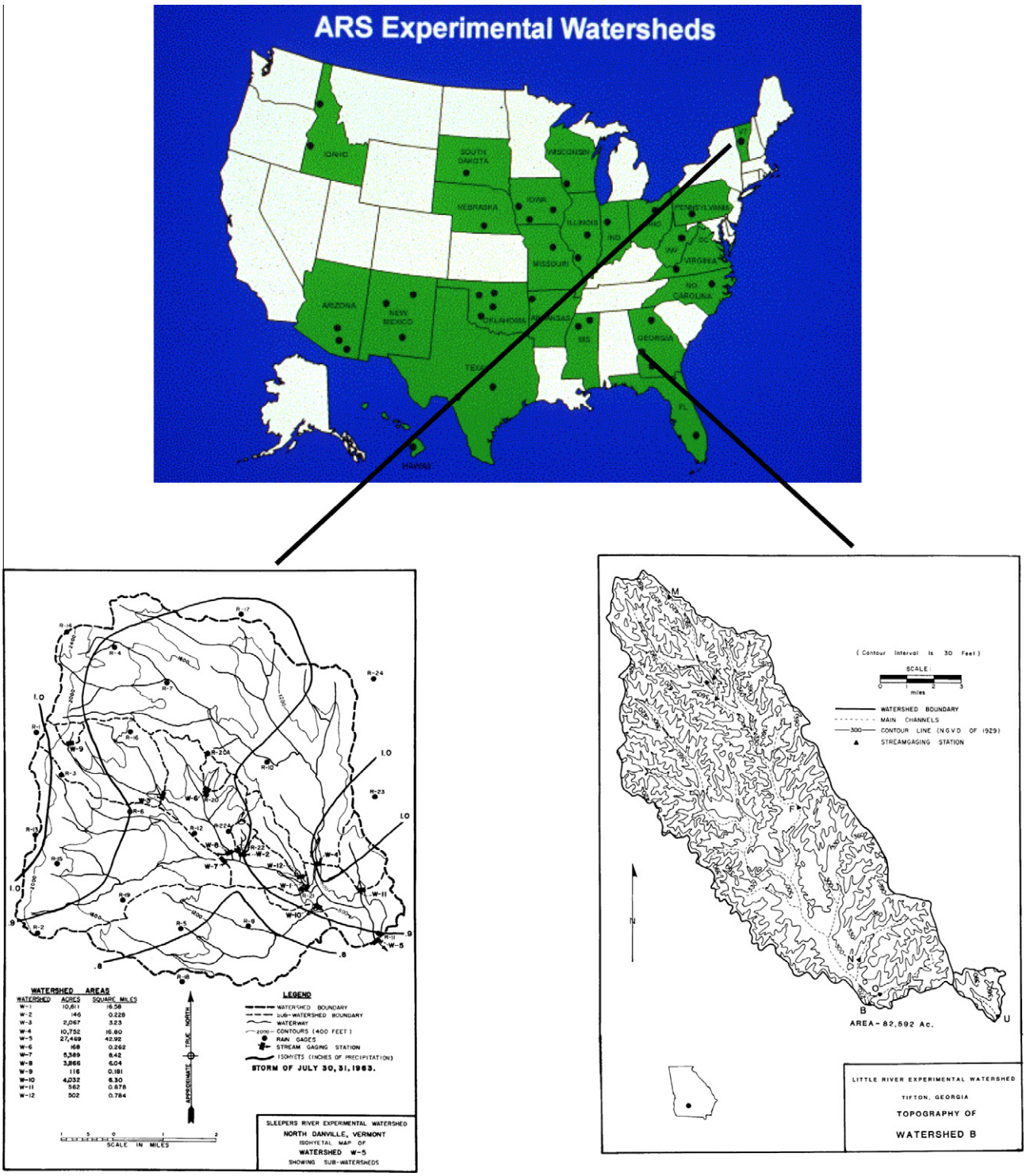

Fig. 1. Maps of Sleepers River experiment watershed (SREW), North Danville, Vermont and Little River experimental watershed (LREW), Tifton, Georgia.

or predictable (e. g, Chang et al., 2009). The metric entropy, H_u, is the Shannon's entropy divided by the word length, and this normalization results in the value of the information independent of the word length. The metric entropy represents the extent of the disorder in the sequence of symbols. The metric entropy vanishes for constant sequences, increases monotonically when the disorder increases, and reaches its maximum of one for uniformly distributed random sequences with complete randomness.The mean information gain, H_G, as yet another measure of the information content, quantifies the additional information that can be gained on the average for the whole symbol sequence from knowing the next symbol. It is defined as:

$$H_G(L) = \sum_{i,j=1}^{2^L} p_{L,ij} \log_2 p_{L,i\to j}. \tag{2}$$

The mean information gain includes the probabilities of state changes in a time series. The larger values of mean information gain indicate the larger possibility of state changes from one to another, and potential higher randomness or less predictability of a

Table 1
The information content and complexity measures of 5-year daily, half-daily, and quarter-daily streamflow time series at Sleepers River (SREW) and Little River experimental watershed (LREW) and their sub-watersheds.

SREW		Daily				Half-Daily				Quarter-Daily			
W	Area	ME	MIG	EMC	FC	ME	MIG	EMC	FC	ME	MIG	EMC	FC
W-1	42.9	0.73	0.43	0.61	1.35	0.67	0.32	0.69	1.48	0.61	0.21	0.79	1.42
W-2	0.6	0.72	0.38	0.68	1.31	0.66	0.30	0.72	1.45	0.63	0.24	0.78	1.32
W-3	8.4	0.74	0.43	0.62	1.30	0.69	0.34	0.69	1.34	0.62	0.24	0.76	1.46
W-4	43.5	0.73	0.43	0.61	1.37	0.66	0.31	0.70	1.54	0.60	0.20	0.80	1.42
W-5	111.1	0.73	0.43	0.60	1.44	0.67	0.32	0.69	1.53	0.61	0.22	0.78	1.43
W-6	0.7	0.68	0.30	0.76	1.12	0.65	0.25	0.80	1.14	0.61	0.20	0.81	1.28
W-7	21.8	0.70	0.37	0.66	1.32	0.64	0.29	0.71	1.47	0.59	0.19	0.80	1.35
W-8	15.6	0.75	0.46	0.58	1.34	0.70	0.36	0.67	1.35	0.63	0.25	0.76	1.44
W-9	0.5	0.69	0.35	0.69	1.41	0.63	0.26	0.75	1.43	0.60	0.18	0.82	1.26
W-10	16.3	0.74	0.45	0.59	1.35	0.67	0.33	0.69	1.45	0.61	0.22	0.78	1.40
W-11	2.3	0.69	0.34	0.71	1.31	0.65	0.27	0.76	1.27	0.61	0.21	0.80	1.32
W-12	2.0	0.76	0.49	0.55	1.37	0.69	0.37	0.65	1.48	0.63	0.26	0.74	1.51
LREW													
W-B	334.2	0.66	0.31	0.69	1.68	0.60	0.20	0.80	1.46	0.56	0.11	0.89	1.07
W-N	15.7	0.71	0.42	0.59	1.62	0.64	0.28	0.73	1.56	0.59	0.17	0.83	1.30
W-O	15.9	0.70	0.40	0.61	1.64	0.64	0.27	0.73	1.53	0.59	0.17	0.83	1.31
W-F	114.8	0.70	0.39	0.62	1.63	0.63	0.26	0.74	1.58	0.58	0.16	0.84	1.26
W-I	49.9	0.70	0.40	0.61	1.64	0.64	0.28	0.72	1.56	0.59	0.18	0.83	1.33
W-J	22.1	0.69	0.36	0.66	1.54	0.62	0.23	0.77	1.45	0.57	0.15	0.85	1.22
W-K	16.7	0.74	0.46	0.57	1.48	0.67	0.32	0.68	1.55	0.61	0.21	0.79	1.46
W-M	2.6	0.67	0.34	0.66	1.75	0.61	0.22	0.78	1.51	0.57	0.14	0.86	1.13

Note: SREW – Sleepers River experimental watershed; LREW – Little River experimental watershed; W – Name of sub-watersheds; Area – in km^2; ME – Metric entropy; MIG – Mean information gain; EMC – Effective measure complexity; FC – Fluctuation complexity.

time series.The complexity in this work reflects the extent of an internal structure in a time series. Two measures of such complexity – fluctuation complexity and effective measure complexity – were used to quantify the internal structure in symbolic strings. The fluctuation complexity (σ_Γ^2) is the variance of the net information gain, i.e., the differences between information gain and loss, defined as

$$\sigma_\Gamma^2 = \sum_{i,j}^{2^L} p_{L,ij}\left(\log_2 \frac{p_{L,i}}{p_{L,j}}\right)^2, \tag{3}$$

where the information gain is associated to the transition from the ith word to the jth word, $G_{L,ij} = \log_2(1/p_{L,i\to j})$, and the information loss $L_{L,ij} = \log_2(1/p_{L,j\leftarrow i})$ is associated with the occurrence of the jth word occurs after the occurrence of the ith word. The net information gain is the difference $G_{L,ij} - L_{L,ij}$. Thus, the fluctuation complexity characterizes the fluctuations in the system transitions from one state to another. The more the net information gain is fluctuating in the investigated string, the more complex is the string (Bates and Shepard, 1993).

The effective measure complexity (C_{EM}) evaluates the minimum total amount of information that has to be stored at any time for an optimal prediction of the next symbol. This measure can be approximately calculated as (Grassberger, 1986).

$$C_{EM} = \sum_{i,j=1}^{2^L} p_{L,ij}\log_2 \frac{p_{L,i\to j}}{p_{L,i}}. \tag{4}$$

Values of the complexity measures (fluctuation and effective measure complexity) are small for time series that are easy to describe such as constant or periodic sequences, or completely random data. Larger values of the complexity measures are observed in time series that are not amenable to an easy description involving only a few parameters (Pachepsky et al., 2006; Wolf, 1999).

All information theory-based measures of information content and complexity are computed in this study using the SYMDYN software (Wolf, 1999). Wolf (1999) presented the required lengths of a time series using binary strings (2-letter alphabet) for different number of symbols in order to estimate the information content and complexity measures with 5% relative error or better accuracy. The required lengths of a time series are 723 for 4-symbol words using a 2-letter alphabet to estimate the effective measure complexity. For 3-, and 4-letter alphabet, the required lengths of a time series could significantly increase. The 2-letter alphabet is used in this study.

2.3. Temporal and spatial scale effects on information theory-based measures of information content and complexity

The temporal scale of observations is defined as the time interval over which the streamflow or precipitation are averaged. The spatial scale is characterized by the area of a watershed or its sub-watersheds. Pearson correlation coefficients between information theory measures and scale measures are computed.

3. Results and discussion

3.1. Precipitation and streamflow patterns

The 5-year daily time series of precipitation and streamflow for SREW and LREW are illustrated in Fig. 2. The daily precipitation and streamflow exhibit significant temporal variability in the two watersheds (Fig. 2). In general, the precipitation and streamflow at SREW are relatively less variable than the data for LREW. For example, the 5-year accumulated precipitation is 5864 mm at SREW, and 5839 mm at LREW but their variances are 39.6, and 82.7 mm^2 at SREW and LREW, respectively. The streamflow time series exhibits large variation among sub-watersheds. For instance, the mean values of 5-year daily streamflow time series vary from 0.01 to 1.86 m^3/s at the 12 sub-watersheds of SREW, and are from 0.03 to 4.26 m^3/s at the 8 sub-watersheds of LREW.

Fig. 3 displays the information content and complexity measures of 5-year and 1-year daily precipitation and streamflow time series from 1965 to 1969 at SREW, and from 1974 to 1978 at LREW. Daily precipitation time series has larger values of the information content measures (mean information gain and metric entropy) than daily streamflow time series but smaller values of the complexity measures (fluctuation and effective measure complexity),

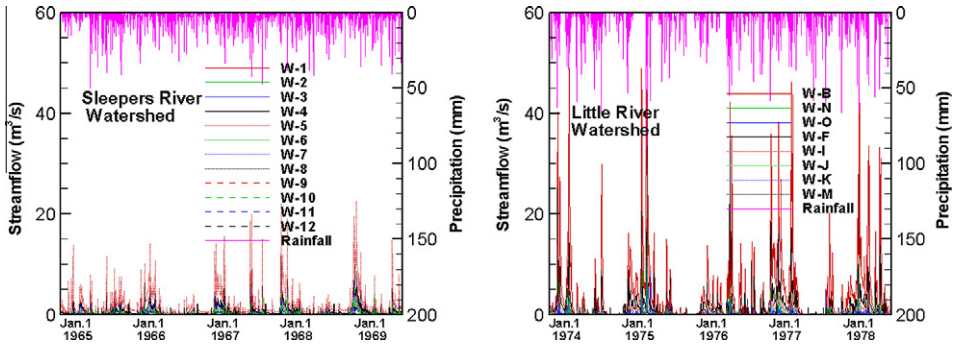

Fig. 2. The 5-year daily time series of streamflow and precipitation at Sleepers River (SREW) and Little River experiment watershed (LREW) and their sub-watersheds.

indicating relatively higher randomness and lower complexity of daily precipitation time series compared to daily streamflow data. The state probabilities of "00", "01", "10", and "11" for daily precipitation time series at sub-watershed W-5 of SREW are 0.324, 0.177, 0.177, and 0.322, respectively and corresponding values for daily streamflow time series are 0.459, 0.050, 0.051, and 0.440, respectively. As discussed in Section 2, the metric entropy reaches maximum with uniform distribution of a time series. Comparisons show the relatively smaller differences among the state probabilities for daily precipitation time series than corresponding values of streamflow time series, leading to higher information content measures and randomness of precipitation compared to streamflow. These data also reveal that precipitation, as one of many forcings driving the hydrologic cycle, is highly random, and meteorological controls of precipitation do not create systematic structure in precipitation time series at the study sites. The smaller information content and the higher complexity of daily streamflow time series compared with precipitation can also be explained by the fact that precipitation conversion to streamflow is controlled by watershed characteristics such as landscape, geology, geomorphology, and soil properties, and these controls impose additional structure on the streamflow time series as compared with precipitation.

Comparison of information content and complexity measures for 5-year time series of precipitation and streamflow between the two watersheds (Fig. 3) show that the precipitation and streamflow time series at SREW have larger information content and lower complexity than the same data at LREW. The higher randomness of precipitation at SREW in terms of probabilities of different words is associated with the fact that precipitation at SREW is less variable than at LREW (Fig. 2). The lower streamflow complexity at SREW as compared with LREW could be related to the lower complexity of precipitation. Another possible reason of lower streamflow complexity at SREW compared with LREW is that soils at LREW are shallower than at LREW, and the lateral flow at SREW can transmit the part of precipitation to streams with adding the less structure to the flow than the one at LREW.

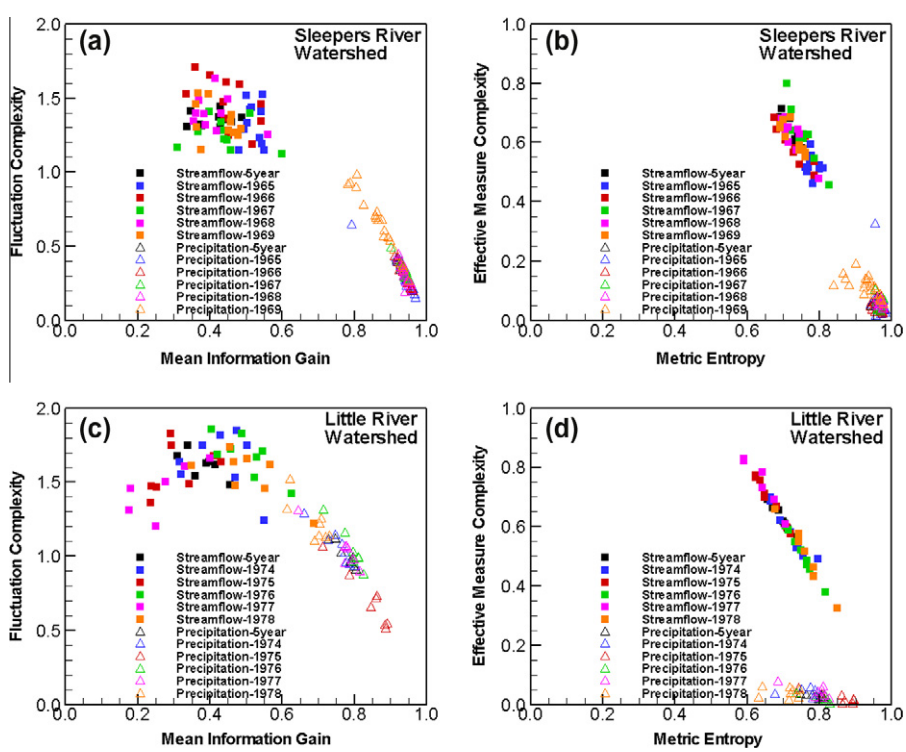

Fig. 3. The information content and complexity measures of 5-year and 1-year daily streamflow and precipitation time series at Sleepers River (a and b) and Little River (c and d) experimental watersheds.

The information content and complexity measures of 5-year time series are generally in the middle of their ranges of 5 yearly time series (Fig. 3). The coefficient of variation (CV) values of the measures in 5 yearly streamflow time series are 0.038, 0.118, 0.089, and 0.067 at sub-watershed W-5 of SREW for metric entropy, mean information gain, effective measure complexity, and fluctuation complexity, respectively. The corresponding values are 0.065, 0.275, 0.120, and 0.132 at sub-watershed W-B of LREW. The small CV values at both watersheds indicate small interannual variability of the measures, thus leading to trivial variation of the streamflow patterns among the 5 years. The CV values of the information content and complexity measures for 5 yearly streamflow time series at LREW are larger than the corresponding data of SREW, indicating the relatively larger variations of streamflow patterns among the 5 years at LREW than the ones at SREW. The average values of the measures from 5 one-year streamflow time series differ from the measures in 5-year time series less than by 3%.

3.2. Temporal scale effects on information content and complexity measures

Fig. 4 illustrates the information content and complexity measures of daily, half-daily, and quarter-daily streamflow time series at SREW and LREW. The information content and complexity measures for 5-year streamflow time series are also listed in Table 1. The mean information gain and metric entropy of streamflow time series increase with the increase of time interval from 6, 12, to 24 h at both watersheds and their sub-watersheds (Fig. 4). For example, at sub-watershed W-B of LREW, the mean information gain and metric entropy of 5-year streamflow time series increase from 0.11 to 0.31, and from 0.56 to 0.66, respectively when the time interval is from 6, 12, to 24 h. As discussed in Session 2, the metric entropy represents the extent of the disorder in the sequence of symbols in a time series and the mean information gain represents the probability of state changes in a time series. Since the large time interval averages the variations in the streamflow time series, this leads to the larger probability of state changes from one to another in the time series. This explains why the information content measures increase with the increase of the time interval. It also illustrates that the relatively smaller ratios among possible state probabilities (i.e., close to uniform distribution) with larger time interval of a time series lead to larger information content measures of the time series.

The effective measure complexity decrease with increase of time interval at both watersheds (Fig. 4). The smaller the time interval of a time series, the larger the effective measure complexity. Since the effective measure complexity describes the minimum information required for an optimal prediction of next symbol, the time series with a smaller time interval is more complex and requires more information for prediction, leading to larger effective measure complexity.

The temporal scale effects on the fluctuation complexity vary among different watersheds. The fluctuation complexity of 5-year streamflow time series increases with an increase of time interval at most sub-watersheds of LREW (Table 1). On the contrary, the fluctuation complexity has close values among the daily, half-daily, and quarter-daily time series at SREW (Table 1). For example, the fluctuation complexity with daily, half-daily, and quarter-daily streamflow time series are 1.68, 1.46, and 1.07 at the sub-watershed W-B of LREW, respectively. The corresponding values are 1.44, 1.53, and 1.43 at the sub-watershed W-5 of SREW, respectively. The fluctuation complexity represents the fluctuations in the system transition from one state to another and is dependent on the transition probability and the ratios of two state probabilities according to Eq. (3). Although the ratios of two state probabilities decrease with the increase of the time interval, the increase of the transition probability could lead to an increase of the fluctuation complexity according to Eq. (3).

Fig. 5 depicts the mean information gain and the effective measure complexity of 5-year streamflow time series at daily, half-daily, and quarter-daily time series at SREW and LREW. The mean information gain increases and effective measure complexity

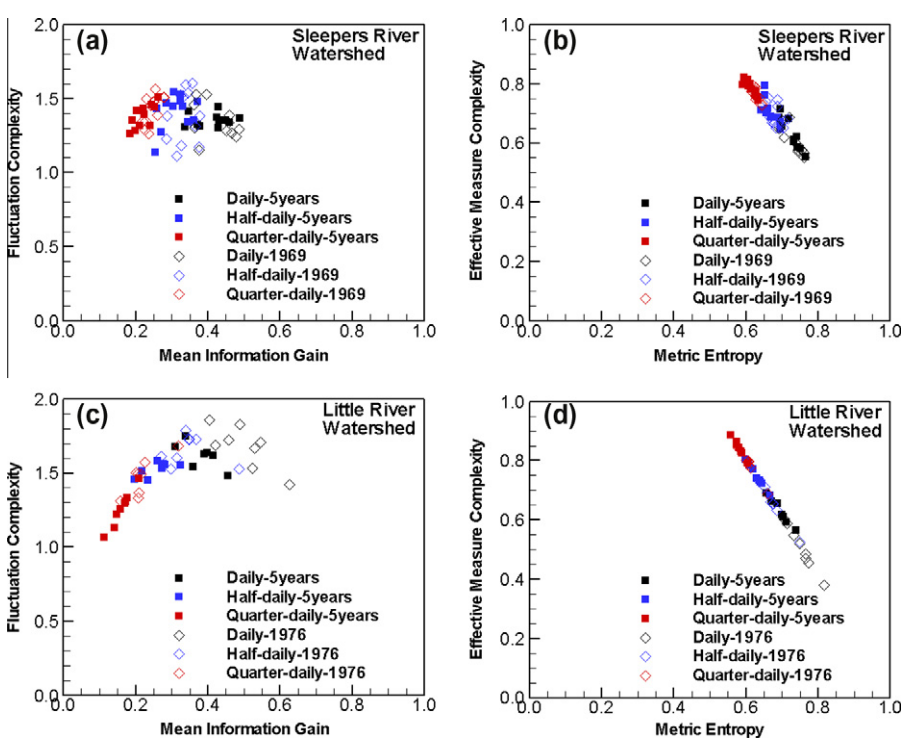

Fig. 4. The information content and complexity measures of daily, half-daily, and quarter-daily streamflow time series at Sleepers River (a and b), and Little River (c and d) experimental watersheds.

decreases as the time scale increases. The temporal scale effects can also be evaluated by the correlation coefficients listed in Table 2. The correlation coefficients between information content measures and time interval are 0.90, and 0.92 at SREW and LREW for the metric entropy, respectively and are 0.87, and 0.92 for the mean information gain, respectively. The strong negative linear relationships between the effective measure complexity and time interval are observed according to their correlation coefficients of −0.79, and −0.93 at SREW and LREW, respectively. The temporal scale effects on the fluctuation complexity are uncertain at different watersheds resulting from the correlation coefficients between the fluctuation complexity and the time interval being −0.26 at SREW and 0.79 at LREW, respectively.

3.3. Spatial scale effects on information content and complexity measures

Fig. 6 shows the mean information gain and the effective measure complexity of 5-year and 1-year daily streamflow time series with sub-watershed areas at SREW and LREW. The mean information gain generally decreases and the effective measure complexity has an increase in trend with the increase of sub-watershed areas at LREW. The correlation coefficients between the information-based measures and the sub-watersheds areas are −0.60, −0.60, 0.59, and 0.27 in 5-year streamflow time series for the metric entropy, the mean information gain, the effective measure complexity, and the fluctuation complexity, respectively, indicating the moderate effects of the spatial scale on the information content and complexity measures. The negative correlation of information content measures and sub-watershed area is in agreement with the observation of Hirpa et al. (2010) that large watersheds have more persistent river flow fluctuations and stronger long memory. In such cases, the distribution of states in symbolic strings should become less even and the information content should decrease. The streamflow response in a watershed is controlled by precipitation input and watershed characteristics such as landscape, geology, geomorphology, and soil properties. For smaller watersheds, more information is transferred from highly random precipitation to streamflow. With increase of watershed areas, these control factors impose more structure on the streamflow time series and result in more complexity.

Much weaker relationships between the information-based measures and the sub-watershed size are observed at SREW. Calculated correlation coefficients were 0.22, 0.30, −0.36, and 0.52 in 5-year streamflow time series for the four measures, respectively (see Table 2). The correlation coefficients also vary among the 5 yearly time series. A possible reason for that may be the presence of shallow soils that decrease the water storage of the sub-watersheds. In addition, in forested shallow soils, substantial flow occurs at the soil bedrock interface (e.g. Peters et al., 1995) and the interaction of precipitation with soils and vegetation may be less pronounced. Modeling applied to this watershed (Wolock, 1995) shows that large variations in responses of streamflow to precipitation in this watershed and its sub-watersheds may be caused by the variations in topography.

We realize that the current application shows the potential of the methodology in hydrological applications rather than presents its fit-for-all version. An essential way forward would be to research and test divisions of data into ordinal categories that are meaningful for hydrology. This direction of research would answer the important question posed by the anonymous reviewer of the earlier version of this manuscript: what will happen when the thresholds between the states are based on more hydrologically relevant distinctions? Starting points could be often made distinctions between driven flow vs. non-driven flow or fast flow/high flow versus base flow/low flow, separation of various hydrograph components, rising limbs, falling limbs, knick points on hydrographs where significance of dominant flow paths change.

Different information theory-based measures represent different aspects of the hydrologic system behaviors in a given hydrologic system. Measures derived from information theory not employed in this work, e.g., transfer entropy, mean mutual

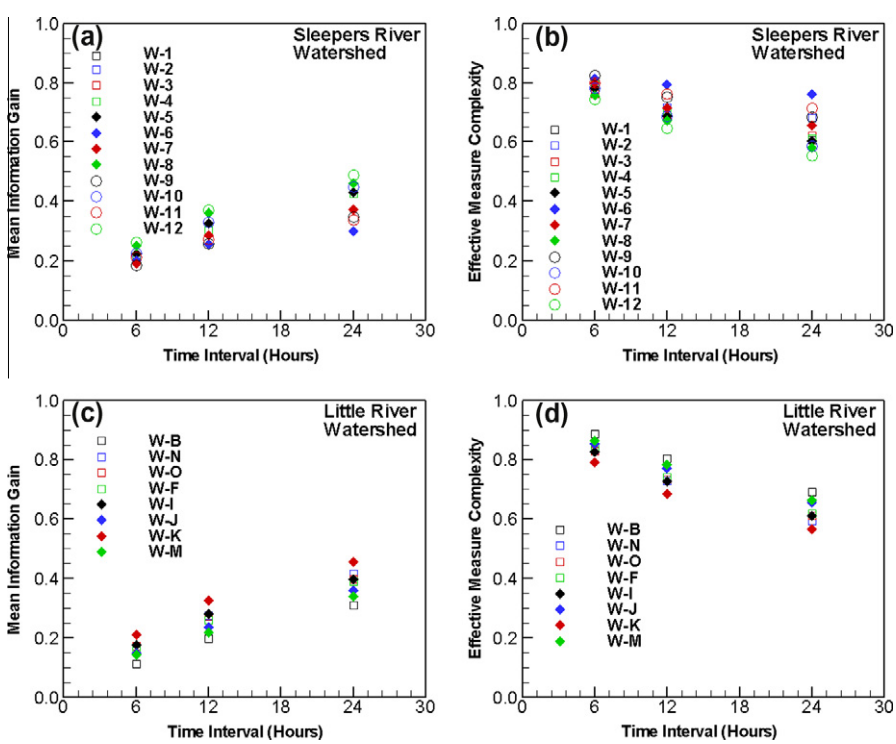

Fig. 5. The mean information gain and effective measure complexity versus time interval of streamflow time series at Sleepers River (a and b), and Little River (c and d) experimental watersheds and their sub-watersheds.

Table 2
The correlation coefficients (r) between information content and complexity measures, and time interval or sub-watershed area for 5-year streamflow time series.

Information metrics	r (Time interval)		r (Sub-watershed area)	
	SREW	LREW	SREW	LREW
Mean information gain	0.87	0.92	0.30	−0.60
Metric entropy	0.90	0.92	0.22	−0.60
Fluctuation complexity	−0.26	0.79	0.52	0.27
Effective measure complexity	−0.79	−0.93	−0.36	0.59

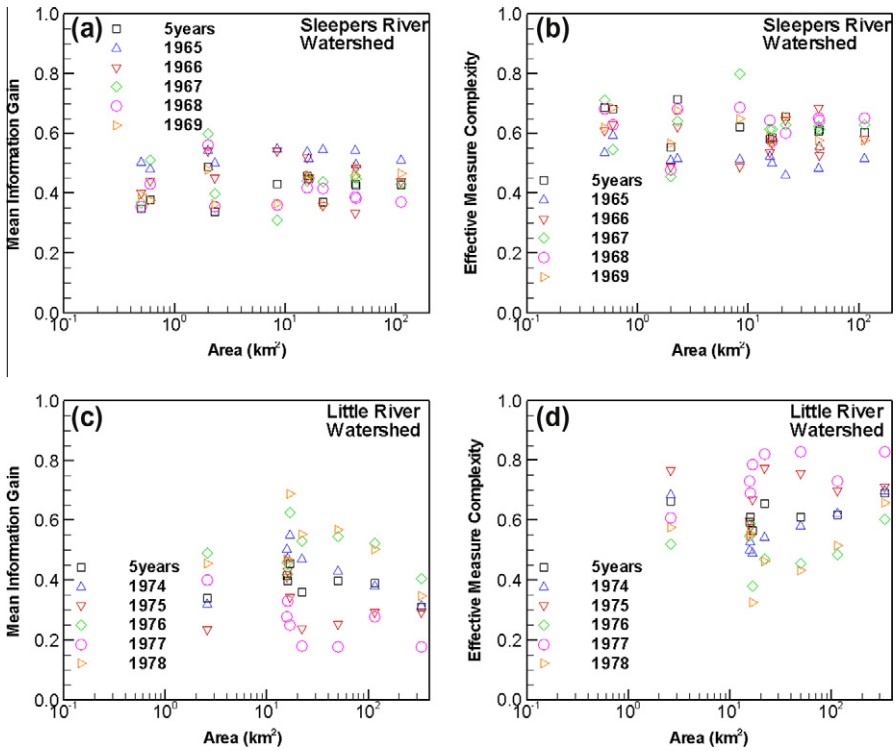

Fig. 6. The mean information gain and fluctuation complexity of streamflow time series versus sub-watershed area at Sleepers River (a and b), and Little River (c and d) experimental watersheds.

information, Rényi entropy, information entropy, relative entropy, maximum entropy, etc., have also been used to measure the information and complexity of a hydrologic time series (Wolf, 1999; Mays et al., 2002; Al-Hamdan and Cruise, 2010; Brunsell, 2010; Singh, 2010a,b; Schreiber, 2000). The transfer entropy (Schreiber, 2000) holds a substantial promise, partly because it is suitable not only for characterization of individual time series but also for establishing relationships between information content and complexity of two or more time series and directions of the information flow between two coupled processes (Kaiser and Schreiber, 2002). Selection of information theory-based measures may depend on the type of study, and the particular time series used in the analysis (Parrott, 2010), and selection of these aspects for streamflow time series presents a meaningful research avenue. Similarly, additional research needs include further exploration of effects of word length and number of symbols in the alphabet on information content and complexity measures, given that these measures have proven to be useful complementary measures in model performance evaluation (Pachepsky et al., 2006).

4. Conclusions

Information content evaluated by mean information gain and metric entropy, and complexity assessed with effective measure and the fluctuation complexity, are analyzed at temporal and spatial scales for two different regions. Results demonstrate the significant temporal scale effects and moderate spatial scale effects on the information content and complexity measures of streamflow.

Comparisons of information content and complexity measures of 5-year daily precipitation and streamflow time series show the relatively lower information content and more complexity in streamflow time series than in precipitation. These results indicate that patterns of streamflow are relatively less random and exhibit higher complexity than the corresponding precipitation data. These data also reveal that precipitation, as one of many forcings driving the hydrologic cycle, is only partially transferred to streamflow through hydrologic processes (e.g., infiltration, surface runoff etc.) in the watersheds, and the watersheds effectively act as filters of the information associated with the precipitation. Conversion from precipitation to streamflow imposes additional structure on the streamflow patterns by controlling factors such as watershed characteristics of landscape, geology, geomorphology, and soil properties, ultimately resulting in higher complexity of streamflow time series than precipitation.

The information content measures increase and the effective measure complexity decreases from quarter-daily, half-daily to daily streamflow time series. Correlation coefficients between the measures and the time interval are around 0.9 (or −0.9), indicating the significant temporal scale effects on the information content and complexity measures and the patterns of streamflow.

The information content measures generally decrease and the trend of complexity measures increases with increase of sub-watershed area at LREW. The correlation coefficients between information theory-based measures and the sub-watershed area have absolute values ranging from 0.27 to 0.60 at LREW, and from 0.22 to 0.52 at SREW, indicating moderate spatial scale effects on the streamflow patterns. Finally, the relationship between the measures and the sub-watershed area also has different characteristics for the two watersheds. Sub-watershed area is one of several spatial scale factors that may have significant effects on streamflow patterns. Other factors include slope, variations in slope, channel lengths, land use, and land cover. All of these factors should be considered and factored in evaluations of the spatial scale effects on the information content and complexity measures and patterns of streamflow. Overall, information theory-based measures of information content and complexity could provide useful complementary knowledge about temporal and spatial patterns and complexity of physical processes in hydrologic systems.

References

Anctil, F., Coulibaly, P., 2004. Wavelet analysis of the interannual variability in southern Québec streamflow. J. Clim. 17 (1), 163–173.

Al-Hamdan, O.Z., Cruise, J.F., 2010. Soil moisture profile development from surface observations by principle of maximum entropy. J. Hydrol. Eng. 15 (5), 327–337.

Baeza Sanz, D., García del Jalón, D., 2005. Characterisation of streamflow regimes in central Spain, based on relevant hydrobiological parameters. J. Hydrol. 310, 266–279.

Black, P.E., 1996. Watershed Hydrology, second ed. CRC Press, Boca Raton, Florida, US.

Bates, J.E., Shepard, H.K., 1993. Measuring complexity using information fluctuation. Phys. Lett. A 172 (6), 416–425.

Bosch, D.D., Sheridan, J.M., Lowrance, R.R., Hubbard, R.K., Strickland, T.C., Feyereisen, G.W., Sullivan, D.G., 2007a. Little river experimental watershed database. Water Resour. Res. 43, W09470. doi:10.1029/2006WR005844.

Bosch, D.D., Sheridan, J.M., Marshall, L.K., 2007b. Precipitation, soil moisture, and climate database, Little River Experimental Watershed, Georgia, United States. Water Resour. Res. 43, W09472. doi:10.1029/2006WR005834.

Bosch, D.D., Sheridan, J.M., 2007. Stream discharge database, Little River Experimental Watershed, Georgia, United States. Water Resources Research 43, W09473. doi:10.1029/2006WR005833.

Brunsell, N.A., 2010. A multiscale information theory approach to assess spatial-temporal variability of daily precipitation. J. Hydrol. 385, 165–172.

Carey, S.K., Tetzlaff, D., Seibert, J., Soulsby, C., Buttle, J., Laudon, H., McDonnell, J., McGuire, K., Caissie, D., Shanley, J., Kennedy, M., Devito, K., Pomeroy, J.W., 2010. Inter-comparison of hydro-climatic regimes across northern catchments: synchronicity, resistance and resilience. Hydrol. Process. 24, 3591–3602.

Chang, W., Fang, B., Yun, X., Wang, S., 2009. The block lossless data compression algorithm. Int. J. Comput. Sci. Network 9 (10), 116–123.

Clausen, B., Biggs, B.J.F., 2000. Flow variables for ecological studies in temperate streams: groupings based on covariance. J. Hydrol. 237, 184–197.

Conrad, P.P., Booth, D.B., 2005. Hydrologic changes in urban streams and their ecological significance. Am. Fish. Soc. Symp. 47, 157–177.

Detenbeck, N.E., Brady, V.J., Taylor, D.L., Snarski, V.M., Batterman, S.L., 2005. Relationship of stream flow regime in the western Lake Superior basin to watershed type characteristics. J. Hydrol. 309, 258–276.

Engelhardt, S., Matyssek, R., Huwe, B., 2009. Complexity and information propagation in hydrological time series of mountain forest catchments. Eur. J. Forest Res. 128, 621–631.

Grassberger, P., 1986. Toward a quantitative theory of self-generated complexity. Int. J. Theor. Phys. 25, 907–938.

Hirpa, F.A., Gebremichael, M., Over, T.M., 2010. River flow fluctuation analysis: effect of watershed area. Water Resour. Res. 46, W12529. doi:10.1029/2009WR009000.

Kahya, E., Dracup, J.A., 1993. US Streamflow patterns in relation to the El Nino/Southern Oscillation. Water Resour. Res. 29 (8), 2491–2503. doi:10.1029/93WR00744.

Kaiser, A., Schreiber, T., 2002. Information transfer in continuous processes. Physica D 166, 43–62.

Lange, H., 1999. Time series analysis of Ecosystem variables with complexity measures. InterJournal for Complex Systems Mauscript #250. New England Complex Systems Institute, Cambridge, MA, US.

Matteau, M., Assani, A.A., Mesfioui, M., 2009. Application of multivariate statistical analysis methods to the dam hydrologic impact studies. J. Hydrol. 371, 120–128.

Mays, D.C., Faybishenko, B.A., Finsterle, S., 2002. Information entropy to measure temporal and spatial complexity of unsaturated flow in heterogeneous media. Water Resour. Res. 38 (12), 1313. doi:10.1029/2001WR001185.

Moliere, D.R., Lowry, J.B.C., Humphrey, C.L., 2009. Classifying the flow regime of data-limited streams in the wet-dry tropical region of Australia. J. Hydrol. 367, 1–13.

Pachepsky, Y.A., Guber, A.K., Jacques, D., Simunek, J., van Genuchten, M.Th., Nicholson, T., Cady, R., 2006. Information content and complexity of simulated soil water fluxes. Geoderma 134, 253–266.

Pan, F., Pachepsky, Y.A., Guber, A.K., Hill, R.L., 2011. Information and complexity measures applied to observed and simulated soil moisture time series. Hydrol. Sci. J. 56 (6), 1027–1039.

Parrott, L., 2010. Measuring ecological complexity. Ecol. Indic. 10, 1069–1076.

Peters, D.L., Buttle, J.M., Taylor, C.H., LaZerte, B.D., 1995. Runoff production in a forested, shallow soil, Canadian Shield Basin. Water Resour. Res. 31 (5), 1291–3104.

Poff, N.L., Ward, J.V., 1989. Implications of streamflow variability and predictability for lotic community structure: a regional analysis of streamflow patterns. Can. J. Fish. Aquat. Sci. 46, 1805–1818.

Poff, N.L., Allan, J.D., 1995. Functional organization of stream fish assemblages in relation to hydrologic variability. Ecology 76, 606–627.

Poff, N.L., 1996. A hydrogeography of unregulated streams in the United States and an examination of scale-dependence in some hydrological descriptors. Freshwater Biol. 36 (1), 71–91.

Richter, B.D., Baumgartner, J.V., Powell, J., Braun, D.P., 1996. A method for assessing hydrologic alteration within ecosystems. Conserv. Biol. 10, 1163–1174.

Sanborn, S.C., Bledsoe, B.P., 2006. Predicting streamflow regime metrics for ungauged streams in Colorado, Washington, and Oregon. J. Hydrol. 325, 241–261.

Schreiber, T., 2000. Measuring information transfer. Phys. Rev. Lett. 85, 461–464.

Shanley, J.B., Kendall, C., Smith, T.E., Wolock, D.M., McDonnell, J.J., 2002. Controls on old and new water contributions to streamflow at some nested catchments in Vermont, USA. Hydrol. Process. 16, 589–609.

Shannon, C.E., 1948. A mathematical theory of communication. AT&T Tech. J. 27, 379–423 (623–656).

Simonovic, S.P., 2009. Managing Water Resources: Methods and Tools for a Systems Approach. UNESCO Publishing, Paris and Earthscan, London.

Singh, V.P., 2010a. Entropy theory for derivation of infiltration equations. Water Resour. Res. 46, W03527. doi:10.1029/2009WR008193.

Singh, V.P., 2010b. Entropy theory for movement of moisture in soils. Water Resour. Res. 46, 03516. doi:10.1029/2009WR008288.

Smith, L.C., Turcotte, D.L., Isacks, B.L., 1998. Streamflow characterization and feature detection using a discrete wavelet transform. Hydrol. Process. 12, 233–249.

Sullivan, D.G., Batten, H.L., Bosch, D., Sheridan, J., Strickland, T., 2007. Little river experimental watershed, Tifton, Georgia, United States: a geographic database. Water Resour. Res. 43, W09471. doi:10.1029/2006WR005836.

USDA-ARS, 2009. ARS water database. <http://www.ars.usda.gov/Main/docs.htm?docid=9696>.

USGS, 2005. Water Resources of New Hampshire and Vermont: Sleepers River Water Research Watershed, Danville, Vermont. <http://nh.water.usgs.gov/projects/sleepers/index.htm>.

Wolock, D.M., 1995. Effects of subbasin size on topographic characteristics and simulated flow paths in Sleepers River watershed, Vermont. Water Resour. Res. 31 (8), 1989–1997.

Wolf, F., 1999. Berechnung von Information und Komplexität von Zeitreihen–Analyse des Wasserhaushaltes von bewaldeten Einzugsgebieten. Bayreuth. Forum Okol. 65, 164 S.