Hydrologic Classification System: A Data Reconstruction Approach

[1]Sivakumar, B., [2]A. W. Jayawardena, and [2]W. K. Li

[1]University of California, Davis, [2]University of Hong Kong, E-Mail: sbellie@ucdavis.edu

Keywords: *Hydrologic systems; Complexity; Classification; Data reconstruction; Phase-space; Attractor*

EXTENDED ABSTRACT

Recent technological advances, such as powerful computers, remote sensors, geographic information systems and worldwide networking facilities, have brought hydrologic research to a whole new level. They have facilitated extensive data collection, better data sharing, formulation of sophisticated methods, and development of complex models to mimic real hydrologic systems. Despite these obvious advances, there are serious concerns about their use in practice and criticisms about our approach to hydrologic modeling. For example: (1) these developments naturally lead to more complex models (having too many parameters requiring too much data) than that may actually be needed; (2) to satisfy the data requirements for such models, we are certainly collecting more and more data, but this does not mean that we are collecting all the relevant data; (3) so, despite their complexity, these models do not perform sufficiently well, even for the situations they are developed for; and (4) since the models are often developed for specific situations, 'translation' of the results to other situations is difficult.

Recent studies have addressed these concerns and criticisms, albeit in different forms, such as dominant processes, thresholds, model integration, and model simplification. A common aspect in these studies is that they recognize the need for a "classification system" in hydrology, so that an appropriate identification as to the model (type) and data requirements can be made. The studies also recognize that, in order to be of general use, such a framework must be: (1) able to provide guidelines for streamlining hydrologic complexity into classes and sub-classes, as appropriate, based on the general/specific information available; and (2) simple enough and commonly agreeable, so that it could provide a "universal" language for communications and discussions in hydrology. They opine that perhaps the identification of dominant governing processes may help in the formation of such a classification system.

The present study explores one potential way to advance this classification system. The exploration involves use of a simple phase-space data reconstruction technique to identify the 'complexity' of hydrologic systems (defined especially in the context of dimensionality of relevant time series). The reconstruction involves representation of the given multi- (often large-) dimensional hydrologic system using only an available (representative) single-variable data series through a delay coordinate embedding procedure. The 'extent of complexity' of the system is identified by the 'region of attraction of trajectories' in the phase-space, which is then used to classify the system as potentially low-, medium- or high-dimensional.

The investigation is carried out in two steps: First, the use of the phase-space concept for system complexity and classification is demonstrated on two artificially generated time series, whose characteristics are known a priori: a high-dimensional purely random series and a low-dimensional deterministic chaotic series. Then, phase-spaces are reconstructed for a host of river flow time series, representing different geographic regions, climatic conditions, river sizes and complexities, and scales. Two specific cases are discussed herein: (1) daily river flow data from different locations; and (2) river flow data of different scales from the same location.

The results for the two artificial time series reveals that direct time series plots and other widely used linear statistical tools (such as autocorrelation function and power spectrum) may not be adequate for studying system complexity and classification. This may be attributed to the inability of these tools to represent the nonlinear properties of the deterministic chaotic series (an inherent property of hydrologic data). The river flow series yield 'attractors' that range from 'very clear' ones to 'moderately clear' to 'very blurry' ones depending on data, indicating the usefulness of this simple phase-space reconstruction concept for studying hydrologic system complexity and classification. The results also reveal the ability of the phase-space to reflect the river basin characteristics and the associated mechanisms, such as basin size, smoothing, and scaling.

1. INTRODUCTION

Hydrologic science has seen an enormous growth, thanks to technological advances that have facilitated collection of extensive data and development of complex models. While continued technological/methodological developments are needed for a more complete understanding, true progress can be evaluated only through a balanced assessment of both positive and negative aspects of these developments. However, encouraged by their positive aspects, we have a tendency to excessively use these technologies and, in the process, to potentially overlook their limitations. For example, there are concerns that: (1) these developments naturally lead to more complex models than that may actually be needed; and (2) the models are often developed for specific situations, making generalization difficult. There are also criticisms that: (1) we may not be collecting all the relevant data, though collecting more data; and (2) the models do not perform sufficiently well, even for the situations they are developed for.

The idea of model simplification is central in many studies, albeit in different forms (e.g. Jakeman and Hornberger, 1993; Grayson and Blöschl, 2000; Sivakumar, 2000); a particularly interesting observation is made by Grayson and Blöschl (2000), who view model simplification and data collection through the prism of "dominant processes." While these studies largely emphasize the possible simplicity in the "apparent" complexity of hydrologic systems, the fact that there also exists "actual" complexity needs careful consideration. Consequently, there is a need for devising a framework that could help in determining the extent of complexity in hydrologic systems, so that an appropriate identification as to the model and data requirements may be made. Such a framework must be able to provide guidelines for streamlining hydrologic complexity into classes and sub-classes, as appropriate, based on the general/specific information available. The framework must also be simple and commonly agreeable, so that it could provide a "universal" language for communications and discussions.

Recent studies raise an increasing concern on the absence of such a framework or a classification system (e.g. McDonnell and Woods, 2004). Such studies suggest that perhaps simplification in data collection and modeling may happen through identification of dominant governing processes and may help in the formation of a classification system. Towards this end, Sivakumar (2004) introduces a classification system based on 'dimensionality' of hydrologic systems (i.e. time series), representing the (number of) dominant processes. The present study explores one potential way to advance this classification using a simple phase-space data reconstruction technique. The basic idea behind this reconstruction is as follows: As a complex looking outcome can *also* be the result of a simple system with a few dominant nonlinear interdependent variables (e.g. Henon, 1976), direct time series plots and linear statistical tools may not be sufficient to identify system complexity. Consequently, one may need to represent the evolution of the system in the form of "trajectories" through reconstruction of the available (often single-variable) data in a multi-dimensional space, so that the "region" of these trajectories (attractor) in the phase-space may be used to obtain useful qualitative information on the extent of complexity of the system dynamics, and may eventually lead to system classification.

With this idea, the potential use of phase-space reconstruction to the study of system complexity (i.e. dimensionality of time series) and classification is investigated in two steps: First, its use is demonstrated on two artificially generated time series, whose characteristics are known a priori: a high-dimensional purely random series and a low-dimensional deterministic chaotic series. Then, phase-spaces are reconstructed for a host of river flow data (representing different conditions and scales) and are interpreted for the underlying system's complexity and classification.

2. DATA RECONSTRUCTION AND SYSTEM CLASSIFICATION

In the "data reconstruction" context, a useful tool for "embedding" the data to represent the system's evolution is the concept of phase-space (e.g. Takens, 1981). Phase-space is essentially a graph, whose coordinates represent the variables necessary to describe the state of the system at any moment. The trajectories of the phase-space diagram describe the evolution of the system from some initial state, and hence represent the history of the system. Phase-space can be reconstructed based on a single- (or multi-) variable series. The idea behind this concept is that a (nonlinear) system is characterized by self-interactions, and that data series of a single variable carries sufficient information about the entire system. Among the methods available for phase-space reconstruction, the method of delays (e.g. Takens, 1981) is the most widely used one. According to this method, with an available single-variable series, X_i, where $i = 1, 2, ..., N$, a multi-dimensional phase-space can be reconstructed as:

$$\boldsymbol{Y_j} = (X_j, X_{j+\tau}, X_{j+2\tau}, ..., X_{j+(m-1)\tau}) \qquad (1)$$

where $j = 1, 2, ..., N-(m-1)\tau$; m is the (embedding) dimension of the vector $\boldsymbol{Y}_j$; and τ is an appropriate delay time. A correct phase-space reconstruction in m allows interpretation of the system dynamics in the form of an m-dimensional map f_T, given by:

$$\mathbf{Y}_{j+T} = f_T(\mathbf{Y}_j) \quad (2)$$

where Y_j and Y_{j+T} are vectors of dimension m, describing the state of the system at times j (current state) and $j+T$ (future state), respectively.

The utility of phase-space reconstruction for system "classification" can be realized from the extent of simplicity/complexity revealed by the phase diagram in the form of the "attractor." For instance, if the attractor is 'clear' (lying on a narrow region), then the dynamics are considered 'simple' and the system low-dimensional. If the attractor is 'blurry' (scattered over a large region), then the dynamics are treated as 'complex' and the system high-dimensional. If the attractor falls somewhere between these two, then the dynamics are assumed to be of 'intermediate complexity' and the system medium-dimensional. If not a convincing 'simple' vs. 'complex' system classification, the reconstruction shows at least the extent of variability within a time series (or difference between two), which is often not possible based on time series plots and linear tools.

To demonstrate the utility of phase-space reconstruction, two 'complex' data sets are artificially generated and studied. Figure 1(a) and (b) presents samples of their variations. It is clear that both exhibit significant variations, with no apparent structure, i.e. both look 'random' and almost indistinguishable. However, the underlying systems (equations) that produce these two sets significantly differ from each other. The first set [Figure 1(a)] is the outcome of a pseudo random number generation function:

$$X_i = rand(\,), \quad (3)$$

which yields independent and identically distributed (IID) numbers. The second [Figure 1(b)], however, is the outcome of a fully deterministic simple two-dimensional map (Henon, 1976):

$$X_{i+1} = a - X_i^2 + bY_i; \quad Y_{i+1} = X_i, \quad (4)$$

which yields irregular solutions for many choices of a and b, but for a = 1.4 and b = 0.3, a typical sequence of X_i is chaotic.

The time series plots are not the only ones that fail to distinguish between the two sets; even the widely used linear tools, such as autocorrelation functions [Figure 1(c) and (d)] and power spectra [Figure 1(e) and (f)] fail as well. This failure is not just in 'visual' or 'qualitative' terms, but also in quantitative terms: for instance, for both sets, the time lag at which the autocorrelation function first crosses the zero line is equal to 1 and the spectral exponent is equal to 0 (indicating randomness in the dynamics of both). It is clear that these linear tools are not sufficient for studies on system complexity, especially when the system possesses nonlinear properties. Consequently, one may need tools that can also represent nonlinear properties. In view of this, these data sets are represented via phase-space reconstruction, and are shown in Figure 1(g) and (h). The phase-space diagrams correspond to reconstruction in two dimensions (m = 2) with delay time τ = 1, i.e. the projection of the attractor on the plane $\{X_i, X_{i+1}\}$. For the first set, the points (of trajectories) are scattered all over the phase-space (i.e. absence of an attractor), a clear indication of a 'complex' and 'random' nature of the underlying dynamics and potentially of a high-dimensional system. On the other hand, the projection for the second set yields a very clear attractor (in a well-defined region), indicating a 'simple' and 'deterministic' nature of the dynamics and potentially of a low-dimensional system. These observations present testimony to the utility of the phase-space reconstruction concept for studying and classifying 'complex' systems.

3. HYDROLOGIC SYSTEMS, COMPLEXITY AND CLASSIFICATION

The utility of phase-space reconstruction is now tested on river flow data. The analysis is presented under two sub-sections, which correspond to: (1) daily river flow data from different locations; and (2) river flow data at different scales from the same location. Due to space limitations, details of these data sets are not reported herein, but relevant references are cited. Also, for brevity, the presentation and interpretation are made only based on time series and phase-space plots. Finally, to facilitate better visualization and comparison, each time series is normalized with respect to its maximum value, so that the data values range between 0 and 1 in all cases.

3.1. Daily Flow: Different Locations

As representations for this case, four flow series, one each from four rivers, are considered: (a) Chao Phraya River in Thailand (Global River Flow Data Center station #2964100); (b) Mississippi River (USGS station #07010000); (c) Kentucky River (USGS station #03284000); and (d) Stillaguamish River in Washington state (USGS station #12167000). These vary in climatic conditions and basin characteristics. For details, see Sivakumar *et al.* (2002) for Chao Phraya River, Sivakumar *et al.* (2004) for Mississippi River and Regonda *et al.* (2004) for Kentucky and Stillaguamish Rivers. Figure 2(a) to (d) shows time series plots of these four series. Except for some noticeable 'peaks' and 'dips,' these series generally show no clear ‘pattern’ in their ‘evolution.’ In short, from a ‘visual inspection’ at least, all four series are complex and irregular, suggesting that the underlying dynamics are random. The data are indistinguishable from a ‘complexity’ perspective and ‘classification’ does not seem possible.

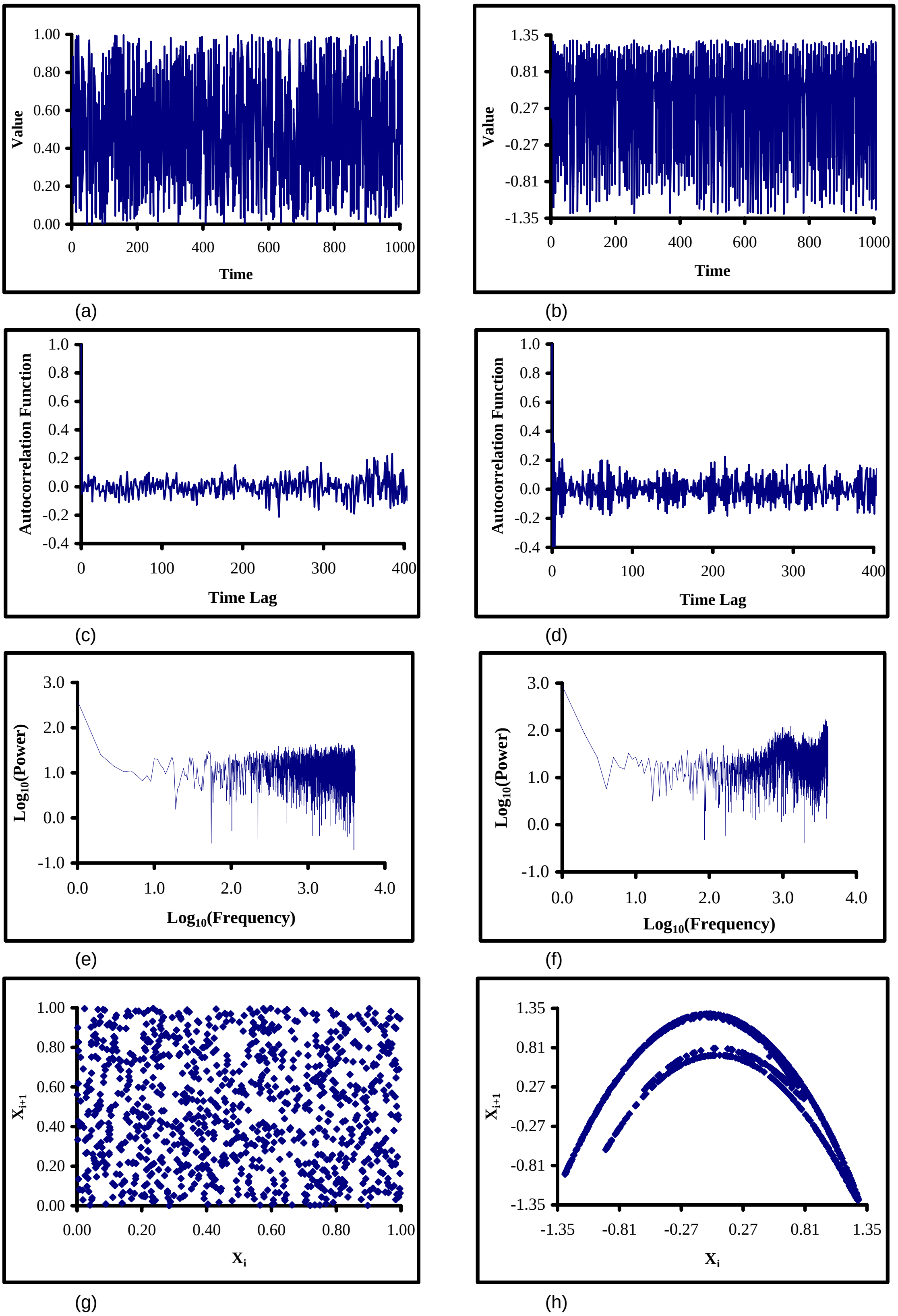

Figure 1. Comparison Between Random and Deterministic Chaotic Sytems: (a) and (b) Time Series; (c) and (d) Autocorrelation Function; (e) and (f) Power Spectrum; and (g) and (f) Phase-space Diagram

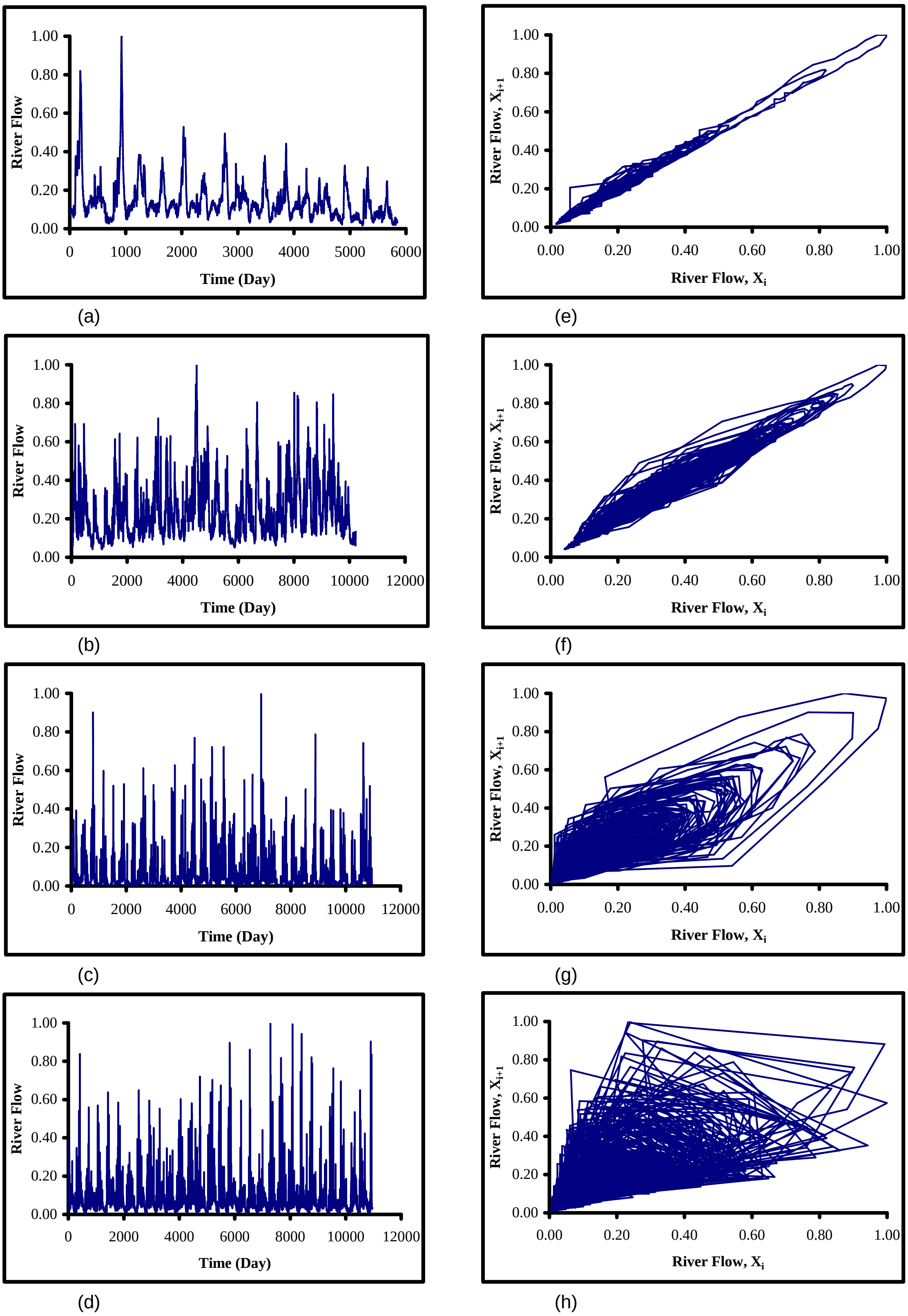

Figure 2. Time Series Plots and Phase-Space Diagrams for Daily River Flow Series: (a) and (e) Chao Phraya River; (b) and (f) Mississippi River; (c) and (g) Kentucky River; and (d) and (h) Stillaguamish River

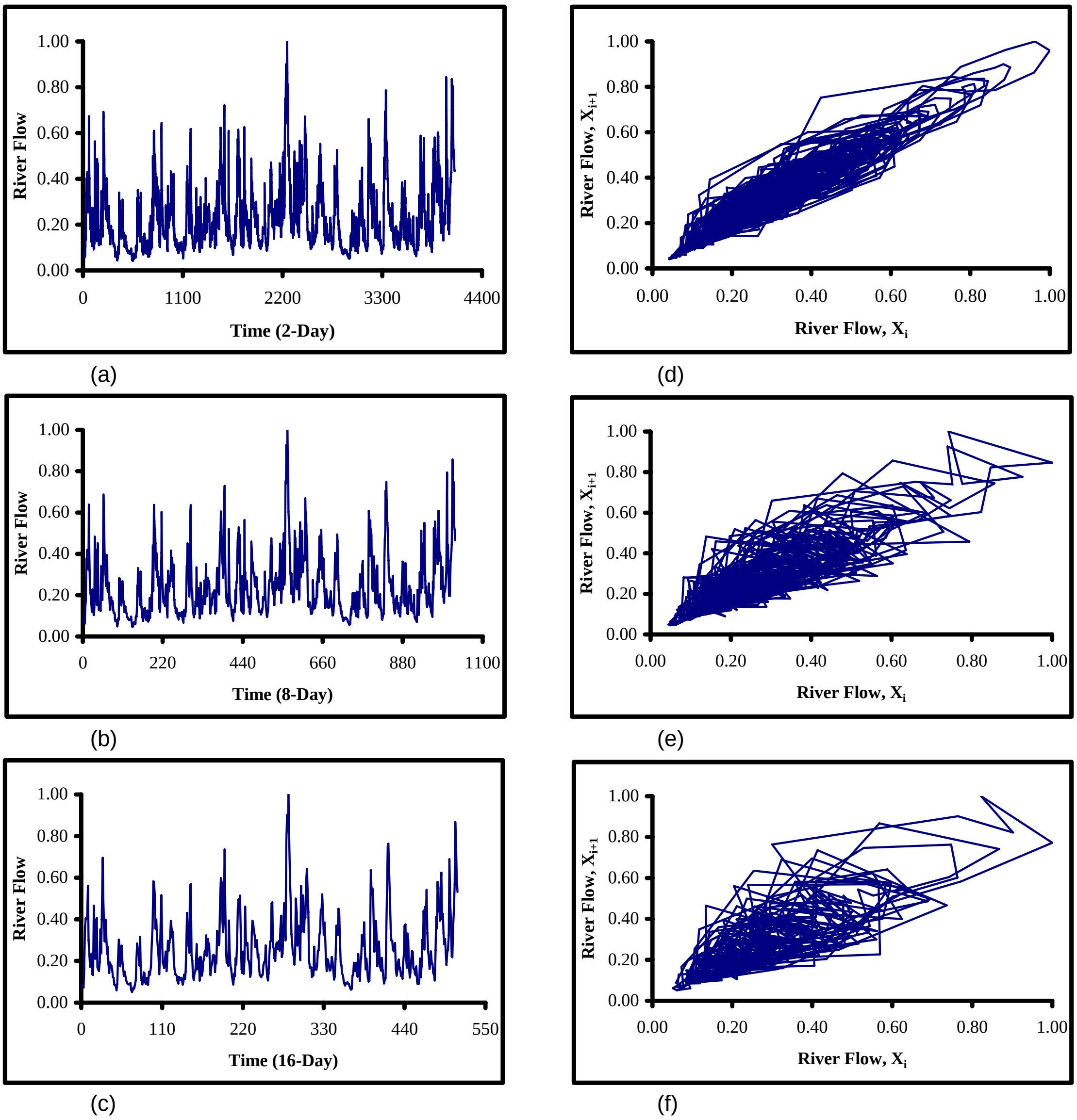

Figure 3. Time Series Plots and Phase-Space Diagrams for Mississippi River Flow at Different Temporal Scales: (a) and (d) 2-day; (b) and (e) 8-day; and (c) and (f) 16-day [see Figure 2(b) and (f) for daily data]

Figure 2(e) to (h) shows phase-space reconstructions of the four series [the points are connected for better visualization of 'evolution']. The differences in the flow dynamics among the four systems are clear. The phase-space diagram for the Chao Phraya River flow exhibits a very clear attractor (flow always evolving within a very well defined 'boundary'), suggesting that the dynamics are simple and predictable and, hence, the system is potentially low-dimensional. The Mississippi River flow series also exhibits a clear attractor in a well-defined region, suggesting possibly simple and predictable dynamics. While the system may be classified as potentially low-dimensional in par with the Chao Phraya River system, the flow dynamics certainly look more complex. The phase-space diagrams for flow from the Kentucky River and the Stillaguamish River show increasing order of complexity. Clearly, the Stillaguamish River flow series, for which points of trajectories are scattered almost all over the phase-space, is highly complex and irregular, suggesting random and unpredictable behavior, and a potentially high-dimensional system. For the Kentucky River flow series, a reasonably clear attractor is still present, though not as clear as that for Chao Phraya River and Mississippi River. Looking at this particular attractor individually, one may interpret that the dynamics fall somewhere between simple and very complex behaviors. Looking at it collectively with the others, one may infer that the dynamics are of intermediate complexity, and the system is potentially medium-dimensional.

3.2. Flow at Different Scales

As river flow at different scales may exhibit different levels of complexity, the utility of phase-space is also studied herein in the context of scale. The analysis is limited only to temporal scale. Flow data corresponding to five different scales from the Mississippi River are considered [same station considered for the daily flow analysis]. These scales range from daily to 16 days at successively doubled resolutions (i.e. daily, 2-day, 4-day, 8-day, and 16-day) [see Sivakumar *et al.* (2004) for details]. Figure 3(a) to (c) shows the time series plots of the 2-day, 8-day, and 16-day flow series, respectively; also see Figure 2(b) for daily series [4-day series not shown for space limitations]. All of these plots show 'peaks' and 'dips', but they do not seem to indicate any clear pattern. While a closer look at these plots indeed reveals certain amount of 'smoothing effect' with temporal aggregation of data, which increases with increasing scale, no obvious pattern that would allow determination of the extent of complexity seems to be present. It may, therefore, be reasonable to say that all of these data sets look just complex and irregular, implying that the underlying dynamics may be random, unpredictable and indistinguishable.

The phase-space reconstructions of these series are shown in Figure 3(d) to (f), respectively; also see Figure 2(f) for daily series. As mentioned above, the daily series exhibits a clear attractor in a well-defined region in the phase-space. As for the remaining series, all of them exhibit reasonably clear attractors as well, suggesting simple and predictable flow dynamics at each of (or across) these scales; however, the region of attraction increases in order with increasing scale of aggregation, with the attractor for the 16-day series being the 'most scattered' among all. Overall, it is reasonable to interpret that the flow dynamics between daily and biweekly scales are simple (may be, approaching intermediate complexity at the biweekly scale). This may be an indication that the systems producing these flows are potentially low-dimensional, or medium-dimensional at worst, at the scale that exhibits the highest complexity. These results also present an interesting observation: that is, increasing scale of temporal aggregation, from finer to coarser, may result in increasing level of complexity in the dynamics (e.g. Sivakumar *et al.*, 2004).

4. CLOSING REMARKS

This study explored the utility of a simple nonlinear phase-space data reconstruction approach for studying hydrologic system complexity (defined by the dimensionality of relevant time series) and classification. The approach was first demonstrated on two artificial time series, and then tested on a host of river flow series, representing different flow conditions and temporal scales. The results for the artificial series revealed the superiority of this approach over direct time series plots and linear statistical tools for system complexity and classification. On the basis of the extent of 'region of attraction' of trajectories in the phase-space, the hydrologic data sets were identified to exhibit 'simple' or 'intermediate' or 'complex' behaviors and, correspondingly, the 'systems' were classified as low- or medium- or high-dimensional. Efforts to advance this classification framework are underway. These include: use of other 'invariants' for verification; and linking the 'dimensions' with the actual physical processes.

5. REFERENCES

Grayson, R.B., and G. Blöschl (2000), *Spatial Patterns in Catchment Hydrology: Observations and Modeling* (ed.), Cambridge University Press, Cambridge, UK.

Henon, M. (1976), A two-dimensional mapping with a strange attractor, *Communications in Mathematical Physics*, 50, 69-77.

Jakeman, A.J., and G.M. Hornberger (1993), How much complexity is warranted in a rainfall-runoff model?, *Water Resources Research*, 29(8), 2637-2650.

McDonnell, J.J., and R. Woods (2004), On the need for catchment classification, *Journal of Hydrology*, 299, 2-3.

Regonda, S., B. Sivakumar, and A. Jain (2004), Temporal scaling in river flow: Can it be chaotic?, *Hydrological Sciences Journal*, 49(3), 373-385.

Sivakumar, B. (2000), Chaos theory in hydrology: important issues and interpretations, *Journal of Hydrology*, 227(1-4), 1-20.

Sivakumar, B. (2004), Dominant processes concept in hydrology: moving forward, *Hydrological Processes*, 18(12), 2349-2353.

Sivakumar, B., A.W. Jayawardena, and T.M.K.G. Fernando (2002), River flow forecasting: use of phase-space reconstruction and artificial neural networks approaches, *Journal of Hydrology*, 265(1-4), 225-245.

Sivakumar, B., W.W. Wallender, C.E. Puente, and M.N. Islam (2004). Streamflow disaggregation: A nonlinear deterministic approach, *Nonlinear Processes in Geophysics*, 11, 383-392.

Takens, F. (1981), Detecting strange attractors in turbulence, In: *Dynamical Systems and Turbulence*, Rand, D.A, and L.S. Young (eds.), Lecture Notes in Mathematics 898, Springer-Verlag, Berlin, Germany, 366-381.