Water Resources Research

RESEARCH ARTICLE
10.1029/2018WR023650

A Nonlinear Dynamical Systems-Based Modeling Approach for Stochastic Simulation of Streamflow and Understanding Predictability

Balaji Rajagopalan[1,2], Solomon Tassew Erkyihun[1,3,4], Upmanu Lall[5], Edith Zagona[1,3], and Kenneth Nowak[6]

[1]Civil, Environmental and Architectural Engineering, University of Colorado Boulder, Boulder, CO, USA, [2]Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado Boulder, Boulder, CO, USA, [3]Center for Advanced Decision Support for Water and Environmental Systems (CADSWES), University of Colorado Boulder, Boulder, CO, USA, [4]Now at Tampa Bay Water, Clearwater, FL, USA, [5]Earth and Environmental Engineering, Columbia University, New York, NY, USA, [6]Bureau of Reclamation, Technical Services Center, Denver, CO, USA

Key Points:
- The dynamics of the multidecadal streamflow signal from long paleo and observed record uncovered by reconstructing the phase space
- Local Lyapunov Exponents are used to understand temporal variability of predictability potentially enabling predictability-based management
- Streamflow simulated by block resampling of trajectories from neighbors in phase space, with skills consistent with predictability

Correspondence to:
B. Rajagopalan,
balajir@colorado.edu

Citation:
Rajagopalan, B., Erkyihun, S. T., Lall, U., Zagona, E., & Nowak, K. (2019). A nonlinear dynamical systems-based modeling approach for stochastic simulation of streamflow and understanding predictability. *Water Resources Research*, *55*, 6268–6284. https://doi.org/10.1029/2018WR023650

Received 9 JUL 2018
Accepted 17 JUN 2019
Accepted article online 26 JUN 2019
Published online 31 JUL 2019

Abstract We propose a time series modeling approach based on nonlinear dynamical systems to recover the underlying dynamics and predictability of streamflow and to produce projections with identifiable skill. First, a wavelet spectral analysis is performed on the time series to identify the dominant quasiperiodic bands. The time series is then reconstructed across these bands and summed to obtain a signal time series. This signal is embedded in a D-dimensional space with an appropriate lag τ to reconstruct the phase space in which the dynamics unfolds. Time-varying predictability is assessed by quantifying the divergence of trajectories in the phase space with time, using Local Lyapunov Exponents. Ensembles of projections from a current time are generated by block resampling trajectories of desired projection length, from the K-nearest neighbors of the current vector in the phase space. This modeling approach was applied to the naturalized historical and paleoreconstructed streamflow at Lees Ferry gauge on the Colorado River, which offered three interesting insights. (i) The flows exhibited significant epochal variations in predictability. (ii) The predictability of the flow quantified by Local Lyapunov Exponent is related to the variance of the flow signal and selected climate indices. (iii) Blind projections of flow during epochs identified as highly predictable showed good skill in capturing the distributional and threshold exceedance statistics and poor performance during low predictability epochs. The ability to assess the potential skill of these long lead projections opens opportunities to perceive hydrologic predictability and consequently water management in a new paradigm.

1. Introduction

With increasing demand for water, such as in the Colorado River Basin (CRB), available water resources have to be managed wisely to address possible supply-demand imbalances. Understanding the variability and predictability of the river flow and the ability to generate realistic flow scenarios is vital in planning and decision making in any river basin.

Linear models or parametric time series modeling techniques have been used traditionally for modeling and simulating time series, especially streamflow (e.g., Salas et al., 1980). Parametric modeling techniques such as Auto Regressive Moving Average model the time series as a sum of mean and random components, with the mean component modeled as a linear function of past values. Simulations from these models reproduce the distributional statistics such as mean, standard deviation, and lag correlations but do not represent non-Gaussian and nonstationary features. Nonparametric time series models such as the K-nearest neighborhood (K-NN) bootstrap (Lall, 1995; Lall & Sharma, 1996; Rajagopalan & Lall, 1998) and kernel density based (Sharma et al., 1997) have been shown to be appropriate for non-Gaussian data but may not adequately model low-frequency variability. For a review of parametric and nonparametric methods for hydrologic time series modeling, we refer the reader to Rajagopalan et al. (2010). Simulation from wavelet spectra has been proposed as a way to model the dominant periodicities (Kwon et al., 2007) and subsequently modified to capture nonstationarity (Nowak et al., 2011).

A nonlinear dynamical system-based time series modeling approach aims to reconstruct the underlying dynamics, which is also referred as a phase space, and exploit it for prediction and simulation. The phase space, or multidimensional space, within which the dynamics is purported to evolve, is reconstructed

from the observed time series (Packard et al., 1980). A reconstructed phase space with appropriate dimension and time delay is a proxy for the true space within which the unknown dynamics of the system unfolds (Takens, 1981). The state of the system at any time point can be mapped on to the phase space, and using local maps (Farmer & Sidorowich, 1987), short-term forecasts are made. The skill of the forecasts depends on the predictability of the current state of the system in the phase space unveiled through Local Lyapunov Exponents (LLEs; Abarbanel et al., 1992; Guégan & Leroux, 2009; Kantz, 1994; Nese, 1989; Wolf et al., 1985). Forecasts from this approach can outperform those from traditional time series approaches (e.g., Casdagli et al., 1990; Elsner & Tsonis, 1992; Grassberger et al., 1991; Regonda et al., 2005; Sangayomi et al., 1996; Tsonis, 1992) if the assumptions for a nonlinear dynamical system generating the process are met. These methods require long time series data and have been applied widely in financial and medical applications (e.g., Kantz & Schreiber, 1997, 1998) and to geophysical data.

One of the early applications to geophysical time series was to model and predict the rise and fall of the Great Salt Lake water levels (Lall et al., 1996; Lall et al., 2006; Sangayomi et al., 1996), forecasting an index of El Niño Southern Oscillation (Regonda et al., 2005) followed by recent hydrologic applications (Kirchner, 2009; Peterson & Western, 2014). These applications, especially of Great Salt Lake water levels, were made possible due to low noise and long time series. For short and noisy time series, filtering is suggested to reduce the noise (Gao et al., 2010; Han et al., 2007; Hansen & Smith, 2001; Jaeger & Kantz, 1996; Porporato & Ridolfi, 1996, 1997; Schreiber & Grassberger, 1991; Smith, 1992) and enable reconstruction of the dynamics.

For noise reduction, wavelet analysis (Torrence & Compo, 1998) is a robust option. We used this to smooth the Colorado River flow to obtain the signal present in the flow series, which is then used in dynamics recovery. Our approach in this paper introduces a blend of K-NN block simulation from the embedding of the system recovered from the wavelet-reconstructed signal of the time series. The embedding requires a determination of the delay time, τ, and embedding dimension, D, through the methods of Average Mutual Information (AMI; Moon et al., 1995) and False Nearest Neighborhood (FNN; Abarbanel & Lall, 1996; Kennel et al., 1992), respectively. The K-NN simulation technique is then applied on the feature vector in the embedding space for the simulation. Details of the embedding and simulation algorithms are presented in the methodology section.

In addition to ensemble forecasts using K-NN, we also identify time series epochs where we can characterize predictability as a function of time through LLE (Abarbanel et al., 1992; Bailey & Nychka, 1995; Guégan & Leroux, 2011). This is similar to the approach (Lall et al., 2006; Moon et al., 2008) where local prediction error criteria such as local generalized cross validation and local generalized cross validation with leverage were introduced and validated as a measure of potential predictability accounting for the predictive error and predictive error accounting for asymmetry of the neighbors, respectively. The LLEs assess how the separation of two initial points in the embedded space diverges or converges exponentially over a finite future time interval. They are thus related to the conditional forecast variance but measure the rate of divergence or change of variance in state space. The cross-validated sum of square of errors of a forecast as a function of forecast lead time is used to assess potential forecast skill. It includes a consideration of forecast bias and variance. The LLEs measure divergence in a d-dimensional state space as the system evolves, and are related to what we expect under forecasting, but are not derived using the forecasting algorithm. They are considered as an intrinsic property of the dynamics that is estimated from the underlying attractor for a given initial condition. Predictability is high when the initial condition is locally stable (LLE < 0, or small positive), and low otherwise. The LLE thus informs the potential predictability at a given system state and hence can inform system management decisions. We provide an application to the Lees Ferry flow data in the CRB and illustrate performance for both predictable and nonpredictable time epochs. There are a number of studies that have used Lyapunov Exponents to assess the chaotic nature of hydroclimate time series and local maps on the reconstructed phase space to enable short term forecasts (to cite a few, see, e.g., Dhanya & Nagesh Kumar, 2010; Elshorbagy et al., 2002; Islam & Sivakumar, 2002; Khatibi et al., 2012; Liu et al., 1998; Rodriguez-Iturbe et al., 1989, and references there in). Here we use LLE for assessing multidecadal predictability, which complements research in literature.

The paper is organized as follows: It starts with a description of the data we used to demonstrate the application of the proposed method; this is followed by the proposed methodology. Then, the results are presented followed by a discussion and summary.

Table 1
Historic and Paleodata Used in the Study and Links to Their Sources

Data	Historic	Paleoreconstructed
AMO (Climate index)	http://www.esrl.noaa.gov/psd/data/timeseries/AMO/	ftp://ftp.ncdc.noaa.gov/pub/data/paleo/treering/reconstructions/amo-gray2004.txt
PDO (Climate index)	http://jisao.washington.edu/pdo/PDO.latest	ftp://ftp.ncdc.noaa.gov/pub/data/paleo/treering/reconstructions/pdo-macdonald2005.txt
Lees Ferry (Streamflow)	http://www.usbr.gov/lc/region/g4000/NaturalFlow/current.html	https://www.treeflow.info/

2. Study Data Sets

Historic and paleo data used in the study are listed in Table 1 along with links to their sources and are described below. The primary data set is the naturalized streamflow at the Lees Ferry gauge on the CRB. This is an important gauge on the river through which 85–90% of the flow in the basin passes. Naturalized flow developed by removing anthropogenic reservoir effects (regulation and consumptive use) is maintained by U.S. Bureau of Reclamation (Prairie & Callejo, 2005) and is available for the historic period 1906 to 2015 at monthly timescale. The annual values are calculated as average of the monthly values. Paleoreconstructed flow (Woodhouse et al., 2006) is available at annual timescale for years 1490 to 1997. The reconstructions are based on tree ring chronologies and capture ~80% of the variance of the historical data at several locations on the CRB. We combined the two data sets—paleoreconstructed flows for the 1490–1905 period and the historic natural flows for 1906–2016—to obtain a long flow time series spanning 1497–2016, which is used for analysis in this research.

The Atlantic Multidecadal Oscillation (AMO) and the Pacific Decadal Oscillation (PDO) have been shown to influence the low-frequency variability of the Colorado River flow (e.g., Bracken et al., 2014; Erkyihun et al., 2016; Nowak et al., 2012). We use these two climate indices to understand the predictability of the Lees Ferry flow. The AMO index (Enfield et al., 2001) is computed as a monthly area weighted average of North Atlantic (0 to 70°N) sea surface temperatures (SSTs), which is subsequently detrended based on 5° × 5° resolution Kaplan SST (Kaplan et al., 1998). Values were accessed from the NOAA Physical Sciences data website for the period 1856 to present. The paleoreconstruction of annual AMO for the period 1650 to 1990, based on reconstructions of annual SST anomalies for the North Atlantic Ocean (0 to 70°N) from tree rings (Gray et al., 2004), was obtained from the NOAA website. Monthly PDO anomalies from 1900 to present are available from University of Washington. The annual data were taken as the average of the monthly values in this analysis. The PDO is calculated as the first principal component of the Northern Pacific SST (Mantua et al., 1997; Zhang et al., 1997). Annual PDO values for the period 993 to 1996, based on tree rings from Pinus flexilis in California and Alberta, Canada, were generated by MacDonald and Case (2005) and are available from the NOAA website.

3. Modeling Approach

Unlike traditional stochastic methods that consider the observed process to be a combination of a mean (i.e., signal) and a random component, the nonlinear dynamical approach considers the observed time series as realizations of a dynamical system. Since the dynamical system is unknown, its recovery from the time series involves reconstructing the phase space from observations. Theoretical results (e.g., Packard et al., 1980; Takens, 1981) suggest a geometrical correspondence between appropriately reconstructed phase space from the observations and the true space in which the unknown dynamics unfolds. The geometry of the reconstructed space can provide insights into predictability, and local maps in the embedded space can be developed for skillful prediction and simulation. The proposed approach consists of four steps: (i) streamflow *signal* extraction, (ii) computation of time delay and embedding dimension, (iii) predictability estimation via local Lypunov exponents, and (iv) ensemble simulation. These steps are described below.

3.1. Streamflow Signal Extraction

Wavelet methods are widely used to decompose time series into signal and noise. An accessible description of these methods is provided by Torrence and Compo (1998) in their review paper. Suppose $x(t), t = 1,2,...,N$ is

the streamflow time series. The signal extraction from the time series involves four steps: (i) compute the wavelet spectrum of the time series, $x(t)$, (ii) identify spectral bands where the global variation meets a statistical significance test of being different from the background noise, (iii) reconstruct or project the time series on to each of the bands that is significant, and (iv) sum each of the reconstructed series to recover the signal. The Morlet is the preferred wavelet function for its boundary properties and its simplicity for discrete time series applications. We used this wavelet to decompose the streamflow time series and for signal extraction. This approach for signal extraction has been applied to modeling and simulating the Colorado River flow series (Erkyihun et al., 2016; Nowak et al., 2011) and for rainfall and temperature in Florida (Kwon et al., 2007). We refer the reader to these papers and the review paper by Torrence and Compo (1998) for details on wavelet analysis and signal extraction. The main steps in obtaining the reconstructed signal are described below. Suppose that the frequency band of interest is defined as the interval $\{j_1, j_2\}$, then the reconstructed time series within this band is obtained as

$$X_t^{'} = \frac{\delta_j \delta_t^{1/2}}{C_\delta \psi_0(0)} \sum_{j=j_1}^{j_2} \frac{R\{W_t(a_j)\}}{a_j^{1/2}} \tag{1}$$

where C_δ is a reconstruction factor and δ_j and δ_t are the scale-averaging coefficient and time factors, respectively. $\psi_0(0) = \pi^{-1/4}$ is the factor that removes the energy scaling specific to the Morlet wavelet function. $R\{.\}$ denotes the real part of W, the wavelet transform of the streamflow data, and a_j is the scale; j_1 and j_2 are the lower and upper scales, respectively. The temporal variance of the reconstructed band is quantified by the Scaled Averaged Wavelet Power (SAWP) and is given by

$$\overline{W}_t^{\,2} = \frac{\delta_j \delta_t}{C_\delta} \sum_{j=j_1}^{j_2} \frac{|W_t(a_j)|^2}{a_j} \tag{2}$$

Reconstruction is done for all the significant bands using the above procedure, and they are summed to obtain the signal of the original time series. Similarly, the temporal variances are summed to obtain the SAWP of the signal.

3.2. Time Delay and Embedding Dimension

Suppose $x'(t)$,$t = 1,2,...,N$ is the streamflow signal time series obtained from the wavelet decomposition and extraction from the previous step. A D-dimensional embedding of this signal time series can be obtained using a time delay of τ, which can be written as

$$\boldsymbol{y}_t = x'(t), x'(t+\tau), ..., x'(t+(D{-}1)\tau); t = 1, ..., \mathrm{N}{-}(\mathrm{D}{-}1)\tau \tag{3}$$

The two parameters D and τ are estimated from the signal time series.

Real data sets are noisy (with dynamical and measurement errors). Therefore, estimates of the parameters can be unreliable (Schreiber & Kantz, 1996) leading to phase space that has poor skill in forecast and simulation. Smoothing can reduce the noise (Porporato & Ridolfi, 1996, 1997; Schreiber & Grassberger, 1991), but if not done properly, it can alter the underlying dynamics (Sivakumar et al., 1999). Hence, we proposed to first smooth the streamflow data using wavelets to extract the signal—described in the previous step—and then estimate D and τ for the signal.

The delay time τ represents the average length of the memory in the system and can be considered as the decorrelation timescale. This can be estimated from mutual information (MI) at various lags using two dimensional histograms (Fraser & Swinney, 1986) or kernel density estimators (Moon et al., 1995). This is tantamount to computing the nonlinear autocorrelation of a time series at various lags. The AMI at a lag m is estimated as

$$I_m = \sum_{k=1}^{N-m} P\left(x'(k), x'(k+m) log_2 \left[\frac{P(x'(k), x'(k+m))}{P(x'(k))P(x'(k+m))}\right]\right. \tag{4}$$

where $P(x'(k),x'(k+m))$ is the joint probability density function (PDF) and $P(x'(k)),P(x'(k+m)$ are the marginal PDFs, estimated using kernel density estimators, and N is the number of observations. If τ is too

small, components in a delay vector are nearly identical, such that adding new components does not provide new information. On the other hand, if τ is too large, successive components are totally unrelated. Popular choices for τ include the first zero crossing of the autocorrelation function and the first minimum of mutual information function (Fraser & Swinney, 1986). Here the average MI is computed for various lags, and the time delay τ is chosen as the lag of the first minimum of the AMI (Abarbanel & Lall, 1996). Typically, the effectiveness of a model is not highly sensitive to the choice of τ (Kantz & Schreiber, 1997) within the range of minimum AMI or zero crossing of autocorrelation.

Two popular methods for estimating the embedding dimension, D, are the Grassberger-Procaccia (Grassberger & Procaccia, 1983a, 1983b) approach, which estimates the dimension mostly as *fractal* or non-integer, and the False Nearest Neighbor (FNN) method (Kennel et al., 1992), which computes an integer dimension suitable for embedding. Here, we apply the FNN technique because it estimates the dimensions as integers, suitable for determining the number of dimensions. This approach assumes that there is a minimum sufficient dimension D that guarantees a one to one relation to the true/original space. For any dimension $D_0 < D$, the one to one map to the original space will not be preserved because two or more points on the original space will be mapped onto the same point (false neighbor) in the reconstructed D_0 dimensional space. In other words, there is not enough room to reconstruct the one to one correspondence. In this approach, the time series $x'(t), t = 1,...,N$ is successively embedded in a D_0 dimensional space. For each vector at a current level of embedding, k-nearest neighbors are identified. Then a check is done as to how many of these nearest neighbors are still nearest neighbors in the D_{0+1} space. Those that fail to be neighbors are called false neighbors. For each D_0, the fraction of false neighbors is computed as

$$r = \frac{\left\|\boldsymbol{y}_{i+1} - \boldsymbol{y}_{j+1}\right\|}{\left\|\boldsymbol{y}_i - \boldsymbol{y}_j\right\|}; r \leq r_t \tag{5}$$

where $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$ are two nearest neighbors in the D_0 dimensional space and $\|\boldsymbol{y}_i - \boldsymbol{y}_j\|$ is the distance between them. $\|\boldsymbol{y}_{i+1} - \boldsymbol{y}_{j+1}\|$ is their distance in dimension D_{0+1}. If the ratio $r > r_t$, where r_t is a heuristic threshold (Kantz & Schreiber, 1997; Kennel et al., 1992), then D_0 is increased and the computation is repeated until the ratio condition is met or D_0 exceeds a reasonable value, in which case the conclusion is that the time series may not correspond to a low dimensional dynamical system. In this study r_t is set to be 10. This threshold declares two neighbors in the current embedding space as false neighbors if their distance exceeds 10 times of that in the previous embedding dimension.

3.3. Predictability: Local Lyapunov Exponents

With the time delay embedding of the time series in the D-dimensional space (equation (3)) and plotting them results in reconstructing the phase space of the system. Since each point in the D-dimensional phase space corresponds to a time, the phase space represents the temporal evolution of the underlying process of the time series. Predictability of the system is quantified using Lyapunov exponents (λ). A Lyapunov exponent (LE) measures the rate at which the initial separation of two points in the phase space ($\delta\boldsymbol{y}_0$) grows or shrinks after an evolution of L time steps. If the separation grows, it indicates that the two initial points follow trajectories that are diverging, indication that the system is unpredictable at L time steps. As mentioned, the rate of growth, LE, is the measure of predictability. On the other hand, if the separation shrinks, the trajectories of the initial points are converging, indicative of good predictability. The Lyapunov exponents $\lambda_d(L)$ can be computed for each dimension d ($d = 1, \ldots D$) and for several evolution periods L. For an evolution period L and dimension d, LEs are calculated from all points in the phase space, and the average LE is computed. Negative values of the exponents indicate convergence of the trajectories toward a locally stable point and thus good predictability, while positive values indicate divergence of trajectories and less predictability. The maximum value of the exponents among all dimensions will provide the predictability limit at a desired evolution period L. For details on the calculations of LE we refer the readers to Abarbanel et al. (1992, 1993), Bailey & Nychka (1995), Abarbanel and Lall (1996), Oseledec (1968), and Wolf et al. (1985).

The average LE described above provides an estimate of average predictability of the system over the entire phase space. However, it is well known that predictability varies across the phase space, and thus, a single LE cannot capture the variability in predictability. To address this, a LLE may be computed over a finite time starting from any point in the time series, which corresponds to a vector in the embedded space. LLE

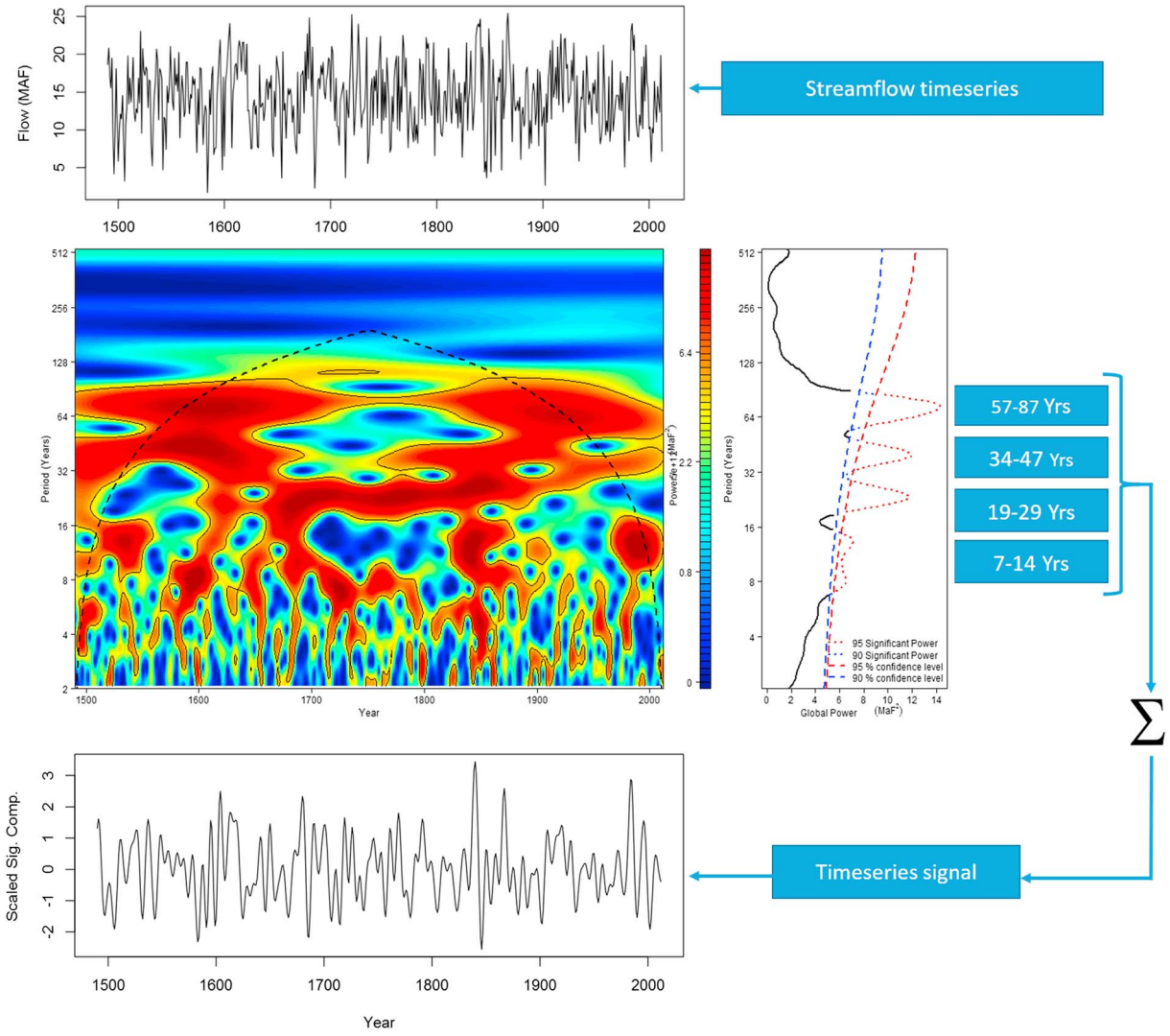

Figure 1. The time series of the reconstructed and observed Lees Ferry Flows is in the top panel. Wavelet spectra of paleoreconstructed and observed Lees Ferry flow are in the middle panel along with the global spectrum to its right. Periods that are significant at 90–95% confidence level are indicated. Signal time series obtained as the summation of band-filtered components is shown in the bottom panel.

considers the evolution forward from nearby points in the embedded space to estimate the divergence or convergence of trajectories in the embedded space corresponding to that initial condition. The LLE, $\lambda_d(\boldsymbol{y}, L)$, can be computed for each point $\boldsymbol{y}$ in the phase space and for any corresponding time step, for a finite evolution period, L, and for each embedding dimension d ($d = 1, \ldots D$). Thus, LLE can be obtained for each L and d. For details on the calculations we refer the readers to the above references. The LLEs provide an estimate of predictability at each point in the phase space (i.e., at each time) for a given L and d, offering rich insight into the temporal variability of the predictability, which will be of immense help in better understanding and modeling the system.

3.4. Ensemble Simulation

The simulation involves identifying nearest neighbors in the phase space of the signal and resampling a flow trajectory. The implementation at any time t proceeds as follows:

(1). Construct the phase space of the flow signal using the embedding dimension and delay obtained from the methods described above.

(2). For the time t, the *feature vector*, $\boldsymbol{y}_t$, that represents the current state of the system in the embedded phase space is selected; $\boldsymbol{y}_t$ represents a point on the attractor.

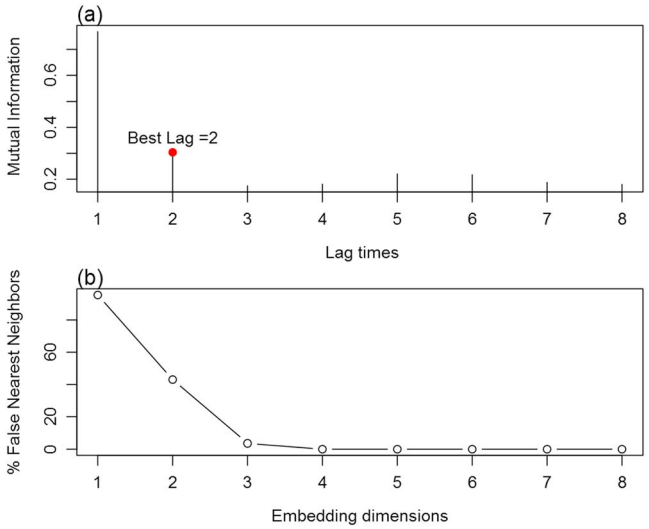

Figure 2. (a) Average Mutual Information of the signal time series of Lees Ferry flow, corresponding to various lag time (in years). (b) Percentage of False Nearest Neighbors corresponding to various embedding dimensions of the signal time series of Lees Ferry flow.

(3). K-nearest neighbors (K-NN) of the feature vector in the embedded phase space are selected. Weights are assigned to the neighbors using a weight function that gives highest weight to the nearest neighbors and least to the farthest (Lall & Sharma, 1996; Rajagopalan & Lall, 1999). A heuristic value of $K = \sqrt{N}$ is shown to work well.

(4). One of the nearest neighbors is selected using the weights. Suppose the selected neighbor in the phase space corresponds to a point in time, *j*.

(5). The sequence or trajectory of the actual flow from time *j+1* to *j+M,* i.e., *x*(*j*+1) to *x*(*j*+*M*) where *M* is the desired length of simulation is selected to be the simulated flow sequence. This can be viewed as a conditional block bootstrap (Efron & Tibishirani, 1993).

(6). This process is repeated 500 times to produce ensembles of trajectories (i.e., projections) of length *M*. Note that this is a blind simulation technique in that only data prior to time *t* are used in the simulation.

Steps 5 and 6 differ from the traditional procedure wherein a local map (e.g., local polynomial), fitted between the K-NN of the phase space vectors and their one-step ahead successors, is used to provide a one-step projection or forecast of the time series (e.g., Asefa et al., 2005; Casdagli et al., 1990; Farmer & Sidorowich, 1987; Grassberger et al., 1991; Lall et al., 1996, 2006; Regonda et al., 2005; Salas et al., 1980; Tsonis, 1992; Wei, 2006). Our block bootstrapping approach allows for a multistep projection or forecast while preserving the trajectory continuity and accounting for the spread across trajectories. This overcomes the problem associated with numerical diffusion that is often associated with iterated one-step ahead forecasts. Ensemble forecasting by using a suite of D and τ have been used for forecasting daily rainfall (Dhanya & Nagesh Kumar, 2010) and using wavelet network model (Dhanya & Nagesh Kumar, 2011). With high degree of noise in daily hydrologic data applied in the above studies, estimates of D and τ can be highly variable; thus, this ensemble approach helps to stabilize the variance in the forecasts. In our modeling approach here, using wavelet-filtered signal data, estimates of D and τ are robust, enabling the application of block resampling to generate ensembles.

4. Results

The nonlinear dynamics system identification, predictability, and simulation are demonstrated on the flow at Lees Ferry in the CRB. Wavelet analysis of the long paleoreconstructed and historical Lees Ferry flow was performed, and time-varying and global spectra are shown in Figure 1. Four spectral bands are found to have variation that is significantly different from background noise at the 95% confidence level: 57–87, 34–47, 19–29, and 7–14 years. Furthermore, these spectral bands exhibit interesting temporal modulation, for example, the 57–87- and 34–47-years bands are dominant during 1500 to 1650 and post 1850; the 19–29-years band is present mainly during 1650 to 1850; and the decadal band of 7–14 years waxes and wanes throughout the length of the record with a strong presence in recent decades. By summing the information in these bands together, we are recovering the joint variation across different organized sources of information that have a low-frequency character and drop the higher-frequency phenomena. The question, then, is whether this low-frequency kernel modeled using nonlinear dynamics is able to provide insights into the low-frequency evolution of the system into the future; that is, given that we used signals with periods from 7 to 87 years, does the modeling approach presented here inform the evolution of the mean flow and/or its variance over the next decade or two? This is what we explore in this section.

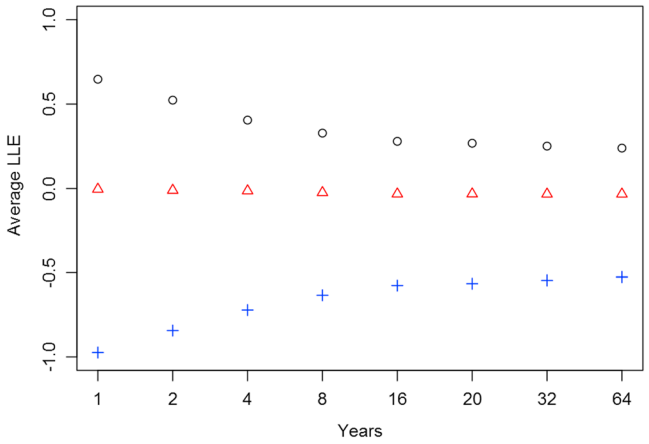

Figure 3. Average Local Lyapunov exponents (LLEs) of the Lees Ferry flow signal for the three dimensions evaluated at various time steps or scale (also referred to as *L*). Each symbol corresponds to a dimension.

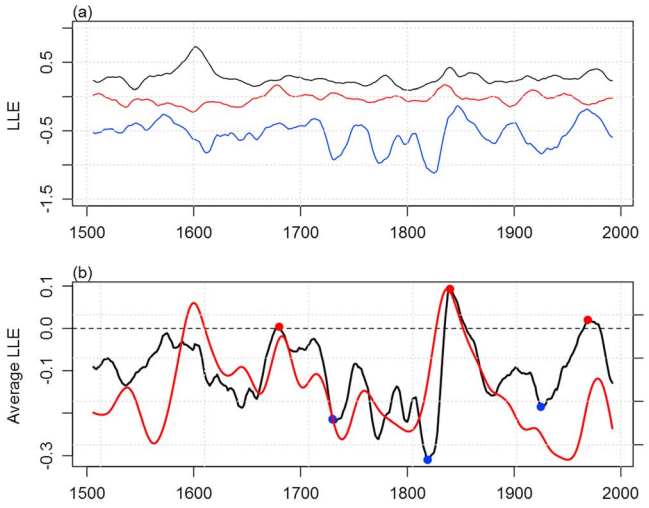

Figure 4. LLEs at each year with a time step of 20 years (a) for the three dimensions for the Lees Ferry flow signal and (b) average LLEs at each year (average of the LLE of the three dimensions in (a)) for the Lees Ferry flow signal (black) and the SAWP of the flow signal (red). The value of the exponent at each year corresponds to the average of following 20-year period —similarly for the SAWP. The selected epochs of high predictability 1731–1750, 1820–1839, and 1926–1945 are shown as blue dots, and the low-predictability epochs 1681–1700, 1841–1860, and 1970–1989 as red dots.

The AMI computed at various lags using the kernel density-based estimators (Moon et al., 1995) is shown in Figure 2a. The first minimum is at lag 3, but we found a lag of 2 to be good at reconstructing the phase space (discussed later)—thus, we selected a time delay τ of 2 years (Figure 2a). As mentioned earlier, the model is generally insensitive to the choice of τ within this range of low AMI (Fraser & Swinney, 1986). To identify the best embedding dimension, D, the percentages of FNNs were computed for various embedding dimensions and the selected D is shown in Figure 2b. The FNN is almost zero for dimension of 3 and zero for 4. We tried both and found no significant difference in the simulation skill and Lyapunov exponents, so we chose D of 3 to be parsimonious.

LLEs of the Lees Ferry flow signal for various time steps or scales (referred to as L) are shown in Figure 3. Note that the exponents remain constant after L of 16 years. The average exponents corresponding to a large L, say 64 years, are the global Lyapunov exponents—which are 0.27, −0.03, and −0.53, respectively, in the three dimensions. Here $\lambda_1 > 0$, which is an indicator of chaos and the first dimension (or direction), inhibits predictability; λ_2 close to 0 tells us that this system can be modeled by a set of differential equations; and $\lambda_3 < 0$ suggests that the third dimension provides predictability. The predictability is dictated by the highest Lyapunov exponent, as that leads to divergence of trajectories and, consequently, reduced predictability (Abarbanel & Lall, 1996). The average global Lyapunov exponent over the three dimensions for the entire phase space is −0.1, indicating that the system is conservative. The value of the largest exponent $\lambda_1 = 0.27 = 1/3.7$ in units of year^{-1} suggests that on an average the errors along the orbit or initial conditions grow as exp[t/3.7], so that after around 4 years or so, the predictability drops rapidly

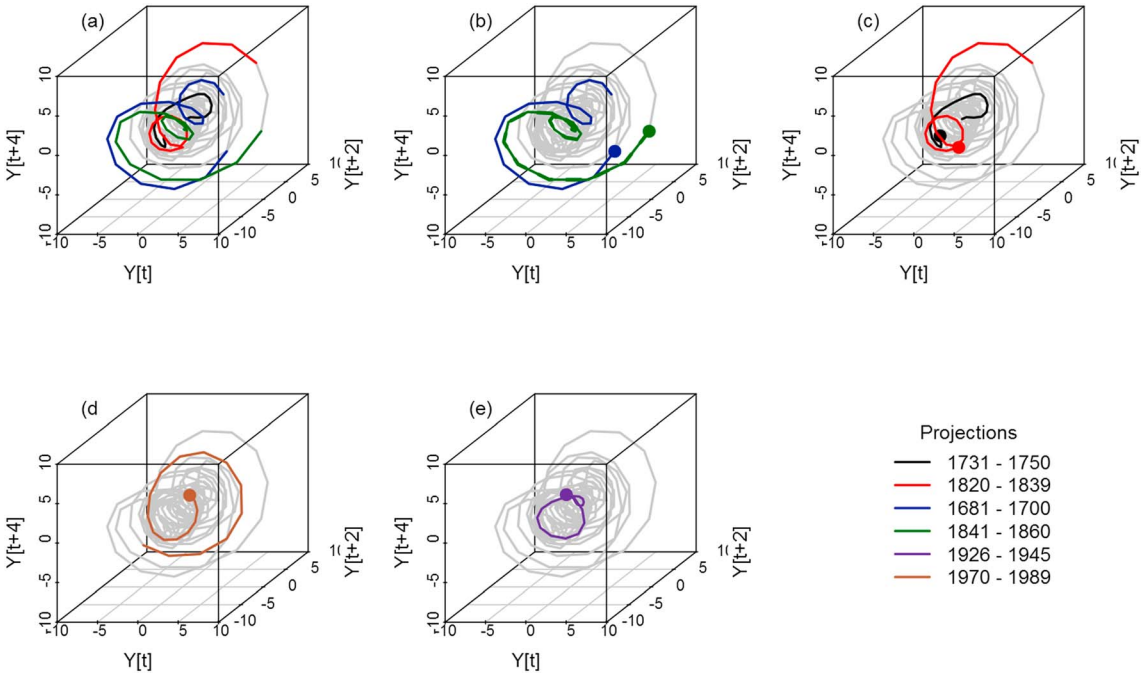

Figure 5. (a) Trajectories of epochs shown in the phase space—high predictable epochs 1788–1807 (red) and 1820–1839 (black), low predictable epochs 1681–1700 (blue) and 1841–1860 (green), and twentieth century epochs 1926–1945 (magenta) and 1970–1989 (brown). (b) and (c) show the trajectories of low and high predictable epochs, respectively, with the starting points of the trajectories shown as dots. (d) and (e) show the trajectories of twentieth century epochs 1970–1989 and 1926–1945, respectively, with the starting points shown as dots.

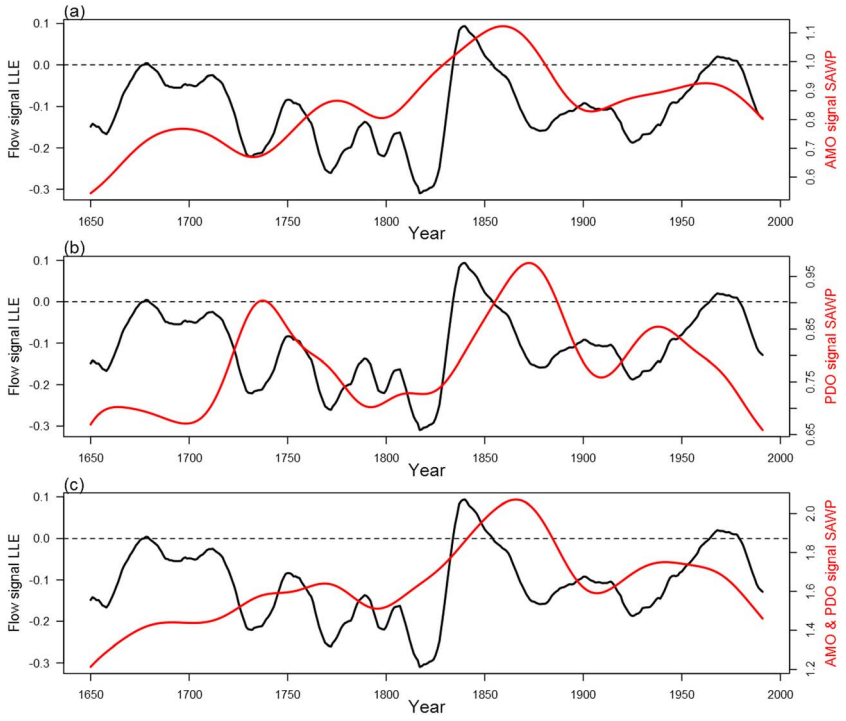

Figure 6. (a) Average LLEs of the Lees Ferry signal (black) and the SAWP of the PDO signal (red). (b) Same as (a) but with SAWP of the AMO signal (red). (c) Same as (a) but with sum of SAWP of AMO and PDO signals (red).

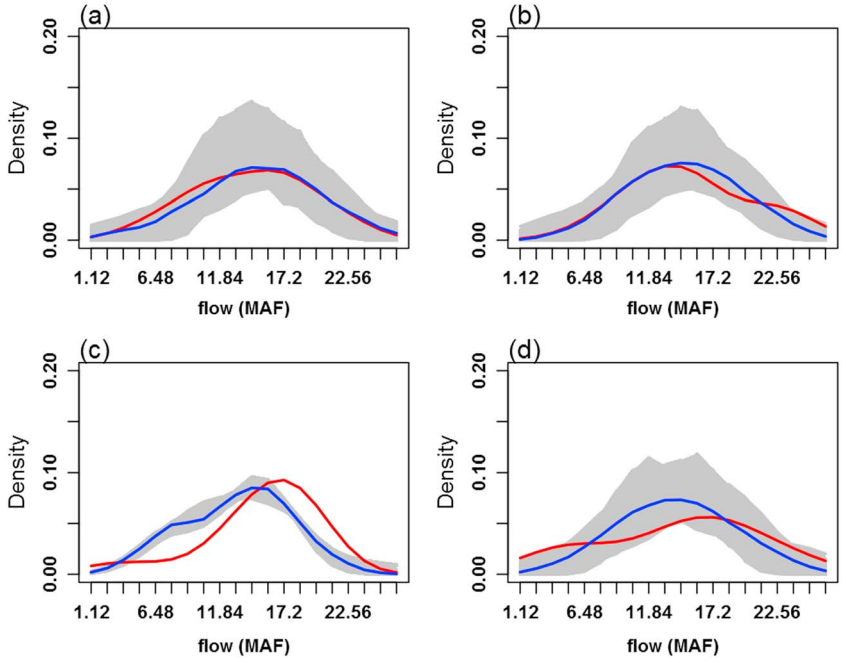

Figure 7. Probability density functions (PDFs) of flow projection ensembles (gray), their median (blue), and that of historic flows (red) for the high-predictability epochs (a) 1731–1750 and (b) 1820–1839 and low predictability epochs (c) 1681–1700 and (d) 1841–1860.

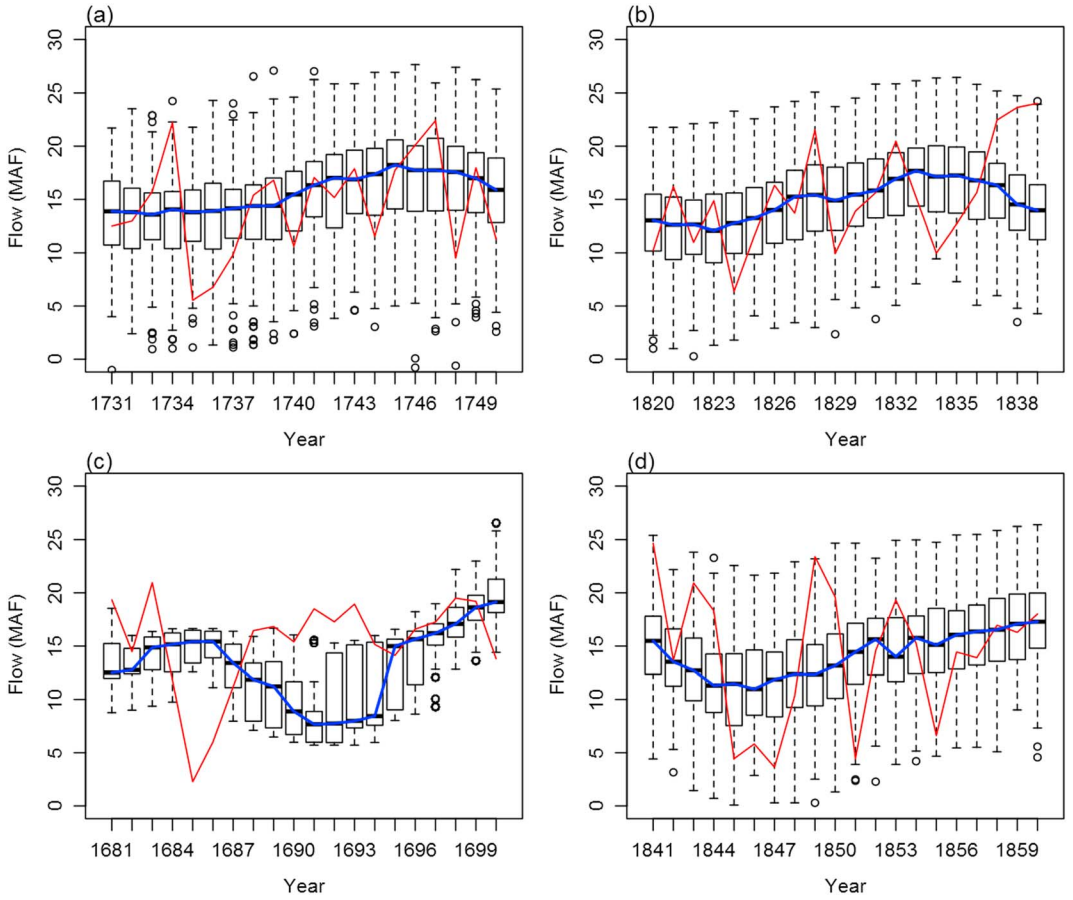

Figure 8. Projection ensembles (boxplot), median (blue), and the historic flows (red) for the high-predictability epochs (a) 1731–1750 and (b) 1820–1839 and low predictability epochs (c) 1681–1700 and (d) 1841–1860.

(Abarbanel & Lall, 1996). However, this predictability varies substantially over time or over parts of the phase space. To understand this, the time-varying LLE for a 20-year period (each value plotted is the average of a following 20-year period) for the three dimensions are shown in Figure 4a and their average is shown in Figure 4b. During epochs when the LLE of the first dimension—which is generally positive and inhibits long lead predictability—is lower, the overall predictability of the system is enhanced. For example, consider the period 1820–1839 when the average LLE is most negative—during this period $\lambda_1$varies in the range of 0.05, which indicates a predictability of ~20 years. At other epochs the predictability varies from 1 to 20 years. For ease of demonstration we use the time-varying average LLE (Figure 4b) in selecting high- and low-predictability epochs for simulation. The selected 20-year high-predictable epochs are 1731–1750 and 1820–1839, and a couple of low-predictability epochs selected are 1681–1700 and 1841–1860.

The temporal variability of predictability is seen from the time-varying average LLE (Figure 4b) along with the temporal variability of the signal variance, SAWP, of the Lees Ferry flow signal. There are distinct epochs where the average exponent is close to zero, indicative of reduced predictability, and where the average exponent is highly negative, suggesting good predictability. Furthermore, the LLEs track the signal SAWP very well, indicating that reduced predictability epochs coincide with increased signal variance and that good predictability epochs are consistent with reduced signal variance, which are quite intuitive. The predictability of the Lees Ferry flow system waxes and wanes over time and is consistent with the variability of the signal, which is very interesting and has significant implications for water resources management.

The embedding in the three dimensions is displayed in Figure 5. The attractor is in the form of a scroll with the trajectories at the center representing the low frequencies, and the outer strands are period excursions at even lower frequencies. This is similar to that seen in Abarbanel and Lall (1996) for the dynamics of the

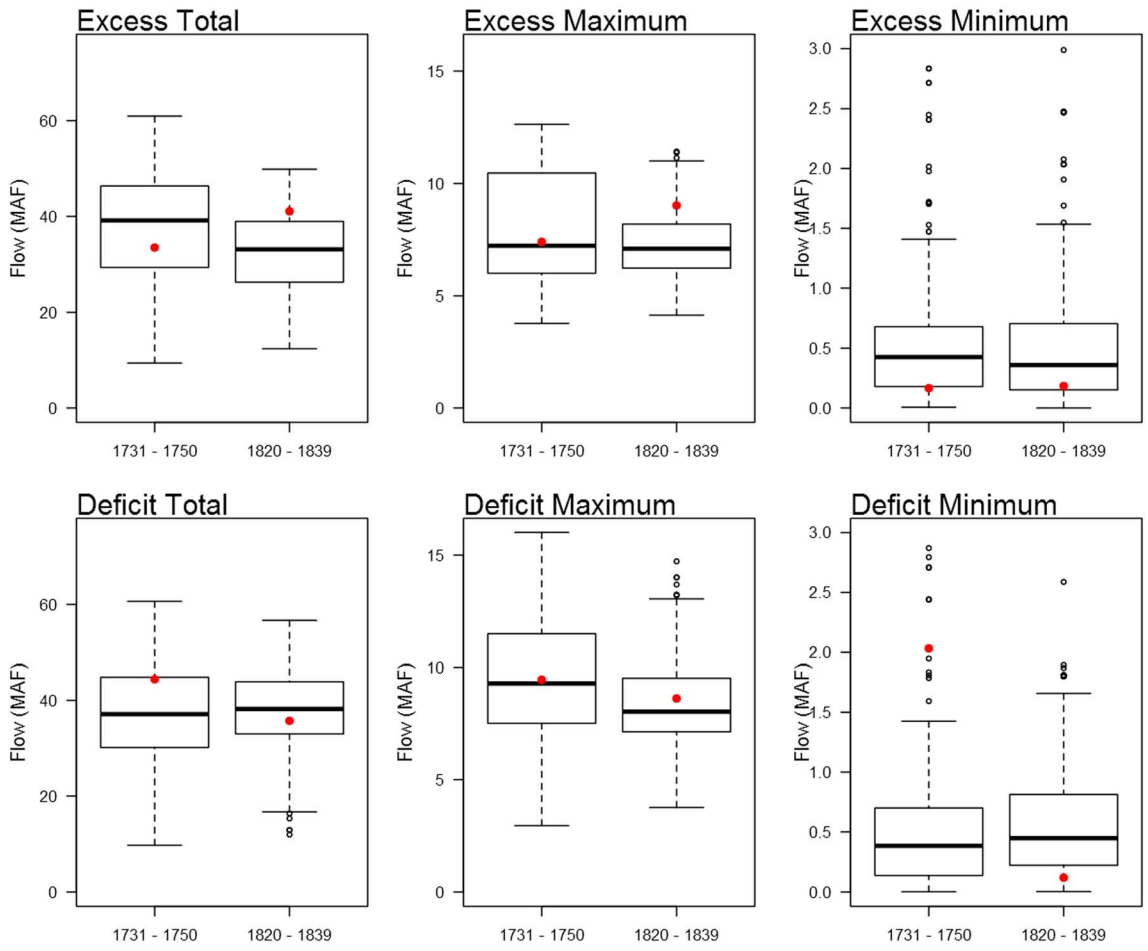

Figure 9. Boxplots of surplus and deficit statistics from flow projections for high-predictability epochs 1731–1750 and 1820–1839—(a) total surplus, (b) maximum surplus, and (c) minimum surplus from the projections. The values from the historic flows are show as red dots. Boxplots of deficit statistics—(d) total deficit, (e) maximum deficit, and (f) minimum deficit.

Great Salt Lake levels. To understand the evolution of these epochs in the phase space, the trajectories of these epochs are plotted on the attractor and are shown in Figure 5. An interesting observation emerges in that the low predictable (Figures 5a and 5b) epochs (1681–1700, 1841–1860, and 1970–1989) have their trajectories excursing to the outer parts of the attractor where the LLEs are higher than inner parts of the attractor, while the trajectories of high predictable (Figures 5b and 5c) epochs (1788–1807, 1820–1839, and 1926–1945) tend to stay within the inner parts of the attractor. However, the trajectory of the high predictable epoch 1820–1839 does an excursion to the outer reaches of the attractor. This is because the system transitions immediately to one of the lowest predictable epochs of 1841–1860; thus, the later part of the 1820–1830 epoch is already in transition to being low predictable. These trajectories seem to suggest that the outer portions of the attractor are less predictable and thus unstable, while the inner parts provide higher predictability. Furthermore, the low predictable epochs (Figure 5b) appear to originate from the outer reaches of the attractor while the high predictable epochs (Figure 5c) from the central part. The trajectory of the epoch of the early part of twentieth century, 1926–1945 (Figure 5e), originates from the center and stays there, while that of the recent epoch 1970–1989 (Figure 5d) has excursions to the outer reaches. The characteristics of trajectories from the reconstructed phase space highlight the ability of the nonlinear dynamical systems-based approach described and demonstrated above to capitalize on the regime dynamics for skillful projections, unlike traditional time series methods.

An interesting question is what is the source of this predictability? Recent studies show the association between AMO and PDO in modulating the variability of Lees Ferry flow (e.g., Bracken et al., 2014;

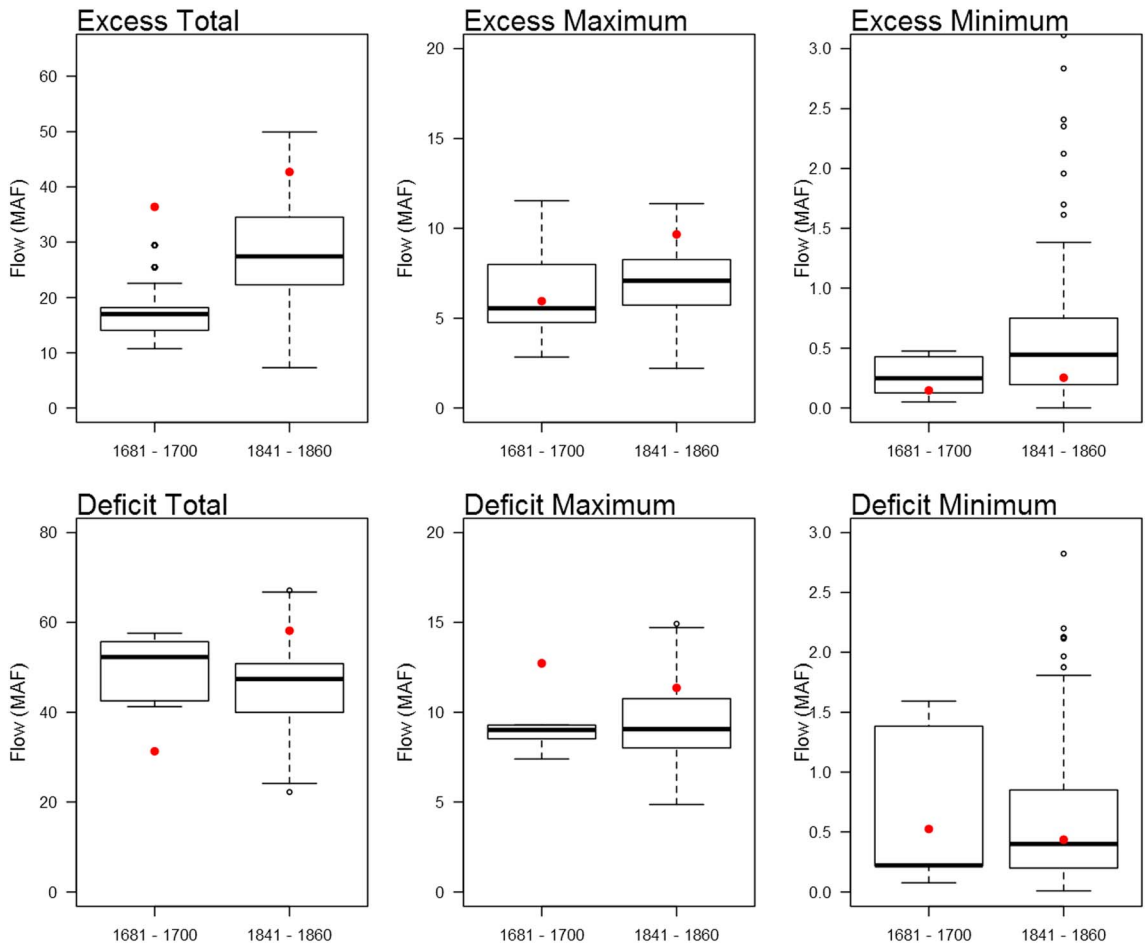

Figure 10. Same as Figure 9 but for low-predictability epochs 1681–1700 and 1841–1860.

Nowak et al., 2012). As an initial rudimentary effort to answering this question we relate the variability of the climate indices to Lees Ferry flow predictability. The SAWP of the signal of the climate indices, computed in the same manner as that of the flow, is shown along with the average LLEs (Figures 6a–6c). By visual inspection, the variability of PDO and AMO seem to be out of phase with the predictability in that a higher variance of climate signal leads to high negative values and, consequently, higher predictability. This is clarified in Figure 6c, which shows the average LLE along with the sum of the SAWP of signals of both of the climate drivers. Spectral coherence between the SAWP of the Lees Ferry Flow and the SAWP of AMO and PDO show strong coherence in the 2–4-year periods (figures not shown) suggesting a strong link at higher frequency. The climate system is more predictable when the drivers have higher variability (e.g., Kirtman & Schopf, 1998)—which then imparts predictability to regional hydrology. We recognize that the linkage between variability in climate drivers and flow predictability is suggestive but not quantitatively sound. This would require detailed analysis of the predictability of the climate drivers and the use of climate models to understand the mechanisms that translate predictability from large-scale climate to the flow. However, this linkage between climate drivers and Colorado River flow signal is quite interesting, suggesting that the epochal natural of the predictability of Lees Ferry flow is orchestrated by the variability in large-scale climate.

We tested the epochal nature of the projection skills on four selected 20-year epochs mentioned above—high-predictability epochs, 1731–1750 and 1820–1839, and low-predictability epochs, 1681–1700 and 1841–1860. Blind projections were made for these 20-year periods using the block bootstrap method (Erkyihun et al., 2016) and as described in the previous section. The PDFs of the flow simulations are shown as gray along with their median PDF and that of the historic flows for the four epochs in Figure 7. The top

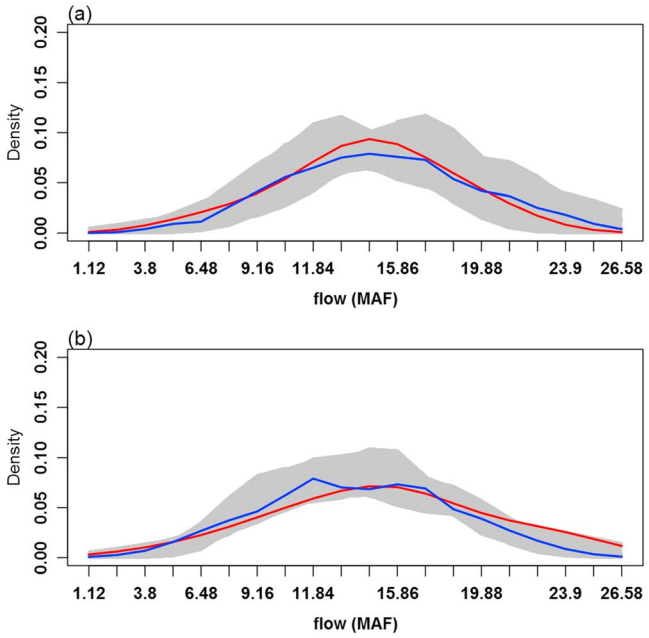

Figure 11. PDFs of flow projection ensembles (gray), their median (blue), and that of historic flows (red) for (a) early twentieth century epoch 1926–1945 and (b) recent epoch 1970–1989.

panels (Figures 7a and 7b) show the PDFs of high-predictability epochs, and it can be seen that the projections capture the PDF of the historic very well. The PDFs of the projections from lower predictability epochs (bottom panel, Figures 7c and 7d) misestimate the historic PDF. We recall that the uncertainty in the PDFs of the epoch 1681–1700 (Figure 7c) is smaller due to the fact that the projections are performed blind, in that data prior to 1681, which is shorter, are used in the projections. A Kolmogrov-Smirnov test suggested that the simulated and historical distributions of the high predictable epochs are indistinguishable and not the case for low predictable epochs. The projection ensembles and the historic flows are shown in Figure 8 in which it can be seen that during the predictable epochs (top panels, Figures 8a and 8b) the median of the ensembles (horizontal line in the boxes) tracks the variability of the historic flows (red line) quite well, but not well during the epochs with lower predictability (bottom panels, Figures 8c and 8d). The uncertainty in the projections for the epoch 1681–1700 (Figure 8c) is smaller due to the reasons explained above.

An important utility of these projections is use in multidecadal water resources planning and management as in Erkyihun (2015). To this end, the ability to capture aspects of sustained wet and dry periods is crucial. Threshold crossing statistics of deficit and excess are computed for each ensemble based on the threshold of 15MAF (the long-term average flow of the modern period). These are total excess and deficit about this threshold and maximum and minimum excess and deficit over the 20-year horizon. Flow values above this threshold indicate excess and below, deficit. For each 20-year horizon all the excess flows are added to get the total excess and, similarly, total deficit. We also select the maximum and minimum excesses, and maximum and minimum deficits, over the 20-year period for each ensemble member. They are shown in Figure 9 as boxplots along with the corresponding values from the historic flows as red dots for the excess (top panels) and deficit (bottom panels) for the high-predictability epochs. The historic values (dots) of all statistics are well captured by the simulations within the interquartile box of the simulations, and the excess and deficit minimum are oversimulated. This indicates that excursions about the thresholds, which are higher-order statistics, are also simulated very well; that will be of immense use in long-term water resources planning. In comparison, the low-predictability epochs show poor performance of the excess and deficit statistics (Figure 10), with most of the historic values outside the interquartile box of the simulations. These results suggest that the predictability of the system dynamics permeates through all aspects of the system statistics. To further assess the performance of the simulations in capturing wet and dry sequences, which are crucial for water resources management, we also computed storage statistics using the sequent peak algorithm (Loucks & Van Beek, 2005). The results are consistent with the performance of deficit and excess statistics in that they are well captured in predictable epochs and poorly in low predictable epochs (figures not shown).

The basin has experienced several impactful events starting in the early twentieth century—especially the water sharing agreements among Basin States in the early part of twentieth century and the recent ongoing drought in the western U.S. and in the basin for over a decade. Interestingly, the Lyapunov exponents are quite negative during early twentieth century, indicative of high predictability, and the recent decades of drought have low predictability (as the exponent values are closer to zero). To highlight these further, projections were made for two epochs from the last century, 1926–1945 and 1970–1989, for which the PDFs are shown in Figure 11. Consistent with high predictability of the early part of twentieth century, the projections capture the historic PDF very well. However, the PDFs of flow projections for the recent decades deviate significantly from the historic PDF, consistent with low predictability. The historic flow variability (red) falls within the interquartile range of the ensembles in most of the years, while in the later epoch (Figure 12b) the historic flow is outside the interquartile range for most of the years, conspicuously so during the high-flow period of 1982–1986.

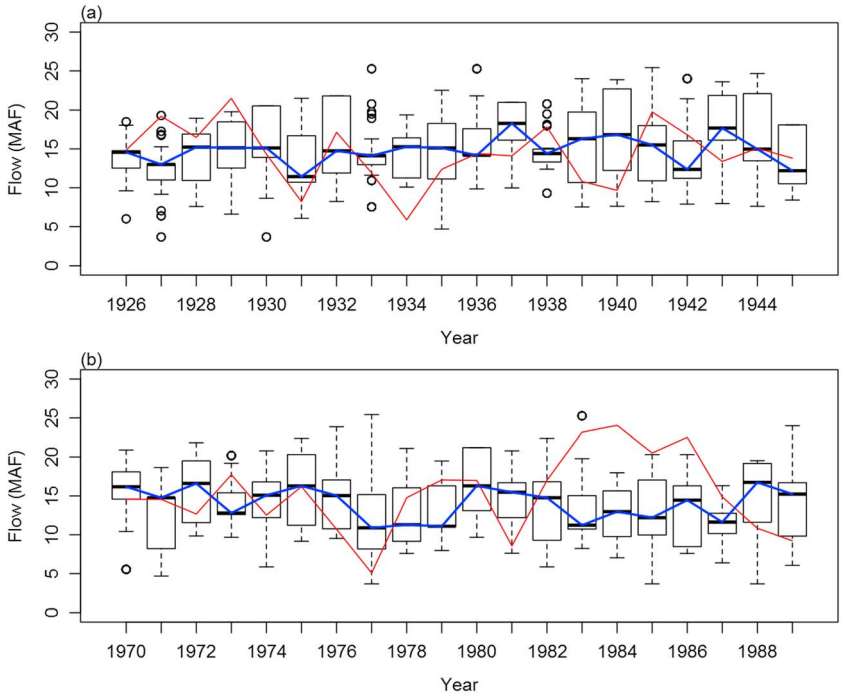

Figure 12. Projection ensembles (boxplot), median (blue), and the historic flows (red) for (a) early twentieth century epoch 1926–1945 and (b) recent epoch 1970–1989.

5. Summary and Discussion

A novel approach to stochastic time series simulation is proposed, which is based on reconstructing the phase space in which the dynamics of the time series unfolds. The approach also provides insights into epochal variability of predictability of the time series. The phase space reconstruction requires embedding the time series in appropriately estimated D dimensions with a time delay of τ. We made the assumption as in Takens (1981) that this maps to the true phase space in which the underlying dynamics unfolds enabling to take advantage of the system predictability to provide skillful projections and forecasts. Estimation of the two parameters D and τ from data with noise, especially of short geophysical time series, is often unreliable, and smoothing is suggested. We proposed obtaining the signal time series by wavelet filtering the original time series within significant frequency bands identified from wavelet spectral analysis. Thus, the reconstructed phase space and the Lyapunov exponents are based on the signal component of the time series. Lyapunov exponents measure the rate of divergence of trajectories in the reconstructed phase space providing an estimate of predictability—higher positive values indicating rapid divergence and thus lower predictability and vice versa. The exponents can be computed locally in the phase space to provide temporal variability of predictability. Projections from a time point t involve (i) mapping the current state (or feature vector) of the system on to the reconstructed phase space; (ii) identifying K-nearest neighbors of the current feature vector; (iii) resampling one of them with a weight function that corresponds to a time j; and (iv) selecting as the simulated flow projection the M-time step sequence (or trajectory) of the original time series from the period j to $j + M$. Identifying the neighbors in the phase space of the smoothed series and resampling the original time series is a new approach to stochastic time series simulation. This is repeated to generate ensembles. Using data prior to time t enables blind projections, which is implemented here.

We applied this modeling approach to the long paleo-reconstructed flow at Lees Ferry gauge, an important location on the Colorado River, which represents 85–90% of its flow. Four dominant period bands in the 7- to 57-year range were identified from the wavelet spectrum analysis; filtering the flow series in these bands provided the signal time series. The reconstructed phase space showed an inner and outer scroll indicative of lower and higher period variations. The global Lyapunov exponents were negative, suggesting that the signal in the flow series is generally predictable. However, the LLEs showed significant epochal variations—with

some epochs exhibiting high predictability (negative exponent values) and some low predictability (exponent values close to zero). The early part of twentieth century when the Colorado River compact agreements were negotiated was a high predictable epoch, and the recent decades with the prolonged and unprecedented drought in the observed record have low predictability. The predictability coincides with the temporal variabilities of the signal variance—suggesting that periods of high signal variance impede predictability and vice versa. Furthermore, the temporal variabilities of the signal variance of large-scale climate indices AMO and PDO also coincide with the temporal variability of the flow signal Lyapunov exponents—indicating that large-scale climate features modulate flow predictability. Blind projections of streamflows during high predictable epochs show good skill and capture all the distributional, drought, and surplus statistics, while the low-predictability epochs had poor performance on these measures.

The method generally performs well for long time series with modest block size. A long time series is essential to adequately sample the underlying attractor, and the block size determines the prediction horizon. Insights as to an appropriate block size to use can be obtained using the LLE and experimentation with the time series. The time-varying predictability of the streamflow offers interesting insight into the system. It suggests that low-predictability epochs are an inherent part of the dynamics of the system and likely resistant to any improvements in modeling efforts—statistical or physical.

It is interesting to find that the recent epoch is likely a low predictable epoch with the confluence of increased variability in the streamflow and in the large-scale climate drivers while the early epoch of the twentieth century when the Colorado River compacts were instituted was a high predictable epoch. This also provokes the tantalizing idea that perhaps water resources management should adapt in a flexible manner to these predictability epochs. The LLEs could be modeled as a time series, and projections of low-predictability epochs and high-predictability epochs could trigger appropriate management and planning responses. This study opens opportunities to perceive hydrologic predictability and consequently water management in a new paradigm.

Acknowledgments
The authors thankfully acknowledge the funding for this research by the Bureau of Reclamation (R12AC30023). To obtain the data used to produce the results in this paper, please follow the links in Table 1 and descriptions in the data section. We thank Dr. T. Lee, one of the reviewers, two anonymous reviewers, and the Editor for their insightful comments, which helped to improve the manuscript significantly.

References

Abarbanel, H. D. I., Brown, R., & Kennel, M. B. (1992). Local Lyapunov Exponents computed from observed data. *Journal of Nonlinear Science*, *2*(3), 343–365. https://doi.org/10.1007/BF01208929

Abarbanel, H. D. I., Brown, R., Sidorowich, J. J., & Tsimring, L. S. (1993). The analysis of observed chaotic data in physical systems. *Reviews of Modern Physics*, *65*(4), 1331–1392. https://doi.org/10.1103/RevModPhys.65.1331

Abarbanel, H. D. I., & Lall, U. (1996). Nonlinear dynamics of the Great Salt Lake: System identification and prediction. *Climate Dynamics*, *12*(4), 287–297. https://doi.org/10.1007/BF00219502

Asefa, T., Kemblowski, M., Lall, U., & Urroz, G. (2005). Support vector machines for nonlinear state space reconstruction: Application to the Great Salt Lake time series. *Water Resources Research*, *41*, W12422. https://doi.org/10.1029/2004WR00378

Bailey, B. A., & Nychka, D. W. (1995). Local Lyapunov Exponents: Predictability depends on where you are. In W. A. Barnett, A. P. Kirman, & M. Salmon (Eds.), *Nonlinear Dynamics and Economics: Proceedings of the Tenth International Symposium in Economic Theory and Econometrics* (pp. 345–360). Cambridge University Press.

Bracken, C., Rajagopalan, B., & Zagona, E. (2014). A hidden Markov Model combined with climate indices for multi-decadal streamflow simulation. *Water Resources Research*, *50*, 7836–7846. https://doi.org/10.1002/2014WR015567

Casdagli, M., Eubank, S., Farmer, J. D., Gibson, J., Des Dardins, D., Hunter, N., & Theiler, J. (1990). *Nonlinear modeling of chaotic time series: Theory and applications*. LA-UR-91-1637: Los Alamos National Laboratories.

Dhanya, C. T., & Nagesh Kumar, D. (2010). Nonlinear ensemble prediction of chaotic daily rainfall. *Advances in Water Resources*, *33*(3), 327–347. https://doi.org/10.1016/j.advwatres.2010.01.001

Dhanya, C. T., & Nagesh Kumar, D. (2011). Predictive uncertainty of chaotic daily streamflow using ensemble wavelet networks approach. *Water Resources Research*, *47*, W06507. https://doi.org/10.1029/2010wr010173

Efron, B., & Tibishirani, R. (1993). *An introduction to the bootstrap*. New York: Chapman and Hall. https://doi.org/10.1007/978-1-4899-4541-9

Elshorbagy, A., Simonovic, S. P., & Panu, U. S. (2002). Estimation of missing streamflow data using principles of chaos theory. *Journal of Hydrology*, *255*(1-4), 123–133. https://doi.org/10.1016/S0022-1694(01)00513-3

Elsner, J. B., & Tsonis, A. A. (1992). Nonlinear prediction, chaos, and noise. *American Meteorological Society*, *73*(1), 49–60. https://doi.org/10.1175/1520-0477(1992)073<0049:NPCAN>2.0.CO;2

Enfield, D. B., Mestas-Nunez, A. M., & Trimble, P. J. (2001). The Atlantic Multidecadal Oscillation and its relation to rainfall and river flows in the continental US. *Geophysical Research Letters*, *28*(10), 2077–2080. https://doi.org/10.1029/2000GL012745

Erkyihun, S. T. (2015). Multi-decadal stochastic streamflow projection: An application to water resources decision making in the Colorado River Basin, civil, environmental, and architectural engineering. PhD thesis, University of Colorado, Boulder, CO.

Erkyihun, S. T., Rajagopalan, B., Zagona, E., Lall, U., & Nowak, K. (2016). Wavelet-based time series bootstrap model for multidecadal streamflow simulation using climate indicators. *Water Resources Research*, *52*, 4061–4077. https://doi.org/10.1002/2016wr018696

Farmer, J. D., & Sidorowich, J. J. (1987). Predicting chaotic time series. *Physical Review Letters*, *59*(8), 845–848. https://doi.org/10.1103/PhysRevLett.59.845

Fraser, A. M., & Swinney, H. L. (1986). *Independent coordinates for strange attractors from mutual information*. Bethesda, MD, The American Physical Society.

Gao, J., Sultan, H., Hu, J., & Tung, W. W. (2010). Denoising nonlinear time series by adaptive filtering and wavelet shrinkage: A comparison. *IEEE Signal Processing Letters*, *17*(3), 237–240. https://doi.org/10.1109/lsp.2009.2037773

Grassberger, P., & Procaccia, I. (1983a). Characterization of strange attractors. *Physical Review Letters*, *50*(5), 346–349. https://doi.org/10.1103/PhysRevLett.50.346

Grassberger, P., & Procaccia, I. (1983b). Measuring the strangeness of strange attractors. *Physica D*, *9*(1-2), 189–208. https://doi.org/10.1016/0167-2789(83)90298-1

Grassberger, P., Schreiber, T., & Schaffrath, C. (1991). Nonlinear time sequence analysis. *International Journal of Bifurcation and Chaos*, *01*(03), 521–547. https://doi.org/10.1142/S0218127491000403

Gray, S. T., Graumlich, L. J., Betancourt, J. L., & Pederson, G. T. (2004). A tree-ring based reconstruction of the Atlantic Multidecadal Oscillation since 1567 A.D. *Geophysical Research Letters*, *31*, L12205. https://doi.org/10.1029/2004GL019932

Guégan D., &Leroux, J. (2009). Forecasting chaotic systems: The role of local Lyapunov exponents. *Chaos, Solitons & Fractals*, *41*(5), 2401–2404.

Guégan, D., & Leroux, J. (2011). Predicting Chaos with Lyapunov Exponents: ZeroPlays no Role in Forecasting Chaotic Systems, Chaotic Systems, Esteban Tlelo-Cuautle, IntechOpen. https://doi.org/10.5772/14349, https://www.intechopen.com/books/chaotic-systems/predicting-chaos-with-lyapunov-exponents-zero-plays-no-role-in-forecasting-chaotic-systems

Han, M., Liu, Y., Xi, J., & Guo, W. (2007). Noise smoothing for nonlinear time series using wavelet soft threshold. *IEEE Signal Processing Letters*, *14*(1), 62–65. https://doi.org/10.1109/LSP.2006.881518

Hansen, J. A., & Smith, L. A. (2001). Probabilistic noise reduction. *Tellus A: Dynamic Meteorology and Oceanography*, *53*(5), 585–598. https://doi.org/10.3402/tellusa.v53i5.12226

Islam, M. N., & Sivakumar, B. (2002). Characterization and prediction of runoff dynamics: A nonlinear dynamical view. *Advances in Water Resources*, *25*(2), 179–190. https://doi.org/10.1016/S0309-1708(01)00053-7

Jaeger, L., & Kantz, H. (1996). Unbiased reconstruction of the dynamics underlying a noisy chaotic time series. *Chaos*, *6*(3), 440. https://doi.org/10.1063/1.166196

Kantz, H. (1994). A robust method to estimate the maximal Lyapunov exponent of a time series. *Physics Letters A*, *185*(1), 77–87. https://doi.org/10.1016/0375-9601(94)90991-1

Kantz, H., & Schreiber, T. (1997). *Nonlinear time series analysis*. New York: Cambridge University Press.

Kantz, H., & Schreiber, T. (1998). Human ECG, Nonlinear deterministic versus stochastic aspects. *IEE Proceedings: Science, Measurement & Technology*, *145*(6), 279–284. https://doi.org/10.1049/ip-smt:19982327

Kaplan, A., Cane, M. A., Kushnir, Y., Clement, A. C., Blumenthal, M. B., & Rajagopalan, B. (1998). Analyses of global sea surface temperature 1856-1991. *Journal of Geophysical Research*, *103*(C9), 18,567–18,589. https://doi.org/10.1029/97JC01736

Kennel, M. B., Brown, R., & Abarbanel, H. D. I. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A*, *45*, 3403–3411.

Khatibi, R., Sivakumar, B., Ghorbani, M. A., Kisi, O., Kocak, K., & Farsadi Zadeh, D. (2012). Investigating chaos in river stage and discharge time series. *Journal of Hydrology*, *414-415*, 108–117. https://doi.org/10.1016/j.jhydrol.2011.10.026

Kirchner, J. W. (2009). Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward. *Water Resources Research*, *45*, W02429. https://doi.org/10.1029/2008WR006912

Kirtman, B. P., & Schopf, P. S. (1998). Decadal variability in ENSO predictability and prediction. *Journal of Climate*, *11*(11), 2804–2822. https://doi.org/10.1175/1520-0442(1998)011<2804:DVIEPA>2.0.CO;2

Kwon, H. H., Lall, U., & Khalil, A. F. (2007). Stochastic simulation model for nonstationary time series using an autoregressive wavelet decomposition: Applications to rainfall and temperature. *Water Resources Research*, *43*, W05407. https://doi.org/10.1029/2006WR005258

Lall, U. (1995). Recent advances in nonparametric function estimation: Hydraulic applications, *U.S. Natl. Rep. Int. Union Geod. Geophys. 1991-1994. Reviews of Geophysics*, *33*(S2), 1093–1102. https://doi.org/10.1029/95rg00343

Lall, U., Moon, Y. I., Kwon, H. H., & Bosworth, K. (2006). Locally weighted polynomial regression: Parameter choice and application to forecasts of the Great Salt Lake. *Water Resources Research*, *42*, W05422. https://doi.org/10.1029/2004WR003782

Lall, U., Sangoyomi, T., & Abarbanel, H. D. I. (1996). Nonlinear dynamics of the Great Salt Lake: Nonparametric short-term forecasting. *Water Resources Research*, *32*(4), 975–985. https://doi.org/10.1029/95WR03402

Lall, U., & Sharma, A. (1996). A nearest neighbor bootstrap for resampling hydrologic time series. *Water Resources Research*, *32*(3), 679–693. https://doi.org/10.1029/95WR02966

Liu, Q., Islam, S., & Rodriguez-Iturbe, I. (1998). Phase-space analysis of daily streamflow: Characterization and prediction. *Advances in Water Resources*, *21*(6), 463–475. https://doi.org/10.1016/S0309-1708(97)00013-4

Loucks, D. P., & Van Beek, E. (2005). Concepts in Probability, Statistics and Stochastic Modelling, in *Water Resources Systems Planning and Management*, pp. 169–229. Paris, The Netherlands: UNESCO PUBLISHING, United Nations Educational, Scientific and Cultural Organization.

MacDonald, G. M., & Case, R. A. (2005). Variations in the Pacific Decadal Oscillation over the past millennium. *Geophysical Research Letters*, *32*, L08703. https://doi.org/10.1029/2005GL022478

Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., & Francis, R. C. (1997). A Pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of the American Meteorological Society*, *78*(6), 1069–1079. https://doi.org/10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2

Moon, Y., Lall, U., & Kwon, H. H. (2008). Non-parametric short-term forecasts of the Great Salt Lake using atmospheric indices. *International Journal of Climatology*, *28*(3), 361–370. https://doi.org/10.1002/joc.1533

Moon, Y., Rajagopalan, B., & Lall, U. (1995). Estimation of mutual information using kernel density estimators. *Physical Review E*, *52*(3), 2318–2321. https://doi.org/10.1103/PhysRevE.52.2318

Nese, J. M. (1989). Quantifying local predictability in phase space. *Physica D: Nonlinear Phenomena*, *35*(1–2), 237–250. https://doi.org/10.1016/0167-2789(89)90105-X

Nowak, K., Hoerling, M., Rajagopalan, B., & Zagona, E. (2012). Colorado River Basin hydroclimatic variability. *Journal of Climate*, *25*(12), 4389–4403. https://doi.org/10.1175/JCLI-D-11-00406.1

Nowak, K., Rajagopalan, B., & Zagona, E. (2011). Wavelet Auto-Regressive Method (WARM) for multi-site streamflow simulation of data with non-stationary spectra. *Journal of Hydrology*, *410*(1-2), 1–12. https://doi.org/10.1016/j.jhydrol.2011.08.051

Oseledec, V. I. (1968). A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems. *Trudy Mosk Mat Obsc 19:197. Moscow Mathematics Society*, *19*, 197.

Packard, N. H., Crutchfield, J. P., Farmer, J. D., & Shaw, R. S. (1980). Geometry from a time series. *Physical Review Letters*, *45*(9), 712–716. https://doi.org/10.1103/PhysRevLett.45.712

Peterson, T. J., & Western, A. W. (2014). Multiple hydrological attractors under stochastic daily forcing: 1. Can multiple attractors exist? *Water Resources Research*, *50*, 3010–3029. https://doi.org/10.1002/2012WR013004

Porporato, A., & Ridolfi, L. (1996). Clues to the existence of deterministic chaos in river flow. *International Journal of Modern Physics B*, *10*(15), 1821–1862. https://doi.org/10.1142/S0217979296000830

Porporato, A., & Ridolfi, L. (1997). Nonlinear analysis of river flow time sequences. *Water Resources Research*, *33*(6), 1353–1367. https://doi.org/10.1029/96WR03535

Prairie, J., & Callejo, R. (2005). *Natural flow and salt computation methods*. Salt Lake City, UT: United States Bureau the Interior.

Rajagopalan, B., & Lall, U. (1998). Low frequency variability in western U.S. precipitation. *Journal of Hydrology*, *210*(1-4), 51–67. https://doi.org/10.1016/S0022-1694(98)00184-X

Rajagopalan, B., & Lall, U. (1999). A nearest neighbor bootstrap resampling scheme for resampling daily precipitation and other weather variables. *Water Resources Research*, *35*(10), 3089–3101. https://doi.org/10.1029/1999WR900028

Rajagopalan, B., Salas, J., & Lall, U. (2010). In B. Sivakumar, & R. Berndtsson (Eds.), *Stochastic methods for modeling precipitation and streamflow, in Advances in data-based approaches for hydrologic modeling and forecasting*, Ed by. Singapore: World Scientific.

Regonda, S., Rajagopalan, B., Lall, U., Clark, M., & Moon, Y. (2005). Local polynomial method for ensemble forecast of time series. *Nonlinear Processes in Geophysics*, *12*, 397–406.

Rodriguez-Iturbe, I., Febres De Power, B., Sharifi, M. B., & Georgakakos, K. P. (1989). Chaos in rainfall. *Water Resources Research*, *25*(7), 1667–1675. https://doi.org/10.1029/WR025i007p01667

Salas, J. D., Delleur, J. W., Yevjevich, V., & Lane, W. L. (1980). *Applied modeling of hydrological time series*. Littleton, Colorado: Water Resources Publication.

Sangayomi, T. B., Lall, U., & Abarbanel, H. D. I. (1996). Nonlinear dynamics of the Great Salt Lake: Dimension estimation. *Water Resources Research*, *32*(1), 149–159. https://doi.org/10.1029/95WR02872

Schreiber, T., & Grassberger, P. (1991). A simple noise-reduction method for real data. *Physics Letters A*, *160*(5), 411–418. https://doi.org/10.1016/0375-9601(91)90237-3

Schreiber, T., & Kantz, H. (1996). Observing and predicting chaotic signals: Is 2% noise too much? In Y. A. Kravtsov & J. B. Kadtke (Eds.), *Predictability of Complex Dynamical Systems* (pp.43–65). New York: Springer. https://doi.org/10.1007/978-3-642-80254-6_3

Sharma, A., Tarboton, D. G., & Lall, U. (1997). Streamflow simulation: A nonparametric approach. *Water Resources Research*, *33*(2), 291–308. https://doi.org/10.1029/96WR02839

Sivakumar, B., Phoon, K.-K., Liong, S.-Y., & Liaw, C.-Y. (1999). Comment on "Nonlinear analysis of river flow time sequences" by A. Porporato and L. Ridolfi. *Water Resources Research*, *35*(3), 895–897. https://doi.org/10.1029/1998WR900033

Smith, L. A. (1992). Identification and prediction of low dimensional dynamics. *Physica D*, *58*(1-4), 50–76. https://doi.org/10.1016/0167-2789(92)90101-R

Takens, F. (1981). Detecting strange attractors in turbulence. In D. Rand, & L. S. Young (Eds.), *Dynamical systems and turbulence, Warwick 1980*, *Lecture Notes in Mathematics*, (Vol. 898, pp. 366–381). Berlin: Springer.

Torrence, C., & Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, *79*(1), 61–78. https://doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2

Tsonis, A. A. (1992). *Chaos: From theory to applications*. NewYork: Springer. https://doi.org/10.1007/978-1-4615-3360-3

Wei, W. W. S. (2006). *Time series analysis: Univariate and multivariate methods*. New York: Temple University, Addison-Wesley.

Wolf, A., Swift, J. B., Swinney, H. L., & Vastano, J. A. (1985). *Determining Lyapunov Exponents from a time series*, (Vol. 16, 285–317). Elsevier Science Publishers. https://doi.org/10.1016/0167-2789(85)90011-9

Woodhouse, C. A., Gray, S. T., & Meko, D. M. (2006). Updated streamflow reconstructions for the Upper Colorado River Basin. *Water Resources Research*, *42*, W05415. https://doi.org/10.1029/2005WR004455

Zhang, Y., Wallace, J. M., & Battisti, D. S. (1997). ENSO-like interdecadal variability: 1900-93. *Journal of Climate*, *10*(5), 1004–1020. https://doi.org/10.1175/1520-0442(1997)010<1004:ELIV>2.0.CO;2