Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Hydrology 312 (2005) 79–94
Journal of Hydrology
www.elsevier.com/locate/jhydrol

Prediction boundaries and forecasting of non linear hydrologic stage data

Joseph Park*, Jayantha Obeysekera, Randy VanZee

Model Development Division (4540), Office of Modeling, South Florida Water Management District, 3301 Gun Club Road, West Palm Beach, FL 33406, USA

Received 28 July 2003; revised 31 December 2004; accepted 11 February 2005

Abstract

This paper analyzes and forecasts hydrologic stage data in the Loxahatchee National Wildlife Refuge, the northernmost extent of the Florida Everglades. Analysis indicates that the process dynamics are chaotic, for which several attractor invariants are evaluated. A connection is made between the Kolmogorov-Sinai (KS) entropy of the phase-space trajectories and limits of temporal stage predictability. Evaluation of the KS entropy establishes a boundary for forecast time periods. Comparisons are presented between inferences produced by linear statistical models and the nonlinear attractor invariants. A nonlinear estimator (feedforward neural network), along with several linear models are employed to perform temporal forecasts of stage data. The observed degradation of forecast accuracy is consistent with the limits inferred from the attractor entropy. The nonlinear estimator is found to have better prediction accuracy than the linear models for prediction intervals beyond several days.

Keywords: Nonlinear dynamics; Stage prediction; Artificial neural network

1. Introduction

Many applications of analytical scrutiny applied to real-world, physical systems are based on assumptions of linearity. These assumptions are often taken for granted, and may lead to inaccuracies as a result of applying linear modeling and forecasting techniques to physical systems that exhibit nonlinear dynamics. Nonlinear dynamics are nherent in chaotic systems: multivariate processes that exhibit internal order, long term stability, sensitive dependence on initial conditions and significant dynamic variations. While the process dynamics are stable in a global sense, the local phase-space trajectories of chaotic system variables are inherently unpredictable as forecast time scale increases. Chaotic systems maintain exquisite balances between the nonlinear forces of dissipation, randomness and internal order, and therefore, are often not best analyzed in the linear domain. It is notable that such realizations evolved from the celebrated work of Lorenz (Lorenz, 1963) who was concerned with forecasting atmospheric flows. Subsequently, evidence of chaotic behavior in hydrological time-series has been found since the late

* Corresponding author. Tel.: +1 561 682 2057; fax: +1 561 682 5750.
E-mail address: jpark@sfwmd.gov (J. Park).

0022-1694/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2005.02.022

1980s. Even though much debate ensued over the veracity of some of these findings, it is now well accepted that low-dimensional determinism is indeed a feature of many hydrological processes (Sivakumar, 2000, 2004). In addition to the detection of chaotic behavior in hydrological processes, application of nonlinear forecasting methods such as artificial neural networks (ANN) (Elshorbagy et al., 2002a,b; Sivakumar et al., 2002; Lambrakis, 2000), k-nearest neighbors (Elshorbagy et al., 2002a), and phase-space reconstruction (Sivakumar et al., 2002) have resulted in viable nonlinear forecasting procedures. The majority of these analysis have focused on rainfall or stream flow dynamics (Sivakumar, 2000, 2004), while to the authors knowledge, few have addressed waterbody volume/stage dynamics (Sangoyomi et al., 1996).

This study applies the methods of nonlinear dynamical systems theory to the analysis of daily hydrologic stage variations in the Loxahatchee National Wildlife Refuge. It is suggested that the process dynamics are chaotic; meaning that the phase-space attractor has a fractal dimension, a positive Lyapunov exponent, and occupies a finite domain. These features imply that linear modeling of the stage evolution could lead to inaccuracies in forecasting. Quantification of several attractor invariants leads to an estimate of the dimensionality of the dynamics, as well as a measure of the local trajectory divergence. Through connection of the trajectory divergence rate and the information capacity of the attractor in terms of the observed state values of the dynamics, a limit on the predictability of fine-scale stage forecasting is established. Comparisons to the predictability inferred from linear correlations are presented.

Having established some properties of the nonlinear dynamics, the study turns to application of a nonlinear model applied to forecasting the stage data. An artificial neural network is constructed to learn, and predict stage evolution within the boundaries of the identified predictability limits. The neural network is employed to forecast stages for future days over a 16 year time period, based on two years of training data. Additionally, several linear models are constructed and their forecasting results compared to those of the ANN. Errors of the forecasts are examined with respect to the predictability derived from analysis of the attractor dynamics.

To provide a basis for subsequent discussion on the features of linear and nonlinear modeling, a brief review of the two approaches is outlined below.

1.1. Linear models

Consider a linear system with f degrees of freedom where the time evolution of each component is described by the vector $\boldsymbol{\mu}(t) = [\mu_1(t), \mu_2(t), \ldots, \mu_f(t)]$ so that

$$\frac{\mathrm{d}\boldsymbol{\mu}(t)}{\mathrm{d}t} = \mathbf{A}\cdot\boldsymbol{\mu}(t) \tag{1}$$

where $\mathbf{A}$ is a constant $f\times f$ matrix. Solution of this equation specifies trajectories of $\boldsymbol{\mu}(t)$ in f-space for which one of three outcomes is possible:

1. Real part of the eigenvalues of $\mathbf{A}$ are positive.
2. Real part of the eigenvalues of $\mathbf{A}$ are negative.
3. Eigenvalues of $\mathbf{A}$ occur in complex-conjugate pairs with either zero or negative real parts.

In the case of negative eigenvalues, the trajectories of $\boldsymbol{\mu}(t)$ eventually collapse to a stable point, such as in the case of a pendulum subjected to friction. If only positive eigenvalues are present the orbits of $\boldsymbol{\mu}(t)$ grow to infinity, a situation which clearly invalidates a stable physical model. The occurrence of positive eigenvalues ensures that the assumption of linear dynamics is incorrect, in which case one may employ nonlinear evolution equations to better approximate the dynamics.

In most modeling applications one has access to sampled versions of the process dynamics at a fixed spatial point

$$s(n) = s(t_0 + n\tau_s) \tag{2}$$

where t_0 is the initial time and τ_s the sampling interval for the nth observation. In linear analysis, it is assumed that the observations are linearly related to previous observations and forcing terms

$$s(n) = \sum_{i=1}^{N} a_i s(n-i) + \sum_{j=1}^{M} b_j \phi(n-j) \tag{3}$$

where the a_i and b_j are model coefficients and ϕ represent the forcing terms. This integral representation is typically cast into an algebraic form through

application of the z-transform (discrete Fourier transform) resulting in

$$S(z) = \sum_n z^n s(n) = \frac{\sum_{j=1}^{M} b_j z^j}{1 - \sum_{i=1}^{N} a_i z^i} \mathbf{\Phi}(z) \tag{4}$$

where $\mathbf{\Phi}(z)$ is the z-transform of ϕ. The task is then one of judiciously selecting the coefficients to conform the model results with the observations. From a dynamical systems perspective, the a_i terms model the linear dynamics (autoregressive), and the b_j address linear averaging of the forcing terms (moving average), resulting in the familiar ARMA models. This form of linear systems model will have zero or negative eigenvalues of $\mathbf{A}$ and will not be the basis of choice for modeling nonlinear, chaotic dynamics.

1.2. Nonlinear models

System modeling and analysis of nonlinear, chaotic processes is centered on the notion of a stable (nondivergent) geometric phase-space attractor from which the observations evolve. It can be assumed that the system dynamics evolve in a multivariate space whose size and structure can be revealed through analysis of the observations. There are no initial assumptions imposed as to the form of the model. The idea is to let the data itself dictate the essential invariant features of the process dynamics. Once the phase-space invariants are identified, interpolation and projection of the process trajectories can provide nonlinear estimators and predictors. The modeler is then faced with the task of coupling physical significance to the revealed invariants.

In order to model a dynamical system it is assumed that a set of differential equations or discrete time evolution rules govern the behavior of the f system variables contained in the vector $\boldsymbol{\mu}(t)$

$$\frac{\mathrm{d}\boldsymbol{\mu}(t)}{\mathrm{d}t} = \mathbf{G}(\boldsymbol{\mu}(t)) \tag{5}$$

where $\mathbf{G}$ is a vector field that is continuous in its variables and which is differentiable as needed. The value of f defines the number of independent equations required to form an orthogonal basis to describe the system dynamics, and corresponds to the number of phase-space dimensions required to completely unfold the attractor so that false projections and crossings of $\boldsymbol{\mu}$ are eliminated. The discrete time extension is specified by a map from vectors in $\mathcal{R}^f$ to other vectors in $\mathcal{R}^f$, each at a discrete time $\boldsymbol{\mu}(n) = \boldsymbol{\mu}(t_0 + n\tau_s)$

$$\boldsymbol{\mu}(n+1) = \mathbf{F}(\boldsymbol{\mu}(n)) \tag{6}$$

This expression defines the evolution equation of the dynamical system, the vector field $\mathbf{F}$ encapsulates parameters which reflect the physical properties of the system as well as external influences of forces and boundary conditions. From a modeling perspective, the idea is to identify parameterized nonlinear functions $\mathbf{F}(\boldsymbol{\alpha})$ which map $\boldsymbol{\mu}(n)$ into $\boldsymbol{\mu}(n+1) = \mathbf{F}(\boldsymbol{\mu}(n), \boldsymbol{\alpha})$ where one has applied appropriate fit criteria to evaluate the parameters $\boldsymbol{\alpha}$.

It should be noted that in the current work, this approach is not explicitly implemented in that $\mathbf{F}$ is not determined by causal relationships to observable parameters. Rather, $\mathbf{F}$ is 'learned' implicitly by the artificial neural network and is represented internally in the organization of the ANN weight matrix. The ANN utilizes this internal representation of $\mathbf{F}$ in forecasting stage evolutions based on current observables.

2. Observed data

The data observations consist of daily recorded stage values at the 1–8C stage monitoring site in the Loxahatchee National Wildlife Refuge (WCA1). The refuge contains one of three water conservation areas in south Florida, and is maintained to provide water storage and flood control, as well as habitat for native fish and wildlife populations. These freshwater storage areas and part of Everglades National Park are what remain of the original freshwater Everglades. The entire refuge comprises about 221 square miles (147,392 acres) and is surrounded by a 57 mile canal and levee. The primary forcing functions for the water levels in the refuge are rainfall, and anthropogenic pumping and drainage for regional water regulation.

The observations cover the period from 1 January 1965 through 31 December 2000, a 35 year period. A plot of the daily time-series of stage observations biased by the arithmetic mean is shown in Fig. 1. Examination of the figure reveals that the data is quasi-periodic, with significant random components,

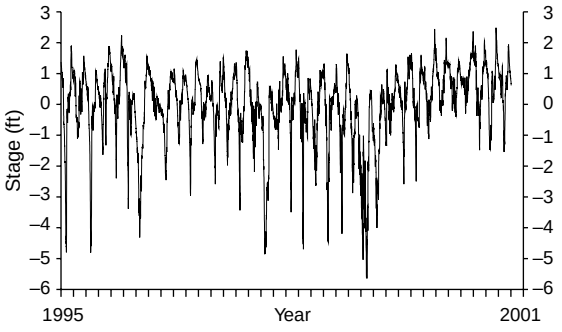

Fig. 1. 35 Year time-series of stage in WCA1.

suggesting that there is an underlying order to the dynamics which interact with random perturbations.

A main source of the stochastic forcing is the rainfall, which for the same period is plotted in Fig. 2.

3. Dynamical invariants

One of the goals of dynamical systems analysis is to identify system invariants which are insensitive to initial conditions. Invariants encode information regarding the structure of the system which imposes order on the dynamics, the signal-in-the-noise. Quantification of system invariants can also lead to physical insights regarding the process dynamics. In this section we examine several invariants: attractor dimension, correlation integrals, and Lyapunov exponents. The former set bounds on the number of degrees of freedom of the differential equation model, while the later provides quantification of the local predictability of system states.

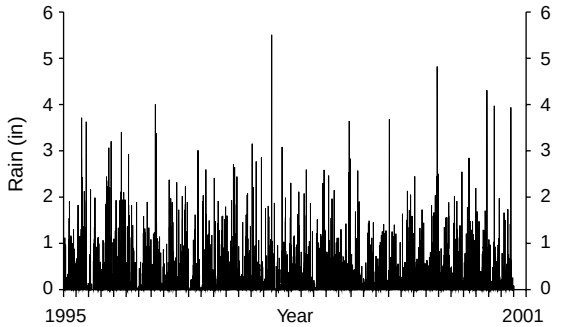

Fig. 2. 35 Year time-series of rainfall in WCA1.

3.1. Dimension

Establishing the dimension of the dynamical process is an important first-step in constructing a viable phase-space representation of the dynamics. There exist several methods for achieving this, each of which holds advantages and disadvantages. Here, we will examine two methods for estimating the dimension: False Nearest Neighbors (FNN), and through dimensional saturation of a system invariant. The FNN method provides a computationally efficient and robust algorithm for establishing an upper-bound on the set dimensionality, however, it does not provide noninteger estimates of dimension. Alternatively, examination of the slope of attractor inter-point correlation integrals can provide an estimate of the dimension in $\mathcal{R}$. Both of these methods rely upon access to a phase-space representation of the data at an arbitrary dimensionality; the particular level of dimensionality corresponding to a phase-space reconstruction is termed the embedding dimension d_E. Given that the observed data are single dimensional (a scalar time-series measured at a single spatial location) the appropriate recourse for reconstruction of phase-space is time-delay embedding.

3.1.1. Time-delay embedding

Time-delay embedding (Elshorbagy et al., 2002a) converts a discrete scalar time-series $s(n)$ into a d dimensional vector time-series:

$$\boldsymbol{\mu}(n) = [s(n), s(n+\tau), s(n+2\tau), \ldots, s(n+(d-1)\tau)]. \quad (7)$$

In order to select a reasonable time-delay τ for the embedding, one can simply compute the autocorrelation of the scalar dataset and use the value of the first zero-crossing of the autocorrelation. While this obviously cannot account for any nonlinearities in the statistical description of the interrelation between points in the time-series, it nonetheless provides a surprisingly robust initial estimate for an appropriate embedding delay as it prescribes on average the delay at which two points become statistically independent. The autocorrelation of the data from Fig. 1 is shown in Fig. 3. It is observed that in a linear sense, there is significant correlation ($R_{ss} \sim 0.8$) within the period of

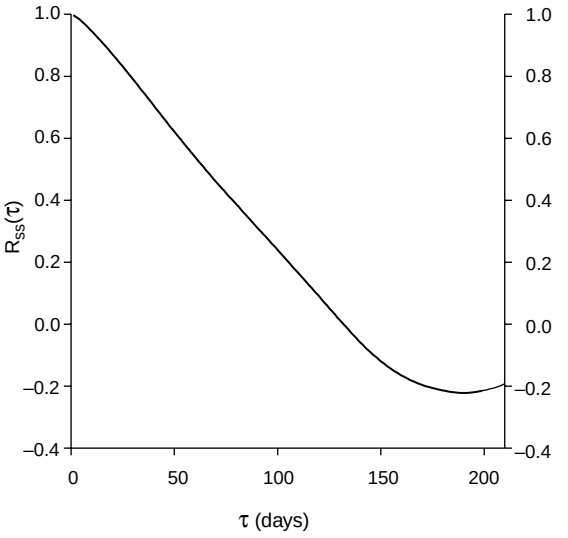

Fig. 3. Autocorrelation of stage data.

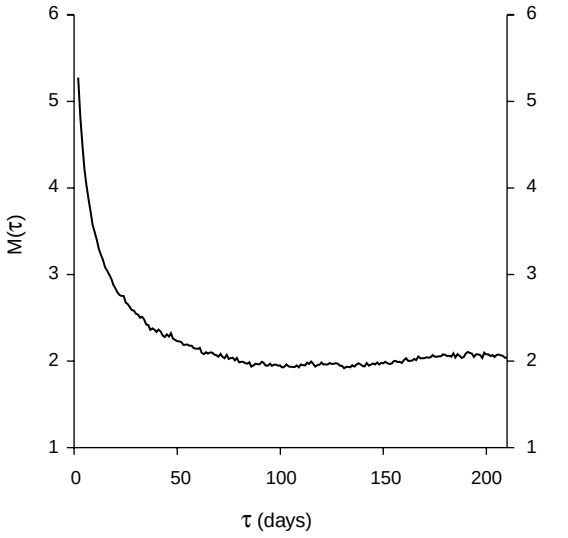

Fig. 4. Mutual information of stage data.

one month, while there is complete time-lag independence near $\tau = 130$ (days).

An alternative approach for establishing a time-delay is to estimate the mutual information, $M(\tau)$, between an observation at a time t and one at some later time $t+\tau$ (Fraser and Swinney, 1986). An advantage of this method is that it considers nonlinearities in the data when prescribing the amount of information that one observation adds after a previous observation. When no new information is apparent, the observations are completely independent and $M(\tau)=0$. Following (Gallagher, 1968), the mutual information can be specified as

$$M(\tau) = \sum_{n=1}^{N} P(s(n), s(n+\tau)) \log_2 \left[\frac{P(s(n), s(n+\tau))}{P(s(n))P(s(n+\tau))} \right] \tag{8}$$

Fig. 4 plots the mutual information of the stage data. The initial rate of amplitude decay is greater than that of the linear correlation, indicating that nonlinearities serve to devolve the mutual information between subsequent stage observations at a greater rate than predicted by a linear correlation. Note that $M(\tau)$ does not approach zero even for large τ, indicating that long-term evolution of stage data is not independent of previous levels, again in contradiction to the linear analysis which finds intervals of near-zero correlation. The usual prescription for choosing a time-delay is to select the first minimum of $M(\tau)$, which in this case may be selected in the area of $\tau = 90$–110 days.

3.1.2. False nearest neighbors

The method of false nearest neighbors (FNN) attempts to directly address the question: what embedding dimension is sufficient to eliminate false crossings of an orbit (phase-space trajectory) with itself as a result of having projected the attractor into too low a dimensional space? The procedure is to define a function of 'nearness' between adjacent points which depends solely on the geometrical arrangement imposed by the coordinate dimensions, and then iteratively increase the number of dimensions until one is satisfied that there are no 'false' nearest neighbors (Kennel et al., 1992). That is, the closest neighboring point has a distance which is not an artifact of having projected the attractor into too low a dimensional phase-space.

If the distance function between the point in question $\mu(n)$ and its nearest neighbor $\mu_{NN}(n)$ is simply a Euclidean distance, then the distance in dimension d is:

$$D_d = \Big[[\mu(n) - \mu_{NN}(n)]^2 + [\mu(n-\tau) - \mu_{NN}(n-\tau)]^2 + \cdots + [\mu(n+(d-1)\tau) - \mu_{NN}(n+(d-1)\tau)]^2 \Big]^{1/2} \tag{9}$$

As the data is embedded in the next higher dimension $(d+1)$, this nearest neighbor distance is changed due to the $(d+1)$ coordinates $\mu(n+d\tau)$ and $\mu_{NN}(n+d\tau)$ to

$$D_{d+1}(n) = \left[D_d(n)^2 + [\mu(n+\mathrm{d}\tau) - \mu_{NN}(n+\mathrm{d}\tau)]^2\right]^{1/2} \tag{10}$$

If $D_{d+1}(n)$ is large, one can assume that the 'nearness' of the two points is a result of the projection from some higher-dimensional attractor down to dimension d, since in going from dimension d to dimension $d+1$, we have 'unprojected' these two points. One is then faced with establishing a criterion to decide when neighbors are false, a normalized distance metric can serve this purpose such that when

$$\frac{\mu(n+\mathrm{d}\tau) - \mu_{NN}(n+\mathrm{d}\tau)}{D_d(n)} > D_{FNN} \tag{11}$$

the nearest neighbors at time index n are classified false. Here we use a threshold value of 15, which lies within the range of $10 < D_{FNN} < 50$ where the criteria is essentially constant (Abarbanel, 1993).

Fig. 5 presents the calculation of the percentage of FNN of all points as a function of embedding dimension d_E with $\tau = 90$ days. The results indicate that embedding the attractor of the scalar time-series into a 4 dimensional phase-space will suffice to ensure

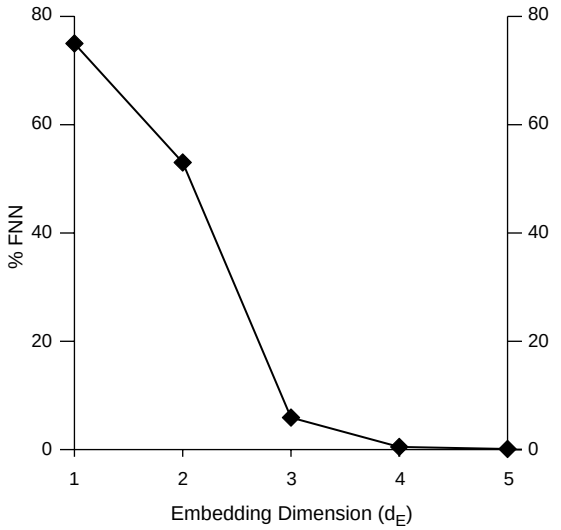

Fig. 5. False nearest neighbors of stage data.

that all trajectory orbits are valid representations of the attractor dynamics.

3.1.3. Correlation integrals

A popular procedure for establishing the dimension of a scalar time-series in $\mathcal{R}$ is to embed the time-series into a multiple of higher dimensions (Abarbanel, 1993) and then search for saturation of a system invariant as the embedding dimension increases. Such an invariant can in principle consist of any property associated with the attractor which depends on distances between points in the phase-space. A popular choice is an average over the attractor of moments of the number density. Define the number density, the number of points on the orbit within a radius of r of point μ in the phase-space, as

$$\eta(r,\mu) = \frac{1}{N}\sum_{n=1}^{N} \Theta(r - |\mu(n) - \mu|) \tag{12}$$

where Θ is the Heaviside function. The average over all points of powers of $\eta(r,\mu)$ defines the well-known correlation integrals (Grassberger and Procaccia, 1983), which for the case of interpoint distances reduces to the two-point correlation integral

$$C_2(r) = \frac{2}{N(N-1)}\sum_{i\neq j}^{N} \Theta\Big(r - |\mu(j) - \mu(i)|\Big) \tag{13}$$

The correlation integral C_2 is assumed to be a geometric invariant of the system, and we compute values of C_2 as a function of embedding dimension in order to establish the dimension at which the integral saturates (Grassberger and Procaccia, 1983). Fig. 6 presents computation of log (C_2) versus log (r) for d_E from 3 to 8. Examination of Fig. 6 reveals that when d_E reaches 4, the curves converge, indicating that the attractor dimension is less than or equal to 4. To estimate the dimension of the attractor the average of the slopes of the curves in regions with converged slopes is computed. It is noticed that there are two such regions in this plot: $\log(r) \in [6,10]$ and $\log(r) \in ([12,15])$. The difference between the two regions may be attributed to differences in scaling behavior of the system dynamics at these two length scales (Abarbanel, 1993;). In the region of $\log(r) \in [6,10]$ the dimension estimate is $d = 1.54$, while over the region of $\log(r) \in [12,15]$ the value is

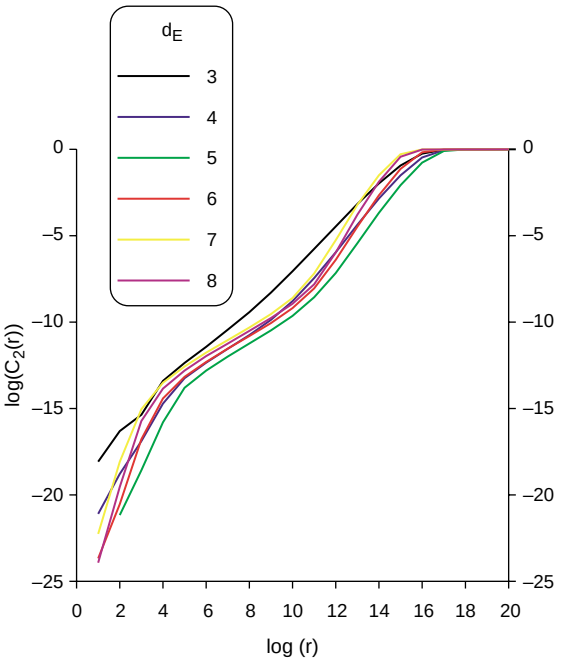

Fig. 6. Correlation integrals of stage data.

$d=3.78$. Based on the results of the FNN computation, which clearly indicated that $d>2$, the value of $d=3.78$ will be estimated to represent the attractor dimensionality. To the extent that the dimensional estimate is accurate, its fractional value provides evidence that the attractor is a fractal set, also known as a 'strange attractor', and therefore, may exhibit chaotic behavior. It may be of interest to note that among the many reported dimensional estimates of hydrologic processes (Sivakumar, 2004), apparently few have applied analysis to hydrologic stage/volume data. The findings of Sangoyomi et al. (1996) indicate a dimension of 3.4 for volume variations in the Great Salt Lake, which is similar to the results we obtained for the Loxahatchee National Wildlife Refuge.

3.2. Lyapunov exponents

While the dimension is useful for characterizing the complexity and distribution of points in the phase-space, it sheds no light on the dynamics of evolving trajectories of such points. For this, one can turn to the Lyapunov exponents (Oselecdec, 1968), λ_i, which quantify the rate of growth of elemental subspaces in the phase-space. λ_1 relates the rate at which linear distances grow between two points on the attractor: two such points separated initially by an infinitesimal distance δ will, on average, have their separation grow as $\delta\, e^{\lambda_1 t}$. The sum $\lambda_1+\lambda_2$ dictates the average growth rate of two-dimensional areas, and in general, the behavior of d-dimensional subspaces is described by the sum of the first d exponents. Accordingly we define the cumulative exponent as

$$\Lambda = \sum_{i=1}^{d} \lambda_i. \tag{14}$$

Since the λ_i govern the rate of attractor expansion and contraction, it is clear that for physically stable systems it not possible for Λ to be positive, this would indicate a globally unbounded behavior. The presence of positive exponents is however, the hallmark of chaos, they ensure that long-term prediction of individual trajectories is limited owing to the exponential divergence of neighboring points from infinitesimal differences in initial conditions. Negative exponents arise when the trajectories of subspaces contract. If the system is a dissipative one, then Λ would be negative, ensuring that as time progresses the attractor remains bounded in a finite phase-space. A Hamiltonian system would exhibit $\Lambda=0$, indicating an energy balance between the expansive and contractive dynamics.

The Lyapunov exponents are computed following Eckmann (Eckmann and Ruelle, 1985; Eckmann et al., 1986), for the data of Fig. 1 embedded into a 4 dimensional attractor. Fig. 7 plots the Lyapunov spectrum of the embedded data and reveals both positive and negative components, as well as $\Lambda<0$. The presence of a positive exponent reveals the sensitive dependence on initial conditions and trajectory divergence which is associated with chaos, further, since $\Lambda<0$ it is known that the attractor is globally stable.

3.3. Phase-space attractor

The global phase space attractor of a nonlinear dynamical process is itself a dynamical invariant. Understanding the structure of the attractor may lead to physical insights regarding the process dynamics, though there is not a prescriptive algorithm for

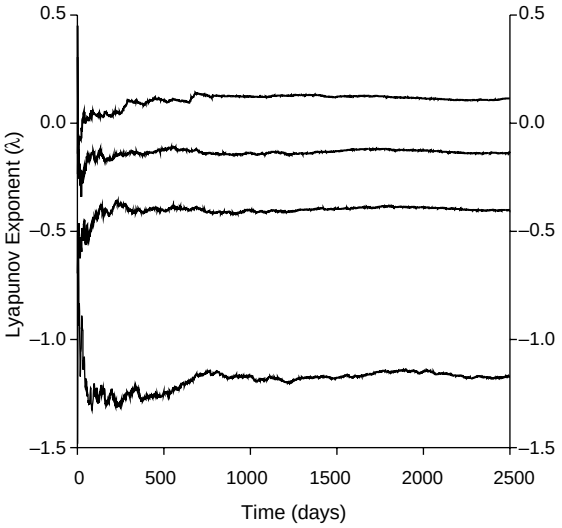

Fig. 7. Lyapunov spectrum of stage data.

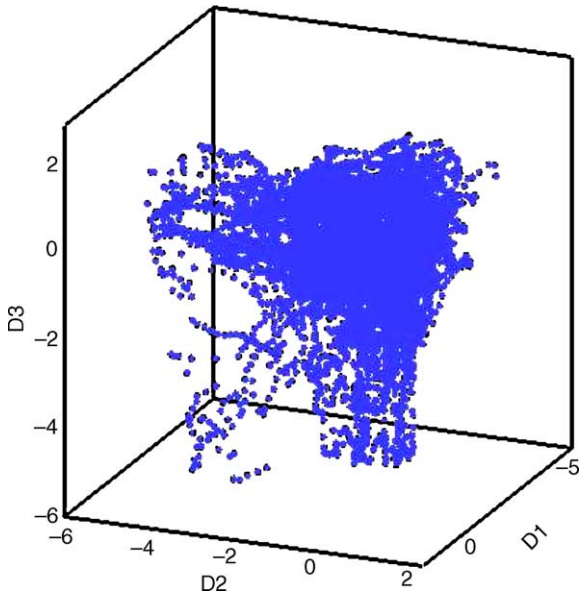

Fig. 8. Phase-space attractor.

extraction of physical variables from those of the phase space. In cases where this is feasible, one stands to gain useful knowledge derived from the data itself, which may then be applied in construction of more accurate dynamical models. In the case of the WCA1 stage data, it is not clear what physical variables correspond to the four necessary dimensions identified by the invariant analysis, although logical candidates would be rainfall, anthropogenic inflows and outflows, previous stage, and local soil hydrology. In order to provide a view of the stage data attractor, a plot of the first 3 phase-space dimension variables from the reconstructed phase space vector $\boldsymbol{\mu}$ of Eq. (7) is shown in Fig. 8, where the phase space variable D_1 corresponds to the component μ_1. This figure indicates that there is an overall structure to the global attractor trajectories, and confirms that the attractor occupies a finite phase space volume required for dynamic stability.

The attractor of Fig. 8 contains some 12,000 points derived from 35 years of daily observations, thus making it difficult to discern any structure inherent in individual trajectories. To provide a view of yearly trajectories, Fig. 9 plots four time slices from the global attractor, each time slice corresponding to a period of 1 year. Examination of this figure reveals a clear structure to the yearly trajectories.

4. Nonlinear estimation

There exist a plethora of nonlinear estimation techniques which can be applied to nonlinear models or data. If one has established a phase space attractor which encapsulates the nonlinear dynamics, an obvious choice is to make use of the attractor trajectories in constructing a model of the attractor, from which phase space forecasts can be made given knowledge of previous state values. The idea would be to define a mapping function $\mathbf{F}(\mu,\alpha,p)$ able to predict p values forward the f trajectory components based on the current and previous orbit points. The α are model parameters fit via a least squares or some other criteria to converge the model with the observed attractor data. In the case where the attractor is embedded in a phase space for which the independent variables are known and connected to physical parameters, one is likely to have success in construction of the model basis functions, and should be in the favorable position of being able to attach physical significance to parametric regimes of α.

In the current analysis however, we are dealing with a reconstructed attractor, derived from a reconstructed phase space based on time-delay

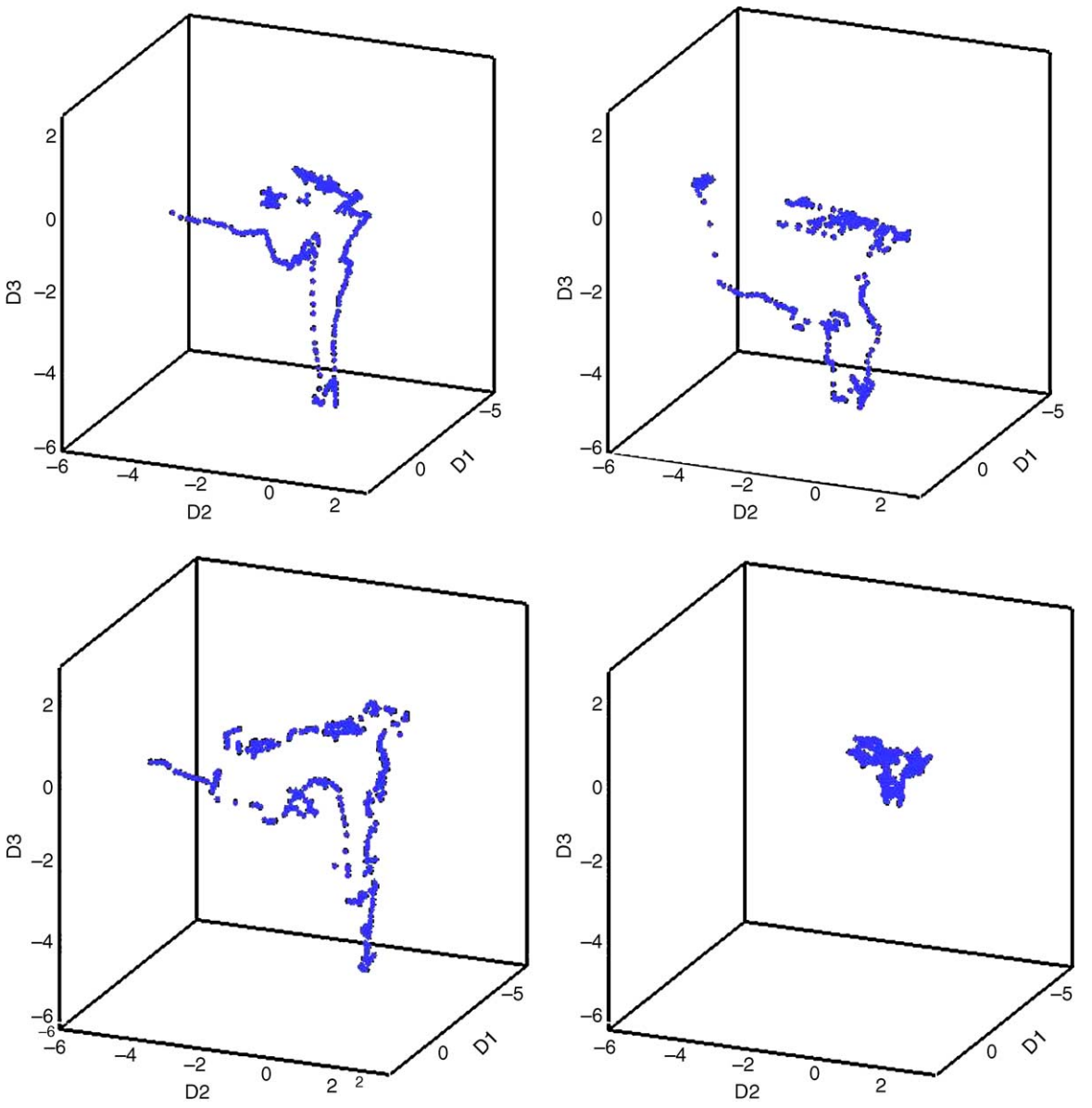

Fig. 9. Phase-space attractors, 4 single years.

embedding of a scalar time-series. It would have been preferable to have access to multi-dimensional observations with a phase space constructed from known independent variables, from which the attractor basis functions and fit parameters could be inferred. Nonetheless, we have still derived important information from the dynamical invariants which can be used to bound the predictability as well as provide clues to the dimensionality for use by nonlinear estimators. The following sections quantify the stage predictability, and implement an artificial neural network to predict the stage given inputs based on physical observations guided by the dimensionality of the attractor.

4.1. Nonlinear predictability

Given knowledge of the of the Lyapunov invariants for a stable chaotic attractor, a natural question to pose concerning the predictability of the nonlinear time-series is: how far in the future can one reasonably expect to be able to make useful predictions of a particular trajectory? To asses this question we can draw on the realization that a dynamical system exhibiting chaotic behavior is an exact analog of Shannon's concept of an ergodic information source. Consider a discrete information source which produces a sequence of symbols drawn from an alphabet $s=[S_1,S_2,\ldots,S_N]$, and denote the distribution of the realized sequence within s as $P(s)$. The amount of information that is transmitted by measurement of the sequence (or alternatively the amount of uncertainty removed at the receiver) can be quantified by an informational entropy as elucidated by Shannon (1948). In the case where a joint set of measurements are performed, the entropy in bits is defined via the joint probability $P(s_1,s_2,\ldots,s_N)$ as

$$H_N(s) = -\sum_{s_j} P(s_1, s_2, \ldots, s_N)\log[P(s_1, s_2, \ldots, s_N)]. \tag{15}$$

As N becomes large, this quantity normalized by N has a finite limit:

$$h(s) = \lim_{N\to\infty} \frac{H_N(s)}{N}. \tag{16}$$

Now consider the variable s to be the possible state-space coordinates of a dynamical system embedded within a finite-dimensional attractor. Let the measurements of this variable S_i define the distribution of attractor trajectories in the state-space. A joint measurement of the S_i corresponds to following the trajectories of N orbits through the dynamics of the attractor. In this situation the source is essentially an abstract statement of the system producing the measurements, and the concept of entropy of a deterministic dynamical system, where no concept of probability is immediately evident, is much the same as in the information-theoretic setting. This connection was made by Kolmogorov (1960) who defined $h(s)$ as above, and this information measure is known as the Kolmogorov-Sinai (KS) entropy of the dynamical system. The KS entropy is an invariant of the system trajectories, and so is independent of the initial conditions or the specific trajectories observed. Further, it can be shown that the sum of positive Lyapunov exponents Λ_+ is equal to the KS entropy (Pesin, 1977). This connection allows one to make a statement regarding the limits of predictability for a nonlinear system in a local sense.

Consider an arbitrary phase-space for a stable dynamical system with sufficient dimension to completely unfold the attractor. Within this phase-space, any realizable statement regarding the system can only be specified to a within a certain accuracy. Within that resolution cell we cannot distinguish between two distinct state-space points. However, any nonlinear system with a positive KS entropy has a degree of intrinsic instability, and the two points occupying the same resolution cell will, after a time T, move to disparate and individually resolvable cells in the state-space. As a rough upper-limit, one can argue that when the number of states occupied by the system is equal to approximately the total number of states available for orbits of the system, that the ability to predict further state-evolution is lost. A measure of the total number of states that a d dimensional system occupies after the passage of time T is $d^{h(s)T}$, therefore, a reasonable assessment for the length of time a particular trajectory remains 'predictable' would be $T\approx 1/h(s)$.

Reference to Fig. 7 reveals a value of $\Lambda_+ \approx 0.12$. Based on this value, the predicted time interval for which evolution of a particular phase-space trajectory is tenable is approximately 8.3 days. This value is considerably lower than one might infer from linear analysis based on stage autocorrelation.

4.2. Neural network stage predictor

Artificial neural networks (ANN) have evolved into a large and diverse class of computational architectures capable of modeling many complex, nonlinear processes. Advantages of these tools include the ability to learn complex, intractable mappings without the requirement of an explicit physical model or functional description of the process. ANN also have the ability to store information, and can serve well as inference engines able to combine learned information in response to

unknown parametric regimes. ANN are computationally efficient predictors with an inherently parallel-processing structure suited to rapid state estimation. Difficulties with the ANN include the lack of assurance that a global error minimum is achieved, input parameter sensitivity issues, and the possibly protracted training intervals. ANN are generally classified as either feedforward, or recurrent networks, and are either supervised or unsupervised in the training phase. In this application, we employ a supervised, feedforward network known as the perceptron (Wasserman, 1989) as a predictor for stage values in the WCA1.

The perceptron requires a training phase prior to use as a predictive algorithm in order to organize the neuron interconnection weight-states into a structure through which the relevant information extraction or recognition can be achieved. The process is essentially a search through the parametric information space encapsulated in the training data so as to produce a minimum error output dictated by the supervisor. The supervisory function consists of the Euclidean error: $\varepsilon = (T_i\text{-}O_i)$, where T indicates the supervisory output of the ith output neuron, and O the actual output of the ith unit during a training cycle. The training data consists of a vector of input-output pairs, each component is presented to the network which is trained with the backpropagation (Wasserman, 1989) gradient-descent algorithm, using a fixed momentum parameter of 0.1, and a learning rate of 0.001. The ANN activation function for the hidden layer neurons is a sigmoidal Bernoulli function with a zero-argument slope of 5.0, while the output layer units employ a simple linear scaling with a unit derivative. Details of the ANN internal computational algorithms can be found in (Park and Abusalah, 1997).

The purpose of the ANN is to predict the stage values at some future time, therefore, the output of the ANN will consist of a single output unit. As discussed in Section 3.3, it is expected that rainfall, time-of-year, and previous stage are physically relevant parameters in determination of the current and future stage values. These variables are available as historic time-series records as depicted in Figs. 1 and 2 and will be used as input data for the ANN. If training the ANN to learn and predict the phase space trajectories were based only on the scalar time-series itself as input data, one would configure the ANN to accept d inputs of delayed vectors $[\boldsymbol{\mu}(k),\boldsymbol{\mu}(k-1),\ldots,\boldsymbol{\mu}(k\text{-}d+1)]$, with the output set to predict $[\boldsymbol{\mu}(k+p)]$ where p is the prediction index (Abarbanel, 1993). This prescription takes advantage of the fact that an upper limit of the dynamical attractor dimension is known, and provides an indication of the number of independent phase-space variables required for analysis of the dynamics. Dimensional analysis of the stage attractor indicated an embedding dimension of $d_E = 4$ was required for accurate phase space reconstruction. Following the recommendations of Abarbanel (1993), we employ the current stage, and three previous stage values as inputs to the ANN. In addition, the day of year and rainfall are also input to the ANN. Based on these assessments, the ANN is configured to have 1 output neuron, 6 input neurons, and 2 hidden layers with 35 neurons per hidden layer. A schematic depiction of the ANN is shown in Fig. 10. The selection of 35 hidden layer neurons was based on trial runs with 10, 15, 25 and 35 hidden layer neurons, from which it was observed that 10 hidden layer neurons were incapable of converging the network to a usable neuron weight-state. The selection

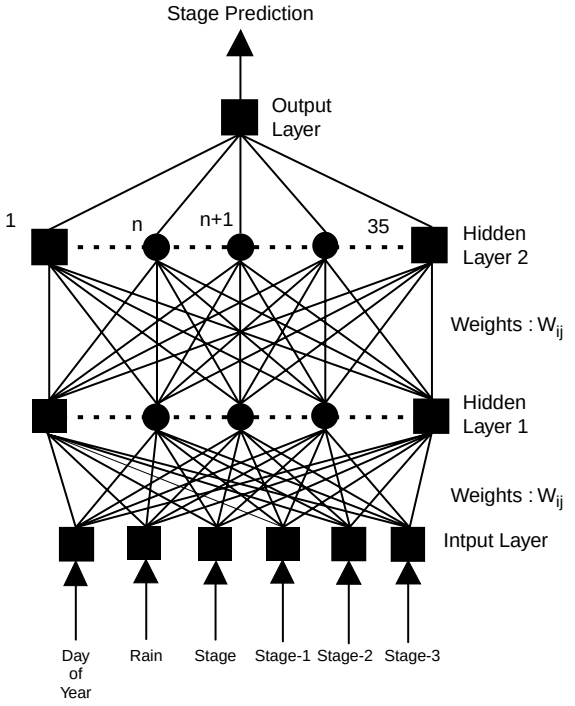

Fig. 10. Nonlinear estimator: artificial neural network.

of 35 is intended to provide ample network weights such that the learning is capable of fully adjusting and converging the network weights to a pragmatic prediction state.

In Section 4.1 it was found that a forecast period of 8.3 days is expected to be a workable limit for accurate stage data predictions. Beyond this, the attractor dynamics indicate that trajectories are capable of transitioning to any point in the phase space, thereby defeating reliable prediction. In accordance with this expectation four training sets were constructed from the available data corresponding to prediction intervals of 1, 2, 8 and 14 days. The training set consists of 730 input–output patterns taken from the beginning of the time-series shown in Figs. 1 and 2, corresponding to a time interval of 2 years. After the ANN was trained, the resultant neural weight matrix was saved for prediction runs. A forecast was made by running the trained network on 6000 days of input data, starting at day index 1000. The network then produces a time-series output of predicted stage data at future intervals of 1, 2, 8 or 14 days. The predicted time-series for the 1 day forecast interval is shown with the actual values in Fig. 11. The prediction is quite good, essentially indiscernible to the eye at the scale shown.

In order to quantify the forecast accuracies, the time-series error magnitude ($\bar{\varepsilon} = |(T_i - O_i)|$) for each ANN forecast is presented in Fig. 12, each forecast is labeled with the RMS error of the time-series, as well as the correlation coefficient between the actual and predicted timeseries. For the forecast periods of 1 and 2 days, the mean error is low, less than 1 inch for the 1 day forecast period. The 2 day forecast mean error is a factor of 1.7 larger than the 1 day forecast mean error. The 8 day forecast interval has a mean error which is 5.3 times that of the 1 day forecast, the 14 day forecast is greater by a factor of 7.5. Based on the analysis of the KS entropy derived from the observed stage data, a value of approximately 8 days was found to represent the point beyond which phase space trajectory forecasting would be problematic. The question of whether or not an mean prediction error in excess of 0.36 feet is acceptable, is an operational decision, nonetheless, the results are consistent with the expectation of forecasting degradation beyond the 8 day forecast interval.

4.3. Linear model stage prediction

In order to compare the nonlinear model forecasts to conventional linear models, two linear models were

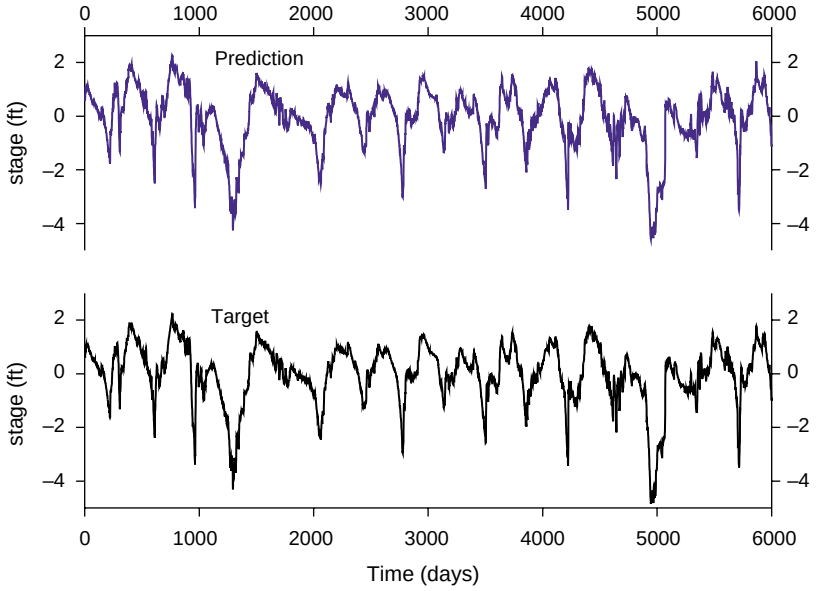

Fig. 11. ANN stage prediction: 1 day forecast.

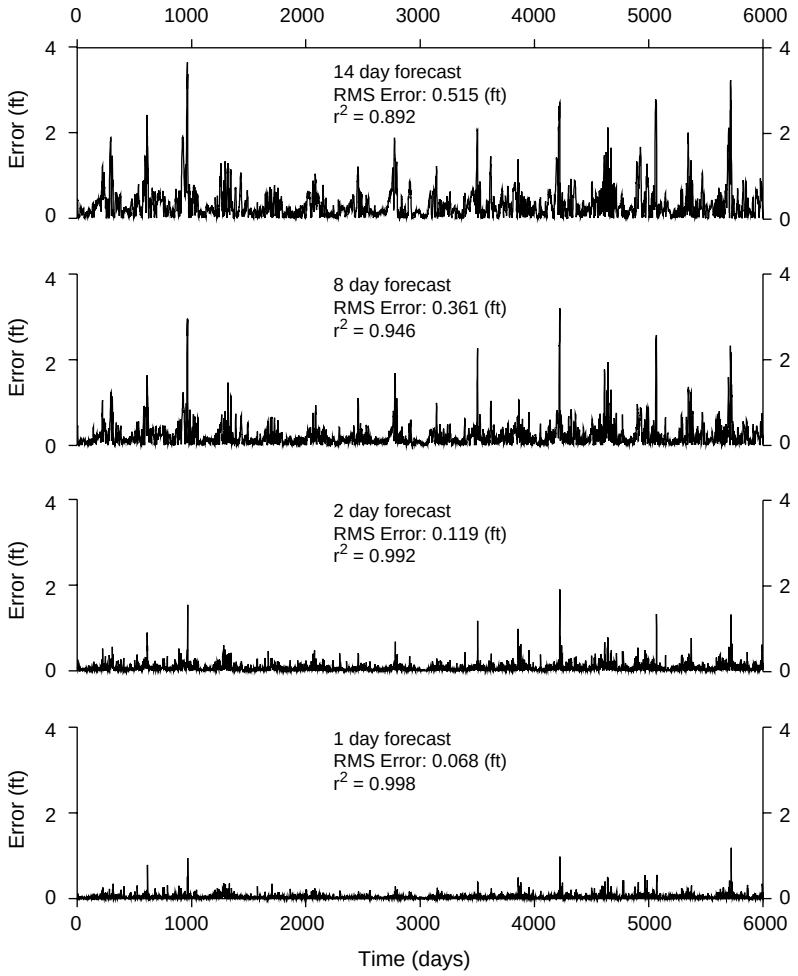

Fig. 12. ANN stage prediction errors.

constructed and used to predict the stage variations: (1) an ARMA model, and (2) a Kalman filter (KF).

4.3.1. ARMA stage prediction

The ARMA model is constructed according to

$$s(t+t_p) = s(t) + \frac{\Delta s}{\Delta t} t_p + R_t \tag{17}$$

where s represents the stage value, t the current time, t_p the prediction time interval, $(\Delta s/\Delta t)$ a best fit slope to the previous stage values $s(t)$, $s(t\text{-}1)$, $s(t-2)$, $s(t\text{-}3)$, and R_t the rainfall contribution to the increase in stage for the current time step.

4.3.2. Kalman filter stage prediction

The Kalman filter is a recursive linear filter, widely used for linear (and in the extended form, for nonlinear) state estimation (Kalman, 1960; Zarchan and Musoff, 2000). The KF estimates the state σ of a d_σ dimensional discrete time process based on measurements m of the d_m dimensional process observables, the state can be expressed via the linear stochastic difference equation:

$$\sigma_i = A\sigma_{i-1} + w_{i-1} \tag{18}$$

where the measurement is represented by

$$m_i = H\sigma_i + v_i \tag{19}$$

A is the $d_\sigma \times d_\sigma$ matrix which relates the previous state to the current one, and w the process noise which is assumed to be a zero mean, normally distributed disturbance with covariance Q. H represents the $d_m \times d_\sigma$ matrix that relates the state σ_i to the observation m_i. The KF attempts to solve an equation which estimates the state $\hat{\sigma}_i$ as a linear combination of a prior state estimate $\hat{\sigma}_{i-1}$ and a weighted difference between an observation m_i and a predicted observation $H\hat{\sigma}_{i-1}$:

$$\hat{\sigma}_i = \hat{\sigma}_{i-1} + K(m_i - H\hat{\sigma}_{i-1}) \tag{20}$$

The $d_\sigma \times d_m$ matrix K is the Kalman gain, which is selected to minimize the covariance of the estimate error $e_i = \sigma_i - \hat{\sigma}_i$.

The KF implementation proceeds by defining the process state vector to consist of the $d_\sigma = 7$ variables: day of year, rain, predicted stage, current stage, and 3 previous stage values. The predicted stage is not considered an observed variable. The A and H are assigned to implement a model wherein each variable consists of a simple linear combination of previous state variables. The process covariance Q is obtained from the stage timeseries of Fig. 1. The observation covariance is set to a low value of 0.01. A training set of the first 730 points of the stage and rain timeseries is applied to the KF in order to 'learn' the linear process dynamics (Digalakis et al., 1993). The configured KF is then run on the 16 years stage data to forecast the stage values.

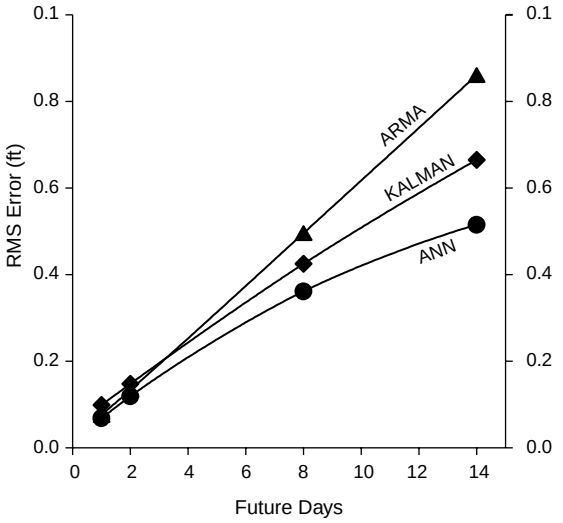

Fig. 13. Comparison of ANN, ARMA and KF prediction errors.

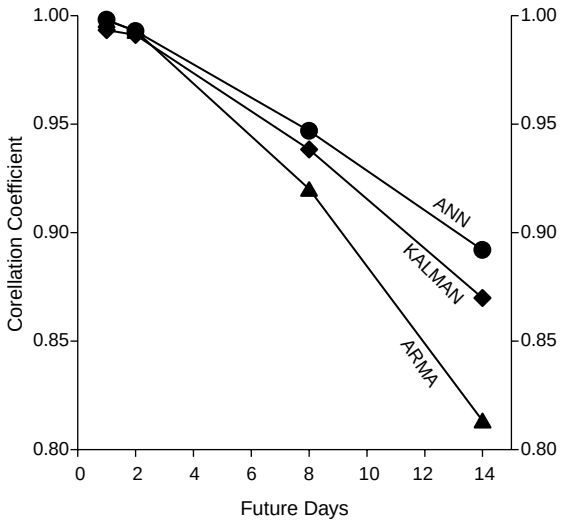

Fig. 14. Comparison of ANN, ARMA and KF correlation coefficients.

The ARMA and KF models were used to forecast the stage values at predictive intervals of 1, 2, 8 and 14 days. Fig. 13 depicts the RMS error of each of the predictive schemes, Fig. 14 plots the correlation coefficient of each of the forecast timeseries with the observed data. It can be observed that for forecast periods of 1–2 days, the models are essentially equivalent in terms of their mean error. As the forecast interval increases, the linear models exhibit a linear increase in forecast error, while the ANN results in a lower RMS error. These results are consistent with the assertion that a purely linear model may be less effective than a nonlinear model in capturing the dynamics of hydrological chaotic processes.

5. Conclusion

In this paper we have applied the methodology of dynamical systems theory to analysis of hydrologic stage data in the Loxahatchee National Wildlife Refuge. Computation of dynamical system invariants indicates that the stage response constitutes a chaotic signal arising from a fractal attractor with a dimension estimated to be 3.78. Evaluation of the global Lyapunov exponents, in conjunction with the

Kolmogorov-Sinai entropy, reveals a stage predictability limit of roughly one week. This assessment is perhaps shorter than one would expect from examination of a linear correlation function, which maintains a 90% correlation out to a time interval of 16 days. Examination of the mutual information function reveals a sharp decline of relative information in the first week of time-lag, indicating that nonlinear interactions limit the predictability in relation to a linear statistical model. Additionally, the mutual information stabilizes at nonzero values for increasing time-lags indicating that significant information is maintained between hydrologic stage values of widely differing time scales. This is in contradiction to the linear autocorrelation that exhibits near zero values indicating statistical independence at large time scales. Such comparisons indicate the importance of examining the assumptions of linearity in the application of statistical analysis, model building and forecasting to natural processes.

To address the issue of nonlinear estimation, an artificial neural network was constructed to learn a two year period of stage dynamics. It was assumed that local rainfall, previous stage, and time of year were important system variables. The ANN was able to successfully learn the system dynamics, and was applied as a nonlinear forecast tool to estimate the stage at forecast intervals of 1, 2, 8 and 14 days over a period of 16 years. Additionally, linear ARMA and Kalman filter models were implemented to forecast the stage at the same forecast intervals. Consistent with the identification of an 8 days limit to predictability based on the Kolmogorov-Sinai entropy of the attractor dynamics, the model results demonstrated significant increases in forecast error at forecast intervals in excess of 8 days. Model results demonstrate essentially no differences in stage forecasting between the linear and nonlinear models for forecast periods of 1 and 2 days. For the 8 and 14 days forecast, the ANN is observed to produce lower mean errors than the linear models. These results can be interpreted as consistent with the notion that nonlinear models can provide fuller encapsulation of chaotic dynamics than simple linear models.

Primary findings of this work are: (1) Hydrologic stage data in the Loxahatchee National Wildlife refuge observed at a daily sample interval arise from a nonlinear dynamical system with a fractal dimension estimated to be 3.78. (2) The dominant Lyapunov exponent of the dynamical system in conjunction with the Kolmogorov-Sinai entropy results in an estimated limit of stage data of 8.3 days. (3) An artificial neural network was implemented to learn the system dynamics and produce forecast values with lower mean errors than linear models for prediction periods in excess of 8 days. Notably, the ANN 16-year daily time-step prediction consumed 27.2 s on a 990 MHz Pentium III based PC. The accuracy and efficiency of these results indicate that forecasting of chaotic stage data based on dynamical systems analysis and application of nonlinear estimators such as an ANN, provides an alternative to linear models which may have difficulty incorporating the relevant nonlinear physical processes.

References

Abarbanel, H., et al., 1993. The analysis of observed chaotic data in physical systems. Rev. Mod. Phys. 65 (4), 1331–1392.

Digalakis, Rohlicek, Ostendorf, 1993. ML Estimation of a stochastic linear system with the EM algorithm and its application to speech recognition. IEEE Trans. Speech Audio Proc. 1 (4), 431–442.

Eckmann, J.P., Ruelle, D., 1985. Ergodic theory of chaos and strange attractors. Rev. Mod. Phys. 57, 617–656.

Eckmann, J.P., Oliffson Kamphorst, S., Ruelle, D., Ciliberto, S., 1986. Lyapunov exponents from a time series. Phys. Rev. A 34, 4971–4979.

Elshorbagy, A., Simonovic, S.P., Panu, U.S., 2002a. Estimation of missing streamflow data using principles of chaos theory. J. Hydrol. 255, 123–133.

Elshorbagy, A., Simonovic, S.P., Panu, U.S., 2002b. Noise reduction in chaotic hydrologic time series: facts and doubts. J. Hydrol. 256, 147–165.

Fraser, A.M., Swinney, H.L., 1986. Independent coordinates for strange attractors from mutual information. Phys. Rev. A 33, 1134–1140.

Gallagher, R.G., 1968. Information theory and reliable communication. Wiley, New York.

Grassberger, P., Procaccia, I., 1983. Characterization of strange attractors. Phys. Rev. Lett. 50, 346–349.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. Trans. ASME J. Basic Eng. 82, 35–45.

Kennel, M.B., Brown, R., Abarbanel, H., 1992. Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys. Rev. A 45, 3403–3411.

Kolmogorov, A.N., 1958. Dokl. Akad. Nauk. SSSR 119, 861 (Math. Rev. 21, p386 (1960)).

Lambrakis, N., et al., 2000. Nonlinear analysis and forecasting of a brackish karstic spring. Water Resour. Res. 36 (4), 875–884.

Lorenz, E., 1963. Deterministic nonperiodic flow. J. Atmos. Sci. 20, 130–141.

Oselecdec, V.I., 1968. A multiplicative ergodic theorem, Lyapunov characteristic numbers for dynamical systems. Trudy. Mosk. Mat. Obsc. 19, 197–231 (Trans. Mosc. Math. Soc.).

Park, J.C., Abusalah, S., 1997. Maximum-entropy: A special case of minimum cross-entropy applied to nonlinear estimation by an artificial neural network. Complex Syst. 11, 289–307.

Pesin, Y.B., 1977. Lyapunov characteristic exponents and smooth ergodic theory. Usp. Mat. Nauk. 32 (4), 55–112 (Russian Math. Survey).

Sangoyomi, T.B., Lall, U., Abarbanel, H.D., 1996. Nonlinear dynamics of the Great Salt Lake: dimension estimation. Water Resour. Res. 32 (9), 2825–2839.

Shannon, C., 1948. A mathematical theory of communication. Bell Sys. Tech. J. 27, 379–423 see also pp. 623–656.

Sivakumar, B., 2000. Chaos theory in hydrology: important issues and interpretations. J. Hydrol. 227, 1–20.

Sivakumar, B., 2004. Choas theory in geophysics: past, present and future. Chaos Solitons and Fractals 19, 441–462.

Sivakumar, B., Jayawardena, A.W., Fernando, T.M.K.G., 2002. River flow forecasting: use of phase-space reconstruction and artificial neural networks approaches. J. Hydrol. 265, 225–245.

Wasserman, P.D., 1989. Neural computing. Van Nostrand Reinhold, New York.

Zarchan, P., Musoff, H., 2000 Fundamentals of kalman filtering: A practical approach. American Institute of Aeronautics and Astronautics (AIAA), ISBN: 1563474557.