

# Aggregation and sampling in deterministic chaos: implications for chaos identification in hydrological processes

J. D. Salas<sup>1</sup>, H. S. Kim<sup>2</sup>, R. Eykholt<sup>3</sup>, P. Burlando<sup>4</sup>, and T. R. Green<sup>5</sup>

<sup>1</sup>Department of Civil Engineering, Colorado State University, Fort Collins, Colorado, USA

<sup>2</sup>Department of Civil Engineering, Inha University, Incheon, Korea

<sup>3</sup>Department of Physics, Colorado State University, Fort Collins, Colorado, USA

<sup>4</sup>Institute of Hydromechanics & Water Resources Management, ETH Hoenggerberg, 8093 Zurich, Switzerland

<sup>5</sup>Agriculture Research Service, USDA, Fort Collins, Colorado, USA

Received: 23 July 2004 – Revised: 18 March 2005 – Accepted: 19 March 2005 – Published: 7 June 2005

Part of Special Issue “Nonlinear deterministic dynamics in hydrologic systems: present activities and future challenges”

**Abstract.** A review of the literature reveals conflicting results regarding the existence and inherent nature of chaos in hydrological processes such as precipitation and streamflow, i.e. whether they are low dimensional chaotic or stochastic. This issue is examined further in this paper, particularly the effect that certain types of transformations, such as aggregation and sampling, may have on the identification of the dynamics of the underlying system. First, we investigate the dynamics of daily streamflows for two rivers in Florida, one with strong surface and groundwater storage contributions and the other with a lesser basin storage contribution. Based on estimates of the delay time, the delay time window, and the correlation integral, our results suggest that the river with the stronger basin storage contribution departs significantly from the behavior of a chaotic system, while the departure is less significant for the river with the smaller basin storage contribution. We pose the hypothesis that the chaotic behavior depicted on continuous precipitation fields or small time-step precipitation series becomes less identifiable as the aggregation (or sampling) time step increases. Similarly, because streamflows result from a complex transformation of precipitation that involves accumulating and routing excess rainfall throughout the basin and adding surface and groundwater flows, the end result may be that streamflows at the outlet of the basin depart from low dimensional chaotic behavior. We also investigate the effect of aggregation and sampling using series derived from the Lorenz equations and show that, as the aggregation and sampling scales increase, the chaotic behavior deteriorates and eventually ceases to show evidence of low dimensional determinism.

## 1 Introduction

Modeling the serendipitous behavior of streamflow series has been an important subject in hydroscience and water resources engineering for several decades, particularly for synthetic simulation and forecasting problems. For this purpose, several concepts and modeling techniques have been suggested in the literature. For example, stationary and periodic autoregressive moving average stochastic models have been used widely (Salas, 1993). Often, the modeling approach for such streamflow series involves: (1) transforming the original series into an approximately normal distributed series, (2) removing any periodicity in the mean and the variance using seasonal standardization (e.g. in the case of monthly data), and (3) fitting a stationary or periodic autoregressive (AR) or autoregressive and moving average (ARMA) model to the residual series. However, while such modeling schemes have been quite successful for annual and seasonal data (e.g. annual and monthly precipitation and streamflows), their use for modeling precipitation and streamflows on smaller time scales, such as daily or hourly, has been less successful. On short time scales, precipitation series are intermittent, so streamflows exhibit the influences of individual rainstorms, which generate rapid rises and slow recessions. Modeling these features stochastically has been more complex, and this has received less attention in the literature (e.g. Cowpertwait and O’Connell, 1992).

In recent years, nonlinear stochastic models (e.g. Tong, 1990) and deterministic nonlinear dynamics and chaos (e.g. Fraedrich, 1987; Rodriguez-Iturbe et al., 1989) have been suggested as modeling alternatives for atmospheric and hydrological processes. Since then, numerous examples for detecting low dimensional chaos using a variety of metrics (e.g.

Correspondence to: J. D. Salas  
(jsalas@engr.colostate.edu)

the correlation dimension and Lyapunov exponents), and applications of deterministic chaos to forecasting various types of hydrologic and atmospheric time series (for time scales of 15 seconds, 15 minutes, hourly, daily, weekly, biweekly, and monthly), such as rainfall, streamflow, lake volumes, tree ring indices, temperature, and southern oscillation index series, can be found in literature (e.g. Rodriguez-Iturbe, 1991; Sharifi et al., 1990; Wilcox et al., 1991; Islam et al., 1993; Jayawardena and Lai, 1994; Berndtsson et al., 1994; Sangoyomi et al., 1996; Lall et al., 1996; Puente and Obregon, 1996; Jeong and Rao, 1996; Porporato and Ridolfi, 1997; Kawamura et al., 1998; Liu et al., 1998; Kim et al., 2001; Sivakumar et al., 2001a; Jayawardena et al., 2002; Zhou et al., 2002).

However, a review of the literature indicates conflicting results regarding the existence and inherent nature of chaos in precipitation and streamflow series, i.e. whether they are low dimensional chaotic or stochastic (e.g. Koutsoyiannis and Pachakis, 1996; Pasternack, 1999; Sivakumar et al., 1999a; Sivakumar, 2000; Sivakumar et al., 2001a; Schertzer et al., 2002). Some hydrologists have discussed the difficulties and uncertainties in discriminating between low dimensional chaotic and stochastic systems using such metrics as Lyapunov exponents and the correlation dimension (e.g. Ghilardi and Rosso, 1990; Sivakumar, 2001). Although other metrics have been suggested and utilized, such as Kolmogorov entropy (e.g. Jayawardena and Lai, 1994; Koutsoyiannis and Pachakis, 1996; Elshorbagy et al., 2002), close returns plots (e.g. Kim et al., 2001), and the C-C method (Kim et al., 1999), the issue of discriminating between low dimensional chaotic and stochastic systems remains complex. The various factors that have been cited in the literature that make such discrimination unclear and inconclusive include primarily the effect of errors (noise), sample size, and intermittence (the presence of zeros). Thus, a number of noise reduction methods have been developed and utilized (e.g. Schreiber and Grassberger, 1991; Schreiber, 1993; Porporato and Ridolfi, 1997; Kantz and Schreiber, 1997; Sivakumar et al., 1999b; Jayawardena and Gurung, 2000; Elshorbagy et al., 2002). Their effects on estimating the correlation dimension have been discussed in some detail (Sivakumar, 2001); and their potential benefits in estimating the chaotic invariants and their practical significance for hydrologic applications have been questioned and critically examined (Elshorbagy et al., 2002).

The intermittence in hydrologic data has been another factor that affects the estimation of chaotic invariants and applications thereof. It has been largely ignored in many studies, but has been recognized by some investigators. For instance, Koutsoyiannis and Pachakis (1996) illustrated the distortion caused by zero rainfall in the estimate of the lower tail of the correlation integral and suggested using the generalized entropy of Cantorian dust to eliminate the problem caused by zeros. Also, Sivakumar (2001), in examining the dynamics of 1-day, 2-day, 4-day, and 8-day rainfall, discussed the effect that data with a large numbers of zeros (e.g. daily rainfall) may have on estimating the correlation dimension. It

appears that a significant underestimation occurs for daily rainfall, and the underestimation diminishes as the time scale increases (or as the number of zeros decreases). Also the effect that zeros (in data) have on identifying low dimensional chaos of hydrologic data, especially in connection with rainfall disaggregation, has been investigated by Sivakumar et al. (2001b). Furthermore, Schertzer et al. (2002) argued that contrary to deterministic chaos, stochastic multifractal processes can easily simulate spatial geophysical systems involving intermittency.

Additionally, the sample size of hydrologic data has always been an issue for estimating many statistical properties. For example, it is well known that estimators of second, third, and higher order moments have large sampling variance and may have significant biases (Mood et al., 1974). To estimate the correlation dimension and to identify chaos large amounts of data may be necessary. While Sivakumar (2001) and Sivakumar et al. (2002b) suggest that the estimation of the correlation dimension and, consequently, the detection of deterministic chaos is reliable even for samples as short as 48 years of monthly data (576 values), on the other hand, Schertzer et al. (2002) indicated that even for deterministic systems the estimates of the correlation dimension is underestimated, i.e. the dimensionality of the dynamics is underestimated. For example, for simulated data ( $N=4096$ ) based on a stochastic cascade process (having a very large dimensional phase space), the correlation dimension gives a low finite value despite that the process has a large dimension. Such underestimation has been attributed to the limited sample size although Sivakumar et al. (2001a; 2002a) do not seem to agree with it.

Furthermore, Sivakumar (2001) analyzed the dynamics of 1-day, 2-day, 4-day, and 8-day rainfall using 25 years of data at the Leaf River basin in Mississippi. Based on the time delay embedding method for calculating the correlation dimension, he concluded that the referred rainfall data exhibited chaotic behavior at all time scales. For the same data, in a related paper focused on rainfall disaggregation, Sivakumar et al. (2001b) investigated the dynamics of rainfall transformations at resolutions of 6, 12, 24, 48, 96, and 192 hours and found finite low correlation dimensions for the distribution of weights that relate the rainfall data for the various time scales. Their results appeared to suggest the existence of chaos and pointed toward a method for disaggregating rainfall under a chaotic framework. The authors (Sivakumar, 2001 and Sivakumar et al., 2001b) recognized, however, the limitations of the method of correlation dimension associated with the effects of noise (e.g. systematic errors) and the large percentage of zeros as the data resolution becomes smaller, and suggested employing "additional techniques to substantiate the existence of chaos in rainfall transformation." In fact, this seems to have been done by Koutsoyiannis and Pachakis (1996), who employed both the delay embedding method and the generalized entropy of Cantorian dust for calculating the correlation dimension and found no evidence of low dimensional determinism for rainfall data at station Ortona Lock 2, Florida at temporal scales of 0.25 h, 1 h, 6 h, and 24 h. In

addition, the authors showed that a stochastic model was capable of preserving important statistical properties of the historical rainfall analyzed.

Furthermore, Sivakumar et al. (2001a) attempting to better understand the dynamics of the rainfall-runoff process analyzed both the precipitation and streamflow time series for the same basin (the Göta River basin in Sweden) from a chaotic perspective. Specifically, they analyzed each time series separately and the runoff coefficient series using the correlation dimension method obtaining values of the correlation dimension as 5.5, 6.4, and 7.8, respectively for runoff, precipitation, and runoff coefficient, thereby concluding that the existence of low dimensional chaos in the precipitation-runoff process for the Göta River basin can not be excluded and suggest further investigation on the subject.

Evidently, further research is needed to elucidate the intricacies of the dynamics of hydrological processes such as precipitation and streamflow. In this paper, we examine the issue of temporal aggregation and sampling in deterministic chaos by using the C-C method (Kim et al., 1999) for estimating the delay time, delay time window, and correlation dimension, and examining the attractors. First, we study the dynamics of daily streamflows for two streams in Florida. Then, the effects of aggregation and sampling are analyzed using time series derived from the Lorenz equations. Specifically, we investigate the effects that temporal aggregation (and/or sampling) has on the capability of identifying chaotic behavior. For convenience, the following section summarizes the basic equations and methods utilized.

## 2 Identifying chaotic behavior

The first step in the analysis of a chaotic time series is the embedding of the scalar time series into an  $m$ -dimensional space. This can be done using the method of delays introduced by Packard et al. (1980) and Takens (1981). A scalar time series,  $\{x_i\}$ ,  $i=1, 2, \dots, N$ , is embedded into  $m$ -dimensional space by constructing the vectors

$$\mathbf{x}_i = (x_i, x_{i+t}, \dots, x_{i+(m-1)t}) , \quad \mathbf{x}_i \in R^m , \quad (1)$$

where  $t$  is the index lag, and  $m$  is the embedding dimension, both of which must be chosen within appropriate ranges. If the sampling time is  $\tau_s$ , then the delay time is  $\tau_d=t\tau_s$ , and the delay time window is  $\tau_w=(m-1)\tau_d$ , which is the entire time spanned by the components of each vector. Since the components of the vector  $\mathbf{x}_i$  are the same quantities, then they have the same noise levels, which is one of the advantages of this method for embedding.

### 2.1 Correlation integral

After the attractor has been reconstructed using Eq. (1), quantitative properties of the chaotic system can be determined. For example, the correlation dimension is widely used in many fields for the quantitative characterization of strange attractors, and it is the only widely used method that

can be used with small data sets such as those analyzed in this paper (Grassberger and Procaccia, 1983). The correlation integral for the embedded time series is the following function:

$$C(m, N, r, t) = \frac{2}{M(M-1)} \sum_{1 \leq i < j \leq M} \Theta(r - \|\mathbf{x}_i - \mathbf{x}_j\|) , \quad (2)$$

$r > 0 ,$

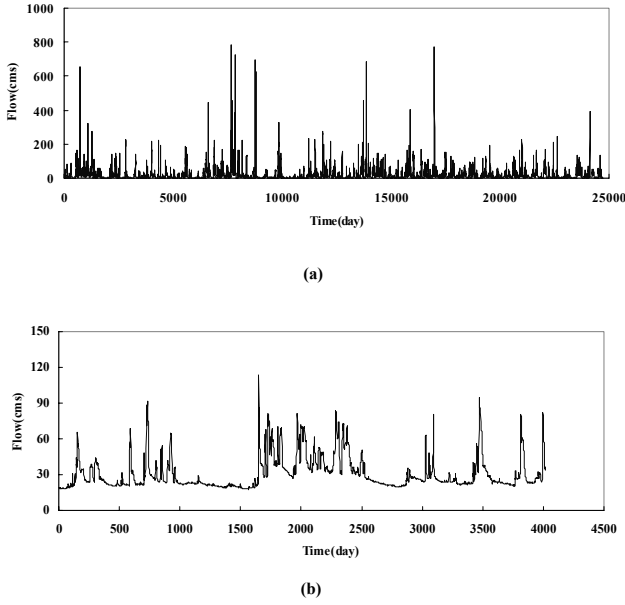
where  $\Theta(a)=0$ , if  $a < 0$  and  $\Theta(a)=1$  if  $a \geq 0$ ;  $N$  is the size of the data set;  $M=N-(m-1)t$  is the number of embedded points in  $m$ -dimensional space; and  $\|\dots\|$  denotes the sup-norm.  $C(m, N, r, t)$  measures the fraction of the pairs of points  $\mathbf{x}_i$ ,  $i=1, 2, \dots, M$ , whose sup-norm separation is no greater than  $r$ . If the limit of  $C(m, N, r, t)$  as  $N \rightarrow \infty$  exists for each  $r$ , the fraction of all state vector points that are within  $r$  of each other may be expressed as  $C(m, r, t) = \lim_{N \rightarrow \infty} C(m, N, r, t)$ , and the correlation dimension is defined as  $D_2(m, t) = \lim_{r \rightarrow 0} [\log C(m, r, t) / \log r]$ .

In practice,  $N$  remains finite, so  $r$  cannot go to zero. If  $r$  is too small, no pairs of points will have a separation less than  $r$ , so  $C(m, N, r, t)$  will be zero. As  $r$  increases,  $C(m, N, r, t)$  will jump in discontinuous steps as pairs of points get included. For larger values of  $r$ ,  $C(m, N, r, t)$  will grow smoothly, and a plot of  $\log C(m, N, r, t)$  vs.  $\log r$  will become linear with slope  $D_2(m, t)$ . If  $r$  becomes too large, so that most pairs of points have a separation less than  $r$ , then  $C(m, N, r, t)$  will begin to level off at its maximum value of 1. Thus, in practice, we must look for the linear region in the plot of  $\log C(m, N, r, t)$  vs.  $\log r$  and compute its slope  $D_2(m, t)$ . If the data set is too small, there may be no linear region, and it may not be possible to determine the correlation dimension  $D_2(m, t)$ .

If the sampling time  $\tau_s$  is considerably less than the appropriately chosen delay time  $\tau_d$ , then successive points  $\mathbf{x}_i$  and  $\mathbf{x}_{i+1}$  of the embedded time series will generally be close, and this will lead to artificial correlations in Eq. (2) (Grassberger, 1990). While it is not possible to make an unambiguous identification of such artificially correlated points (Rosenstein et al., 1994; Grassberger, 1990), a practical way to minimize this effect is to remove from Eq. (2) the contributions of all pairs of points  $\mathbf{x}_i$  and  $\mathbf{x}_j$  where  $|j-i| < t$  and  $t$  is the index lag (Grassberger, 1990).

### 2.2 The delay parameters

Many researchers use a fixed delay time  $\tau_d$  as the embedding dimension  $m$  is increased. Some suggest obtaining  $\tau_d$  from the autocorrelation function (ACF) or the mutual information (MI). An alternative is fixing the delay time window  $\tau_w$ , rather than the delay time  $\tau_d$ , but the estimation of  $\tau_w$  is not well developed. Martinerie et al. (1992) examined the estimation of the delay time window and compared it with the estimated delay times using the ACF and the MI. They concluded that  $\tau_w$  could not be estimated using either of these two methods. Basically,  $\tau_w$  is the optimal time for independence of the data, but these methods estimate the first locally



**Fig. 1.** Daily streamflow for the (a) St. Marys River near Macclenny, Florida, USA, for the period 1927–1993, and (b) Ocklawaha River near Conner, Florida, USA for the period 1978–1988.

optimal time, which is  $\tau_d$ . From this distinction between  $\tau_d$  and  $\tau_w$ , a technique called the  $C-C$  method was developed to estimate both  $\tau_d$  and  $\tau_w$  (Kim et al., 1999). They showed that, for small data sets, as the embedding dimension  $m$  is increased, the correlation dimension  $D_2$  converges more rapidly if  $\tau_w$  is held fixed rather than  $\tau_d$  (Kim et al., 1998). For ease of reference, this method is summarized below.

### 2.3 The $C-C$ method

Brock et al. (1991, 1996) studied the BDS statistic (named for Brock, Dechert, and Scheinkman), which is based on the correlation integral, to test the null hypothesis that the data are independently and identically distributed (*iid*). This test has been particularly useful for chaotic systems and nonlinear stochastic systems. An application of the BDS statistic can be found in Kim et al. (2003). Under the *iid* hypothesis, the BDS statistic for  $m > 1$  is defined as

$$\text{BDS}(m, M, r, t) = \frac{\sqrt{M}}{\sigma(m, M, r, t)} [C(m, M, r, t) - C^m(1, M, r, t)], \quad (3)$$

where

$$C(m, M, r, t) = \frac{2}{M(M-1)} \sum_{1 \leq i < j \leq M} \Theta(r - \|\mathbf{x}_i - \mathbf{x}_j\|) \quad (4)$$

$$K(m, M, r, t) = \frac{6}{M(M-1)(M-2)} \sum_{1 \leq i < j < k \leq M} \Theta(r - \|\mathbf{x}_i - \mathbf{x}_j\|) \Theta(r - \|\mathbf{x}_j - \mathbf{x}_k\|) \quad (5)$$

$$\sigma^2(m, M, r, t) = 4 \left\{ m(m-1)C^{2(m-1)}(K - C^2) + K^m - C^{2m} + 2 \sum_{i=1}^{m-1} \left[ C^{2i} (K^{m-i} - C^{2(m-i)}) - mC^{2(m-i)} (K - C^2) \right] \right\} \quad (6)$$

It may be shown that the BDS statistic converges as  $M \rightarrow \infty$  to a normal distribution  $N(0,1)$ .

In addition, using a concept suggested by Brock et al. (1991), the statistic  $S(m, N, r, t) = C(m, N, r, t) - C^m(1, N, r, t)$  may be interpreted as the serial correlation of a nonlinear time series. Therefore, it can be regarded as a dimensionless measure of nonlinear dependence. Thus, for fixed  $m$ ,  $N$ , and  $r$ , the plot of  $S(m, N, r, t)$  versus  $t$  is a nonlinear analog of the plot of the autocorrelation function versus  $t$ . Brock et al. (1991) suggested that  $m$  should be between 2 and 5, and  $r$  should be between  $\hat{\sigma}/2$  and  $2\hat{\sigma}$ , where  $\hat{\sigma}$  is the standard deviation of the data set. In addition, the asymptotic distributions were well approximated by finite time samples for  $N \geq 500$ . Thus, four values of  $r$  are selected in the range  $\hat{\sigma}/2 \leq r \leq 2\hat{\sigma}$  as representative values:  $r_1 = (0.5)\hat{\sigma}$ ,  $r_2 = (1.0)\hat{\sigma}$ ,  $r_3 = (1.5)\hat{\sigma}$ , and  $r_4 = (2.0)\hat{\sigma}$ .

Kim et al. (1999) considered the following quantities:

$$S(m, r, t) = \frac{1}{t} \sum_{s=1}^t [C_s(m, r, t) - C_s^m(1, r, t)], \quad m=2, 3, \dots \quad (7)$$

$$\Delta S(m, t) = \max\{S(m, r_j, t)\} - \min\{S(m, r_j, t)\} \quad (8)$$

and their averages

$$\bar{S}(t) = \frac{1}{16} \sum_{m=2}^5 \sum_{j=1}^4 S(m, r_j, t), \quad (9)$$

$$\Delta \bar{S}(t) = \frac{1}{4} \sum_{m=2}^5 \Delta S(m, t). \quad (10)$$

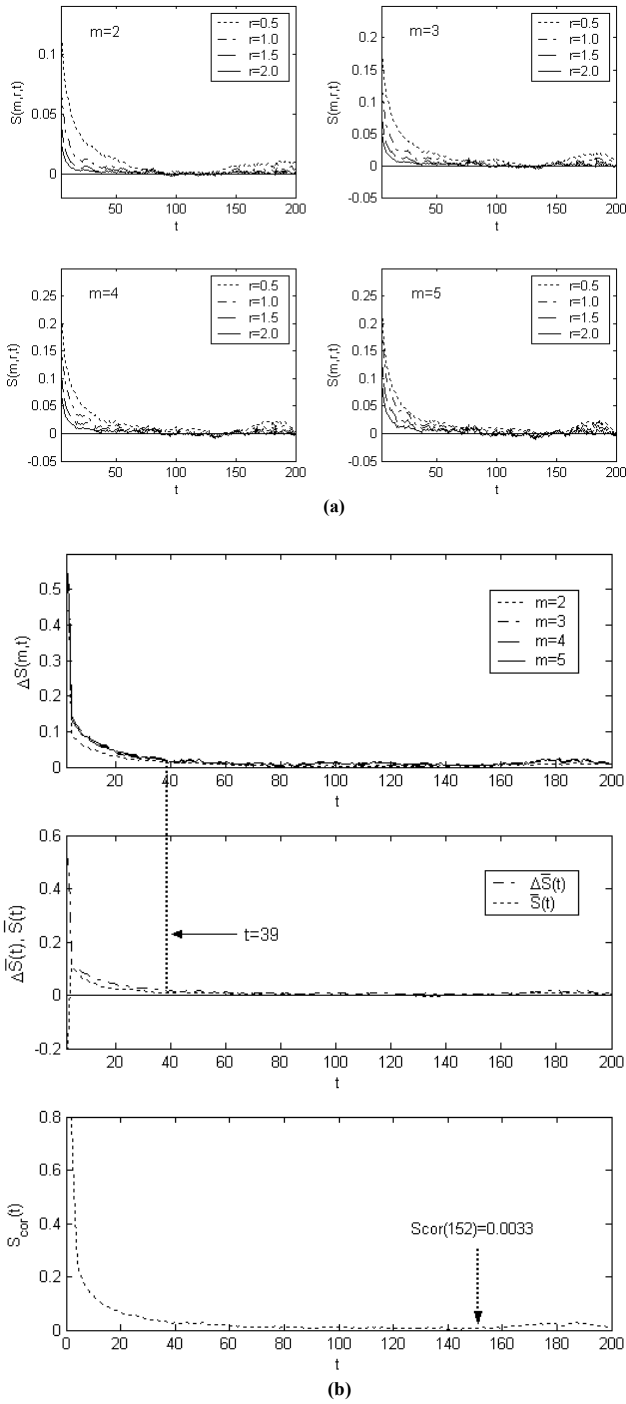
Then the first zero crossing of  $\bar{S}(t)$  or the first local minimum of  $\Delta \bar{S}(t)$  is located to find the first locally optimal time  $t$  for independence of the data, and this gives the delay time  $\tau_d = t \tau_s$ . The optimal time is the index lag  $t'$  for which  $\bar{S}(t)$  and  $\Delta \bar{S}(t)$  are both closest to zero. If one gives equal weight to these two quantities, then one may simply look for the minimum of the quantity

$$S_{cor}(t) = \Delta \bar{S}(t) + |\bar{S}(t)|, \quad (11)$$

and this optimal time gives the delay time window  $\tau_w = t' \tau_s$ .

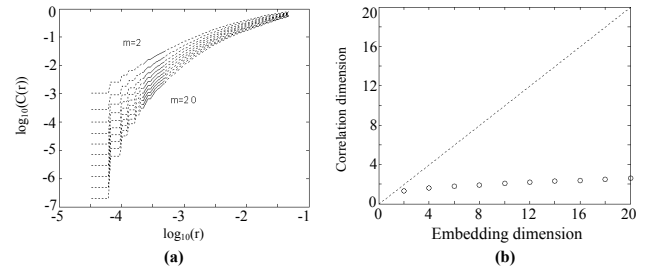
### 3 Investigating the dynamics of daily streamflow data

We search for evidence of low dimensional deterministic chaos for the daily streamflows of the St. Marys River near Macclenny and the Ocklawaha River near Conner, both in Florida, USA. The records consist of 67 years (1927–1993) of daily data (24 472 values) for the St. Marys River and 11

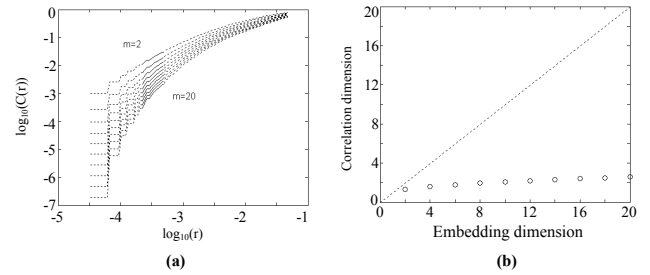


**Fig. 2.** Estimations of  $\tau_d$  and  $\tau_w$  for the St. Marys River, Florida, USA: (a)  $S(m, r, t)$  and (b)  $\Delta S(m, t)$ ,  $\Delta \bar{S}(t)$ ,  $\bar{S}(t)$ , and  $S_{cor}(t)$ .

years (1978–1988) of daily data (4018 values) for the Ocklawaha River. The time series plots for these streamflow data are shown in Fig. 1. Clearly, the two time series display quite different time patterns. The time series for the Ocklawaha River in Fig. 1b shows a significant groundwater contribution with clear streamflow recessions, a characteristic that is not evident in the daily streamflows of the St. Marys River in Fig. 1a.



**Fig. 3.** Estimation of the correlation dimension using the delay time  $\tau_d$  for the daily flows of the St. Marys River, Florida, USA: (a) correlation integral and its linear regions (shown as full lines) and (b) slopes (correlation dimension) vs. embedding dimension.



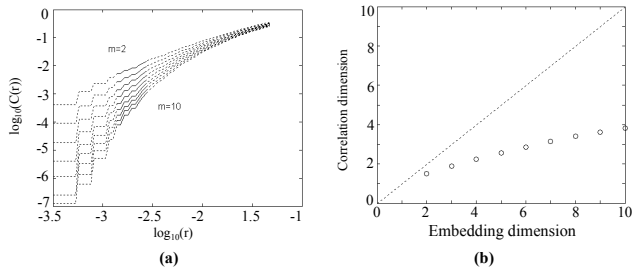
**Fig. 4.** Estimation of the correlation dimension using the delay time window  $\tau_w$  for the daily flows of the St. Marys River, Florida, USA: (a) correlation integral and its linear regions (shown as full lines) and (b) slopes (correlation dimension) vs. embedding dimension.

Figure 2 shows the statistics  $S(m, r, t)$ ,  $\Delta S(m, t)$ ,  $\Delta \bar{S}(t)$ ,  $\bar{S}(t)$ , and  $S_{cor}(t)$  for the daily streamflows of the St. Marys River (similar statistics calculated for the Ocklawaha River are not shown). The time delay  $\tau_d=39$   $\tau_s=39$  days corresponds to the first local minimum of  $\Delta \bar{S}(t)$ , as indicated by the arrow in the middle of Fig. 2b. For the Ocklawaha River,  $\tau_d=33$   $\tau_s=33$  days. Also, as shown in Fig. 2b the minimum of  $S_{cor}$  occurs at  $t=152$ , which gives  $\tau_w=152$  days for the St. Marys River. Similarly,  $\tau_w=187$  days for the Ocklawaha River.

The correlation integral analysis is now conducted on the two daily streamflow series using the values  $\tau_d$  and  $\tau_w$ . Figure 3 shows the plots of  $\log[C(r)]$  versus  $\log(r)$  (base 10 logs) for the reconstructed attractors for the St. Marys River for embedding dimensions  $m=2, 4, \dots, 20$  using  $\tau_d=39$  days. The slopes are determined by a least-squares fit to the linear regions discussed in Sect. 2.1, which are indicated by the solid lines (Barnett, 1993):

$$D_2 = \frac{\sum_{i=2}^N (x_i - x_{i-1})(y_i - y_{i-1})}{\sum_{i=2}^N (x_i - x_{i-1})^2}, \quad (12)$$

where  $x = \log(r)$  and  $y = \log[C(r)]$ . The slopes shown in Fig. 3b slowly increases up to  $m=20$  and the lack of saturation makes the evidence for chaotic behavior doubtful. The analysis is repeated using the delay time window  $\tau_w=152$  days, and the results are shown in Fig. 4. Since the index lag is determined by  $(m-1) t=152$ , it is necessary to round



**Fig. 5.** Estimation of the correlation dimension using the delay time  $\tau_d$  for the daily flows of the Ocklawaha River, Florida, USA: (a) correlation integral and its linear regions (shown as full lines) and (b) slopes (correlation dimension) vs embedding dimension.

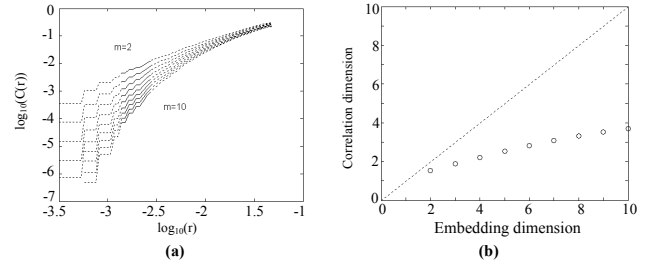
off  $t$  to the nearest integer. The actual values used for  $(m, t)$  are as follows: (2, 152), (4, 51), (6, 30), (8, 22), (10, 15), (12, 14), (14, 12), (16, 10), (18, 9), and (20, 8). Again, it is not clear the value of the correlation dimension that can confirm a chaotic behavior.

A similar analysis is performed for the Ocklawaha River using both  $\tau_d$  and  $\tau_w$ , and the results are shown in Figs. 5 and 6, respectively. In this case, there is not a clear linear region between the steps at small  $r$  and the saturation at large. What we have instead is a linear region disrupted by oscillations that are not fully damped. However, the procedure for dealing with these oscillations is well-established: average the two slopes obtained using the maxima and minima of these oscillations (or, equivalently, obtain a slope using the centers of these oscillations). The correlation dimension fails to saturate in this case, indicating a lack of low-dimensional nonlinear determinism.

The results obtained regarding the dynamics of daily streamflows for the two streams in Florida raises some interesting issues. Although the correlation integral analysis does not show evidence of deterministic chaos in either case, the daily flows for the Ocklawaha River, which has a substantial groundwater contribution, depart significantly from any evidence of low dimensional chaos, i.e. there is a clear lack of saturation in the correlation dimension versus embedding dimension plots. On the other hand, for the daily flows of the St. Marys River, which has weak or less significant groundwater contribution, the correlation dimension vs. embedding dimension plots approach a saturation condition, albeit without reaching it. The question is why the two cases are so different.

#### 4 Dynamics of hydrological processes

Our interpretation of the results described above and the conflicting results found in the literature regarding the inherent dynamics of hydrologic processes, i.e. whether hydrologic data arise from low dimensional deterministic systems or from stochastic systems (e.g. Koutsoyiannis and Pachakis, 1996; Sivakumar, 2001), lead to the following concepts. First of all, let us consider basic hydrologic input data, such as



**Fig. 6.** Estimation of the correlation dimension using the delay time window  $\tau_w$  for the daily flows of the Ocklawaha River, Florida, USA: (a) correlation integral and its linear regions (shown as full lines) and (b) slopes (correlation dimension) vs embedding dimension.

precipitation. Precipitation is measured in different forms particularly using recording and non-recording rain gages (the former provides continuous records of precipitation data, and the latter generally provides accumulated daily records). Clearly one can obtain high frequency and hourly, daily, weekly, etc. records by integration or aggregation from the continuous records. In the same vein, streamflow records can be measured using recording streamgages (limnigraphs) that provide continuous records and nonrecording gages (e.g. staff gages or limnimeters) that record river levels and after transformation through rating curves, provide flow data at a given time of the day, i.e. sampled data (such data are generally assumed to be valid for the entire day). Likewise, other types of hydrologic data such as water quality parameters at streams, lakes, and groundwater, are generally obtained by sampling at regular or irregular time intervals. The point is that a significant amount of hydrological data becomes available in either aggregated form or sampled form. Then, assuming that continuous time hydrologic data, e.g. precipitation data, are chaotic in nature, one could ask the question, what is the effect of aggregation and sampling? More specifically, given that continuous time precipitation is chaotic, are daily and weekly precipitation series chaotic?

In addition, the case of streamflow data deserves some further analysis. Assuming that the input (precipitation) to a watershed system (or river basin) is low dimensional chaotic, it becomes transformed in various ways as it travels down to the basin outlet. First of all, the continuous precipitation field is subjected to a number of losses, such as interception, depression storage, and infiltration. Secondly, the rainfall excess travels throughout the basin as overland flow, thereby being affected by time delay, attenuation, and further losses. At the same time, the direct and delayed infiltration increase the soil moisture, part of which is lost by evapotranspiration, another part may come back to the land surface as interflow, and some other part may further percolate through the unsaturated and saturated components of the basin. Eventually, the groundwater storage feeds the stream and is added to the surface runoff to form the total streamflow. Thus, the end result, i.e. streamflow, is a process that has been transformed by a complex sequence of linear and nonlinear mechanisms.

**Table 1.** The delay times and the delay time windows for the aggregated and sampled Lorenz series.

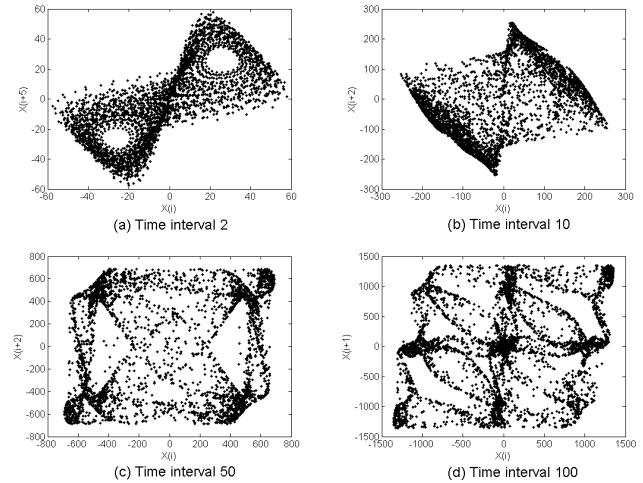
Time interval	$\tau_d$		$\tau_w$	
	aggregated	sampled	aggregated	sampled
2	5	5	134	134
10	2	2	91	91
50	2	2	21	26
100	1	2	5	24

In fact, these types of transformations are significantly more complex than the simple transformations alluded to in the previous paragraph, where precipitation (or some other process) is simply aggregated or sampled. Then the question is, given that precipitation is chaotic, is the resulting streamflows also chaotic?

We pose the following two hypotheses: (1) Assuming that a continuous hydrologic process is low dimensional chaotic (e.g. precipitation), as it is transformed in time and space by aggregation or sampling, the chaotic behavior of the resulting process (e.g. weekly precipitation) becomes more difficult to identify and as the aggregation or sampling scale increases, eventually, the referred hydrologic process may no longer exhibit chaotic features. (2) Assuming that the continuous time precipitation input to a basin is low dimensional chaotic, the streamflows at the outlet of the basin (such as daily streamflows) may depart from chaotic behavior due to linear and nonlinear filtering and interactions occurring throughout the basin. Proving these two hypotheses may not be simple, because it requires that the underlying continuous hydrological process is in fact chaotic. Nevertheless, one can make some experiments that may shed some light onto the ongoing debate and controversy. In the following section, we address the hypothesis (1) using a limited experiment and illustrate how effective the chaos detection algorithm utilized in this paper is for identifying low dimensional chaos in a series that has been aggregated or sampled. The hypothesis (2) is related to the effect of the complex behavior of a hydrologic system that may result in system output that does not exhibit chaotic behavior. Experiments related to hypothesis (2) are beyond the scope of the present manuscript and will be reported elsewhere.

### 5 Aggregation and sampling of chaotic systems

We will examine the effect of temporal aggregation and the effect of sampling on a system that is chaotic. They are simple transformations yet realistic from the hydrologic viewpoint as discussed in the previous section. Again, if the underlying input series is chaotic, temporal aggregation or sampling may hinder the identification of nonlinear determinism, as we will show in this section.



**Fig. 7.** Attractors of the aggregated time series from the Lorenz system for time intervals ranging from 2 to 100.

As an example, we examine the effects of aggregation and sampling on a time series of the variable  $x$  obtained from the Lorenz equations (Abarbanel et al., 1993):

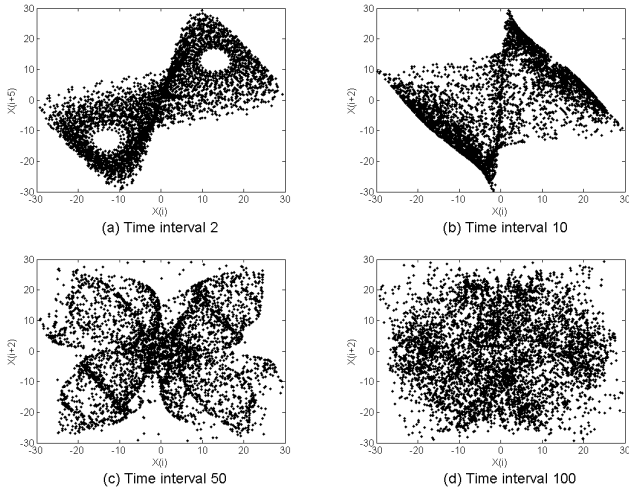
$$\begin{aligned} \dot{x} &= -a(x - y), \\ \dot{y} &= -xz + cx - y, \\ \dot{z} &= xy - bz, \end{aligned} \tag{13}$$

using the parameters  $a=16.0$ ,  $b=4.0$ , and  $c=45.92$ . Each time series used in the analyses below will consist of 15 000 values. The Lorenz equations are solved numerically using a time step  $\tau_s=0.01$ . We then apply the referred  $C-C$  method to this time series and obtain the delay time  $\tau_d=10 \tau_s=0.1$  and the delay time window  $\tau_w=100 \tau_s=1.0$ . An analysis of the correlation integral yields clear evidence of low dimensional chaos (as it should) with the dimension  $D_2=2.05$ .

The aggregated series is obtained using

$$\bar{x}_i = \frac{1}{n}(x_{1+(i-1)n} + x_{2+(i-1)n} + \dots + x_{n+(i-1)n}), \quad i=1, 2, \dots, \tag{14}$$

where  $n=2, 10, 50$ , and  $100$ . We also perform sampling from the series by keeping only the values at every  $n$ th time step, with  $n=2, 10, 50$ , and  $100$  (again in all cases our total sample series after aggregation or sampling result in 15 000 values). For each of the four aggregated and four sampled time series, we use the  $C-C$  method to compute the delay time  $\tau_d$ , and the results are shown in Table 1. We then construct embeddings of the attractors using these delay times, and the results for the aggregated and sampled series are shown in Figs. 7 and 8, respectively. The shapes of the attractors are modified and lose their structure as  $n$  increases. For instance, for  $n=100$ , the attractors are not recognizable for both the aggregated and sampled series. We also compute the correlation dimensions, and these results are shown in Figs. 9 and 10, respectively. The convergence of the correlation dimension degrades slowly as  $n$  increases, and, this convergence is lost



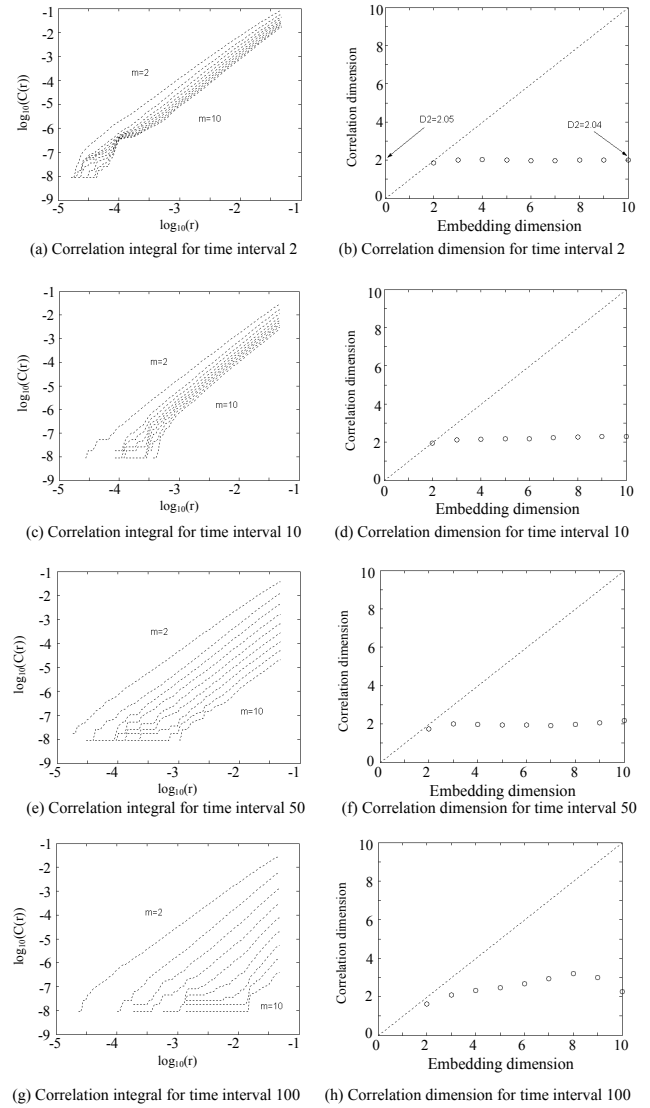
**Fig. 8.** Attractors of the sampled time series from the Lorenz system for time steps ranging from 2 to 100.

for  $n=100$ . Thus, we observe that both aggregation and sampling can hinder the evidence of nonlinear determinism from a chaotic time series.

For the larger values of  $n$ , there are fewer data points available. Also, as the embedding dimension  $m$  increases, these data points become more sparse (reducing  $m$  is equivalent to projecting onto fewer dimensions, which increases the density of the embedded points). Thus, for large values of both  $n$  and  $m$ , it is not possible to find a linear region in the plot of  $\log[C(r)]$  vs.  $\log(r)$ . This is why the plots for  $n=100$  bend at  $m=8$  or 9.

Furthermore, note that, as the time interval  $n$  is increased, it becomes less evident that the time series is chaotic; conversely the transformed time series appears more like a stochastic series. As a result, it is more difficult to find the best optimal point of  $S_{cor}(t)$  for  $\tau_w$ . Thus, we have not used the delay time window in the estimation of the correlation dimension for the aggregated and sampled series.

In the above analysis, we investigated the effects of aggregation and sampling on the time series by keeping only the aggregated (or averaged) values or sampling the series at every  $n$ th time step. This is equivalent to making aggregated and sampled measurements with a regular time interval  $\tau_m = n\tau_s$ . As this time interval  $\tau_m$  increases, the successive aggregated and sampled measurements will eventually become irrelevant, and any nonlinear determinism (exhibited in the original system) will be lost. When this happens, the data will appear to be stochastic, rather than chaotic. Since the delay time  $\tau_d$  is a basic correlation time, then irrelevance should begin to become apparent when  $\tau_m \sim \tau_d$ . However,  $\tau_w$  is the maximum time for correlations, so stochastic like behavior should not occur until  $\tau_m \sim \tau_w$ . For the Lorenz system, the conditions  $\tau_m \sim \tau_d$  and  $\tau_m \sim \tau_w$  become  $n \sim 10$  and  $n \sim 100$ , respectively. In Figs. 7 and 8, we do indeed see that the reconstructed attractors begin to lose their structure when  $n \sim 10$ , and that this structure is not recognizable when  $n \sim 100$ .



**Fig. 9.** Correlation integral and slopes (correlation dimension) vs. embedding dimension for the aggregated time series from the Lorenz system for aggregation time scales ranging from 2 to 100.

In order to explain this more concretely, consider a deterministic nonlinear dynamical system governed by an iterative map  $x_s = f(x_{s-1})$ . Then let us consider the aggregated quantities  $\bar{x}_s = x_s + x_{s+1} + \dots + x_{s+n-1}$ . Since the map is nonlinear, we do not have the relation  $\bar{x}_s = f(\bar{x}_{s-1})$  or any simple deterministic relation between  $\bar{x}_s$  and  $\bar{x}_{s-1}$ . Of course, if  $n$  is small, then  $\bar{x}_s$  will not differ much from  $x_s$ , so we will have an approximate relation  $\bar{x}_s \cong f(\bar{x}_{s-1})$ . However, as  $n$  increases, this approximate relation will degrade, and the characteristic time for this degradation is the correlation time  $\tau_d$  for the values of  $x_s$  to change substantially.

The aggregation and sampling experiment described in this section is by no means complete and does not answer all relevant questions. For example, the experiment has been limited to a sample of 15 000 values for all cases after aggregation and sampling and to the referred identifica-



tion method. Our analysis illustrated that aggregation and sampling could hinder the identification of chaos using the algorithm presented in this paper. The results are relevant from the practical standpoint but a more complete experiment and analysis should include the effect of sample size and other identification methods, which were not considered in our study.

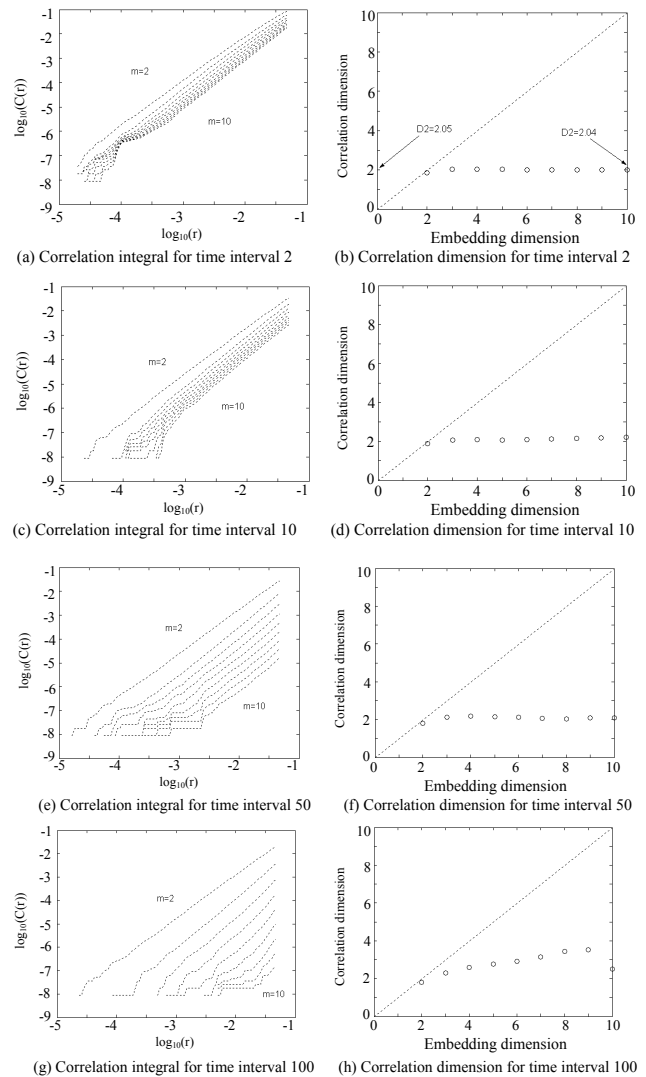
### 6 Final remarks and conclusions

Several stochastic and non-stochastic approaches have been suggested in the literature for representing the dynamics of precipitation and streamflow processes on various time scales. Among the non-stochastic approaches, nonlinear dynamics and chaotic systems have been investigated and applied to various types of hydrologic series by many water scientists and hydrologists. Our review of the literature produced conflicting results regarding the inherent dynamics of hydrological processes such as precipitation and streamflows, i.e. whether they exhibit low dimensional deterministic (chaotic) behavior or whether they behave as stochastic systems.

We suggested two hypotheses that may help clarify the debate. Assuming that the underlying continuous time hydrologic process (e.g. precipitation) is low dimensional chaotic: (1) As the (low dimensional chaotic) system is subjected to transformations such as aggregation and sampling, its behavior becomes more difficult to identify and as the scale of aggregation and sampling increase, the system may no longer exhibit features of low dimensional determinism; instead, it will display stochastic features. (2) As the chaotic precipitation input to a basin is subjected to complex linear and nonlinear transformations involving many effects from losses, travel time, attenuation, ponding, and surface and subsurface storages, the resulting streamflow at the outlet of the basin may depart from chaotic behavior. The significance of such departure from a chaotic system will depend on the extent and impact of the referred transformations.

Our investigation using simple aggregation and sampling of a chaotic system (the Lorenz system) supports that the first hypothesis can be true, i.e. aggregation and sampling hinders the identification of a chaotic behavior. This highlights an interesting phenomenon where the application of a simple linear filter may cause a nonlinear deterministic system to be not recognizable and instead exhibit features of a stochastic system. It may be worth performing similar experiments using other types of low dimensional chaotic systems and other tests and methods for detecting chaotic behavior to further understand this phenomenon.

We also investigated whether the time series of daily streamflows at two river sites in Florida show evidence of being characterized by a low dimensional chaotic system. The time series pattern of one of the series shows the effect of strong basin storage, especially groundwater contribution, while the other one appears to have minimal effects of surface and subsurface storages, and the streamflow appears to



**Fig. 10.** Correlation integral and slopes (correlation dimension) vs. embedding dimension for the sampled time series from the Lorenz system for sampling time scales ranging from 2 to 100.

consist mostly of routed excess precipitation. It turns out that the daily streamflow series with significant effects from surface and groundwater storages departs significantly from any evidence of a chaotic system, while the one with lesser contribution from basin storage appears to be closer to being low dimensional chaotic. One may argue that since groundwater storage is a form of accumulation (aggregation) from past occurrences of precipitation, it may explain why the streamflow series with significant groundwater contribution (Ocklawaha River) departs significantly from any evidence of low dimensional chaos. While we did not perform any scientific experiment to test whether the referred second hypothesis is true, the analysis made using the daily streamflows of the two rivers in Florida suggests that the complex transformations that precipitation is subjected in a basin may further affect the detection of chaotic behavior.

*Acknowledgements.* Part of this work has been undertaken in connection to the Specific Cooperative Agreement (SCA) between the USDA Agriculture Research Service and Colorado State University entitled “Quantifying Space-Time Variability in Agricultural Landscapes”. In addition, we thank Demetris Koutsoyiannis and an anonymous reviewer for their meticulous and insightful review and comments that improved the final paper. Furthermore, the first author thanks the Editor of the special issue for his patience and support.

Edited by: B. Sivakumar

Reviewed by: D. Koutsoyiannis and another referee

## References

- Abarbanel, H. D. I., Brown, R., Sidorowich, J. J., and Tsimring, L. S.: Analysis of observed chaotic data in physical systems, *Rev. Mod. Phys.*, 65, 1331–1392, 1993.
- Barnett, K. D.: On the estimation of the correlation dimension and its application to radar reflector discrimination, NASA Rep. 4564, DOT/FAA/RD-93/41, 1993.
- Berndtsson, R., Jinno, K., Kawamura, A., Olsson, J., and Xu, S.: Dynamical systems theory applied to long-term temperature and precipitation time series, *Trends Hydrol.* 1, 291–297, 1994.
- Brock, W. A., Dechert, W. D., Scheinkman, J. A., and LeBaron, B.: A test for independence based on the correlation dimension, *Econ. Rev.*, 15, 197–235, 1996.
- Brock, W. A., Hsieh, D. A., and LeBaron, B.: *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence*, The MIT Press, 1991.
- Cowpertwait, P. S. P. and O’Connell, P. E.: A Neyman-Scott shot noise model for the generation of daily streamflow time series, in: *Advances in Theoretical Hydrology*, edited by: O’Kane, J. P., Elsevier, 75–94, 1992.
- Elshorbagy, A., Simonovic, S. P., and Panu, U. S.: Noise reduction in chaotic hydrologic time series: facts and doubts, *J. Hydrol.*, 256, 147–165, 2002.
- Fraedrich, K.: Estimating weather and climate predictability on attractors, *J. Atmos. Sci.*, 44, 4, 722–728, 1987.
- Ghilardi, P. and Rosso, R.: Comment on Chaos in rainfall by Rodriguez-Iturbe, I. et al., *Water Resour. Res.*, 26, 1, 1837–1839, 1990.
- Gilmore, C. G.: A New Test for Chaos, *J. Economic Behavior and Organization*, 22, 209–237, 1993.
- Grassberger, P.: An optimized box-assisted algorithm for fractal dimensions, *Phys. Lett. A*, 148, 1–2, 63–68, 1990.
- Grassberger, P. and Procaccia, I.: Measuring the strangeness of strange attractors, *Physica D*, 7, 153–180, 1983.
- Islam, S., Bras, R., and Rodriguez-Iturbe, I.: A possible explanation for low correlation dimension estimates for the atmosphere, *J. Appl. Meteor.*, 32, 203–208, 1993.
- Jayawardena, A. W. and Lai, F. Z.: Analysis and prediction of chaos in rainfall and stream flow time series, *J. Hydrol.*, 153, 23–52, 1994.
- Jayawardena, A. W. and Gurung, A. B.: Noise reduction and prediction of hydrometeorological time series: dynamical systems approach vs. stochastic approach, *J. Hydrol.*, 228, 242–264, 2000.
- Jayawardena, A. W., Li, W. K., and Xu, P.: Neighborhood selection for local modeling and prediction of hydrological time series, *J. Hydrol.*, 258, 40–57, 2002.
- Jeong, G. D. and Rao, A. R.: Chaos characteristics of tree ring series, *J. Hydrol.*, 182, 239–257, 1996.
- Kantz, H. and Schreiber, T.: *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge, 1997.
- Kawamura, A., McKerchar, A. L., Spigel, A. L., and Jinno, K.: Chaotic characteristics of the Southern Oscillation Index time series, *J. Hydrol.*, 204, 168–181, 1998.
- Kim, H. S., Eykholt, R., and Salas, J. D.: Nonlinear dynamics, delay times, and embedding windows, *Physica D*, 127, 48–60, 1999.
- Kim, H. S., Eykholt, R., and Salas, J. D.: Delay time window and plateau onset of the correlation dimension for small data sets, *Phys. Rev. E*, 58, 5, 5676–5682, 1998.
- Kim, H. S., Yoon, Y. N., Kim, J. H., and Kim, J. H.: Searching for strange attractor in wastewater flow, *Stoch. Env. Res. Risk Assess.*, 15, 5, 399–413, 2001.
- Kim, H. S., Kang, D. S., and Kim, J. H.: The BDS statistic and residual test, *Stoch. Env. Res. Risk Assess.*, 17, 1–2, 104–115, 2003.
- Koutsoyiannis, D. and Pachakis, D.: Deterministic chaos versus stochasticity in analysis and modeling of point rainfall series, *J. Geophys. Res.*, 101, D21, 26 441–26 451, 1996.
- Lall, U., Sangoyomi, T. B., and Abarbanel, H. D. I.: Nonlinear dynamics of the Great Salt Lake: nonparametric short term forecasting, *Water Resour. Res.*, 32, 4, 975–985, 1996.
- Liu, Q., Islam, S., Rodriguez-Iturbe, I., and Le, Y.: Phase-space analysis of daily streamflow: characterization and prediction, *Adv. Water Resour.*, 21, 463–475, 1998.
- Martinerie, J. M., Albano, A. M., Mees, A. I., and Rapp, P. E.: Mutual information, strange attractors, and the optimal estimation of dimension, *Phys. Rev. A*, 45, 7058–7064, 1992.
- Mood, A. M., Graybill, F. A., and Boes, D. C.: *Introduction to the Theory of Statistics*, McGraw Hill, New York, 564 pages, 1974.
- Packard, N. H., Crutchfield, J. P., Farmer, J. D., and Shaw, R. S.: “Geometry from a time series, *Phys. Rev. Lett.*, 45, 9, 712–716, 1980.
- Pasternack, G. B.: Does the river run wild? Assessing chaos in hydrological systems, *Adv. Water Resour.*, 23, 253–260, 1999.
- Porporato, A. and Ridolfi, L.: Nonlinear analysis of river flow time sequences, *Water Resour. Res.*, 33, 6, 1353–1367, 1997.
- Puente, C. E. and Obregon, N.: A deterministic geometric representation of temporal rainfall: results for a storm in Boston, *Water Resour. Res.*, 32, 9, 2825–2839, 1996.
- Rodriguez-Iturbe, I., Power, B. F. D., Sharifi, M. B., and Georgakakos, K. P.: Chaos in rainfall, *Water Resour. Res.*, 25, 7, 1667–1675, 1989.
- Rodriguez-Iturbe, I.: Exploring complexity in the structure of rainfall, *Adv. Water Resour.*, 14, 4, 162–167, 1991.
- Rosenstein, M. T., Collins, J. J., and De Luca, C. J.: Reconstruction expansion as a geometry-based framework for choosing proper delay times, *Physica D*, 73, 82–98, 1994.
- Salas, J. D.: Analysis and modelling of hydrologic time series, in: *Handbook of Hydrology*, edited by: Maidment, D. R., McGraw Hill Book Co., New York, 1993.
- Sangoyomi, T. B., Lall, U., and Abarbanel, H. D. I.: Nonlinear dynamics of the Great Salt Lake: dimension estimation, *Water Resour. Res.*, 32, 1, 149–159, 1996.
- Schertzer, D., Tchigurinskaia, I., Lovejoy, S., Hubert, P., Bendjoudi, H., and Larcheveque, M.: Which chaos in the rainfall-runoff process?, Discussion of “Evidence of chaos in the rainfall-runoff process” by Sivakumar B., et al., *Hydrol. Sci. J.*, 47, 1, 139–148, 2002.

- Schreiber, T. and Grassberger, P.: A simple noise-reduction method for real data, *Phys. Lett. A*, 160, 5, 411–418, 1991.
- Schreiber, T.: Extremely simple nonlinear noise-reduction method, *Phys. Rev. E*, 47, 4, 2401–2404, 1993.
- Sharifi, M. B., Georgakakos, K. P., and Rodriguez-Iturbe, I.: Evidence of deterministic chaos in the pulse of storm rainfall, *J. Atmos. Sci.*, 47, 7, 888–893, 1990.
- Sivakumar, B.: Chaos theory in hydrology: important issues and interpretations, *J. Hydrol.*, 227, 1–20, 2000.
- Sivakumar, B.: Rainfall dynamics at different temporal scales: A chaotic perspective, *Hydrol. Earth Sys. Sci.*, 5, 645–651, 2001, **SRef-ID: 1607-7938/hess/2001-5-645**.
- Sivakumar, B., Liang, S. Y., Liaw, C. Y., and Phoon, K. K.: Singapore rainfall behavior: chaotic?, *J. Hydro. Eng.*, 4, 1, 38–48, 1999a.
- Sivakumar, B., Phoon, K. K., Liang, S. Y., and Liaw, C. Y.: A systematic approach to noise reduction in chaotic hydrological time series, *J. Hydrol.*, 219, 103–135, 1999b.
- Sivakumar, B., Berndtsson, R., Olsson, J., and Jinno, K.: Evidence of chaos in the rainfall-runoff process, *Hydrol. Sci. J.*, 46, 1, 131–145, 2001a.
- Sivakumar, B., Sorooshian, S., Gupta, H. V., and Gao, X.: A chaotic approach to rainfall disaggregation, *Water Resour. Res.*, 37, 1, 61–72, 2001b.
- Sivakumar, B., Berndtsson, R., Olsson, J., and Jinno, K.: Reply to which chaos in the rainfall-runoff process, *Hydrol. Sci. J.*, 47, 1, 149–158, 2002a.
- Sivakumar, B., Persson, M., Berndtsson, R., and Bertacchi-Uvo, C.: Is correlation dimension a reliable indicator of low-dimensional chaos in short hydrologic time series?, *Water Resour. Res.*, 38, 2, doi:10.1029/2001WR000333, 2002b.
- Takens, F.: Detecting strange attractors in turbulence, in: *Dynamical Systems and Turbulence*, edited by: Rand, D. A. and Young, L. S., 336–381, Springer-Verlag, 1981.
- Tong, H.: *Non-Linear Time Series: A Dynamical System Approach*, Clarendon Press, 1990.
- Wilcox, B. P., Seyfried, M. S., and Matison, T. H.: Searching for chaotic dynamics in snowmelt runoff, *Water Resour. Res.*, 27, 6, 1005–1010, 1991.
- Zhou, Y., Ma, Z., and Wang, L.: Chaotic dynamics of the flood series in the Huaihe River Basin for the last 500 years, *J. Hydrol.*, 258, 100–110, 2002.