Stoch Environ Res Risk Assess (2008) 22:777–787
DOI 10.1007/s00477-007-0186-2

ORIGINAL PAPER

Signatures of low-dimensional chaos in hourly water level measurements at coastal site of Mariupol, Ukraine

V. Khokhlov · A. Glushkov · N. Loboda · N. Serbov · K. Zhurbenko

Published online: 11 September 2007
© Springer-Verlag 2007

Abstract Variations of water levels in ports and estuaries are important for ship guidance and navigation and are influenced by a variety of factors. The hourly data that was collected from the coastal site at the Port of Mariupol, Ukraine during January–December 2005 were analysed with an objective to reveal features of chaotic behaviour. The concepts and methods of chaos theory (average mutual information, correlation dimension, false nearest neighbours, Lyapunov exponents) were applied. The manifestation of low-dimensional chaos was found in the time series. The possibility of nonlinear prediction was concluded.

Keywords Coastal water level · Chaos theory · Nonlinear modelling · Predictability

1 Introduction

Coastal water levels are influenced by a variety of astronomical, meteorological, oceanographical and tectonic factors, the most readily apparent being the tides (e.g. Madsen and Jakobsen 2004; Benavente et al. 2006). At times, these factors interact in a complex way to elevate water levels significantly above normal tide level. Storms, which develop low atmospheric pressure, strong onshore winds and large waves, are the most common cause of elevated water levels.

The elevated water levels are of concern because they intensify damage to the coastline and to coastal developments. The elevated water levels allow larger waves to cross offshore bars and break closer to the beach, which in turn increases beach erosion and the threat to coastal developments. The elevated water levels can inundate low-lying areas of the coastline and around estuaries.

The elevated water level is not the only troublesome outcome of coastal wind events. Its companion, depressed sea level, can render navigation of coastal bays and harbours difficult and hazardous. Misleadingly called blow-out tides in the marine vernacular, sea level depression can have the same magnitude in height as floods.

The water level variability can be studied using statistical and mathematical models. For any attempt on mathematical explanatory modelling (e.g. Surkov et al. 1990; Bates et al. 2005), a detailed knowledge of the local oceanography and meteorology is required. When such data are not available, there is room for statistical modelling of mean water level time series. Srinivas et al. (2005) investigated the suitability of some statistical models for their predictive potential for the monthly sea level. They found that the exponentially weighted moving average technique gives the lowest root mean square errors relative to the verifying observations. Tilburg and Garvine (2004) developed and tested an empirical model based on principles of coastal Ekman circulation. Their model employs locally observed or forecast coastal winds and pressure and uses regression analysis where the subtidal frequency

V. Khokhlov (✉) · N. Loboda
Hydrometeorological Institute,
Odessa State Environmental University,
PO Box 116, Odessa 65009, Ukraine
e-mail: vkhokhlov@ukr.net

A. Glushkov · N. Serbov
Institute of Applied Mathematics,
Odessa State Environmental University,
PO Box 116, Odessa 65009, Ukraine

K. Zhurbenko
Hydrometeorological Observatory,
Lunin av. 5, Mariupol 87510, Ukraine

response is set by the along-shelf wind, the across-shelf wind stress, and the sea level atmospheric pressure. Sobey (2006) shown that the normal mode decomposition approach is a useful analysis methodology in the evaluation of storm tide and tsunami hazard at a coastal site. During the last years, the techniques based on wavelet decompositions (Różyński and Reeve 2005), cardinal B-spline basic functions (Wei and Billings 2006), artificial neural networks (Chang and Lin 2006) were used in the investigations of water level variations at different time scales. Among these methods, the chaotic time series analysis of water level occupies a fitting place.

Chaos theory establishes that apparently complex irregular behaviour could be the outcome of a simple deterministic system with a few dominant nonlinear interdependent variables. The past decade has witnessed a large number of studies employing the ideas gained from the science of chaos to characterize, model, and predict the dynamics of various geophysical phenomena (e.g. Sivakumar 2004). The outcomes of such studies are very encouraging, as they not only revealed that the dynamics of the apparently irregular geophysical phenomena could be understood from a chaotic deterministic point of view but also reported very good predictions using such an approach, in particular for water level variations (Frison et al. 1999; Zaldivar et al. 2000).

The present study attempts to employ a variety of techniques for characterizing the dynamics of hourly water level measured at the coastal site of Mariupol, Ukraine. More specifically, we attempt to identify the possible presence of chaotic dynamics in this time series. The techniques employed range from standard statistical techniques that can provide general indications regarding the dynamics of the phenomenon to specific ones that can provide comprehensive characterization of the dynamics. The standard statistical techniques used are the autocorrelation function and the Fourier power spectrum, whereas the mutual information approach, the correlation integral analysis, the false nearest neighbour algorithm, the Lyapunov exponents analysis, and the surrogate data method are employed for comprehensive characterization. Note that we give brief description only for the above methods of chaos theory; more exhaustive information can be found in the reviews of Abarbanel et al. (1993) and Schreiber (1999).

The organization of this paper is as follows. Details on the data considered for investigation are provided in the next section. Further, the methods used in this study and results are presented. Finally, conclusions drawn from the present study are discussed.

2 Data

In the present study, water level data observed at the Port of Mariupol, Ukraine are used. The port is located at 47°03′N, 37°30′E (north-western part of Taganrog Bay of the Azov Sea, 14 miles from the entry into the Bay). Average depth in the roads is 12 m and it allows handling ships with draught up to 8 m. Navigation in the Port of Mariupol is all year round.

For the present investigation, 1 year hourly water level data from 2005, consisting of a total of 8,760 data points, are used. Figure 1 shows the variation of this time series, and Table 1 presents some of the important statistics of the series. As it can be seen in Fig. 1, the water level exhibits significant variations without any apparent cyclicity. In the figure, the horizontal dashed lines indicate the levels corresponding dangerous elevated (531 cm) and depressed (411 cm) water levels. During 2005, the five events with dangerous elevated water level and single event of dangerous depressed water level were observed. It is clear that a visual inspection of the (irregular) water level series does not provide any clues regarding its dynamical behaviour, whether chaotic or stochastic.

To detect some regularity (or irregularity) in the time series, the Fourier power spectrum is often analyzed. For a

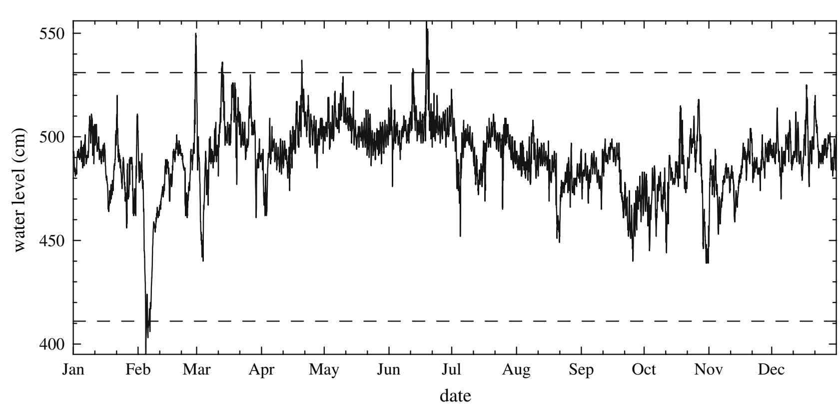

Fig. 1 Time series plot for 1-year water level data from 2005 at coastal site of Mariupol, Ukraine. Horizontal *dashed lines* indicate dangerous elevated (531 cm) and depressed (411 cm) water levels

Table 1 Some statistics of water level at coastal site of Mariupol during January–December 2005

Statistics	Value
Number of data	8,760
Mean (cm)	489.81
Maximum value (cm)	556
Minimum value (cm)	395
Standard deviation (cm)	16.93
Skewness	–0.94
Kurtosis	3.25

purely random process, the power spectrum oscillates randomly about a constant value, indicating that no frequency explains any more of the variance of the sequence than any other frequency. For a periodic or quasi-periodic sequence, only peaks at certain frequencies exist; measurement noise adds a continuous floor to the spectrum. Chaotic signals may also have sharp spectral lines but even in the absence of noise there will be continuous part (broadband) of the spectrum. The broad power spectrum falling as a power of frequency is a first indication of chaotic behaviour, though it alone does not characterize chaos (Abarbanel et al. 1993). From this point of view, the time series analyzed in this study is presumably chaotic (Fig. 2). However, more well-defined conclusion on the dynamics of the time series can be made after the data will be treated by methods of chaos theory.

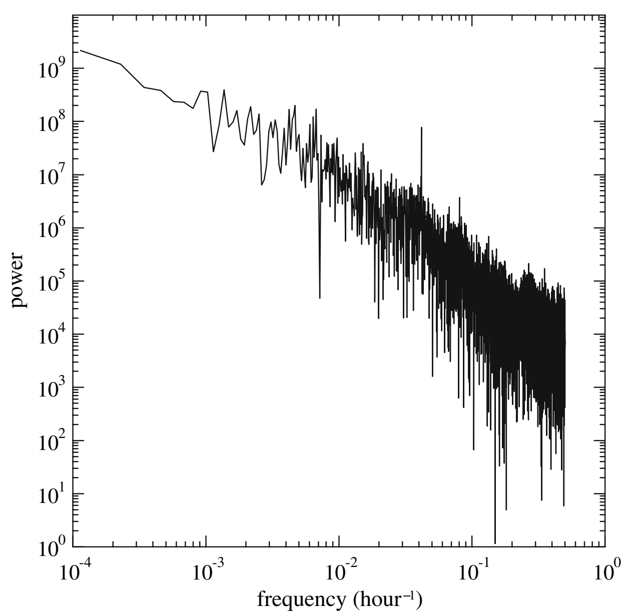

Fig. 2 Fourier power spectrum for hourly water level data at coastal site of Mariupol

3 Investigation of chaos in water level time series

Let us consider scalar measurements $s(n) = s(t_0 + n\Delta t) = s(n)$, where t_0 is the start time, Δt is the time step, and n is the number of measurements. In a general case, $s(n)$ is any time series, particularly the water level. Since processes resulting in the chaotic behaviour are fundamentally multivariate, it is necessary to reconstruct phase space using as well as possible information contained in the $s(n)$. Such a reconstruction results in a certain set of d-dimensional vectors $\mathbf{y}(n)$ replacing the scalar measurements. Packard et al. (1980) introduced the method of using time-delay coordinates to reconstruct the phase space of an observed dynamical system. The main idea is that the direct use of the lagged variables $s(n + \tau)$, where τ is some integer to be determined, results in a coordinate system in which the structure of orbits in phase space can be captured. Then using a collection of time lags to create a vector in d dimensions,

$$\mathbf{y}(n) = [s(n), s(n+\tau), s(n+2\tau), \ldots, s(n+(d-1)\tau)], \quad (1)$$

the required coordinates are provided. In a nonlinear system, the $s(n + j\tau)$ are some unknown nonlinear combination of the actual physical variables that comprise the source of the measurements. The dimension d is also called the embedding dimension d_E. The example of the Lorenz attractor given by Abarbanel et al. (1993) is a good choice to illustrate the efficiency of the method.

3.1 Choosing time lag

The statement of Mañé (1981) and Takens (1981) that any time lag will be acceptable is not terribly useful for extracting physics from data. If τ is chosen too small, then the coordinates $s(n + j\tau)$ and $s(n + (j + 1)\tau)$ are so close to each other in numerical value that they cannot be distinguished from each other. Similarly, if τ is too large, then $s(n + j\tau)$ and $s(n + (j + 1)\tau)$ are completely independent of each other in a statistical sense. Also, if τ is too small or too large, then the correlation dimension of attractor can be under- or overestimated, respectively (Havstad and Ehlers 1989). It is therefore necessary to choose some intermediate (and more appropriate) position between above cases.

First approach is to compute the linear autocorrelation function

$$C_L(\delta) = \frac{\frac{1}{N}\sum_{m=1}^{N}[s(m+\delta)-\bar{s}][s(m)-\bar{s}]}{\frac{1}{N}\sum_{m=1}^{N}[s(m)-\bar{s}]^2}, \quad (2)$$

where $\bar{s} = \frac{1}{N}\sum_{m=1}^{N} s(m)$ and N is the number of data measurements, and to look for that time lag where $C_L(\delta)$ first passes through zero (Holzfuss and Mayer-Kress 1986). This gives a good hint of choice for τ at that $s(n + j\tau)$ and $s(n + (j + 1)\tau)$ are linearly independent. However, a linear independence of two variables does not mean that these variables are nonlinearly independent since a nonlinear relationship can differs from linear one. It is therefore preferably to utilize approach with a nonlinear concept of independence, e.g. the average mutual information.

Briefly, the concept of mutual information can be described as follows. Let there are two systems, A and B, with measurements a_i and b_k. The amount one learns in bits about a measurement of a_i from a measurement of b_k is given by the arguments of information theory (Gallager 1968) as

$$I_{AB}(a_i, b_k) = \log_2\left(\frac{P_{AB}(a_i, b_k)}{P_A(a_i)P_B(b_k)}\right), \tag{3}$$

where the probability of observing a out of the set of all A is $P_A(a_i)$, and the probability of finding b in a measurement B is $P_B(b_i)$, and the joint probability of the measurement of a and b is $P_{AB}(a_i, b_k)$. The mutual information I of two measurements a_i and b_k is symmetric and non-negative, and equals to zero if only the systems are independent. The average mutual information between any value a_i from system A and b_k from B is the average over all possible measurements of $I_{AB}(a_i, b_k)$,

$$I_{AB}(\tau) = \sum_{a_i, b_k} P_{AB}(a_i, b_k) I_{AB}(a_i, b_k). \tag{4}$$

To place this definition into the context of observations from a certain physical system, let us think of the sets of measurements $s(n)$ as the A and of the measurements a time lag τ later, $s(n + \tau)$, as the B set. The average mutual information between observations at n and $n + \tau$ is then

$$I_{AB}(\tau) = \sum_{a_i, b_k} P_{AB}(a_i, b_k) I_{AB}(a_i, b_k). \tag{5}$$

Now we have to decide what property of $I(\tau)$ we should select, in order to establish which among the various values of τ we should use in making the data vectors $\mathbf{y}(n)$. Fraser and Swinney (1986) suggest, as a prescription, that it is necessary to choose that τ where the first minimum of $I(\tau)$ occurs.

Figure 3a presents the variations of the autocorrelation coefficient for the hourly water level observed at the Port of Mariupol in 2005 up to the lag time equal to 1,000 h. As it can be seen, the autocorrelation function exhibits some kind of exponential decay up to a lag time of about 100 h. Such an exponential decay can be an indication of the presence of chaotic dynamics in the process of water level variations. On the other hand, the autocorrelation coefficient failed to achieve zero, i.e. the autocorrelation function analysis not provides us with any value of τ. Such an analysis can be certainly extended to values exceeding 1,000, but Islam and Sivakumar (2002) showed that an attractor cannot be adequately reconstructed for very large values of τ.

Figure 3b shows the variation of the mutual information function against the lag time. The mutual information function exhibits an initial rapid decay (up to a lag time of about 10 h) followed more slow decrease before attaining near-saturation at the first minimum. Thus, we can use in following investigations the value of τ equals to 40 h that is obtained by using the average mutual information analysis.

Let us also note that the autocorrelation function and average mutual information can be to some extent considered as analogues of the linear redundancy and general redundancy, respectively, which was applied by Paluš (1995) in the test for nonlinearity. If a time series under consideration have an n-dimensional Gaussian distribution, these statistics are theoretically equivalent (Paluš 1995). The general redundancies detect all dependences in the time series, while the linear redundancies are sensitive only to linear structures. Although we do not perform the test for nonlinearity of Paluš (1995) in full, the simple comparison of the curves in Fig. 3a and b shows that most of features observed in the autocorrelation function values are missing in the average mutual information. In other words, the nature of curves in Fig. 3a and b is substantially different. From this fact, a possible nonlinear nature of process resulting in the water level variations can be concluded.

3.2 Choosing embedding dimension

The goal of the embedding dimension determination is to reconstruct a Euclidean space R^d large enough so that the set of points d_A can be unfolded without ambiguity. In accordance with the embedding theorem, the embedding dimension d_E must be greater, or at least equal, than a dimension of attractor d_A, i.e. $d_E > d_A$. In other words, we can choose a fortiori large dimension d_E, e.g. 10 or 15, since the previous analysis provides us prospects that the dynamics of our system is probably chaotic. However, two problems arise with working in dimensions larger than really required by the data and time-delay embedding (Abarbanel et al. 1993). First, many of computations for extracting interesting properties from the data require searches and other operations in R^d whose computational cost rises exponentially with d. Second, but more

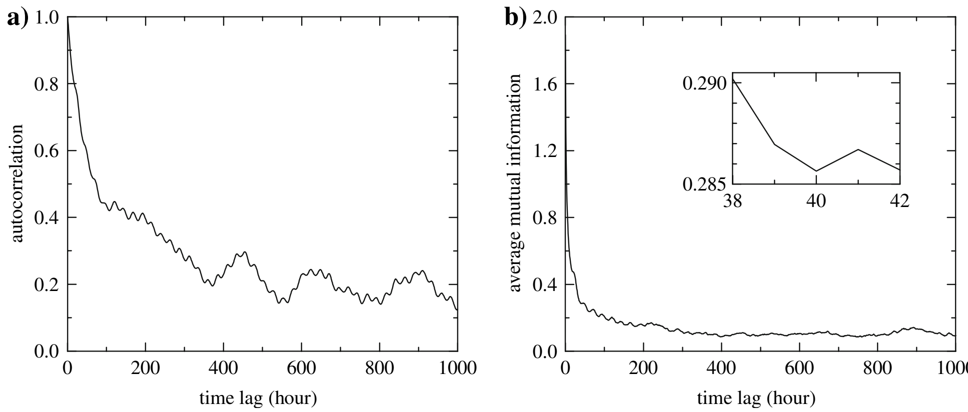

Fig. 3 (**a**) Autocorrelation function and (**b**) average mutual information for hourly water level at coastal site of Mariupol

significant from the physical point of view, in the presence of noise or other high dimensional contamination of the observations, the extra dimensions are not populated by dynamics, already captured by a smaller dimension, but entirely by the contaminating signal. In too large an embedding space one is unnecessarily spending time working around aspects of a bad representation of the observations which are solely filled with noise. It is therefore necessary to determine the dimension d_A.

There are several standard approaches to reconstruct the attractor dimension (see, e.g., Abarbanel et al. 1993; Schreiber 1999), but let us consider in this study two methods only.

The correlation integral analysis is one of the widely used techniques to investigate the signatures of chaos in a time series. The analysis uses the correlation integral $C(r)$, to distinguish between chaotic and stochastic systems. To compute the correlation integral, the algorithm of Grassberger and Procaccia (1983) is the most commonly used approach. According to this algorithm, the correlation integral is computed as

$$C(r) = \lim_{N\to\infty} \frac{2}{N(n-1)} \sum_{\substack{i,j \\ (1\le i<j\le N)}} H\left(r - \|\mathbf{y}_i - \mathbf{y}_j\|\right), \qquad (6)$$

where H is the Heaviside step function with $H(u) = 1$ for $u > 0$ and $H(u) = 0$ for $u \le 0$, r is the radius of sphere centered on $\mathbf{y}_i$ or $\mathbf{y}_j$. If the time series is characterized by an attractor, then the correlation integral $C(r)$ is related to the radius r given by

$$d = \lim_{\substack{r\to 0 \\ N\to\infty}} \frac{\log C(r)}{\log r}, \qquad (7)$$

where d is the correlation exponent that can be determined as the slop of line in the coordinates log $C(r)$ versus log r by a least-squares fit of a straight line over a certain range of r, called the scaling region.

If the correlation exponent attains saturation with an increase in the embedding dimension, then the system is generally considered to exhibit chaotic dynamics. The saturation value of the correlation exponent is defined as the correlation dimension (d_2) of the attractor. The nearest integer above the saturation value provides the minimum or optimum embedding dimension for reconstructing the phase–space or the number of variables necessary to model the dynamics of the system. On the other hand, if the correlation exponent increases without bound with increase in the embedding dimension, the system under investigation is generally considered stochastic.

In this study, the correlation functions and the exponents was computed for the hourly water level. Figure 4 shows the correlation dimension results, i.e. the relationship between the correlation exponent and embedding dimension values. As it can be seen, the correlation exponent value increases with embedding dimension up to a certain value, and then saturates beyond that value. The saturation of the correlation exponent beyond a certain embedding dimension is an indication of the existence of deterministic dynamics. The saturation value of the correlation exponent, i.e. correlation dimension of attractor, for the water level series is about 3.46 and occurs at the embedding dimension value of 6. The low, non-integer correlation dimension value indicates the existence of low-dimensional chaos in the hourly water level data of Mariupol.

The nearest integer above the correlation dimension value can be considered equal to the minimum dimension of the phase–space essential to embed the attractor. The value of the embedding dimension at which the saturation of the correlation dimension occurs is considered to provide the upper bound on the dimension of the phase–space sufficient to describe the motion of the attractor. Furthermore, the dimension of the embedding phase–space is equal to the number of variables present in the evolution of the system dynamics. Therefore, the results from the present study indicate that to model the dynamics of process resulting in the water level variations the minimum number of variables essential is equal to 4 and the number of variables sufficient is equal to 6. Therefore, the water level attractor should be embedded at least in a four-dimensional phase–space. The results also indicate that the upper bound on the dimension of the phase–space sufficient to describe the motion of the attractor, and hence the number of variables sufficient to model the dynamics of process resulting in the water level variations is equal to 6.

There are certain important limitations in the use of the correlation integral analysis in the search for chaos. For instance, the selection of inappropriate values for the parameters involved in the method may result in an underestimation (or overestimation) of the attractor dimension (Havstad and Ehlers 1989). Consequently, finite and low correlation dimensions could be observed even for a stochastic process (Osborne and Provenzale 1989). To verify the results obtained by the correlation integral analysis, we use surrogate data method.

The method of surrogate data (Theiler et al. 1992) is an approach that makes use of the substitute data generated in accordance to the probabilistic structure underlying the original data. This means that the surrogate data possess some of the properties, such as the mean, the standard deviation, the cumulative distribution function, the power spectrum, etc., but are otherwise postulated as random, generated according to a specific null hypothesis. Here, the null hypothesis consists of a candidate linear process, and the goal is to reject the hypothesis that the original data have come from a linear stochastic process. One reasonable statistic suggested by Theiler et al. (1992) is obtained as follows.

Let Q_{orig} denote the statistic computed for the original time series and $Q_{\mathrm{s}i}$ for the ith surrogate series generated under the null hypothesis. Let μ_s and σ_s denote, respectively, the mean and standard deviation of the distribution of Q_s. Then the measure of significance S is given by

$$S = \frac{|Q_{\mathrm{orig}} - \mu_s|}{\sigma_s}. \tag{8}$$

An S value of ~ 2 cannot be considered very significant, whereas an S value of ~ 10 is highly significant (Theiler et al. 1992). The details on the null hypothesis and surrogate data generation are described by Schreiber (1999).

To detect nonlinearity in the water level data, the one hundred realizations of surrogate data sets were generated according to a null hypothesis in accordance to the probabilistic structure underlying the original data. The correlation integrals and the correlation exponents, for embedding dimension values from 1 to 20, were computed for each of the surrogate data sets using the Grassberger–Procaccia algorithm as explained earlier. Figure 4 shows the relationship between the correlation exponent values and the embedding dimension values for the original data set and mean values of the surrogate data sets as well as for one surrogate realization. As it can be seen from Fig. 4, a significant difference in the estimates of the correlation exponents, between the original and surrogate data sets, is observed. In the case of the original data, a saturation of the correlation exponent is observed after a certain embedding dimension value (i.e., 6), whereas the correlation exponents computed for the surrogate data sets continue increasing with the increasing embedding dimension. The significance values (S) of the correlation exponent are computed for each embedding dimension and are shown in Fig. 5. The significance values lie mostly in the range between 10 and 50. The high significance values of the statistic indicate that the null hypothesis (the data arise from a linear stochastic process) can be rejected and hence the original data might have come from a nonlinear process.

Let us consider another method for determining d_{E} that comes from asking the basic question addressed in the embedding theorem: when has one eliminated false crossing of the orbit with itself which arose by virtue of having projected the attractor into a very low dimensional space? In other words, when points in dimension d are neighbours of one other? By examining this question in dimension one, then dimension two, etc. until there are no incorrect or false neighbours remaining, one should be able to establish, from geometrical consideration alone, a value for the

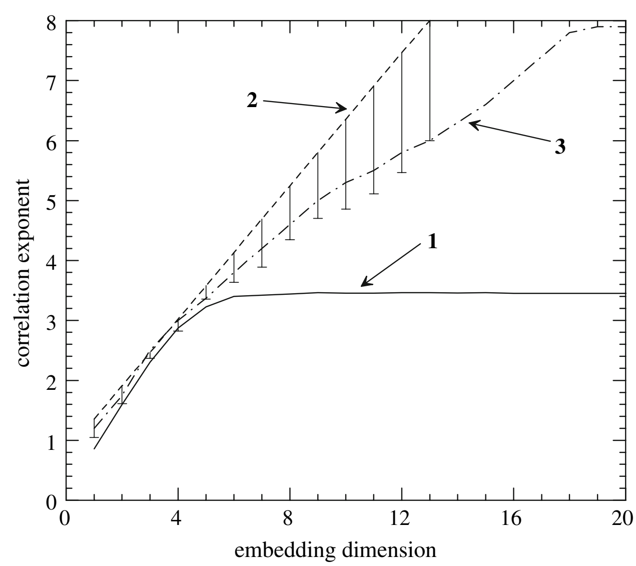

Fig. 4 Relationship between correlation exponent and embedding dimension for hourly water level data at coastal site of Mariupol for original time series (*line 1*), mean values of surrogate data sets (*line 2*), and one surrogate realization (*line 3*). Error bars indicate minimal values of correlation exponent among all realizations of surrogate data

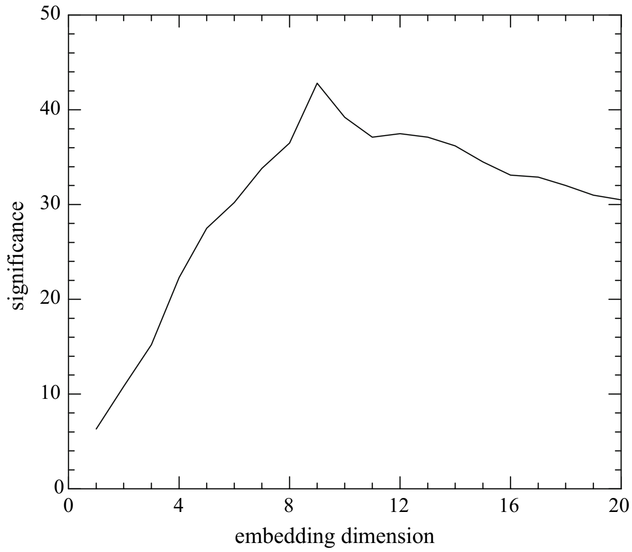

Fig. 5 Relationship between significance values of correlation dimension and embedding dimension

necessary embedding dimension. Such an approach was described by Kennel et al. (1992).

In dimension d each vector

$$\mathbf{y}(k) = [s(k), s(k+\tau), s(k+2\tau), \ldots, s(k+(d-1)\tau)] \quad (9)$$

has a nearest neighbour $\mathbf{y}^{\mathrm{NN}}(k)$ with nearness in the sense of some distance function. The Euclidean distance in dimension d between $\mathbf{y}(k)$ and $\mathbf{y}^{\mathrm{NN}}(k)$ we call $R_d(k)$:

$$R_d^2(k) = \left[s(k) - s^{\mathrm{NN}}(k)\right]^2 + \left[s(k+\tau) - s^{\mathrm{NN}}(k+\tau)\right]^2 + \cdots + \left[s(k+\tau(d-1)) - s^{\mathrm{NN}}(k+\tau(d-1))\right]^2 \quad (10)$$

$R_d(k)$ is presumably small when one has a lot a data, and for a dataset with N measurements, this distance is of order $1/N^{1/d}$. In dimension $d + 1$ this nearest-neighbour distance is changed due to the $(d + 1)$st coordinates $s(k + d\tau)$ and $s^{\mathrm{NN}}(k + d\tau)$ to

$$R_{d+1}^2(k) = R_d^2(k) + [s(k+d\tau) - s^{NN}(k+d\tau)]^2. \quad (11)$$

We can define some threshold size R_{T} to decide when neighbours are false. Then if

$$\frac{|s(k+d\tau) - s^{\mathrm{NN}}(k+d\tau)|}{R_d(k)} > R_{\mathrm{T}}, \quad (12)$$

the nearest neighbours at time point k are declared false. Kennel et al. (1992) showed that for values in the range $10 \le R_{\mathrm{T}} \le 50$ the number of false neighbours identified by this criterion is constant. In practice, the percentage of false nearest neighbours is determined for each dimension d. A value at which the percentage is almost equal to zero can be considered as the embedding dimension.

Figure 6 displays the percentage of false nearest neighbours that was determined for the water level series, for phase-spaces reconstructed with embedding dimensions from 1 to 20. As it can be seen, the percentage of false neighbours drops to almost zero at 4 or 5. This indicates that a four or five-dimensional phase–space is necessary to represent the dynamics (or unfold the attractor) of the water level series. From the other hand, the mean percentage of false nearest neighbours computed for the surrogate data sets decreases steadily but at 20 is about 35%. Such a result seems to be in close agreement with that was obtained from the correlation integral analysis, providing further support to the observation made earlier regarding the presence of low-dimensional chaotic dynamics in the water level variations.

3.3 Lyapunov exponents

Lyapunov exponents are the dynamical invariants of the nonlinear system. They are very useful when physics of process is considered. Using the spectrum of Lyapunov exponents, the average predictability of nonlinear system can be estimated. In a general case, the orbits of chaotic attractors are unpredictable, but there is the limited predictability of chaotic physical system, which is defined by the global and local Lyapunov exponents.

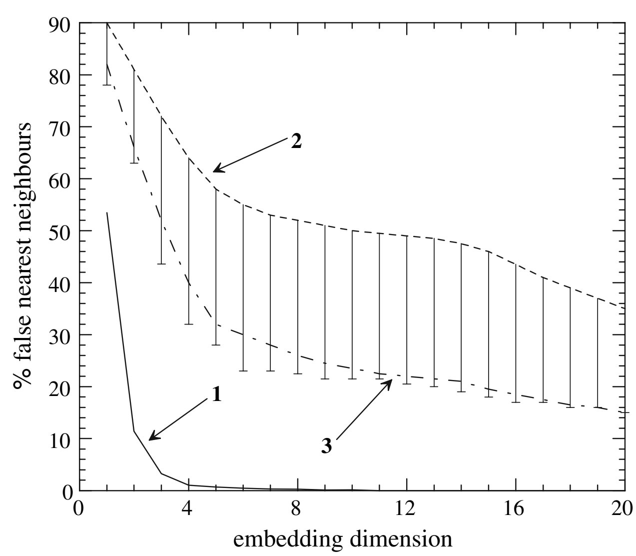

Fig. 6 Embedding dimension estimation using false nearest neighbour method for hourly water level data at coastal site of Mariupol for original time series (*line 1*), mean values of surrogate data sets (*line 2*), and one surrogate realization (*line 3*). Error bars indicate minimal percentage of false nearest neighbour among all realizations of surrogate data

A concept of Lyapunov exponents existed long before the establishment of chaos theory, and was developed to characterize the stability of absolute value of the eigenvalues of the linearized dynamics averaged over the attractor. A negative exponent indicates a local average rate of contraction while a positive value indicates a local average rate of expansion. In the chaos theory, the spectrum of Lyapunov exponents is considered a measure of the effect of perturbing the initial conditions of a dynamical system. Note that both positive and negative Lyapunov exponents can coexist in a dissipative system, which is then chaotic.

Since the Lyapunov exponents are defined as asymptotic average rates, they are independent of the initial conditions, and therefore they do comprise an invariant measure of attractor. In fact, if one manages to derive the whole spectrum of Lyapunov exponents, other invariants of the system, i.e. Kolmogorov entropy and attractor's dimension can be found. The Kolmogorov entropy K, measures the average rate at which information about the state is lost with time. An estimate of this measure is the sum of the positive Lyapunov exponents. The inverse of the Kolmogorov entropy is equal to the average predictability. The estimate of the dimension of the attractor is provided by the Kaplan and Yorke (1979) conjecture:

$$d_{\mathrm{L}} = j + \frac{\sum_{\alpha=1}^{j} \lambda_\alpha}{|\lambda_{j+1}|}, \tag{13}$$

where j is such that $\sum_{\alpha=1}^{j} \lambda_\alpha > 0$ and $\sum_{\alpha=1}^{j+1} \lambda_\alpha < 0$, and the Lyapunov exponents λ_α are taken in descending order.

There are several approaches to computing the Lyapunov exponents (see, e.g., Abarbanel et al. 1993; Schreiber 1999); in this paper, we use one which computes the whole spectrum and is based on the Jacobi matrix of the system function (Sano and Sawada 1985).

To calculate the spectrum of Lyapunov exponents from the water level data, we use the time delay $\tau = 40$ and embed the data in the four-dimensional space. Such a choice of the input parameters is the result of the previous calculations. Table 2 summarizes the results of the Lyapunov exponent analysis. For the time series under consideration, there exist two positive exponents (indicating expansion along two directions) and two negative ones (indicating contraction along remaining directions). The Kaplan–Yorke dimension is equal to 3.23; this value is very close to the correlation dimension which was determined by the Grassberger–Procaccia algorithm. The estimations of the Kolmogorov entropy and average predictability show that a limit, up to which the water level data can be on average predicted, is equal to 110.6 h or 4.6 days.

3.4 Statistical significance of results

It is known from experience that results of state–space reconstruction are highly sensitive to the length of data set (i.e. it must be sufficiently large) as well as to the time lag and embedding dimension determined.

Indeed, there are limitations on the applicability of chaos theory for observed (finite) hydrometeorological time series arising from the basic assumptions that the time series must be infinite. A finite and small data set may probably results in an underestimation of the actual dimension of the process. There are two opposite views on the sufficient length of data set. Smith (1988) concluded that the minimum number of data points ($N_{\min}$), must be equal to $42^{d_{\mathrm{E}}}$. In our case, $d_{\mathrm{E}} = 4$ and $N_{\min} = 3{,}111{,}696$, i.e. the length of data used in current study is certainly insufficient. From the other hand, Nerenberg and Essex (1990) showed that the minimum number of points is $N_{\min} \sim 10^{2+0.4 d_{\mathrm{E}}}$, i.e. $N_{\min}$ is about 4,000 for $d_{\mathrm{E}} = 4$. Sivakumar (2000) argued also that the phase space reconstruction can be successful for smaller observational data sets. Nevertheless, we check the robustness of our results with respect to the size of time series by dividing the data in 2 sets of 4,380 points and in 4 sets of 2,190 points and using the methods described in Sects. 3.2 and 3.3 for these subsets. The main assumption is that the results obtained for the subsets are close to the results obtained for the whole time series.

Table 3 presents a summary of the results achieved for the whole data set and the subsets. The correlation dimensions for the subsets are slightly smaller in comparison that for the whole data set. It is noteworthy that when the length of data sets is 2,190 points only, the correlation dimension is still comparable with $d_2 = 3.46$ obtained for the whole time series. The percentages of false nearest

Table 2 Results of Laypunov exponents analysis for water level data at coastal site of Mariupol during January–December 2005

λ_1	λ_2	λ_3	λ_4	d_{L}	K	P
0.0077	0.0013	–0.0051	–0.0175	3.23	0.0090	110.6

λ_1–λ_4 are the Lyapunov exponents in descending order, d_{L} is the Kaplan–Yorke attractor dimension, K is the Kolmogorov entropy, and P is the average predictability (hours)

Table 3 Correlation dimensions (d_2), embedding dimensions determined by false nearest neighbours method (d_N) with percentage of false neighbours (in parentheses), Lyapunov exponents in descending order (λ_1–λ_4), and Kolmogorov entropy (K) in different ranges of data points for water level data at coastal site of Mariupol during January–December 2005

Data points	d_2	d_N	λ_1	λ_2	λ_3	λ_4	K
1–8,760	3.46	4 (1.2)	0.0077	0.0013	–0.0051	–0.0175	0.0090
1–4,380	3.43	4 (2.0)	0.0072	0.0012	–0.0059	–0.0180	0.0084
4,381–8,760	3.45	4 (1.7)	0.0069	0.0012	–0.0055	–0.0168	0.0081
1–2,190	3.41	4 (2.1)	0.0094	0.0004	–0.0060	–0.0172	0.0098
2,191–4,380	3.42	4 (2.2)	0.0072	0.0006	–0.0066	–0.0176	0.0078
4,381–6,570	3.44	4 (1.8)	0.0066	0.0010	–0.0062	–0.0176	0.0076
6,571–8,760	3.43	4 (2.3)	0.0080	0.0003	–0.0058	–0.0172	0.0083

neighbours for the subsets are also insignificantly changed. Furthermore, there still exist two positive Lyapunov exponents for all subsets, and their values vary slightly especially for the first exponents. It is therefore suggested that, to reconstruct accurately the phase–space, the length of our time series is sufficient.

As it was mentioned above, an appropriate time lag is necessary because an optimum selection of τ gives best separation of neighboring trajectories within the minimum embedding phase–space. Note that the mutual information and autocorrelation function for some attractors behave in a different way. For example, these approaches applied to the Mackey–Glass system (Mackey and Glass 1977) provide equal values of τ, i.e. it really does not matter whether the autocorrelation function or the mutual information is used. On the other hand, for the system of Lorenz (1963) the mutual information method provides τ which is one order lesser than that determined by the autocorrelation function. In this study, we determine the time lag as a value at which the autocorrelation function first crosses the zero (Holzfuss and Mayer-Kress 1986). Other approaches consider the time lag at which the autocorrelation function attains a certain value, say 0.1 (Tsonis and Elsner 1988) or 0.5 (Schuster 1989). For observational time series, a practical approach is to experiment with different τ to ascertain its effect on the correlation dimension (e.g. Tsonis et al. 1993; Sivakumar 2000; Islam and Sivakumar 2002).

The dimensions determined for various time lags are presented in Table 4. Here, $\tau = 40$ is the time lag provided by the mutual information approach, and the correlation dimension is 3.46 due to the saturation at the embedding dimension 6 (see Fig. 4). Using other τ values of 30, 35, 45, and 50 h, an underestimation or overestimation of the dimensions is observed when τ is smaller or larger than 40 h, but the embedding dimension which is provided by the false nearest neighbours algorithm is still 4 (except for $\tau = 50$). If ever the criterion proposed by Schuster (1989) is used, $\tau = 75$ and $d_2 = 4.08$ with the embedding dimension $d_N = 6$. Moreover, if the time lag at which the autocorrelation function first crosses the zero (i.e. $\tau = 1{,}521$) or time lag at which the autocorrelation function attains the value of 0.1 (i.e. $\tau = 1{,}371$) is applied, the saturation is not observed at low dimensions. The value of $\tau = 40$ is therefore seems to be optimal for the time series discussed in this study.

Table 4 Correlation exponents (d_2) and embedding dimensions determined by false nearest neighbours method (d_N) with percentage of false neighbours (in parentheses) calculated for various time lags (τ) from hourly water level data at coastal site of Mariupol during January–December 2005

τ	d_2	d_N
30	3.22	4 (1.3)
35	3.29	4 (1.4)
40	3.46	4 (1.2)
45	3.56	4 (1.2)
50	3.78	5 (1.4)
75	4.08	6 (2.5)
1,371	No saturation	8 (4.2)
1,521	No saturation	8 (4.0)

Thus, the statistical convergence tests that are described in this subsection together with the surrogate data approach that was applied in Sect. 3.2 provide the satisfactory significance of our results regarding the state–space reconstruction.

4 Conclusions and discussion

This paper investigated the existence of chaotic behaviour in the hourly water level data at the coastal site of Mariupol, Ukraine. The mutual information approach, the correlation integral analysis, the false nearest neighbour algorithm, the Lyapunov exponents analysis, and the surrogate data method were used in the analysis.

The mutual information approach provided a time lag which is needed to reconstruct phase space. Such an approach allowed concluding the possible nonlinear nature of process resulting in the water level variations.

The correlation dimension method provided a low fractal-dimensional attractor thus suggesting a possibility of the existence of chaotic behaviour. Based on the attractor dimensions, the minimum number of variables essential to model the hourly water level dynamics at the Port of Mariupol was identified as 4 and the number of variables sufficient as 6. This implies that it is impossible to model the dynamics of the hourly process resulting in the water level variations with fewer than 4 variables. Significant improvement can be achieved when additional variables, up to the number of variables sufficient (6), are included in the model.

The method of surrogate data, for detecting nonlinearity, provided significant differences in the correlation exponents between the original data series and the surrogate data sets. This finding indicates that the null hypothesis (linear stochastic process) can be rejected.

The results from the aforementioned methods indicate that the hourly water level at the Port of Mariupol exhibits a nonlinear behaviour and possibly low-dimensional chaos. Thus, a short-term prediction based on nonlinear dynamics is possible. The Lyapunov exponents analysis supported this conclusion. It can be noted that the nonleading exponents are notoriously difficult to estimate from time series data. Moreover, the interpretation of inverse Lyapunov exponents as predictability times can results in ambiguous conclusions. In fact, the degree of instability and predictability can vary considerably throughout phase space (Schreiber 1999). Zaldivar et al. (2000) showed that nonlinear forecasting produces adequate results for the 'normal' dynamic behaviour of the water level of Venice Lagoon, outperforming linear algorithms, however, both methods fail to forecast the 'high water' phenomenon more than 2–3 h ahead.

Though a large number of studies employed the ideas gained from the science of chaos, there have also been widespread criticisms on the application of chaos theory. Important reasons for this are: (1) the assumptions with which the chaos identification methods have been developed, i.e. infinite and noise-free time series; and (2) the inability of the investigative methods to provide irrefutable proof regarding the existence of chaos. The fact that observational time series are almost always finite and are inherently contaminated by noise, such as errors arising from measurements, necessitate addressing the above issues in the application of chaos theory.

On the one hand, the basis for the criticisms of studies investigating and reporting existence of chaos in water level variations is our strong belief that they are influenced by a large number of variables and, therefore, are stochastic. On the other hand, the outcomes of the present study provide support to the claims that the (seemingly) highly irregular processes could be the result of simple deterministic systems with a few degrees of freedom. Therefore, the hypothesis of chaos in water level variations is reasonable and can provide an alternative approach for characterizing and modelling the dynamics of processes resulting in the water level variations. The evidence that chaos theory can be applied to observational time series was in detail adduced by Sivakumar (2000).

From our point of view, future investigations can be realized as follows. First, the adaptation of linear or empirical models such as that of Tilburg and Garvine (2004) is needed. Next, the nonlinear prediction method or artificial neural network approach can be applied to predict the water level variations. Comparing outcomes from the above methodologies, the best strategy for the short-term forecasting of water level in the Port of Mariupol can be achieved.

Acknowledgments The authors would like to thank the two anonymous reviewers for their valuable suggestions, which resulted in a more technically sound and complete presentation of the work.

References

Abarbanel HDI, Brown R, Sidorowich JJ, Tsimring LSh (1993) The analysis of observed chaotic data in physical systems. Rev Mod Phys 65:1331–1392

Bates PD, Dawson RJ, Hall JW, Horritt MS, Nicholls RJ, Wicks J, Hassan MAAM (2005) Simplified two-dimensional numerical modelling of coastal flooding and example applications. Coast Eng 52:793–810

Benavente J, Del Río L, Gracia FJ, Martínez-del-Pozo JA (2006) Coastal flooding hazard related to storms and coastal evolution in Valdelagrana spit (Cadiz Bay Natural Park, SW Spain). Cont Shelf Res 26:1061–1076

Chang H-K, Lin L-C (2006) Multi-point tidal prediction using artificial neural network with tide-generating forces. Coast Eng 53:857–864

Fraser AM, Swinney HL (1986) Independent coordinates for strange attractors from mutual information. Phys Rev A 33:1134–1140

Frison T, Abarbanel H, Earle M, Schultz J, Sheerer W (1999) Chaos and predictability in ocean water level measurements. J Geophys Res Ocean 104:7935–7951

Gallager RG (1968) Information theory and reliable communication. Wiley, New York

Grassberger P, Procaccia I (1983) Measuring the strangeness of strange attractors. Physica D 9:189–208

Havstad JW, Ehlers CL (1989) Attractor dimension of nonstationary dynamical systems from small data sets. Phys Rev A 39:845–853

Holzfuss J, Mayer-Kress G (1986) An approach to error-estimation in the application of dimension algorithms. In: Mayer-Kress G (ed) Dimensions and entropies in chaotic systems. Springer, Berlin, pp 114–122

Islam MN, Sivakumar B (2002) Characterization and prediction of runoff dynamics: a nonlinear dynamical view. Adv Water Resour 25:179–190

Kaplan JL, Yorke JA (1979) Chaotic behavior of multidimensional difference equations. In: Peitgen HO, Walter HO (eds)

Functional differential equations and approximations of fixed points. (Lecture notes in mathematics no 730). Springer, Berlin, pp 204–227
Kennel MB, Brown R, Abarbanel HDI (1992) Determining embedding dimension for phase–space reconstruction using a geometrical construction. Phys Rev A 45:3403–3411
Lorenz EN (1963) Deterministic nonperiodic flow. J Atmos Sci 20:130–141
Mackey M, Glass L (1977) Oscillation and chaos in physiological control systems. Science 197:287–289
Madsen H, Jakobsen F (2004) Cyclone induced storm surge and flood forecasting in the northern Bay of Bengal. Coast Eng 51:277–296
Mañé R (1981) On the dimensions of the compact invariant sets of certain non-linear maps. In: Rand DA, Young LS (eds) Dynamical systems and turbulence, Warwick 1980. (Lecture notes in mathematics No 898). Springer, Berlin, pp 230–242
Nerenberg MAH, Essex C (1990) Correlation dimension and systematic geometric effects. Phys Rev A 42:7065–7074
Osborne AR, Provenzale A (1989) Finite correlation dimension for stochastic systems with power-law spectra. Physica D 35:357–381
Packard NH, Crutchfield JP, Farmer JD, Shaw RS (1980) Geometry from a time series. Phys Rev Lett 45:712–716
Paluš M (1995) Testing for nonlinearity using redundancies: quantitative and qualitative aspects. Physica D 80:186–205
Różyński G, Reeve D (2005) Multi-resolution analysis of nearshore hydrodynamics using discrete wavelet transforms. Coast Eng 52:771–792
Sano M, Sawada Y (1985) Measurement of the Lyapunov spectrum from a chaotic time series. Phys Rev Lett 55:1082–1085
Schreiber T (1999) Interdisciplinary application of nonlinear time series methods. Phys Rep 308:1–64
Schuster HG (1989) Deterministic chaos: an introduction. Verlagsgesellschaft, Weinheim
Sivakumar B (2000) Chaos theory in hydrology: important issues and interpretations. J Hydrol 227:1–20
Sivakumar B (2004) Chaos theory in geophysics: past, present and future. Chaos Solitons Fractals 19:441–462
Smith LA (1988) Intrinsic limits on dimension calculations. Phys Lett A 133:283–288
Sobey RJ (2006) Normal mode decomposition for identification of storm tide and tsunami hazard. Coast Eng 53:289–301
Srinivas K, Kesava Das V, Dinesh Kumar PK (2005) Statistical modelling of monthly man sea level at coastal tide gauge stations along the Indian subcontinent. Ind J Mar Sci 34:212–224
Surkov FA, Krukier LA, Muratova GV (1990) Numerical modelling of the Sea of Azov's dynamics resulting from narrowing of the mouth of Taganrog Bay. Phys Oceanogr 1:551–559
Takens F (1981) Detecting strange attractors in turbulence. In: Rand DA, Young LS (eds) Dynamical systems and turbulence, Warwick 1980. (Lecture notes in mathematics No 898). Springer, Berlin, pp 366–381
Theiler J, Eubank S, Longtin A, Galdrikian B, Farmer JD (1992) Testing for nonlinearity in time series: the method of surrogate data. Physica D 58:77–94
Tilburg CE, Garvine RW (2004) A simple model for coastal sea level prediction. Weather Forecast 19:511–519
Tsonis AA, Elsner JB (1988) The weather attractor over very short timescale. Nature 333:545–547
Tsonis AA, Elsner JB, Georgakakos KP (1993) Estimating the dimension of weather and climate attractors: important issues about the procedure and interpretation. J Atmos Sci 50:2549–2555
Wei HL, Billings SA (2006) An efficient cardinal B-spline method for high tide forecasts at the Venice Lagoon. Nonlinear Process Geophys 13:577–584
Zaldivar JM, Gitiérrez E, Galván IM, Strozzi F, Tomasin A (2000) Forecasting high waters at Venice Lagoon using chaotic time series analysis and nonlinear neural networks. J Hydroinform 2:61–84