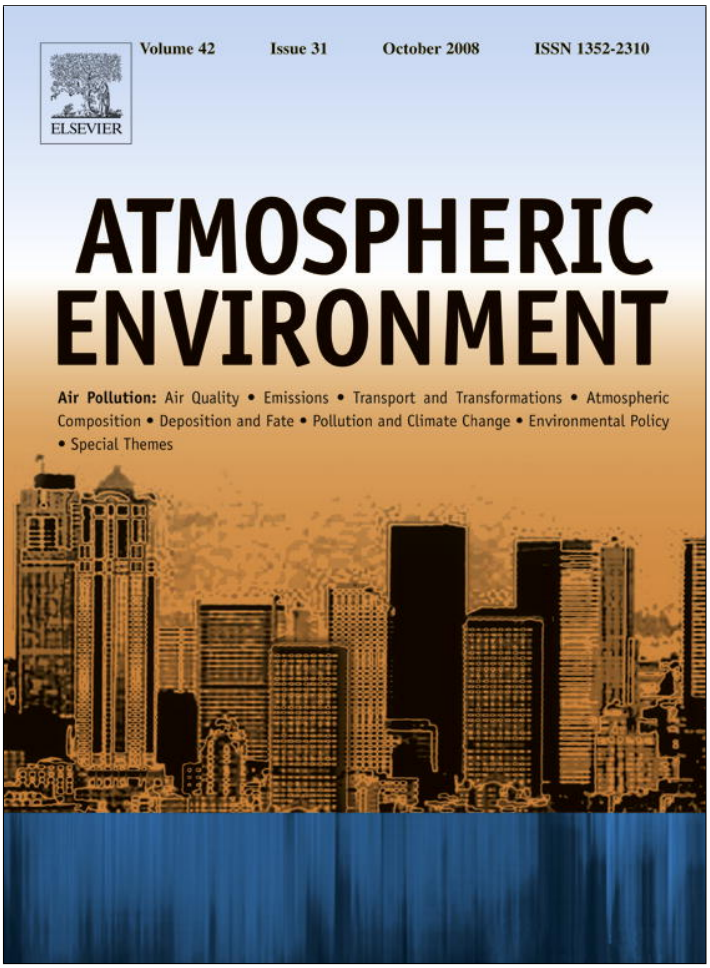

Atmospheric Environment 42 (2008) 7284–7292

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

Short-range forecast of atmospheric pollutants using non-linear prediction method

Valeriy N. Khokhlov [a,*], Alexander V. Glushkov [b], Nataliya S. Loboda [a], Yulia Y. Bunyakova [b]

[a] Hydrometeorological Institute, Odessa State Environmental University, 15 Lvovskaya Street, 65016 Odessa, Ukraine
[b] Institute of Applied Mathematics, Odessa State Environmental University, PO Box 116, 65009 Odessa, Ukraine

ARTICLE INFO

Article history:
Received 26 March 2008
Received in revised form 21 June 2008
Accepted 26 June 2008

Keywords:
Atmospheric pollutants
Chaos theory
Lyapunov exponents
Non-linear modelling

ABSTRACT

In this paper chaotic behavior in the nitrogen dioxide and sulphurous anhydride concentration time series at two sites in Gdansk region is investigated. To reconstruct an attractor, the time delay and embedding dimension are needed. The former is determined by the methods of autocorrelation function and average mutual information, and the latter is calculated by means of correlation dimension method and algorithm of false nearest neighbors. It was shown that the low-dimensional chaos existed in the time series under investigation. The spectrum of Lyapunov exponents was reconstructed as well as both Kaplan–Yorke dimension and Kolmogorov entropy that inversely proportional to the predictability limit are calculated. Non-linear prediction method is used for the time series. It is shown that even though the simple procedure is used to construct the non-linear model, the results are quite satisfactory.

1. Introduction

Every science purposes predicting a future state of system under consideration. Consequently, the main problem of science can be defined as: "Is it possible to predict a future behavior of process using its past states?" Conventional approach applied to resolve this problem consists in building an explanatory model using an initial data and parameterizing sources and interactions between process properties. Unfortunately that kind of approach is realized with difficulties, and its outcomes are insufficiently correct; moreover, sources and/or interactions of process cannot always be exactly defined. For example, the development of atmospheric numerical models is in progress during at least half a century, but short-range weather forecasts can be only considered as successful and forecast skill decreases with increasing of forecast range (see e.g. Kalnay, 2003). Most models, that are currently used to estimate the air pollution level, are either deterministic or statistical (e.g. Boy et al., 2006; Dirks et al., 2006), but their skilfulness are still limited due to both inability for describing non-linearities in pollutant time series and lack of understanding involved physical and/or chemical processes. In Gdansk region, the Agency of Regional Air Quality Monitoring (ARMAAG) provides presently the 24-h forecasts of air quality levels using the model called CALMET/CALPUFF (CALPUFF).

According to the modern theory of prediction, time series is considered as random realization, when the randomness is caused by a complicated motion with many independent degrees of freedom. Chaos is alternative of randomness and occurs in very simple deterministic systems. Although chaos theory places fundamental limitations for long-range prediction (see e.g. Abarbanel et al., 1993), it can be used for short-range prediction since ex facte random data can contain simple deterministic relationships with only a few degrees of freedom.

The systematic study of chaos is of recent date, originating in the 1960s. One important reason for this is that linear techniques, so long dominant within applied mathematics and the natural sciences, are inadequate when

* Corresponding author. Tel./fax: +380 482 637 227.
E-mail address: vkhokhlov@ukr.net (V.N. Khokhlov).

1352-2310/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.atmosenv.2008.06.023

considering chaotic phenomena since the amazingly irregular behavior of some non-linear deterministic systems was not appreciated and when such behavior was manifest in observations, it was typically explained as stochastic. Starting from the meteorologist Edward Lorenz, who observed extreme sensitivity to changes to initial conditions of a simple non-linear model simulating atmospheric convection (Lorenz, 1963), the experimental approach relies heavily on the computational study of chaotic systems and includes methods for investigating potential chaotic behavior in observational time series (see Tsonis, 1992; Wiggins, 1997; Ott, 2002).

During the last two decades, many studies in various fields of science have appeared, in which chaos theory was applied to a great number of dynamical systems, including those are originated from nature (e.g. Frison et al., 1999; Faure and Korn, 2001; Baas, 2002; Sivakumar, 2004, and a lot of another). Ecologists have also paid attention to the methodology with potentially wide scope (May, 1995; Letellier and Aziz-Alaoui, 2002; Sprott et al., 2005; Facchini et al., 2007; Mandal et al., 2007). On the other hand, the studies concerning non-linear behavior in the time series of atmospheric constituent concentrations are sparse, and their outcomes are ambiguous. For example, Lanfredi and Macchiato (1997) have investigated time series of NO_2, CO, and O_3 in Bristol and New Castle (Pennsylvania, USA), and have not concluded evidence of chaos. On the contrary, both Chen et al. (1998) and Koçak et al. (2000) have shown that O_3 concentrations in Cincinnati area (Ohio, USA) and Istanbul City (Turkey), respectively, are evidently chaotic, and non-linear prediction method provides quite satisfactory results. Moreover, Paluš et al. (2001) have denoted possible improvement of forecast skill if non-linear methods are applied. Using artificial neural network, where input data are the results of attractor dimension reconstruction, Chelani (2005) have obtained satisfactory short-range forecasts of PM10 concentration.

The above-mentioned studies concerning the atmospheric constituent concentrations allow concluding that methodology from chaos theory can be applied and the short-range forecast by the non-linear prediction method can be satisfactory. In contrast to conventional approach utilizing explanatory models, methodology from chaos theory do not require an in-depth knowledge of driving processes, nor do they require the form of the model to be specified a priori. In other words, the non-linear model looks much more 'simpler' than deterministic or statistical ones since it requires only that measurements must be comparatively long and continuous. From the other hand, time series are, however, not always chaotic, and chaotic behavior must be examined for each time series. Therefore, we shall (i) study the concentration of atmospheric constituents in Gdanck region (Poland) to select only those measurements, which are defined as chaotic, and (ii) build non-linear prediction model for the selected time series.

2. Data

In the present study, nitrogen dioxide (NO_2) and sulphurous anhydride (SO_2) concentration data observed at the sites of Gdansk region (Poland) during 2003 year are used. There are 10 sites in the region (Fig. 1), but the time series are continuous at two ones only, Sopot (site 6) and Gdynia (site 9). We use 1-year hourly concentrations, consisting of a total of 8760 data points. Fig. 2 shows the variation of these time series, and Table 1 presents some of the important statistics of the series. As can be seen in Fig. 2, the concentrations exhibit significant variations without any apparent cyclicity, i.e. a visual inspection of the (irregular) concentration time series does not provide any clues regarding its dynamical behavior, whether chaotic or stochastic. In the figure, horizontal dashed lines indicate the lower limits (96 $\mu g/m^3$ for NO_2 and 89 $\mu g/m^3$ for SO_2) for the index 2 of air pollution as provided by the EU Directives on Air Quality. A few events with values exceeding these limits have occurred during the year, and all the events were registered in winter or early spring, whereas in summer and autumn the concentrations of the pollutants were on the average lower especially for the sulphurous anhydride.

Although the distance between the sites 6 and 9 is only about 7 km and they are located on the beach of Gdansk Bay (the coordinates of sites 6 and 9 are 54°24′54″N, 18°34′47″E and 54°29′40″N, 18°33′15″E, respectively), both Table 1 and Fig. 2 show significant differences. For example, the absolute maxima (minima) at the site 6 are slightly higher (lower) than those at the site 9 (see Table 1). Also, there is a considerable increase of SO_2 concentration was registered only at the site 6 during early December, whereas the middle November peak has occurred only at the site 9. The overall variations of the pollutants at the two sites are nearly coherent with the exception of the above-mentioned cases and the lack of SO_2 concentration increase at the site 6 at the end of March. It is noteworthy that the time series of the concentrations not obey the Gaussian distribution (see Table 1).

3. Testing for chaos in time series

Let us consider scalar measurements $s(n) = s(t_0 + n\Delta t)$, where t_0 is the start time, Δt is the time step, and n is the

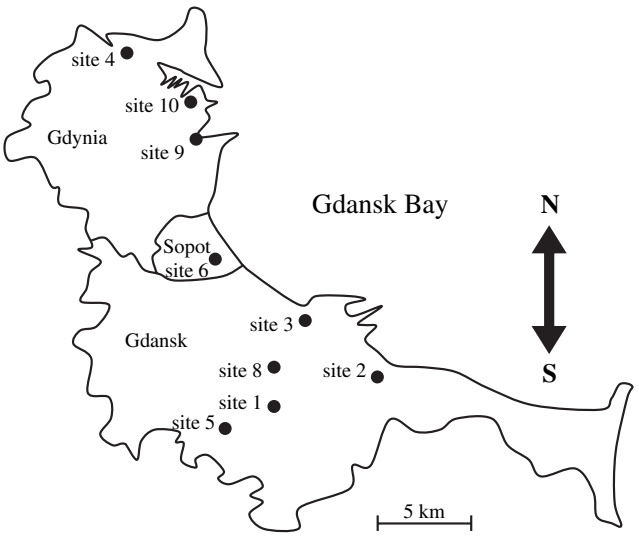

Fig. 1. Location of sites in Gdansk region, Poland. Site 7 is located southward and not included in figure.

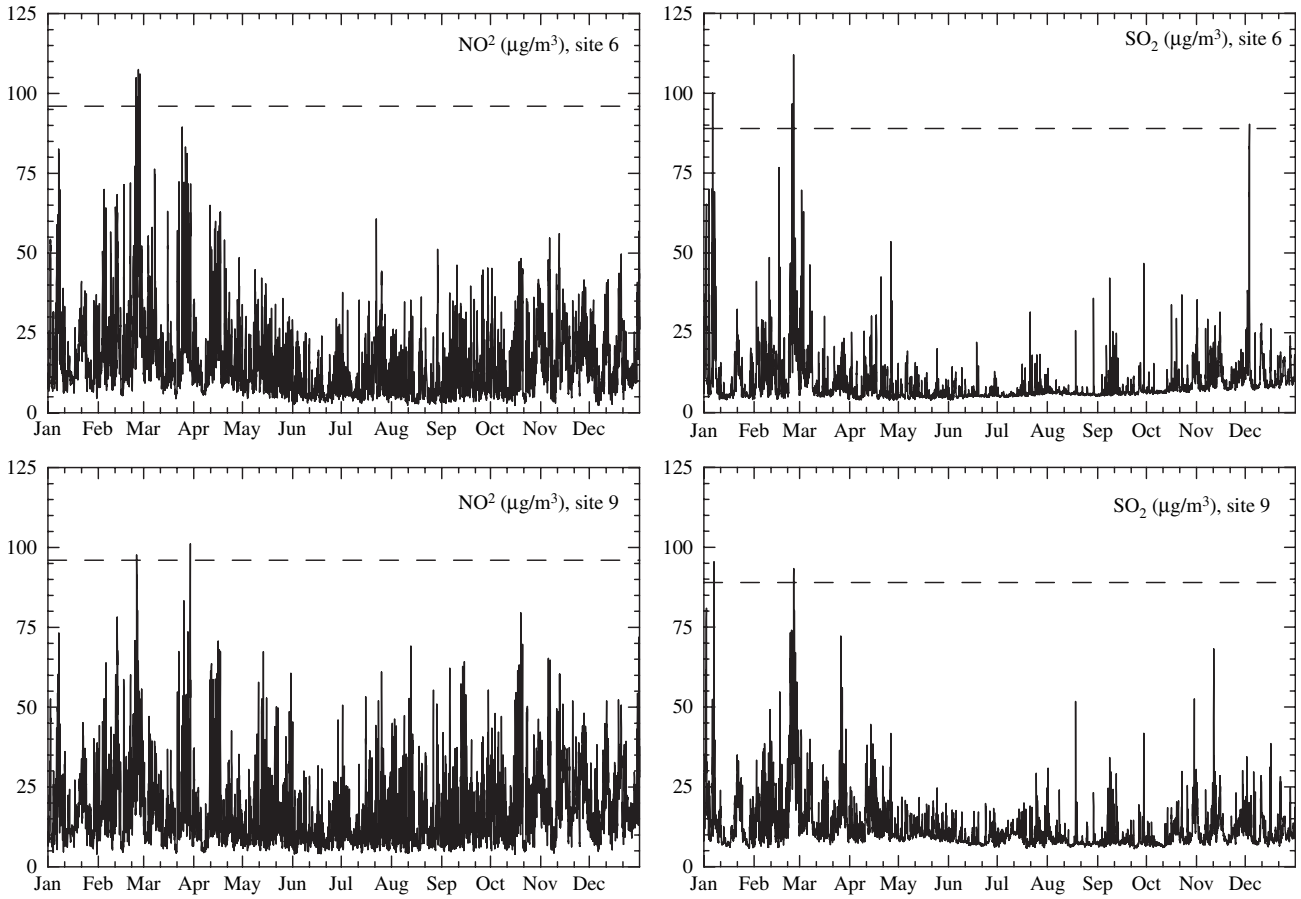

Fig. 2. Time series plot for 1-year atmospheric pollutant data from 2003 at two sites of Gdansk region, Poland. Horizontal dashed lines indicate lower limits (96 μg/m^3 for NO_2 and 89 μg/m^3 for SO_2) for the index 2 of air pollution as provided by the EU Directives on Air Quality.

number of the measurements. In a general case, $s(n)$ is any time series, particularly the concentrations of atmospheric pollutants. Since processes resulting in the chaotic behavior 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,

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

the required coordinates are provided. In a non-linear system, the $s(n+j\tau)$ are some unknown non-linear 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.

Table 1
Some statistics of air pollutant concentrations at the sites of Gdansk region (Poland) during January–December 2003

Statistics	Site 6 (Sopot)		Site 9 (Gdynia)	
	NO_2	SO_2	NO_2	SO_2
Number of data	8760	8760	8760	8760
Mean (μg/m^3)	15.46	9.13	17.04	11.84
Maximum value (μg/m^3)	107.53	111.99	101.13	95.47
Minimum value (μg/m^3)	2.29	3.99	3.92	5.59
Standard deviation (μg/m^3)	11.99	6.94	11.22	7.19
Skewness	2.26	4.79	1.81	3.89
Kurtosis	7.61	38.15	4.43	22.78

3.1. Time lag

The choice of proper time lag is very important for the subsequent reconstruction of phase space. 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

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

where $\bar{s} = (1/N)\sum_{m=1}^{N} s(m)$, and to look for that time lag where $C_{\mathrm{L}}(\delta)$ first passes through 0 (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 non-linearly independent since a non-linear relationship can differs from linear one. It is therefore preferably to utilize approach with a non-linear 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

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

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 0 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 $I_{AB}(a_i, b_k)$.

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, $s(n+t)$, as the B set. The average mutual information between observations at n and $n+\tau$ is then

$$I_{AB}(\tau) = \sum_{a_i, b_k} P_{AB}(a_i, b_k) I_{AB}(a_i, b_k). \quad (2)$$

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) have suggested, as a prescription, that it is necessary to choose that τ where the first minimum of $I(\tau)$ occurs.

3.2. 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. $d_{\mathrm{E}} \geq d_{\mathrm{A}}$. 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 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

$$C(r) = \lim_{N\to\infty} \frac{2}{N(n-1)} \sum_{\substack{i,j \\ (1\leq i<j\leq N)}} H(r - \|\mathbf{y}_i - \mathbf{y}_j\|), \quad (3)$$

where H is the Heaviside step function with $H(u)=1$ for $u>0$ and $H(u)=0$ for $u\leq 0$, r is the radius of sphere centered on y_i or y_j, and N is the number of data measurements. If the time series is characterized by an attractor, then the correlation integral $C(r)$ is related to the radius r given by

$$d = \lim_{\substack{r\to 0 \\ N\to\infty}} \frac{\log C(r)}{\log r}, \quad (4)$$

where d is correlation exponent that can be determined as the slope 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.

Another method for determining d_{E} 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 too low-dimensional space? In other words, when points in dimension d are neighbors of one other? By examining this question in dimension one, then dimension two, etc. until there are no incorrect or false neighbors remaining, one should be able to establish, from geometrical consideration alone, a value for the necessary embedding dimension. Such an approach was described by Kennel et al. (1992).

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

$$R_d^2(k) = \left[s(k) - s^{NN}(k)\right]^2 + \left[s(k+\tau) - s^{NN}(k+\tau)\right]^2 + \cdots + \left[s(k+\tau(d-1)) - s^{NN}(k+\tau(d-1))\right]^2. \quad (5)$$

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

$$R_{d+1}^2(k) = R_d^2(k) + \left[s(k+d\tau) - s^{NN}(k+d\tau)\right]^2. \quad (6)$$

We can define some threshold size R_T to decide when neighbors are false. Then if

$$\frac{\left|s(k+d\tau) - s^{NN}(k+d\tau)\right|}{R_d(k)} > R_T, \quad (7)$$

the nearest neighbors at time point k are declared false. Kennel et al. (1992) showed that for values in the range $10 \le R_T \le 50$ the number of false neighbors identified by this criterion is constant. In practice, the percentage of false nearest neighbors is determined for each dimension d. A value at which the percentage is almost equal to 0 can be considered as the embedding dimension.

3.3. Results for atmospheric pollutant time series

As it was mentioned above, the first step, which is needed to reconstruct the phase space, is to chose an appropriate time lag τ. For that purpose, we have used methods from Section 3.1. 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. However, before making up any final decision, let us make some remarks on using the autocorrelation function. Holzfuss and Mayer-Kress (1986) have determined the time lag as a value at which the autocorrelation function first crosses the 0. 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, 2005). An ambiguity is what value of autocorrelation function must be accepted? For observational time series, a practical approach is to experiment with different τ to ascertain its effect on the dimension of attractor (e.g. Sivakumar, 2000; Islam and Sivakumar, 2002).

Table 2 summarizes the results for the time lag calculated for the first thousand values of the time series. It is noteworthy that the autocorrelation function crosses the 0 only for the NO_2 time series at the site 9, whereas this statistic for other time series remains positive. Of course, the values, where the autocorrelation function first crosses 0.1, can be chosen as τ, but Islam and Sivakumar (2002) showed that an attractor cannot be adequately reconstructed for very large values of τ. Before making up final decision we therefore calculate the dimension of attractor for all values in Table 2 using methods described in Section 3.2.

The large values of τ result in impossibility to determine both the correlation exponents and the attractor dimensions (see Table 3) using the method of Grassberger and Procaccia (1983). Such an outcome can be explained not only inappropriate values of τ but also shortcomings of correlation dimension method (see e.g. Sivakumar, 2000; Khokhlov et al., 2008). Moreover, if the algorithm of Kennel et al. (1992) is used, then the percentages of false nearest neighbors are comparatively large in the case of large τ. On the other hand, if the time lags determined by the average mutual information are used, then the algorithm of false nearest neighbors provides the embedding dimension $d_E = 6$ for all air pollutants. Since we assume that the dynamics of system is preferably non-linear, the latter value will be used in the following investigations.

4. Model of non-linear prediction

Before building a model of non-linear prediction, it is important to define how predictable is a chaotic system? The predictability can be estimated by the Kolmogorov entropy, which is proportional to the sum of positive Lyapunov exponents. The spectrum of Lyapunov exponents is one of dynamical invariants for non-linear system with chaotic behavior. The chaotic motion of physical system is not unpredictable even all the instabilities one encounters in phase space. The limited predictability of the chaos is quantified by the local and global Lyapunov exponents, which can be determined from the measurements.

4.1. Lyapunov exponents

The Lyapunov exponents measure the average local rates of expansion and contraction of chaotic system. The

Table 2
Time lags (h) subject to different values of autocorrelation function, C_L, and first minima of average mutual information, I_{min1}, for the time series of NO_2 and SO_2 at the sites of Gdansk region (Poland) during January–December 2003

	Site 6 (Sopot)		Site 9 (Gdynia)	
	NO_2	SO_2	NO_2	SO_2
$C_L = 0$	–	–	102	–
$C_L = 0.1$	136	232	53	147
$C_L = 0.5$	6	12	4	26
I_{min1}	9	19	8	17

Table 3
Correlation exponents (d_2) and embedding dimensions determined by false nearest neighbors method (d_N) with percentage of false neighbors (in parentheses) calculated for various time lags (τ) for the time series of NO_2 and SO_2 at the sites of Gdansk region (Poland) during January–December 2003

Site 6 (Sopot)						Site 9 (Gdynia)					
NO_2			SO_2			NO_2			SO_2		
τ	d_2	d_N	τ	d_2	d_N	τ	d_2	d_N	τ	d_2	d_N
136	n/s[a]	11 (6.2)	232	n/s[a]	10 (8.8)	53	7.62	9 (9.2)	147	n/s[a]	10 (9.8)
6	5.42	6 (1.3)	12	1.64	6 (1.2)	4	5.29	6 (1.1)	26	3.95	6 (1.1)
9	5.31	6 (1.2)	19	1.58	6 (1.2)	8	5.31	6 (1.1)	17	3.40	6 (1.2)

[a] The saturation is not registered.

Lyapunov exponents are related to the eigenvalues of the linearized dynamics across the attractor. Negative values show stable behavior while positive values show local unstable behavior. For chaotic systems, being both stable and unstable, the Lyapunov exponents indicate the complexity of the dynamics. The largest positive value determines some "average" prediction limit. High-dimensional chaotic systems tend to have very large positive exponents and predictions may be of little use. Since the advent of chaos, the spectrum of Lyapunov exponents can be considered a measure of the effect of perturbing the initial conditions of a dynamical system. 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 hence the choice of trajectory, and therefore they do comprise an invariant measure of the attractor. In fact, if one manages to derive the whole spectrum of Lyapunov exponents, other invariants of the system, i.e. the Kolmogorov entropy, K, and the attractor's dimension can be found. The Kolmogorov entropy 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

$$K = \sum_{\alpha=1}^{j} \lambda_\alpha. \tag{8}$$

The average limit of predictability, Pr_{max}, is equal to the inverse of the Kolmogorov entropy, i.e. $Pr_{max} = 1/K$, since the units of Lyapunov exponents are inverse times. It is naturally that predictability can vary considerably throughout phase space, and the Lyapunov exponents constitute a particular way of averaging over these variations (Schreiber, 1999).

The estimate of the dimension of the attractor is provided by the Kaplan and Yorke (1979) conjecture

$$d_L = j + \frac{\sum_{\alpha=1}^{j} \lambda_\alpha}{|\lambda_{j+1}|}. \tag{9}$$

Note that in Eqs. (8)–(10) j is such that $\sum_{\alpha=1}^{j} \lambda_\alpha > 0$ and $\sum_{\alpha=1}^{j+1} \lambda_\alpha < 0$, and the Lyapunov exponents are taken in descending order. The dimension d_L gives values close to the dimension estimates discussed earlier and is preferable when estimating high dimensions.

There are a few approaches to computing the Lyapunov exponents. One of them computes the whole spectrum and is based on the Jacobi matrix of system. In the case where only observations are given and the system function is unknown, the matrix has to be estimated from the data. In this case, all the suggested methods approximate the matrix by fitting a local map to a sufficient number of nearby points. In this study, we use the method with the linear fitted map proposed by Sano and Sawada, 1985, although the maps with higher order polynomials can be also used.

4.2. Non-linear modelling

Non-linear modelling of chaotic processes is based on the concept of compact geometric attractor on which observations evolve. Since the orbit is continually folded back on itself by the dissipative forces and the non-linear part of the dynamics, some orbit points $\mathbf{y}^r(k)$, $r = 1, 2, \ldots, N_B$ can be found in the neighborhood of any orbit point $\mathbf{y}(k)$, at that the points $\mathbf{y}^r(k)$ arrive in the neighborhood of $\mathbf{y}(k)$ at quite different times than k. One can then choose some interpolation functions, which account for whole neighborhoods of phase space and how they evolve from near $\mathbf{y}(k)$ to whole set of points near $\mathbf{y}(k+1)$ (Abarbanel et al., 1993).

The implementation of this concept is to build parametrized non-linear functions $\mathbf{F}(\mathbf{x}, \mathbf{a})$ which take $\mathbf{y}(k)$ into $\mathbf{y}(k+1) = \mathbf{F}(\mathbf{y}(k), \mathbf{a})$ and then use various criteria to determine parameter $\mathbf{a}$. Further, since one has the notion of local neighborhoods, one can build up one's model of the process neighborhood by neighborhood and, by piecing together these local models, produce a global non-linear model that capture much of the structure in the attractor itself.

It is noteworthy that no clue is given by the data as to the kind of model that would be appropriate for the source of chaotic data. It is likely there is no algorithmic solution to how to choose models from data alone (Rissanen, 1989). Schreiber (1999) showed that most widespread form of local model is quite simple:

$$s(k + \Delta k) = a_0^{(k)} + \sum_{j=1}^{d_A} a_j^{(k)} s(k - (j-1)\tau), \tag{10}$$

where Δk is the time over which predictions are being made. The coefficients $a_j^{(k)}$, may be determined by a least-squares procedure, involving only points $s(k)$ within a small neighborhood around the reference point. Thus, the coefficients will vary throughout phase space. The fit procedure

Table 4
First two Lyapunov exponents (λ_1 and λ_2), Kaplan–Yorke dimension (d_L), and average limit of predictability (Pr_{max}, h) for the time series of NO_2 and SO_2 at the sites of Gdansk region (Poland) during January–December 2003

	Site 6 (Sopot)		Site 9 (Gdynia)	
	NO_2	SO_2	NO_2	SO_2
λ_1	0.0184	0.0164	0.0189	0.0150
λ_2	0.0061	0.0066	0.0052	0.0052
d_L	4.11	5.01	3.85	4.60
Pr_{max}	40	43	41	49

amounts to solving $(d_A + 1)$ linear equations for the $(d_A + 1)$ unknowns.

When fitting the parameters a, several problems are encountered that seem purely technical in the first place but are related to the non-linear properties of the system. If the system is low-dimensional, the data that can be used for fitting will locally not span all the available dimensions but only a subspace, typically. Therefore, the linear system of equations to be solved for the fit will be ill-conditioned. However, in the presence of noise the equations are not formally ill-conditioned but still the part of the solution that relates the noise directions to the future point is meaningless (Schreiber, 1999).

4.3. Short-range forecast of atmospheric pollutant time series

At first, Table 4 shows the global Lyapunov exponents. It can note that the Kaplan–Yorke dimensions, which are also the attractor dimensions, are smaller than the dimensions obtained by the algorithm of false nearest neighbors. Also, there are the two positive λ_i for the each time series under consideration. Since the Lyapunov exponents determine conversion rate from a sphere into an ellipsoid, then the smaller sum of positive exponents results in the more stable dynamical system and, correspondingly, the higher predictability. The presence of the two (from six) positive λ_i suggests the system broadens in the line of two axes and converges along four axes that in the six-dimensional space. The time series of SO_2 at the site 9 have the highest predictability (slightly more than 2 days), and other time series have the predictabilities slightly less than 2 days. Such the predictability is quite sufficient.

To use the non-linear prediction method, it is necessary to solve another one problem which can be defined as how much exactly nearest neighbors, NN, must be considered to obtain satisfactory results of the forecasts? This problem springs from the methodology used here for local modelling, when a collection of near neighbors of the point $\mathbf{y}(k)$ is taken and an averaged value of their images is considered as the 'prediction' (Abarbanel et al., 1993). The solution can be arrived at if a few forecasts are accomplished with various number of NN, and their results are compared with original data. As a rule, the coefficient of correlation rises to a maximum (Islam and Sivakumar, 2002). In the current study, we use this approach the last 100 points of the time series; the 24-h forecasts are accomplished on the basis of previous 8660 data points. Table 5 summarizes the results of our experiments. As it was expected, the coefficients of correlation rise to the maxima at some number of NN. It is noteworthy that these coefficients are both large and significant. Thus, we further use NN = 180 for NO_2 and NN = 260 for SO_2 at the site 6, as well as NN = 210 for NO_2 and NN = 250 for SO_2 at the site 9.

Fig. 3 shows, as an example, the original data and 24-h forecasts of pollutants at the site 6. As can be seen, almost all the maxima in the original data have the counterparts in the forecast, but the differences between them can be sufficiently great. This discrepancy is well illustrated by Fig. 4; all values exceeding $\sim 35\ \mu g/m^3$ for NO_2 and $\sim 15\ \mu g/m^3$ for SO_2 were underestimated by the non-linear model. Some possible causes for such insufficient forecasts of comparatively large concentrations will be considered in the next section. Nevertheless, the root-mean-square errors are 6.025 and 5.636 for NO_2 and SO_2, respectively. The results of the short-range forecasts can be thus considered as quite acceptable.

5. Discussion and conclusions

In this study, we have analyzed the time series of pollutants in the atmosphere using methodology from chaos theory. Our results show that these time series are resulted from the low-dimensional chaos. In spite of the fact that the embedding dimensions for each time series are identical ($d_N = 6$), this outcome is not regularity (see, e.g. Chen et al., 1998; Koçak et al., 2000). Also, the correlation dimensions are calculated using the algorithm of Grassberger and Procaccia (1983). It is noteworthy that 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. This concept can be applied to the dynamics of NO_2 time series, since the embedding dimension determined by both the correlation dimension method and the algorithm of false nearest neighbors is identical. In this case, the number of variables necessary to model the dynamics of the system equals six. From the other hand, the analysis of correlation dimension provides only the number of variables, but not their physical meaning.

In the previous section, we have mentioned that our non-linear model underestimate comparatively large

Table 5
Coefficient correlation (r) between actual data and 24-h forecast subject to number of neighbors (NN) for last 100 points of the time series of NO_2 and SO_2 at the sites of Gdansk region (Poland) during January–December 2003

	Site 6 (Sopot)						Site 9 (Gdynia)					
	NO_2			SO_2			NO_2			SO_2		
NN	80	180	200	80	260	280	80	210	230	80	250	270
r	0.95	0.96	0.96	0.91	0.94	0.94	0.96	0.97	0.97	0.93	0.94	0.94

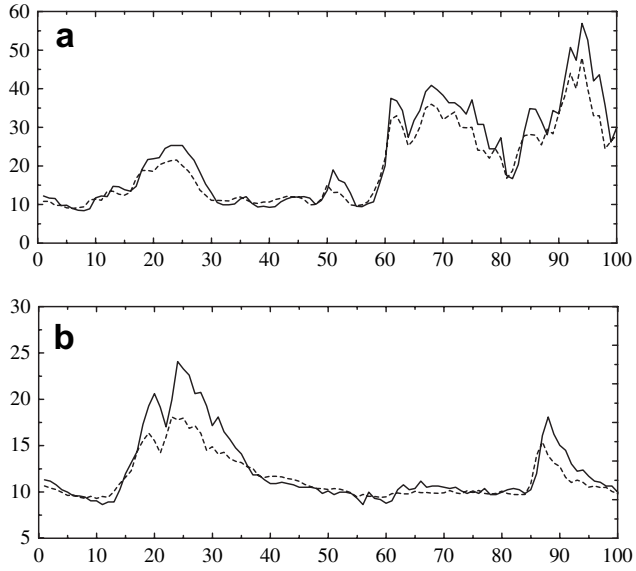

Fig. 3. Original data (solid lines) and 24-h forecasts (dashed lines) for last 100 data points of (a) NO_2 and (b) SO_2 from 2003 at site 6 of Gdansk region, Poland.

values of concentrations. In our opinion, if unavoidable simulation errors are not taken into account, there are at least two causes for such discrepancy. First one springs from the fact that we use relatively simple method of local modelling. If more complicated procedure will be applied then the forecasts can improve. Second one is resulted from the ubiquity of noise in measurements.

Kantz and Schreiber, 2003 have summarized the following examples of the noise effects: (1) self-similarity of the attractor is broken; (2) phase-space reconstruction appears as high-dimensional on small length scales; (3) nearby trajectories diverge diffusively rather than exponentially; and (4) prediction error is found to be bounded from below no matter which prediction method is used and to how many digits the data are recorded. The noise limits the accuracy of predictions in three possible ways: (1) the prediction error cannot be smaller than the noise level, since the noise part of the future measurement cannot be predicted; (2) the values on which the predictions are based are themselves noisy, inducing an error proportional to and of the order of the noise level; and (3) in the generic case, where the dynamical evolution has to be estimated from the data, this estimate will be affected by noise (Schreiber and Kantz, 1996).

Conceivably, knowing the noise level in a time series can help one at least to understand why errors occur in the forecast. In this study, we used methodology proposed by Hu et al. (2004) to estimate the signal-to-noise ratio (SNR) that can be represented the relative variance of the signal and the noise. For the segments of the time series shown in Fig. 3, the noise levels are estimated to be about $3.6 \pm 0.6\%$ and $4.2 \pm 0.5\%$ in terms of variance. Thus, the presence of noise has effects on the prediction results. The magnitudes of the estimated noise levels seem to be quite reasonable considering the fact that the measurements of air pollutant concentration are influenced by a large number of meteorological and human-induced factors.

Thus, our results can be considered as the example of quite successful short-range forecast for the concentrations of atmospheric pollutants. It is noteworthy that the non-linear prediction method provides the satisfactory results in the case, when the concentrations are sharply rising; at least, all the tendencies to the rising were revealed by the method. In addition, we have used the simplest approach for the approximation of local model and more complicated methodology can provide hopefully the better forecasts.

The non-linear model possesses some advantages over explanatory models. Particularly, the errors due to parameterizations of sources and interactions, which are inherent in explanatory models, are absent in the non-linear model. Further, it is not necessary to estimate the characteristics described in Sections 3 and 4.1 for every forecasting procedure since they define certain dynamics of system. The latter, in turn, depends on some meteorological and human-induced factors, and if these factors are unchangeable then the parameters (i.e. time lag and dimension of attractor) needed to model the dynamics of

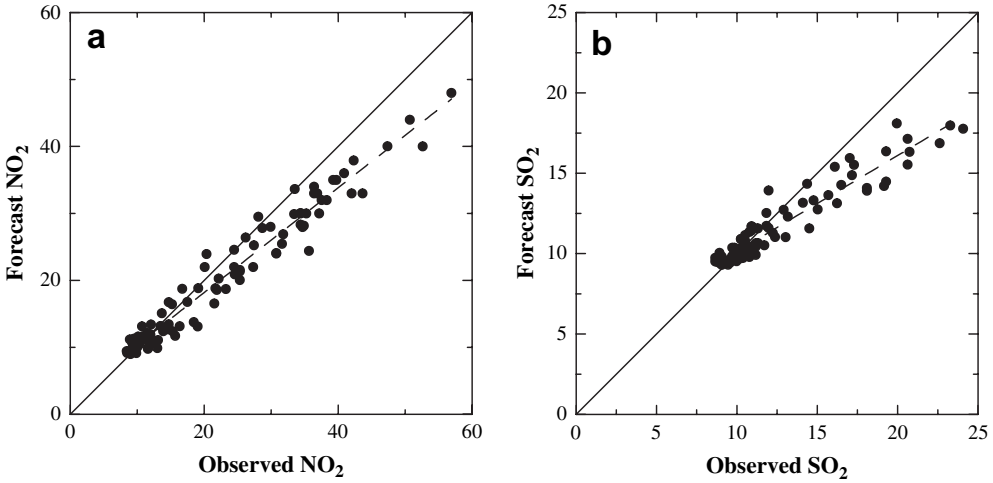

Fig. 4. Scatter plots of observation (abscissa) versus 24-h forecast (ordinate) for last 100 data points of (a) NO_2 and (b) SO_2 from 2003 at site 6 of Gdansk region, Poland. Dashed lines are linear fits.

system are also unchangeable. Moreover, the calculations of above-mentioned parameters take the most part of time, whereas the time required for the forecasting procedure by personal computer is a few minutes only. Finally, the nonlinear model can be considered as a good alternative of conventional methods, especially for regions where deterministic or statistical models cannot be used because of one or another reasons.

Acknowledgements

We wish to thank two anonymous reviewers for their insightful comments which noticeably improved the original manuscript.

References

Abarbanel, H.D.I., Brown, R., Sidorowich, J.J., Tsimring, L.Sh., 1993. The analysis of observed chaotic data in physical systems. Reviews of Modern Physics 65, 1331–1392.

ARMAAG. Available from: http://armaag.gda.pl/en/index.htm.

Baas, A.C.W., 2002. Chaos, fractals and self-organization in coastal geomorphology: simulating dune landscapes in vegetated environments. Geomorphology 48, 309–328.

Boy, M., Hellmuth, O., Korhonen, H., Nilsson, E.D., ReVelle, D., Turnipseed, A., Arnold, F., Kulmala, M., 2006. MALTE – model to predict new aerosol formation in the lower troposphere. Atmospheric Chemistry and Physics 6, 4499–4517.

CALPUFF. Available from: http://www.src.com/calpuff/calpuff1.htm.

Chelani, A.B., 2005. Predicting chaotic time series of PM10 concentration using artificial neural network. International Journal of Environmental Studies 62, 181–191.

Chen, J.-L., Islam, S., Biswas, P., 1998. Nonlinear dynamics of hourly ozone concentrations: nonparametric short term prediction. Atmospheric Environment 32, 1839–1848.

Dirks, K.N., Nanni, A., Dirks, V.I., 2006. Modelling and predicting urban atmospheric pollutants in the Aosta Valley region of Italy using a site-optimised model. Atmospheric Science Letters 7, 15–20.

Facchini, A., Mocenni, C., Marwan, N., Vicino, A., Tiezzi, E., 2007. Nonlinear time series analysis of dissolved oxygen in the Orbetello Lagoon (Italy). Ecological Modelling 203, 339–348.

Faure, P., Korn, H., 2001. Is there chaos in the brain? I. Concepts of nonlinear dynamics and methods of investigation. Life Sciences 324, 773–793.

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

Frison, T.W., Abarbanel, H.D.I., Earle, M.D., Schultz, J.R., Scherer, W.D., 1999. Chaos and predictability in ocean water levels. Journal of Geophysical Research C 104, 7935–7951.

Gallager, R.G., 1968. Information Theory and Reliable Communication. Wiley, New York, 608 pp.

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

Havstad, J.W., Ehlers, C.L., 1989. Attractor dimension of nonstationary dynamical systems from small data sets. Physical Review 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 Heidelberg New York, pp. 114–122.

Hu, J., Gao, J.B., White, K.D., 2004. Estimating measurement noise in a time series by exploiting nonstationarity. Chaos, Solitons & Fractals 22, 807–819.

Islam, M.N., Sivakumar, B., 2002. Characterization and prediction of runoff dynamics: a nonlinear dynamical view. Advances in Water Resources 25, 179–190.

Kalnay, E., 2003. Atmospheric Modeling, Data Assimilation and Predictability. Cambridge University Press, Cambridge, 340 pp.

Kantz, H., Schreiber, T., 2003. Nonlinear Time Series Analysis. Cambridge University Press, Cambridge. 386 pp.

Kaplan, J.L., Yorke, J.A., 1979. Chaotic behavior of multidimensional difference equations. In: Peitgen, H.-O., Walter, H.-O. (Eds.), Functional Differential Equations and Approximations of Fixed Points. Lecture Notes in Mathematics No. 730. Springer, Berlin, pp. 204–227.

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

Khokhlov, V., Glushkov, A., Loboda, N., Serbov, N., Zhurbenko, K., 2008. Signatures of low-dimensional chaos in hourly water level measurements at coastal site of Mariupol, Ukraine. Stochastic Environmental Research and Risk Assessment 22, 777–787.

Koçak, K., Şaylan, L., Şen, O., 2000. Nonlinear time series prediction of O_3 concentration in Istanbul. Atmospheric Environment 34, 1267–1271.

Lanfredi, M., Macchiato, M., 1997. Searching for low dimensionality in air pollution time series. Europhysics Letters 1997, 589–594.

Letellier, C., Aziz-Alaoui, M.A., 2002. Analysis of the dynamics of a realistic ecological model. Chaos, Solitons & Fractals 13, 95–107.

Lorenz, E.N., 1963. Deterministic nonperiodic flow. Journal of the Atmospheric Sciences 20, 130–141.

Mackey, M., Glass, L., 1977. Oscillation and chaos in physiological control systems. Science 197, 287–289.

Mandal, S., Ray, S., Roy, S., Jørgensen, S.E., 2007. Investigation of thermodynamic properties in an ecological model developing from ordered to chaotic states. Ecological Modelling 204, 40–46.

May, R.M., 1995. Necessity and chance: deterministic chaos in ecology and evolution. Bulletin of the American Mathematical Society 32, 291–308.

Ott, E., 2002. Chaos in Dynamical Systems. Cambridge University Press, Cambridge, 490 pp.

Packard, N.H., Crutchfield, J.P., Farmer, J.D., Shaw, R.S., 1980. Geometry from a time series. Physical Review Letters 45, 712–716.

Paluš, M., Pelikán, E., Eben, K., Krejčíř, P., Juruš, P., 2001. Nonlinearity and prediction of air pollution. In: Kurkova, V. (Ed.), Artificial Neural Nets and Genetic Algorithms. Springer, Wien, pp. 473–476.

Rissanen, J., 1989. Stochastic Complexity in Statistical Inquiry. World Scientific, Singapore, 177 pp.

Sano, M., Sawada, Y., 1985. Measurement of the lyapunov spectrum from chaotic time series. Physical Review Letters 55, 1082–1085.

Schreiber, T., 1999. Interdisciplinary application of nonlinear time series methods. Physics Reports 308, 1–64.

Schreiber, T., Kantz, H., 1996. Observing and predicting chaotic signals: is 2% noise too much? In: Kravtsov Yu, A., Kadtke, J.B. (Eds.), Predictability of Complex Dynamical Systems. Springer Series in Synergetics. Springer, Berlin, pp. 43–65.

Schuster, H.G., 2005. Deterministic Chaos: an Introduction. Wiley, New York, 312 pp.

Sivakumar, B., 2000. Chaos theory in hydrology: important issues and interpretations. Journal of Hydrology 227, 1–20.

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

Sprott, J.C., Vano, J.A., Wildenberg, J.C., Anderson, M.B., Noel, J.K., 2005. Coexistence and chaos in complex ecologies. Physics Letters A 335, 207–212.

Tsonis, A.A., 1992. Chaos: From Theory to Application. Plenum Press, New York, 274 pp.

Tsonis, A.A., Elsner, J.B., 1988. The weather attractor over very short timescale. Nature 333, 545–547.

Wiggins, S., 1997. Introduction to Applied Nonlinear Dynamical Systems and Chaos. Springer-Verlag, New York, 688 pp.