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
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

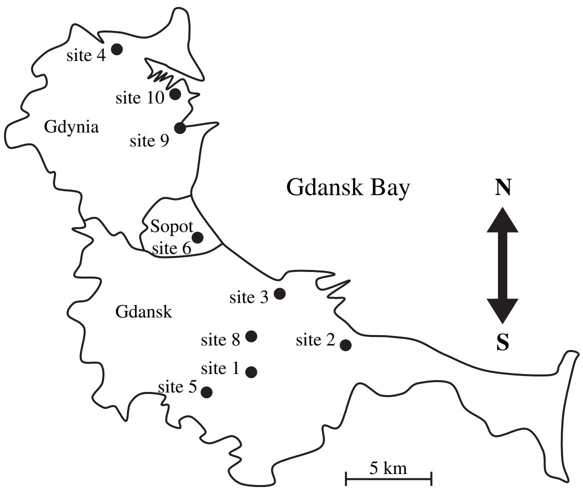

Fig. 1. Location of sites in Gdansk region, Poland. Site 7 is located southward and not included in figure.