UDC 519.24.001.57

**A.V. Glushkov,** *d.ph.-m.n.,* **N.G. Serbov,** *c.geogr.n.,* **I.A. Shakhman**, *c.geogr.n.,*
**A.K. Balan,** *prep.,* **V.F. Mansarliyskii**, *prep .,* **A.A. Dudinov***, as.*

*Odessa State Environmental University*

# CHAOS-GEOMETRIC METHOD IN SHORT-RANGE FORECAST OF HYRDO- ATMOSPHERIC POLLUTANTS: ADVANCED ESTIMATES

*Within the chaos-geometric approach there are obtained improved data on the analysis and forecasting chaotic fluctuations in the time series of concentrations of nitrogen dioxide and sulfur dioxide in the atmosphere of the Gdansk region.*
***Keywords:*** *hydro-and atmosphere pollution, modeling, chaos-geometric approach*

## 1. Introduction

It is known that a chaos is alternative of randomness and occurs in very simple deterministic systems. Although chaos theory places fundamental limitations for long-rage prediction [1-8], 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. Many studies in various fields of science have appeared, where chaos theory was applied to a great number of dynamical systems. The studies concerning non-linear behaviour in the time series of atmospheric constituent concentrations are sparse, and their outcomes are ambiguous.

In refs. [5,6] there is an analysis of the $NO_2$, CO, $O_3$ concentrations time series and is not received an evidence of chaos. Also, it was shown that $O_3$ concentrations in Cincinnati (Ohio) and Istanbul are evidently chaotic, and non-linear approach provides satisfactory results. In ref. [2,3] there were developed a new approach to modelling, analyzing and forecasting hydro-and air pollution and presented geographically reasonable data, illustrating its high effectiveness and accuracy.

These studies show that chaos theory methodology can be applied and the short-range forecast by the non-linear prediction method can be satisfactory. Time series of concentrations are however not always chaotic, and chaotic behaviour must be examined for each air pollutants time series.

In this paper, we will present more advanced (improved) studying the concentration of the atmospheric and in principle (in full analogy) hydrological constituents in nature mediums. As example, we will use air pollution in the Gdansk region (Poland). Besides, we will use only those measurements, which are defined as chaotic. At last, we will present advanced non-linear prediction modelling for selected time series.

## 2. Advanced chaos-geometric approach to modeling hydro- and atmospheric pollution

As, in our previous papers, we will use the nitrogen dioxide ($NO_2$) and sulphurous anhydride ($SO_2$) concentration data observed at the sites of Gdansk region during 2003-2005. There are ten sites in the region, but time series are continuous at 2 ones only, Sopot (site 6) and Gdynia (site 9). We use one year hourly concentrations (total of 8760 data points).

Table 1 presents some of the important statistics (coordinates of sites 6 and 9 are 54°24'54"N, 18°34'47"E and 54°29'40"N, 18°33'15"E) [2].

Table 1 - Some statistics of air pollutant concentrations at the Gdansk region (Jan.-Dec.2003)

| Statistics | Site 6 (Sopot) | | Site 9 (Gdynia) | |
|---|---|---|---|---|
| | $NO_2$ | $SO_2$ | $NO_2$ | $SO_2$ |
| Number of data | 8760 | 8760 | 8760 | 8760 |
| Mean ($\mu g/m^3$) | 15.46 | 9.13 | 17.04 | 11.84 |
| Maximum value ($\mu g/m^3$) | 107.53 | 111.99 | 101.13 | 95.47 |
| Minimum value ($\mu g/m^3$) | 2.29 | 3.99 | 3.92 | 5.59 |
| Standard deviation ($\mu g/m^3$) | 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 |

As usually, let us consider scalar measurements $S(n)=s(t_0+ n\Delta t) = s(n)$, where $t_0$ is a start time, $\Delta t$ is time step, and $n$ is number of the measurements. In a general case, $s(n)$ is any time series (f.e. hydrological or atmospheric pollutants concentration). As processes resulting in a chaotic behaviour are fundamentally multivariate, one needs to reconstruct phase space using as well as possible information contained in $s(n)$. Such reconstruction results in set of $d$-dimensional vectors $\mathbf{y}(n)$ replacing scalar measurements. The main idea is that direct use of lagged variables $s(n+\tau)$, where $\tau$ is some integer to be defined, results in a coordinate system where a structure of orbits in phase space can be captured. Using a collection of time lags to create a vector in $d$ dimensions, $\mathbf{y}(n)=[s(n),s(n + \tau),s(n + 2\tau),..,s(n +(d-1)\tau)]$, the required coordinates are provided. In a nonlinear system, $s(n + j\tau)$ are some unknown nonlinear combination of the actual physical variables. The dimension $d$ is the embedding dimension, $d_E$.
The next important step is the choice of proper time lag. Obvious that it is important for the subsequent reconstruction of phase space. If $\tau$ is chosen too small, then the coordinates $s(n + j\tau)$, $s(n +(j +1)\tau)$ are so close to each other in numerical value that they cannot be distinguished from each other. If $\tau$ is too large, then $s(n+j\tau)$, $s(n+(j+1)\tau)$ are completely independent of each other in a statistical sense. If $\tau$ is too small or too large, then the correlation dimension of attractor can be under-or overestimated. One needs to choose some intermediate position between above cases. First approach is to compute the linear autocorrelation function $C_L(\delta)$ and to look for that time lag where $C_L(\delta)$ first passes through 0. This gives a good hint of choice for $\tau$ at that $s(n+j\tau)$ and $s(n+(j +1)\tau)$ are linearly independent. It's better to use approach with a nonlinear concept of independence, e.g. an average mutual information. 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 $I_{AB}(a_i, b_k)$. In ref. [4] it is suggested, as a prescription, that it is necessary to choose that $\tau$ where the first minimum of $I(\tau)$ occurs.

Further let us consider an advanced approach to the embedding dimension determination. 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. 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. 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. According to [8], it is computed the correlation integral $C(r)$. If the time series is characterized by an attractor, then the correlation integral $C(r)$ is related to the radius $r$ as

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

where $d$ is correlation exponent. 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 correlation exponent is defined as the correlation dimension ($d_2$) of the attractor (see details in refs. [2,8]).

The most fundamental and important block of the whole approach is formulation of comprehensive advanced non-linear prediction model. Here, let us begin from the rhetorical question. Namely, first of all, it's important to define how predictable is a chaotic system? The predictability can be estimated by the Kolmogorov entropy, which is proportional to a sum of positive LE. The spectrum of LE is one of dynamical invariants for non-linear system with chaotic behaviour. The limited predictability of the chaos is quantified by the local and global LE, which can be determined from measurements. The LE are related to the eigenvalues of the linearized dynamics across the attractor. Negative values show stable behaviour while positive values show local unstable behaviour. For chaotic systems, being both stable and unstable, LE indicate the complexity of the dynamics. The largest positive value determines some average prediction limit. Since the LE are defined as asymptotic average rates, they are independent of the initial conditions, and hence the choice of trajectory, and they do comprise an invariant measure of the attractor. An estimate of this measure is a
sum of the positive LE. The estimate of the attractor dimension is provided by the conjecture $d_L$ and the LE 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. To compute LE, we use a method with linear fitted map [1,2], although the maps with higher order polynomials can be used too. Non-linear model of chaotic processes is based on the concept of compact geometric attractor on which observations evolve. Since an orbit is continually folded back on itself by dissipative forces and the non-linear part of dynamics, some orbit points $\mathbf{y}^r(k)$, $r$ =1, 2, ..,$N_B$ can be found in the neighbourhood of any orbit point $\mathbf{y}(k)$, at that the points $\mathbf{y}^r(k)$ arrive in the neigh-bourhood of $\mathbf{y}(k)$ at quite different times than $k$. One can then choose some interpolation functions, which account for whole neighbourhoods of phase space and how they evolve from near $\mathbf{y}(k)$ to whole set of points near $\mathbf{y}(k+1)$. The implementation of this concept is to build parameterized 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 use various criteria to determine parameters $\mathbf{a}$. Since one has the notion of local neighbourhoods, one can build up one's model of the process neighbourhood by neighbourhood and, by piecing together these local models, produce a global non-linear model that capture much of the structure in an attractor itself. The other model and computation details can be found in [4-8].

## 3. Some illustrative results for hudro-and atmospheric pollutant time series

Table 2 summarizes the results for the time lag calculated for first $10^3$ values of time series. The autocorrelation function crosses 0 only for the $NO_2$ time series at the site 9, whereas this statistic for other time series remains positive. The values, where the autocorrelation function first crosses 0.1, can be chosen as $\tau$, but in [1] it's showed that an attractor cannot be adequately reconstructed for very large values of $\tau$. So, before making up final decision we calculate the dimension of attractor for all values in Table 2. The large values of $\tau$ result in impossibility to determine both the correlation exponents and attractor dimensions (Table 3) using Grassberger-Procaccia method [8].

Table 2 - Time lags (hours) subject to different values of $C_L$, and first minima of average mutual information, $I_{min1}$, for the time series of $NO_2$ , $SO_2$ at the sites of Gdansk reg. (Jan.-Dec. 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 |

The outcome is explained not only inappropriate values of τ but also shortcomings of correlation dimension method [2]. If algorithm [1] is used, then a percentages of false nearest neighbours are comparatively large in a case of large τ. If time lags determined by average mutual information are used, then algorithm of false nearest neighbours provides $d_E = 6$ for all air pollutants.

Table 3 - 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) for time series of $NO_2$, $SO_2$ of Gdansk reg. (Jan.-Dec. 2003)

| Site 6 (Sopot) $NO_2$ | | | Site 6 (Sopot) $SO_2$ | | | Site 9 (Gdynia) $NO_2$ | | | Site 9 (Gdynia) $SO_2$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| τ | $d_2$ | $d_N$ | τ | $d_2$ | $d_N$ | τ | $d_2$ | $d_N$ | τ | $d_2$ | $d_N$ |
| 136 | - | 11 (6.2) | 232 | - | 10 (8.8) | 53 | 7.62 | 9 (9.2) | 147 | - | 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) |

Table 4 shows the global LE. 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 neighbours. The presence of the two (from six) positive $\lambda_i$ 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 (more than 2 days), and other time series have the predictabilities slightly less than 2 days.

To use the non-linear prediction method, it is necessary to solve another one problem which can be defined as how much exactly nearest neighbours, *NN*, must be considered to obtain satisfactory results of the forecasts? Table 5 summarizes the results of our experiments. The coefficients of correlation rise to the maxima at some number of *NN*. These coefficients are both large and significant. So, we further use $NN = 180$ for $NO_2$ and $NN = 260$ for $SO_2$ at the site 6, as well as $NN = 2100$ for $NO_2$ and $NN = 250$ for $SO_2$ at the site 9

In conclusion let us underline that we have investigated a chaotic behaviour in the nitrogen dioxide and sulphurous anhydride concentration time series at 2 sites in Gdansk region and proved an existence of the low-dimensional chaos in these series. We have presented an effective nonlinear prediction model and realized a successful short-range forecast of atmospheric pollutant time series.

Acknowledgements. The work is carried out in accordance with the NDS-159 project of Ministry of Education and Science of Ukraine.

Table 4 - First two LE($\lambda_1$,$\lambda_2$), Kaplan-Yorke dimension ($d_L$), and average limit of predictability ($Pr_{max}$, hours) for time series of $NO_2$,$SO_2$ at sites of Gdansk reg. (Jan.-Dec. 2003)

| | Site 6 (Sopot) $NO_2$ | Site 6 (Sopot) $SO_2$ | Site 9 (Gdynia) $NO_2$ | Site 9 (Gdynia) $SO_2$ |
|---|---|---|---|---|
| $\lambda_1$ | 0.0184 | 0.0164 | 0.0189 | 0.0150 |
| $\lambda_2$ | 0.0061 | 0.0066 | 0.0052 | 0.0052 |
| $d_L$ | 4.11 | 5.01 | 3.85 | 4.60 |
| $Pr_{max}$ | 40 | 43 | 41 | 49 |

Table 5. Coefficient correlation (*r*) between actual data and 24-hour forecast subject to *NN* for last 100 points of time series of $NO_2$ and $SO_2$ at the sites of Gdansk reg. during Jan.-Dec. 2003

| | Site 6 (Sopot) $NO_2$ | | | Site 6 (Sopot) $SO_2$ | | | Site 9 (Gdynia) $NO_2$ | | | Site 9 (Gdynia) $SO_2$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *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 |

**List of literature**


1. *Sivakumar B.* Chaos theory in geophysics: past, present and future // Chaos, Solitons & Fractals.-2004.-Vol.19(2).-P.441-462.
2. *Kennel M.B., Brown R., Abarbanel H.* Determining embedding dimension for phase-space reconstruction using a geometrical construction//*Phys.Review* A.-1992.-Vol.45.- 3403-3411.
3. *Bunyakova Yu.Ya, Glushkov A.V.* , Analysis and forecast of the impact of anthropogenic factors on air industrial goroda.-Odessa: Ecology.-2010.-256C.
4. *Khokhlov V.N., Glushkov A.V., Loboda N.S., Bunyakova Yu.Ya.* Short-range forecast of atmospheric pollutants using non-linear prediction method// *Atmospheric Environment* (Elsevier).-2008.-Vol. 42.-P.7774-7788.
5. *Lanfredi M., Macchiato M.*, Searching for low dimensionality in air pollution time series// *Europhys.Lett.*-1997.-Vol.97.-P.589-594.
6. *Koçak K., Şaylan L., Şen O.,* Nonlinear time series prediction of $O_3$ concentration in Istanbul//Atmospheric Environment (Elsevier).-2000.-Vol.34.-P.1267-1271.
7. *Gallager R.G.*, Information theory and reliable communication.-N-Y.: Wiley.-1986.-320P.
8. *Grassberger P., Procaccia I.,* Measuring the strangeness of strange attractors//Physica D.-1983.-Vol.9.-P.189–208.
9. *Sivakumar B.* Chaos theory in geophysics: past, present and future // Chaos, Solitons & Fractals.-2004.-Vol.19(2).-P.441-462.
10. *Chelani A.B.* Predicting chaotic time series of PM10 concentration using artificial neural network // Int. J. Environ. Stud.-2005.-Vol.62(2).-P.181-191.
11. *Gottwald G.A., Melbourne I.* A new test for chaos in deterministic systems// Proc. Roy. Soc. London. Ser. A. Mathemat. Phys. Sci.-2004.-Vol.460.-P.603-611.
12. *Packard N.H., Crutchfield J.P., Farmer J.D., Shaw R.S.* Geometry from a time series// Phys. Rev. Lett. – 1980. – Vol. 45. – P. 712-716.
13. *Abarbanel H.D.I., Brown R., Sidorowich J.J., Tsimring L.Sh.* The analysis of observed chaotic data in physical systems // Rev. Mod. Phys.-1993.-Vol.65.- P.1331-1392.
14. *Mañé R.* On the dimensions of the compact invariant sets of certain non-linear maps// Dynamical systems and turbulence, Warwick 1980. Lecture Notes in Mathematics No. 898 / D.A. Rand, L.S. Young (Eds.). – Berlin: Springer, 1981. – P. 230-242.

15. *Takens F.* Detecting strange attractors in turbulence // Dynamical systems and turbulence, Warwick 1980. Lecture Notes in Mathematics No. 898 / D.A. Rand, L.S. Young (Eds.). – Berlin: Springer, 1981. – P. 366-381.
16. *Fraser A.M., Swinney H.* Independent coordinates for strange attractors from mutual information// Phys. Rev. A. – 1986. – Vol. 33. – P. 1134-1140.
17. *Schreiber T.* Interdisciplinary application of nonlinear time series methods // Phys. Rep. – 1999. – Vol. 308. – P. 1-64.
18. *Glushkov A.V., Khokhlov V.N., Tsenenko I.A.,* Atmospheric teleconnection patterns: wavelet analysis//*Nonlinear Processes in Geophysics*.-2004.-Vol.11(3).-P.285-293.
19. *Glushkov A.V., Khokhlov V.N., Serbov N.G., Bunyakova Y.Y., Balan A.K., Balanyuk E.P.* Low-dimensional chaos in time series of concentrations of pollutants in the atmosphere and hydrosphere // Herald Odessa Environment. Univ.-2007.-N4.-C.337-348
20. *Glushkov A.V., Svinarenko A.A., Buyadzhi V.V., Zaichko P.A., Ternovsky V.B.* Chaos-Geometric Attractor and Quantum Neural Networks Approach to Simulation Chaotic Evolutionary Dynamics During Perception Process// Advances in Neural Networks, Fuzzy Systems and Artificial Intelligence, Series: Recent Advances in Computer Engineering, Ed. J.Balicki.(Gdansk, WSEAS Pub.).-2014.-Vol.21.-P.143-150
21. *Glushkov A.V., Khetselius O.Y., Brusentseva S.V., Zaichko P.A., Ternovsky V.B.* Studying interaction dynamics of chaotic systems within a non-linear prediction method: application to neurophysiology// Advances in Neural Networks, Fuzzy Systems and Artificial Intelligence, Series: Recent Advances in Computer Engineering, Ed. J.Balicki.(Gdansk, WSEAS Pub.).-2014.-Vol.21.-P.69-75.
22. *Glushkov A.V., Loboda N.S., Khokhlov V.N.* Using meteorological data for reconstruction of annual runoff series over an ungauged area: Empirical orthogonal functions approach to Moldova-Southwest Ukraine region//Atmospheric Research (Elseiver).-2005.-Vol.77.-P.100-113.
23. *Glushkov A.V., Loboda N.S., Khokhlov V.N., Lovett L.* Using non-decimated wavelet decomposition to analyse time variations of North Atlantic Oscillation, eddy kinetic energy, and Ukrainian precipitation // Journal of Hydrology (Elseiver).-2006.-Vol. 322. N1-4.-P.14-24.

**Хаос-геометричний метод в короткостроковому прогнозі рівня забруднень гідро-і атмосфери: поліпшені оцінки**

**Глушков А.В., Сербов Н.Г., Шахман І.А., Балан Г.К., Мансарлійскій В.Ф ., Дудінов А.А.**

*В рамках хаос-геометричного підходу отримані поліпшені дані з аналізу та прогнозу хаотичних флуктуацій для часових рядів концентрацій діоксиду азоту та сірчистого ангідриду в атмосфері Гданського регіону.*

***Ключові слова:*** *забруднення гідро-і атмосфери, моделювання, хаос-геометричний підхід.*

**Хаос-геометрический метод в краткосрочном прогнозе уровня загрязенния гидро-и атмосферы: Улучшенные оценки**

**Глушков А.В., Сербов Н.Г., Шахман И.А., Балан А.К., Мансарлийский В.Ф., Дудинов А.А.**

*В рамках хаос-геометрического подхода получены улучшенные данные по анализу и прогнозу хаотических флуктуаций для временных рядов концентраций диоксида азота и сернистого ангидрида в атмосфере Гданьского региона.*

***Ключевые слова****: загрязнение гидро-и атмосферы, моделирование, хаос-геометрический подход.*