



A Study of the Dynamic Behaviour of Fine Particulate Matter in Santiago, Chile

Giovanni A. Salini^{1*}, Patricio Pérez²

¹ *Departamento de Matemática y Física Aplicadas, Universidad Católica de la Santísima Concepción, Concepción, Chile*

² *Departamento de Física, Universidad de Santiago de Chile, Santiago, Chile*

ABSTRACT

We present here a study about the limitations found when trying to develop an accurate atmospheric particulate matter forecasting model based on real data, and evidence that the time series of fine particulate matter concentration exhibit deterministic chaotic behavior. We have calculated the Lyapunov exponents of PM_{2.5} time series obtained from measurements from four monitoring stations located in the city of Santiago, Chile, in recent years. Values obtained for the largest Lyapunov exponents turned out to be positive and ranging between 0.3 and 0.5 which, according to the theory of chaos, is a condition for the presence of deterministic chaos and random behavior in time series. Given the shape of decay of autocorrelation functions and values of correlation dimension and Hurst exponents, random behavior can be discarded: we therefore conclude that the series are chaotic and very sensitive to initial conditions. The study presented here can be replicated in other mid-sized cities that present similar situations to the city of Santiago, where complexity of topography, meteorology and seasonal trends favor the generation of high concentration episodes of atmospheric particulate matter and where a reliable air quality forecasting model may be important for environmental management.

Keywords: Chaotic time series; Air quality forecasting; PM_{2.5}; Time series analysis.

INTRODUCTION

In urban centers around the world it is observed that on some days the population is exposed to unhealthy concentrations of atmospheric particulate matter. For this reason, it seems convenient that environmental authorities operate a forecasting system which can be used to prevent the population or to enforce emissions restrictions.

Atmospheric particulate matter concentration forecasting models may be classified as statistical or deterministic. Statistical models are based on historical information from the particulate matter (PM) time series and associated meteorology. During a training stage, using samples of an appropriate input-output relation, a number of adjustable parameters are calculated. During operational mode, values of pollutant concentration and meteorological variables measured within the last few hours are introduced in the algorithm that generates an estimation of the future value of the variable of interest.

Particulate matter forecasts using statistical models, in which historical values are used for parameter adjustment, may be of reasonable accuracy when estimation of

concentrations a few hours ahead in time is the goal. However with these models in general it is not possible to generate spatial distribution of pollutant concentrations. Besides, errors become considerably large when forecasting more than one day into the future (Pérez *et al.*, 2000; Ordieres *et al.*, 2005; Grivas and Chaloulakou, 2006). Since these models are trained with data from the past, they usually tend to be less accurate for extreme cases (very low or very high pollutant concentration) that are observed with a relative low frequency. From an operational point of view, poor long-range forecasting will be a disadvantage, because high particulate matter episodes are of crucial interest for air pollution management and they are normally observed on few occasions during the year. Among statistical models used for air pollution forecasting we find linear regressions (Thomas and Jacko, 2007), artificial neural networks (Pérez and Reyes, 2006), discriminant analysis (Silva *et al.*, 2001) and Kalman filtering (Van der Wal and Janssen, 2000).

On the other hand, deterministic models are chemistry-transport models (Grell *et al.*, 1994; Byun and Ching, 1999; Tie *et al.*, 2007; Jorba *et al.*, 2008; Stern *et al.*, 2008) for which, given the rate of emission of a variety of pollutants, meteorological fields and chemical boundary conditions, the time evolution of several chemical compounds and aerosols is calculated. In general, these models are composed of a meteorological driver (MD) and a chemical and transport module (CTM). Limitations on the forecasting accuracy of this type of model may be attributed to the difficulty to

* Corresponding author.

Tel.: 56-41-2345549; Fax: 56-2345001

E-mail address: gsalini@ucsc.cl

consider in detail all the relevant sources of air pollution and to take into account all the physical-chemical processes that have an effect on the evolution of particulate matter concentrations (Pool, 1989). Accuracy of forecasting is strongly dependent on the precise estimation of initial conditions (Manders *et al.*, 2009). Kononov *et al.* (2009) presented a study that combines the properties of a statistical model and a deterministic model, which seems to optimally combine the accuracy of the former with the capability of longer time forecasting of the later.

Average PM concentration forecasting errors with one day anticipation are not significantly lower than 20% with any model, whether statistical or deterministic. It is possible that intrinsic properties of the PM time series do not allow for forecasting accuracies much better than those obtained with existing models, due in part to the evidence that these series are chaotic in nature (Chelani and Devotta, 2006a; Kumar *et al.*, 2008; Lee and Lin, 2008).

According to Lorenz (1963) nonlinearity is inherent to atmospheric systems. Sivakumar *et al.* (2007) points out the existence of nonlinear deterministic behavior in the air quality index in time series such as ozone. Raga and Le Moine (1996) have applied nonlinear dynamic tool on air quality data (NO, CO, SO₂ and O₃).

In the decade of the 80's, the analysis of time series through dynamical systems theory (Takens, 1981) has established that relevant information may be obtained by the calculation of a set of parameters (or indexes) as time lag, mutual information (Fraser and Swinney, 1986), embedding dimension, capacity dimension, correlation dimension, entropy of information, Hurst exponent and Lyapunov exponents, among others (Grassberger and Procaccia, 1983).

Among these parameters, the largest Lyapunov exponent is one of the most widely used indicator to detect chaotic behavior in a dynamical system consisting of a time series (Eckmann *et al.*, 1986).

In this paper the PM_{2.5} data recorded in four monitoring stations located in the urban area of Santiago, Chile during 2000–2006 is analyzed. The significantly positive Lyapunov exponents obtained are an indication that the evolution of concentrations is very sensitive to initial conditions and the estimation of capacity dimension, correlation dimension and Hurst exponent indicate that they are not random but chaotic.

Chaotic Behavior in Air Pollution Time Series

Linear methods for analysis and prediction of time series may be a starting point in the study of behavior of air pollution and meteorological data. However, the most relevant results have been obtained with the introduction of nonlinear tools. The reason for this is that in this type of data, small causes may have large effects in future values, which is the essence of nonlinearity.

Let us consider the following time series:

$$x(t), x(t + \Delta), x(t + 2\Delta), x(t + 3\Delta), \dots \quad (1)$$

where Δ is the time difference between measurements of variable x . We can write this series in simpler form by

calling $x(t + k\Delta) = x_k$. A distance d_0 between two arbitrary samples of the series x_i and x_j may be defined as:

$$d_0 = \|x_i - x_j\| \quad (2)$$

This corresponds to absolute value of difference in one dimension and to Euclidean distance in higher dimensions.

The distance between the respective following values is called d_1 :

$$d_1 = \|x_{i+1} - x_{j+1}\| \quad (3)$$

If we assume that the distance $d_N = \|x_{i+N} - x_{j+N}\|$ increases exponentially with N , the exponent λ defined by Eq. (4) is called the Lyapunov exponent (Hilborn, 2000):

$$d_N = d_0 e^{\lambda N} \quad (4)$$

Taking natural logarithm:

$$\lambda = \frac{1}{N} \ln \frac{d_N}{d_0} \quad (5)$$

In practice, the Lyapunov exponent is obtained from Eq. (5) in the limit of large N , for which saturation of $\log(d_N)$ becomes evident (see Fig. 3). What is found is that there are as many different Lyapunov exponents as the embedding dimension D_e of the time series, where D_e is the integer dimension of phase-space required to reproduce the steady state of the system.

Nonlinear systems are described by multidimensional vectors labeled by a time coordinate. The space in which these vectors lie is known as state space or phase space. What is shown in Eqs. (2)–(5) corresponds to the case with $D_e = 1$. These equations describe the way to calculate the single Lyapunov exponent for this simple case. In the general case, points must be represented by a D_e -dimensional vector:

$$\{x(t), x(t + \Delta), x(t + 2\Delta), \dots, x(t + (D_e - 1)\Delta)\} \quad (6)$$

and distances d_N must be understood as Euclidian distances.

For a time series of scalar observations, which may be the case of the series of one hour average PM_{2.5} concentrations measured at a monitoring station, Pérez and Reyes (2001) have shown a methodology for calculating the embedding dimension and the optimal time lag Δ . In this case, the embedding dimension corresponds to the length of a sequence of observations used as input to a multilayer neural network which are necessary to predict the next step value, where the time separation between inputs is Δ .

Lyapunov exponents may be positive, zero or negative. If the maximal Lyapunov exponent is negative, we are in the presence of a time series representing a dissipative system with a stable fixed point. If the motion settles down onto a limit cycle, the maximal Lyapunov exponent is zero. If a predominantly deterministic system is affected by random noise, the maximal Lyapunov exponent is infinite (Kantz

and Schreiber, 2005). A positive finite maximal Lyapunov exponent would be an indication of chaos. For a given time series, the sum of all positive Lyapunov exponents defines its Kolmogorov entropy K (Kolmogorov, 1959), and its inverse quantifies the average time of predictability, $T_p = 1/K$. In general, the largest positive Lyapunov exponent of a system λ_L is obtained in a straightforward manner, and then we can establish that

$$T_p \leq \frac{1}{\lambda_L} \quad (7)$$

providing in this way a strong restriction for modeling (Liu *et al.*, 2004).

Lanfredi and Macchiato (1997) have found that random behavior dominates over the chaotic behavior in time series of pollutants like NO_x , O_3 and CO . Raga and Le Moine (1996) have applied nonlinear analysis to O_3 , NO_2 , SO_2 and CO time series obtained from 13 monitoring stations in Mexico City, finding low-dimensional chaotic behavior in those time. Related to total ozone concentration, Chattopadhyay and Chattopadhyay (2008) have found the existence of chaotic behavior in Arosa, using the correlation dimension method. On the other hand, Haase *et al.* (2002) have indicated that if there is any nonlinearity in the ozone time series it may be very weak. But different to that result, Chelani (2010) has observed chaotic signatures in ozone concentration time series in Delhi, India, for 1 h and 4 h sampling rates. Koçak *et al.* (2000) applied a local method to predict observed O_3 time series data in Istanbul. The attractor in a multidimensional space was reconstructed, concluding that O_3 concentration is governed by a deterministic chaotic system.

Chelani and Devotta (2006b) implemented a hybrid methodology in order to predict one step ahead air pollutant concentration time series using nonlinear dynamic modeling and applied to NO_2 concentrations observed at a site in Delhi that was more effective than an autoregressive model. Also, using neural networks and nonlinear dynamical systems theory, Chelani *et al.* (2005) have characterized and predicted ambient nitrogen dioxide time series concentration at three sites in Kolkata, India. At two sites in the Gdansk region in Poland, Khokhlov *et al.* (2008) have investigated chaotic behavior in the nitrogen dioxide and sulphurous anhydride concentration time series finding low-dimensional chaos (low-dimensional chaos is characterized by only one positive Lyapunov exponent). Using artificial neural networks and a new scheme to predict chaotic time series of air pollutant concentrations (applied to Lorenz map and ozone observations), Gautam *et al.* (2008) got better predictions as compared to standard backpropagation algorithm. Yu *et al.* (2011) have found clear evidence of chaotic behavior of PM_{10} time series over a time span of 10 years in Lanzhou, China.

An indication of the presence of chaos in a time series is the value of the Lyapunov exponent, for which we describe a calculation method below.

METHODOLOGY

In what follows, the largest Lyapunov exponents for $\text{PM}_{2.5}$ series are calculated using the algorithm proposed by Wolf *et al.* (1985). In order to use this algorithm, an estimation of the embedding dimension of the system is necessary. One way to calculate embedding dimension is by analyzing the correlation dimension (Grassberger and Procaccia, 1983) as a function of embedding dimension D_e . Correlation dimension is calculated as follows: for each data point generate a hyper dimensional sphere of embedding dimension D_e and radius r with the point at its center. The fraction of subsequent points within the sphere is calculated for different values of r . Then the log of this number versus the log of the radius is plotted. The correlation dimension is defined as the average slope of the cumulative curve. As the embedding dimension increases, the correlation dimension also increases, but for most systems it saturates at some value. The smallest integer value of D_e within the saturated region corresponds to the correct value of embedding dimension D_e .

Once we have estimated the embedding dimension, it is possible to construct the state space from the experimental data record (Small, 2005; Takens, 1981).

Let us consider

$$y(t) = x(t), x(t + \Delta), x(t + 2\Delta), \dots, x(t + (D_e - 1)\Delta) \quad (8)$$

where $y(t)$ is the reconstructed D_e dimensional state vector, $x(t)$ is the observed variable, τ is a time lag, calculated from the first minimum of the Average Mutual Information function (Fraser and Swinney, 1986; Abarbanel, 1996). This quantity is a set theoretic concept and as a function of time delay estimates the mutual information content of two sets of measurements.

It is worth mention that nearly random time series has also large Lyapunov exponents. However, in that case a signature is the linear autocorrelation function, defined as (Abarbanel, 1996):

$$C(T) = \sum_n [x(n) - \bar{x}] \cdot [x(n+T) - \bar{x}] \quad (9)$$

This correlogram is an image of correlation statistics and may be used to investigate if a sequence of data is random or if adjacent observations are related.

For a random series, $C(t)$ goes abruptly to zero, while for chaotic deterministic data it shows slower decay. The calculation of additional statistical parameters as capacity dimension, correlation dimension, Hurst exponent and entropy may be related to the determination of whether the analyzed time series are chaotic.

The capacity dimension is the fractional dimension of any attractor of the time series obtained when data are plotted in a reconstructed phase space with a given embedding dimension. It is calculated by iteratively dividing the phase space with embedding dimension D_e into equal hypercubes and plotting the log of the fraction of hypercubes containing data points against the log of the size of the edge of the hypercube. The average slope of the central part of the curve is taken as the capacity dimension. A dimension greater than five is an indication of essentially random data.

The correlation dimension is calculated by counting data points inside hyper-spheres of different radii centered on each data point in a reconstructed phase space with a given embedding dimension. It corresponds to the average slope of the cumulative curve obtained from the log of the record against the log of the radius. Also, a correlation dimension greater than about five implies that the time series represents data from a random process.

The Hurst exponent is a measure of the degree to which the data can be represented by a random walk. Here, the root-mean-square displacement is calculated as a function of time, where each point in the time series is taken as an initial condition. The slope of the curve obtained is the Hurst exponent. For ordinary Brownian motion the exponent is 0.5. Values between 0.5 and 1.0 indicate deterministic behavior and persistence. Hurst exponent significantly less than 0.5 are typical for deterministic behavior and anti-persistence (Kantz and Schreiber, 2005).

Chaos in Santiago $PM_{2.5}$ Time Series

$PM_{2.5}$ series in several monitoring stations in Santiago, Chile show annual averages that are high relative to national standards. In addition, a seasonal effect is observed, according to which concentrations are significantly higher during fall and winter as compared to the rest of the year. This seasonal increase in concentrations, results from a combination of unfavorable meteorology dominated by severe temperature inversion, the presence of urban emission sources, and a confined geographical area (Ruttland and Garreaud, 1995). This pattern of atmospheric pollution is not unique to Santiago, but is also observed in other cities around the world, like Lohan, Utah, USA (Silva *et al.*, 2007), Tucson, Arizona, USA (Bailey *et al.*, 2011), Graz, Austria (Stadlober *et al.*, 2008), Gothenburg, Sweden (Olofson *et al.*, 2000). When episodes of high concentrations occur in the cold period, harmful effects for the population may be prevented in part by the use of a forecasting model. However, when chaos is present in $PM_{2.5}$ time series, important limitations to the accuracy of the forecast are unavoidable.

The data from four $PM_{2.5}$ monitoring stations in Santiago is analyzed in this study. Their location within the urban area is shown in Fig. 1. Station L is 604 m over sea level and is located near a commercial area. An important contribution to $PM_{2.5}$ concentrations measured in this station may be attributed to traffic congestion produced at rush hours. Station M is 785 m over mean sea level (MSL) and it is located in a residential area. Fine particles measured at this station are expected to be produced by vehicle traffic and wood burning (in winter). This station registers $PM_{2.5}$ concentrations that in general are lower than those measured by the other three analyzed stations. A reason for this is that its altitude is often higher than altitude of thermal inversions during episodes (between 600 m and 700 m MSL). Station N is 550 m over sea level and is near down town and close to freeway. Station O is 481 m MSL. It is located in a residential area, and being located towards the extreme west and the lowest part of the city, it is expected to measure not only local emissions but also particles transported from other areas. During high concentration episodes this station

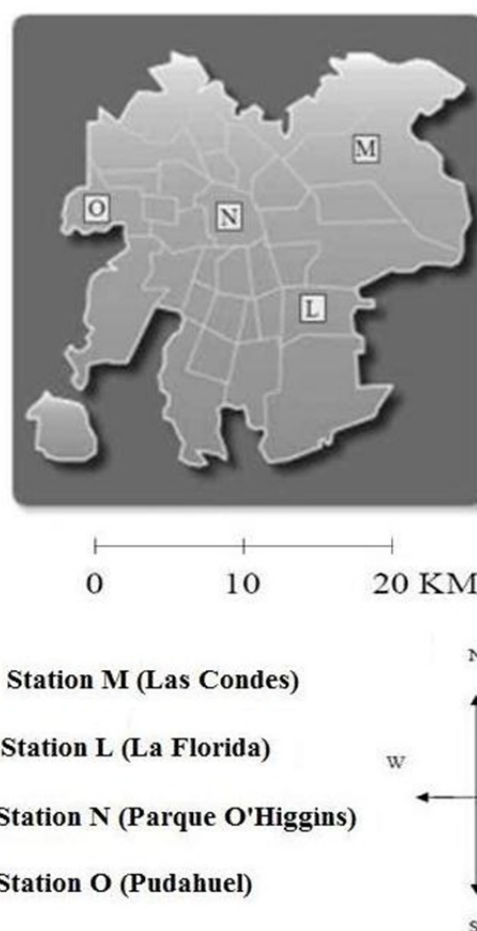


Fig. 1. Location of $PM_{2.5}$ monitoring stations in Santiago urban area.

registers the highest values. The accumulation of air pollution in this area is enhanced by the presence of a night breeze that blows from east to west, which is blocked by a coastal range with elevations over 1000 m that stands between the city and the Pacific Ocean. Given the high $PM_{2.5}$ concentrations observed in Santiago, environmental authorities have implemented control strategies, like restrictions to wood burning, vehicle circulation and closing of industries during episodes. We must have in mind that in urban areas an important fraction of the measured $PM_{2.5}$ is produced through chemical reactions that occur in the atmosphere (Wang *et al.*, 2008; Seguel *et al.*, 2009).

About 2% of our time series data are missing. For small gaps linear interpolation was applied and for bigger gaps, a cubic spline with single imputation technique was used (Junninen *et al.*, 2004).

One hour average $PM_{2.5}$ concentrations during one year are shown in Fig. 2. Seasonal variation is evident from data in the four stations. Being Santiago located in the southern hemisphere, autumn starts in March and winter ends in September. Average concentrations are significantly higher between April and August during which more episodes of high $PM_{2.5}$ are observed. During this period, a greater number of episodes are observed in station O as compared with the rest of the year (notice the difference of scale in

Fig. 2). Statistical properties of the $PM_{2.5}$ series for years from 2000 to 2006 are displayed in Tables 1 and 2. Table 1 shows statistics for the complete year and Table 2 for the period between April and August, which is the season with the worst air quality. From the period averages for all stations and all years it becomes evident the difference between the cold season and the rest of the year. By comparing maximum values in Tables 1 and 2 we can verify that most of them occur between April and August and that the highest values are observed in station O. Large standard deviations suggest that forecasting is not easy.

Skewness indicates the symmetry of the probability density

function (PDF) of a time series. A time series with an equal number of large and small values has a skewness of zero (Joanes and Gill, 1998). The positive values of skewness in Tables 1 and 2 is an indication that there are relatively fewer large $PM_{2.5}$ concentrations in all series. This will be inconvenient for the construction of precise statistical forecasting models because parameters in them are calculated from historical cases and a low representation of high concentration cases will imply in general a tendency to under-predict extreme values. From an operational point of view, for air quality management in a city, detection of high concentrations situations are particularly important.

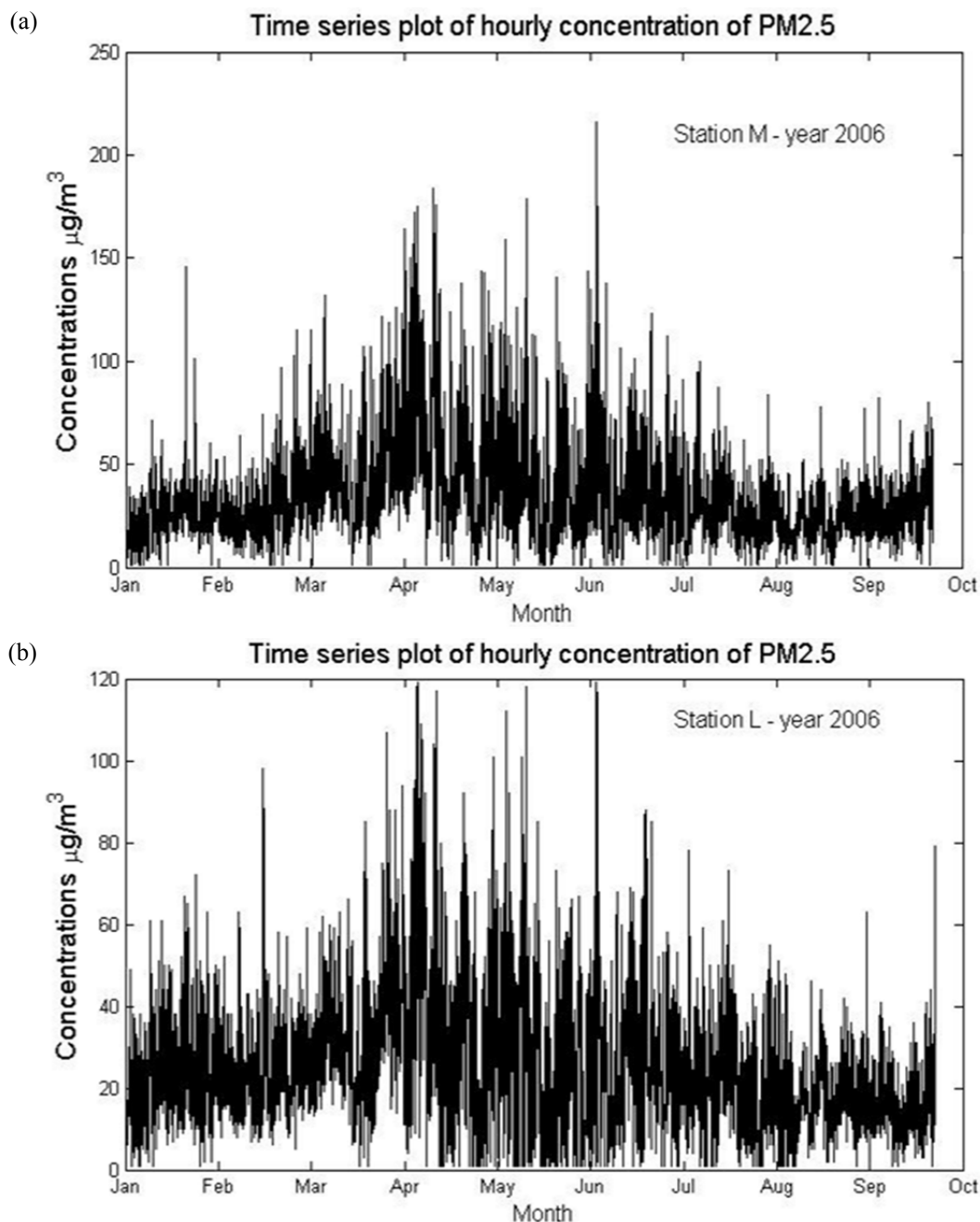


Fig. 2. One hour average of $PM_{2.5}$ concentrations measured in Santiago monitoring stations during 2006. (a) Station M, (b) station L, (c) station N, (d) station O.

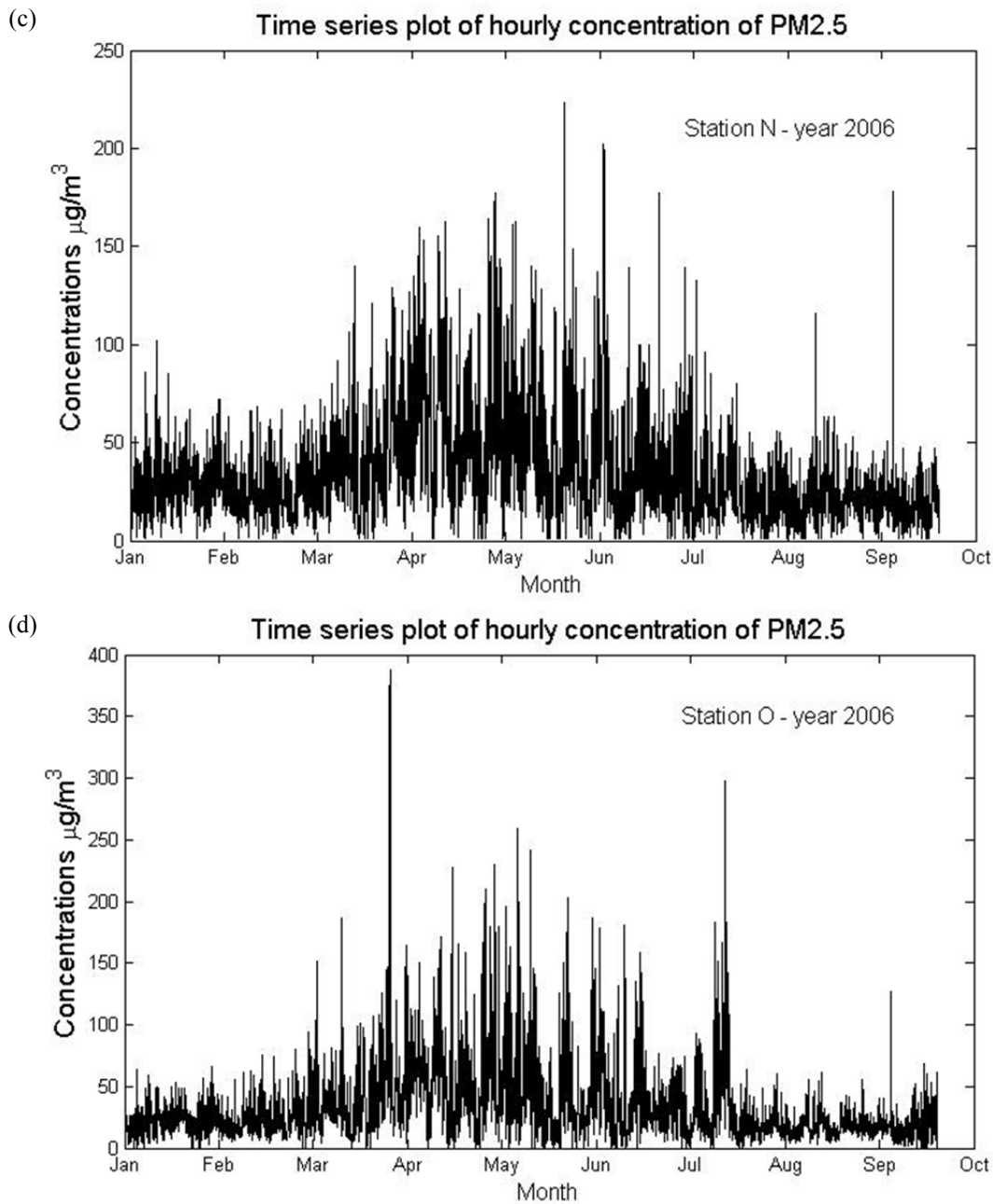


Fig. 2. (continued).

Average skewness for series including one year data is 2.06, while the average for data from the cold season decreases to 1.54, which is consistent with the observation that higher concentrations occur in the fall and winter periods.

Kurtosis is a measure of the peakedness of the PDF of a time series. A kurtosis value close to three indicates a Gaussian like peakedness. PDFs with relatively sharp peaks have kurtosis greater than three. PDFs with flat peaks have kurtosis less than three. From the values of this parameter obtained from Santiago data, we observe significant differences from one monitoring station to another. It is not unexpected that largest kurtosis corresponds to station O, where higher concentrations within the city are verified,

and where under unstable atmospheric conditions, the lowest concentrations are observed.

Since it is of interest to investigate the presence of chaos in PM_{2.5} series, we proceeded, as mentioned in the previous section, to calculate the Lyapunov exponent using the Wolff algorithm. As a previous step it was necessary to estimate the time lag τ and the embedding dimension D_e . The values obtained for our series of interest remain rather constant, and for all of them we estimated values $\tau = 11$ and $D_e = 5$.

Table 3 shows the largest Lyapunov exponents for PM_{2.5} time series (one hour spacing between data) at four monitoring stations in Santiago, between years 2000 and 2006. We can verify that these values are relatively large, and then we may conclude, based on Eq. (7), that predictability

Table 1. Statistics of PM_{2.5} series for years 2000–2006 as measured in four monitoring stations. Data from January through December.

PM _{2.5} Year	Number of data	Mean ($\mu\text{g}/\text{m}^3$)	Maximum Value (^a) ($\mu\text{g}/\text{m}^3$)	Standard deviation ($\mu\text{g}/\text{m}^3$)	Skewness ($\mu\text{g}/\text{m}^3$)	Kurtosis ($\mu\text{g}/\text{m}^3$)
Station M						
2000	8687	28.17	190	20.74	1.66	3.91
2001	8280	25.84	157	17.56	1.77	5.78
2002	8736	24.96	191	18.37	1.96	6.81
2003	8760	27.56	273	19.26	1.96	8.28
2004	8784	25.69	151	17.91	1.58	4.32
2005	8736	26.57	159	17.16	1.30	2.99
2006	8760	24.80	119	16.60	1.51	3.58
Station L						
2000	8737	35.55	182	24.91	1.46	2.65
2001	8760	34.81	224	22.73	1.82	5.59
2002	8736	32.27	316	24.20	1.94	6.68
2003	8736	34.27	231	25.48	1.72	4.29
2004	8760	36.39	209	26.23	1.55	3.60
2005	8760	33.58	208	23.82	1.55	3.45
2006	8760	34.51	216	23.91	1.86	5.08
Station N						
2000	8734	35.53	222	29.57	1.85	4.25
2001	8687	33.94	269	27.38	2.12	6.30
2002	8639	35.04	327	30.03	2.15	6.46
2003	8639	40.15	251	33.38	1.94	4.60
2004	8616	35.16	213	27.71	1.96	5.22
2005	8313	32.71	180	23.33	1.79	4.60
2006	8664	34.09	223	26.00	1.82	4.44
Station O						
2000	8759	33.27	319	31.06	2.66	9.65
2001	8736	33.20	328	32.25	2.89	11.82
2002	8664	34.26	425	37.15	3.30	15.72
2003	8759	36.46	351	35.06	2.85	11.77
2004	8712	33.66	434	33.01	3.41	18.69
2005	8760	30.68	287	25.91	2.63	10.81
2006	8664	33.03	388	30.37	2.85	13.69

(^a) The minimum value for all time series was $1 \mu\text{g}/\text{m}^3$.

based on information contained in the series is not greater than 3 hours. The average values for all the stations are similar except for station O, which are somewhat smaller. The reason for this may be attributed to the fact that this station is located in an area with the lowest altitude in the city, where concentrations tend to have extreme values, very low under conditions of unstable atmosphere and very high under stable atmosphere, so their values may be easier to predict under specific conditions. One must take into account that this time limit may be increased if we incorporate information from relevant exogenous variables (meteorological for example, which however may also show chaotic behavior). Calculated Lyapunov exponents are between 0.274 and 0.504. These values are of the same order of magnitude as those obtained for PM₁₀ series in Mexico (Vasquez *et al.*, 2012) and Taiwan (Lee and Lin, 2008). The Lyapunov exponents in Table 3 were calculated using an embedding dimension $D_e = 5$. For the same embedding dimension, the Lyapunov exponent for PM₁₀ series in Taipei was 0.277 and they found a correlation

dimension of 4.32. Both of these values are indications of chaos. Lyapunov exponents for PM₁₀ series in Mexico city were between 0.4 and 0.6, correlation dimension between 3.6 and 4.4 and a Hurst exponent between 0.2 and 0.25. The values of these three parameters are consistent with attribution of chaotic behavior to the PM₁₀ time series. Lyapunov exponents for PM₁₀ time series in Lanzhou are positive but much smaller (Yu *et al.*, 2011), which we believe is due to an erroneous estimation of the time scale.

In our case we can test the exponential divergence by constructing five dimensional vectors, where components are five successive measurements in PM_{2.5} series.

For an arbitrary measurement at time t we would have:

$$\mathbf{X}_t = (x(t), x(t + \Delta), x(t + 2\Delta), x(t + 3\Delta), x(t + 4\Delta)) \quad (10)$$

In order to study the evolution of state vectors we calculated the Euclidian distance between points separated by time $n\Delta$. We selected at random points for which the distance to $\mathbf{X}_{t+\Delta}$ was smaller or of the order of $10 \mu\text{g}/\text{m}^3$

Table 2. Statistics of PM_{2.5} series for years 2000–2006 as measured in four monitoring stations. Data from April through August.

PM _{2.5} Year	Number of data	Mean ($\mu\text{g}/\text{m}^3$)	Maximum Value (^b) ($\mu\text{g}/\text{m}^3$)	Standard deviation ($\mu\text{g}/\text{m}^3$)	Skewness ($\mu\text{g}/\text{m}^3$)	Kurtosis ($\mu\text{g}/\text{m}^3$)
Station M						
2000	3676	33.08	190	25.75	1.19	1.16
2001	3676	28.84	157	21.34	1.50	3.89
2002	3676	28.61	170	24.26	1.41	2.62
2003	3676	31.19	145	23.29	1.20	1.90
2004	3676	29.68	151	22.26	1.16	2.00
2005	3676	27.68	159	21.53	1.12	1.59
2006	3676	28.63	119	20.57	1.12	1.60
Station L						
2000	3676	43.75	182	30.23	0.95	0.79
2001	3676	42.29	224	29.18	1.29	2.51
2002	3676	40.92	192	30.54	1.11	1.29
2003	3676	44.99	231	30.75	1.17	1.84
2004	3676	45.70	209	32.00	1.11	1.60
2005	3676	38.90	208	29.01	1.24	1.78
2006	3676	43.35	216	29.62	1.32	2.36
Station N						
2000	3676	35.53	222	29.57	1.85	4.25
2001	3676	33.94	269	27.38	2.12	6.30
2002	3676	35.04	327	30.03	2.15	6.46
2003	3676	40.15	251	33.38	1.94	4.60
2004	3676	35.16	213	27.71	1.96	5.22
2005	3676	32.71	180	23.33	1.79	4.60
2006	3676	34.09	223	26.00	1.82	4.44
Station O						
2000	3676	44.58	319	40.13	1.87	4.47
2001	3676	45.84	328	42.56	1.98	5.35
2002	3676	48.66	425	49.58	2.30	7.40
2003	3676	50.89	351	44.91	1.948	5.46
2004	3676	47.26	434	43.39	2.46	9.81
2005	3676	37.05	287	32.87	2.08	6.50
2006	3676	45.11	298	36.77	1.67	3.74

(^b) The minimum value for all time series was $1 \mu\text{g}/\text{m}^3$

and then proceeded to calculate the Euclidian distance to successive five dimensional vectors. Fig. 3 shows the logarithm of the distance as a function of time interval for different starting point. The fact that most of the curves show a linear increase may be seen as a verification of exponential divergence that in turn tells us the presence of chaos.

Correlograms provide a useful method of visualizing the spatial dependence between data points in relation to distance (Fortin and Dale, 2005). Furthermore, the correlogram is a tool for checking randomness, rising or declining trend, oscillation, etc., of a time series. If a time series is random, its autocorrelation function should be near zero for all times greater than zero, and if non-random, the autocorrelation function will decay slowly to zero. The correlogram for PM_{2.5} series during 2005 is shown in Fig. 4, where significant correlation between measurements close in time is observed. Curves for the other years are not significantly different. This has to do with the presence of a daily cycle, where in general, higher concentrations are observed at times where

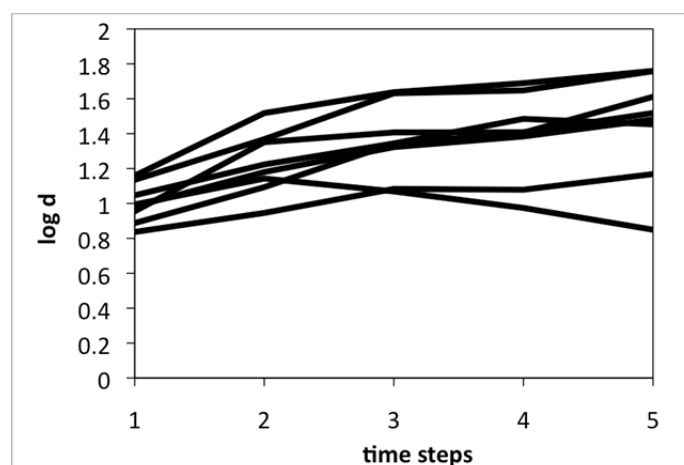
anthropogenic activity is high and meteorological conditions for pollutant dispersion are unfavorable (around 8 a.m. in the morning and 8 p.m. in the evening). Table 4 shows the result of the calculation of some additional indexes for the case of station N, now for years 2000 through 2008. We observe that for all years capacity dimension and correlation dimension are less than five, which is an indication of non random behavior. Hurst exponent is for all years, except 2001, significantly smaller than 0.5, which is an indication of deterministic behavior and antipersistence (predominance of oscillation between high and low values).

The values of the parameters calculated for station N and shown in Table 4 confirm the conclusion that PM_{2.5} series are chaotic. Values for stations L, M, and O are the same order of magnitude and are not shown.

Given the values of the largest Lyapunov exponents displayed in Table 3, the shape of the autocorrelation function shown in Fig. 4 and the values of the additional time parameter characterization presented in Table 4, we can conclude that the analyzed PM_{2.5} time series are chaotic.

Table 3. Lyapunov exponents for PM_{2.5} annual and winter time series in four monitoring station in Santiago, Chile.

	STATION			
	M	N	L	O
Annual				
2000	0.428	0.504	0.425	0.351
2001	0.482	0.484	0.399	0.325
2002	0.413	0.394	0.340	0.276
2003	0.365	0.472	0.429	0.351
2004	0.482	0.472	0.456	0.274
2005	0.462	0.503	0.487	0.365
2006	0.481	0.478	0.456	0.275
Mean	0.445	0.472	0.427	0.317
Winter				
2000	0.440	0.471	0.447	0.412
2001	0.453	0.452	0.423	0.398
2002	0.436	0.455	0.417	0.383
2003	0.455	0.452	0.465	0.411
2004	0.442	0.427	0.466	0.370
2005	0.414	0.443	0.452	0.389
2006	0.439	0.431	0.464	0.361
Mean	0.44	0.447	0.448	0.389

**Fig. 3.** Logarithm of Euclidian distance between successive five dimensional vectors constructed from 2007 PM_{2.5} time series in station O.

DISCUSSION

The evidence that the time series of atmospheric concentrations of particulate matter from four monitoring stations in Santiago, Chile, exhibit deterministic chaotic behavior imposes severe restrictions about the predictability of air quality in the city. Although a positive finite maximal Lyapunov exponent is not by itself an unambiguous indication of chaos. In our case, the results of calculation of additional statistical parameters confirm this behavior, which means that the series are deterministic and very sensitive to initial conditions. Statistical forecasting models, which constitute a pragmatic approach, adjusting parameters on the basis of historical data, would appear as a reliable tool for operational forecasting of particulate matter concentrations in urban regions only for a few hours into the future, provided relevant meteorological information is

included as input (Klingner and Sähn, 2008; Pérez and Salini, 2008; Stadobler *et al.*, 2008; Stern *et al.*, 2008). Deterministic photo-chemical forecasting models (Kononov *et al.*, 2009) would require detailed consideration of topography, emission sources and atmospheric processes. Any spatial or temporal averaging oriented to simplify the analysis may be critical if they leave out conditions that are relevant for estimating future patterns. A combination of deterministic and statistical approaches may be an option for more accurate results and longer time forecasting (Kononov *et al.*, 2009). For high concentration episodes, which in general are of great interest for air pollution management, special effort must be put towards overcoming shortcomings of prognostic weather prediction models for the simulation of stagnant weather situations. The results of this study show that particulate matter forecasting in a large city like Santiago will exhibit intrinsic limitations on

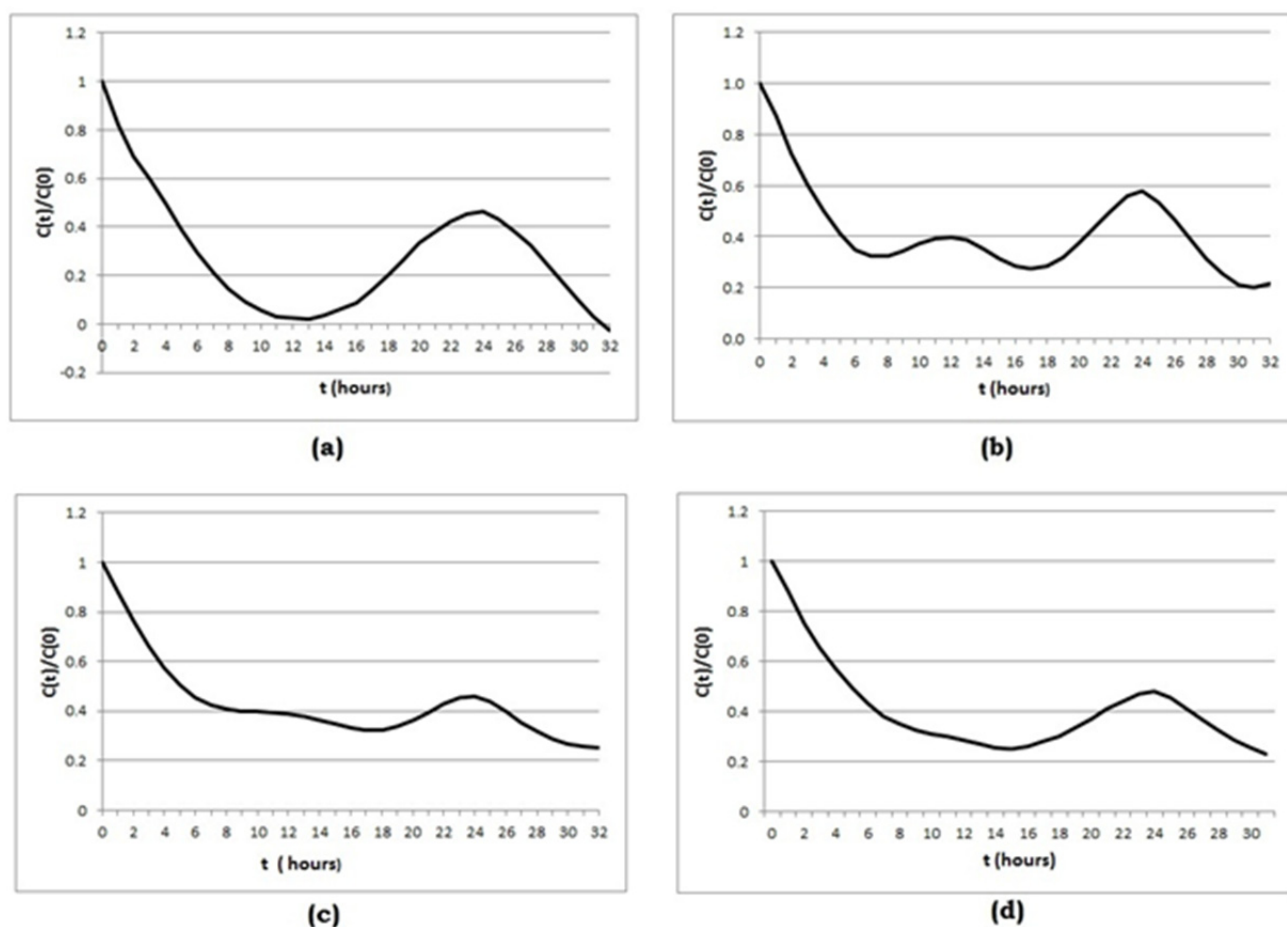


Fig. 4. Correlogram for 2005 $PM_{2.5}$ time series obtained from 4 stations in Santiago. (a) Station M. (b) Station L. (c) Station N. (d) Station O.

Table 4. Statistics of $PM_{2.5}$ series for years 2000–2006 as measured in four monitoring stations. Data from April through August are considered.

Year	Capacity Dimension	Correlation Dimension	Embedding Dimension D_e	Hurst exponent H	Entropy (1/hours)	Maximum Prediction (hours)
2000	2.355	3.455	5	0.265	0.525	1.900
2001	2.115	3.601	5	0.584	0.559	1.789
2002	2.285	3.520	5	0.257	0.489	2.045
2003	2.503	3.671	5	0.258	0.544	1.840
2004	2.476	3.776	5	0.247	0.416	2.400
2005	2.170	3.744	5	0.260	0.510	1.960
2006	2.202	3.824	5	0.241	0.488	2.050
2007	2.194	3.943	5	0.258	0.421	2.380
2008	2.644	3.641	5	0.185	0.405	2.470

accuracy. However, this conclusion may not discourage us from the necessary task of modeling in order to provide the environmental authorities with a tool that nevertheless will help to protect the health of citizens. The situation described here may apply in other places such as mid-sized cities, where the complexity of topography, multiple emission sources, seasonal trends and meteorology impose limitations to the estimation of future pollutant concentrations.

ACKNOWLEDGMENTS

We would like to thank the support from Facultad de Ingeniería, Universidad Católica de la Sma. Concepción, and the support from Universidad de Santiago de Chile through Dicyt Project number 091331PJ.

REFERENCES

- Abarbanel, H.D.I. (1996). *Analysis of Observed Chaotic Data*, Springer Verlag, New York.
 Bailey, A., Chase, T. Cassano, J. and Noone, D. (2011).

- Changing Temperature Inversion Characteristics in the U.S. Southwest and Relationship to Large-Scale Atmospheric Circulation. *J. Appl. Meteorol. Climatol.* 50: 1307–1323, doi: 10.1175/2011JAMC2584.1.
- Byun, D.W. and Ching, J.K.S. (Eds.) (1999). *Science Algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) Modeling System*, (US EPA Report No. EPA/600/R-99/030, Office of research and Development, Washington, DC.)
- Chattopadhyay, G. and Chattopadhyay, S. (2008). A Probe into the Chaotic Nature of Total Ozone Time Series by Correlation Dimension Method. *Soft Comput.* 12: 1007–1012, doi: 10.1007/S00500-007-267-7.
- Chelani, A.B., Singh, R.N. and Devotta, S. (2005). Nonlinear Dynamical Characterization and Prediction of Ambient Nitrogen Dioxide Concentration. *Water Air Soil Pollut.* 166: 121–138.
- Chelani, A.B. and Devotta, S. (2006a). Nonlinear Analysis and Prediction of Coarse Particulate Matter Concentration in Ambient Air. *J. Air Waste Manage. Assoc.* 56: 78–84, doi: 10.1080/10473289.2006.10464432.
- Chelani, A.B. and Devotta, S. (2006b). Air Quality Forecasting Using a Hybrid Autoregressive and Nonlinear Model. *Atmos. Environ.* 40: 1774–1780, doi: 10.1016/j.atmosenv.2005.11.019.
- Chelani, A.B. (2010). Nonlinear Dynamical Analysis of Ground Level Ozone Concentrations at Different Temporal Scales. *Atmos. Environ.* 44: 4318–4324, doi: 10.1016/j.atmosenv.2010.07.028.
- Eckmann, J.P., Oliffson, S., Kamphorst, S., Ruelle, D. and Ciliberto, C. (1986). Liapunov Exponents from Time Series. *Phys. Rev. A* 34: 4971–4979.
- Fortin, F.J. and Dale, M.R.T. (2005). *Spatial Analysis: A Guide for Ecologist*, Cambridge University Press, Cambridge.
- Fraser, A.M. and Swinney, H.L. (1986). Independent Coordinates for Strange Attractors from Mutual Information. *Phys. Rev. A* 33: 1134–1140.
- Gautam, A.K., Chelani, A.B., Jain, V.K. and Devotta, S. (2008). A New Scheme to Predict Chaotic Time Series of Air Pollutant Concentrations Using Artificial Neural Network and Nearest Neighbor Searching. *Atmos. Environ.* 42: 4409–4417, doi: 10.1016/j.atmosenv.2008.01.005.
- Grassberger, P. and Procaccia, L. (1983). Characterization of Strange Attractors. *Phys. Rev. Lett.* 50: 346–349.
- Grell, G., Dudhia, J. and Stauffer, D. (1994). *A Description of the Fifth Generation Penn State/NCA Mesoscale Model (MM5)*. (NCAR Technical Note, CAR/TN 398+ STR.
- Grivas, G. and Chaloulakou, A. (2006). Artificial Neural Network Models for Prediction of PM₁₀ Hourly Concentrations in the Greater Area of Athens, Greece. *Atmos. Environ.* 40, 1216–1229, doi: 10.1016/j.atmosenv.2005.10.036.
- Haase, P., Schlink, U. and Richter, M. (2002). Critical Reconsideration of Phase Space Embedding and Local Non-Parametric Prediction of Ozone Time Series. *Water Air Soil Pollut.* 2: 513–524.
- Hilborn, R. (2000). *Chaos and Nonlinear Dynamics: An Introduction for Scientists and Engineers*, Second Edition, Oxford University Press Inc., New York.
- Joanes, D.N. and Gill, C.A. (1998). Comparing Measures of Sample Skewness and Kurtosis. *J. R. Stat. Soc. Ser. D* 47: 183–189.
- Jorba, O., Loridan, T., Jiménez-Guerrero, P., Pérez, C. and Baldasano, J.M. (2008). Linkage the Advanced Research WRF Meteorological Model with the CHIMERE Chemistry-transport Model. *Environ. Modell. Softw.* 23: 1092–1094, doi: 10.1016/j.envsoft.2008.02.002.
- Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J. and Kolehmainen, M. (2004). Methods of Imputation of Missing Values in Air Quality Data Sets. *Atmos. Environ.* 38: 2895–2907, doi: 10.1016/j.atmosenv.2004.02.026.
- Kantz, H. and T. Schreiber (2004). *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge.
- Khokhlov, V.N., Glushkov, A.V., Loboda, N.S. and Bunyakova, Y.Y. (2008). Short-range Forecast of Atmospheric Pollutants Using Non-linear Prediction Method. *Atmos. Environ.* 42: 7284–7292, doi: 10.1016/j.atmosenv.2008.08.023.
- Klingner, M. and Sähn, E. (2008). Prediction of PM₁₀ Concentration on the Basis of High Resolution Weather Forecasting. *Meteorol. Z.* 17: 263–272.
- Koçak, K., Şaylan, L. and Şen, O. (2000). Nonlinear Time Series Prediction of O₃ Concentration in Istanbul. *Atmos. Environ.* 34: 1267–1271, doi: 10.1016/S1352-2310(99)00323-4.
- Kolmogorov, A.N. (1959). Entropy per Unit Time as a Metric Invariant of Automorphisms. *Dokl. Akad. Nauk SSSR* 124: 754–755.
- Konovalov, I.B., Beekmann, M., Meleux, F., Dutot, A. and Foret, G. (2009). Combining Deterministic and Statistical Approaches for PM₁₀ Forecasting in Europe. *Atmos. Environ.* 43: 6425–6434, doi: 10.1016/j.atmosenv.2009.06.039.
- Kumar, U., Prakash, A. and Jain, V.K. (2008). Characterization of Chaos in Air Pollutants: A Volterra-Wiener-Korenberg Series and Numerical Titration Approach. *Atmos. Environ.* 42: 1537–1551, doi: 10.1016/j.atmosenv.2007.10.076.
- Lanfredi, M. and Macchiato, M. (1997). Searching for Low Dimensionality in Air Pollution Time Series. *Europhys. Lett.* 40: 589–594.
- Lee, C.K. and Lin, S.C. (2008). Chaos in Air Pollutant Concentration (APC) Time Series. *Aerosol Air Qual. Res.* 8: 381–391, doi: 10.4209/aaqr.2008.09.0039.
- Liu, J.Q., Zhang, T.Q. and Yu, S.K. (2004). Chaotic Phenomenon and the Maximum Predictable Time Scale of Observation Series of Urban Hourly Water Consumption. *J. Zhejiang Univ. Sci.* 5: 1053–1059.
- Lorenz, E.N. (1963). Deterministic Nonperiodic Flow. *J. Atmos. Sci.* 20: 130–141.
- Manders, A.M.M., Schaap, M. and Hoogerbrugge, R.E. (2009). Testing the Capability of the Chemistry Transport Model Lotos-Euros to Forecast PM₁₀ Levels in the Netherlands. *Atmos. Environ.* 43: 4050–4059, doi: 10.1016/j.atmosenv.2009.05.006.
- Olofson, K., Andersson, P., Hallquist, M., Ljungström, E.,

- Tang, L., Chen, D. and Petteersson, J. (2009). Urban Aerosol Evolution and Particle Formation during Wintertime Temperature Inversions. *Atmos. Environ.* 43: 340–346, doi:10.1016/j.atmosenv.2008.09.080.
- Ordieres, J.B., Vergara, E.R., Capuz, R.S. and Salazar, R.E. (2005). Neural Network Prediction Model for Fine Particulate Matter (PM_{2.5}) on the US-Mexico Border in El Paso (Texas) and Ciudad Juarez (Chihuahua). *Environ. Modell. Softw.* 20, 547–559, doi: 10.1016/j.envsoft.2004.03.010.
- Pérez, P., Trier, A. and Reyes, J. (2000). Prediction of PM_{2.5} Concentrations Several Hours in Advance Using Neural Networks in Santiago, Chile. *Atmos. Environ.* 34: 1189–1196, doi: 10.1016/S1352-2310(99)00316-7.
- Pérez, P. and Reyes, J. (2001). Prediction of Particulate Air Pollution Using Neural Techniques. *Neural Comput. Appl.* 10: 165–171, doi: 10.1007/s005210170008.
- Pérez, P. and Reyes, J. (2006). An Integrated Neural Network Model for PM₁₀ Forecasting. *Atmos. Environ.* 40: 2845–2851, doi: 10.1016/j.atmosenv.2006.01.010.
- Pérez, P. and Salini, G.A. (2008). PM_{2.5} Forecasting in a Large City: Comparison of Three Methods. *Atmos. Environ.* 42: 8219–8224, doi: 10.1016/j.atmosenv.2008.07.035.
- Pool, R. (1998). Is Something Strange about the Weather? *Science* 243: 1290–1293.
- Raga, G.B. and Le Moyne, L. (1996). On the Nature of Air Pollution Dynamics in Mexico City—I. Nonlinear Analysis. *Atmos. Environ.* 30: 3987–3993, doi: 10.1016/1352-2310(96)00122-7.
- Rutllant, J. and Garreaud, R. (1995). Meteorological Air Pollution Potential for Santiago, Chile: Towards an Objective Episode Forecasting. *Environ. Monit. Assess.* 34: 223–244, doi: 10.1007/BF00554796.
- Seguel, R., Morales, R. and Leiva, M. (2009). Estimations of Primary and Secondary Organic Carbon Formation in PM_{2.5} Aerosols of Santiago City, Chile. *Atmos. Environ.* 43: 2125–2131, doi: 10.1016/j.atmosenv.2009.01.029.
- Silva, C., Pérez, P. and Trier, A. (2001). Statistical Modelling and Prediction of Atmospheric Pollution by Particulate Material: Two Nonparametric Approaches. *Environmetrics* 12: 147–159, doi: 10.1002/1099-095X(200103)12:2<147::AID-ENV451>3.0.CO;2-3.
- Silva, P., Vawdrey, E., Corbett, M. and Erupe, M. (2007). Fine Particle Concentrations and Composition during Wintertime Inversions in Logan, Utah, USA. *Atmos. Environ.* 41: 5410–5422, doi: 10.1016/j.atmosenv.2007.02.016.
- Sivakumar, B., Wallender, W., Horwath, W. and Mitchell, J. (2007). Nonlinear Deterministic Analysis of Air Pollution Dynamics in a Rural and Agricultural Setting. *Adv. Complex Syst.* 10: 581–597, doi: 10.1142/S0219525907001288.
- Small, M. (2005). *Applied Nonlinear Time Series Analysis*, Series A, Vol. 52, World Scientific Pub. Co. Pte. Ltd. Series on Nonlinear Science.
- Stadlober, E., Hörmann, S. and Pfeiler, B. (2008). Quality and Performance of a PM₁₀ daily Forecasting Model. *Atmos. Environ.* 42: 1098–1109, doi: 10.1016/j.atmosenv.2007.10.073.
- Stern, R., Builtjes, P.M., Schaap, R., Timmermans, R., Vautard, A., Hodzic, M., Memmesheimer, M., Feldmann, H., Renner, E., Wolke, R. and Kerschbaumer, A. (2008). A Model Intercomparison Study Focusing on Episodes with Elevated PM₁₀ Concentrations. *Atmos. Environ.* 42: 4567–4588, doi: 10.1016/j.atmosenv.2008.01.068.
- Takens, F. (1981). Detecting Strange Attractors in Turbulence, In *Dynamical Systems and Turbulence, Lecture Notes in Mathematics*, Rand, D.A. and Young, L.S. (Eds.), Vol. 898. Springer-Verlag: Berlin Heidelberg, Germany.
- Thomas, S. and Jacko, R.B. (2007). Model for Forecasting Expressway Fine Particulate Matter and Carbon Monoxide Concentrations: Application of Regression and Neural Network Models. *J. Air Waste Manage. Assoc.* 57: 480–488, doi: 10.3155/1047-3289.57.4.480.
- Tie, X., Madronich, S., Li, G., Ying, Z., Zhang, R., García, A.R., Lee-Taylor, J. and Liu, Y. (2007). Characterization of Chemical Oxidants in Mexico City: A Regional Chemical Dynamical model (WRF-Model) Study. *Atmos. Environ.* 41: 1989–2208, doi: 10.1016/j.atmosenv.2006.10.053.
- Van der Wal and J.T., Janssen, L.H.J.M. (2000). Analysis of Spatial and Temporal Variations of PM₁₀ Concentrations in the Netherlands Using Kalman Filtering. *Atmos. Environ.* 34: 3675–3687, doi: 10.1016/S1352-2310(00)00085-6.
- Vásquez, C., Beristain, F. and Audelo, E. (2012). Characterization of the Behavioral Pattern of Suspended Particles in the Valley of Mexico. *Inf. Tecnol.* 23: 3–10, doi: 10.4067/S0718-07642012000300002.
- Wang, W.C., Chen, K.S., Chen, S.J., Lin, C.C., Tsai, J.H., Lai, C.H. and Wang, S.K. (2008). Characteristics and Receptor Modeling of Atmospheric PM₁₀ and PM_{2.5} at Urban and Rural Sites in Pingtung, Taiwan. *Aerosol Air Qual. Res.* 8: 112–129, doi: 10.420/aaqr.2007.09.0039.
- Wolf, A., Swift, J.B., Swinney, H.L. and Vastano, J.A. (1985). Determining Lyapunov Exponents from a Time Series. *Physica D* 16: 285–317.
- Yu, B., Huang, C., Liu, Z., Wang, H. and Wang, I. (2011). A Chaotic Analysis on Air Pollution Index over the Past Ten Years in Lanzhou, Norwest China. *Stoch. Environ. Res. Risk Assess.* 25: 643–653, doi: 10.1007/s00477-011-0471-y.

Received for review, October 23, 2013

Revised, July 29, 2014

Accepted, September 26, 2014