Chapter 5

PM_{10} Time Series Analysis Through Geostatistical Techniques

Claudia Cappello, Sabrina Maggio,

Daniela Pellegrino and Donato Posa

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/60115

1. Introduction

Particulate matter (PM) is an air pollutant comes from vehicular traffic, industrial activities and street dust, or from the atmosphere, by transformation of the gaseous emissions. In recent years the interest in the health effects of this pollutant have increased, since high concentration levels in urban area have been measured.

Several studies suggest an association between fine particulate air pollution and the increase of the mortality rate [1]. In particular, PM up to 10 micrometers in size (PM_{10}) could cause negative health effects such as respiratory illness or cardiovascular problems. Hence, the analysis of temporal evolution of this pollutant could be useful in decision-making process for environmental policy.

Typically, in time series analysis, the Box-Jenkins methodology is widely applied and the autocorrelation function (ACF) is used as a standard exploratory tool to identify the model structure [3, 4]. In this context, the use of geostatistical techniques could also be convenient, nevertheless these techniques are usually applied to analyze, through the variogram, spatial relationships among sample data measured at some locations in a domain and to predict the corresponding spatial phenomena [6, 18, 22, 29]. In particular, the variogram could represent a complementary exploratory tool for assessing stationarity in time series [2, 19] and it has the considerable advantage that it is defined in much wider circumstances than the autocovariance and the autocorrelation. Moreover, this analytical tool is appropriate to identify trends and periodicity exhibited by the data and to obtain kriging predictions of the variable under study, either for temporal intervals with missing values (interpolation mode) and in time points after the last available data (extrapolation mode).

Different studies have suggested the use of geostatistical methods in time domain [7, 19]. In particular, De Iaco et al. [12] illustrated the role of variogram in this context for different purposes.

The aim of this paper is to analyze PM_{10} air pollution in an area of South Italy characterized by high levels of industrial emissions and vehicular traffic, through geostatistical techniques.

Thus, after a brief review on stochastic processes and geostatistical methods in time series analysis, the temporal evolution of PM_{10} daily concentrations, for the period 2010-2013 has been assessed. After the identification of trend and periodicity, the reconstruction of the analyzed time series by estimation of missing values has been discussed, and predictions of PM_{10} daily concentrations at some unsampled points have been produced. Moreover, the probability distributions of the variable under study have been estimated for future time points.

For interpolation and prediction purposes, a modified version of *GSLib* kriging routine has been used.

2. Theoretical framework

In time series analysis the observed values of a variable for different time points or intervals can be reasonably considered as a finite realization of a real-valued random process, denoted with $\{X_t, t \in T \subseteq \mathbb{R}\}$.

Besides the common second-order moments used to describe the random process $\{X_t, t \in T\}$, such as the autocovariance function and the autocorrelation function, the variogram can also be considered and even preferred with respect to covariance function [10, 19].

Given a stochastic process $\{X_t, t \in T\}$ over a temporal domain $T \subseteq \mathbb{R}$, the corresponding variogram is defined as follows

$$\gamma(t, t+h_t) = 0.5Var\left[X_t - X_{t+h_t}\right], \qquad t, t+h_t \in T. \tag{1}$$

Note that a function $\gamma(\cdot)$ is a variogram if and only if it is conditionally strictly negative definite [23].

As known in the literature [3–5], time series analysis is based on the theory of stationary processes. It is worth highlighting that the second-order stationarity implies the intrinsic stationarity, but the converse is not true [18, 22].

In particular, the stochastic process $\{X_t, t \in T\}$ is intrinsically stationary if its variogram $\gamma(t, t+h_t)$ depends solely on the temporal lag h_t and the expected value of the difference $(X_t - X_{t+h_t})$ is constant.

The variogram, widely used in geostatistical context, could be applied efficiently in time series analysis [14, 15], since

- it can describe a wider class of stochastic processes, i.e. the class of intrinsic stochastic processes, which includes the class of second-order stationary stochastic processes,

http://dx.doi.org/10.5772/60115

- its estimation does not require the knowledge of the expected value of the associated stochastic process,
- it is appropriate to identify trend and periodicity exhibited by data,
- it can be used for prediction purposes.

Regarding this last aspect, geostatistical techniques provide different parametric and nonparametric prediction methods, among these the sample and ordinary kriging, the universal kriging and the indicator kriging. Further details can be found in the specialized literature [7, 12, 19]. Thus, the estimation of the unknown value x_t of the stochastic process $\{X_t, t \in T\}$, using the data observed in the past (extrapolation mode), or the data observed before and after the time point t (interpolation mode) can be easily supported by geostatistical tools.

In the following, the ordinary kriging method and the indicator kriging approach are briefly reviewed, since these geostatistical tools are used for analyzing the variable under study.

Let $\widehat{X}_t$ the linear predictor of the intrinsic stationary process $\{X_t, t \in T\}$:

$$\widehat{X}_t = \sum_{i=1}^{n} \lambda_i(t) X_{t_i}, \tag{2}$$

where $\lambda_i(t)$, $i = 1, 2, \ldots, n$, are unknown real coefficients and X_{t_i} are random variables of the process X at the sampled time points t_i. The unknown weights $\lambda_i(t)$, $i = 1, 2, \ldots, n$, of (2) are obtained by solving the following kriging system

$$\begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1n} & -1 \\ \gamma_{21} & \cdots & \gamma_{2n} & -1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma_{n1} & \cdots & \gamma_{nn} & -1 \\ 1 & \ldots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \\ \vdots \\ \gamma_{n0} \\ 1 \end{bmatrix}, \tag{3}$$

where $\gamma_{ij} = 0.5\, Var(X_{t_i} - X_{t_j})$, $\gamma_{i0} = 0.5 Var(X_{t_i} - X_t)$, μ is the Lagrange multiplier. If γ is conditionally strictly negative definite, then the above system presents one and only one solution.

The ordinary kriging [22] requires only the knowledge of the variogram model and it is used when the expected value of the process is constant and unknown. Since the kriging system can be expressed in terms of the variogram, as in (3), the kriging predictor can be used even when the stochastic process under study satisfies the intrinsic hypothesis. Moreover, using a predictor based on a variogram, rather than on a covariance, avoids the estimation of the expected value, if this last is unknown.

The usefulness of geostatistical techniques in time series analysis can be appreciated through nonparametric estimation of the variable under study.

The kriging approach, based on the knowledge of variogram, leads naturally to nonparametric estimation [17]. Indicator kriging is a nonparametric approach to estimate the posterior cumulative distribution function (c.d.f.) of the variable under study at an unsampled point [16, 25, 26].

In this context, given the observed time series $x_{t_i},\ t_i,\ i = 1, 2, ..., n$, the conditional probability $Prob\{X_t \leq x|\mathcal{H}_n\}$, with $\mathcal{H}_n = \{x_{t_i},\ t_i,\ i = 1, 2, ..., n\}$, is interpreted as conditional expectation of an indicator random field $I(t; x)$ [27], that is

$$Prob\{X_t \leq x|\mathcal{H}_n\} = E[I(t; x|\mathcal{H}_n)]$$

where

$$I(t; x) = \begin{cases} 1, & if \ \ X_t \leq x \\ 0, & if \ \ X_t > x. \end{cases}$$

In the case study presented hereafter, ordinary kriging and indicator kriging are applied for interpolation and prediction purposes of an environmental variable. Note that a *GSLib* routine for kriging, named "KT3DP" [12], has been used in order to define appropriate temporal search neighborhoods in presence of periodicity, since environmental time series, such as the ones for air pollution data, usually are characterized by a periodic behavior.
Hence, the use of periodic and nonperiodic variogram models have been proposed through two different approaches:

- the periodic component has been factored out using the moving average method [5] and nonperiodic variogram model has been fitted;
- the periodicity has been retained and described by a periodic variogram model.

3. PM_{10} time series

In the present case study, the analysis of daily concentrations of PM_{10} ($\mu g/m^3$), measured at one of the monitoring stations of Brindisi district during the period 2010-2013, has been conducted through geostatistical techniques.

These data have been collected by the Environmental Protection Agency of Apulian region (ARPA Puglia) which controls the air quality of urban, suburban, and industrialized areas of the region.

Note that PM_{10} monitoring stations are classified in the following three categories:

- traffic stations, located in areas with heavy traffic;
- industrial stations, located close to industrialized areas;
- background stations, located in peripheral areas.

The analyzed station, named "Torchiarolo" is located in the municipality of Torchiarolo (Brindisi district), as shown in Fig. 1. It is classified as industrial station, since it is strictly close to an industrial site, i.e. the thermoelectric power station "Enel-Federico II" in Cerano (Brindisi district).

http://dx.doi.org/10.5772/60115

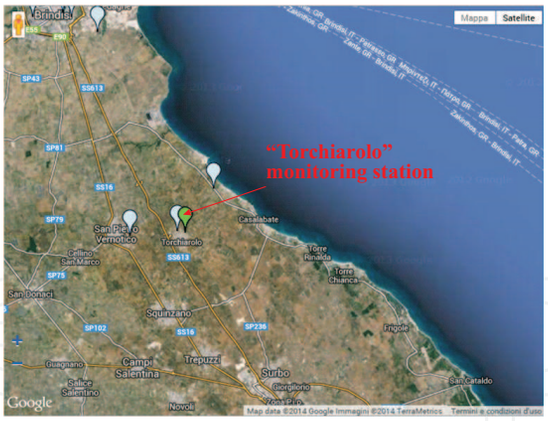


Figure 1. "Torchiarolo" monitoring station belonging to the Environmental Protection Agency of Apulian region (ARPA Puglia)

3.1. Exploratory Data Analysis

In order to assess the statistical properties of PM_{10} measured at the "Torchiarolo" station in the period 2010-2013, an exploratory data analysis has been performed. Some results are shown in Tab. 1.

Year	Min	Max	Mean	Standard Deviation	Number of exceedances
2010	8	114	35.10	20.09	67
2011	8	147	36.08	21.57	66
2012	10	108	32.83	17.01	49
2013	8	146	35.83	22.71	61
2010-2013	8	147	34.97	20.47	243

Table 1. Descriptive statistics of PM_{10} ($\mu g/m^3$), measured in the period 2010-2013 at the "Torchiarolo" monitoring station

According to the National Law concerning the human health protection, PM_{10} daily average concentrations cannot be greater than 50 $\mu g/m^3$ for more than 35 times per year. During the period under study, the PM_{10} daily values exceeded the threshold 243 times; in addition, the station has measured more than 35 exceedances per year.

The box plot in Fig. 2 shows that the observed time series is characterized by a seasonal component. During summertime, particle pollution shows lower levels compared to those recorded during wintertime; in particular, in summertime, PM_{10} doesn't exceed the limit value fixed by the National Law.
On the other hand, in wintertime changes in the lower layer of the troposphere determine PM_{10} stagnation. Hence, high levels of this pollutant are recorded.
In the following sections, the study of the temporal evolution of PM_{10} at the analyzed station has been conducted by performing

- structural analysis,
- estimation of some consecutive values assumed as missing,
- prediction of PM_{10} daily averages,
- estimation of the c.d.f. of PM_{10} daily concentrations at some unsampled time points.

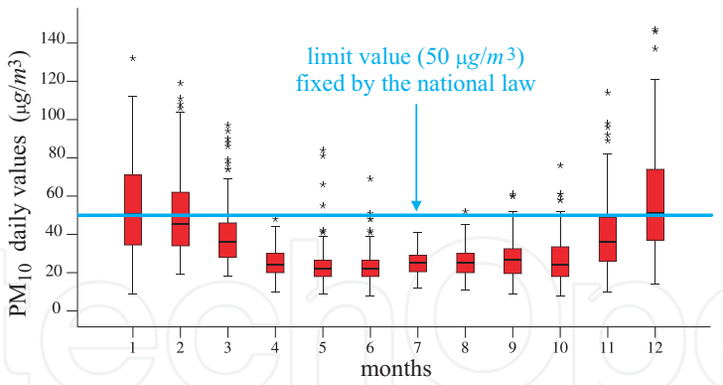


Figure 2. Box plot of PM_{10} daily concentrations, grouped by month, and limit value fixed by the National Decree (Decree-law 60/2002)

4. Structural Analysis

As previously pointed out, the variogram can describe a wider class of stochastic processes, that is the class of intrinsic stochastic processes and is usually preferred to the use of the covariance function.

In structural analysis, before modeling the temporal correlation described by the variogram, its estimation from data is required. The following classical estimator [9] is often used:

$$\hat{\gamma}(r_t) = \frac{1}{2\,|M(r_t)|} \sum_{M(r_t)} \left[X_{t+h_t} - X_t\right]^2 \tag{4}$$

where r_t is the temporal lag, $M(r_t) = \{t + h_t \in H \text{ and } t \in H, \text{ such that } \|r_t - h_t\| < \delta_t\}$, δ_t is the tolerance, H is the set of data at different time points (not necessarily equally-spaced) and $|M(r_t)|$ is the cardinality of this set.

In the present case study the variogram has been used as an exploratory tool to assess stationarity and periodicity. In particular, sample temporal variogram for PM_{10} daily observations, shown in Figure 3-a), reproduces the seasonal behavior of the variable under study, which presents an annual periodicity at 365 days. In equation (5) the analytic expression of the periodic variogram model, fitted to the sample variogram for the observed values, is proposed:

$$\gamma(h_t) = 265\,Exp(|h_t|;10) + 130\,Cos(|h_t|;365); \tag{5}$$

where $Exp(\cdot)$ and $Cos(\cdot)$ are the exponential and the cosine variogram models [29], respectively.

On the other hand, since the variable under study, is characterized by periodicity, this seasonal component could be factored out. Moving average and monthly averages techniques have been applied in order to obtain PM_{10} residuals. Note that the FORTRAN program "REMOVE" [11] has been used to apply moving average techniques.

http://dx.doi.org/10.5772/60115

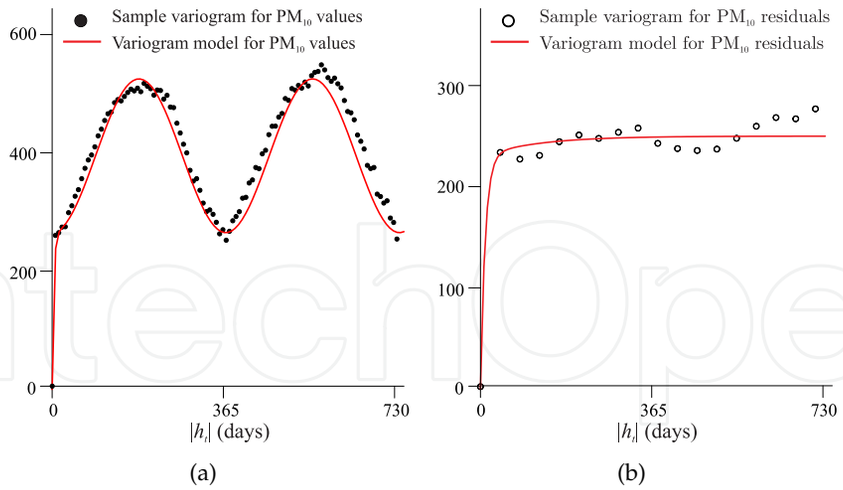


Figure 3. Sample temporal variograms and fitted models. (a) Variogram for PM_{10} daily concentrations (b) Variogram for PM_{10} residuals

The sample variogram of the residuals has been computed and modelled and the following nonperiodic variogram model has been chosen:

$$\tilde{\gamma}(h_t) = 348\, Exp(|h_t|;30) + 30\, Exp(|h_t|;365); \tag{6}$$

where $Exp(\cdot)$ is the exponential variogram model [8].

The sample temporal variogram for PM_{10} daily residuals and the corresponding nonperiodic fitted model (6) are illustrated in Fig. 3-b).

In both cases (original data and residuals), the behavior of the variogram functions near the origin is assumed to be linear with no nugget effect.

The goodness of variogram models (5) and (6) has been evaluated through cross-validation, which allows the estimation for PM_{10} daily concentrations and PM_{10} residuals, respectively, at all data points. Figure 4 shows the scatter plots of PM_{10} observed values (a) and PM_{10} residuals (b) towards the corresponding estimated values. The high values of the linear correlation coefficients (0.783 and 0.780, respectively) confirm the goodness of the above fitted models.

It is important to point out that the variogram model (5) has been validated using a modified version of the *GSLib* program "KT3D" [13], named "KT3DP". This program has been developed in order to properly define the neighborhood, i.e. the subset of available data used in the kriging system.

By taking into account the main features of the analyzed pollutant and its temporal behavior (periodicity at 365 days), the kriging routine has been modified and the value at time t is estimated by considering data observed

- at the two adjacent time points $(t-1)$ and $(t+1)$,
- at the same day of the year before and/or later, $(t-d)$ and $(t+d)$, with $d = 365$ and some days before and/or later, $(t-d\pm k)$ and $(t+d\pm k)$, with $k = 1, 2, 3$,

- at the same day of two years before and/or later, $(t - 2d)$ and $(t + 2d)$, with $d = 365$ and some days before and/or later, $(t - 2d \pm k)$ and $(t + 2d \pm k)$, with $k = 1, 2, 3$,

up to a maximum number of eight values.

The variogram model (6), which describes the temporal correlation for PM_{10} residuals, has been validated using the *GSLib* program "KT3D".

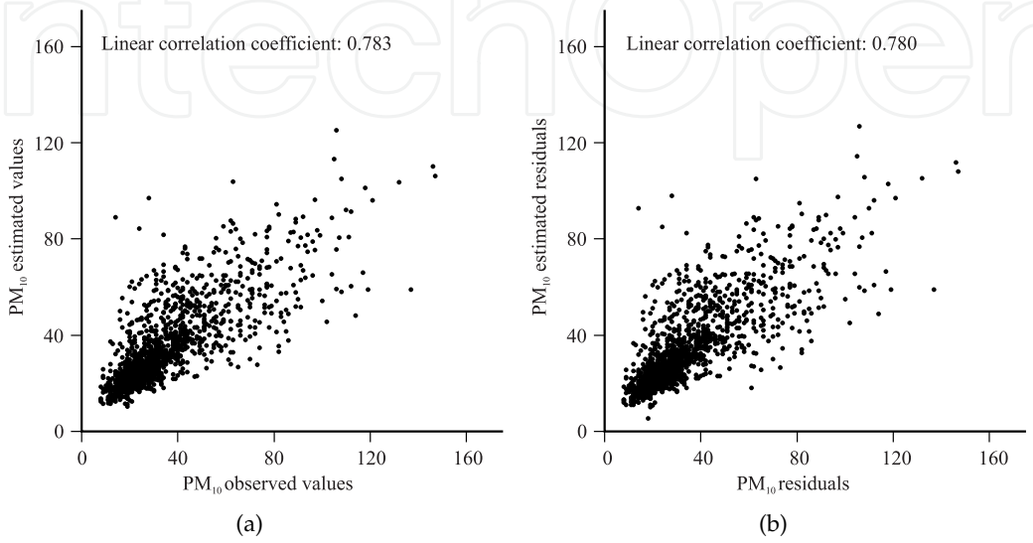


Figure 4. Scatter plots between observed and estimated values. (a) Diagram of PM_{10} daily concentrations towards the estimated ones (b) Diagram of PM_{10} residuals towards the estimated ones

5. Estimation of missing values

In this section the reconstruction of PM_{10}, by using the kriging technique, has been discussed [20, 21, 30, 31].

The reconstruction of temporal data is required if a time series is incomplete. This problem could be due to a malfunction of the monitoring station or the presence of invalid data.

With this aim, six consecutive PM_{10} values from the 12th to the 17th of June 2011, have been considered as missing, both for the observed time series with a 365-day periodic behavior and the deseasonalized values.

Kriging daily estimations for these missing values have been obtained using, alternatively

1. the periodic variogram model (5), which describes the temporal correlation for PM_{10} daily concentrations,
2. the nonperiodic variogram model (6), which describes the temporal correlation for PM_{10} daily residuals.

Since the time series of the observed values is characterized by a periodic behavior, *GSLib* routine "KT3DP", properly modified in order to define an appropriate neighborhood, has been used with the aim to estimate PM_{10} daily measurements.

http://dx.doi.org/10.5772/60115

On the other hand, for the deseasonalized time series, PM_{10} residuals have been estimated by the original version of "KT3D". Finally the periodic component, previously estimated by the moving average and monthly averages techniques, has been added to the estimated residuals, in order to obtain estimates of PM_{10} daily concentrations.

Time series of estimated missing values, obtained with the periodic variogram model (5) and the nonperiodic variogram model (6), are shown in Fig. 5, together with the time series of PM_{10} values, observed on June 2011.

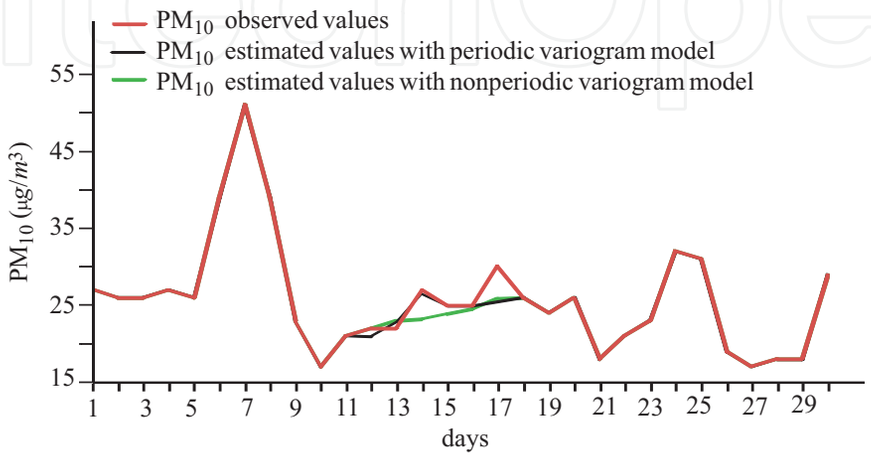


Figure 5. Time plot of PM_{10} estimated missing values and PM_{10} daily concentrations ($\mu g/m^3$), from the 12th to the 17th of June 2011

In order to test the validity of the estimation procedure, the linear correlation coefficients have been computed. In particular, the linear correlation coefficient between the PM_{10} observed values and the corresponding estimates, obtained with the periodic variogram model, is equal to 0.805. On the other hand, the linear correlation coefficient between the residuals and the corresponding estimates, obtained with the nonperiodic variogram model, is equal to 0.831. These results confirm the goodness of the kriging technique as estimator of missing values.

In Table 2 some results of estimation procedure are shown. Note that the mean value of the kriging standard error is lower if the periodic variogram model is used, compared with the kriging results based on the nonperiodic variogram model.

Therefore, the flexibility of kriging to reconstruct the time series has been demonstrated even when the periodic component is not factored out and the temporal correlation is described by a periodic variogram model.

6. Prediction of PM_{10} values

In this section, predictions for the variable under study in time points after the last available data are discussed [23, 24, 28].

The periodic variogram model (5) of PM_{10} concentrations and the nonperiodic variogram model (6) of PM_{10} residuals, have been used in order to predict six time points after the last available data, i.e. the 31st of December 2013. In particular, kriging predictions have

June 2011	PM_{10} value	PM_{10} Est. value[a]	Est. Error[a]	Est. value[b]	Est. Error[b]
12th	22	20.970	-1.030	22.029	0.029
13th	22	22.821	0.821	22.940	0.940
14th	27	26.437	-0.563	23.178	-3.822
15th	25	24.867	-0.133	23.878	-1.122
16th	25	24.926	-0.074	24.473	-0.527
17th	30	25.466	-4.534	25.825	-4.175
Mean values	21.167	24.248	-0.919	23.720	-1.446

[a] Results obtained by using the periodic variogram model (5) [b] Results obtained by using the nonperiodic variogram model (6)

Table 2. Kriging estimations of a sequence of 6 missing values, from the 12th to the 17th of June 2011 and corresponding errors for periodic and nonperiodic variogram models

been computed for the period ranging from the 1st to the 6th of January 2014, by using, alternatively

1. the available data, the variogram model (5) and the modified *GSLib* routine "KT3DP" which builds the searching neighborhood taking into account the periodicity exhibited by the data,
2. the deseasonalized PM_{10} observations, the variogram model (6) and the original *GSLib* routine "KT3D" which produces PM_{10} predicted residuals at which the diurnal component of the day before has been added to obtain predictions of PM_{10} daily concentrations.

In Fig. 6, the time series of PM_{10} daily concentrations measured from the 9th of December

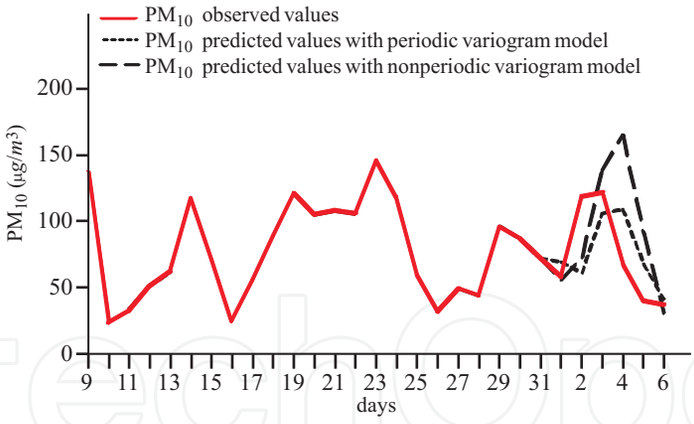


Figure 6. Time plot of PM_{10} predicted values and PM_{10} daily concentrations ($\mu g/m^3$), from the 1st to the 6th of January 2014

2013 to the 6th of January 2014 is shown together with the predicted PM_{10} values for the period ranging from the 1st to the 6th of January 2014. Note that the kriging procedure using the nonperiodic variogram model (6) related to PM_{10} residuals has produced overestimates of the pollution levels.

Moreover, in Table 3 some results of the performance of the prediction procedure are presented. The mean value of the kriging standard error is lower if the periodic variogram model is used, compared with the case of kriging based on the nonperiodic variogram model.

http://dx.doi.org/10.5772/60115

January 2014	PM_{10} Obs. value	PM_{10} Est. value[a]	Est. Error[a]	Est. value[b]	Est. Error[b]
1st	58	69.415	11.415	55.774	-2.226
2nd	119	61.031	-57.969	72.985	-46.015
3rd	122	106.051	-15.949	137.731	15.731
4th	67	108.397	41.397	164.909	97.909
5th	40	67.545	27.545	91.090	51.090
6th	37	41.498	4.498	30.738	-6.262
Mean values	73.833	75.656	1.823	92.205	18.371

[a] Results obtained by using the periodic variogram model (5) [b] Results obtained by using the nonperiodic variogram model (6)

Table 3. Kriging predictions of a sequence of six days, from the 1st to the 6th of January 2014 and corresponding errors for periodic and nonperiodic variogram models

It is important to highlight that in the period 1-5 January 2014 predicted values greater than 50 $\mu g/m^3$ (i.e. the limit value fixed by the National Law) have been obtained.

Note that in the period 1-4 January 2014, PM_{10} values greater than this threshold have been measured. On the other hand, at day 5th, the kriging procedure produces overestimate of the variable under study.

7. Estimation of the c.d.f.

For a given time series of PM_{10}, it might be useful to estimate the probability that the variable under study exceeds a fixed limit, so that appropriate and prompt solutions might be adopted if necessary.

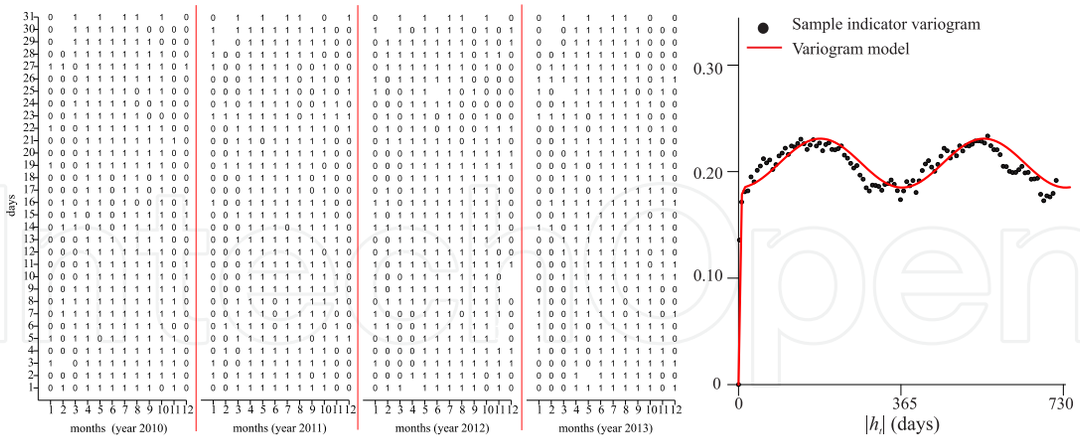
In this section, estimation of c.d.f. of PM_{10} daily concentrations ($\mu g/m^3$) has been conducted.

In particular, the c.d.f. of PM_{10} at unsampled time points has been estimated by indicator kriging [17].

Six threshold values for PM_{10} (22, 35, 50, 78, 98, and 108 $\mu g/m^3$) have been properly chosen, and six indicator variables according to the fixed thresholds have been defined as follows

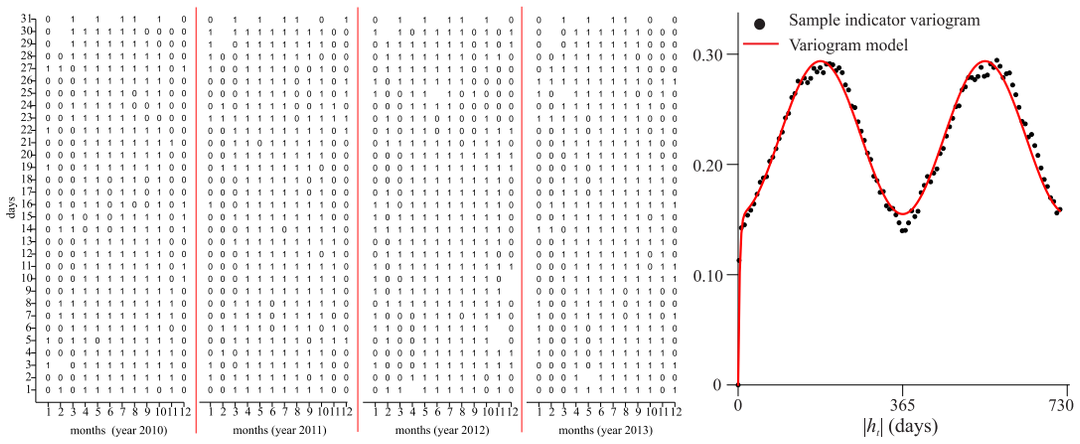
- $I_1(t;22) = \begin{cases} 1, & if \ \ PM_{10} \leq 22 \\ 0, & if \ \ otherwise \end{cases}$
- $I_2(t;35) = \begin{cases} 1, & if \ \ PM_{10} \leq 35 \\ 0, & if \ \ otherwise \end{cases}$
- $I_3(t;50) = \begin{cases} 1, & if \ \ PM_{10} \leq 50 \\ 0, & if \ \ otherwise \end{cases}$
- $I_4(t;78) = \begin{cases} 1, & if \ \ PM_{10} \leq 78 \\ 0, & if \ \ otherwise \end{cases}$
- $I_5(t;98) = \begin{cases} 1, & if \ \ PM_{10} \leq 98 \\ 0, & if \ \ otherwise \end{cases}$
- $I_6(t;108) = \begin{cases} 1, & if \ \ PM_{10} \leq 108 \\ 0, & if \ \ otherwise \end{cases}$

with $t \in T$. Note that indicator data are equal to 1 if the values of the variable under study are not greater than the considered threshold and they are equal to 0 otherwise. For each threshold, the temporal indicator variogram has been computed and modelled (Figs. 7, 8).



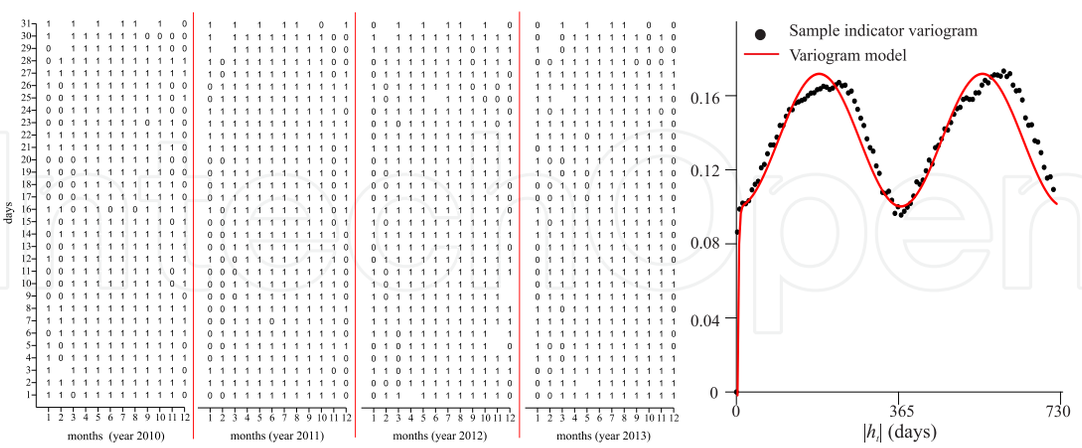
(a) $x_1 = 22\,\mu g/m^3$

(b) $x_1 = 22\,\mu g/m^3$



(c) $x_2 = 35\,\mu g/m^3$

(d) $x_2 = 35\,\mu g/m^3$

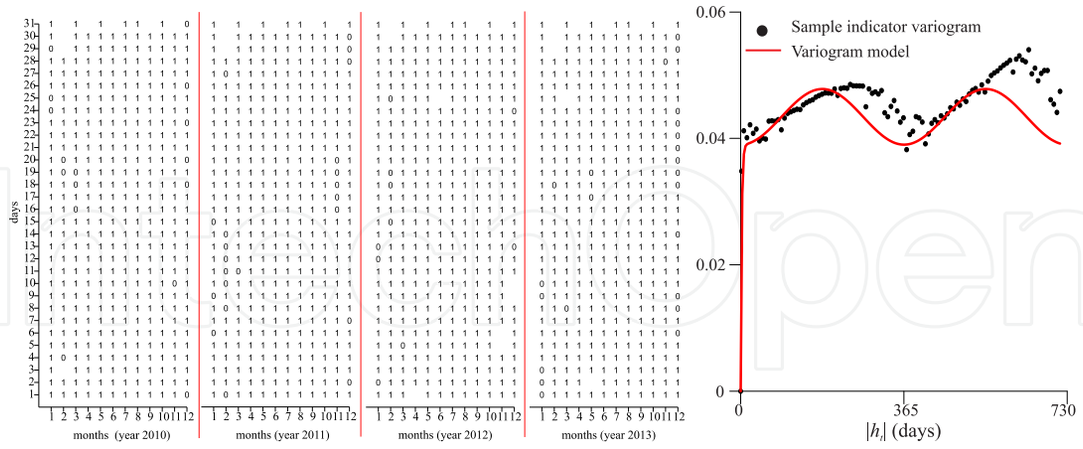


(e) $x_3 = 50\,\mu g/m^3$

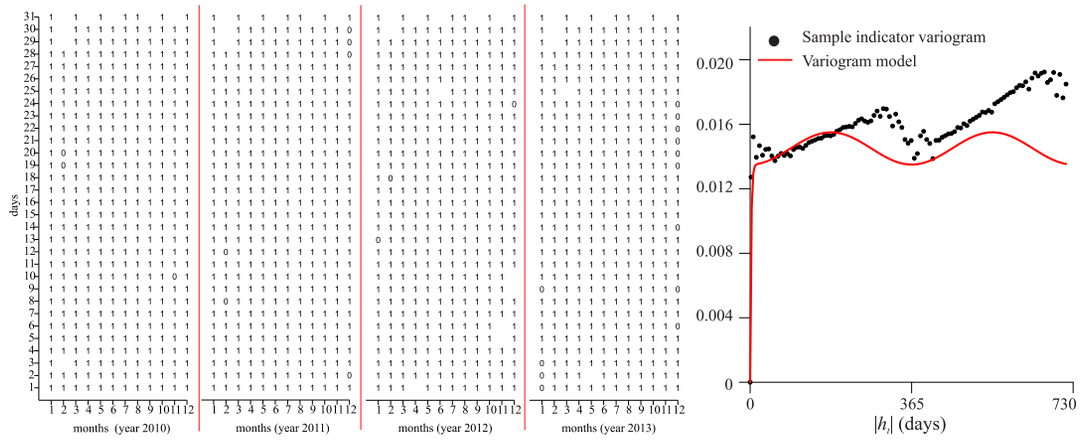
(f) $x_3 = 50\,\mu g/m^3$

Figure 7. Indicator maps of PM_{10} daily concentrations and their sample indicator variograms with the fitted models, for three threshold values. (a) Indicator map and (b) sample variogram indicator for the threshold $x_1 = 22\,\mu g/m^3$. (c) Indicator map and (d) sample variogram indicator for the threshold $x_2 = 35\,\mu g/m^3$. (e) Indicator map and (f) sample variogram indicator for the threshold $x_3 = 50\,\mu g/m^3$

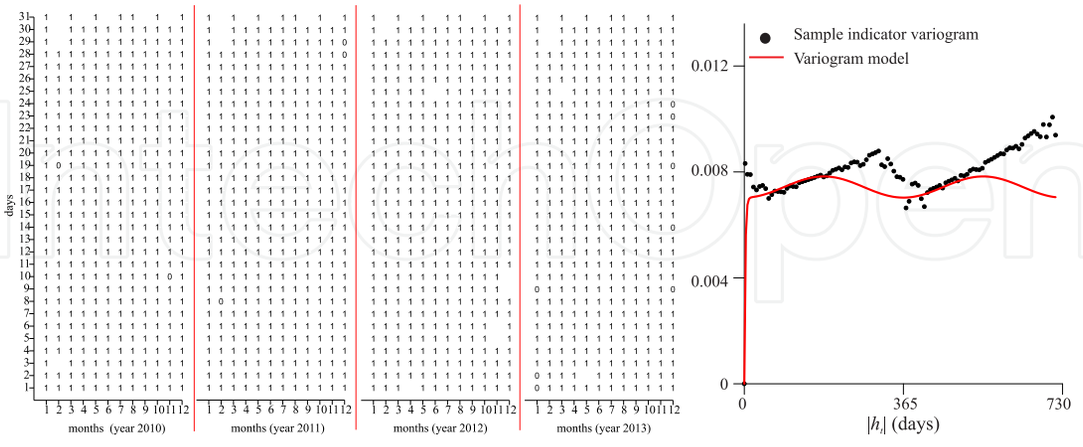
http://dx.doi.org/10.5772/60115



(a) $x_4 = 78\,\mu g/m^3$ (b) $x_4 = 78\,\mu g/m^3$



(c) $x_5 = 98\,\mu g/m^3$ (d) $x_5 = 98\,\mu g/m^3$



(e) $x_6 = 108\,\mu g/m^3$ (f) $x_6 = 108\,\mu g/m^3$

Figure 8. Indicator maps of PM_{10} daily concentrations and their sample indicator variograms with the fitted models, for three threshold values. (a) Indicator map and b) variogram for the threshold $x_4 = 78\,\mu g/m^3$. (c) Indicator map and (d) variogram for the threshold $x_5 = 98\,\mu g/m^3$. (e) Indicator map and (f) variogram for the threshold $x_6 = 108\,\mu g/m^3$

In particular the following models have been fitted

- $\gamma_{I_1}(h_t;22) = 0.185\,Exp(|h_t|;10) + 0.023\,Cos(|h_t|;365)$,
- $\gamma_{I_2}(h_t;35) = 0.15\,Exp(|h_t|;10) + 0.07\,Cos(|h_t|;365)$,
- $\gamma_{I_3}(h_t;50) = 0.102\,Exp(|h_t|;10) + 0.036\,Cos(|h_t|;365)$,
- $\gamma_{I_4}(h_t;78) = 0.039\,Exp(|h_t|;10) + 0.0004\,Cos(|h_t|;365)$,
- $\gamma_{I_5}(h_t;98) = 0.013\,Exp(|h_t|;10) + 0.001\,Cos(|h_t|;365)$,
- $\gamma_{I_6}(h_t;108) = 0.007\,Exp(|h_t|;10) + 0.0004\,Cos(|h_t|;365)$.

Thus, the c.d.f.s corresponding to six different unsampled time points, i.e. the days 1-6 of January 2014, have been estimated by using the "KT3DP" routine.

For each day of interest, the c.d.f. has been estimated by solving as many kriging systems as the number of threshold values considered. For each threshold, the corresponding indicator variogram model has been used for the kriging procedure.

Figure 9 shows the c.d.f.s estimated at days 1-6 of January 2014. It is clear that the probability of not exceeding a fixed threshold increases gradually from the 1st to the 6th of January 2014. For example, the estimated probability that PM_{10} concentrations, on the 1st of January 2014, do not exceed 22 $\mu g/m^3$ is lower than the estimated probability on the 3rd or the 6th of the considered month.

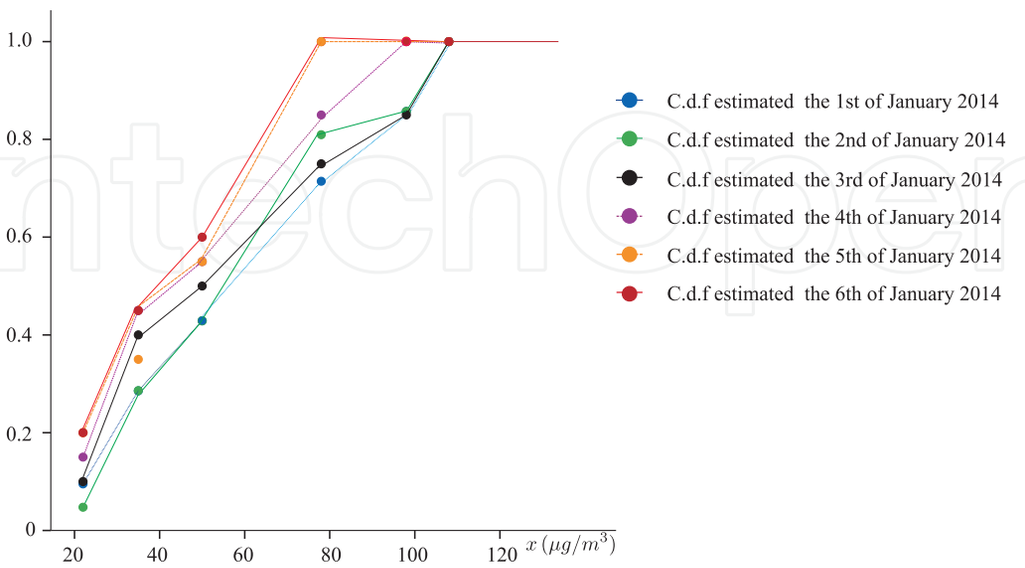


Figure 9. C.d.f.s estimated for PM_{10} daily concentrations ($\mu g/m^3$) at days 1-6 of January 2014

http://dx.doi.org/10.5772/60115

Thresholds	January 2014					
	1st	2nd	3rd	4th	5th	6th
22	0.095	0.047	0.100	0.150	0.200	0.200
35	0.285	0.286	0.400	0.450	0.350	0.450
50	0.429	0.429	0.500	0.550	0.550	0.600
78	0.714	0.810	0.750	0.850	0.900	0.900
98	0.857	0.857	0.850	0.900	0.950	1
108	1	1	1	1	1	1

Table 4. Estimated values for c.d.f. at days 1-6 of January 2014, for fixed thresholds

Moreover, note that it is almost sure that PM_{10} concentrations do not exceed the cutoff 78 $\mu g/m^3$ at days the 5th and 6th (Table 4).

The probability that the variable under study doesn't exceed the threshold fixed by the National Law (50 $\mu g/m^3$) is high (equal to 60%) for the last day of interest. In fact, the 6th of January is a non-working day (Epiphany) characterized by low traffic and low industrial emissions.

The local government could use these results in order to carry out environmental policies for the control of high levels of PM_{10}, since it is well known that high concentrations of this pollutant are dangerous for the human health.

Indeed, the estimation of the c.d.f. is a very powerful tool since any action of environmental protection might be adopted in advance by taking into account the actual likelihood of dangerous PM_{10} exceeding. For example, decisions about traffic limitation in high traffic urban area might be supported by the knowledge of the probability that a hazardous pollutant exceeds the level of attention.

8. Conclusions

In this paper, PM_{10} time series analysis, by using geostatistical techniques, has been discussed and the importance of appropriate tools of Geostatistics to study the temporal evolution of this environmental phenomena has been highlighted.

The seasonal behavior of PM_{10} levels has been evaluated through the variogram, that is the basic tool of Geostatistics. Moreover, estimation and prediction problems in the analysis of the time series of this pollutant, characterized by a periodic behavior, have been solved through kriging geostatistical techniques.

The computational aspects have been performed through the use of "KT3D" for the observed values and "KT3DP" for the residuals obtained after removing the periodic component.

Finally, the indicator approach and its capability for assessing the probability that PM_{10} exceeds the specific threshold values have been demonstrated.

The results obtained in this paper by applying geostatistical techniques to analyze PM_{10} time series could be useful to support national policies for environmental and health protection. Governments' activity must be oriented to control that the concentrations of the analyzed pollutant don't exceed specific thresholds according to national or international directives, since it has been demonstrated that particulate matter is dangerous for the human health.

Author details

Claudia Cappello, Sabrina Maggio*, Daniela Pellegrino and Donato Posa

*Address all correspondence to: sabrina.maggio@unisalento.it

University of Salento, Dept. of Management, Economics, Mathematics and Statistics, Italy

References

[1] Arden Pope III C., Dockery D. W. Health effects of fine particulate air pollution: lines that connect. Journal of Air and Waste Management Association 2006; 56 709–742.

[2] Bisgaard S., Khachatryan D. Asymptotic confidence intervals for variograms of stationary time series. Quality and Reliability Engineering International 2010; 26(No. 3) 259–265.

[3] Bloomfield P. Fourier analysis of time series: An Introduction. USA:Wiley & Sons Inc.; 2000.

[4] Box GEP., Jenkins GM. Time series analysis: Forecasting and Control. San Francisco:Holden Day; 1976.

[5] Brockwell P.J., Davis R.A. Time Series: Theory and Methods. New York: Springer; 1987.

[6] Chilés J.P., Delfiner P. Geostatistics: Modeling Spatial Uncertainty. New York: Wiley; 1999.

[7] Cressie, N. (1988). A graphical procedure for determining nonstationarity in time series, Journal of the American Statistical Association 1988; 83(No. 44) 1108–1116.

[8] Cressie, N. (1993). Statistics for Spatial Data, Wiley Series in Probability and Mathematical Statistics, Wiley, New York.

[9] Cressie N., Hawkins D.M. Robust estimation of the variogram. Mathematical Geology 1980;12(2) 115–125.

[10] Cressie N., Grondona M.O. A comparison of variogram estimation with covariogram estimation. In: Mardia K.V. (ed.) The Art of Statistical Science. Chichester: Wiley; 1992.

[11] De Cesare L., Myers D.E., Posa D. FORTRAN 77 programs for space-time modeling. Computers & Geosciences 2002;28(2) 205–212.

[12] De Iaco S., Palma M., Posa, D. Geostatistics and the role of variogram in time series analysis: a critical review. In: Montrone S. & Perchinunno P. (ed.) Statistical Methods for Spatial Planning and Monitoring. Italia: Springer-Verlag; 2013. p47–75.

[13] Deutsch C.V., Journel, A.G. GSLib: Geostatistical Software Library and User's Guide. New York: Oxford University Press; 1998.

[14] Gevers M. On the use of variograms for the prediction of time series. Systems & Control Letters 1985;6(1) 15–21.

[15] Haslett J. On the sample variogram and sample autocovariance for non-stationary time series. Statistician 1997;46(4) 475–485.

[16] Janis M.J., Robeson S.M. Determining the spatial representativeness of air-temperature records using variogram-nugget time series. Physical Geography 2004;25(6) 513–530.

[17] Journel A.G. Nonparametric estimation of spatial distributions. Mathematical Geology 1983;15(3) 445–468.

[18] Journel A.G., Huijbregts C.J. Mining Geostatistics. London: Academic; 1981.

[19] Khachatryan D., Bisgaard S. Some results on the variogram in time series analysis. Quality and Reliability Engineering International 2009;25 947–960.

[20] Little R.J.A., Rubin D.B. Statistical Analysis with Missing Data. New York: Wiley; 2002.

[21] Luceno A. Estimation of missing values in possibly partially nonstationary vector time series. Biometrika 1997;84(2) 495–499.

[22] Matheron G. Principles of Geostatistics. Economic Geology 1963;58(8) 1246–1266.

[23] Matheron G. Les variables régionalisées et leur estimation. Paris: Masson; 1965.

[24] Matheron G. The intrinsic random functions and their applications. Advances in Applied Probability 1973;5(3) 439–468.

[25] Myers D.E. Interpolation with positive definite functions. Science de la Terre 1988;28 251–265.

[26] Myers D.E. Kriging, cokriging, radial basis functions and the role of positive definiteness. Computers & Mathematics with Applications 1992;24(12) 139–148.

[27] Posa D. The indicator formalism in spatial data analysis. Journal of Applied Statistics 1992;19(1) 83–101.

[28] Posa D., Rossi M. Applying stationary and non-stationary kriging. Metron XLVII 1989;(1-4) 295–312.

[29] Posa D., De Iaco S. Geostatistica: teoria e applicazioni. Torino: Giappichelli editore; 2009.

[30] Uysal M. Reconstruction of time series data with missing values. Journal of Applied Sciences 2007;7(6) 922–925.

[31] Weerasinghe S. A missing values imputation method for time series data: an efficient method to investigate the health effects of sulphur dioxide levels. Environmetrics 2010;21(2) 162–172.

IntechOpen