Environmental Research 160 (2018) 27–34

Contents lists available at ScienceDirect

Environmental Research

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

Spatiotemporal interpolation of air pollutants in the Greater Cairo and the Delta, Egypt

Sara Osama Ahmed[a,*], Reda Mazloum[a], Hala Abou-Ali[b,c]

[a] Department of Statistics, Faculty of Economics and Political Science, Cairo University, Giza, Egypt
[b] Department of Economics, Faculty of Economics and Political Science, Cairo University, Giza, Egypt
[c] Economic Research Forum(ERF), Egypt

ARTICLE INFO

Keywords:
Air pollution
Spatiotemporal interpolation
Sum-metric variogram
Tans-Gaussian kriging
Monitoring network

ABSTRACT

This paper analyses the spatiotemporal variability of air pollutants in Egypt using monthly averages from the air quality monitoring network from 2011 to 2015. Particulate Matters (PM_{10}) Nitrogen Dioxide (NO_2) and Sulfur Dioxide (SO_2), measured by the monitoring stations network are studied. A log transformation is applied for the three pollutants to achieve normality. The sum-metric function is utilized for modelling the spatiotemporal variogram as it gave the smallest Mean Squared Error (MSE) compared to other forms namely separable, metric, and product sum models. Therefore, employing the gstat package in R together with the trans-Gaussian spatiotemporal kriging, the maps are generated for the interpolated surfaces for the monthly averages of 2015 and the corresponding standard error values. These maps will help the decision maker to understand and visualize the spatial and temporal variability of the measured pollutants and hence undertake the necessary policies and decisions. The results show that the down town area has the highest pollutants levels. As concerns the temporal dimension, the highest values are depicted during the month of February as compared to the rest of the year. Furthermore, Egypt is suffering from a serious PM_{10} problem for the area and period under study that extremely exceed the WHO and Egyptian guidelines.

1. Introduction

Air pollution is a crucial phenomenon that is targeted by scientists from different specialties using different approaches. Several studies attempted to analyze air pollution in different nations, using some advanced spatiotemporal techniques including spatiotemporal hierarchical Bayesian models, and spatiotemporal regression models (Holland et al., 2004; Ignaccolo et al., 2004; Sahu et al., 2006; Giannitrapani et al., 2007; Mariappan et al., 2013 and Del Sarto et al., 2014). While the literature review reveals that to date most Egyptian studies, depending on the available data set about air pollution, have only relied on descriptive methods and rarely on some elementary inferential methods like hypothesis testing and correlation (EL-Dars et al., 2004; Egyptian Environmental Affairs Agency, 2008a; Abu-Allaban et al., 2007; Moussa and Abdelkhalek, 2007; Zakey et al., 2008; Gad et al., 2009; World Bank, 2013; Aboel Fetouh et al., 2013).

Osama (2011) has studied air pollution in Greater Cairo by applying different spatial interpolation techniques which are Inverse Distance Weighting (IDW), Splines, Trend Surface Model, Kriging and Cokriging, for three air pollutants; Particulate Matters (PM_{10}), Nitrogen Dioxide (NO_2), and Sulfur Dioxide (SO_2). The study concluded that the Kriging approach should be used for analyzing air pollution in Egypt as it gave the lower root mean squared errors (RMSE) values and it recommended that the temporal effect should also be taken into consideration.

The present study is trying to incorporate the temporal effect as well as the spatial effect, by applying spatiotemporal techniques instead of pure spatial ones. Since air pollution is composed of many processes that exhibit complicated variability over a vast range of spatial and temporal scales. Moreover, urban areas are heterogeneous and air pollution concentration may vary over time and space. On the one hand, the spatial variation provides evidence of patterns of dependence and level of noise in the data. On the other hand, time series analyses have been used to examine air pollution over time. Spatiotemporal analyses have the additional benefits over purely spatial or time series analyses because they allow the simultaneous study of the persistence of patterns over time and helps explaining any unusual patterns. Prediction in this case depends not only on the information obtained from spatial data or time series data alone but on both of them. Separate analyses using pure spatial (temporal) techniques allow predictions in space (time) only. However interpolation of unknown values in a

* Corresponding author.
E-mail addresses: sarosama@feps.edu.eg (S.O. Ahmed), rmazloum@feps.edu.eg, rmazloum@aucegypt.edu (R. Mazloum), habouali@feps.edu.eg (H. Abou-Ali).

http://dx.doi.org/10.1016/j.envres.2017.09.005
Received 24 February 2017; Received in revised form 21 August 2017; Accepted 9 September 2017
0013-9351/

continuous space-time process should take into account the interactions between spatial and temporal components allowing for predictions in time and space. In sum, using data from air monitoring stations for spatial interpolation and creating interpolated surfaces can be done either using pure spatial techniques or spatiotemporal ones. Spatial interpolation techniques make use of the known values at measured locations at the current time point only to interpolate the unknown points at the same time point. Alternatively, spatiotemporal interpolation methods exploit the known values at the current and previous time points for interpolating the unknown values taking into consideration the structure of the temporal, spatial, and joint covariance.

The network of monitoring stations is a rich source of data that should be used efficiently to help and guide in the decision making process. These data mainly consist of measurements of specific pollutant(s) at some points in space (sites) collected on (ir) regular time points. Such data are called spatiotemporal data sets, where both the spatial and temporal effects are taken into consideration while analyzing it.

The aim of this paper is to apply the spatiotemporal kriging as an important statistical technique to study the spatial and temporal variability of $PM_{10,}$ NO_2, and SO_2 in the Greater Cairo and the Delta for the year 2015 using monthly averages from the network of monitoring stations during the period 2011–2015. This helps in studying the spatial and temporal variability of the pollutants, allowing the interpolation at unknown locations and creating interpolated surfaces for the whole area under study. This gives a great help to decision maker for taking the required actions and decisions and making best use of data from the monitoring network. The subsequent part of the paper covers data and study area in which we discuss the monitoring network and provide a basic description of the three air pollutants data. Spatiotemporal statistical techniques and the various definitions behind the methodology are illustrated in Section 3. Section 4 presents the main results and findings from applying the suggested approach. Finally Section 5 gives discussions with a summary of conclusions and recommendations in terms of research needs and priorities.

2. Data and study area

The data used in the analysis are the monthly averages from the measurements of Particulate Matters (PM_{10}), Nitrogen Dioxide (NO_2) and Sulfur Dioxide (SO_2) during the period 2011–2015 from air quality stations distributed through the Greater Cairo area and the Delta (Fig. 1). These stations are operated under the supervision of the Egyptian Environmental Affairs Agency (EEAA). The stations can be classified into industrial, urban, residential, traffic, remote and mixed. Each station measures one or more of the previously stated pollutants.

The location of each monitoring station is specified by mentioning the coordinates using longitude and latitude. The study area is depicted in Fig. 1 where the analysis is bounded by Giza governorate in the south and the Mediterranean Sea in the north. The locations of monitoring stations are spread across the Greater Cairo region and several Lower Egypt governorates where the major cities are Elmahala, Mansoura and Damietta. Furthermore, Fig. 1 shows the positions of each monitoring station for the three pollutants where the PM_{10}, SO_2 and NO_2 are measured using 26, 23 and 20 stations, respectively. Table 1 summarizes the available data for the three pollutants. Although the NO_2 is measured in shorter time interval, it has the largest percentage of missing data (20.67%). The main advantage of using the spatiotemporal kriging is the ability of making spatiotemporal prediction for the fitted surfaces. A Gridded surface of (0.05*0.05) grid is used for creating the interpolated surfaces presented later.

3. Methodology

In general, spatial interpolation is the procedure of estimating the value of properties at un-sampled sites within the area covered by existing observations. Now by taking the temporal effect into consideration the interpolation problem becomes; given $z_i = Z(s, t), i = 1,2,\dots,n \times T$, where n is the number of stations, and T the number of time points, predict $z_0 = Z(s_0, t_0)$ for a new location s_0 at a specific time point t_0. The random fields $Z(s, t)$ can be modelled as follows (Montero and Fernandez-Aviles, 2015):

$$Z(s, t) = \mu(s, t) + \varepsilon(s, t) \tag{1}$$

where $\mu(s, t)$ is the trend representing the deterministic part of the model. In the ordinary kriging method it is assumed that $(s, t) = \mu 1$, so the trend of the spatiotemporal process is unknown but constant, ε is a zero mean stochastic errors with a spatiotemporal covariance Σ_ε. The space-time interpolation problem makes use of such data to predict $Z(s_0, t_0)$.

The Best Linear Unbiased Predictor (BLUP) $Z^*_{ok}(s_0, t_0)$ for $Z(s_0, t_0)$ is given by (Cressie and Wikle, 2011):

$$Z^*_{ok}(s_0, t_0) = \hat{\mu} + \boldsymbol{\nu}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}(s, t) - \hat{\mu}\mathbf{1}) \tag{2}$$

where $\hat{\mu} = (\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^{-1}\mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}(s, t)$ is the generalized least square estimator, $var[Z(s, t)] = \boldsymbol{\Sigma}$, $var[Z(s_0, t_0)] = \sigma^2 = C(0, 0)$, $cov[Z(s, t), Z(s_0, t_0)] = \boldsymbol{\nu}$, and the corresponding ordinary kriging variance is given by

$$\sigma^2_{ok}(s_0, t_0) = C(0, 0) - \boldsymbol{\nu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu} + \frac{(1-\mathbf{1}'^{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\nu})^2}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}} \tag{3}$$

In some cases researchers use some transformations to achieve the normality assumption. The most common one is the natural log; however the usual exponential back-transformation $p(Z; s_0, t_0) = \exp\{Z^*_{ok}(s_0, t_0)\}$ is unfortunately a biased predictor of $Z(s_0, t_0)$ (Schabenberger and Gotway, 2005). Cressie (1993) has proposed using the trans-Gaussian kriging with any Box-Cox transformed variables. The new biased corrected ordinary kriging predictor and the corresponding standard error will be obtained by the krigeSTTg function in gstat package (Denby et al., 2008; Liu et al., 2016; Gräler et al., 2012; Pebesma, 2014)

The ordinary kriging predictor and the corresponding variance depend on the spatiotemporal covariance (semivariogram) structure, which needed to be estimated first. There are several approaches for defining the parametric variogram (covariance) model $\gamma(h, u; \theta)$, the most common ones are the separable, metric, sum-metric, and product-sum models (Gräler et al., 2012). Throughout this work the sum-metric covariance function will be used, which is given by

$$\gamma_{st}(h, u) = \gamma_s(h) + \gamma_t(u) + \gamma_{joint}(\sqrt{h^2 + (\kappa.\, u)^2}) \tag{4}$$

or equivalently

$$C_{st}(h, u) = C_s(h) + C_t(u) + C_{joint}(\sqrt{h^2 + (\kappa.\, u)^2}) \tag{5}$$

where h, and u are the spatial distance and time lag, respectively. The spatiotemporal anisotropy parameter κ is given as spatial unit per temporal unit. γ_s, γ_t and γ_{joint} represent the spatial, the temporal, and the joint variograms, respectively. C_s, C_t, and C_{joint} represent the spatial, the temporal, and the joint covariance functions, respectively.

4. Results

The data analysis is divided into two main parts: the exploratory data analysis including some data visualization and description, and the spatiotemporal interpolation from applying the spatiotemporal kriging.

4.1. Data exploration

As previously mentioned the data consist of monthly averages from the monitoring stations. The unit of measurement of the three pollutants is the microgram per cubic meter (μg/m^3). The Egyptian environmental law no. 4/1994 specifies the permissible limits for SO_2,