ENVIRONMETRICS
Environmetrics (2009)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/env.1007

Comparative spatiotemporal analysis of fine particulate matter pollution

W. Pang[1], G. Christakos[1*,†] and J-F Wang[2]

[1] *Department of Geography, San Diego State University, San Diego, CA, USA*
[2] *Institute of Geographic Sciences and Nature Resources Research, Chinese Academy of Sciences, Beijing, China*

SUMMARY

The prime focus of this work is the comparative investigation, theoretical and numerical, of spatiotemporal techniques used in air pollution studies. Space-time statistics techniques are classified on the basis of a set of criteria and the relative theoretical merits of each technique are discussed accordingly. The numerical comparison involves the applications of two representative techniques. For this purpose, the popular spatiotemporal epistemic knowledge synthesis and graphical user interface (SEKS-GUI) software of spatiotemporal statistics is used together with a dataset of $PM_{2.5}$ daily measurements obtained at monitoring stations geographically distributed over the state of North Carolina, USA. The analysis offers valuable insight concerning the choice of an appropriate spatiotemporal technique in air pollution studies. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: spatiotemporal analysis; air pollution; particulate matter; Bayesian maximum entropy; kriging; geostatistics

1. INTRODUCTION

Concern over ambient fine particulate matter (PM) pollution is becoming more prevalent in the modern world due to its potentially harmful effects on the human health and the environment. $PM_{2.5}$ air pollution studies, in particular, have investigated its possible association with certain adverse health effects (Dockery *et al*., 1993; Pope *et al*., 1995; Gauderman *et al*., 2000; Samet *et al*., 2000; Tainio *et al*., 2005; Mascarenhas *et al*., 2008). Environmental organizations, regulatory groups, and local governments have launched projects that monitor general PM air pollution levels, in an effort to adequately represent their patterns across space-time under conditions of uncertainty (Christakos and Serre, 2000a, 2000b; Christakos *et al*., 2001; Kibria *et al*., 2002; Smith *et al*., 2003; Serre *et al*., 2004; Allshouse *et al*., 2006; Liao *et al*., 2006; Bell *et al*., 2007; Cocchi *et al*., 2007; Bogaert *et al*., 2009; Yu *et al*., 2009). The latter task calls on the need for rigorous methods that can provide informative space-time estimates and dynamic visualizations (maps) of $PM_{2.5}$ pollution throughout a geographical region.

*Correspondence to: G. Christakos, Department of Geography, San Diego State University, San Diego, CA, USA.
†E-mail: gchrista@mail.sdsu.edu

Received 16 March 2009
Accepted 7 April 2009

Mainstream spatial statistics and geostatistics methods have been widely used to assess geographical dependence and generate maps of various physical attributes (Gandin, 1963; Olea, 1974, 1999; Journel and Huijbregts, 1978; Christakos, 1985; Haining, 1990; Cressie, 1991). Among the best known spatial estimation techniques is kriging (or spatial regression) in its various forms, including ordinary kriging, OK; simple kriging, SK; indicator kriging, InK; universal kriging, UK; and intrinsic kriging, IK (e.g., Isaaks and Srivastava, 1989; Deutsch and Journel, 1992; Chiles and Delfiner, 1999; Dalezios *et al*., 2002; Bayraktar and Turalioglu, 2005). As part of the scientific evolution process, the earlier development of spatial statistics and geostatistics was followed by the introduction of a theoretical framework of spatiotemporal statistics and geostatistics (Christakos, 1991a, 1991b, 1992; Bogaert, 1996). This framework extended many of the earlier techniques in a composite space-time domain, including the introduction of space-time kriging for heterogeneous variations, in general (non-homogeneous, non-stationary etc.) and the construction of space-time dependence models (covariance and variograms, ordinary and generalized, separable and non-separable). In addition, this framework introduced concepts and tools that could effectively handle space-time problems, which previous methods were not able to study due to lack of the adequate conceptual and technical support. With the advent of modern spatiotemporal statistics and geostatistics, a set of new techniques of space-time modeling and estimation were proposed, including, Bayesian maximum entropy, Bayesian data fusion, information-theoretic analysis, Radonian space transforms, non-Bayesian stochastic logic, differential geometric, and space-time diagrammatic techniques (Christakos, 1984, 1990, 1992; Christakos and Li, 1998; Hristopulos *et al*., 1999; Bogaert, 2002; Christakos *et al*., 2002; Kolovos *et al*., 2004; Douaik *et al*., 2005; Wibrin *et al*., 2006; Orton and Lark, 2007; Fasbender *et al*., 2008; Lee *et al*., 2008).

In science-based spatiotemporal analysis, one distinguishes between two major knowledge bases (KB), as follows:

(i) The core (or general) KB, denoted by *G*-KB, which refers to what is already known about the situation under study. As such, the *G*-KB may include physical laws, primitive equations, reasoning schemes, and theoretical models of space-time dependence.
(ii) The specificatory KB, *S*-KB, which refers to the characteristics of the specific site under consideration. As such, the *S*-KB may include hard data obtained across the site (exact numerical measurements with no uncertainty for all practical purposes), and soft information (data with fair or considerable degrees of uncertainty).

The soft information component of the *S*-KB may take the form, e.g., of intervals (there is not a unique data value available at a location but, instead, an interval of possible values), probability functions (the datum at the specified space-time location has the form of a probability distribution), and fuzzy data (gradual assessment of uncertain information sources). Naturally, the total KB is denoted by $K = G \cup S$, i.e., it includes both the core and the site-specific KBs.

Modern spatiotemporal analysis can account for various kinds of KBs as described above and provide valuable tools for ambient air pollution monitoring and mapping, especially in regions with fine PM pollution. In this work, representative space-time data analysis techniques are used to study the geographical distribution of $PM_{2.5}$ concentrations in the state of North Carolina during the year 2000. The relative merits of the techniques are compared on both theoretical and practical grounds, thus offering valuable insight concerning the choice of an appropriate spatiotemporal analysis technique in real-world studies.

DOI: 10.1002/env

2. BACKGROUND AND STUDY AREA

$PM_{2.5}$ air pollution is a class of particulate pollution that comprises particles that have effective aerodynamic diameters of 2.5 micrometers (μm) or less, allowing them to remain suspended in the atmosphere and thus, settle out slowly, as well as potentially penetrate deeper into human bodies through passages such as the mouth or nose (Godish, 2004; Tainio *et al*., 2005). In terms of their origins, $PM_{2.5}$ air pollution can arise from natural and anthropogenic sources (Artnano *et al*., 2003; Zhang *et al*., 2007).

Natural $PM_{2.5}$ air pollution is generated based on certain unavoidable, but transient, atmospheric conditions (Godish, 2004). Pollutant concentrations are affected by emissions, topography, and land cover. The $PM_{2.5}$ distribution tends to vary geographically and seasonally. Besides major events such as forest fires and volcano eruptions, events producing natural $PM_{2.5}$ air pollution occur from animal and plant decomposition, pollen and spores, volatile hydrocarbon emissions from vegetation, ocean spray, soil erosion and mineral weathering, gas-phase substance emissions from soil and water surfaces, and ozone and nitrogen oxide emissions from electrical storms (Goldberg *et al*., 2000). While little can be done to mitigate natural $PM_{2.5}$ air pollution, a lot of things could be done to lower its anthropogenic counterpart.

Anthropogenic $PM_{2.5}$ air pollution is viewed as a serious environmental and public health problem (Bernstein and Abelson, 2005; Mascarenhas *et al*., 2008; Yu *et al*., 2008). Its seriousness lies in the fact that elevated pollutant levels are produced in environments where harm to human health and welfare is more likely (Pope *et al*., 1995; Godish, 2004). Some of the most common sources of anthropogenic $PM_{2.5}$ air pollution include transportation (such as cars, trains, and airplanes), stationary fuel combustion, industrial processes, waste disposal, and secondary chemical reactions in the atmosphere. The potential that ambient $PM_{2.5}$ air pollution has in regard to its adverse effects on both the environment and human health is what makes it the significant concern (Zidek, 1997; Kibria *et al*., 2002; Bell *et al*., 2007). Thus, research on this phenomenon is of high value to many academic, industrial, and governmental sectors.

In the present work, the study area is the state of North Carolina (Eastern USA). North Carolina has a latitudinal span from 33°50' to 36°35' N and a longitudinal span from 75°28' to 84°19' W. Its geography consists of three main regions: the coastal plain, the Piedmont region, and the Appalachian mountains and foothills. The coastal plain's relative flatness makes it prime land for agriculture. The Piedmont region is the most urbanized and densely populated region, but still has gently rolling countryside frequently broken by hills or low mountain ranges. The Appalachian mountains and foothills section of the state has some of the tallest peaks in the Eastern USA, making it a hub for tourism. Thus, air pollution monitoring and control has population health and financial consequences as well. While North Carolina is located in a warm temperate zone, its diverse regions can experience a variety of weather conditions. Locations on the mountains may see average temperatures of 30°F in January and 65°F in August, whereas certain locations in the coastal plains often experience averages in the mid 40s in January and in the 90s°F in August.

Space-time $PM_{2.5}$ hard datasets were acquired from the United States Environmental Protection Agency's (USEPA) Air Quality System (AQS) database. The data were compiled from 38 $PM_{2.5}$ monitoring stations geographically distributed throughout the study area (Figure 1), which provide information about the $PM_{2.5}$ concentration levels (measured in $\mu g/m^3$), spatial coordinates, collection time, sampling duration, and sampling frequency. The period of temporal data collection was from January 1 to December 31, 2000. During the preprocessing stage of the dataset, the relevant column fields included the latitude and longitude coordinates, date, hour, and measurement value. Due to the

DOI: 10.1002/env

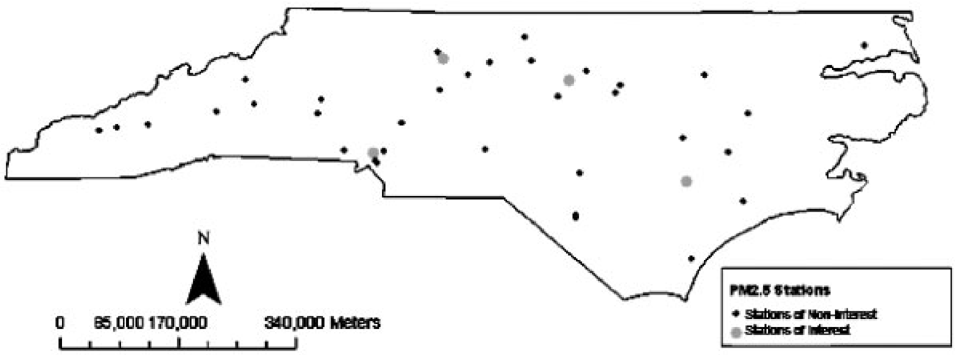

Figure 1. Geographical distribution of the $PM_{2.5}$ monitoring stations in the state of North Carolina; the four bigger (gray) circles indicate the control stations

requirements of subsequent analysis, a unique temporal coordinate column field had to be created using the values from the date and hour column fields. Also, instead of dealing with daily average measurements, weekly average measurements were used in order to reduce the calculation overhead (this process involved aggregating the daily data into weekly averages).

Table 1 shows the statistics of the weekly adjusted averages obtained from the $PM_{2.5}$ hard dataset ($\mu g/m^3$, wherever applicable). Ideally there should have been 1976 data (52 weekly average measurements for each of the 38 $PM_{2.5}$ monitoring stations). However, the $PM_{2.5}$ dataset had missing data (which is common in practice), but not enough to inhibit analysis or cause the results to be affected in any significant way. It is worth noting that modern spatiotemporal analysis techniques can account for missing data, if necessary, in a physically meaningful and mathematically rigorous manner (see the next section).

Since the prime goal of the present study is to compare the performances of different space-time analysis techniques, the weekly adjusted averages for the $PM_{2.5}$ air pollution dataset were used as a reference to create two additional files: (i) a hard dataset with intentionally omitted space-time data at four $PM_{2.5}$ monitoring stations (herein called "control stations," CS_i, $i = 1, \ldots, 4$; stations are numbered from left to right in Figure 1) and (ii) a soft dataset containing secondary space-time information about the omitted data. In particular, for comparative analysis purposes the soft dataset was generated by replacing the previously omitted hard data at the control stations with interval data, where the upper and lower bounds at each space-time point are assumed to be varying percentages of the original hard datum. So while the modified hard dataset had one column for the measurement value, the soft dataset has two columns: one with a lower bound and one with an upper bound at each space-time point.

Table 1. $PM_{2.5}$ dataset statistics

Count	1776	Standard deviation	5.61
Minimum	3.70	Median	14.30
Maximum	49.35	Skewness	0.89
Mean	15.14	Kurtosis	4.74

3. SPATIOTEMPORAL ANALYSIS

This work used the spatiotemporal OK and BME techniques to study $PM_{2.5}$ patterns in the state of North Carolina during the year 2000 using space-time information (i.e., geographically distributed, including data from other years). Although spatial OK has been used for several decades, spatiotemporal OK is a much more recent development. Also, spatiotemporal OK is one of the most widely used kriging techniques in space-time analysis (Bogaert, 1996), which is why it has been used in the present study. Both the OK and BME techniques are described in detail in the relevant geostatistics and spatiotemporal statistics literature (see references above), so that there is no need for us here to delve into mathematical and other technical details. Instead, we start with a brief presentation of the theoretical differences of these two representative space-time techniques, and then proceed with their application in the dataset of interest.

On theoretical grounds, the basic concepts and assumptions of OK (spatial or spatiotemporal) can be compared with those of BME on the basis of the following criteria (similar theoretical comparisons are also valid between BME and techniques other than OK, including SK, UK, InK, IK, and Kalman filters):

Estimator form: The OK is a linear estimator (Dowd, 1992; Bogaert, 1996). BME, on the other hand, makes no restrictive assumptions concerning the linearity of the estimator (Law *et al*., 2006; Lee *et al*., 2008).

Shape of probability law: A basic assumption of OK is normality, i.e., the underlying random fields are assumed to be Gaussian (Olea, 1999; Chiles and Delfiner, 1999). In the case of BME, however, non-Gaussian laws are automatically incorporated (Hristopulos and Christakos, 2001; Papantonopoulos and Modis, 2006; Orton and Lark, 2007).

KBs processed: In the case of OK, the associated site-specific KB basically processes hard data, which can be limiting, especially in situations where some potentially informative data might be soft (Haining, 1990; Gundogdu and Guney, 2007). In some special applications, kriging has relied on arbitrary and rather ad hoc tricks to account for certain soft data forms, but these tricks often lack mathematical rigor and scientific substance (Douaik *et al*., 2005; Saito *et al*., 2005). Also, OK does not make use of core physical knowledge that proves to be significant in certain situations (Kolovos *et al*., 2002; Bayraktar and Turalioglu, 2005). The BME, on the other hand, can integrate various kinds of core and site-specific KBs in a general and unified manner, and it can even assimilate uncertain yet valuable information at the estimation points themselves, when available (Serre *et al*., 2004; Parkin *et al*., 2005; Christakos *et al*., 2005). This also allows BME to efficiently account for missing data by means of the nonlinear integral formulation of the probability density function (pdf) at each space-time point (Christakos, 2000).

Estimation characterization: The OK generates a single estimated value at each geographical grid node and the associated statistical estimation variance (Isaaks and Srivastava, 1989; Haining, 1990). On the other hand, BME offers a more sound characterization in terms of the complete estimation pdf at every node. Each pdf may have a different shape (non-Gaussian, in general) at each space-time node; and from each pdf one can chose a number of possible estimates with their associated probabilities, accuracies, confidence intervals etc. (Serre *et al*., 2003; Puangthongthub *et al*., 2007; Querido *et al*., 2007).

Generalization power: In theory, BME derives several mainstream geostatistics and space-time statistics techniques as its special cases, a fact that amply demonstrates BME's generalization power. For example, it can be shown (Christakos and Hristopulos, 1998; Christakos, 2000) that under certain limiting conditions on the KB and the space-time dependence functions considered the BME

DOI: 10.1002/env

obtains OK, SK, and IK as its special cases. Also, BME accounts for important physical cross-correlations in the spatiotemporal domain that are not considered by mainstream techniques (Kolovos *et al*., 2002, 2004). Various extensions of BME are possible, including the generalized BME (GBME; Yu *et al*., 2007a, 2008) that processes directly heterogeneous space-time variations of any degree, vectorial BME (Choi *et al*., 1998) that simultaneously incorporates several space-time attributes linked via a physical law or an empirical relationship, and functional BME (Christakos, 2000) that accounts for different space-time attribute supports.

On practical grounds, the spatiotemporal epistemic knowledge synthesis and graphical user interface software library (SEKS-GUI; Kolovos *et al*., 2006) was the primary tool used in the present spatiotemporal $PM_{2.5}$ data analysis. A recent version of SEKS-GUI can be found in http://homepage.ntu.edu.tw/~hlyu/software/SEKSGUI/SEKSHome.html. Since this kind of spatiotemporal statistics software has been routinely used in scientific applications for about two decades, no mathematical or technical details are presented here concerning the relevant models and methods (BME, GBME, space-time kriging etc.). Interested readers are referred to the website above, the SEKS-GUI Users Manual (Kolovos *et al*., 2006), the review paper by Yu *et al.* (2007a), and references therein.

In SEKS-GUI, the distribution of the $PM_{2.5}$ concentrations is mathematically represented as a spatiotemporal random field X_p (S/TRF). The S/TRF domain is denoted by $p = (s, t)$, in which $s = (s_1, s_2)$ refers to two-dimensional spatial coordinates and t is time. For example, one may chose to express spatial distance in kilometers (km) and time in weeks (wks). In order to generate maps of the $PM_{2.5}$ concentration distribution over the study area, an output grid containing evenly distributed nodes p was defined. This output grid followed a simple format where each axis was determined by supplying an upper and lower bound and a parameter that signified the number of spacing units that exist between nodes. The dimensions of the output grid allowed for a sufficient number of concentration estimates; the total number of nodes along the $s_1 \times s_2 \times t$-axes is $(32 \times 90 \times 52) = 149\,760$ nodes.

As was mentioned before, two different spatiotemporal analyses were conducted using the SEKS-GUI software: the OK and BME analyses of space-time statistics and modern geostatistics. For numerical comparison purposes, it was assumed that the techniques shared in common certain stages of data preparation. Space-time trends in the $PM_{2.5}$ variation were identified and removed. Experimental space-time covariance values were calculated and theoretical space-time models were fitted to these experimental values. The models were selected from a list of space-time covariance models available in SEKS-GUI so that they offered best fit to the experimental values and, at the same time, represented adequately the composite space-time dependence (correlation) structure of the $PM_{2.5}$ distribution. ‘‘Composite’’ is meant in the sense that the full spatiotemporal structure was taken into account (including heterogeneities and cross-dependences), whereas no simplifying assumptions were made, such as spatial independence and temporally uncorrelated $PM_{2.5}$ components used in previous studies (Daniels *et al*., 2001; Smith *et al*., 2003). Then, the two techniques generated $PM_{2.5}$ estimates and error variances (or standard deviations) at each output grid node across space-time.

The visualization of the results of the spatiotemporal analysis above (estimates across space-time, estimation errors) in terms of informative maps involved an essential geographical element provided by external GIS software. In particular, ESRI-ArcGIS and MapWindow’s CSV-to-Shapefile plug-in were used to convert the output $PM_{2.5}$ estimation files into shapefiles and reproduce maps that were superimposed over a shapefile of the study area. Maps of $PM_{2.5}$ estimates help visualize the spatiotemporal distribution and identify pollution patterns during the course of the 52-week period (1 year) for the area of interest. The corresponding $PM_{2.5}$ estimation error variance maps and resulting summary error statistics served as a measure of comparison between the performances of the space-time estimation techniques. The error variances at the four control stations were analyzed and the actual errors were calculated at these stations as well as the statistical estimation error variances at every output grid node. Maps such as the above offer valuable insight concerning decision-making, resource allocation, and emission control strategies.

4. NUMERICAL RESULTS

4.1. $PM_{2.5}$ pollution analysis and mapping

By means of the SEKS-GUI software, the space-time OK analysis calculated the $PM_{2.5}$ statistics, space-time dependence functions, and $PM_{2.5}$ estimates on a space-time grid; and then used these estimates to generate a set of informative pollution maps by means of the ESRI-ArcGIS visualization tools. In particular, following the initial data processing stage (data detrending, normalization etc.), the space-time dependence stage was conducted next, in which OK used hard weekly data at 34 geographical locations during the year 2000. Experimental covariance values were calculated along different directions, and the dependence structure of the $PM_{2.5}$ dataset was subsequently represented in terms of theoretical nested models. Using the interactive SEKS-GUI screens, the first component of the nested model consisted of a space-time combination of spherical-exponential functions, whereas the second component included a combination of spherical-Gaussian functions. The space-time covariance can be presented as a moving three-dimensional plot (two spatial and one temporal coordinates) or by a variety of two-dimensional plots. As was also observed in previous $PM_{2.5}$ studies, some spatial covariance plots may include a nugget effect. Certain differences were observed between spatial covariance plots considered at different time lags (corresponding to different seasons) and between temporal covariance plots considered at different spatial lags, whereas the overall covariance shape reflects the different scales of the physical processes underlying the $PM_{2.5}$ distribution across space and time (Yu *et al*., 2009).

At the space-time OK estimation stage, the data included the weekly adjusted $PM_{2.5}$ hard measurements at the 34 monitoring stations together with the means of the interval (soft) data at the four control stations (which is a usual kriging practice known as soft data ‘‘hardening’’), see also Section 2. Detailed $PM_{2.5}$ maps were generated for every week of the year 2000. In Figure 2, we show a subset of these OK maps of $PM_{2.5}$ estimates for the wks 0, 13, 38, and 51. The black dots in each map represent the locations of the monitoring stations, with the larger white dots designated as the four $PM_{2.5}$ control stations (also shown in Figure 1). A considerable $PM_{2.5}$ variation is observed across space and time, which may be due to a number of factors discussed in the relevant literature (e.g., Holland *et al*., 2000; Artnano *et al*., 2003; Godish, 2004).

Next the air pollution study focused on space-time BME analysis. By means of SEKS-GUI, concentration estimates and maps were generated using the $PM_{2.5}$ dataset available. In the case of missing data at critical locations, BME can handle the problem by generating soft data (e.g., in the form of probability distributions or fuzzy sets) on the basis of secondary information (qualitative assessment, expert opinions, empirical relationships, fuzzy logic etc.; Bogaert, 2002; Kovitz and Christakos, 2004; Parkin *et al*., 2005; Law *et al*., 2006). Also, it is worth noting that if the space-time GBME technique is selected, certain activities of the OK preprocessing stage (data detrending, normalization) are not needed, which could reduce possible preprocessing errors (Yu *et al*., 2007a, 2008). At the space-time dependence stage of BME, a technique proposed in Christakos *et al.* (2002) was used to include some soft data into covariance calculations. Note that there is a variety of space-time dependence functions

DOI: 10.1002/env

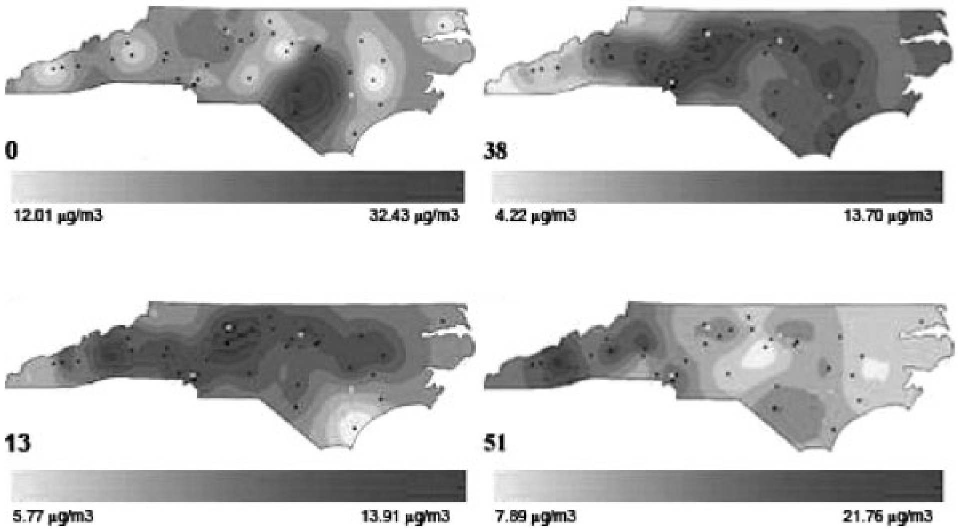

Figure 2. A subset of OK-based $PM_{2.5}$ maps (weeks 0, 13, 38, and 51)

(covariance and variograms; separable and non-separable; ordinary and generalized) that can be used in this stage (e.g., Kolovos *et al*., 2004; Mateu *et al*., 2007; Porcu *et al*., 2008). Nevertheless, for comparison purposes the space-time dependence $PM_{2.5}$ structure was represented using the theoretical nested models as in the case of space-time OK above, although different model parameters were calculated when the models were fitted to the experimental values (primarily due to the fact that in the BME case some soft data were also included in the calculation of the experimental covariances).

As with space-time OK analysis, at the space-time BME estimation stage the data available were the $PM_{2.5}$ hard dataset and the soft dataset that contained interval data at the control stations. However, unlike space-time OK (which used the so-called ''hardening'' scheme based on middle soft values at these points), the space-time BME formulation processed directly the actual functional form of the soft data at the same points (thus accounting for all possible values of each interval). Note that the shape of the soft data affects the nonlinearity of the space-time BME estimator; e.g., if the soft datum has an interval or probabilistic shape, the corresponding expression for the integrated pdf at each space-time grid node has a nonlinear integral expression, in general. As before, detailed maps of the space-time $PM_{2.5}$ distribution were generated for every week of the year 2000. A subset of these $PM_{2.5}$ maps is shown in Figure 3. The maps provide valuable space-time visualizations of the considerable geographical variation and seasonality of the $PM_{2.5}$ distribution in the state of North Carolina. These $PM_{2.5}$ maps, just like those of Figure 2, can offer valuable insight concerning decision-making, resource allocation and emission control strategies. The theoretical differences of the two space-time analysis techniques (estimator form, underlying pdf, internal processing of KBs etc.) have led to some quantitative differences between the space-time OK and BME maps of $PM_{2.5}$ distribution, which will be assessed numerically in the following section.

4.2. Numerical comparison of space-time $PM_{2.5}$ analyses

The $PM_{2.5}$ concentration values X_p are known at the control stations $CS_i(i = 1, \ldots, 4$; p denotes the space-time coordinates of each station (Figure 1). Given the X_p at $CS_i(i = 1, \ldots, 4)$, the actual errors

DOI: 10.1002/env

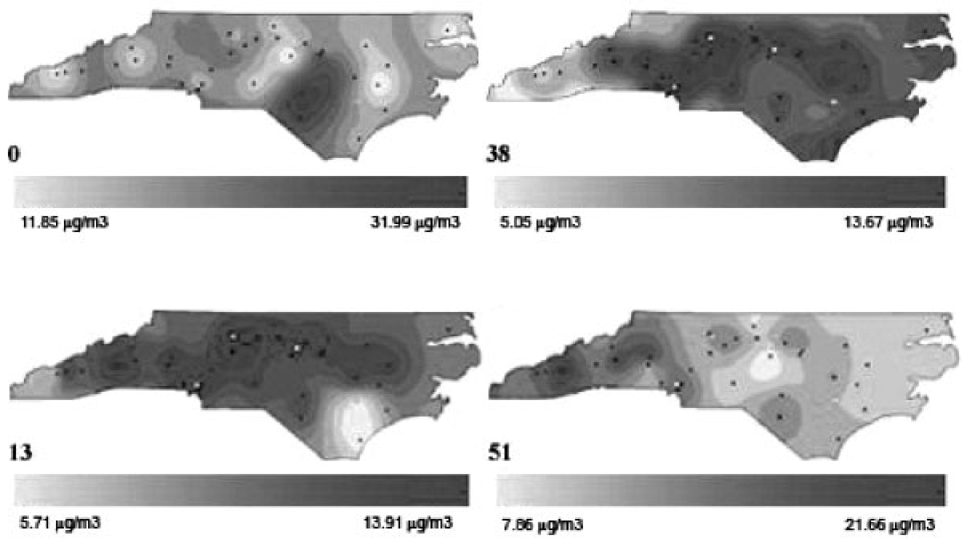

Figure 3. A subset of BME-based $PM_{2.5}$ maps (weeks 0, 13, 38, and 51)

at the control stations can be calculated in terms of the absolute differences, $e^{\mathrm{OK}} = |X_p - \hat{X}_p^{\mathrm{OK}}|$ and $e^{\mathrm{BME}} = |X_p - \hat{X}_p^{\mathrm{BME}}|$, where $\hat{X}_p^{\mathrm{OK}}$ and $\hat{X}_p^{\mathrm{BME}}$ are the corresponding OK and BME estimates at p, respectively. The average e^{OK} and e^{BME} values during the entire year 2000 are tabulated in Table 2. Clearly, at all control stations considered, the space-time BME estimates were superior to the space-time OK estimates.

Furhermore, the $PM_{2.5}$ estimation error statistics at the control stations $\mathsf{CS}_i(i = 1, \ldots, 4)$ are shown in Table 3. The mean, maximum, and minimum error variances $(\mu g/m^3)^2$ over all 52 wks of

Table 2. Actual space-time OK and BME estimation errors ($\mu g/m^3$) at control stations $\mathsf{CS}_i(i = 1, \ldots, 4)$ averaged over the year 2000

Control station	OK	BME
CS_1	2.02	0.42
CS_2	2.10	1.03
CS_3	1.21	0.50
CS_4	2.30	0.59

Table 3. Error variances $(mg/m^3)^2$ at control stations $\mathsf{CS}_i(i = 1, \ldots, 4)$ and the corresponding ratios between the two space-time estimation techniques

	CS_1			CS_2			CS_3			CS_4		
	OK	BME	$\frac{OK}{BME}$	OK	BME	$\frac{OK}{BME}$	OK	BME	$\frac{OK}{BME}$	OK	BME	$\frac{OK}{BME}$
Mean	6.07	1.82	3.34	1.36	0.83	1.64	2.34	1.22	1.91	1.92	0.90	2.14
Maximum	6.72	1.96	3.42	1.45	0.95	1.52	2.43	1.38	1.76	2.81	1.09	2.57
Minimum	6.05	1.73	3.50	1.16	0.74	1.57	2.33	1.15	2.03	1.89	0.83	2.28

DOI: 10.1002/env

2000 were calculated at each control station for space-time OK and BME, and were subsequently compared via the ratios between them. Once again, the space-time BME analysis produced $PM_{2.5}$ estimation statistics that were better than those produced by the space-time OK analysis.

5. DISCUSSION AND CONCLUSIONS

As was mentioned in the introduction, the principal focus of this work was the comparative analysis of different spatiotemporal techniques on theoretical and numerical grounds. Therefore, the paper did not deal with otherwise important issues, such as the effect on space-time air pollution distributions of weather conditions, residential features, scales, topography, and land use.

A brief yet critical review of the literature outlined the main methodological differences between certain techniques of spatiotemporal statistics and geostatistics. It was shown that space-time BME has a number of theoretical advantages over space-time OK and other techniques. The numerical comparison between space-time OK and BME was made using the SEKS-GUI spatiotemporal analysis software, which also demonstrated certain important practical differences when the two space-time techniques are applied in practice. For example, they handle soft information in a different manner: space-time OK used an ad hoc criterion of considering the middle value of the soft datum at each space-time point, whereas BME accounted for all possible values of the soft datum (BME can also account for useful soft information at the estimation points themselves). As a matter of fact, many techniques routinely assume the mean value of the interval or probability distribution at each (soft data) point as the ‘‘hardened’’ value to be used in space-time estimation. However, there may be sound reasons to select different soft data values at certain points (say, the mean, highest, lowest, most probable, or most improbable values) rather than insisting on the mean value at all points. These reasons include the fact that the former provides sufficient flexibility and it often is a more realistic approach than the latter (Yu *et al*., 2008). For example, a spatiotemporal analysis that, on occasion, accounts for low-probability values, may turn out to be very informative: when these values occur, they can be highly consequential (as the financial markets know, it does not matter how rare an event is if its occurrence is too costly to bear).

The space-time $PM_{2.5}$ maps generally provide interesting information about the composite space-time features of pollutant distribution, including spatial variation and seasonal variations. At a fine scale, the $PM_{2.5}$ estimates for the space-time OK and BME techniques are significantly different from each other, with the estimates produced by the space-time BME analysis being more accurate than the space-time OK analysis. This observation is further emphasized through the comparison of the $PM_{2.5}$ estimation errors and error variances obtained by these two techniques. Comparative analysis quantitatively and visually demonstrated that for each of the four control stations the BME technique yielded lower $PM_{2.5}$ estimation errors and error variances than OK. Also, at the control stations, the values of the OK/BME estimation error ratios favored space-time BME over OK by a factor of up to 3.5.

Summarizing, the comparative analysis of the two techniques has demonstrated that space-time BME is superior to space-time OK (and this is true for other forms of kriging and mainstream statistical regression, for the theoretical reasons discussed above). Furthermore, the analysis showed that accounting for useful information at certain space-time points, even if the information is soft and uncertain, it is often better than ignoring it. Several useful extensions of the present air pollution analysis are possible. For example, given the space-time heterogeneous variations of certain pollutants, the implementation of generalized spatiotemporal analysis may be a more adequate approach in such

DOI: 10.1002/env

cases. This approach, which combines heterogeneity characterization in terms of generalized space-time dependence models (covariances, variograms, and structure functions) and air pollution mapping in terms of the generalized space-time kriging system, has already been used with considerable success in previous studies (Vyas and Christakos, 1997; Christakos and Vyas, 1998). In recent years, generalized spatiotemporal analysis has been rediscovered and proposed as a potentially useful method in the study of space-time heterogeneous air pollution variations, including PM datasets (e.g., Smith *et al*., 2003). Another useful extension is a combination of generalized spatiotemporal analysis with BME modeling that leads to various forms of GBME analysis with interesting air pollution applications (Christakos and Hristopulos, 1998; Christakos and Kolovos, 1999; Yu *et al*., 2007a, 2007b, 2008). Finally, incorporating multiple-point spatiotemporal statistics and accounting for space-time support and scale effects may further improve the air pollution analysis.

ACKNOWLEDGEMENTS

Partial support for the research was provided by the California Air Resources Board, USA (55245A).

REFERENCES

Allshouse WB, Serre ML, Pleil J, Rappaport S. 2006. A space/time particulate matter mass fraction framework for the assessment of outdoor exposure to polycyclic aromatic hydrocarbons after 9/11 in New York city. *Epidemiology* **17**(6): Suppl: S268–S269.

Artnano B, Salvador P, Alonso DG, Querol X, Alastuey A. 2003. Anthropogenic and natural influence on the PM10 and PM2.5 aerosol in Madrid (Spain). Analysis of high concentration episodes. *Environmental Pollution* **125**(3): 453–465.

Bayraktar H, Turalioglu FS. 2005. A kriging-based approach for locating a sampling site in the assessment of air quality. *Stochastic Environmental Research and Risk Assessment* **19**(4): 301–305.

Bell ML, Dominici F, Ebisu K, Zeger SL, Samet JM. 2007. Spatial and temporal variation in PM(2.5) chemical composition in the United States for health effects studies. *Environmental Health Perspectives* **115**(7): 989–995.

Bernstein AS, Abelson HT. 2005. PM 2.5: a killer in our midst. *Archives of Pediatrics & Adolescent Medicine* **159**(8): 786.

Bogaert P. 1996. Comparison of kriging techniques in a space-time context. *Mathematical Geology* **28**: 73–786.

Bogaert P. 2002. Spatial prediction of categorical variables: the BME approach. *Stochastic Environmental Research and Risk Assessment* **18**: 425–448.

Bogaert P, Christakos G, Jerrett M, Yu H-L. Spatiotemperol modelling of ozone distribution in the State of California. *Athmospheric Environment* **43**: 2471–2480.

Chiles J-P, Delfiner P. 1999. *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons: New York, NY.

Choi K-M, Christakos G, Serre ML. 1998. Recent developments in vectorial and multi-point BME analysis. In *Proceedings of 4th Annual International Association for Mathematical Geology (IAMG) Conference*, Buccianti A, Nardi G, Potenza R (eds). De Frede Editore: Naples, Italy; Vol. 1: 91–96.

Christakos G. 1984. The space transformations and their applications in systems modelling and simulation. In *Proceeding 12th International Confererence on Modelling and Simulation-AMSE*. C 1(3): 49–68 Athens, Greece.

Christakos G. 1985. Modern statistical analysis and optimal estimation of geotechnical data. *Engineering Geology* **22**(2): 175–200.

Christakos G. 1990. A Bayesian/maximum entropy view to the spatial estimation problem. *Mathematical Geology* **22**(7): 763–777.

Christakos G. 1991a. On certain classes of spatiotemporal random fields with application to space-time data processing. *IEEE Transactions on Systems, Man, and Cybernetics* **21**(4): 861–875.

Christakos G. 1991b. Some applications of the Bayesian maximum entropy concept in geostatistics. In *Maximum Entropy and Bayesian Methods*. Kluwer: Boston, MA; 215–229.

Christakos G. 1992. *Random Field Models in Earth Sciences*. Academic Press: San Diego, CA.

Christakos G. 2000. *Modern Spatiotemporal Geostatistics*. Oxford: New York, NY.

Christakos G, Bogaert P, Serre ML. 2002. *Temporal GIS*, With CD-ROM, Springer-Verlag: New York, NY.

Christakos G, Olea RA, Serre ML, Yu HL, Wang L-L. 2005. *Interdisciplinary Public Health Reasoning and Epidemic Modelling: The Case of Black Death*. Springer: New York, NY.

Christakos G, Serre ML, Kovitz J. 2001. BME representation of particulate matter distributions in the state of California on the basis of uncertain measurements. *Journal of Geophysical Research* **106**(D9): 9717–9731.

DOI: 10.1002/env

Christakos G, Hristopulos DT. 1998. *Spatiotemporal Environmental Health Modelling: A Tractatus Stochasticus*. Kluwer Academic Publishers: Boston, MA.
Christakos G, Li X. 1998. Bayesian maximum entropy analysis and mapping: a farewell to kriging estimators. *Mathematical Geology* **30**(4): 435–462.
Christakos G, Kolovos A. 1999. A study of the spatiotemporal health impacts of ozone exposure. *Journal of Exposure Analysis & Environmental Epidemiology* **9**(4): 322–335.
Christakos G, Serre ML. 2000. A spatiotemporal study of exposure-health effect associations. *Journal of Exposure Analysis & Environmental Epidemiology* **10**(2): 168–187.
Christakos G, Serre ML. 2000. BME analysis of spatiotemporal particulate matter distributions in North Carolina. *Atmospheric Environment* **34**(20): 3393–3406.
Christakos G, Vyas VM. 1998. A composite space/time approach to studying ozone distribution over the Eastern United States. *Atmospheric Environment* **32**: 2845–2857.
Cocchi D, Greco F, Trivisano C. 2007. Hierarchical space-time modelling of PM10 pollution. *Atmospheric Environment* **41**(3): 532–542.
Cressie N. 1991. *Statistics for Spatial Data*. Wiley: New York, NY.
Dalezios NR, Loukas A, Bampzelis D. 2002. Universal kriging of hail impact energy in Greece. *Physics and Chemistry of the Earth* **27**(23): 1039–1043.
Daniels MJ, Lee Y-D, M. Kaiser M. 2001. Assessing sources of variability in measurement of ambient particulate matter. *Environmetrics* **12**: 547–558.
Deutsch CV, Journel AG. 1992. *Geostatistical Software Library and User's Guide*. Oxford University Press: Oxford, UK.
Dockery DW, Pope CA, Xu X, Spengler JD, Ware JH, Fay ME, Ferris BG, Speizer F. 1993. An association between air pollution and mortality in six U.S. cities. *New England Journal of Medicine* **329**: 1753–1759.
Douaik A, Van Meirvenne M, Toth T. 2005. Soil salinity mapping using spatio-temporal kriging and Bayesian maximum entropy with interval soft data. *Geoderma* **128**: 234–248.
Dowd PA. 1992. A review of recent developments in geostatistics. *Computers & Geosciences* **17**(10): 1481–1500.
Fasbender D, Radoux J, Bogaert P. 2008. Bayesian data fusion for adaptable image pansharpening. *IEEE Transactions on Geoscience Remote Sensing* **46**(6): 1847–1857.
Gandin LS. 1963. *Objective Analysis of Meteorological Fields*. Gidrometeorologicheskoe Izdatel'stvo, Leningrad, USSR.
Gauderman JW, McConnell R, Gilliland F, London S, Thomas D, Avol E, Berhane K, Rappaport E, Lurmann F, Margolis H, Peters J. 2000. Association between air pollution and lung function growth in Southern-California school-aged children. *American Journal of Respiratory and Critical Care Medicine* **162**(4): 1–8.
Godish T. 2004. *Air Quality*. Lewis Publishers: Boca Raton, FL.
Goldberg MS, Bailar JC, Burnett RT, Brook JR, Tamblyn R, Bonvalot Y, Ernst P, Flegel KM, Singh R, Valois M-F. 2000. Identifying subgroups of the general population that may be susceptible to short-term increases in particulate air pollution: a time-series study in Montreal, Quebec *Report 97*. Health Effects Institute, Cambridge, MA.
Gundogdu KS, Guney I. 2007. Spatial analyses of groundwater levels using universal kriging. *Journal Earth System Science* **116**(1): 49–55.
Haining R. 1990. *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge University Press: Cambridge, UK.
Holland DM, De Oliveira V, Cox LH, R. L. Smith RL. 2000. Estimation of regional trends in sulfur dioxide over the eastern United States. *Environmetrics* **11**: 373–393.
Hristopulos DT, Christakos G, Serre M. 1999. Implementation of a space transformation approach for solving the three-dimensional flow equation. *Society for Industrial & Applied Mathematics (SIAM)-Scientific Computing* **20**(2): 619–647.
Hristopulos DT, Christakos G. 2001. Practical calculation of non-Gaussian multivariate moments in BME analysis. *Mathematical Geology* **33**(5): 543–568.
Isaaks EH, Srivastava RM. 1989. *An Introduction to Applied Geostatistics*. Oxford University Press: New York, NY.
Journel AG, Huijbregts ChJ. 1978. *Mining Geostatistics*. Academic Press: London, UK.
Kibria GBM, Sun L, Zidek JV, Le ND. 2002. Bayesian spatial prediction of random space-time fields with application to mapping PM2.5 exposure. *Journal of the American Statistical Association* **97**: 112–124.
Kolovos A, Christakos G, Serre ML, Miller CT. 2002. Computational BME solution of a stochastic advection-reaction equation in the light of site-specific information. *Water Resources Research* **38**(12): 1318–1334.
Kolovos A, Christakos G, Hristopulos DT, Serre ML. 2004. Methods for generating non-separable spatiotemporal covariance models with potential environmental applications. *Advances in Water Resources* **27**: 815–830.
Kolovos A, Yu H-L, Christakos G. 2006. *SEKS-GUI v.0.6 User Manual*. Department of Geography, San Diego State University: San Diego, CA.
Kovitz J, Christakos G. 2004. Assimilation of fuzzy data by the BME method. *Stochastic Environmental Research & Risk Assessment* **18**(2): 79–90.
Law DC, Bernstein K, Serre ML, Schumacher CM, Leone PA, Zenilman JM, Miller WC, Rompalo AM. 2006. Modelling an early syphilis outbreak through space and time using the Bayesian maximum entropy approach. *Annals of Epidemiology* **16**(11): 797–804.
Lee S-J, Balling R, Gober P. 2008. Bayesian maximum entropy mapping and the soft data problem in urban climate research. *Annals of the Association of American Geographers* **98**(2): 309–322.

Liao D, Peuquet DJ, Duan Y, Whitsel EA, Dou J, Smith RL, Lin H-M, Chen J-C, Heiss G. 2006. GIS approaches for the estimation of residential-level ambient PM concentrations. *Environmental Health Perspectives* **114**(9): 1374–1380.

Mascarenhas MDM, Vieira LC, Lanzieri TM, Leal APPR, Duarte AF, Hatch DL. 2008. Anthropogenic air pollution and respiratory disease-related emergency room visits in Rio Branco, Brazil—September, 2005. *Jornal Brasileiro de Pneumologia* **34**(1): 42–46.

Mateu J, Porcu E, Gregori P. 2007. Recent advances to model anisotropic space-time data. *Statistical Methods & Applications* **17**: 209–223.

Olea RA. 1974. Optimal contour mapping using universal kriging. *Journal of Geophysical Research* **79**(5): 695–702.

Olea RA. 1999. *Geostatistics for Engineers and Earth Scientists*. Kluwer Academic Publishers: Boston, MA.

Orton TG, Lark RM. 2007. Accounting for the uncertainty in the local mean in spatial prediction by BME. *Stochastic Environmental Research and Risk Assessment* **21**(6): 773–784.

Papantonopoulos G, Modis K. 2006. A BME solution of the stochastic three-dimensional Laplace equation representing a geothermal field subject to site-specific information. *Stochastic Environmental Research and Risk Assessment* **20**(1–2): 23–32.

Parkin R, Savelieva E, Serre ML. 2005. Soft geostatistical analysis of radioactive soil contamination. In *GeoENV V Geostatistics for Environmental Applications*. Renard Ph (ed.). Kluwer Academic Publishers: Dordrecht, the Netherlands.

Pope CA, Thun MJ, Namboodiri MM, Dockery DW, Evans JS, Speizer FE, Heath CW. 1995. Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. *American Journal of Respiratory Critical Care Medicine* **151**: 669–674.

Porcu E, Mateu J, Saura F. 2008. New classes of covariance and spectral density functions for spatio-temporal modelling. *Stochastic Environmental Research and Risk Assessment* **22**(1): 65–79.

Puangthongthub S, Wangwongwatana S, Kamens RM, Serre ML. 2007. Modelling the space/time distribution of particulate matter in Thailand and optimizing its monitoring network. *Atmospheric Environment* **41**(36): 7788–7805.

Querido A, Yost R, Traore S, Doumbia MD, Kablan R, Konare H, Ballo A. 2007. Spatiotemporal mapping of total carbon stock in agroforestry systems of sub-Saharan Africa.In *Proceedings of ASA-CSSA-SSSA International Annual Meetings*. Nov. 4–8, 2007, New Orleans, Louisiana.

Saito H, McKenna SA, Zimmerman DA, Coburn TC. 2005. Geostatistical interpolation of object counts collected from multiple strip transects: Ordinary kriging versus finite domain kriging. *Stochastic Environmental Research and Risk Assessment* **19**(1): 71–85.

Samet JM, Dominici F, Zeger SL, Schwartz J, Dockery DW. 2000. The national morbidity, mortality, and air pollution study. Part II: morbidity, mortality and air pollution in the United States. *Report 94*. Health Effects Institute, Cambridge, MA.

Serre ML, Kolovos A, Christakos G, Modis K. 2003. An application of the holistochastic human exposure methodology to naturally occurring arsenic in Bangladesh drinking water. *Risk Analysis* **23**(3): 515–528.

Serre ML, Christakos G, Lee SJ. 2004. Soft data space/time mapping of coarse particulate matter annual arithmetic average over the US. In *Geoenv IV*. Kluwer Academic Publishers: Dordrecht, The Netherlands.

Smith RL, Kolenikov S, Cox LH. 2003. Spatio-temporal modeling of PM2.5 data with missing values. *Journal of Geophysical Research-Atmospheres* **108**(D24): STS11.1–STS11.11.

Tainio M, Tuomisto JT, Hänninen O, Päivi A, Koistinen KJ, Jantunen MJ, Pekkanen J. 2005. Health effects caused by primary fine particulate matter (PM 2.5) emitted from buses in the Helsinki metropolitan area, Finland. *Risk Analysis* **25**(1): 151–160.

Vyas V, Christakos G. 1997. Spatiotemporal analysis and mapping of sulfate deposition data over the conterminous USA. *Atmospheric Environment* **31**(21): 3623–3633.

Wibrin M-A, Bogaert P, Fasbender D. 2006. Combining categorical and continuous spatial information within the Bayesian Maximum Entropy paradigm. *Stochastic Environmental Research and Risk Assessment* **20**: 423–434.

Yu H-L, Christakos G, Bogaert P. 2008. Dealing with spatiotemporal heterogeneity: the generalized BME model. In *Progress in Spatial Analysis: Theory and Computation, and Thematic Applications*, Páez A, Le Gallo J, Buliung R, Dall'Erba S (eds). Springer Series on Advances in Spatial Sciences: New York, NY.

Yu H-L, Christakos G, Chen J-C. 2007b. Spatiotemporal air pollution modeling and prediction in epidemiologic research. In *Air Pollution Research Trends*, Bodine CG (ed.). Nova Science Publishers, Inc.: Hauppauge, NY; 57–75.

Yu H-L, Chen J-C, Christakos G, Jerrett M. 2009. BME estimation of residential exposure to ambient PM10 and Ozone at multiple time-scales. *Environmental Health Perspectives* **117**(4): 537–544.

Yu HL, Kolovos A, Christakos G, Chen JC, Warmerdam S, Dev B. 2007a. Interactive spatiotemporal modelling of health systems: the SEKS-GUI framework. *Stochastic Environmental Research and Risk Assessment* **21**(5): 555–572.

Zhang Q, Streets DG, He K, Klimont Z. 2007. Major components of China's anthropogenic primary particulate emissions. *Environmental Research Letters* **2**: 1–7.

Zidek JV. 1997. Interpolating air pollution for health impact assessment. In *Statistics in Environment 3: Pollution Assessment and Control*, Barnett V, Turkman KF (eds). Wiley: New York, NY; 251–268.

DOI: 10.1002/env