Stochastic modeling of atmospheric pollution: a spatial
time-series framework. Part II: application to monitoring
monthly sulfate deposition over Europe

Phaedon C. Kyriakidis*, André G. Journel

Department of Geological and Environmental Sciences, Stanford University, Stanford, CA 94305-2115, USA

Received 23 July 1999; accepted 11 November 2000

Abstract

A spatial time-series framework is adopted for stochastic modeling of monthly averaged sulfate deposition over Europe. The sulfur concentration data used in this study were measured at the European Monitoring and Evaluation Program (EMEP) monitoring network from January 1980 to December 1988. Parametric temporal trend and residual models, associated with long-term (linear trend or annual periodicity) and short-term (seasonal) concentration variability, respectively, are first established at monitoring stations. The resulting model parameters are regionalized in space to arrive at parametric trend and residual models at any unmonitored location. Stochastic simulation is performed for prediction and modeling of joint uncertainty regarding unknown sulfur concentration levels at unmonitored spatial locations and time instants. The case study illustrates the applicability of the proposed spatial time series framework to a real-world data set.

Keywords: Geostatistics; Space-time modeling; Stochastic simulation; Sulfate deposition; European Monitoring and Evaluation Program

1. Introduction

The spatial time-series framework proposed in the companion paper for modeling the spatiotemporal distribution of atmospheric pollution (Kyriakidis and Journel, 2001) is applied to the study of sulfate deposition. The pollutant concentration measurements used in this study consist of monthly averaged daily values of particulate sulfate SO$_4^{2-}$ dry deposition over Europe. The data were provided by the Norwegian Institute for Air Research (NILU), and were collected through the Cooperative Program for Monitoring and Evaluation of the Long-Range Transmission of Air Pollutants (EMEP) in Europe. A detailed description of the data set, which includes measurements of other chemicals, can be found in Schaug et al. (1987). Host et al. (1995) also used a portion of these data to illustrate a joint spatiotemporal modeling approach.

Sulfur concentration measurements (in $\mu$g(S) m$^{-3}$) are available at $n = 60$ monitoring stations from January 1980 to December 1988. The monitoring station locations, and the number of informed months per station, are shown in Fig. 1. The length of the time series at each station varies from 60 to 108 months (9 yr); approximately half of the concentration profiles have no missing values. The highest measurement density, both in space and time, is in Germany; fewer stations are in operation in southern Europe.

Three representative temporal concentration profiles are shown in Fig. 1. The vertical axis is the cumulative month, with the earlier (1st) month January 1980 plotted at the bottom, and the last (108th) month December 1988 plotted at the top of each graph. The horizontal axis is
sulfur concentration ranging from 0.0 to 10.0 μg(S) m⁻³. Periodicity can be detected in the concentration profile of station 43 (Norway), whereas such periodic variation is not evident in the other two profiles. The temporal profile at station 33 (France) exhibits a long-term decrease in concentration. Such reduction could be attributed to pollution control policies, which reduce emissions of sulfur dioxide in the atmosphere, and hence the levels of associated sulfates (Boubel et al., 1994).

The monthly concentration values are transformed into a data set following a normal distribution, using the rank preserving normal scores transform (Deutsch and Journel, 1998). Stochastic simulation is performed in this normal scores space. Sulfur concentration realizations are then obtained by back-transforming the simulated normal score values to the original variable space.

2. Spatiotemporal modeling

Sulfur concentration levels are modeled as a joint realization of a collection of spatially correlated time series \( \{Z(u,t), t \in T\} \), one per location \( u \in D \), with \( u = (x, y) \) denoting the 2D spatial coordinate vector, \( D \) the study area, and \( T \) the period of interest. Each time-series model is decomposed into a parametric trend \( \{M(u,t), t \in T\} \) and a parametric residual \( \{R(u,t), t \in T\} \) component. Analysis proceeds by first establishing component parameters at monitoring station locations \( u_i \), and then regionalizing these parameters in space (Kyriakidis and Journel, 2001). The trend component accounts for long-term depositional patterns, such as a linear decrease in concentration due to local emission-control policies or annual periodicity linked to

![Image](image_url)
seasonality in emissions and meteorological patterns; the residual component accounts for the remainder concentration variability.

2.1. Temporal trend models at monitoring stations

From spectral analysis (e.g., Chatfield, 1996) performed independently on each sample concentration profile \( \{c(u_i, t_i), t_i \in T_s \} \), the annual cycle was identified as the most dominant periodic component. The trend model for the normal scores of sulfur concentration profile at each station location \( u_i \) is thus written as

\[
m(u_i, t_i) = \sum_{k=0}^{3} a_k (u_i) f_k(t_i), \quad \forall \in (n), \quad t_i \in T_s
\]

\[
= b_0(u_i) + b_1(u_i) t_i + b_2(u_i) \cos \left( \frac{2\pi}{12} t_i \right)
\]

\[
+ b_3(u_i) \sin \left( \frac{2\pi}{12} t_i \right).
\]

where coefficients \( b_0(u_i) \) and \( b_1(u_i) \) are the intercept and slope of the temporal trend model \( m(u_i, t_i), i = 1, \ldots, T_s \). Intercept \( b_0 \)-values are associated with the magnitude of the temporal trend at the initial month January 1980, whereas slope \( b_1 \)-values quantify the rate of (linear) long-term decrease or increase in concentration.

Coefficients \( b_2(u_i) \) and \( b_3(u_i) \) are linked to the amplitude \( \phi(u_i) \) and the phase \( \phi(u_i) \) of the periodic component of the temporal trend model, as

\[
a(u_i) = \sqrt{[b_2(u_i)]^2 + [b_3(u_i)]^2},
\]

\[
\phi(u_i) = \tan^{-1} \left( -\frac{b_3(u_i)}{b_2(u_i)} \right).
\]

where the amplitude value \( a(u_i) \) is associated with the magnitude of the annual period component at location \( u_i \), and the phase value \( \phi(u_i) \) is associated with the month at which the trend-related concentration is minimum or maximum at that location.

Intensity coefficients \( b_k(u_i) \) are established via ordinary least squares (OLS), independently at each monitoring station location \( u_i \). Note that the specific algorithm (OLS) adopted fully defines the trend values, which are hereafter regarded as precise data. Plots of normal scores concentration profiles and trend model series at selected stations are shown in Fig. 2. Temporal trend models capture important aspects of the concentration temporal variability, and differ significantly from one station to another. The negative slope of the temporal trend profile in station 33 indicates a long-term decrease in sulfur concentration at that location over the 108 months. Correlation coefficient values between intensity \( b_k \)-data are generally low, except from a typical negative correlation between slope \( b_1 \) and intercept \( b_0 \)-values \( (\rho_{b_1b_0} = -0.56) \). This is inherent to any line fitting procedure: the greater the slope, the smaller the intercept of the fitted line.

Once temporal trend models \( \{m(u_i, t_i), t_i \in T \} \) are established at station locations via Eq. (1), the resulting trend \( b_k \)-coefficients are regionalized in space. This regionalization yields trend-related concentration profiles at any unmonitored location \( u \).

2.2. Simulation of spatiotemporal trend

A set of \( S = 50 \times 4 \) realizations \( \{b_k^{(s)}(u), u \in D\}, \)

\( s = 1, \ldots, S, k = 0, \ldots, K \), of temporal trend coefficients were generated on a \( 140 \times 180 \) grid, of cell size \( 25 \times 25 \) km\(^2\), via the procedure outlined in Section 2.3 of Kyriakidis and Journel (2001). Principal components of the four sets of coefficient \( b_k \)-values (not shown) were found to be orthogonal for all spatial lags. The hyperparameters (nugget, sill, and range), which specify the spatial semivariogram model \( \gamma_X(h) \) of each factor field \( \{X_k(u), u \in D\} \) were established via cross-validation (Deutsch and Journel, 1998); here \( h \) denotes a spatial lag separation vector. Simulated coefficient realizations were conditioned to the respective trend coefficient \( b_k \)-data at the 60 monitoring stations. One such joint realization of the four coefficients is shown in Fig. 3. Presentation of simulation results is limited to a polygonal area approximately enclosing points within the convex hull of the monitoring station locations.

Intercept \( b_0 \)-values in Fig. 3A are relatively higher in central/eastern Europe than in Scandinavia. Concentration profiles in some countries, e.g., France, exhibit long-term decrease (negative slope \( b_1 \)-values in Fig. 3B), whereas concentration profiles in Ireland show a long-term increase (positive slope \( b_1 \)-values). The average slope value over Europe (based on the 60 monitoring stations) is negative at \(-0.002\) month\(^{-1}\), indicating a slight long-term decrease of sulfur concentration levels for the time span 1980–88, possibly due to emission-control policies. Amplitude \( a \)-values (Fig. 3C) are more pronounced in Scandinavia than in central Europe, indicating a more significant annual periodic component in those areas. Trend-related concentration peaks typically in October–December (dark-colored phase \( \phi \)-values in Fig. 3D), whereas in southern Europe such peaks occur earlier in March–May (light-colored phase \( \phi \)-values). The sample correlation coefficient between any pair of local temporal trend coefficients is reproduced by the simulated realizations. For example, high simulated intercept values are associated with low simulated slope values, indicating a negative correlation between them.

2.2.1. Ensemble trend characteristics

At each node \( u \), the average \( \overline{b_k(u)} \) of \( S = 50 \) simulated coefficient values \( \{b_k^{(s)}(u), s = 1, \ldots, S\} \), generated via
sequential Gaussian simulation (SGSIM), is calculated as

$$\bar{b}_1(u) = \frac{1}{S} \sum_{s=1}^{S} b^{(s)}_1(u), \quad u \in D.$$ 

Maps of these average $\bar{b}_1$ -values are shown for the slope $b_1$ coefficient in Fig. 4A. Relatively fast rates of concentration decrease (low $\bar{b}_1$-values) can be found in northern France, Hungary, Romania, and Latvia; relatively fast rates of increase are more limited in space, and are found in Ireland, southern Spain, Italy, Czechia, and eastern Poland.

Similarly, at each grid node $u$, the local standard deviation $\sigma_1(u)$ of the $S = 50$ simulated coefficient values $\{b^{(s)}_1(u), s = 1, \ldots, S\}$ is calculated as

$$\sigma_1(u) = \sqrt{\frac{1}{S} \sum_{s=1}^{S} [b^{(s)}_1(u)]^2 - \bar{b}_1^2(u), \quad u \in D.}$$

These standard deviation $\sigma_1$-values for the slope $b_1$ coefficient are mapped in Fig. 4B. Standard deviation values increase away from the data locations, indicating higher uncertainty about the unknown trend coefficient at these nodes. More reliable estimates of either intercept or slope (lower standard deviation values) are found in central Europe, and Scandinavia, while such estimates in eastern Europe are less certain.

The local probability $p_b(u, b)$ for the unknown coefficient $b_1(u)$ to exceed any threshold value $b$ is calculated at each node $u$ as

$$p_b(u, b) = \frac{1}{S} \sum_{s=1}^{S} i_b^{(s)}(u, b), \quad u \in D,$$

where $i_b^{(s)}(u, b)$ is the indicator transform of the simulated coefficient $b^{(s)}_1(u)$, defined as $i_b^{(s)}(u, b) = 1$, if $b^{(s)}_1(u) > b$, zero otherwise. For example, $p_0(u; 0)$ is the probability for long-term concentration increase, given the available data and the parameters adopted for the simulation.

Such local probability $p_1$-values (for $b_1 = 0$) are mapped in Fig. 4C. Relatively, such high probabilities for positive slope are found in Ireland, southern Spain,
central Italy, Czechia, eastern Poland, and parts of central/northern Scandinavia. Areas like France, the Baltic, and the Balkans, exhibit low probability for long-term concentration increase.

On average over all stations, only 25% of the variance of the concentration profiles is accounted for by the temporal trend model. In certain areas in Scandinavia, the contribution of the trend model rises to 60% of the total variance of the sulfur concentration profile. This implies a significant importance of the residual component, which is spatially varying and should be accounted for in the modeling procedure.

2.3. Spatiotemporal residual modeling

The spatiotemporal residual field is modeled as a collection of spatially correlated residual time series (TS) \( \{ R(\mathbf{u}, t), t \in T \} \), one for each location \( \mathbf{u} \in D \). Simulation of the spatiotemporal residual field proceeds by generating simulated realizations of 1D residual profiles at any unmonitored location \( \mathbf{u} \) (Kyriakidis and Journel, 2001). Uncertainty regarding the temporal semivariogram model parameters at unmonitored locations is not accounted for in this study. The spatially varying importance of the residual component is quantified by
regionalizing in space the variance $s_R(\mathbf{u}_i)$ of the residual concentration profiles $\{r(\mathbf{u}_i, t), t \in T \}$ at the monitoring stations (Kyriakidis and Journel, 2001). Analysis then proceeds by quantifying the temporal correlation of the resulting residual profiles.

A set of $n = 60$ temporal semivariograms of standardized residual profiles are computed at each of the $n = 60$ monitoring stations, and modeled as

$$\gamma_R(\tau; \theta(\mathbf{u}_s)) = p_0(\mathbf{u}_s) + (1.0 - p_0(\mathbf{u}_s))\exp \left( \frac{-\tau}{q(\mathbf{u}_s)} \right),$$

where $q(\mathbf{u}_s)$ and $p_0(\mathbf{u}_s)$ denote the location $\mathbf{u}_s$-specific temporal range and nugget variance, respectively; $\tau = t - t'$ denotes a temporal lag distance. The parameter vector $\theta(\mathbf{u}_s)$ is comprised of the nugget component $p_0(\mathbf{u}_s)$ and the range coefficient $q(\mathbf{u}_s)$, which quantify the proportion of purely random temporal variability and the correlation period of the concentration profile at station location $\mathbf{u}_s$, respectively.

Simple kriging (SK) is performed for estimating the unknown parameter vector $\theta(\mathbf{u})$ at any unmonitored location $\mathbf{u}$, i.e., for deriving location-specific temporal
semivariogram models $\gamma_K(\tau; \theta^*(u))$. An estimate $s_K(u)$ of the standard deviation of the unknown residual profile $\{r(u,t), t \in T\}$ at any unmonitored location $u$ is also derived via SK. The hyper-parameters (nugget, sill, and range), which specify the spatial semivariogram models $\gamma_S(h)$, $\gamma_P(h)$, and $\gamma_Q(h)$ of the spatiotemporal residual parameter fields $\{S_K(u), u \in D\}$, $\{P_0(u), u \in D\}$, and $\{Q_1(u), u \in D\}$, were again established via cross-validation.

Maps of the SK-derived parameter fields are given in Fig. 5: note the typical smoothing effect of kriging. Residual concentration profiles in southwest Scandinavia and northern Baltic exhibit small $s_K$-values (Fig. 5A), i.e., local trend models in these areas account for a larger proportion of temporal variability in the sample concentration profiles. Highly variable residual profiles are located in Spain, and to a lesser extent in eastern Europe. A high nugget variance $p_0(u)$ in Fig. 5B, when associated with a small range $q_1(u)$ in Fig. 5C, indicates that, once the local temporal trend component $\{m(u,t), t \in T\}$ is removed, the variability of the resulting residual component is almost purely random. Such residual randomness is apparent in Scandinavia and northern France/southern England.

Cross-validation was also performed to check whether the assumption of multivariate normality for the standard deviation field $\{S_K(u), u \in D\}$ is approximately valid. The normal probability plot of the cross-validation errors (SK-derived $s_K$-estimate minus true $s_K$-value) at the $n = 60$ station locations is shown in Fig. 5D. A nearly

Fig. 5. Location-specific parameters of residual concentration profiles estimated by simple kriging (SK): (A) residual standard deviation $s_K(u)$, (B) nugget variance $p_0(u)$, (C) range $q_1(u)$ in months, and (D) normal probability plot of SK cross-validation errors for $s_K$. 
Fig. 6. Maps of mean simulated concentration in μg(S) m⁻³ (A,C), and probability for high sulfur concentration z(u) > 3.5 μg(S) m⁻³ (B,D), calculated from S = 50 simulated realizations for January and July 1982. Sample concentration histogram and semivariogram reproduction of a single realization for July 1982, (E) quantile–quantile plot of simulated versus sample concentration, (F) semivariogram of simulated (dotted line) and sample (bullets) concentration, both standardized to unit sill.
normal distribution can be postulated for such errors: the multivariate normal assumption for random field \( \{ S_{E}(u), u \in D \} \) does not appear refutable. Similar graphs were prepared (not shown) for all the parameters of the trend and residual components established at the monitoring stations, which lead to similar conclusions for the corresponding cross-validation error distributions. More elaborate procedures also exist in the literature for checking bivariate normality (Deutsch and Journel, 1998).

Conditional simulated realizations \( \{ \hat{p}^{(i)}(u, t_i), i = 1, \ldots, T \} \) of the standardized residual process \( \{ \hat{R}(u, t), t \in T \} \) at each node \( u \) are generated using SGSIM and spatially correlated standard normal deviates for Monte Carlo drawing at each “point” \( (u, t) \in D \times T \) (Kyriakidis and Journel, 2001). Simulated profiles \( \{ \hat{p}^{(i)}(u, t_i), i = 1, \ldots, T \} \) at each location \( u \) are then transformed into simulated non-standardized residual profiles \( \{ p^{(i)}(u, t_i), i = 1, \ldots, T \} \) at that location, by multiplying each simulated value \( \hat{p}^{(i)}(u, t_i) \) with the

---

**Fig. 7.** Sample sulfur concentration profiles (in \( \mu g(S) \ m^{-1} \)) at stations 45, 60, and 8 (Norway, Sweden, Germany), and simulated concentration profiles at nearby test nodes 1, 2, and 3.
corresponding SK-derived standard deviation $\sigma_g(\mathbf{u})$ shown in Fig. 5.

2.4. Final realizations of sulfur concentration

The independently simulated trend and residual realizations are added to produce realizations of the spatiotemporal normal scores field $\{Z(\mathbf{u}, t), \mathbf{u} \in D, t \in T\}$. These latter realizations are finally back-transformed into realizations of sulfur concentration (in units of $\mu$g($S$) m$^{-3}$), using the sample data histogram over all stations and time instants (see Deutsch and Journel (1998) for details on the back-transformation procedure).

The local means of $S = 50$ simulated concentration values $\{z(s)(\mathbf{u}, t), s = 1, \ldots, S\}$ generated at each node $\mathbf{u}$, and the local probability for concentration to exceed 3.5 $\mu$g($S$) m$^{-3}$ (corresponding to 90th percentile of the overall sample distribution) for January and July 1982, is mapped in Fig. 6. Relatively high concentrations for January (Fig. 6A) are found in northern Germany, Denmark, and to a lesser extent in Hungary. The spatial extent of such high-concentration areas is reduced in July 1982 (Fig. 6C), as expected from the limited emissions during summer months. This reduction in spatial extent is more evident on the probability maps, where the probability for sulfur concentration to exceed 3.5 $\mu$g($S$) m$^{-3}$ in July (Fig. 6D) is overall less than in January (Fig. 6B). The close reproduction of the sample concentration histogram and semivariogram from a single concentration realization for July 1982 is shown in Figs. 6E–F.

Realizations of the final back-transformed simulated concentration profiles for selected test nodes and nearby stations are shown in Fig. 7. The shape of the simulated profiles varies in space, as a result of the spatially varying trend and residual fields. Note the similarity of the simulated profiles with the measured concentration profiles at nearby station locations. Annual periodicity and long-term trends in concentration profiles, as quantified at monitoring stations, are reproduced by the simulated concentration profiles.

3. Discussion

The methodology proposed in the companion paper of Kyriakidis and Journel (2001) has been applied to the stochastic modeling of monthly averaged sulfate deposition over Europe during the years 1980–1988. Stochastic simulation within a spatial time series framework is performed for prediction and modeling of joint uncertainty regarding unsampled sulfur concentration levels at unmonitored spatial locations and time instants.

Simulated concentration profiles reproduce important spatiotemporal characteristics of the concentration field, as quantified at monitoring stations. The alternative realizations of sulfur concentration provide a model of the joint space–time uncertainty regarding the unknown levels of sulfate deposition over the domain and period of interest. In conjunction with additional information, such an uncertainty model can be used for improving data-collection strategies and for evaluating the risk associated with alternative decisions.

Acknowledgements

The authors would like to thank Dr. Gudmund Høst and the Norwegian Institute for Air Research (NILU) for providing the sulfate deposition data used in this study. Funding was provided by the Stanford Center for Reservoir Forecasting.

References


