SPATIAL-TEMPORAL MODELS

1 Introduction

Spatial-temporal models arise when data are collected across time as well as space. A typical example would be that of a monitoring network (of an atmospheric pollutant, say, or a network of meteorological stations) on which data are collected at regular intervals, say every day or every week. Thus the data analysis has to take account of spatial dependence among the monitors, but also that the observations at each monitor typically are not independent but form a time series. In other words, one must take account of temporal correlations as well as spatial correlations. Among the complications in such an analysis, few networks have exactly the same length of observation at each monitors, so one must take account of missing data, whether these are isolated missing data arising from the inevitable occasional failure to record a data value, or more systematic effects such as a deliberate decision to open a new monitor or close an existing monitor part way through the observation period; in that case, one may choose to treat all the observations before the monitor opened, or after it closed, as a block of missing data at that monitor.

Until recently, there has not been a theory of spatial-temporal processes separate from the already well established theories of spatial statistics and time series analysis. For example, Cressie (1993) devoted only 4 pages, out of nearly 900 in his book, to spatial-temporal models. However there has been a very rapid outgrowth of research in spatial-temporal data over the last decade, and our aim in this chapter is to survey some of that.

2 Background on spatial-temporal modeling

In this section we discuss a number of examples of spatial-temporal modeling, mostly taken from papers analyzing environmental data. These illustrate how the theory evolved essentially though numerous *ad hoc* analyses, paving the way for more recent studies that have resulted in more systematic models.

2.1 Acid rain in New York: Initial temporal analysis

Among the earliest papers on spatial-temporal statistics were a series of papers by Bilonick and Nichols (1983) and Bilonick (1983, 1985, 1988).

Bilonick and Nichols (1983) considered the analysis of rainfall data from 22 stations in or near New York State, collected from 1965 to 1979, and pooled into monthly values at each station. Variables measured included acidity (pH), and concentrations of sulfates, nitrates and calcium, as well as the amount of rainfall. The main question of interest was to determine whether the pattern of average monthly rainfall acidity had shown an increasing trend over the time period. In this paper, the main emphasis was not on spatial correlations among the stations, but on performing a time series analysis at each site to determine the trend.

To create a monthly time series, the pH values were first converted into hydrogen ion concentrations in milliequivalents per liter (meq ℓ^{-1}). Similar conversions were applied to the sulfate, nitrate and calcium concentrations. Then, the concentration in each month of each of these variables was multiplied by the total amount of rainfall to obtain a "total deposition series". The basic data set therefore consists of four series:

- $H_{x,t}$: total deposition of hydrogen ion in month t at location x,

- $S_{x,t}$: total deposition of sulfate in month t at location x,
- $N_{x,t}$: total deposition of nitrate in month t at location x,
- $C_{x,t}$: total deposition of calcium in month t at location x.

The units of all four series were cm meq ℓ^{-1}.
These were then aggregated across monitors into monthly temporal totals:

$$\begin{aligned} H_t &= \frac{1}{n_{H,t}} \sum_x H_{x,t}, \\ S_t &= \frac{1}{n_{S,t}} \sum_x S_{x,t}, \\ N_t &= \frac{1}{n_{N,t}} \sum_x N_{x,t}, \\ C_t &= \frac{1}{n_{C,t}} \sum_x C_{x,t}, \end{aligned}$$

where $n_{H,t}$ is the number of stations reporting hydrogen ion depositions in month t, and similarly for $n_{S,t}$, $n_{N,t}$ and $n_{C,t}$. In this way, a single time series was created for each variable. Although there were variations in the number of stations reporting in each month, these were considered to be of minor importance, and the aggregated series had the advantage of being a continuous time series extending across most of the 14.5 years observation period. For example, the H_t series consisted of 175 pooled monthly measurements derived from 1,436 total observations at the individual monitors, with similar sample sizes for the other three variables. Analysis of correlations among the series showed positive correlation between each pair of series except H_t and C_t for which the correlation was negative; however, the negative correlation was not statistically significant.

A significant part of their paper was devoted to the subject of "data screening". Plots of the raw data showed clear evidence of outliers, especially in the H_t series where there were several unaccountably large values in the latter part of the series. It was pointed out that the chemical analysis method had changed in 1976, with samples subsequently being analyzed in Atlanta (GA) rather than Albany (NY). It was suggested that contaminants getting into the samples might have been responsible for discrepant values. Based on the comparisons between different methods of chemical analyses, the authors proposed a method of data screening, to remove suspect values from the data set. The effect of this was to reduce the number of individual monthly values (for example, in the $H_{x,t}$ series the total number of observations dropped from 1,436 to 1,272, but there were still 175 monthly H_t values). The correlations among the four series were not much changed, except that the negative correlation between H_t and C_t was now larger in magnitude and statisticially significant.

After completing the data screening process, the authors analyzed all four series by Box-Jenkins time series methods. The fitted models were

$$\begin{aligned} H_t &= 0.14 + 0.37H_{t-1} + 0.30H_{t-12} + \epsilon_t, \\ S_t &= 0.19 + 0.41S_{t-1} + 0.19S_{t-11} + 0.17S_{t-12} + \epsilon_t, \\ N_t &= 0.24 + 0.36N_{t-1} + \epsilon_t, \\ C_t &= 0.16 + 0.20C_{t-1} + 0.23C_{t-9} + 0.25C_{t-14} + \epsilon_t, \end{aligned}$$

where the standard deviations of the residual series ϵ_t were respectively 0.21, 0.28, 0.14 and 0.18 for the H_t, S_t, N_t and C_t series. The inclusion of terms at or near lag 12 (months) reflects seasonality

in the data. All four series passed various standard tests for stationarity except that the N_t series analysis was confined to the period after June 1969, in which month, there had been a change of analysis method for N_t.

At the end of the paper, the authors concluded "The USGS data, we believe, overwhelmingly support the conclusion that there is no evidence for a long-term change in the mean level of acidity. The observed patterns in the hydrogen ion data can be completely explained in terms of a stationary ARIMA model." It should be pointed out, however, that they never formally tested for a long-term trend. For example, in Chapter 1 we studied methods for testing a linear trend in series of the form

$$X_t = \alpha + \beta t + \eta_t,$$

with $\{\eta_t\}$ a stationary time series, and we develop such techniques further in Chapter 7. Bilonick and Nichols carried out several tests for "interventions" (in other words, sudden jump shifts in the level of the process — they used such a test formally to identify the shift in the N_t process in 1969), but did not consider the possible consequences of a long-term linear (or non-linear) trend. Also, they apparently did not make any attempt to determine whether there were differences among the mean levels at the individual monitoring stations and whether this factor, combined with using different stations in computing different monthly values, could have contributed any bias of its own.

2.2 Acid rain in New York: Spatial analysis

In contrast with the preceding paper that just looked at temporal effects and ignored spatial differences, Bilonick (1983) considered a purely spatial analysis whose purpose was to construct, for each month within the data period, a map of mean H^+ (hydrogen ion) concentrations in units of meq ℓ^{-1}. Kriging methodology was used for this. The main steps in the calculation were as follows:

1. Time series were calculated for monthly total precipitation and deposition, based on the screened data described above. They were tested for stationarity and found to be stationary.

2. For the monthly precipitation series $P_{x,t}$ and deposition series $D_{x,t}$, semivariograms were calculated using the standard (method of moments) calculation of the semivariogram. Because each month typically contained data from only 5–10 stations, the semivariograms were aggregated across all months so as to use a single semivariogram for all the time periods. This aggregation ignores possible changes in the semivariogram with time, including seasonal changes. A very small number of individual $D_{x,t}$ values were found to be outliers and omitted from the analysis.

3. The spherical model was chosen as a parametric form for the semivariogram and was fitted to the empirical semivariograms, separately for the precipitation and deposition series, resulting in estimates for the nugget (non-zero), sill and range. The ranges were of the order of 150–200 km. The paper did not describe the method of fitting.

4. Kriging was attempted for each month of data, using the semivariogram that had just been fitted to all months of data. Both point and block averages estimates were computed, with correponding mean squared prediction errors based on the fitted semivariogram. By block averages, we mean here the average across 80 km^2 blocks, in which kriging was first performed on a grid of points within the block and then averaged over all grid points, as described in Chapter 2. A specific set of 12 blocks was used for most of the discussion. Most of the

discussion was for the block averages rather than the pointwise predictions as the latter were considered too variable to produce meaningful results.

5. Yearly estimates of the predicted block means were computed for each year and each month, along with corresponding mean squared prediction errors. For the latter, it is natural to consider variance estimates of the form

$$Var\left\{\frac{1}{12}\sum_{t=1}^{12} Y_t\right\} = f\sum_{t=1}^{12}\left\{\frac{Var(Y_t)}{12}\right\}, \tag{1}$$

where Y_t is the predicted value on a given block in month t. The constant f would be $\frac{1}{12}$ if the data were independent in time, but is presumably something greater than $\frac{1}{12}$ for a stationary time series. After an empirical comparison of yearly versus monthly variation, the paper used $f = 0.25$.

6. The results were then converted to results for H^+ concentration, defined to be D/P where D is deposition and P is precipitation. An approximate formula was used for the variance of the ratio.

7. The calculations of mean squared prediction errors for the block means were repeated on the assumption that the network consisted of just 12 stations, one centered at each of the blocks. This was viewed as a rough attempt at a network design problem, in other words, to compute how the kriging variances would change if the network was indeed redesigned as just described. In the present case, this resulted in reductions of mean squared prediction error of the order of 30%.

The resulting maps showed a "weak tendency" for predicted H^+ concentrations to decrease in moving from west to east, though rough tests of significance (based on a comparison of differences in predicted H^+ values with the mean squared prediction errors) suggested this was barely significant. As with the earlier discussion of time trends, no attempt was made at a formal test of trend, such as could have been made by means of a universal kriging procedure (including latitude and longitude as spatial covariates) in comparision with ordinary kriging.

2.3 Spatial-temporal analysis of sulfate depositions

Bilonick (1985) continued the discussion of Bilonick and Nichols (1983) and Bilonick (1983), by extending the earlier analysis to a fully spatial-temporal analysis. The initial steps in the analysis were are follows:

1. Two USGS networks were used — the 1965–1979 network for New York State, mentioned previously, supplemented by three months of data collected during 1980–1981, on 199 stations over the whole northeastern United States.

2. The variable of interest is either sulfate concentration, measured in milligrams per litre (mg ℓ^{-1}) or sulfate deposition, converted to kilograms per hectare per year (kg ha^{-1} y^{-1}). The same "data screening" method as previously was used to eliminate suspect measurements.

3. If the objective is to produce averaged or interpolated maps of sulfate concentration, the point was made that it is better first to average or interpolate total deposition and total rainfall, and then to divide the former by the latter, than to average or interpolate concentration directly.

The reason is that it would not be sensible to calculate an average concentration across many sites by unweighted averaging of the concentrations at the individual sites, when the amounts of rainfall at each site are greatly different. This point was reinforced by scatterplots of sulfate concentration against rainfall amounts, showing a clear negative correlation between the two variables.

4. For mapping purposes, the authors used the gnomonic projection (Richardus and Alder 1972), which converts great circles into straight lines. The point with latitude 42° N and longitude 82° W was used as the center of this projection.

5. For the 1965–1979 deposition series over New York State, a space-time semivariogram was computed of the form $\gamma(h,t)$, defined by

$$2\gamma(h,t) = E\{Z_{x_1,t_1} - Z_{x_2,t_2} : \ ||x_1 - x_2|| = h, \ |t_1 - t_2| = t\}, \tag{2}$$

where $||\cdot||$ denotes the usual Euclidean distance and it is part of the assumption that (2) is both stationary and isotropic in space, and stationary in time. To estimate (2), the standard sample semivariogram estimator was adapted in the obvious way, i.e. pairs of observations $(Z_{x_1,t_1}, Z_{x_2,t_2})$ were grouped into bins according to the values of h and t, and the semivariogram computed by the Method of Moments.

6. Some observations were deleted as apparent outliers and the semivariograms recomputed. In particular, one entire station (Allegany in western New York) was deleted after finding the seasonal pattern there to be completely different from all the other stations.

The empirical space-time semivariogram was then examined graphically to determine a suitable parametric form. The proposed form was

$$\gamma = \gamma_0 + \gamma_P + \gamma_S + \gamma_L \tag{3}$$

where

$$\begin{aligned}
\gamma_0 &= C_0, \\
\gamma_P &= C_P \left\{ 1 - 0.5 \cos\left(\frac{2\pi t}{365}\right) \right\}, \\
\gamma_S &= \begin{cases} 0 & \text{if } t = 0, \\ C_S\{3t(2a_S)^{-1} - 2^{-1}t^3 a_S^{-3}\}, & \text{if } 0 < t < a_S, \\ C_S & \text{if } t \geq a_S, \end{cases} \\
\gamma_L &= k_L h.
\end{aligned}$$

To interpret the different components of (3):

- γ_0 is the nugget effect
- γ_P represents a periodic effect in time, assuming time t is measured in days (Bilonick left out the divisor 365 but this was presumably an oversight),
- γ_S represented a residual (aperiodic) effect in time. It is interesting to note the departure from the Bilonick-Nichols (1983) paper, where temporal dependence was modeled using autoregressive processes: here, the authors decided to use a semivariogram of "spherical" structure.

- The spatial part of the variogram, γ_L, was modeled as linear in h based on the appearance of the empirical variogram.

Specific estimates were $C_0 = 100$, $C_P = 300$, $C_S = 150$, $a_S = 30$, $k_L = 1.0$. The resulting fitted semivariogram is illustrated in Fig. 1.

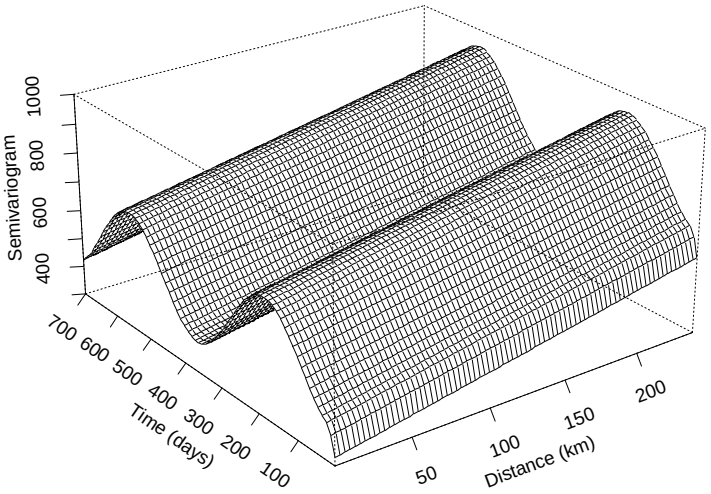

Fig. 1. Fitted spatial-temporal semivariogram.

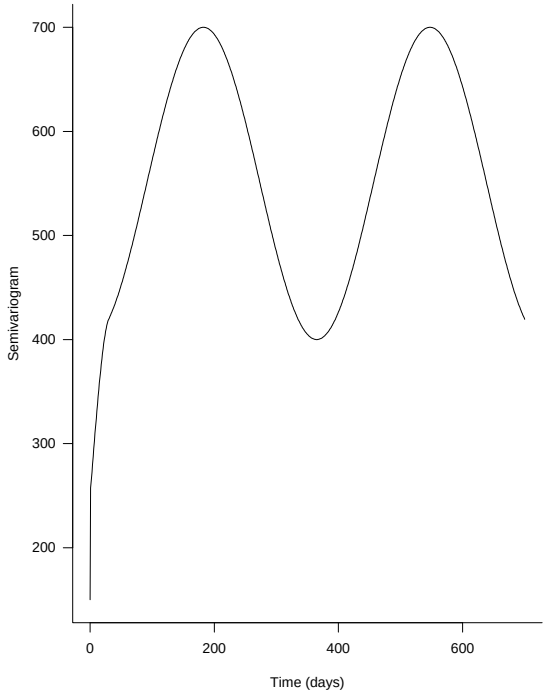

Fig. 2. The temporal variogram for a single location, $\gamma(0, t)$

To create interpolated maps from (3), the standard kriging equations were applied, but interpretated in a space-time setting, to calculate estimated annual averages over each of the 80 km^2

grid blocks mentioned earlier. This means that in principle, the prediction for any one grid block should be a weighted average over all space and time points of the entire data set. In practice, observations considered to contribute very small weight were excluded: the actual prediction for sulfate deposition for a particular year in a particular grid block was based on observations within 240 km. of the center of the block and within 180 days of the beginning and ending time points of the block.

The resulting maps for each of the years 1966–1978 showed clear differences in the spatial pattern of deposition from year to year, but still no evidence of an overall temporal trend. A plot of spatially and temporally weighted yearly averages showed a small peak around 1972, but no long-term increasing or decreasing trend.

The last part of Bilonick (1985) considered the 1980–1981 data set over the whole northeastern United States, but we shall not consider this in detail since it was a purely spatial analysis (based on a single time snapshot of the data) and is therefore of less methodological interest. A universal kriging algorithm was employed including a linear trend in latitude and longitude, to allow for an apparent north-to-south drift in the level of sulfate deposition. The residual spatial component was modeled with a combination of nugget, a spherical spatial variogram, and a linear spatial variogram. Based on this, maps of predicted sulfate deposition (aggregated on 80 km^2 blocks), and the corresponding mean squared prediction errors, were produced.

2.4 Space-time maps of H^+ depositions via indicator kriging

Bilonick (1988) analyzed data from another network (the Precipitation Quality Network, or PCN) which collected data on a wide variety of pollutants, across much of the U.S., but with the greatest density of monitors in the northeastern United States. This particular analysis was based on 35 stations in the states of NY, PA, WV, VA, OH, IN, KY, IL plus Ontario (Canada), and the time period from July 1982 to September 1984. About 20 stations were reporting in any given week. The main variable of interest was total hydrogen ion (H^+) deposition, computed from acidity (pH) measurement by first converting to H^+ concentration and then multiplying by the total volume of rainfall to calculate a deposition level.

Initial data analysis consistent of calculating weekly time series of mean precipitation and H^+ deposition averaged over the whole network. To correct for the inhomogeneous distribution of monitors, a "spatial declustering" technique was used: in essence, monitors were grouped into 300 km^2 blocks, block averages were computed by averaging the data over all monitors within each block, and then the block averages were equally weighted to compute overall spatial averages. A similar technique was used to compute an overall CDF (cumulative distribution function) for H^+ deposition. For the latter, a 2-year period was used (instead of the full 26 months of data) to avoid bias caused by seasonal variation, though it is questionable whether the correct way to deal with seasonal variation (in calculating a CDF) is simply to aggregate over all seasons. However, the main purpose of this was to divide the ordered H^+ data into ten equal-sized bins in order to calculate deciles, i.e. the quantiles corresponding to the 0.1, 0.2,...,0.9 cumulative probabilities, and for this purpose, aggegrating across seasons is probably reasonable.

The next step in the analysis was an application of "indicator kriging" (Cressie 1993, pp. 281–283, Chilès and Delfiner 1999, pp. 382–384). The basic idea of this method is to construct indicator variables

$$I_k(x,t) = \begin{cases} 1 & \text{if } D_{x,t} \le d_k, \\ 0 & \text{otherwise,} \end{cases}$$

for $k = 1, 2, ..., 9$, where $D_{x,t}$ is the H^+ deposition at location x and time t, and $d_1, ..., d_9$ are the nine decile values estimated from the aggregated CDF as just described.

The basic idea of indicator kriging is to construct predicted values $I_k^*(x,t)$ say, at all space-time coordinates (x,t), using standard kriging methodology applied to the indicator variables $I_k(x,t)$ for each value of k. There are some disadvantages and inconsistencies of this procedure: for example, if ordinary linear kriging is used, there is no *a priori* guarantee that $0 \leq I_k^*(x,t) \leq 1$ for all (x,t), or that $I_{k'}^*(x,t) \leq I_k^*(x,t)$ whenever $k' < k$. Advocates of the method argue that these disadvantages are outweighted by the advantages of being able to estimate a full distribution function for the predicted variables, the absence of any assumption of a Gaussian process, and the resistence of the method to outliers.

For the application in Bilonick (1988), it was necessary first to estimate a space-time variogram of $I_k(x,t)$ for each $k = 1, ..., 9$. This was constructed, as in Section 2.3, by estimating a semivariogram $\hat{\gamma}_k(h,t)$ by the usual binning technique, where h is spatial distance and t is time, but from the indicator variables $I_k(x,t)$ rather than the raw data points $D_{x,t}$. It is an implicit (untested) assumption here that the semivariogram is isotropic in space. At this point, however, Bilonick made a number of plots of the spatial-temporal $\hat{\gamma}_k(h,t)$ in order to determine whether *this* function was isotropic — in other words, whether $\hat{\gamma}_k(h,t)$ was a function of $\sqrt{h^2+t^2}$. The conclusion was that it was not, except possibly for $k = 9$ (the highest level). After further trial and error, for each k Bilonick (1988) fitted a model of the form

$$\gamma = C_r\gamma_r + C_h\gamma_h + C_t\gamma_t \tag{4}$$

where γ_r is a spherical semivariogram in the variable $r = \sqrt{gh^2+t^2}$ (g is an estimated anisotropy parameter), γ_h is a spherical semivariogram in h, γ_t is a spherical semivariogram in t, and C_r, C_h and C_t are constants corresponding to the sills of the three semivariograms. Thus the fitted semivariogram is a sum of a pure spatial semivariogram, a pure temporal semivariogram, and one that is effectively of "geometrically anisotropic" form in the space and time variables.

As an example, Fig. 3 shows the fitted surface (4) for two of the threshold levels conisdered by Bilonick: (a) $k = 5$, for which the parameters are $g = 3$, $C_r = 0.2$, $\phi_r = 300$, $C_h = 0.02$, $\phi_h = 300$, $C_t = 0.03$, $\phi_t = 300$, (b) $k = 9$, for which the parameters are $g = 0.5$, $C_r = 0.1$, $\phi_r = 65$, $C_h = 0.01$, $\phi_h = 65$, $C_t = 0.06$, $\phi_t = 65$. Here ϕ_r, ϕ_h and ϕ_t are the ranges of the respective semivariograms γ_r, γ_h and γ_t. As can be seen, the variogram in (a) increases slowly and almost linearly, whereas that in (b) (plotted across the same range of spatial and temporal distances) reaches its sill much more quickly.

From the point of view of the general theory of variograms outlined in Chapter 2, (4) is of the "zonal" form in three-dimensional space-time coordinates, the fitted semivariogram being represented as a sum of three semivariograms each corresponding to a different linear combination of the coordinates. Viewing the problem in this way, there are many possible generalizations of (4). Another comment that cannot be avoided at this point is that Bilonick plotted a number of empirical space-time variograms in his paper, and they do not have the appearance of smooth monotone functions that Fig. 3 suggests! So the question of whether this model actually fits the data, still more whether it is an optimal fit, must be regarded as not fully resolved.

Following the indicator kriging procedure, the nine values of $I_k^*(x,t)$ for $k = 1, ..., 9$ for a given (x,t) may be interpolated to construct an estimated CDF of H^+ values at that (x,t). In practice the x values were aggregated into 50 km^2 grid blocks and the t values into months. Based on the estimated CDFs, an estimate for any desired quantile of the distribution may be obtained at any (x,t). In this study, the median was used — in other words, for each month, an estimate was constructed of the median H^+ level in each 50 km^2 grid block, and maps of the resulting estimates were constructed.

The final question considered by Bilonick (1988) was whether it was possible to improve the location of monitors. The following procedure was suggested. First, pick a threshold that represents

a critical level of risk. Then, choose an optimal estimator at each point based on a suitable loss function. For example, one such estimator might be to apply indicator kriging to construct predictors $I_k^*(x,t)$ of the probability that H^+ is below a threshold d_k, and to classify the point (x,t) as below the threshold if $I_k^*(x,t) > \frac{1}{2}$ and above the threshold if $I_k^*(x,t) \leq \frac{1}{2}$. We use the estimated CDFs to determine the probability that (x,t) is misclassified. According to Bilonick, "areas with a high probability of misclassification should be candidates for additional sampling."

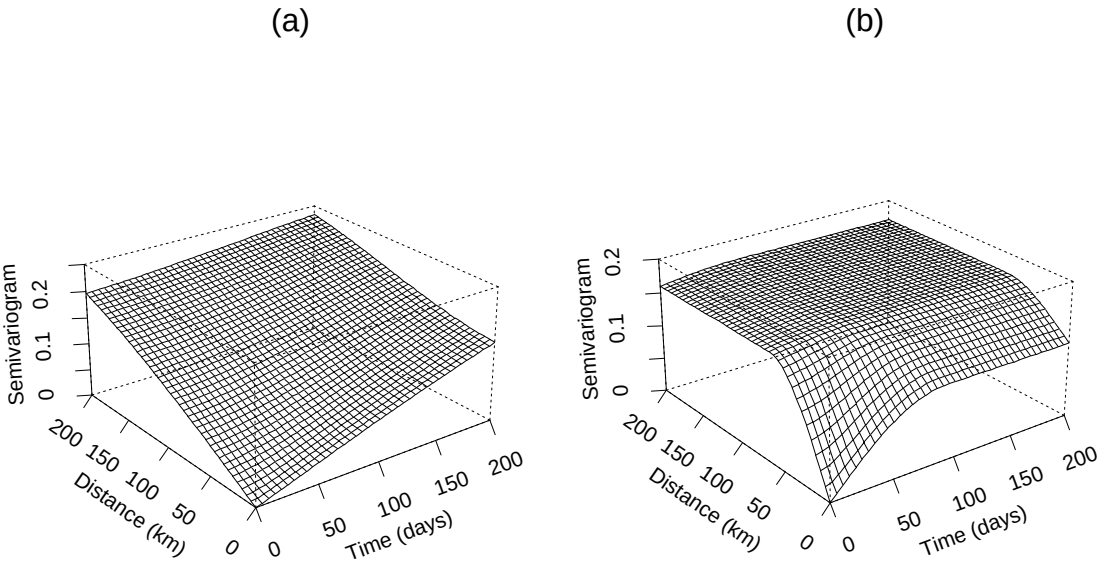

Fig. 3. Fitted spatial-temporal semivariogram for Bilonick's example of indicator kriging via (4). (a) $k = 5$ (indicator threshold is median of distribution), (b) $k = 9$ (indicator threshold is 90th percentile of distribution).

To summarize this whole sequence of papers:

1. Bilonick and Nichols (1983) performed a time series analysis of spatially aggregated H^+ and other deposition series, primarily to examine whether there was any trend. They did not formally test for a trend, but concluded that the time series was stationary, which they interpreted as absence of trend. The analysis did not take account of possible biases due to spatial inhomogeneities.

2. Bilonick (1983) performed a spatial analysis month by month, and also aggregated into yearly averages. The main tool was kriging based on the spherical model of a spatial semivariogram, applied separately to the deposition and precipitation series, and then to their quotient to form estimates of concentration. Some simple rules for network design were also given.

3. Bilonick (1985) presented the first truly spatial-temporal analysis, in which the space-time variogram was fitted by the function (3) (see also Fig. 1). This was used to construct annual maps of sulfate depositions for each year of the study.

4. Finally, Bilonick (1988) provided a similar analysis of H^+ depositions, but using indicator kriging. In this case the estimated spatial-temporal variogram took the form (4), and was used to construct maps of the estimated median H^+ deposition (but could equally well have been used to construct maps of any other quantile of the distribution). The "design of network" implications were also noted in this case.

2.5 The model of Egbert and Lettenmaier

Egbert and Lettenmaier (1986) developed a rather general class of multivariate space-time models, motivated by the National Atmospheric Deposition Program (NADP). This is a spatial network consisting of weekly observations of 10 ionic species. Features of the data set include:

- Temporal structure at both short and long time scales;
- Both the long-term and the short-term time averages exhibit spatial dependence;
- Seasonal effects;
- Multiple components — need for multivariate models;
- Many short-term time series, lots of missing data.

For dealing with seasonal effects, they divided each year into four 3-month seasons (pointing out that alternative definitions of "seasons" are possible) and fitted the following temporally stationary model to the data in each season.

The basic model equation is

$$Z_{st}^p(x) = W_{st}^p(x) + Y_s^p(x) + M^p(x), \tag{5}$$

in which $Z_{st}^p(x)$ represents the pth component of the observed process in year s, week t and location x (assume $1 \le p \le P,\ 1 \le s \le S,\ 1 \le t \le T$), $W_{st}^p(x)$ is a process of weekly variation, $Y_s^p(x)$ is a process of yearly variation, and $M^p(x)$ represents long-term effects. We assume $W_{st}^p(x)$ and $Y_s^p(x)$ have mean 0, and

$$E\{W_{st}^p(x)W_{st}^{p'}(x')\} = \kappa_W^{pp'}(x, x') \tag{6}$$

assumed to be a smooth function of either $x - x'$ (spatially stationary case) or $|x - x'|$ (stationary and isotropic).

For the full spatial-temporal covariance structure, Egbert and Lettenmaier assumed

$$E\{W_{st}^p(x)W_{s't'}^{p'}(x')\} = \delta_{ss'}\kappa_W^{pp'}(x - x', t - t'), \tag{7}$$

$$E\{Y_s^p(x)Y_{s'}^{p'}(x')\} = \delta_{ss'}\kappa_Y^{pp'}(x - x'), \tag{8}$$

where $\delta_{ss'}$ is the Kronecker delta function (=1 if $s = s'$, otherwise 0). Thus, it is assumed that the W and Y processes in different years are uncorrelated, and it is also assumed that the two processes are uncorrelated with each other.

For the long-term mean process $M^p(x)$, it is assumed

$$E\{M^p(x)\} = \mu^p, \tag{9}$$

$$\frac{1}{2}E[\{M^p(x) - M^p(x')\}\{M^{p'}(x) - M^{p'}(x')\}] = \gamma^{pp'}(x - x'). \tag{10}$$

Next we discuss the estimation technique. This is in several parts:

First model: time-independent weekly effects

In (7), assume

$$\kappa_W^{pp'}(x, t) = \begin{cases} \kappa_W^{pp'}(x), & \text{for } t \neq 0, \\ 0, & \text{for } t \neq 0. \end{cases} \tag{11}$$

Egbert and Lettenmaier commented that this seems to be quite realistic for the weekly NADP data.

Writing Z^p_{sti} in place of $Z^p_{st}(x_i)$ where x_i is the ith sampling location, they defined temporal averages in the obvious way (allowing for the possibility of missing data)

$$\begin{aligned} \bar{Z}^p_{si} &= \frac{\sum_t Z^p_{sti}}{T_{si}}, \\ \bar{Z}^p_i &= \frac{\sum_s \sum_t Z^p_{sti}}{\sum_s T_{si}}, \end{aligned}$$

where T_{si} is the number of observed data points at x_i in year s. The structure of the data is similar to that of a three-way analysis of variance and the estimation method exploits that analogy:

1. For each pair of points $x_i, x_{i'}$, estimate $\kappa_W^{pp'}(x_i - x_{i'})$ by

$$\hat{\kappa}_W^{pp'}(x_i - x_{i'}) = \frac{\sum_s \sum_t (Z^p_{sti} - \bar{Z}^p_{si})(Z^{p'}_{sti'} - \bar{Z}^{p'}_{si'})}{a^{pp'}_{ii'}} \tag{12}$$

   where the sum is taken over all s, t pairs for which the summand is available and $a^{pp'}_{ii'}$ is a corresponding degrees of freedom calculation, defined so as to make (12) unbiased;

2. For yearly deviations from the long-term mean, we have

$$E\left\{\sum_s (\bar{Z}^p_{si} - \bar{Z}^p_i)(\bar{Z}^{p'}_{si'} - \bar{Z}^{p'}_{i'})\right\} = b^{pp'}_{ii'} \kappa_Y^{pp'}(x_i - x_{i'}) + c^{pp'}_{ii'} \kappa_W^{pp'}(x_i - x_{i'}) \tag{13}$$

   where $b^{pp'}_{ii'}$ and $c^{pp'}_{ii'}$ may again be explicitly calculated; hence an unbiased estimate of $\kappa_Y^{pp'}(x_i - x_{i'})$ is

$$\hat{\kappa}_Y^{pp'}(x_i - x_{i'}) = \frac{\sum_s (\bar{Z}^p_{si} - \bar{Z}^p_i)(\bar{Z}^{p'}_{si'} - \bar{Z}^{p'}_{i'}) - c^{pp'}_{ii'} \hat{\kappa}_W^{pp'}(x_i - x_{i'})}{b^{pp'}_{ii'}} \tag{14}$$

3. Similarly for the long-term means, one can calculate

$$\begin{aligned} E\left\{(\bar{Z}^p_i - \bar{Z}^p_{i'})(\bar{Z}^{p'}_i - \bar{Z}^{p'}_{i'})\right\} &= 2\gamma^{pp'}(x_i - x_{i'}) + d^{pp'}_{ii'} \kappa_W^{pp'}(0) + e^{pp'}_{ii'} \kappa_W^{pp'}(x_i - x_{i'}) \\ &\quad + f^{pp'}_{ii'} \kappa_Y^{pp'}(0) + g^{pp'}_{ii'} \kappa_Y^{pp'}(x_i - x_{i'}) \end{aligned} \tag{15}$$

   for explicit constants $d^{pp'}_{ii'}$, $e^{pp'}_{ii'}$, $f^{pp'}_{ii'}$, $g^{pp'}_{ii'}$, and hence calculate an unbiased estimate of $\gamma^{pp'}(x_i - x_{i'})$ in similar fashion to the way (14) was derived from (13).

Smoothing the autocovariances

The approach described so far has the disadvantage that each value of $\kappa_W^{pp'}(x_i - x_{i'})$ (and similarly for κ_Y and γ) is estimated entirely seperately from every other value, so a plot of $\kappa_W^{pp'}(x_i - x_{i'})$ against $|x_i - x_{i'}|$ will not be a smooth function of $|x_i - x_{i'}|$. To remedy this, Egbert and Lettenmaier proposed what was in effect a kernel smoother; for example,

$$\hat{\kappa}_W^{pp'}(r) = \frac{\sum_{i,i',s,t} \phi(r - |x_i - x_{i'}|)(Z^p_{sti} - \bar{Z}^p_{si})(Z^{p'}_{sti'} - \bar{Z}^{p'}_{si'})}{\sum_{i,i'} \phi(r - |x_i - x_{i'}|) a^{pp'}_{ii'}} \tag{16}$$

where ϕ is some kernel function, typically a function that has a peak at 0 and is 0 outside some finite interval, say $(-r_0, r_0)$ for some r_0. As kernel functions they considered

$$\phi(r) = \begin{cases} 1, & \text{if } |r| \le r_0, \\ 0, & \text{otherwise,} \end{cases}$$

or

$$\phi(r) = \begin{cases} 1 - \frac{|r|}{r_0}, & \text{if } |r| \le r_0, \\ 0, & \text{otherwise,} \end{cases}$$

though of course they could have considered other kernel functions.

The estimator (16) assumes isotropy; there is an obvious analog for anisotropic variograms, which we shall not state explicitly. Also, the method extends to smoothed estimation of κ_Y and γ in similar fashion. In their discussion, Egbert and Lettenmaier suggested that it might be feasible to make an anisotropic estimate of κ_W but not of κ_Y and γ for which there are far fewer data points.

Second model: including temporal autocorrelation

For the case in which it is not assumed that the weekly effects are temporally independent, Egbert and Lettenmaier nevertheless assumed that they are independent at all lags greater than somne T_0. They then derived a series of equations of the form

$$E\left\{\sum_{s=1}^{S}\sum_{t'=1}^{T-t}(Z^p_{st'i} - \bar{Z}^p_{si})(Z^{p'}_{s(t+t')i'} - \bar{Z}^{p'}_{si'})\right\} = \sum_{t'=0}^{T} a^{pp'}_{ii'tt'}\kappa^{pp'}_W(x_i - x_{i'}, t) \tag{17}$$

for $0 \le t \le T_0$, where $a^{pp'}_{ii'tt'}$ are explicit constants; by treating (17) as a system of $T_0 + 1$ linear equations in $T_0 + 1$ unknowns, they were then able to derive unbiased estimators of $\kappa^{pp'}_W(x_i - x_{i'}, t),\ 0 \le t \le T_0$ as linear combinations of the quantities

$$\sum_{s=1}^{S}\sum_{t'=1}^{T-t}(Z^p_{st'i} - \bar{Z}^p_{si})(Z^{p'}_{s(t+t')i'} - \bar{Z}^{p'}_{si'})$$

for $t = 0, ..., T_0$.

The estimators of κ_Y and γ also need some adjustment to allow for the autocorrelation of the weekly effects but we shall not indicate that explicitly here.

This procedure could be followed by kernel smoothing of the estimates, exactly as for the time-independent case.

Discussion and Application

This direct approach to decomposing a spatial-temporal covariance function has the advantage of relative simplicity — at least in comparison with full maximum likelihood and Bayesian approaches — and the usual merits of a nonparametric approach, in that the method does not require or depend on the assumption of a parametric model and hence we do not have to worry about misspecification of the model. However, there are some disadvantages. The use of simple moment estimates as in (14) can sometimes result in negative estimates of variance — as pointed out by Egbert and Lettenmaier, this is analogous to random effects ANOVA in which the method of moments does sometimes produce negative estimates (this could be avoided by using maximum likelihood estimates but at a corresponding increase in computational complexity). A second difficulty, not directly discussed by Egbert and Lettenmaier, is that since there is no guarantee that the resulting covariance estimates are positive definite, there is a danger that negative prediction variances might occur when doing

kriging with the estimated covariances. This could be avoided by using alternative nonparametric estimators guaranteed to produce positive definite covariance estimates (see Section 3.3.2), but again at the cost of increased computational complexity.

In the application of this method, Egbert and Lettenmaier discussed a network of 51 sites in the northeast U.S., with two years of data (1980 and 1981) subdivided into four seasons — January to March as winter, April to June as spring, and so on. They considered three variables: pH, precipitation and sulfate acidity. They discussed data quality issues and proposed a criterion to eliminate unreliable observations (recall earlier discussion by Bilonick and Nichols (1983) of the same point). The main steps of the rest of the analysis were as follows:

- Transformation of data — square roots were taken of both precipitation and sulfate concentration to improve fit to a normal distribution. No transformation was taken of pH but, as they remarked, pH is already impicitly on a logarithmic scale.
- Temporal autocorrelation — some mild temporal autocorrelation was seen in the precipitation data, modeled using (17) with $T_0 = 1$. No autocorrelation was seen for the other two variables.
- For the weekly spatial autocovariances, both isotropic and anisotropic estimates were computed but it was remarked that "basic features of the spatial structure are seem most simply from the isotropic estimates".
- There was little "yearly" effect once the weekly and long-term spatial covariances were computed, and it was suggested that the yearly component could be dropped (for this data set — a difference conclusion might be drawn with more than two years' data).
- There were significant differences among seasons – for example, in most cases the longest spatial ranges were seen in the winter data, the shortest in summer, with spring and fall estimates in between.
- Spatial correlations for sulfate concentrations were stronger than for pH (possibly due to measurement error or laboratory bias in the pH analyses).
- There was some anisotropy in the preciptation data with a tendency for spatial correlations to be higher in the east-west direction. This effect was less pronounced in summer than the other three seasons, possibly because of different patterns of precipitation in the different seasons (more frontal storms in winter, more convective rain events in summer).

The preceding conclusions are for looking at the three variables one at a time. For multivariate results, the authors considered the following:

- Direct correlations among the three variables — precipitation and sulfate concentration were negatively correlated, as might be expected because precipitation tends to dilute the sulfate concentration. No correlation was observed between pH and the other two variables.
- The question then arose of whether the spatial correlation in sulfate concentration could be *solely* explained by a rainfall dilution hypothesis. To examine this, the authors constructed a 4×4 covariance matrix (superscipts p and s denote precipitation and sulfate concentration respectively),

$$\Sigma = \begin{pmatrix} \kappa_W^{pp}(0) & & & \\ \kappa_W^{pp}(r) & \kappa_W^{pp}(0) & & \\ \kappa_W^{ps}(0) & \kappa_W^{ps}(r) & \kappa_W^{ss}(0) & \\ \kappa_W^{ps}(r) & \kappa_W^{ps}(0) & \kappa_W^{ss}(r) & \kappa_W^{ss}(0) \end{pmatrix}$$

which they also partitioned as

$$\Sigma = \begin{pmatrix} \Sigma_{pp} & \Sigma_{ps} \\ \Sigma_{sp} & \Sigma_{ss} \end{pmatrix}.$$

The conditional covariance matrix of the sulfate measurements given precipitation is therefore

$$\Sigma_{ss} - \Sigma_{sp}\Sigma_{pp}^{-1}\Sigma_{ps}. \tag{18}$$

Examination of (18) for different x showed that there was still significant correlation between $Z^s(x)$ and $Z^s(0)$ even after conditioning on $Z^p(x)$ and $Z^p(0)$; in other words, the spatial structure of sulfate concentration could not be explained solely by the dilution effect due to precipitation.

- Another aspect of this comparison is the effect on prediction of say $Z^s(x)$ given (a) $Z^p(x)$ alone, (b) $Z^s(0)$ and $Z^p(0)$ (c) all three of $Z^p(x)$, $Z^s(0)$ and $Z^p(0)$. Comparisons of prediction variances showed that (c) is about as good as (b) for small $|x|$ and about as good as (a) for large $|x|$, but that for intermediate $|x|$, (c) was clearly better than both (a) and (b). The message here was that optimal prediction of sulfate at a site should depend both on rainfall at the same site, and on measurements of rainfall and sulfate at a remote site — neither source of variability (spatial or that due to the sulfate-precipitation correlation) could be "explained" in terms of the other.

2.6 Stein's model for spatial-temporal processes

Stein (1986) considered a model of the form

$$z(x,t) = m(x) + \mu(t) + e(x,t) \tag{19}$$

observed at kn space-time points (x_i, t_j), $i = 1, ..., n$, $j = 1, ..., k$. In Stein's formulation, $\mu(t_1), ..., \mu(t_k)$ are fixed unknown time constants (not modeled by any random process) while $e(x,t)$ and $m(x)$ are both random processes satisfying

$$\begin{aligned} E\{e(x,t)\} &= 0, \\ \frac{1}{2}E[\{e(x,t) - e(x',t)\}^2] &= \gamma(x - x'), \\ E[\{e(x,t_i) - e(x',t_i)\}\{e(x'',t_j) - e(x''',t_j)\}] &= 0,\ i \neq j, \\ E\{m(x)\} &= 0, \\ \frac{1}{2}E[\{m(x) - m(x')\}^2] &= \eta(x - x'). \end{aligned}$$

Thus the basic point is that we have a superposition of two zero-mean noise processes, one a spatial process m with semivariogram η that is common to all time points, the other a spatial process e with semivariogram γ that is generated independently at each time point. The semivariograms η and γ are assumed known.

Within this framework, Stein considered two prediction problems,

1. Predict $z(x_0, t_\beta)$ for any arbitrary $x_0 \notin \{x_1, ..., x_n\}$ for any observed time point t_β, $\beta \in \{1, ..., k\}$.

2. Predict a difference of spatial averages at two time points

$$\frac{1}{|R|}\int_R \{z(x,t_\alpha) - z(x,t_\beta)\}dx, \quad \alpha,\beta \in \{1,...,k\}, \tag{20}$$

where R is some region and $|R|$ denotes the area of R.

A key point is that although direct solution of the kriging equations would require inversion of a $(kn) \times (kn)$ matrix, the specific structure of (19) actually permits the problem to be solved with the inversion of only $n \times n$ matrices.

For the solution of problem 1, Stein showed that the optimal predictor $\hat{z}(x_0, t_\beta)$ is a linear combination of z_{t_β}, the vector of observations at time t_β, and the vector of time-averaged responses, $\bar{z} = \frac{1}{k}\sum_{i=1}^{k} z(t_i)$. Under certain circumstances (if the dominant source of spatial variability is m rather than e), this can perform substantially better than an estimator based on z_{t_β} alone.

For problem 2, Stein pointed out that the optimal kriging solution is based only on pairwise differences, $z(x_i,t_\alpha) - z(x_i,t_\beta),\ i = 1,...,n$, and he compared this with the alternative solution in which both $\frac{1}{|R|}\int_R\{z(x,t_\alpha)$ and $\frac{1}{|R|}\int_R\{z(x,t_\beta)$ are predicted from z_{t_α} and z_{t_β} respectively, in effect treating z_{t_α} and z_{t_β} as independent. The latter procedure often gives a reasonable predictor but may completely misstate the prediction variance. Stein likened this to the difference between a two-sample t test and a paired comparison t-test: optimal prediction of (20) is like a paired comparison t test in which we take advantage of the possibility that two observations at the same location but different time points may be very similar, whereas the procedure using separate predictors of $\frac{1}{|R|}\int_R z(x,t_\alpha)dx$ and $\frac{1}{|R|}\int_R z(x,t_\beta)dx$ is like a two-sample t test where this feature is ignored.

3 The Repeated Measurements and Separable Models

A widely studied class of spatial-temporal models (usually attributed to Rodriguez-Iturbe and Mejia (1974)) is the so-called separable model, in which the space-time covariance function or variogram can be written as a product of two functions, one solely a function of space and the other solely a function of time. In covariance terms and assuming stationarity in both time and space, the model may be written

$$C(h,u) = C_0(h)\gamma(u) \tag{21}$$

where $C(h,u)$ denotes the covariance between two space-time coordinates with spatial separation h and temporal separation u, $C_0(h)$ is a pure spatial covariance and $\gamma(u)$ is a temporal autocovariance. Since we may always transfer a constant between the functions C_0 and γ, there is no loss of generality in assuming $\gamma(0) = 1$, in other words, that γ is a temporal autocorrelation function.

Mardia and Goodall (1993) gave a detailed survey of results for spatial-temporal processes concentrating on the separable case. They also considered the generalization to multivariate processes.

In some cases it is possible to adopt an even simpler assumption, in which $\gamma(u) = 0$ for $u \neq 0$; in other words, the situation consists of repeated (in time) independent sampling of spatially dependent vectors. Mardia and Goodall call this the "repeated measurements" model, though it should be pointed out that their interpretation of this phrase is subtly different from what is often called repeated measurements analysis: in a standard interpretation, in biostatistics for example, repeated measurements consist of repeated observations on the same individual at different points in time, but the individuals are independent. In other words, the observations are dependent in time but (if we regard the individuals as having spatial coordinates) independent in space. Mardia and Goodall had in mind the opposite interpretation.

A recent paper employing a repeated measurements analysis (in the Mardia-Goodall sense) to the analysis of environmental pollution data is Smith *et al.* (2003).

4 The Cressie-Huang approach

Cressie and Huang (1999) proposed a generic approach to develop parametric models for spatial-temporal processes. The method relies heavily on spectral representations for the theoretical space-time covariance structure, and generalizes results of Matérn (1986) for pure spatial processes. In essence, Matérn constructed a number of parametric families for spatial processes (including the family that now bears his name) by direct inversion of spectral densities, and Cressie and Huang showed that the same ideas, and very often the same formulae, can be used to construct families of spatial-temporal covariances.

The basic formulae develop from the representation

$$C(h,u) = \int\int e^{i(h^T\omega + u\tau)} g(\omega,\tau) d\omega d\tau \tag{22}$$

where $C(h,u)$ is a stationary spatial-temporal covariance function in which h represents a d-dimensional spatial vector (the spatial separation of two monitor locations) and u is a scalar time component. The function $g(\omega,\tau)$, where ω is d-dimensional and τ is scalar, is then the spectral density of the covariance function C.

We may also write g as a scalar Fourier transform in τ,

$$g(\omega,\tau) = \frac{1}{2\pi}\int e^{-iu\tau} h(\omega,u) du$$

with inverse

$$h(\omega,u) = \int e^{iu\tau} g(\omega,\tau) d\omega. \tag{23}$$

Putting (22) and (23) together,

$$C(h,u) = \int e^{ih^T\omega} h(\omega,u) du. \tag{24}$$

The next step is to write

$$h(\omega,u) = k(\omega)\rho(\omega,u) \tag{25}$$

where $k(\omega)$ is the spectral density of a pure spatial process and $\rho(\omega,u)$ for each ω is a valid temporal autocorrelation function in u. Cressie and Huang remarked that any smooth space-time covariance function can be written in the form (24) and (25) and also imposed the conditions:

(C1) For each ω, $\rho(\omega,\cdot)$ is a continuous temporal autocorrelation function, $\int \rho(\omega,u)du < \infty$ and $k(\omega) <> 0$;

(C2) $\int k(\omega)d\omega < \infty$.

Under those conditions, the generic formula for $C(h,u)$ becomes

$$C(h,u) = \int e^{ih^T\omega} k(\omega)\rho(\omega,u) d\omega. \tag{26}$$

When $\rho(\omega,u)$ is independent of ω, (26) reduces again to a separable model, but of course the main idea of Cressie and Huang was to avoid that case.

Cressie and Huang developed seven special cases of (26).

Model 1:

$$\begin{aligned}\rho(\omega,u) &= \exp\left(-\frac{||\omega||^2u^2}{4}\right)\exp\left(-\delta u^2\right), \qquad (\delta>0)\\ k(\omega) &= \exp\left(-\frac{c_0||\omega||^2}{4}\right), \qquad (c_0>0)\end{aligned}$$

which lead to

$$C(h,u) \propto \frac{1}{(u^2+c_0)^{d/2}}\exp\left(-\frac{||h||^2}{u^2+c_0}\right)\exp\left(-\delta u^2\right). \tag{27}$$

The condition $\delta > 0$ is needed to ensure condition (C1) is satisfied at $\omega = 0$, but the limit of (27) as $\delta \to 0$ is also a valid spatial-temporal covariance function, leading to the three-parameter family

$$C(h,u) = \frac{\sigma^2}{(a^2u^2+1)^{d/2}}\exp\left(-\frac{b^2||h||^2}{a^2u^2+1}\right). \tag{28}$$

Model 2:

$$\begin{aligned}\rho(\omega,u) &= \exp\left(-\frac{||\omega||^2|u|}{4}\right)\exp\left(-\delta u^2\right), \qquad (\delta>0)\\ k(\omega) &= \exp\left(-\frac{c_0||\omega||^2}{4}\right), \qquad (c_0>0),\\ C(h,u) &\propto \frac{1}{(|u|+c_0)^{d/2}}\exp\left(-\frac{||h||^2}{|u|+c_0}\right)\exp\left(-\delta u^2\right),\end{aligned}$$

which again leads to a limit as $\delta \to 0$,

$$C(h,u) = \frac{\sigma^2}{(a^2|u|+1)^{d/2}}\exp\left(-\frac{b^2||h||^2}{a^2|u|+1}\right). \tag{29}$$

Model 3:

$$\begin{aligned}\rho(\omega,u) &= \exp\left(-\frac{||\omega||u^2}{4}\right)\exp\left(-\delta u^2\right), \qquad (\delta>0)\\ k(\omega) &= \exp\left(-c_0||\omega||\right), \qquad (c_0>0),\\ C(h,u) &\propto \frac{1}{(u^2+c_0)^d}\left\{1+\frac{||h||^2}{(u^2+c_0)^2}\right\}^{-(d+1)/2}\exp\left(-\delta u^2\right).\end{aligned}$$

In this case the limit as $\delta \to 0$ is

$$C(h,u) = \frac{\sigma^2(a^2u^2+1)}{\{(a^2u^2+1)^2+b^2||h||^2\}^{(d+1)/2}}. \tag{30}$$

Model 4:

$$\begin{aligned}\rho(\omega,u) &= \exp\left(-||\omega||u|\right)\exp\left(-\delta u^2\right), \qquad (\delta>0)\\ k(\omega) &= \exp\left(-c_0||\omega||\right), \qquad (c_0>0),\\ C(h,u) &\propto \frac{1}{(|u|+c_0)^d}\left\{1+\frac{||h||^2}{(|u|+c_0)^2}\right\}^{-(d+1)/2}\exp\left(-\delta u^2\right),\end{aligned}$$

with limit as $\delta \to 0$,

$$C(h,u) = \frac{\sigma^2(a|u|+1)}{\{(a|u|+1)^2 + b^2||h||^2\}^{(d+1)/2}}. \tag{31}$$

Model 5:

$$\begin{aligned}
\rho(\omega,u) &= \frac{c_0^{d/2}}{(u^2+c_0)^{d/2}} \exp\left(-\frac{||\omega||^2}{4(u+2+c_0)} + \frac{||\omega||^2}{4c_0}\right), \\
k(\omega) &= \exp\left(-\frac{||\omega||^2}{4c_0}\right), \qquad (c_0 > 0) \\
C(h,u) &\propto \exp\left\{-(u^2+c_0)||h||^2 - a_0 u^2\right\} \qquad (a_0 > 0,\ c_0 > 0),
\end{aligned}$$

leading to parametric family

$$C(h,u) = \sigma^2 \exp\left(-a^2u^2 - b^2||h||^2 - cu^2||h||^2\right). \tag{32}$$

Model 6:

$$\begin{aligned}
\rho(\omega,u) &= \frac{c_0^{d/2}}{(|u|+c_0)^{d/2}} \exp\left(-\frac{||\omega||^2}{4(|u|+c_0)} + \frac{||\omega||^2}{4c_0}\right) \exp\left(-\delta u^2\right), \qquad (\delta > 0) \\
k(\omega) &= \exp\left(-\frac{||\omega||^2}{4c_0}\right), \qquad (c_0 > 0) \\
C(h,u) &\propto \exp\left\{-(|u|+c_0)||h||^2 - a_0|u|\right\} \exp\left(-\delta u^2\right), \qquad (a_0 > 0,\ c_0 > 0,\ \delta > 0),
\end{aligned}$$

and again we can let $\delta \to 0$ to get the parametric family

$$C(h,u) = \sigma^2 \exp\left(-a^2|u| - b^2||h||^2 - c|u|||h||^2\right). \tag{33}$$

Model 7:

$$\begin{aligned}
\rho(\omega,u) &= \left\{u^2 + 1 + (u^2+c)||\omega||^2\right\}^{-\nu-d/2} \left(1 + c||\omega||^2\right)^{\nu+d/2}, \qquad (c > 0,\ \nu > 0) \\
k(\omega) &= \left(1 + c||\omega||^2\right)^{-\nu-d/2}, \qquad (c > 0,\ \nu > 0) \\
C(h,u) &\propto \begin{cases} \frac{1}{(u^2+1)^\nu (u^2+c)^{d/2}} \left\{\left(\frac{u^2+1}{u^2+c}\right)^{1/2} ||h||\right\}^\nu K_\nu\left(\left(\frac{u^2+1}{u^2+c}\right)^{1/2} ||h||\right) & \text{if } ||h|| > 0, \\ \frac{1}{(u^2+1)^\nu (u^2+c)^{d/2}} & \text{if } ||h|| = 0, \end{cases}
\end{aligned}$$

where K_ν is the modified Bessel function of the second kind of order ν (exactly as in the usual definition of the Matérn covariance function). In this case, the resulting five-parameter nonseparable spatial-temporal covariance function is

$$C(h,u) = \begin{cases} \frac{\sigma^2(2c^{d/2})}{(a^2u^2+1)^\nu (a^2u^2+c)^{d/2}\Gamma(\nu)} \left\{\frac{b}{2}\left(\frac{a^2u^2+1}{a^2u^2+c}\right)^{1/2} ||h||\right\}^\nu K_\nu\left(b\left(\frac{a^2u^2+1}{a^2u^2+c}\right)^{1/2} ||h||\right) & \text{if } ||h|| > 0, \\ \frac{\sigma^2(2c^{d/2})}{(a^2u^2+1)^\nu (a^2u^2+c)^{d/2}\Gamma(\nu)} & \text{if } ||h|| = 0. \end{cases} \tag{34}$$

We shall not give the detailed proofs of all these formulae, but just that for Model 1 (Model 2 is similar, but the others depend on different Fourier transform formulae). From elementary formulae for characteristic functions, if $X_1, ..., X_d$ are independent $N[0, \sigma^2]$ random variables,

$$E\left\{e^{i(h_1X_1+...+h_dX_d)}\right\} = \exp\left(-\frac{\sigma^2}{2}\sum h_j^2\right).$$

Hence writing $k = 1/\sigma^2$,

$$\int \exp\left(ih^t\omega - \frac{k}{2}||\omega||^2\right) d\omega = \left(\frac{2\pi}{k}\right)^{d/2} \exp\left(-\frac{||h||^2}{2k}\right) \tag{35}$$

Using (26) and setting $k = (c_0 + u)^2/2$, (27) follows directly from (35).

5 References

Bilonick, R.A. (1983), Risk qualified maps of hydrogen ion concntration for the New York Stat area for 1966–1978. *Atmospheric Environment* **17**, 2513–2524.

Bilonick, R.A. (1985), The space-time distribution of sulfate deposition in the Northeastern United States. *Atmospheric Environment* **19**, 1829–1845.

Bilonick, R.A. (1988), Monthly hydrogen ion deposition maps for the Northeastern U.S. from July 1982 to September 1984. *Atmospheric Environment* **22**, 1909–1924.

Bilonick, R.A. and Nichols, D.G. (1983), Temporal variations in acid precipitation over New York State — What the 1965–1979 USGS data reveal. *Atmospheric Environment* **17**, 1063–1072.

Chilès, J.-P. and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty.* John Wiley, New York.

Cressie, N. (1993), *Statistics for Spatial Data.* Second edition, John Wiley, New York.

Cressie, N. and Huang, H.-C. (1999), Classes of nonseparable, spatio-temporal stationary covariance functions. *J. Amer. Statist. Assoc.* **94**, 1330–1340.

Egbert, G.D. and Lettenmaier, D.P. (1986), Stochastic modelingof the space-time structure of atmospheric chemical deposition. *Water Resources Research* **22**, 165–179.

Handcock, M.S. and Wallis, J.R. (1994), An approach to statistical spatial-temporal modeling of meteorological fields (with discussion). *J. Amer. Statist. Assoc.* **89**, 368–390.

Haslett, J. and Raftery, A.E. (1989), Space-time modelling with long-memory dependence: assessing Ireland's wind power resource (with discussion). *Applied Statistics* **38**, 1-50.

Mardia, K.V. and Goodall, C.R. (1993), Spatial-temporal analysis of mulitvariate environmental monitoring data. In *Multivariate Environmental Statistics*, eds. G.P. Patil and C.R. Rao, Elsevier Science Publishers, pp. 347–386.

Matérn, B. (1986), *Spatial Variation.* Lecture Notes in Statistics, Number 36, Springer Varlag, New York. (Second edition: originally published in 1960).

Richardus, R. and Adler, R.K. (1972), *Map Projections.* American Elsevier, New York.

Rodriguez-Iturbe, I. and Mejia, J.M. (1974), The design of rainfall networks in time and space. *Water Resources Research* **10**, 713–729.

Smith, R.L., Kolenikov, S. and Cox, L.H. (2003), Spatio-temporal modeling of $PM_{2.5}$ data with missing values. *Journal of Geophysical Research–Atmospheres*, to appear.

Stein, M.L. (1986), A simple model for spatial-temporal processes. *Water Resources Research* **22**, 2107–2110.