DEVELOPMENTS IN THE MODELLING OF NONSTATIONARY SPATIAL COVARIANCE STRUCTURE FROM SPACE-TIME MONITORING DATA

W. MEIRING
National Center for Atmospheric Research, Boulder, Colorado, USA

P. MONESTIEZ
Unité de Biométrie, INRA, Domaine St Paul, Site Agroparc, 84914 Avignon Cedex 9, France

AND

P.D. SAMPSON, P. GUTTORP
Department of Statistics, University of Washington, Seattle, USA

1. Introduction

Spatio-temporal environmental monitoring is producing vast amounts of data world-wide. These data are being used for purposes such as spatial estimation, (re)design of environmental monitoring networks, evaluation of geophysical space-time models, and definition of inputs to climatological and general circulation models (GCMs). Fundamental to all these purposes is a characterization of the spatial-temporal covariance structure of the environmental processes underlying these data. For most environmental space-time processes there is little reason to expect spatial (and sometimes temporal) stationarity of the covariance structure. This paper presents the most recent methodological developments for an approach to modelling nonstationary spatial covariance structure through deformations of the geographic coordinate system that was first introduced in a technical report 10 years ago (Sampson, 1986).

We address primarily the problem of estimating the spatial covariance structure in levels of an environmental process at arbitrary locations (both monitored and unmonitored), based on records from N point monitoring sites $x_1, \ldots, x_N$, which recorded observations at each of T time points. We denote these records by $Z(x_i, t)$, $i \in \{1, \ldots, N\}$, $t \in \{1, \ldots, T\}$; and consider them to be a sample from the space-time process $Z(x, t)$. For most of this paper we assume independence in time, although we discuss some extensions for modelling spatial and temporal correlation in Section 5.

The fundamental idea of the spatial covariance estimation approach is to compute a deformation of the geographic plane so that the spatial covariance structure can be considered stationary (and isotropic) in terms of a new spatial coordinate system. We model the *spatial dispersion* function, defined as $D(x,y) = \text{Var}\left(Z(x,t) - Z(y,t)\right)$ for each pair of geographic locations x and y, by

$$D(x,y) = \gamma_\theta\left(\|f(x) - f(y)\|\right), \tag{1}$$

where $f(\cdot)$ represents a smooth (bijective) transformation of the geographic coordinate system, and γ_θ represents an isotropic variogram function with parameters θ. In recent years the estimation methodology has been developed by a number of authors and applied in analyses of solar radiation (Sampson and Guttorp, 1992), acid precipitation (Guttorp et al., 1992; Guttorp et al., 1993; Guttorp and Sampson, 1994; Mardia and Goodall, 1993), rainfall (Monestiez et al., 1993; Meiring, 1995), air pollution (Brown et al., 1994), and tropospheric ozone (Sampson et al., 1994; Guttorp et al., 1994; Meiring, 1995).

We denote the sample estimate of the dispersion between the i^{th} and j^{th} monitoring sites by d_{ij}; and the fitted dispersion between the i^{th} and j^{th} monitoring sites by $\widehat{d_{ij}} = \gamma_{\hat\theta}\left(\left\|\hat f(x_i) - \hat f(x_j)\right\|\right)$. For analysis and modelling we usually standardize the site-specific time series to variance 1, in which case we can compute $d_{ij} = 2(1 - r_{ij})$, where r_{ij} is the empirical correlation between the observations at sites x_i and x_j. We refer to the geographic coordinate system as the G-plane, and the transformed coordinate representation as the D-plane. In principle, we may require a D-space of dimension greater than two, but in all applications to date a two-dimensional D-plane representation has been used.

We calculate the G-plane to D-plane transformation using a pair of thin-plate splines. The first spline maps the two-dimensional G-plane into the first coordinate of the D-plane representation and the second spline maps the G-plane coordinates into the second coordinate of the D-plane representation. The splines are fully determined by the D-plane locations of the monitoring sites, so model-fitting involves estimation of these N D-plane locations together with parameters of the isotropic variogram model γ_θ. We minimize the objective criterion

$$C_{\theta,f,\lambda} = \sum_{j=2}^{N}\sum_{i=1}^{j-1}\left[\frac{d_{ij} - \widehat{d_{ij}}}{\widehat{d_{ij}}}\right]^2 + \lambda\ \text{BEP}, \tag{2}$$

with respect to θ and the D-plane coordinates, where BEP denotes a *bending energy penalty* for the transformation. Two of the N monitoring site locations are held at their geographic locations in order to fix the location, scale and orientation of the D-plane representation. The weighted least squares term accounts for the heterogeneity in variance of the sample dispersions d_{ij} (see Cressie, 1985), while the bending energy penalty controls the smoothness of the computed deformation. The so-called bending energy of an $\Re^2 \to \Re$ thin-plate spline mapping g, is

$$\int_{\Re^2} \left[\left(\frac{\partial^2 g}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 g}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 g}{\partial y^2} \right)^2 \right] dx\, dy. \tag{3}$$

For each of the thin-plate spline mappings, this integral can be expressed simply as a quadratic form in the D-plane coordinates. The total BEP of equation (2) is the sum of the bending energies for the two thin-plate spline mappings, a sum which is invariant under translation and rotation of the coordinate systems. See Wahba (1990) for more details on thin-plate splines. Guttorp and Sampson (1994) discuss their interpretation and visualization using biorthogonal grids. These depict the spatially varying geographic directions of greatest and weakest spatial correlation.

Meiring (1995) proves that the model (1) is identifiable in the sense that two strictly increasing D-plane variograms and two bijective deformation maps give the same fitted dispersions between all locations in a plane, if and only if the variograms and deformation maps are the same. (This does not imply identifiability for a finite number of sites.) She shows that the model is invariant under translation, rotation, and scaling of the D-plane coordinates and she provides a discussion of optimization complexities such as multiple modes encountered when optimizing over the variogram parameters and D-plane coordinates.

The choice of smoothing parameter is extremely important. For low values of λ we may "over-fit" the sample dispersions (in view of their sampling variability) and find that the G-plane to D-plane mapping "folds" (is not bijective), something which is generally uninterpretable. For high values of λ, the G-plane to D-plane mapping approaches an affine transformation, corresponding to a stationary or homogeneous model with elliptical anisotropy. Figure 1 shows fitted dispersion models for time series of altitude-adjusted 10-day aggregated precipitation records

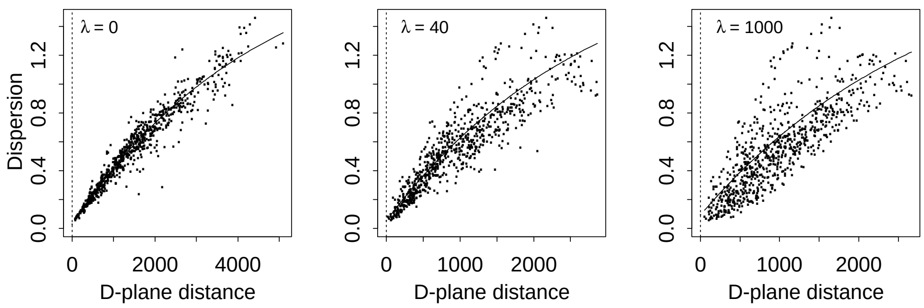

Figure 1. Fitted exponential variograms with nugget, as a function of D-plane distance, for 3 values of λ. The spatial dispersions of the vertical axis correspond to spatial correlations ranging from 1 (for $d_{ij} = 0$) to about 0.3 (for d_{ij}=1.4).

from November and December of 1975 through 1992 (108 records per site) for 39 sites from the Languedoc-Roussillon region of southern France. Results for three values of λ are shown, using an exponential variogram model with nugget. Figure 2 shows the corresponding G-plane to D-plane mapping, $f()$, for $\lambda = 40$. Sites in the northeast region are relatively compressed in the D-plane representation because of relatively higher spatial correlation there.

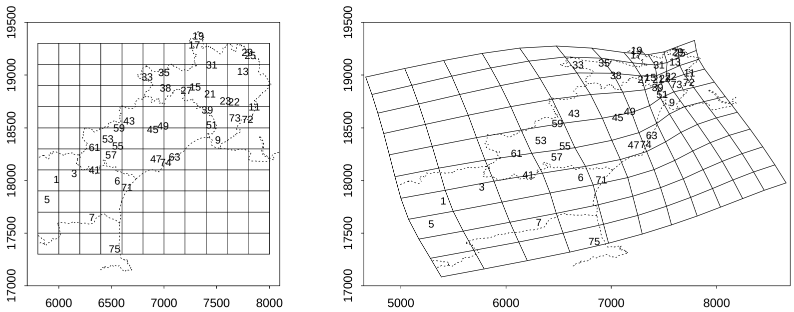

Figure 2. Left panel: Locations of the 39 sites used in the estimation. Site numbers are those used in Monestiez et al. (1993). Right panel: deformation map for $\lambda = 40$. The boundaries of three French departments are drawn in each panel.

These precipitation data are a subset of those previously used and described in Monestiez et al. (1993), where the authors proposed a cross-validation approach for choosing the smoothing parameter (although their objective criterion and G-plane to D-plane mapping approach were different from those just discussed).

An outline of the remainder of the paper is as follows. In Section 2 we present simulations illustrating the bias-variance trade-off based on different choices of the smoothing parameter λ. Sample-based uncertainty in the sample correlations or dispersions d_{ij} contributes to variability of model-based point estimates $\widehat{d_{ij}}$. Section 3 presents a bootstrap approach for assessment of variability of the resulting correlation estimates. In Section 4 a cross-validation study, along the lines of Monestiez et al. (1993), is presented for the penalized weighted least squares objective criterion and the deformation mapping using thin-plate splines. Finally Section 5 discusses spatio-temporal extensions and future directions for research.

2. Bias-variance trade-off

Figure 3 shows differences between estimated and true correlations r_{ij} (computed from the spatial dispersions d_{ij}) for eight Gaussian simulations of spatial time series. For four of the simulations, we simulated $T = 100$ independent realizations of a multivariate Gaussian at $N = 39$ sites of the French precipitation network, with the nonstationary spatial correlation structure corresponding to the model for $\lambda = 40$ represented in Figures 1 and 2. This yielded four sets of 100×39 values, for each of which correlations between pairs of monitoring sites could be estimated. The other four spatial time series were simulated in the same manner, but were of length T=1000.

The deviations of the solid curves from the dotted lines in the figure represent the errors in the estimated correlations between site 45 and the site indicated on the left axis. The dotted lines are labeled on the right with the true underlying correlations (for example, cor($Z(x_{45}, \cdot), Z(x_{59}, \cdot)$) = 0.8), while the spacing between these lines represents a correlation difference of 0.2, so that we can judge the magnitude of the errors of estimation. (See Figure 2 for relative site locations.) The value of λ ranges from 0 (corresponding to no smoothing), to over 20000 (essentially an affine transformation).

As expected, the plots show that the variability of the differences between the true and the estimated correlations is smaller for the longer time series. They also indicate that the estimation bias may be substantial (positive or negative) when an affine transformation is fitted and the true correlation structure is not stationary. For the pairs of sites represented in Figure 3, $\lambda = 40$ appears reasonable. Smaller values of λ result in better fits, but at the expense of less smooth mappings and poorer cross-validatory predictions, as shown in section 4.

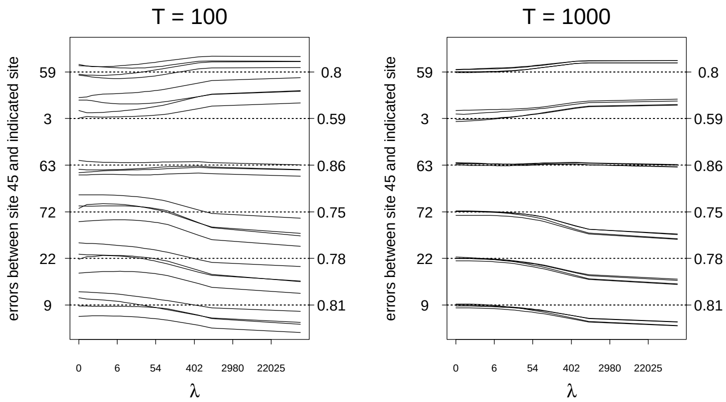

Figure 3. Differences between estimated and true correlation values for correlations between site 45 and other sites in the region (indicated on left vertical axis), for different values of smoothing parameter λ. The left panel plots the errors vs λ for 4 simulations of time series of length 100 with a known nonstationary spatial correlation structure. The spacing between the dotted horizontal lines corresponds to a correlation difference of 0.2. True pairwise correlations are shown on the right vertical axis. The right panel is a similar plot for simulations of time series of length 1000.

3. Assessment of variability

When assessing variability of kriging estimates, researchers most often assume a known covariance structure; i.e., they ignore the uncertainty in the estimated covariance structure. (Some Bayesian approaches to spatial estimation provide an exception. See, for example, Brown et al., 1994.) For the approach described in this paper, it is important to assess whether some of the structure in the covariance field may be due to a few influential points, or whether this structure is truly representative of the underlying process. A bootstrap approach for assessment of variability of this estimator was proposed in Meiring (1995), and is based on work by Beran and Srivastava (1985) on bootstrapping functions of covariance matrices. Whole time slices are resampled with replacement from the original data, to generate M bootstrap samples each composed of T resampled vectors of length N. By resampling whole time slices, we maintain the spatial structure in the original data. The spatial correlation structure is estimated based on each of these M bootstrap samples. Kriging estimates could be obtained using each of these estimated covariance structures, to assess better the additional variability due to estimation of the covariance.

Figure 4 shows: sample (observed) dispersions, fitted dispersions (for $\lambda = 40$), and bootstrapped dispersions between site 45 and the other 38 sites used in this study. The boxplots of the bootstrapped dispersions are based on 50 bootstrap samples from the original data. Figure 4 indicates which sample dispersions are smoothed by the spatial correlation model. The sample dispersions are not always covered by the bootstrap distributions.

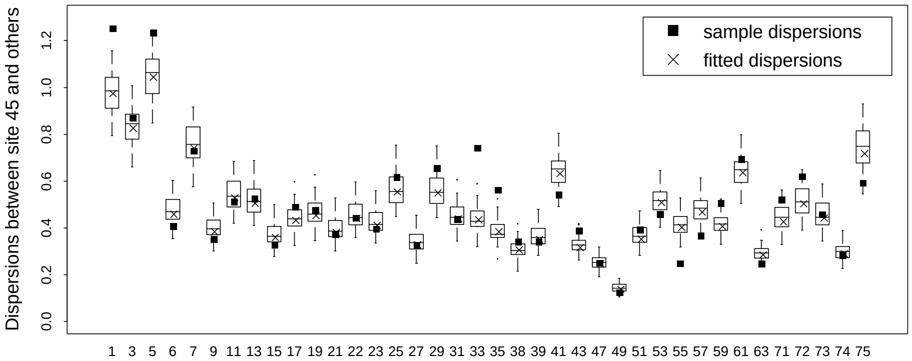

Figure 4. Boxplots of fitted dispersions between site 45 and sites indicated by horizontal axis labels, based on 50 bootstrap samples each of length 108 from the original data, and using $\lambda = 40$. The sample and fitted (shown in figure 1, $\lambda = 40$) pairwise dispersions are also indicated.

4. Choosing the smoothing parameter

Sections 1 and 2 illustrated the role of the smoothing parameter in the estimation approach. In this section we address the choice of λ using a simulation experiment. We compare two different approaches for the specification of λ, along with the extreme cases of $\lambda = 0$ (no smoothing) and $\lambda = \infty$ (yielding an affine G-plane to D-plane mapping, corresponding to a stationary, elliptically anisotropic model).

The first of the two procedures uses a simple stopping rule for a series of decreasing values of λ. The first term of equation (2) decreases monotonically as λ decreases, reaching a minimum value at $\lambda = 0$, corresponding to the fitting of a model without any smoothing constraint. This case generally results in overfitting of the deformation with values of the weighted least squares component of the criterion less than the expected value under the true underlying model. In the case of Gaussian data where the underlying model (θ and $f(\cdot)$) is known (and without standardizing the time series to variance 1), the d_{ij} are simply empirical variances so that the expected value of the first term of the criterion is:

$$E\left[\sum_{i\neq j\in\mathcal{N}}\left(\frac{d_{ij}-\gamma_\theta(f(x_i)-f(x_j))}{\gamma_\theta(f(x_i)-f(x_j))}\right)^2\right]=N(N-1)/T. \tag{4}$$

where $\mathcal{N}=\{1,\ldots,N\}$. In order to avoid overfitting we choose that value of λ in a decreasing sequence, for which the first term of the criterion reaches this expected value. This stopping rule is easy to implement and is computationally simpler than our second approach which follows.

The second procedure is a cross validation minimization. Each monitoring site x_i is eliminated in turn from the data set. Let $C_{\theta,f,\lambda}(i)$ denote the criterion $C_{\theta,f,\lambda}$ given by (2) evaluated on the $N-1$ monitoring sites after excluding site x_i. (Both the sum of squares and BEP in (2) change when excluding a site.) A deformation $\hat{f}_{i\lambda}$ and a variogram $\gamma_{\hat{\theta}_{i\lambda}}$ are computed to minimize $C_{\theta,f,\lambda}(i)$ for each value of λ under consideration. λ is then chosen to minimize the cross validation criterion

$$C_{cv\lambda}=\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}-\{i\}}\left(\frac{d_{ij}-\gamma_{\hat{\theta}_{i\lambda}}(\hat{f}_{i\lambda}(x_i)-\hat{f}_{i\lambda}(x_j))}{\gamma_{\hat{\theta}_{i\lambda}}(\hat{f}_{i\lambda}(x_i)-\hat{f}_{i\lambda}(x_j))}\right)^2. \tag{5}$$

4.1. SIMULATION EXPERIMENT

The function f used in this simulation experiment was computed using a pair thin-plate splines, as described in Section 1. Nine geographic locations and their D-plane locations were used to define a case-study transformation which was

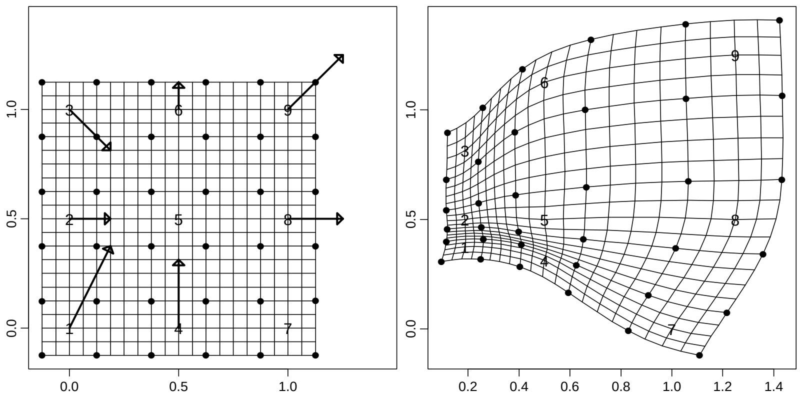

Figure 5. Left: Geographic space, figure shows the nine points and their translations used to define f; the 36 observation sites are dots. Right : Transformed D-space in which the process is isotropic.

clearly non-linear, but smooth (Figure 5). A grid of $N = 36$ points is taken to define the observation sites in the G-plane and γ_θ was specified as an isotropic stationary exponential variogram model with no nugget.

Using the computed image of the $N = 36$ sampled sites in the D-plane together with the variogram model, we computed the true underlying covariance matrix according to the model of equation (1). We then simulated T independent sets of 36 multivariate normal variables. Simulations were run for time series of length T = 25, 100, and 400.

Twenty simulations were run for each of the three values of T. For each simulation, f and γ_θ were estimated using each of 27 different values of λ between 0 and 10,000. An example is given in Figure 6 for one simulation of length $T = 25$.

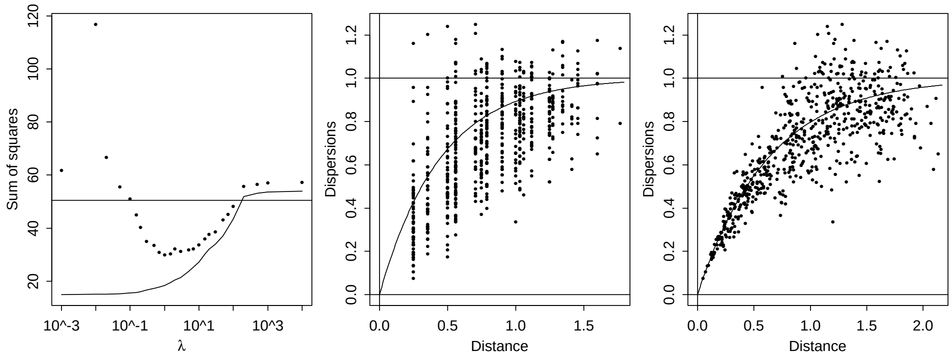

Figure 6. (a) Criterion versus λ : solid curve shows the minimization criterion without the bending energy term, dots show the cross validation criterion. (b) Variogram plot for $\lambda = \infty$ (homogeneous affine deformation). (c) Variogram plot for $\lambda = 1$, close to minimum of cross validation criterion.

4.2. COMPARING THE APPROACHES FOR CHOOSING λ

We used a larger set of validation sites in order to compare the different choices of λ. The test grid was defined as a 9×9 grid of points with corner points (0,0), (1,0), (1,1), and (0,1). (These grid points are at the intersections of every other horizontal and vertical grid line in Figure 5.) The D-plane locations of the 81 points were computed using the estimated f for each value of λ, and the dispersions were then estimated for all pairs of combinations of the 81 sites. The validation criterion is the Mean Square Error comparing the predicted and the theoretical dispersions computed using the theoretical f and variogram model from which we simulated. The mean square is computed over all 3240 pairwise dispersions separately for each of the 20 simulations. For T=25, estimated values of the pairwise dispersions are plotted against the theoretical values in Figure 7 for three values of λ: zero, the value minimizing the cross-validation criterion, and the value resulting from the simple stopping rule.

Each of these three values of λ resulted in better predictions of the true dispersions (i.e., a lower validation criterion) than a model with a homogeneous

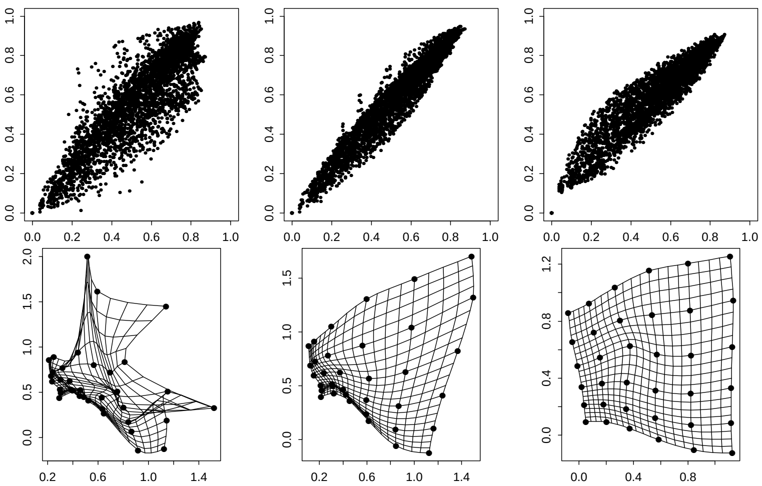

Figure 7. First row from left to right: predicted dispersions (vertical axis) versus true dispersions (horizontal axis) for all pairs of points from the test grid, (i) without smoothing ($\lambda = 0$), (ii) for λ minimizing the cross validation criterion, (iii) for λ chosen by the stopping rule on the expected value of the criterion (4). Second line: Corresponding estimated deformations f.

affine transformation ($\lambda = \infty$). The fits with no smoothing ($\lambda = 0$) resulted in deformations f that folded (as demonstrated in the lower left panel of Figure 7). Nonetheless, dispersions are reasonably well predicted at short geographic distances, which are not much affected by the folding. Much larger cross-validation errors may be introduced for sites further apart in the D-plane.

The simpler stopping rule generally stopped too early in the decreasing sequence, resulting in too much smoothing and a corresponding bias in the estimates, especially for short spatial lags. The shortest D-plane distances were, relatively speaking, not as small as they should have been according to the true model (compare the right panel of Figure 5 and the lower right panel of Figure 7). This oversmoothing did not affect the accuracy of the predictions of dispersions at longer range, but these are less important in kriging calculations. The value of λ determined by cross-validation resulted in more accurate predictions.

Comparisons of the results of the cross-validation criterion (Mean Square Error) across the 20 simulations are summarized in Figure 8 for each of the values of T. The main result is that the determination of λ by cross-validation yields nearly the optimal value (as determined by minimum of the validation Mean Square Error). The simple stopping rule derived from the approximate expected value of the weighted least squares criterion performed poorly, even for the longest time series length ($T = 400$). As expected, for sufficiently large T, no smoothing is necessary, but for short time series, the computationally intensive cross-validation calculation appears necessary.

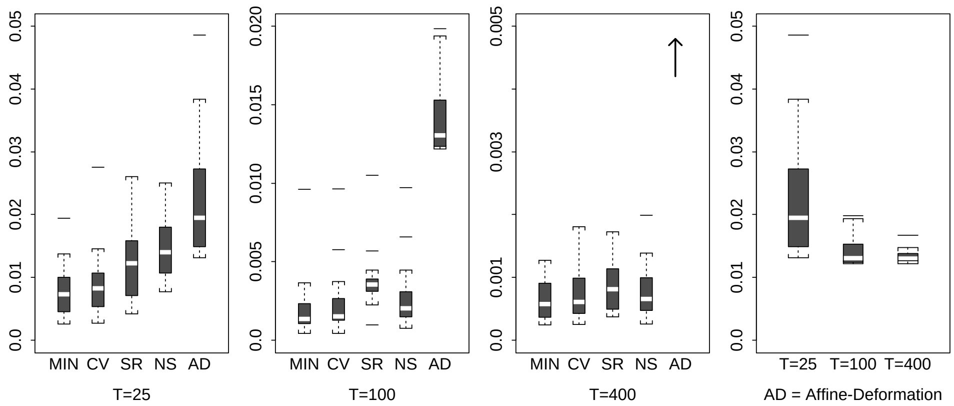

Figure 8. Boxplots of the Mean Square Error validation criterion computed for the 9×9 test grid for each of 20 simulations for T is 25, 100, 400, respectively. MIN: λ yielding the absolute minimum MSE; CV: λ chosen by the cross-validation criterion; SR: λ chosen by the simple stopping rule; NS: $\lambda = 0$ (no smoothing/bending energy penalty); AD: $\lambda = \infty$ (infinite smoothing–homogeneous affine model). Note that the vertical axes differ from plot to plot due to the greater accuracy of the fitted values for large T. The fourth plot on the right shows, on a common scale, the MSE criteria for the AD case.

5. Discussion

In this discussion we wish to note a number of research questions, including variations and extensions of the modelling approach just described. Space limitations prohibit us from going into any detail.

Our spatial deformation model seems heuristically justified for the many areas of application that we have considered and referenced. However, both of the components of the model, the deformation mapping and the variogram function, need not be restricted to the classes of functions that we currently use. The thin-plate spline was chosen for convenience. Smoothing using the bending energy penalty in order to eliminate folding also results in mappings that may be smoother than necessary at small spatial scales. The bending energy penalizes most the small spatial scale deviations from globally smooth maps. A different class of mappings, or equivalently, a different approach to smoothing or regularization may prove useful. It would be interesting to compare our approach to that of Hall et al. (1994), provided the latter can be extended to higher dimensions.

Similarly, in these applications there appears to be no substantive or theoretical rationale for choosing the form of the D-plane variogram, γ_θ (as there is in some traditional geostatistical applications assuming (intrinsic) stationarity). Except for our initial application to solar radiation data (Sampson and Guttorp, 1992), our variogram scatters have been sufficiently noisy that very simple exponential or Gaussian models, or mixtures of these, have sufficed. However, we have previously suggested consideration of a larger nonparametric class of variogram models represented by continuous mixtures of Bessel functions (Guttorp and Sampson, 1994). It is also of interest to investigate statistical properties (such as consistency) of our estimator. This would have to be done using infill asymptotics, i.e., assuming an ever denser network of monitoring stations.

The fitting of our models is a challenging numerical problem with dimensionality roughly proportional to the number of fixed monitoring sites. This makes the cross-validation and bootstrapping procedures presented here extremely demanding computationally. The dimensionality of the problem leads us to ask whether likelihood methods (as proposed by R.L. Smith, unpublished manuscript) or explicitly Bayesian approaches to this deformation model might be fitted using Markov Chain Monte Carlo techniques, which have proved so useful in similarly high-dimensional problems of image analysis (see for example Besag, 1989; and Besag et al., 1995).

One obvious extension of this modelling framework concerns multivariate spatio-temporal data. This has been addressed by Mardia and Goodall (1993) and by Brown et al. (1994). However, the explicit deformation modelling of a cross-variogram has yet to be developed.

We have recently addressed the modelling of spatio-temporal processes having important temporal as well as spatial temporal correlation structure. We have considered two different approaches to spatio-temporal estimation for hourly ozone measurements from a large monitoring network in California. The first approach considers spatially varying time series models. We first pre-whiten these individual time series before applying the deformation modelling methods discussed here. Spatial estimation is carried out using the pre-whitened residuals and then the resulting estimates are post-coloured (Meiring, 1995). An alternative approach is to apply the spatial deformation modelling method to the space-time covariance structure using the bilinear model of coregionalization as suggested by Rouhani and Wackernagel (1990). This means considering the space-time process as a multivariate temporal process, with the multivariate dimension being the number of monitoring sites (Sampson et al., 1994). More recent analyses consider the observations at the 24 hours of the day as a 24-variate spatial process. Extensions of the bilinear model of coregionalization to deal with asymmetric space-time cross-variograms, such as those recently introduced by Grzebyk and Wackernagel (1994) and Ver Hoef and Barry (1994) are called for here.

Finally, we note that a tremendous amount of spatial-temporal environmental monitoring is now being carried out by remote sensing. Data are acquired as images with typically irregular time series coverage of given locations on earth—that is, most satellites are asynchronous and do not yield regular time series of images of a fixed array of pixels. The deformation model for nonstationary spatial covariance structure is clearly still relevant for the processes being monitored, but new methods of fitting or estimation must be developed.

References

Beran, R., and M.S. Srivastava (1985). Bootstrap tests and confidence regions for functions of a covariance matrix. *The Annals of Statistics*, **13**: 95-115. (Correction given, 1987, Vol **15**: 470-471.)

Besag, J. E. (1989). Towards Bayesian image analysis. *Appl. Statist.*, **16**: 395-407.

Besag, J., P. Green, D. Higdon, and K. Mengersen (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science*, **10**: 3-66.

Brown, P.J., N.D. Le, and J.V. Zidek (1994). Multivariate spatial interpolation and exposure to air pollutants. *The Canadian Journal of Statistics* **22**: 489-505.

Cressie, N.A.C. (1985). Fitting variogram models by weighted least squares. *Mathematical Geology*, **17**: 563-586.

Grzebyk, M. and Wackernagel, H. (1994). Multivariate analysis and spatial/temporal scales: Real and complex models. *Proceedings of the XVIIth International Biometric Conference*, Hamilton, Ontario.

Guttorp, P., W. Meiring, and P.D. Sampson (1994). A space-time analysis of ground-level ozone data. *Environmetrics* **5**: 241-254.

Guttorp, P., and P.D. Sampson (1994). Methods for Estimating Heterogeneous Spatial Covariance Functions with Environmental Applications. *In* Patil, G.P. and C.R. Rao (eds), *Handbook of Statistics XII: Environmental Statistics*, Elsevier/North Holland, New York, pp. 663-690.

Guttorp, P., P.D. Sampson, and K. Newman (1992). Nonparametric estimation of spatial covariance with application to monitoring network evaluation. *In* Guttorp, P. and A. Walden (eds), *Statistics in the environmental and earth sciences*, pp. 39-57.

Guttorp, P., N.D. Le, P.D. Sampson, and J.V. Zidek (1993). Using entropy in the redesign of an environmental monitoring network. *In* Patil, G.P. and C.R. Rao (eds), *Multivariate Environmental Statistics*, North-Holland Elsevier Science Publishers, pp. 175-202.

Hall, P., N.I. Fisher, B. Hoffmann (1994). On the nonparametric estimation of covariance functions. *Annals of Statistics*, **22**: 2115-2134.

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

Meiring, W. (1995). *Estimation of Heterogeneous Space-Time Covariance.* Ph.D. Dissertation, University of Washington, Seattle, 192 pp.

Monestiez, P., P.D. Sampson, and P. Guttorp (1993). Modelling of heterogeneous spatial correlation structure by spatial deformation. *Cahiers de Geostatistique*, Fascicule 3, Compte Rendu des Journees de Geostatistique, 25-26 May 1993, Fontainebleau. Published by the École Nationale Supérieure des Mines de Paris.

Rouhani, S., and H. Wackernagel, (1990). Multivariate Geostatistical Approach to Space-Time Data Analysis. *Water Resources Research* **26:** No. 4: 585-591.

Sampson, P.D. (1986). Spatial covariance estimation by scaled-metric scaling and biorthogonal grids. Technical Report No. 91, Department of Statistics, University of Washington, and SIMS Technical Report No. 102, Department of Statistics, University of British Columbia.

Sampson, P.D., and P. Guttorp (1992). Nonparametric estimation of nonstationary spatial covariance structure. *JASA*, **87:** 108-119.

Sampson, P.D., P. Guttorp, and W. Meiring (1994). Spatio-temporal analysis of regional ozone data for operational evaluation of an air quality model. *Proceedings of the Section on Statistics and the Environment, American Statistical Association.*

Ver Hoef, J.M., and Barry, R.P. (1994). Modeling crossvariograms for cokriging and multivariable spatial prediction, (submitted manuscript).

Wahba, G. (1990). *Spline models for observational data.* SIAM:PA.

Acknowledgements: We thank two reviewers for their comments which helped improve this paper. Funding from the Electric Power Research Institute, INRA, and the NCAR Geophysical Statistics Project is gratefully acknowledged. The NCAR Geophysical Statistics Project is sponsored by the National Science Foundation under Grant #DMS93-12686.