Analysis of multivariate spatial data using latent variables

William Fredrick Christensen

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Statistics and Probability Commons

Recommended Citation

https://lib.dr.iastate.edu/rtd/12559
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Analysis of multivariate spatial data using latent variables

by

William Fredrick Christensen

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics
Major Professor: Yasuo Amemiya

Iowa State University
Ames, Iowa
1999
This is to certify that the Doctoral dissertation of

William Fredrick Christensen

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program

Signature was redacted for privacy.

For the Graduate College
To my wife, Mary Elizabeth,
who believed in my ability to get a Ph.D. long before I did
# TABLE OF CONTENTS

**GENERAL INTRODUCTION** ................................................................. 1
1 Introduction ......................................................................................... 1
2 Dissertation organization ................................................................. 2
References .............................................................................................. 3

**GENERALIZED SHIFTED-FACTOR ANALYSIS METHOD FOR MULTIVARIATE GEO-REFERENCED DATA** ................................................................. 5
Abstract ................................................................................................ 5
1 Introduction ......................................................................................... 5
2 The generalized shifted-factor model ................................................ 8
3 Identification and interpretation of shifts ............................................ 10
4 Estimation, inference, and prediction .................................................. 12
5 Simulation results ............................................................................... 23
References .............................................................................................. 26

**LATENT VARIABLE MODELING OF MULTIVARIATE SPATIAL DATA WITH POTENTIAL LAGGED-DEPENDENCIES** ................................................................. 29
Abstract ................................................................................................ 29
1 Introduction ......................................................................................... 29
2 The model and proposed approach ...................................................... 31
3 Estimation issues ............................................................................... 36
4 Inference and prediction .................................................................... 39
5 Simulation results ............................................................................... 43
5.1 Evaluation of estimation approaches ............................................... 43
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2</td>
<td>Evaluation of estimation properties under several data models</td>
<td>51</td>
</tr>
<tr>
<td>6</td>
<td>Analysis of precision farming data</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Appendix</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td><strong>MODELING AND PREDICTION FOR MULTIVARIATE SPATIAL DATA USING SPATIALLY-CORRELATED LATENT VARIABLES</strong></td>
<td>76</td>
</tr>
<tr>
<td>Abstract</td>
<td></td>
<td>76</td>
</tr>
<tr>
<td>1</td>
<td>Introduction</td>
<td>76</td>
</tr>
<tr>
<td>2</td>
<td>Estimation and inference procedures</td>
<td>78</td>
</tr>
<tr>
<td>3</td>
<td>Tests for model fit</td>
<td>84</td>
</tr>
<tr>
<td>4</td>
<td>Prediction of latent variables</td>
<td>88</td>
</tr>
<tr>
<td>5</td>
<td>Simulation results</td>
<td>91</td>
</tr>
<tr>
<td>5.1</td>
<td>Inference and model building procedures</td>
<td>91</td>
</tr>
<tr>
<td>5.2</td>
<td>Comparison of parameter estimators and latent variable predictors</td>
<td>95</td>
</tr>
<tr>
<td>6</td>
<td>Illustrative example</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td><strong>GENERAL CONCLUSION</strong></td>
<td>102</td>
</tr>
<tr>
<td></td>
<td><strong>ACKNOWLEDGMENTS</strong></td>
<td>103</td>
</tr>
</tbody>
</table>
1 Introduction

Latent variable modeling is among the most widely used techniques in multivariate analysis. This model-based approach to the analysis of multi-dimensional data is particularly popular in the behavioral and social sciences where researchers are often most interested in abstract concepts such as feelings, attitudes, and aptitudes. Because these variables of interest are not observed directly, researchers rely on measurements of several observable quantities that are believed to be closely related to the latent variables of interest. Latent variable models such as the linear factor analysis model allow one to explore relationships among the observed variables and incorporate subject matter knowledge directly into the model. A wide variety of statistical software packages are available for data collected under the usual assumption of random sampling or independent individual observations.

Many disciplines, however, are primarily interested in the analysis of multivariate correlated data. Factor analysis for multivariate time series data has been discussed by Cattell (1957, 1963), Anderson (1963), Box and Tiao (1977), Peña and Box (1987), and Molenaar (1985). A different kind of correlated data is regularly of interest in ecology, agriculture, and the environmental sciences. In these disciplines, data are gathered from locations throughout a geographic region and are thus spatially correlated. When fitting a latent variable model to multivariate, spatially correlated data, and when predicting the spatially correlated latent processes, we wish to exploit not only the dependence among the elements of the multivariate observation vector, but also the spatial depen-
Several authors have proposed dimension-reduction techniques for multivariate spatial data including Wackernagel (1988), Goovaerts, Sonnet, and Navarre (1993), Grunsky and Agterberg (1991, 1992), and Cook, et al. (1994). This dissertation extends the work of these authors by developing a formal, statistically sound methodology for latent variable modeling of spatially correlated data and for latent variable prediction.

2 Dissertation organization

This dissertation is composed of three papers, each focusing upon a different aspect of latent variable modeling for multivariate spatial data. The first paper targets the audience in applied sciences such as geology and agriculture where multivariate georeferenced data are often collected. In this paper, the generalized shifted-factor model is introduced for modeling spatial data which may exhibit a wide variety of dependence characteristics. The model incorporates potential lagged dependencies between factors and observed variables, representing asymmetric spatial dependencies observed in practice. Identification of asymmetric lagged dependencies or "shifted factors" is discussed and tools for practical implementation of the approach are given. Unlike many geostatistical techniques, this methodology does not depend upon the fitting of parametric (co)variogram functions to the data. Methods are presented for parameter estimation and prediction of latent variables.

The second paper explores analytically and empirically the statistical properties of the augmented observation model-fitting approach for the generalized shifted-factor model. Modeling tools are discussed and theorems related to the large-sample properties of the parameter estimators are stated and proved. Extensive simulation verifies the theoretical properties of the estimation and inference procedures. A simple example from precision farming is given.

In the third paper, a systematic procedure for modeling the underlying structure of multivariate spatial data is proposed. Aspects of parameter estimation, model fitting and assessing, and inference are discussed. Given spatially dependent observations with
unspecified distributions, very unrestrictive and practical conditions are given under which an inference procedure based on normal-theory maximum-likelihood estimation is valid. Procedures for assessing such conditions and for checking and comparing model fits are discussed. For multivariate prediction, a procedure combining the latent variable modeling and a measurement-error-free kriging technique is introduced and compared to other methods. An example using agricultural data is given.

Following the three papers is a general conclusion which summarizes the contributions of these papers to the body of spatial multivariate analysis literature.

References


GENERALIZED SHIFTED-FACTOR ANALYSIS METHOD
FOR MULTIVARIATE GEO-REFERENCED DATA

A paper to be submitted to *Mathematical Geology*

William F. Christensen and Yasuo Amemiya

Abstract

Multivariate data with spatial dependencies arise in many areas of application, including environmental sciences, precision agriculture, and ecology. For analysis of such data, a methodology based on a generalized shifted-factor model is developed. The model incorporates potential lagged dependencies between factors and observed variables, representing asymmetric spatial dependencies observed in practice. Identification and estimation issues are discussed, and inference and prediction procedures are proposed.

1 Introduction

Factor analysis is a widely-used technique for modeling high-dimensional data as a function of a smaller number of latent variables or factors. Classical factor analysis, as used in the social sciences and other disciplines, typically assumes that the observations form a random sample from some population. The factor analysis model exploits this independence among observations in order to estimate the factors and factor loadings. The result is often a model that is able to both simplify an often complex correlation structure among variables and yield inference that is meaningful for the subject matter.
In many areas of application the data are collected in space and, as a result, the observations are correlated. As a general rule, two observations that are close together tend to be more highly correlated than two observations that are far apart. When multivariate observations are collected in space, a classical factor analysis model is not appropriate, but a factor analytic approach is still attractive. For example, a common practice in ecology is to lay out a grid over a geographic region and then evaluate the abundance of each of many species at each location on the grid. In such situations, the ecologist often has in mind a model which states that at each location, the abundance vector of a large dimension depends on some small number of latent variables or factors—perhaps a water abundance factor, an elevation factor, or soil factor. In general, it is often desirable to model multivariate geo-referenced observed variables in terms of a smaller number of underlying factors based on either subject matter theory or practical reasons.

Several authors have discussed approaches to latent variable modeling of spatial data. Wartenberg (1985) discussed a factor analytic approach in two steps. The first step involves a dimension reduction ignoring spatial correlation structure and the second involves an analysis of the spatial structure of each factor. Wackernagel (1988) used a model with a restrictive assumption that lagged cross-covariances between different variables are symmetric. Grunsky and Agterberg (1991, 1992) proposed a technique which defines factors after finding linear combinations of the variables that are maximally correlated at specific lags. Goovaerts, Sonnet, and Navarre (1993) employ the concept of “coregionalization,” wherein the researcher makes an *a priori* selection of a base set of variograms. All of the variograms and cross-variograms for the variables in the observation vector are then modeled as a linear combination of the base set of variograms. Factors are then extracted by analyzing the structure of these linear coefficients. Cook et al. (1994) introduced the shifted-lag model, which states that each observed variable depends on a factor at a location that is possibly shifted away from the observational location. Majure and Cressie (1997) developed exploratory data analysis tools that are helpful in investigating lagged relationships among variables in space. These last two
papers did not discuss statistical model fitting and estimation procedures.

Because of the obvious similarities between spatial and time series data analyses, a review of factor analytic research in the time series literature is useful. Cattell (1957, 1963) described the $P$-technique, which finds the time shift between pairs of variables that will maximize the time series' correlation. These lagged correlations are then used to extract factors. This approach was criticized by Anderson (1963) and others for two main reasons. First, this model is unable to sufficiently incorporate important lagged dependencies among the variables into the model. Second, the model-fitting procedure advocated by Cattell will be inconsistent with the model itself if the pattern of pairwise lagged relationships among the variables do not “fit together” in a cohesive manner. Box and Tiao (1977) and Peña and Box (1987) used eigenvectors as a canonical transformation to create new variables that can be considered as factors. Molenaar (1985) pointed out that variables will often need to be correlated with the latent factors at multiple lags. He developed the so-called dynamic factor model, which incorporates multiple lagged dependencies among variables, and captures most of the important correlation structure among the components of the multivariate time series.

In this paper, we develop a factor analytic methodology which provides an internally consistent, realistic latent variable approach for modeling multivariate spatial data. Our approach is based on a new model called the generalized shifted-factor model. This model is a generalization of the shifted-lag model of Cook et al. (1994), but incorporates multiple lagged dependencies among variables collected in space in a manner similar to that of Molenaar's dynamic factor analysis model for time series data. Also, the new model allows for the asymmetric covariance properties missing in Wackernagel's model. Model fitting and parameter estimation methods are proposed, and statistical inference and prediction procedures for addressing scientifically relevant issues are developed. Our overall approach is to develop a broadly applicable statistical methodology without requiring assumptions about the distribution of data or the form of covariance functions.

Section 2 describes the generalized shifted-factor model. Section 3 further discusses
the concept of the shifted-factor, and proposes methods for selecting shifts to be incor-
porated in a model. Section 4 discusses estimation, inference, and prediction procedures,
and Section 5 evaluates the performance of this methodology via computer simulation.

2 The generalized shifted-factor model

Define $Z^2$ to be a two-dimensional rectangular grid. Although each location on
the grid is denoted with $(i, j)$, $i \in \{..., -1, 0, 1, ..., \}$, $j \in \{..., -1, 0, 1, ..., \}$, we
allow the distance between neighboring sites within a column of the grid to be different from the
distance between neighboring sites within a row of the grid. Suppose that we observe
$Z(s) = (Z_1(s), ..., Z_p(s))'$, a $p$-dimensional second-order stationary process for $s \in D$ and
$D \subset Z^2$. Assume that the $p$ observed variables $Z_i$ can be expressed as linear functions of
$k (< p)$ unobservable factors $W_j(s), j = 1, ..., k$, except for errors $e_i(s)$. To represent the
general dependence structure of $Z(s)$ in factors, it is assumed that each $Z_i(s)$ linearly
depends on the values of each $W_j(s)$ at $m_{ij}$ locations or shifts. Then, the generalized
shifted-factor model for $Z(s)$ is

$$Z_i(s) = \gamma_i + \sum_{j=1}^{k} \sum_{g=0}^{m_{ij}} \lambda_{ijg} W_j(s + \Delta_{ijg}) + e_i(s), \quad i = 1, ..., p, \quad (2.1)$$

where $\Delta_{ij0} = 0$, and $\mathbb{E}\{e_i(s)\} = 0$. Since all relationships among the $p$ observed variables
$Z_i(s)$ are to be explained by the $k$-dimensional process $W(s) = (W_1(s), ..., W_k(s))'$, it is
assumed that the $p + 1$ spatial processes $W(s), e_i(s), ..., e_p(s)$ are independent. We refer
to $W_j(s + \Delta_{ijg})$ as a "shifted factor" when $\Delta_{ijg} \neq 0$. Asymmetric lagged dependencies
among the observed variables can be incorporated by including shifted factors in the
model.

This model generalizes the shifted-lag model of Cook et al. (1994) by incorporating
ideas of Anderson (1963) and Molenaar (1985) and the general shifted-factor structure.
Model (2.1) expresses an internally coherent model that allows modeling issues to be
addressed by statistical inference. For example, the dependence of $Z_i(s)$ on particular
shifted factor $W_j(s + \Delta_{ijg})$ can be checked by testing for $\lambda_{ijg} = 0$. Model (2.1) has some
similarity with the spatial auto-regressive models discussed by several authors including
Whittle (1954) and Basu and Reinsel (1993). However, in model (2.1), $Z(s)$ depends on a linear combination of shifted (and zero-shifted) latent variables rather than a linear combination of neighboring values of the $Z$-process.

To use the general representation (2.1) for statistical analysis, two indeterminacy issues need to be addressed. First, as in a classical factor analysis, model (2.1) has the factor indeterminacy problem that the unobservable $k \times 1$ factors $W(s)$ can be linearly transformed without altering the model form. Second, the geo-referenced location of the unobservable $W(s)$ is only arbitrarily determined, and thus, the shift quantities $\Delta_{ijg}$ are not uniquely identified. Both indeterminacies can be removed using a simple and practical parameterization of model (2.1) obtained by fixing $k$ of the observed variables to be equal to a factor plus error, i.e., by fixing the values of certain $\gamma_i$ and $\lambda_{ijg}$ to be zeros and ones. The generalized shifted-factor model with identifiable parameterization is

$$Z_i(s) = \gamma_i + \sum_{j=1}^{k} \sum_{g=0}^{m_{ij}} \lambda_{ijg} W_j(s + \Delta_{ijg}) + e_i(s), \quad \text{for } i = 1, \ldots, p - k,$$

$$Z_i(s) = W_{i-p+k}(s) + e_i(s), \quad \text{for } i = p - k + 1, \ldots, p,$$

(2.2)

where $\Delta_{ij0} \equiv 0$, $i = 1, \ldots, p - k$, $j = 1, \ldots, k$. In this parameterization, $k$ factors $W_j(s)$ cannot be transformed, and the non-zero shifts $\Delta_{ij0}$ are uniquely identified. Also, the $k$ factors $W_j(s)$ can be interpreted as the true values of the last $k$ observed variables $Z_i(s)$.

Model (2.2) with no indeterminacy can be used meaningfully in discussing statistical model fitting and parameter estimation. For example, if $p = 3$, $k = 1$, and $m_{11} = m_{21} = 1$, then we write the model as follows:

$$Z_1(s) = \gamma_1 + \lambda_{110} W_1(s) + \lambda_{111} W_1(s + \Delta_{111}) + e_1(s),$$

$$Z_2(s) = \gamma_2 + \lambda_{210} W_1(s) + \lambda_{211} W_1(s + \Delta_{211}) + e_2(s),$$

$$Z_3(s) = W_1(s) + e_3(s).$$

(2.3)

As in the classical factor analysis, there is an additional identification issue related to the number $k$ of factors in the model. In the generalized shifted-factor model (2.2), the existence of the shifted terms makes this issue somewhat difficult to address comprehensively. In practice, we suggest the following guideline. Given a model in form (2.2),
first consider the corresponding classical factor model obtained by deleting all shifted factor terms and by treating $W(s)$ and $e_i(s)$ to be spatially-independent processes. We consider the proposed generalized shifted-factor model to be identified only if the corresponding classical model is identified in the usual factor analysis sense. This practical approach is justified, because our model fitting and estimation methods in Section 4 work properly under this guideline for a broad class of models.

3 Identification and interpretation of shifts

The existence of the shifted factor terms $W_j(s + \Delta_{ij})$ in the generalized shifted-factor model (2.2) makes it possible to model underlying asymmetric spatial dependencies in multivariate data. The shift or lag parameters $\Delta_{ij}$ to be included in a model can be selected based on either subject matter meaning or an exploratory model building approach. For example, because of the movement of pollutants downstream over time, one might have reason to believe that the amount of nitrate observed at a given location will depend heavily upon the amount of nitrate (or some other pollutant) observed some distance upstream. In such situations, the shifts suspected by the subject matter meaning should be included in the model. In other situations, researchers may not be hypothesizing the existence of particular shifts, but may wish to explore possible asymmetric dependence structure. For the exploratory purpose, certain nearest neighbors can be included as potential shifts in the generalized shifted-factor model (2.2). For example, if the data is collected on a grid, a subset of the four nearest neighbors (i.e., $\{(-1,0), (1,0), (0,-1), (0,1)\}$) or the eight nearest neighbors (i.e., $\{(-1,0), (1,0), (0,-1), (0,1), (-1,-1), (1,-1), (-1,1), (1,1)\}$) on the grid can be included as $\Delta_{ij}$'s. Once potential shifts are included in the model, our statistical procedure to be developed in Section 4 can be used to test the significance or existence of shifts.

In addition, there is a graphical exploratory method for selecting the shifts. For a simple shifted-factor model, Majure and Cressie (1997) suggested a method based on a cross-variogram-cloud plot. They considered a model that can be considered a special case of (2.1) with simple structure. In their model, only one factor exists, and each
$Z_i(s)$ depends on only one term of the form $W_j(s + \Delta)$ with possibly nonzero $\Delta$. Our parameterization (2.2) allows us to extend their approach to more general models. Given $Z_i(s), i = 1, ..., p$, with a possible model of the form (2.2), we make a number of cross-variogram plots. For every pair $(l_1, l_2), l_1 = 1, ..., p - k, l_2 = p - k + 1, ..., p$, consider the cross-variogram-cloud plot of points

$$\{(||s_r - s_q||, |Z_{l_2}(s_r) - Z_{l_1}(s_q)|^{1/2}) : (s_q, s_r) \in A_\theta\}, \quad (3.1)$$

and

$$\{(-||s_r - s_q||, |Z_{l_2}(s_r) - Z_{l_1}(s_q)|^{1/2}) : (s_q, s_r) \in A_{\theta + \pi}\}, \quad (3.2)$$

where $A_\theta$ is an angle class centered at $\theta$ radians. The first coordinate of each point in the cross-variogram-cloud plot is the signed distance between the two locations. When the angle from $s_q$ to $s_r$ is in the interval $\theta \pm \epsilon$, we define the signed distance for the pair of locations as the distance from $s_q$ to $s_r$. When the angle from $s_q$ to $s_r$ is in the interval $\theta + \pi \pm \epsilon$, we define the signed distance for the pair of locations as the negative distance from $s_q$ to $s_r$. We next fit a line such as a smoothed spline or a median smoother (Majure and Cressie, 1997) through the points on the cross-variogram-cloud plot. and look for outstanding dips or peaks in the line. If a shifted factor $W_i(s + \Delta_{l_1l_2})$ exists in this angle class, then the signed distance associated with the dip/peak and the angle $\theta$ together denote the value of the potential shift parameter $\Delta_{l_1l_2}$. A dip (peak) in the smoothed line indicates that $Z_{l_1}(s)$ is positively (negatively) correlated with $W_i(s + \Delta_{l_1l_2})$. After considering cross-variogram-cloud plots corresponding to each possible angle class $A_\theta$, the researcher may choose to include several shifted factors in an exploratory model. Statistical inference procedures developed in the next section can be used to statistically determine which shifted factors are important in the model.

Examples of cross-variogram-cloud plots generated from simulated data are given in Figures 1 and 2. Normal factors and errors with spherical covariance functions were generated with factor range $= 4$, factor sill $= 4$, factor nugget $= 0$, error range $= 2,$
error sill = 1, and error nugget = 0. Figure 1 uses data generated according to
\[
Z_1(s) = 0.5 W_1(s) + 1.0 W_1(s + \Delta_{111}) + e_1(s), \\
Z_2(s) = W_1(s) + e_2(s),
\]
and Figure 2 uses data generated according to
\[
Z_1(s) = 0.5 W_1(s) + 1.0 W_1(s + \Delta_{111}) + 1.0 W_1(s + \Delta_{112}) + e_1(s), \\
Z_2(s) = W_1(s) + e_2(s),
\]
where \( \Delta_{111} = (2,0) \) and \( \Delta_{112} = (0,4) \) in rectangular coordinates.

Note that the most pronounced nonzero shift apparent on the cross-variogram-cloud plots in Figure 1 is at the distance of 2 units on the 0-radian-centered angle class, indicating a potential shifted factor at \( \Delta_{111} = (2,0) \) in rectangular coordinates. In Figure 2, a shift of distance 2 units is evident on the plot for the 0-radian-centered angle class, and a shift of distance 4 units is evident on the \( \pi/2 \)-radian plot, indicating potential shifted factors at \( \Delta_{111} = (2,0) \) and \( \Delta_{112} = (0,4) \).

4 Estimation, inference, and prediction

In this section, a model fitting method for the generalized shifted-factor model (2.2) and associated statistical inference procedures are developed. Our approach here is to present procedures that can be used for a broad class of problems without specifying or checking particular distributional or covariance function forms. For this, we consider model (2.2) with any distribution for \( W(s) \) and \( e_i(s) \), and without any parametric covariance function models for any processes. Because we make no distributional assumptions, a likelihood-based approach is not possible, and we employ a distribution-free moment-based approach that we will refer to as the “augmented observation” model-fitting approach.

Given model (2.2), first the \( p \)-dimensional observation vector \( Z(s) \) is expanded into a \( p^* \)-vector \( Z^*(s) \) which includes measurements of lagged variables. The new observation \( Z^*(s) \) can be created using the following “simple expansion rule”:
Figure 1 Cross-variogram-cloud plots for 0 (top), $\pi/4$ (middle), and $\pi/2$ (bottom) radian angle classes (singly-shifted).
Figure 2 Cross-variogram-cloud plots for 0 (top), π/4 (middle), and π/2 (bottom) radian angle classes (multiply-shifted).
1. For each $\Delta_{ijg}$ in the model ($i = 1, \ldots, p - k, j = 1, \ldots, k, g = 0, \ldots, m_{ij}$), include the shifted variable $Z_i(s + \Delta_{ijg})$ in $Z^*(s)$.

2. For all $W_j(s + \Delta_{ijg})$ involved in the model for each of the variables added to $Z^*(s)$ in step 1, include $Z_{p-k+j}(s + \Delta_{ijg})$ in $Z^*(s)$.

For example, for model (2.3):

$$Z^*(s) = \begin{bmatrix} Z_1(s) \\ Z_1(s + \Delta_{111}) \\ Z_2(s) \\ Z_2(s + \Delta_{211}) \\ Z_3(s) \\ Z_3(s + \Delta_{111}) \\ Z_3(s + \Delta_{211}) \\ Z_3(s + 2\Delta_{111}) \\ Z_3(s + 2\Delta_{211}) \end{bmatrix}.$$  

Let $p^*$ denote the dimension of $Z^*(s)$. Model (2.2) for the $p \times 1$ $Z(s)$ induces a similar model for $Z^*(s)$ in the form

$$Z^*(s) = \gamma + \Lambda W^*(s) + e^*(s),$$  

where $W^*(s)$ and $e^*(s)$ contain the original $W(s)$ and $e(s)$ and their lagged values, respectively, the elements of $\Lambda$ are the original $\lambda_{ijg}$'s, zeros and ones, and $\gamma$ consists of $\gamma_i$'s and zeros. Let the dimension of $W^*(s)$ be denoted by $k^*$. Then, the $p^* \times p^*$ $\Sigma = \text{var}\{Z^*(s)\}$ can be written as

$$\Sigma = \Lambda \Phi \Lambda' + \Psi,$$

where $\Lambda$ is $p^* \times k^*$, $\Phi = \text{var}\{W^*(s)\}$ is $k^* \times k^*$, and $\Psi = \text{var}\{e^*(s)\}$ is $p^* \times p^*$. In writing down $\Phi$ and $\Psi$, we do not assume any parametric models for covariance functions of $W(s)$ and $e_i(s)$, and treat all variances and covariances to be unknown and unrestricted. Since $W^*(s)$ contains lagged values of $W(s)$, and since $W(s)$ is second-order stationary,
the number of unique elements in $\Phi$ is less than $\frac{k^2(k^2+1)}{2}$. Since $e_i(s)$, $i = 1, ..., p$, are independent, the number of unique elements in $\Psi$ is much less than $\frac{p^2(p^2+1)}{2}$. But, $\Psi$ contains spatial covariances of the type $\text{cov}\{e_i(s), e_i(s + \Delta_{ij})\}$, and $\Psi$ is a (possibly scrambled) block diagonal matrix with $p$ blocks. Let $\theta$ denote the parameter vector consisting of all distinct parameters in $\Lambda$, $\Phi$, and $\Psi$, i.e., all $\lambda_{ij}$'s and all unique variances and covariances for $W^*(s)$ and $e^*(s)$. Using this $\theta$, we write $\Sigma$ in (4.3) as $\Sigma(\theta)$, a function of $\theta$.

In the example with model (2.3) and $Z^*(s)$ defined in (4.1), we have

$$\Sigma(\theta) = \Lambda \Phi \Lambda' + \Psi,$$

where

$$W^*(s) = \begin{bmatrix} W_1(s) \\ W_1(s + \Delta_{111}) \\ W_1(s + \Delta_{211}) \\ W_1(s + 2\Delta_{111}) \\ W_1(s + 2\Delta_{211}) \end{bmatrix},$$

$$\Lambda = \begin{pmatrix} B \\ I_5 \end{pmatrix} = \begin{bmatrix} \lambda_{110} & \lambda_{111} & 0 & 0 & 0 \\ 0 & \lambda_{110} & 0 & \lambda_{111} & 0 \\ \lambda_{210} & 0 & \lambda_{211} & 0 & 0 \\ 0 & 0 & \lambda_{210} & 0 & \lambda_{211} \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\Phi = \begin{bmatrix} \phi_1(0) & \phi_1(\Delta_{111}) & \phi_1(\Delta_{211}) & \phi_1(2\Delta_{111}) & \phi_1(2\Delta_{211}) \\ \phi_1(\Delta_{111}) & \phi_1(0) & \phi_1(\Delta_{211}-\Delta_{111}) & \phi_1(\Delta_{111}) & \phi_1(2\Delta_{211}-\Delta_{111}) \\ \phi_1(\Delta_{211}) & \phi_1(\Delta_{111}) & \phi_1(\Delta_{211}-\Delta_{111}) & \phi_1(0) & \phi_1(2\Delta_{111}-\Delta_{211}) \\ \phi_1(2\Delta_{111}) & \phi_1(\Delta_{111}) & \phi_1(2\Delta_{111}-\Delta_{211}) & \phi_1(0) & \phi_1(2\Delta_{211}-\Delta_{111}) \\ \phi_1(2\Delta_{211}) & \phi_1(2\Delta_{211}-\Delta_{111}) & \phi_1(\Delta_{211}) & \phi_1(2\Delta_{211}-\Delta_{111}) & \phi_1(0) \end{bmatrix}.$$
\[ \Psi = \begin{bmatrix} \Psi_1 & 0 & 0 \\ 0 & \Psi_2 & 0 \\ 0 & 0 & \Psi_3 \end{bmatrix}, \]

\[ \Psi_1 = \begin{bmatrix} \psi_1(0) & \psi_1(\Delta_{111}) \\ \psi_1(\Delta_{111}) & \psi_1(0) \end{bmatrix}, \]

\[ \Psi_2 = \begin{bmatrix} \psi_2(0) & \psi_2(\Delta_{211}) \\ \psi_2(\Delta_{211}) & \psi_2(0) \end{bmatrix}, \]

and

\[ \Psi_3 = \begin{bmatrix} \psi_3(0) & \psi_3(\Delta_{111}) & \psi_3(\Delta_{211}) & \psi_3(2\Delta_{111}) & \psi_3(2\Delta_{211}) \\ \psi_3(\Delta_{111}) & \psi_3(0) & \psi_3(\Delta_{211} - \Delta_{111}) & \psi_3(\Delta_{111}) & \psi_3(2\Delta_{211} - \Delta_{111}) \\ \psi_3(\Delta_{211}) & \psi_3(\Delta_{211} - \Delta_{111}) & \psi_3(0) & \psi_3(\Delta_{211} - \Delta_{211}) & \psi_3(\Delta_{211}) \\ \psi_3(2\Delta_{111}) & \psi_3(\Delta_{111}) & \psi_3(2\Delta_{111} - \Delta_{211}) & \psi_3(0) & \psi_3(2\Delta_{211} - 2\Delta_{111}) \\ \psi_3(2\Delta_{211}) & \psi_3(2\Delta_{211} - \Delta_{111}) & \psi_3(2\Delta_{211}) & \psi_3(2\Delta_{211} - 2\Delta_{111}) & \psi_3(0) \end{bmatrix}, \]

where \( \phi_1(h) = \text{cov}\{W_i(s), W_i(s + h)\} \) and \( \psi_i(h) = \text{cov}\{e_i(s), e_i(s + h)\} \).

The next step in our augmented observation model-fitting approach is to obtain a sample estimate of \( \Sigma(\theta) \). Since \( Z^*(s) \) consists of \( Z(s) \) and some of its lagged values, the number of unique elements in \( \Sigma(\theta) \) is less than \( \frac{n^2(n+1)}{2} \). Let \( n_u \) be the number of unique elements in \( \Sigma(\theta) \) and let \( \sigma(\theta) \) be the \( n_u \times 1 \) vector of the unique elements. Each element of \( \sigma(\theta) \) can be estimated by the corresponding straightforward sample variance or covariance. Given observations \( Z(s), s \in D \), we can maximize the efficiency of each covariance estimate by using

\[ \text{cov}\{Z_i(s + h_1), Z_{i'}(s + h_2)\} = \frac{1}{N_{b_2 - b_1} - 1} \sum_{s_m - s_l = b_2 - b_1} (Z_i(s_l) - \bar{Z}_i)(Z_{i'}(s_m) - \bar{Z}_{i'}), \]

(4.4)

where \( N_{b_2 - b_1} \) is the number of pairs \( (s_l, s_m) \) such that \( s_m - s_l = h_2 - h_1 \). Let \( u \) be the \( n_u \times 1 \) sample estimate of \( \sigma(\theta) \) that can be obtained using (4.4) or some other approach. Because of the spatial dependence in \( Z(s) \), the covariances in (4.4) are biased. An approach for obtaining a bias-adjusted estimate of the variances and covariances is described and evaluated in Christensen and Amemiya (1999).
We propose estimating model parameters in $\theta$ by a least squares approach. Our estimator of $\theta$ is $\hat{\theta}$, the value of $\theta$ minimizing

$$SS(\theta) = (u - \sigma(\theta))' \hat{\Xi} (u - \sigma(\theta)) \quad (4.5)$$

over the parameter space for $\theta$, where $\hat{\Xi}$ is some estimated weight matrix. The logical choice for the weight matrix $\hat{\Xi}$ is the inverse of an estimate of $\text{var}\{u\}$. But, because of the artificial expansion in $Z^*$ and because of the large dimension of $u$, an estimated $\text{var}\{u\}$ is often singular. Alternately, any $\hat{\Xi}$ that approaches a positive definite matrix in large samples can be used. For simplicity, we use the identity matrix as $\hat{\Xi}$ in our simulation. Minimization of (4.5) needs to be carried out numerically using an iterative algorithm. We propose a modification of the Gauss-Newton algorithm which incorporates the constraint that each variance estimate is nonnegative and each covariance estimate satisfies the Cauchy-Schwarz inequality. Additional Marquardt-like modifications to the standard Gauss-Newton algorithm can be employed to improve the robustness of the algorithm in the presence of near-singular matrix inversions.

The $i$th step of such an algorithm for obtaining $\hat{\theta}^{(i)}$ from $\hat{\theta}^{(i-1)}$ is illustrated in Figure 3. This algorithm requires that the user select a range of values for $\lambda$, a parameter which specifies the potential "directions" to be considered at each step. These values for $\lambda$ are bounded below by $\lambda_{\text{start}}$ and bounded above by $\lambda_{\text{finish}}$. Additionally, the user selects a value for $h_{\text{finish}}$ which indicates the maximum number of times a step will be halved when seeking to find a value for $\hat{\theta}^{(i)}$ such that $SS(\hat{\theta}^{(i)}) < SS(\hat{\theta}^{(i-1)})$.

At the $i$th iteration of the algorithm, we tentatively choose our new estimate of $\theta$ to be

$$\hat{\theta}_{\lambda,h} = \hat{\theta}^{(i-1)} + \text{step}_{\lambda} \cdot 2^{-h}, \quad (4.6)$$

where $\text{step}_{\lambda}$ is calculated using

$$\text{step}_{\lambda} = \begin{cases} 
\left(F' \hat{\Xi} \hat{F}\right)^{-1} F' \hat{\Xi} (u - \sigma(\hat{\theta}^{(i-1)})), & \lambda = \lambda_{\text{start}}, \\
\left(F' \hat{\Xi} \hat{F} + 10^\lambda \cdot \text{diag}(F' \hat{\Xi} \hat{F})\right)^{-1} F' \hat{\Xi} (u - \sigma(\hat{\theta}^{(i-1)})), & \lambda \neq \lambda_{\text{start}},
\end{cases} \quad (4.7)$$
Figure 3 A modified Gauss-Newton algorithm for calculating $\hat{\theta}^{(i)}$. 

\[ \lambda = \lambda_{\text{start}} \\
\text{\[ h = 0 \]}

Calculate $\text{step}_\lambda$ using eqn. (4.7)

Let $\hat{\theta}_{\lambda, h} = \hat{\theta}^{(i-1)} + \text{step}_\lambda \cdot 2^{-h}$

Let $\hat{\theta}_{\lambda, h} = \text{Constrain}(\hat{\theta}_{\lambda, h})$

\[ \text{SS}(\hat{\theta}_{\lambda, h}) < \text{SS}(\hat{\theta}^{(i-1)})? \]
\[ \text{yes} \]
\[ \text{Let } \hat{\theta}^{(i)} = \hat{\theta}_{\lambda, h} \]

Increment $h$ by 1

\[ h = h_{\text{finish}}? \]
\[ \text{no} \]

Increment $\lambda$ by 1; $h = 0$

\[ \lambda = \lambda_{\text{finish}}? \]
\[ \text{no} \]

No improvement from $\hat{\theta}^{(i-1)}$ can be made

\[ \text{yes} \]
Note that when $\lambda = \lambda_{\text{start}}$, the step direction is the usual Gauss-Newton step direction. This is equivalent to the case where $\lambda = -\infty$. Note that as $\lambda \to \infty$, the direction of the step approaches the steepest descent direction (Marquardt, 1963). Thus, we begin by looking for an improved estimate of $\theta$ by using the Gauss-Newton step direction. If this approach yields no improvement, we look at step directions that approach the steepest descent step direction.

If step$_{\lambda}$ has already been found to be too large to yield $SS(\hat{\theta}_{\lambda,A}) < SS(\hat{\theta}^{(i-1)})$, then the step size is cut in half and this can be repeated until the step size has become $2^{-h_{\text{finish}}}$ of its original size. At this point, the “Constrain” function is applied to the estimate $\hat{\theta}_{\lambda,A}$ found using equation (4.6). Constrain($\hat{\theta}_{\lambda,A}$) will return $\hat{\theta}_{\lambda,A}$ which is equal to $\hat{\theta}_{\lambda,A}$ with the following modifications: (1) all variance parameter estimates in $\hat{\theta}_{\lambda,A}$ that were negative are set equal to zero, and (2) all covariances are then constrained to satisfy the Cauchy-Schwarz inequality (for example, $|\text{cov}\{e_1(s), e_1(s + \Delta)\}| \leq (\text{var}\{e_1(s)\} \cdot \text{var}\{e_1(s + \Delta)\})^{1/2} = \text{var}\{e_1(s)\}$).

If $SS(\hat{\theta}_{\lambda,A}) < SS(\hat{\theta}^{(i-1)})$, then $\hat{\theta}^{(i)} = \hat{\theta}_{\lambda,A}$. Otherwise, the step size is halved again and the resulting $\hat{\theta}_{\lambda,A}$ is checked for improvement over $\hat{\theta}^{(i-1)}$ in the same manner. When the number of times the step has been halved ($h$) reaches $h_{\text{finish}}$, then $\lambda$ is incremented, $h$ is reset to 0, and the process begins again from the calculation of step$_{\lambda}$ in equation (4.7). If all possible combinations of $\lambda$ and $h$ have been investigated and none yields an estimate $\hat{\theta}_{\lambda,A}$ such that $SS(\hat{\theta}_{\lambda,A}) < SS(\hat{\theta}^{(i-1)})$, then the procedure is terminated. The entire algorithm begins with the starting value $\theta^{(0)}$ and continues until either: (1) no improvement can be made ($SS(\hat{\theta}_{\lambda,A}) \geq SS(\hat{\theta}^{(i-1)})$ for all $\lambda \in \{\lambda_{\text{start}}, \lambda_{\text{start}} + 1, \ldots, \lambda_{\text{finish}}\}$ and $h \in \{0, 1, \ldots, h_{\text{finish}}\}$), (2) a maximum number of iterations ($i_{\max}$) is reached, or (3) $SS(\hat{\theta}^{(i-1)}) - SS(\hat{\theta}^{(i)}) < \delta$, where $\delta$ is some small tolerance value.

The estimator $\hat{\theta}$ can also be used for making inference about the parameter $\theta$, e.g., constructing confidence intervals and performing tests of significance. Recall that $\theta$ contains all original factor loadings $\lambda_{ijg}$. Thus, the selection of important shifts $\Delta_{ijg}$ can now be made statistically by testing $\lambda_{ijg} = 0$. For statistical inference using $\hat{\theta}$, we
need to obtain an estimated covariance matrix of \( \hat{\theta} \), which in turn requires an estimate of \( \Gamma = \text{var}\{u\} \). For this purpose, we use a practical assumption that the original \( p \)-dimensional process \( Z(s) \) has no spatial dependency beyond a certain range \( \tau \). We first let \( A(s) = \text{vecu} (Z^*(s) - \bar{Z}^*)(Z^*(s) - \bar{Z}^*)' \), where \( \text{vecu} \) is the operator such that \( \sigma(\theta) = \text{vecu} \Sigma(\theta) \). The assumption of limited spatial dependence for \( Z(s) \) implies \( \text{cov}\{A(s_i), A(s_j)\} = 0 \) for \( \|s_i - s_j\| > \tau_A \), where \( \tau_A \) depends on \( \tau \). In practice, \( \tau_A \) can be estimated by either inspecting each of the variograms and cross-variograms for the elements of \( A(s) \), or can be assumed based on subject matter knowledge. Then,

\[
\Gamma = \text{var}\{u\} = \text{var}\left\{ \frac{1}{N-1} \sum_{i=1}^{N} A(s_i) \right\} = \frac{1}{(N-1)^2} \left( N \text{var}\{A(s)\} + \sum_{i \neq j, \|s_i - s_j\| \leq \tau_A} \text{cov}\{A(s_i), A(s_j)\} \right) = X + Y,
\]

where

\[
X = \frac{N}{(N-1)^2} \text{var}\{A(s)\}, \quad Y = \frac{1}{(N-1)^2} \sum_{i \neq j, \|s_i - s_j\| \leq \tau_A} \text{cov}\{A(s_i), A(s_j)\}.
\]

We can estimate \( X \) and \( Y \) with

\[
\hat{X} = \frac{N}{(N-1)^3} \sum_s (A(s) - \bar{A})(A(s) - \bar{A})', \quad \hat{Y} = \frac{1}{(N-1)^2} \sum_{s_i \neq s_j, \|s_i - s_j\| \leq \tau_A} (A(s_i) - \bar{A})(A(s_j) - \bar{A})'.
\]

To obtain a nonnegative definite estimate of \( \Gamma = X + Y \), let the spectral decomposition of

\[
H = \hat{X}^{-1/2} \hat{Y} \hat{X}^{-1/2}
\]
be $Q^T Y Q'$. Write
\[
Y = \begin{bmatrix} Y_m & 0 \\ 0 & Y_{n_a - m} \end{bmatrix},
\]
where the $m$ diagonal elements of $Y_m$ are the $m$ eigenvalues of $H$ that are greater than or equal to $-1$, and $Y_{n_a - m}$ contains the $n_a - m$ eigenvalues of $H$ that are less than $-1$. Then an estimator $\hat{\Gamma}$ of $\Gamma$ is

\[
\hat{\Gamma} = \hat{X}^{1/2} \left( I_{n_a} + Q \begin{bmatrix} Y_m & 0 \\ 0 & -I_{n_a - m} \end{bmatrix} Q' \right) \hat{X}^{1/2}.
\] (4.8)

By replacing $Y_{n_a - m}$ with $-I_{n_a - m}$ in (4.8), $\hat{\Gamma}$ is constrained to be nonnegative definite. Using this $\hat{\Gamma}$, an estimate of the approximate variance of $\hat{\theta}$ is

\[
\text{var}\{\hat{\theta}\} = (\hat{P}' \hat{F} \hat{P})^{-1} \hat{P}' \hat{F} \hat{P} (\hat{P}' \hat{F} \hat{P})^{-1},
\] (4.9)

where
\[
\hat{F} = \frac{\partial \sigma(\theta)}{\partial \theta'}.
\]

This $\text{var}\{\hat{\theta}\}$ and $\hat{\theta}$ can be used in the standard way to perform statistical inference for $\theta$.

The researcher may also be interested in predicting or estimating the value of the underlying factors $W(s)$ at some location in the domain. The standard method of prediction in statistics is the best linear unbiased prediction (BLUP) under an estimated model. Since we do not assume the forms of distributions or covariance functions, and since we estimate only covariances of $W(s)$ and $e_i(s)$ appearing in $\Sigma(\theta) = \text{var}\{Z^*\}$, we propose a “limited-information BLUP” of $W(s)$ based on $Z^*(s)$ instead of using all $Z(s), s \in D$. Our predictor of $W(s)$ at a particular location $s \in D$ is

\[
\hat{W}(s) = \hat{E}\{W(s)\} + \hat{\text{cov}}\{W(s), Z^*(s)\} [\text{var}\{Z^*(s)\}]^{-1} (Z^*(s) - \hat{E}\{Z(s)\}),
\] (4.10)

where all quantities with “hats” are obtained from $\hat{\theta}$ and $\hat{Z}$. 
5 Simulation results

In order to evaluate the performance of this methodology, a simulation study was conducted. For a \(30 \times 30\) square lattice domain \(\mathbf{D}\) with the same row and column grid distance, 3-dimensional data were generated from

\[
\begin{align*}
Z_1(s) &= \gamma_1 + \lambda_{110} W_1(s) + \lambda_{111} W_1(s + \Delta_{111}) + e_1(s), \\
Z_2(s) &= \gamma_2 + \lambda_{210} W_1(s) + \lambda_{211} W_1(s + \Delta_{211}) + e_2(s), \\
Z_3(s) &= W_1(s) + e_3(s),
\end{align*}
\]

(5.1)

where the shift parameters are \(\Delta_{111} = (2,2)\) and \(\Delta_{211} = (2,0)\), the true parameter values are \(\gamma_1 = \gamma_2 = 0\), \(\lambda_{110} = 0.5\), \(\lambda_{111} = 1\), \(\lambda_{210} = 1\), \(\lambda_{211} = 0.5\), and \(W_1(s)\) and \(e_i(s)\), \(i = 1,2,3\), are independent zero-mean spatial processes having isotropic spherical variograms with zero nugget. These error processes \(e_i(s)\), \(i = 1,2,3\), have common range \(\tau_c\) and partial sill \(\sigma_e^2\). Denote the range and partial sill of \(W_1(s)\) by \(\tau_{W_1}\) and \(\sigma_{W_1}^2\). The error processes \(e_i(s)\) were generated as normally distributed processes. For the factor process \(W_1(s)\), we considered two distributional forms, normal and linearly transformed lognormal. Four different data generation scenarios considered were:

(a) normally distributed factor, \(\tau_{W_1} = 4\), \(\tau_c = 2\), \(\sigma_{W_1}^2 = 7\), and \(\sigma_e^2 = 3\);
(b) normally distributed factor, \(\tau_{W_1} = 10\), \(\tau_c = 5\), \(\sigma_{W_1}^2 = 9\), and \(\sigma_e^2 = 1\);
(c) lognormally distributed factor, \(\tau_{W_1} = 4\), \(\tau_c = 2\), \(\sigma_{W_1}^2 = 9\), and \(\sigma_e^2 = 1\);
(d) lognormally distributed factor, \(\tau_{W_1} = 10\), \(\tau_c = 5\), \(\sigma_{W_1}^2 = 7\), and \(\sigma_e^2 = 3\).

We expect our statistical procedures to have more difficulty in scenarios (b) and (d) with larger ranges of spatial correlation than (a) and (c).

For each of the 4 scenarios, 1000 samples were generated using model (5.1). From each sample, the least squares estimator \(\hat{\theta}\) of (4.5) was obtained using an identity matrix as the weight in (4.5), and assuming \(\Delta_{111}\) and \(\Delta_{211}\) are given. The corresponding approximate asymptotic standard errors were calculated using (4.9), and an approximate 95% confidence interval was calculated for \(\lambda_{110}\), \(\lambda_{111}\), \(\lambda_{210}\), and \(\lambda_{211}\). Figure 4 gives boxplots of \(\hat{\lambda}_{110}\), \(\hat{\lambda}_{111}\), \(\hat{\lambda}_{210}\), and \(\hat{\lambda}_{211}\) after subtracting the true value, and the empirical
coverage probability for the approximate 95% confidence interval in parentheses below each $\lambda$ label. In all four scenarios, the least squares estimators of the $\lambda_{ij}$'s are nearly median unbiased and are roughly symmetrically distributed around the true value. Thus, our estimation and model fitting approach works well for nonnormally distributed data and a fairly large range of spatial dependence. Under scenarios (a), (b), and (c), the actual coverage probabilities are all very close to the nominal 95%. Even under scenario (d) with low reliability ratio $\left(\frac{\sigma^2_{\nu}}{\sigma^2_{\nu} + \sigma^2_e}\right)$, nonnormal factor distribution, and large range of spatial dependence, the actual coverage probabilities are still reasonably close to the nominal level. Thus, our inference procedures using the least squares estimator and the asymptotic covariance matrix (4.9) are reasonably accurate for practical use.

To illustrate the use of the limited-information BLUP (4.10), we used one of the data sets generated from scenario (d), and obtained a predicted value $\hat{W}_i(s)$ of $W_i(s)$ for every $s \in D$. Figure 5 shows image plots of $Z_1(s)$, $Z_2(s)$, $Z_3(s)$, $W_1(s)$, and $\hat{W}_1(s)$. The predicted factor process $\hat{W}_1(s)$ is a smooth process which clearly delineates the high and low areas of the true $W_1(s)$ process and has less variability than any $Z_i(s)$.

Note that although $Z_3(s) = W_1(s) + e_3(s)$, $\hat{W}_1(s)$ (combining $Z_1(s)$, $Z_2(s)$, and $Z_3(s)$) is actually a much better predictor of the true factor process $W_1(s)$ than $Z_3(s)$. In fact, $\frac{1}{900} \sum_{s \in D} (\hat{W}_1(s) - W_1(s))^2 = 1.05$ while $\frac{1}{900} \sum_{s \in D} (Z_3(s) - W_1(s))^2 = 2.72$.

The improvement of $\hat{W}_1(s)$ over $Z_3(s)$ in predicting $W_1(s)$ is more pronounced when the reliability ratio is low and the error processes are dominated by measurement error as opposed to spatial dependence. Figure 6 shows image plots of $Z_1(s)$, $Z_2(s)$, $Z_3(s)$, $W_1(s)$, and $\hat{W}_1(s)$ under a scenario where the $W_1(s)$ process is as given in scenario (d), but the $e_i(s)$ processes have $\sigma^2_e = 7$ and nugget = 5. Thus, this process has reliability ratio $\left(\frac{\sigma^2_{\nu}}{\sigma^2_{\nu} + \sigma^2_e}\right) = 0.5$ and the error processes are dominated by measurement error. Here, we see that $\hat{W}_1(s)$ is much less noisy than any $Z_i(s)$ and is vastly superior to $Z_3(s)$ in predicting $W_1(s)$. For these data, $\frac{1}{900} \sum_{s \in D} (\hat{W}_1(s) - W_1(s))^2 = 1.72$ while $\frac{1}{900} \sum_{s \in D} (Z_3(s) - W_1(s))^2 = 7.50$.

We conclude that the use of the generalized shifted-factor model can be of great benefit to researchers involved in the modeling of multivariate spatial data, and that the
Figure 4: Boxplots of $\hat{\lambda}_{ij} - \lambda_{ij}$ in 4 scenarios with empirical coverage probabilities in parentheses.
Figure 5 Image plots of $Z_1(s)$, $Z_2(s)$, $Z_3(s)$, $W_1(s)$, and $\hat{W}_1(s)$ for data where the reliability ratio is 0.7 and error processes contain no measurement error. Black regions have high values and white regions have low values.

methodology discussed herein provides appropriate inference and prediction tools for a variety of problems.

References


Figure 6 Image plots of $Z_1(s)$, $Z_2(s)$, $Z_3(s)$, $W_1(s)$, and $\hat{W}_1(s)$ for data where the reliability ratio is 0.5 and error processes are dominated by measurement error. Black regions have high values and white regions have low values.


LATENT VARIABLE MODELING OF MULTIVARIATE SPATIAL DATA WITH POTENTIAL LAGGED-DEPENDENCIES

A paper to be submitted to Journal of the American Statistical Association

William F. Christensen and Yasuo Amemiya

Abstract

Multivariate spatial or geo-referenced data arise naturally in disciplines such as ecology, agriculture, geology, and atmospheric sciences. Statistical analysis of such data using a latent variable model is considered. We propose a general model that incorporates spatial correlation and potential lagged or shifted dependencies. Identification and interpretation of the model are discussed. Model fitting, parameter estimation, and inference procedures are developed without restrictive assumptions on distribution and covariance function forms. Also, a method for predicting the underlying latent variable process is introduced. The properties and usefulness of the proposed approaches are assessed by asymptotic theory and simulation. An example from precision agriculture is also discussed.

1 Introduction

Latent variable analysis is a statistical tool for describing and modeling the underlying structure in multivariate data. The analysis is widely used in situations where theoretical quantities can be measured only with errors and/or observed variables can be explained (except for error) by a smaller number of unobservable factors. For example, in the social and behavioral sciences researchers are often interested in modeling
abstract concepts such as abilities, attitudes, or tendencies. While these concepts are not directly observable, multi-item questionnaires, tests, or other measurement processes can be employed in order to measure quantities that are related to the latent concepts of interest. In such situations, the latent variable model describes the general relationship among the observed variables and the unobservable factors of interest.

This type of modeling can also be of interest in the physical and biological sciences, where a large number of variables related to some underlying quantities or trend variables can be measured easily. Multivariate data in these areas may be geo-referenced, exhibiting spatial correlation. In the area of precision farming, researchers often collect high-dimensional data sets including measurements of numerous soil characteristics, topological features, and yield parameters at each location of interest in a field. The precise geographic information provided by the now commercially available Global Positioning System satellites has given farmers the ability to implement practices such as variable-rate fertilizer and pesticide application. Such tools offer the potential for maximizing profit and minimizing damage to the environment. However, while rich and increasingly precise geo-referenced agricultural data are now available, the existing models for multivariate spatial data are few in number and often limited in scope.

Latent variable analysis is a model-based approach in the usual statistical sense—that is, statistical analysis is performed by fitting and assessing various aspects of a model formulated either to represent subject-matter theory or for exploratory modeling. Thus, unlike in purely descriptive methods, scientific or model building issues can be addressed meaningfully by performing appropriate statistical analysis. Development of such model fitting and inference procedures for spatial problems requires some care, because of the nonstandard spatial dependencies. Our approach to this problem begins with introducing a general model that is meaningful in applications and allows useful statistical inferences. Then, we develop estimation, prediction, and inference procedures that are broadly applicable and justified without restrictive assumptions about distributional forms and spatial correlation structure.

Some approaches for latent variable analysis of spatial data have been discussed in
the literature. Wartenberg (1985) discussed a factor analytic approach in two steps. The first step involves a dimension reduction ignoring spatial correlation structure and the second involves an analysis of the spatial structure of each factor. Wackernagel (1988) used models that require the covariances among observed variables to be symmetric. Grunsky and Agterberg (1991, 1992) provided a technique for exploring lagged relationships among variables. Goovaerts, Sonnet, and Navarre (1993) and others employ the “coregionalization” technique which places strong constraints on the forms of the variograms and cross-variograms for the data. Cook et al. (1994) introduce the shifted-lag model which allows for asymmetric relationships among observed variables. For such a model, Majure and Cressie (1997) discussed graphical methods for identifying shifts. The models used in the literature have been rather restrictive, and formal statistical model fitting and inference procedures have not been addressed.

We present an approach for latent variable modeling of multivariate geo-referenced data which differs considerably from the previous work. In Section 2, we introduce a very general model similar to the dynamic factor analysis model (Molenaar, 1985) in the time series factor analysis literature. Statistical issues related to estimation and model fitting are discussed in Section 3. Inference and prediction procedures are developed in Section 4 along with associated asymptotic theory. Section 5 presents extensive simulation results. An example from precision agriculture is given in Section 6.

2 The model and proposed approach

Define $Z^2$ to be a two-dimensional rectangular grid. Suppose that we observe $Z(s) = (Z_1(s), ..., Z_p(s))^\prime$, a $p$-dimensional second-order stationary process with $s \in D$ and $D \subset Z^2$. Consider a situation where observed variables $Z_i(s), i = 1, ..., p$, can be expressed as linear functions of an unobservable $k \times 1$ factor process $W(s) = (W_1(s), ..., W_k(s))^\prime$ except for the zero-mean error processes $e_i(s), i = 1, ..., p$, and where all the inter-relationships among the $p Z_i(s)$ processes are explained by $W(s)$, not by the $e_i(s)$ processes. A natural model for such a situation is a factor analysis model satisfying that each $Z_i(s)$ is some linear function of $W(s)$ plus $e_i(s)$, and that the $p + 1$ processes $W(s), e_1(s), ..., e_p(s)$ are
independent processes. However, unlike the standard independent sample case, each of the p+1 spatial processes $W(s), e_1(s), ..., e_p(s)$ is assumed to be a second-order stationary process with an unspecified general spatial covariance structure for $s \in D$. Note that none of the processes is assumed to be isotropic. In the context of a simple model, Cook et al. (1994) suggested that an observation $Z_i(s)$ at a particular location may be related to a value $W_j(s + \Delta)$ at a different location $s + \Delta$ ($\Delta \neq 0$). Inclusion of shifted-factor terms of the form $W_j(s + \Delta)$ allows the model to represent possible asymmetric lagged dependencies among the $Z_i(s)$ processes, i.e., $\text{cov}\{Z_i(s), Z_l(s + h)\} \neq \text{cov}\{Z_i(s + h), Z_l(s)\}, i \neq l$. Following the ideas of Anderson (1963) and Molenaar (1985) for multiple-lag time series factor analysis models, we generalize the simple shifted-lag model of Cook et al. (1994) to an internally consistent model that allows systematic modeling and inference regarding the existence and type of shifts. Such a general model is

$$Z_i(s) = \gamma_i + \sum_{j=1}^{k} \sum_{g=0}^{m_{ij}} \lambda_{ijg} W_j(s + \Delta_{ijg}) + e_i(s), \quad i = 1, ..., p,$$

(2.1)

where $\Delta_{ij0} \equiv 0$, and $m_{ij}$ is the number of shifted values of the jth factor $W_j(s)$ assumed to be related to the ith observed variable $Z_i(s)$.

As in the standard factor analysis model, model (2.1) has some indeterminacy related to the fact that the $k \times 1$ unrestricted factor process $W(s)$ is unobservable. Note that $W(s)$ can be linearly transformed without altering the form of the model. In addition, the shifts $\Delta_{ijg}$ are not uniquely defined due to the arbitrary nature of the spatial reference for an unobservable process. To remove this factor/shift indeterminacy, we use the so-called errors-in-variables parameterization to obtain an identifiable form of model (2.1). Thus, with possible re-ordering of the p variables $Z_i(s)$, we consider the identifiable generalized shifted-factor model written as

$$Z_i(s) = \gamma_i + \sum_{j=1}^{k} \sum_{g=0}^{m_{ij}} \lambda_{ijg} W_j(s + \Delta_{ijg}) + e_i(s), \quad i = 1, ..., p - k,$$

$$Z_i(s) = W_{i-p+k}(s) + e_i(s), \quad i = p - k + 1, ..., p,$$

(2.2)

where $\Delta_{ij0} \equiv 0, i = 1, ..., p - k, j = 1, ..., k$. In addition to removing the factor/shift indeterminacy, this model permits a meaningful interpretation of a factor $W_j(s)$ as the
true (error-free) value of $Z_{p-k+j}(s)$. We refer to $W_j(s + \Delta_{ij})$ as a "shifted factor" when $\Delta_{ij} \neq 0$. For example, if $p = 3$, $k = 1$, and $m_{11} = m_{21} = 1$, then we write the model as

$$
\begin{align*}
Z_1(s) &= \gamma_1 + \lambda_{110}W_1(s) + \lambda_{111}W_1(s + \Delta_{111}) + e_1(s), \\
Z_2(s) &= \gamma_2 + \lambda_{210}W_1(s) + \lambda_{211}W_1(s + \Delta_{211}) + e_2(s), \\
Z_3(s) &= W_1(s) + e_3(s).
\end{align*}
$$

(2.3)

By including multiple shifted terms along with $W_j(s)$, model (2.2) can be considered a very flexible explanatory model that also allows checking for the existence of shifts $\Delta_{ij}$ by testing $\lambda_{ij} = 0$. Thus, we consider model (2.2) with known, given values of the $m_{ij}$'s and $\Delta_{ij}$'s. The $m_{ij}$'s and $\Delta_{ij}$'s that are suspected or proposed can be chosen either based on subject-matter meaning or by including certain neighboring shifts for exploratory model building purposes. See the illustration in Section 6. Majure and Cressie (1997) and Christensen and Amemiya (1999) discuss graphical methods for identifying possible $\Delta_{ij}$'s.

Given $m_{ij}$ and $\Delta_{ij}$, we can consider fitting model (2.2) by developing procedures for estimating the identified parameters $\gamma_i$, $\lambda_{ij}$, and distributional parameters for $W(s)$ and $e_i(s)$. To develop a broadly applicable exploratory procedure, we do not assume particular distributional forms of $W(s)$ and $e_i(s)$. Thus, the use of some moment-based methods is reasonable. In the standard random sample factor analysis, the first two sample moments of the $p \times 1$ observation $Z(s)$ are used. However, with the shifted factors $W_j(s + \Delta_{ij})$ in the model, the first two moments of $Z(s)$ may not be sufficient to identify a fit of the model. Also, because of the spatial dependency of $Z(s)$, the use of the first two moments of $Z(s)$ may not be most efficient even if the model can be estimated. Thus, some lagged sample cross-covariances need to be included in the moment-based analysis in addition to the contemporaneous second moments of $Z(s)$. Natural lags to include for fitting model (2.2) are those related to the $\Delta_{ij}$ appearing in the model or neighboring lags.

To express all the second moments to be used in a manner consistent with model (2.2), we augment $Z(s)$ with lagged values of the form $Z_i(s + \Delta)$ to create an expanded vector
$Z^*(s)$ of dimension $p^*$ ($> p$), and consider the elements of $\Sigma = \text{var}\{Z^*(s)\}$ to be the second moments to be fitted. Under model (2.2), $Z^*(s)$ also satisfies a factor analytic model

$$Z^*(s) = \gamma + \Lambda W^*(s) + e^*(s), \quad (2.4)$$

where $\gamma$ consists of $\gamma_i$'s and zeros, $\Lambda$ consists of $\lambda_{ij}$'s and constants, $W^*(s)$ is a $k^* \times 1$ expanded version of $W(s)$, and $e^*(s)$ contains $e_i(s)$'s and their lagged values. Hence, the $p^* \times p^*$ $\Sigma$ can be written as

$$\Sigma = \Lambda \Phi \Lambda' + \Psi, \quad (2.5)$$

where the $k^* \times k^*$ $\Phi = \text{var}\{W^*(s)\}$ and the $p^* \times p^*$ $\Psi = \text{var}\{e^*(s)\}$. We do not assume any special structure for $\Phi$ and $\Psi$. But, because of the second-order stationarity of $Z(s)$, $W(s)$, and $e(s)$, $\Sigma$, $\Phi$, and $\Psi$ contain many duplicate elements (beyond symmetry). For example, $\text{var}\{Z_i(s)\} = \text{var}\{Z_i(s + \Delta)\}$. Also, due to the independence of the $e_i(s)$'s, $\Psi$ contains many zeros. We write $\Sigma = \Sigma(\theta)$ where $\theta$ is a vector containing all $\lambda_{ij}$ in model (2.2) and all distinct elements in $\Phi$ and $\Psi$.

Our moment-based approach obtains an estimator of $\theta$ using some sample second moment estimators of the unique elements of $\Sigma(\theta)$. Once $\hat{\theta}$ containing $\hat{\lambda}_{ij}$ is obtained, we estimate the remaining parameters by

$$E\{W_j(s)\} = \bar{Z}_{p-k+j} = \frac{1}{N} \sum_{s \in D} Z_{p-k+j}(s), \quad j = 1, \ldots, k.$$  

$$\hat{\gamma}_i = \bar{Z}_i - \sum_{j=1}^{k} \sum_{g=0}^{m_j} \hat{\lambda}_{ijg} \bar{Z}_{p-k+j}, \quad i = 1, \ldots, p - k, \quad (2.6)$$

where $N$ is the sample size. Hence, the estimator development in Section 3 concentrates on obtaining $\hat{\theta}$ based on an estimator of $\Sigma(\theta)$.

Though several approaches to creating the expanded observation vector $Z^*(s)$ are possible, we present a systematic one called "the simple expansion" which is appropriate for model (2.2). The augmented observation vector $Z^*(s)$ is obtained as the simple expansion of $Z(s)$ following the steps:
1. For each $\Delta_{ijg}$ in the model $(i = 1, \ldots, p - k, j = 1, \ldots, k, g = 0, \ldots, m_{ij})$, include the shifted variable $Z_i(s + \Delta_{ijg})$ in $Z^*(s)$.

2. For all $W_j(s + \Delta_{ijg})$ appearing in the model for each of the variables added in 1, include the corresponding $Z_{p-k+j}(s + \Delta_{ijg})$ in $Z^*(s)$.

As an example, consider applying the simple expansion to model (2.3). In step 1, we include $Z_1(s)$, $Z_1(s + \Delta_{111})$, $Z_2(s)$, and $Z_2(s + \Delta_{211})$. Note that model (2.3) written for $Z_1(s + \Delta_{111})$ has $W_1(s + 2\Delta_{111})$ in the right-hand side, and that $W_1(s + 2\Delta_{211})$ appears in the model for $Z_2(s + \Delta_{211})$. Then, in step 2, we include $Z_3(s)$, $Z_3(s + \Delta_{111})$, $Z_3(s + \Delta_{211})$, $Z_3(s + 2\Delta_{111})$, and $Z_3(s + 2\Delta_{211})$. Thus, the simple expansion $Z^*(s)$ for this model is

$$Z^*(s) = \begin{bmatrix}
Z_1(s) \\
Z_1(s + \Delta_{111}) \\
Z_2(s) \\
Z_2(s + \Delta_{211}) \\
Z_3(s) \\
Z_3(s + \Delta_{111}) \\
Z_3(s + \Delta_{211}) \\
Z_3(s + 2\Delta_{111}) \\
Z_3(s + 2\Delta_{211})
\end{bmatrix}.$$

Note that when a model contains no shifted factors (no nonzero values of $\Delta_{ijg}$) then the expanded observation vector $Z^*(s)$ under the simple expansion is equivalent to the original observation vector $Z(s)$. For this and other models, we may choose to create an expanded observation vector that is larger than what the simple expansion would prescribe. This can be carried out following the approach given in the simple expansion steps. That is, when we add shifted versions of any of the first $p - k$ variables to $Z^*(s)$, we also add the shifted versions of the final $k$ variables corresponding to shifted factor variables appearing in the model for the added observations.
For a given augmented observation vector $\mathbf{Z}^*(s)$, we obtain $\Sigma(\theta)$ in (2.5). If the dimension of $\theta$ is less than or equal to the number of distinct elements in $\Sigma(\theta)$, the model fitting and estimation can be carried out using a moment-based method. This "counting rule" identification is simple to verify in practice. For models employing a reasonable number of shifted factors, the simple expansion vector $\mathbf{Z}(s)$ can be shown to provide the counting rule identification if the $\Delta_{ijg}$'s in model (2.2) do not depend on $i$ (common over $i$), and if $k \leq (2p + 1 - \sqrt{8p + 1})/2$. The latter condition is equivalent to the identification of model (2.2) with $\Delta_{ijg} = 0$ for $g > 0$, i.e., the corresponding standard factor analysis model with no shifts. Moment-based estimation methods based on $\mathbf{Z}^*(s)$ are introduced in the next section.

3 Estimation issues

To develop a moment-based estimation procedure for model (2.2), we first consider estimation of $\Sigma(\theta)$ in (2.5) based on the augmented observation vector $\mathbf{Z}^*(s)$. As noted in the previous section, $\Sigma(\theta)$ contains redundant elements beyond the symmetry. We denote a vector consisting of distinct elements in $\Sigma(\theta)$ by $\text{veu} \Sigma(\theta)$. Each element of $\text{veu} \Sigma(\theta)$ is a variance or lagged/cross covariance for the elements of $\mathbf{Z}(s)$, and thus can be estimated by the naive sample corrected-product-moment covariance using all possible terms in the sample. For example,

$$\text{cov}\{Z_i(s + h_1), Z_{i'}(s + h_2)\} = \frac{1}{N_{h_2-h_1} - 1} \sum_{s_i,s_i' \in h_2-h_1, s_m - s_i = h_2 - h_1} (Z_i(s_i) - \bar{Z}_i)(Z_{i'}(s_m) - \bar{Z}_{i'}),$$

where $N_{h_2-h_1}$ is the number of pairs $(s_i, s_m)$ such that $s_m - s_i = h_2 - h_1$. For notational simplicity, we ignore the different number of terms used for different lagged covariances and write such a naive estimator of $\text{veu} \Sigma(\theta)$ by $\text{veu} \Sigma_{un}$, where

$$\Sigma_{un} = \frac{1}{N - 1} \sum_{i=1}^{N} (\mathbf{Z}^*(s_i) - \bar{\mathbf{Z}}^*)(\mathbf{Z}^*(s_i) - \bar{\mathbf{Z}}^*)',$$

$\bar{\mathbf{Z}}^* = \frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}^*(s_i)$, and $N$ is the number of observations in the random field. In any discussion related to asymptotics as $N \to \infty$, we use the expression of type (3.1) to represent the actual estimator of $\text{veu} \Sigma(\theta)$. 
Because of the spatial dependence, $S_{un}$ in (3.1) is known to have nonnegligible bias, i.e.,

$$E\{S_{un}\} = \Sigma(\theta) - \frac{1}{N(N - 1)} \sum_{i \neq j} \text{cov}\{Z^*(s_i), Z^*(s_j)\}$$

$$= \Sigma(\theta) + O(1/N),$$

(3.2)

where the bias is related to $\text{var}\{Z^*\}$. See, e.g., Cressie (1993, Section 3.5.4). Since our moment-based approach is based on a good estimator of $\text{vec}\Sigma(\theta)$, we may be interested in reducing the bias of $S_{un}$. For this, assume a practical condition that $\text{cov}\{Z(s), Z(s + h)\} = 0$ if $||h|| > \tau_0$. In practice, it is possible to estimate $\tau_0$ by graphically inspecting the variograms and cross-variograms for the observed $Z(s)$. Then, for some $\tau$ related to $\tau_0$, $\text{cov}\{Z^*(s), Z^*(s + h)\} = 0$ if $||h|| > \tau$. Under this condition, the bias term in (3.2) can be estimated by

$$\hat{B} = \frac{1}{N(N - 1)} \sum_{i \neq j} \frac{1}{||s_i - s_j|| \leq \tau} (Z^*(s_i) - \bar{Z}^*)(Z^*(s_j) - \bar{Z}^*'),$$

(3.3)

and we can consider a bias-adjusted estimator $\text{vec} S_{ad}$ of $\text{vec} \Sigma(\theta)$ by

$$S_{ad} = S_{un} - \hat{B}.$$

(3.4)

It can be shown that

$$E\{S_{ad}\} = \Sigma(\theta) + O(1/N^2).$$

Note that $S_{ad}$ in (3.4) may not be nonnegative definite. To assure the nonnegative definiteness of our estimator, we suggest the use of a slightly modified version

$$S_{ad} = S_{un} + S_{un}^{1/2} \left( I_{p^*} + \bar{Q} \begin{bmatrix} \Upsilon_m & 0 \\ 0 & -I_{p^*-m} \end{bmatrix} \bar{Q}' \right) S_{un}^{1/2},$$

(3.5)

where $p^*$ is the dimension of $Z^*(s)$, $\bar{Q}\Upsilon\bar{Q}'$ is the spectral decomposition of $\bar{H} = S_{un}^{-1/2}(\hat{B})S_{un}^{-1/2}$, and $\Upsilon_m$ is a diagonal matrix containing the $m$ eigenvalues of $\bar{H}$ that are greater than or equal to -1. In Section 5.1, a simulation study is used to compare the use of $S_{un}$ and $S_{ad}$. 
To obtain the augmented observation moment-based estimator, let $\text{vec}uS$ denote the sample estimator of $\text{vec} \Sigma(\theta)$ to be used in our method, where $S$ is either $S_{\text{un}}$ or $S_{\text{ad}}$. Our moment-based approach is to use some least squares estimate of $\theta$ by minimizing a distance between $\text{vec}uS$ and $\text{vec} \Sigma(\theta)$. Let $\Theta$ be the parameter space for $\theta$. Then, our least squares estimator $\hat{\theta}$ is defined to be the value of $\theta \in \Theta$ minimizing

$$SS(\theta) = [\text{vec}uS - \text{vec} \Sigma(\theta)]' \hat{\Sigma} [\text{vec}uS - \text{vec} \Sigma(\theta)],$$

(3.6)

where $\hat{\Sigma}$ is some weight matrix possibly depending on the sample. We consider three choices for $\hat{\Sigma}$.

A natural choice for the weight matrix $\hat{\Sigma}$ in (3.6) is that matrix yielding the estimated generalized least squares (GLS) estimator of $\theta$. Note that the difference between $S_{\text{un}}$ and $S_{\text{ad}}$ is of small order. Thus, acting as if only $S_{\text{un}}$ is being used as $S$, we can write

$$\Gamma = \text{var}\{\text{vec}uS\}$$

$$= \text{var}\left\{\frac{1}{N-1} \sum_{i=1}^{N} A(s_i)\right\},$$

(3.7)

where $A(s_i) = \text{vec}(Z^*(s_i) - \bar{Z}^*)(Z^*(s_i) - \bar{Z}^*)'$. If the observed $Z(s)$ has no spatial dependency beyond the range $\tau_0$, then $A(s)$ has zero spatial correlation beyond some range $\tau_A$. Alternatively, $\tau_A$ can be directly estimated by inspecting the variograms and cross-variograms for the components of $A(s)$. Given such a $\tau_A$, we can estimate $\Gamma$ in (3.7) by

$$\frac{1}{(N-1)^2} \sum_{i \neq j, \|s_i - s_j\| \leq \tau_A} (A(s_i) - \bar{A})(A(s_j) - \bar{A})',$$

where $\bar{A} = \frac{1}{N} \sum_{i=1}^{N} A(s_i)$. To insure the nonnegative definiteness of our estimate, we use

$$\hat{\Gamma} = \hat{\Sigma}^{1/2} \left( I_{nu} + Q^{(A)} \begin{bmatrix} Y_m^{(A)} & 0 \\ 0 & -I_{nu-m} \end{bmatrix} Q^{(A)} \right) \hat{\Sigma}^{1/2}$$

(3.8)

where

$$\hat{\Sigma} = \frac{N}{(N-1)^3} \sum_{s} (A(s) - \bar{A})(A(s) - \bar{A})',$$

$$\hat{Y} = \frac{1}{(N-1)^2} \sum_{s_i \neq s_j, \|s_i - s_j\| \leq \tau_A} (A(s_i) - \bar{A})(A(s_j) - \bar{A})',$$
\(Q^{(A)} Y^{(A)} Q^{(A)'}\) is the spectral decomposition of \(H^{(A)} = \hat{X}^{-1/2} \hat{Y} \hat{X}^{-1/2}\), and \(Y^{(A)}_m\) is an \(m \times m\) diagonal matrix containing the \(m\) eigenvalues of \(H^{(A)}\) that are greater than or equal to \(-1\). Since this \(\hat{\Gamma}\) may not be nonsingular, we use \(\hat{\Gamma}^+\) or the inverse of a finitely modified \(\hat{\Gamma}\) as the GLS weight \(\hat{\Sigma}_{GLS}\). The use of \(\hat{\Sigma}_{GLS}\), the inverse of the fourth-order sample moment matrix, may lead to large variability of \(\hat{\theta}\) in finite samples.

The second approach to choosing \(\hat{\Sigma}\) in (3.6) is to use some kind of iteratively re-estimated weight. If \(Z^*(s)\) is independent and identically distributed (i.i.d.) normal, the fourth moment matrix \(V[\Sigma(\theta)] = \text{var}\{\text{vec} S_{un}\}\) is just a simple function of second moments. Thus, although the i.i.d. normal assumption is not valid, this approach provides a simple second-moment-based iteratively estimated weight.

In order to obtain an explicit expression for \(V[\Sigma(\theta)]\), we define some additional notation. After \(\text{vec} E\) has been defined for a given \(p^* \times p^*\) symmetric matrix \(E\), we can find a \(p^*^2 \times n_u\) matrix \(\kappa_u\) such that \(\text{vec} E = \kappa_u \text{vec} E\). Define the Moore-Penrose generalized inverse of \(\kappa_u\) to be \(\kappa_u^+ = (\kappa_u' \kappa_u)^{-1} \kappa_u'\). Under the assumption that \(Z^*(s)\) is i.i.d. normal, we can use the proof of Lemma 4.A.1 of Fuller (1987) to show that

\[
V[\Sigma(\theta)] = \text{var}\{\text{vec} S_{un}\} = \frac{2}{N-1} \kappa_u^+ [\Sigma(\theta) \otimes \Sigma(\theta)] \kappa_u^{+/'}.
\]

Thus, our iteratively re-weighted pseudo-normal estimation procedure obtains the \(j\)th step estimate minimizing (3.6) with \(\hat{\Sigma} = (V[\Sigma(\theta)^{[j-1]}])^{-1}\). The starting value \(\hat{\theta}^{[0]}\) is obtained by minimizing \(SS(\theta)\) with \(\hat{\Sigma} = (V[S])^{-1}\), where the function \(V[\cdot]\) is defined in (3.9).

The third choice of \(\hat{\Sigma}\) in (3.6) is the identity matrix \(I_{n_u}\), where \(n_u\) is the dimension of \(\text{vec} S\). This choice not depending on the sample provides a simplicity that may be attractive in small to moderate samples.

### 4 Inference and prediction

Our general model (2.2) allows us to address relevant scientific or modeling questions by testing hypotheses concerning parameters. For example, the existence or inclusion
of a certain shifted factor term can be examined by testing the corresponding \( \lambda_{ijg} = 0 \).
Also, studying the spatial or lagged covariances of \( W(s) \) and \( \varepsilon_i(s) \) appearing in \( \Sigma(\theta) \) may be of modeling interest, and confidence intervals for parameters may be useful. Such inference can be obtained based on the least squares estimator \( \hat{\theta} \) in Section 3 and its asymptotic distribution.

To derive the limiting distribution of \( \hat{\theta} \), we use a central limit theorem for a sample mean in a random field as the field increases in domain. For our use, we need a version for multivariate data with explicitly stated normalization constant and limiting variance expressions. The following theorem presents such a result under assumptions appropriate for our practical situations. In this sense, Theorem 1 differs from some central limit theorems in the literature, e.g., Rosenblatt (1985) and Bolthausen (1982). The proof of Theorem 1 is given in the Appendix. We define \( C_X(h) = \text{cov}\{X(s), X(s + h)\} \).

**Theorem 1 (Central limit theorem for multivariate random fields)** Let \( X(s) = (X_1(s), ..., X_p(s)) \) be a \( p \)-variate zero-mean second-order stationary random field with \( s \in D \), where \( D \) is an \( n_1 \times n_2 \) subset of a rectangular grid. Assume

(i) there exists a finite \( b \) such that \( E\{ |X_i(s)|^3 \} \leq b. i = 1, ..., p. \)

(ii) there exists a \( \tau \) such that \( \{X(s), s \in D_1\} \) is independent of \( \{X(s), s \in D_2\} \) for any \( D_1 \subset D \) and \( D_2 \subset D \) satisfying

\[
\min_{s_1 \in D_1 \cap D_2} \|s_1 - s_2\| > \tau.
\]

Then, as \( n_1 \to \infty, n_2 \to \infty, \)

\[
\frac{1}{\sqrt{n_1 n_2}} \sum_{s \in D} X(s) \xrightarrow{L} N(0, \Omega),
\]

where \( \Omega = \sum_{h \in M} C_X(h) \),

\[
M \equiv \{(i, j) : \|i\alpha_c, j\alpha_r\| \leq \tau; i, j \in (-\infty, 0, 1, \ldots)\},
\]

\( \alpha_c \) is the distance between columns of the grid, and \( \alpha_r \) is the distance between rows of the grid.
Using Theorem 1, we can explicitly derive the limiting distribution for \( \hat{\theta} \) in the next theorem. This result applies to \( \hat{\theta} \) using any choice of \( \hat{\mathbf{Z}} \) given in Section 3 and for either \( \mathbf{S} = \mathbf{S}_{un} \) or \( \mathbf{S} = \mathbf{S}_{sd} \).

**Theorem 2 (Limiting distribution of \( \hat{\theta} \))** Let \( \mathbf{Z}(s) = (Z_1(s), ..., Z_p(s)) \) be a \( p \)-variate fourth-order stationary random field satisfying model (2.2) with \( s \in \mathbf{D} \), where \( \mathbf{D} \) is an \( n_1 \times n_2 \) subset of a rectangular grid. Let \( \theta_0 \) denote the true value of \( \theta \). Assume

(i) there exists a finite \( b \) such that \( \mathbb{E}\{(Z_i(s))^6\} \leq b, \ i = 1, ..., p, \)

(ii) for \( \mathbf{D}_1 \subset \mathbf{D} \) and \( \mathbf{D}_2 \subset \mathbf{D} \),

\[
\min_{s_1 \in \mathbf{D}_1, \ s_2 \in \mathbf{D}_2} ||s_1 - s_2|| > \tau \implies \{\mathbf{Z}(s), s \in \mathbf{D}_1\} \text{ is independent of } \{\mathbf{Z}(s), s \in \mathbf{D}_2\}.
\]

(iii) as \( n_1 \to \infty, n_2 \to \infty, \hat{\mathbf{Z}} \overset{p}{\to} \mathbf{\Xi}, \text{ where } \mathbf{\Xi} \text{ is positive definite},

(iv) \( \theta_0 \) is an interior point of \( \Theta \),

(v) For any \( \epsilon > 0 \), there is a \( \delta_\epsilon > 0 \) such that for any \( \theta \in \Theta \),

\[
||\theta - \theta_0|| > \epsilon \implies ||\text{vec} \Sigma(\theta) - \text{vec} \Sigma(\theta_0)|| > \delta_\epsilon,
\]

(vi) \( \mathbf{F} = \frac{\partial \text{vec} \Sigma(\theta)}{\partial \theta'} \bigg|_{\theta = \theta_0} \) has full column rank.

Then, as \( n_1 \to \infty, n_2 \to \infty, \)

\[
\sqrt{n_1 n_2} (\hat{\theta} - \theta_0) \overset{L}{\to} N \left[ 0, (\mathbf{F} \mathbf{\Xi} \mathbf{F})^{-1} \mathbf{F} \mathbf{\Xi} \mathbf{\Sigma} \mathbf{\Sigma} \mathbf{\Sigma} \mathbf{\Xi} \mathbf{F} (\mathbf{F} \mathbf{\Xi} \mathbf{F})^{-1} \right], \tag{4.1}
\]

where \( \mathbf{\Sigma} = \underset{n_1 \to \infty}{\text{plim}} \ n_1 n_2 \text{var}\{\text{vec} \mathbf{S}\} \).

**Proof:** Under the conditions, \( \hat{\mathbf{Z}} = \mathbb{E}\{\mathbf{Z}(s)\} + O_p(1/\sqrt{n_1 n_2}) \), and by Theorem 1,

\[
\sqrt{n_1 n_2} (\text{vec} \mathbf{S} - \text{vec} \Sigma(\theta_0)) \overset{L}{\to} N(0, \mathbf{\Omega}). \tag{4.2}
\]

Combining this with (iii), \( \mathbf{SS}(\theta) \) in (3.6) satisfies

\[
\mathbf{SS}(\theta) \overset{p}{\to} \mathbf{SS}_0(\theta) = [\text{vec} \Sigma(\theta_0) - \text{vec} \Sigma(\theta)]' \mathbf{\Xi} [\text{vec} \Sigma(\theta_0) - \text{vec} \Sigma(\theta)].
\]
By the identifiability condition \((v)\), \(SS_0(\theta)\) is uniquely minimized at \(\theta = \theta_0\). Hence, it follows from the standard subsequence argument that

\[
\hat{\theta} \overset{p}{\rightarrow} \theta_0. 
\] (4.3)

Note that, by definition, \(\Sigma(\theta)\) is twice continuously differentiable. By (4.3) and (\(iv)\), with probability approaching one as \(n_1 \rightarrow \infty, n_2 \rightarrow \infty,\)

\[
\frac{\partial}{\partial \theta} SS(\hat{\theta}) = -2 \left( \frac{\partial vecu \Sigma(\theta)}{\partial \theta'} \right)_{\theta = \hat{\theta}} \hat{\Sigma} (vecu S - vecu \Sigma(\hat{\theta})) = 0.
\]

Thus, it can be shown using (4.2), (4.3), (\(vi)\), and the standard Taylor expansion argument that

\[
\hat{\theta} - \theta_0 = (F' \hat{\Sigma} F)^{-1} F' \hat{\Sigma} \hat{\theta} F (F' H F)^{-1} + o_p(1/\sqrt{n_1 n_2}).
\]

Hence, the result follows from (4.2).

An estimator of the asymptotic variance of \(\hat{\theta}\) based on (4.1) is

\[
\overline{\text{var}}\{\hat{\theta}\} = (F' \hat{\Sigma} F)^{-1} \hat{F}' \hat{\Sigma} \hat{F} (F' H F)^{-1},
\] (4.4)

where \(\hat{\Gamma}\) is given in (3.8), and

\[
\hat{F} = \frac{\partial vecu \Sigma(\theta)}{\partial \theta'}_{\theta = \hat{\theta}}.
\]

Using (4.4), we can test hypotheses about \(\theta\) and construct confidence intervals.

In (4.4), \(\hat{\Gamma}\) of (3.8) is used as an estimator of \(\Gamma = \frac{1}{n_1 n_2} \Omega\), where \(N = n_1 n_2\). Since \(\hat{\theta}\) is obtained before computing \(\overline{\text{var}}\{\hat{\theta}\}\), we can suggest an alternative estimator of \(\Gamma = \frac{1}{n_1 n_2} \Omega\). Let \(\tilde{\Gamma}\) be \(\hat{\Gamma}\) in (3.8) except that \(\hat{A}\) is replaced by \(vecu \Sigma(\hat{\theta})\) in the formulas for \(\tilde{X}\) and \(\tilde{Y}\). While both \(\hat{\Gamma}\) and \(\hat{\Sigma}\) may underestimate \(\Gamma\), we might expect that the bias of \(\hat{\Gamma}\) will be less than the bias of \(\hat{\Sigma}\). Section 5.1 compares \(\hat{\Gamma}\) and \(\hat{\Sigma}\) in \(\overline{\text{var}}\{\hat{\theta}\}\).

Once model (2.2) is fitted and associated inferences are made, the researcher may be interested in predicting the value of the underlying factor process \(W(s)\) at one or more locations in the domain. The standard approach for predicting an unobservable random
variable is to use the best linear unbiased predictor (BLUP) based on all observations. However, this approach is not applicable in our problem, because we neither assume particular covariance function forms nor estimate all second moments. Instead, we propose a limited information approach based on $Z^*(s)$ of (2.4) used for model fitting. This approach is possible and practical under our assumption of general unspecified covariance structure, because all quantities necessary for prediction are obtained in our estimation procedure based on $Z^*(s)$. The limited-information BLUP (LIBLUP) of $W(s)$ in model (2.2) is

$$\hat{W}^{(LIBLUP)}(s) = \hat{E}\{W(s)\} + \text{cov}\{W(s), Z^*(s)\} \left[\text{var}\{Z^*(s)\}\right]^{-1} (Z^*(s) - \hat{E}\{Z^*(s)\}),$$

(4.5)

where $\hat{E}\{W(s)\}$ and $\hat{E}\{Z^*(s)\}$ are as developed in (2.6), $\text{var}\{Z^*(s)\} = \Sigma(\hat{\theta})$, and $\text{cov}\{W(s), Z^*(s)\}$ is found using $\hat{\lambda}_{ijg}$ and estimated second moments of $W(s)$.

5 Simulation results

5.1 Evaluation of estimation approaches

The purpose of the first simulation study is to compare various estimation options in order to recommend an approach or approaches for estimating $\theta$ and carrying out inference. We consider options for estimating $\Sigma(\theta)$, for selecting a weight matrix to be used in the least squares minimization (3.6), and for obtaining an estimate of $\Gamma$ to be used in the estimated covariance matrix (4.4).

The data were generated from model (2.3) with the domain $D$ being a $20 \times 20$ grid with a common row and column distance. The factor process $W_i(s)$ was generated as a normal random field with an isotropic spherical variogram function (partial sill $= \sigma_w^2 = 4$, range $= \tau_w$, and nugget $= 0$). Each of the error processes $e_i(s), i = 1, 2, 3,$ was generated as a normal random field with an isotropic spherical variogram function (partial sill $= \sigma_e^2 = 1$, range $= \tau_e$, and nugget $= 0$). For model fitting, the true values of the shift parameters $\Delta_{111} = (2, 2)$ and $\Delta_{121} = (2, 0)$ are assumed given.
Three different data generation scenarios were studied:

Scenario 0: \[
\begin{bmatrix}
\lambda_{110} & \lambda_{111} \\
\lambda_{210} & \lambda_{211}
\end{bmatrix} =
\begin{bmatrix}
1.5 & 0 \\
1.5 & 0
\end{bmatrix}, \quad \tau_{W_1} = 0, \quad \tau_e = 0,
\]

Scenario 1: \[
\begin{bmatrix}
\lambda_{110} & \lambda_{111} \\
\lambda_{210} & \lambda_{211}
\end{bmatrix} =
\begin{bmatrix}
0.5 & 1 \\
1 & 0.5
\end{bmatrix}, \quad \tau_{W_1} = 4, \quad \tau_e = 2,
\]

Scenario 2: \[
\begin{bmatrix}
\lambda_{110} & \lambda_{111} \\
\lambda_{210} & \lambda_{211}
\end{bmatrix} =
\begin{bmatrix}
0 & 1.5 \\
0 & 1.5
\end{bmatrix}, \quad \tau_{W_1} = 10, \quad \tau_e = 5.
\]

We expect estimation and inference properties to be good under Scenario 0 since there is no spatial dependence and no shifted factors. In contrast, the range of spatial dependence in Scenario 2 is large relative to the domain size, and \(Z_1(s)\) and \(Z_2(s)\) depend only on shifted factors. Thus, parameter estimation and inference are expected to be problematic in Scenario 2.

For each scenario, 500 samples were generated. From each sample, \(\hat{\theta}, \text{var}(\hat{\theta})\) of (4.4), and \(\hat{W}^{\text{LIBLUP}}(s)\) of (4.5) were obtained using different estimation and inference approaches. For each data generation scenario, all combinations of 3 weight matrices \((I_n, (\Sigma(\hat{\theta}))^{-1}, \hat{\Gamma}^{-1})\), and 2 estimates of \(\Sigma(\theta)\) (\(S_{un}\) and \(S_{ad}\)) were evaluated. Table 1 summarizes the four estimated factor loadings’ \((\hat{\lambda}_{110}, \hat{\lambda}_{111}, \hat{\lambda}_{210}, \hat{\lambda}_{211})\) empirical absolute bias, MSE, and coverage probability for a nominal 95% confidence interval using each of \(\hat{\Gamma}\) and \(\hat{\Gamma}\) in (4.4). Each entry in the table is the average value over the factor loadings. Additionally, Table 1 gives the median of \(\text{ASE}_{\hat{W}_1}\) where

\[
\text{ASE}_{\hat{W}_1} = \frac{1}{N} \sum_{i=1}^{N} (\hat{W}_1(s_i) - W_1(s_i))^2,
\]

and \(\hat{W}_1(s)\) is the “LIBLUP” estimator of \(W_1(s)\) defined in (4.5) and the median of (5.1) is taken over the 500 samples.

Table 1 shows that the use of \(S_{un}\) or \(S_{ad}\) for estimating \(\Sigma(\theta)\) makes very little difference for the factor loading estimation. In fact, the only way in which \(S_{ad}\) provides
Table 1 Summary results for factor loading estimates and $\text{ASE}_{\hat{\theta}_1}$. 

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Weight matrix $\Sigma(\theta)$</th>
<th>Average absolute bias</th>
<th>Average MSE $(\text{using } \hat{\theta})$</th>
<th>Average coverage prob. $(\text{using } \hat{\theta})$</th>
<th>Average coverage prob. $(\text{using } \hat{\theta})$</th>
<th>Median $\text{ASE}_{\hat{\theta}_1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$I_{n_u}$ $S_{un}$</td>
<td>0.0036</td>
<td>0.0026</td>
<td>0.9640</td>
<td>0.9745</td>
<td>0.1896</td>
</tr>
<tr>
<td>0</td>
<td>$I_{n_a}$ $S_{ad}$</td>
<td>0.0035</td>
<td>0.0026</td>
<td>0.9645</td>
<td>0.9730</td>
<td>0.1895</td>
</tr>
<tr>
<td>0</td>
<td>$(V[\Sigma(\hat{\theta})])^{-1}$ $S_{un}$</td>
<td>0.0032</td>
<td>0.0026</td>
<td>0.9420</td>
<td>0.9610</td>
<td>0.1874</td>
</tr>
<tr>
<td>0</td>
<td>$(V[\Sigma(\hat{\theta})])^{-1}$ $S_{ad}$</td>
<td>0.0030</td>
<td>0.0026</td>
<td>0.9435</td>
<td>0.9605</td>
<td>0.1873</td>
</tr>
<tr>
<td>0</td>
<td>$\hat{\theta}^{-1}$ $S_{un}$</td>
<td>0.0916</td>
<td>0.0173</td>
<td>0.3110</td>
<td>0.8035</td>
<td>0.2792</td>
</tr>
<tr>
<td>0</td>
<td>$\hat{\theta}^{-1}$ $S_{ad}$</td>
<td>0.0917</td>
<td>0.0173</td>
<td>0.3100</td>
<td>0.8030</td>
<td>0.2795</td>
</tr>
<tr>
<td>1</td>
<td>$I_{n_u}$ $S_{un}$</td>
<td>0.0070</td>
<td>0.0112</td>
<td>0.9350</td>
<td>0.9425</td>
<td>0.5328</td>
</tr>
<tr>
<td>1</td>
<td>$I_{n_a}$ $S_{ad}$</td>
<td>0.0071</td>
<td>0.0113</td>
<td>0.9275</td>
<td>0.9375</td>
<td>0.5331</td>
</tr>
<tr>
<td>1</td>
<td>$(V[\Sigma(\hat{\theta})])^{-1}$ $S_{un}$</td>
<td>0.0104</td>
<td>0.0106</td>
<td>0.9215</td>
<td>0.9365</td>
<td>0.5260</td>
</tr>
<tr>
<td>1</td>
<td>$(V[\Sigma(\hat{\theta})])^{-1}$ $S_{ad}$</td>
<td>0.0088</td>
<td>0.0105</td>
<td>0.9190</td>
<td>0.9310</td>
<td>0.5283</td>
</tr>
<tr>
<td>1</td>
<td>$\hat{\theta}^{-1}$ $S_{un}$</td>
<td>0.0283</td>
<td>0.0145</td>
<td>0.7105</td>
<td>0.9105</td>
<td>0.5744</td>
</tr>
<tr>
<td>1</td>
<td>$\hat{\theta}^{-1}$ $S_{ad}$</td>
<td>0.0308</td>
<td>0.0150</td>
<td>0.6985</td>
<td>0.9015</td>
<td>0.5751</td>
</tr>
<tr>
<td>2</td>
<td>$I_{n_u}$ $S_{un}$</td>
<td>0.0350</td>
<td>0.1484</td>
<td>0.8705</td>
<td>0.8930</td>
<td>0.9252</td>
</tr>
<tr>
<td>2</td>
<td>$I_{n_a}$ $S_{ad}$</td>
<td>0.0313</td>
<td>0.1508</td>
<td>0.8615</td>
<td>0.8815</td>
<td>0.9415</td>
</tr>
<tr>
<td>2</td>
<td>$(V[\Sigma(\hat{\theta})])^{-1}$ $S_{un}$</td>
<td>0.0360</td>
<td>0.1436</td>
<td>0.8415</td>
<td>0.8808</td>
<td>0.9083</td>
</tr>
<tr>
<td>2</td>
<td>$(V[\Sigma(\hat{\theta})])^{-1}$ $S_{ad}$</td>
<td>0.0342</td>
<td>0.1461</td>
<td>0.8294</td>
<td>0.8618</td>
<td>0.9052</td>
</tr>
<tr>
<td>2</td>
<td>$\hat{\theta}^{-1}$ $S_{un}$</td>
<td>0.0867</td>
<td>0.1858</td>
<td>0.5070</td>
<td>0.6770</td>
<td>1.4428</td>
</tr>
<tr>
<td>2</td>
<td>$\hat{\theta}^{-1}$ $S_{ad}$</td>
<td>0.0753</td>
<td>0.1786</td>
<td>0.5080</td>
<td>0.6720</td>
<td>1.4180</td>
</tr>
</tbody>
</table>
a non-trivial improvement over $S_{un}$ is in the bias of the factor and error variances and covariances. In general, using $S_{ad}$ yields variance and covariance parameter estimates that are slightly less biased than the estimates obtained using $S_{un}$. This is not surprising since the bias of $S_{un}$ is greater than the bias of $S_{ad}$. Nevertheless, the estimates of the factor loadings appear to be relatively unaffected by a biased estimate of $\Sigma(\theta)$.

Because the estimate of $\Sigma(\theta)$ has little influence on the performance of the factor loading estimation and inference and on the performance of $\text{ASE}_{\hat{W}}$, Figures 1 through 5 present results averaged over the two levels of “estimate of $\Sigma(\theta)$.” These figures illustrate the fact that using $\hat{\Gamma}$ as a weight matrix in the GLS minimization of equation (3.6) is unacceptable because it yields estimates of $\theta$ that are poor (in terms of both bias and variance).

Figures 1 and 2 plot the average absolute bias and the average MSE for the factor loadings, respectively. It is clear from these figures that Scenario 2 is the most difficult scenario for obtaining good estimates of the factor loadings. This is not surprising since the size of the domain $D$ (20x20) is relatively small compared to the range of spatial dependence for $W_1(s)$ ($\tau W_1 = 10$). Additionally, $Z_1(s)$ and $Z_2(s)$ are only related to $Z_3(s)$ through the shifted factor. In terms of bias, the $I_{na}$ does slightly better than the $(V[\Sigma(\hat{\theta})])^{-1}$ weight. The bias when using either $I_{na}$ or $(V[\Sigma(\hat{\theta})])^{-1}$ is quite small in Scenarios 0 and 1. In terms of MSE, the $I_{na}$ and $(V[\Sigma(\hat{\theta})])^{-1}$ weights are almost identical in Scenarios 0 and 1, and $(V[\Sigma(\hat{\theta})])^{-1}$ does only slightly better than $I_{na}$ in Scenario 2. The bias and MSE of the estimates using $\hat{\Gamma}^{-1}$ as a weight are poor compared to the bias and MSE of the estimates obtained using the other two weight matrices.

Figures 3 and 4 plot coverage probabilities when estimating $\text{var}\{\hat{\theta}\}$ using $\hat{\Gamma}$ and $\hat{\Gamma}$ as the matrix in the middle of the sandwich formula (4.4), respectively. These figures show that using $\hat{\Gamma}$ increases the coverage probability, particularly when using $\hat{\Gamma}^{-1}$ as the weight matrix. For Scenarios 0 and 1, the coverage probabilities of the nominal 95% confidence intervals are very close to 95% and the coverage probabilities associated with the $I_{na}$ weight matrix are always slightly larger than the coverage probabilities associated with the $(V[\Sigma(\hat{\theta})])^{-1}$ weight matrix. Coverage probabilities with the $\hat{\Gamma}^{-1}$
Figure 1 Absolute bias for the factor loading estimates for each combination of three scenarios (0, 1, 2) and the three weight matrices ($I_{nu} = \Gamma$ with a solid line, $(V[\Sigma(\hat{\theta})])^{-1} = V$ with a dotted line, $\hat{\Gamma}^{-1} = G$ with a dashed line).
Figure 2 MSE for the factor loading estimates for each combination of the three scenarios (0, 1, 2) and the three weight matrices ($I_n = "I"$ with a solid line, $(V[\Sigma(\hat{\theta})])^{-1} = "V"$ with a dotted line, $\tilde{V}^{-1} = "G"$ with a dashed line).
Figure 3 Coverage probability for the factor loading estimates when estimating $\text{var}(\hat{\theta})$ using $\hat{\Gamma}$ for each combination of the three scenarios (0, 1, 2) and the three weight matrices ($I_n$ = "I" with a solid line, $(V[\Sigma(\hat{\theta})])^{-1}$ = "V" with a dotted line, $\hat{\Gamma}^{-1}$ = "G" with a dashed line). The nominal 95% coverage is denoted with a line.
Figure 4 Coverage probability for the factor loading estimates when estimating \( \text{var}\{\hat{\theta}\} \) using \( \hat{\Gamma} \) for each combination of the three scenarios (0, 1, 2) and the three weight matrices (\( \Gamma_n = \text{"I"} \) with a solid line, \( (V[\Sigma(\hat{\theta})])^{-1} = \text{"V"} \) with a dotted line, \( \hat{\Gamma}^{-1} = \text{"G"} \) with a dashed line). The nominal 95% coverage is denoted with a line.
weight are poor, but using $\hat{\Gamma}$ in the calculation of the confidence interval improves the coverage dramatically. Figure 5 illustrates that, with respect to $\text{ASE}_{\hat{\mu}_1}$, $(V[\Sigma(\hat{\theta})])^{-1}$ and $I_{na}$ give similar performance, and are superior to $\hat{\Gamma}^{-1}$.

We conclude from this simulation that when there is interest in accurate estimation of factor and error variances and covariances, $S_{ad}$ in (3.5) should be used as an estimate of $\Sigma(\theta)$. When the primary interest is in estimation and inference for the factor loadings, one can use unadjusted sample moments to estimate the unique moments in $\Sigma(\theta)$ with no noticeable loss of estimation accuracy or inference validity. Use of $\hat{\Gamma}^{-1}$ as the weight matrix in (3.6) should be avoided. Rather, use of $I_{na}$ is recommended because of its simplicity and good performance. For the associated inferences, the appropriate covariance matrix (4.4) using $\hat{\Gamma}$ gives very accurate results.

5.2 Evaluation of estimation properties under several data models

The purpose of the simulation study in Section 5.1 was to compare several different approaches for estimation and inference for $\theta$. In this section, we assess the performance of a particular approach under several data models. We use the procedure with $S_{un}$ as an estimate of $\Sigma(\theta)$, $I_{na}$ as the weight, and $\hat{\Gamma}$ in the asymptotic covariance matrix. The data-model experimental factors considered are the factor loading structure, the factor process distribution, the spatial dependence range, the signal-to-noise ratio, and the sample field size.

Two factor loading structures

\[
\begin{bmatrix}
\lambda_{110} & \lambda_{111} \\
\lambda_{210} & \lambda_{211}
\end{bmatrix} = \begin{bmatrix}
0 & 1.5 \\
1.5 & 0
\end{bmatrix}
\]

\[
\begin{bmatrix}
\lambda_{110} & \lambda_{111} \\
\lambda_{210} & \lambda_{211}
\end{bmatrix} = \begin{bmatrix}
0.5 & 1 \\
1 & 0.5
\end{bmatrix}
\]

were considered. For the factor process, normal and linearly transformed lognormal distributions were considered. Each of these two processes were generated to have an isotropic spherical variogram with range = $\nu_W$, partial sill = $\sigma^2_W$, and nugget = 0. The
Figure 5 Median $\text{ASE}_{W1}$ for each combination of the three scenarios (0, 1, 2) and the three weight matrices ($I_{na} = \Gamma$ with a solid line, $(V[\Sigma(\hat{\theta})])^{-1} = V$ with a dotted line, $\hat{\Gamma}^{-1} = G$ with a dashed line).
error processes always have a common normal distribution with a spherical variogram with range = \( \tau_z \), partial sill = \( \sigma_z^2 \), and nugget = 0. For the range of spatial dependence of the factor and error processes, we considered a "small" range with \( \tau_{W_i} = 4 \) and \( \tau_z = 2 \), and a "large" range with \( \tau_{W_i} = 10 \) and \( \tau_z = 5 \). For the estimation of \( \Gamma \), it was assumed that \( \tau_A = 4 \) for the small range and \( \tau_A = 6 \) for the large range. For the signal-to-noise ratio, a "low reliability" case with \( \sigma_{W_i}^2 = 7 \) and \( \sigma_z^2 = 3 \), and a "high reliability" case with \( \sigma_{W_i}^2 = 9 \) and \( \sigma_z^2 = 1 \) were considered. Two sample fields, 20 \( \times \) 20 and 30 \( \times \) 30 grids, were used, where both grids are squares, i.e., have a common row and column grid distance.

This simulation study, assessing five experimental factors with two levels each, was run using a half-fractional factorial treatment structure, with two-way interactions confounded with three-way interactions. Consequently, we can discuss only main effects and two-way interactions. For each factor combination in the experiment, 1000 samples were generated. Following the same basic structure of Table 1 from Section 5.1, for each experimental factor combination, Table 2 gives values of the following quantities averaged over the 4 factor loadings in the model: absolute bias, MSE, and coverage probability for a nominal 95% confidence interval. Also given is the median of ASE\(_{W_i}\).

Figures 6 and 7 depict the main effects and two-way interactions of the 5 experimental factors for the absolute bias averaged over the 4 factor loadings. We first note that bias is usually quite small relative to the average size of the factor loadings. The main effect with the greatest impact on bias is the spatial dependence range, followed by the sample size. When the range of spatial dependence is large, there is a tendency for the factor loadings associated with \( W_i(s) \) to be overestimated and for the factor loadings associated with the shifted factor to be underestimated. This tendency is particularly pronounced under factor loading structure A, where both \( Z_1(s) \) and \( Z_2(s) \) depend on the factor at only one of the two potential locations (see Figure 8).

Note also from Figure 7 that the improvement of the bias due to an increased sample size is much greater when the processes have a large range of spatial dependence than when the processes have a small range of spatial dependence. The bias of the factor loadings is equally low for the two factor loading structures when the range of spatial
Table 2 Summary results for factor loading estimates and $\hat{W}^{(LIBLUP)}(s)$.

<table>
<thead>
<tr>
<th>Factor loading structure</th>
<th>Factor distribution</th>
<th>Range</th>
<th>Reliability</th>
<th>Sample size</th>
<th>Avg. abs. bias</th>
<th>Avg. MSE</th>
<th>Avg. cov. prob.</th>
<th>Median $\text{ASE}_{\hat{W}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>normal</td>
<td>Small</td>
<td>High</td>
<td>$30 \times 30$</td>
<td>0.0074</td>
<td>0.0051</td>
<td>0.9650</td>
<td>0.3892</td>
</tr>
<tr>
<td>A</td>
<td>normal</td>
<td>Small</td>
<td>Low</td>
<td>$20 \times 20$</td>
<td>0.0047</td>
<td>0.0413</td>
<td>0.9668</td>
<td>1.0169</td>
</tr>
<tr>
<td>A</td>
<td>normal</td>
<td>Large</td>
<td>High</td>
<td>$20 \times 20$</td>
<td>0.0364</td>
<td>0.0519</td>
<td>0.9035</td>
<td>0.5436</td>
</tr>
<tr>
<td>A</td>
<td>normal</td>
<td>Large</td>
<td>Low</td>
<td>$30 \times 30$</td>
<td>0.0223</td>
<td>0.0758</td>
<td>0.9280</td>
<td>0.9628</td>
</tr>
<tr>
<td>A</td>
<td>lognor</td>
<td>Small</td>
<td>High</td>
<td>$20 \times 20$</td>
<td>0.0138</td>
<td>0.0211</td>
<td>0.9597</td>
<td>0.4966</td>
</tr>
<tr>
<td>A</td>
<td>lognor</td>
<td>Small</td>
<td>Low</td>
<td>$30 \times 30$</td>
<td>0.0028</td>
<td>0.0220</td>
<td>0.9625</td>
<td>0.9046</td>
</tr>
<tr>
<td>A</td>
<td>lognor</td>
<td>Large</td>
<td>High</td>
<td>$30 \times 30$</td>
<td>0.0151</td>
<td>0.0294</td>
<td>0.9352</td>
<td>0.4281</td>
</tr>
<tr>
<td>A</td>
<td>lognor</td>
<td>Large</td>
<td>Low</td>
<td>$20 \times 20$</td>
<td>0.0461</td>
<td>0.2592</td>
<td>0.8815</td>
<td>1.1469</td>
</tr>
<tr>
<td>B</td>
<td>normal</td>
<td>Small</td>
<td>High</td>
<td>$20 \times 20$</td>
<td>0.0030</td>
<td>0.0062</td>
<td>0.9635</td>
<td>0.7038</td>
</tr>
<tr>
<td>B</td>
<td>normal</td>
<td>Small</td>
<td>Low</td>
<td>$30 \times 30$</td>
<td>0.0020</td>
<td>0.0080</td>
<td>0.9550</td>
<td>1.2588</td>
</tr>
<tr>
<td>B</td>
<td>normal</td>
<td>Large</td>
<td>High</td>
<td>$30 \times 30$</td>
<td>0.0129</td>
<td>0.0168</td>
<td>0.9322</td>
<td>0.5424</td>
</tr>
<tr>
<td>B</td>
<td>normal</td>
<td>Large</td>
<td>Low</td>
<td>$20 \times 20$</td>
<td>0.0193</td>
<td>0.1771</td>
<td>0.8570</td>
<td>1.4475</td>
</tr>
<tr>
<td>B</td>
<td>lognor</td>
<td>Small</td>
<td>High</td>
<td>$30 \times 30$</td>
<td>0.0018</td>
<td>0.0040</td>
<td>0.9768</td>
<td>0.5593</td>
</tr>
<tr>
<td>B</td>
<td>lognor</td>
<td>Small</td>
<td>Low</td>
<td>$20 \times 20$</td>
<td>0.0166</td>
<td>0.0290</td>
<td>0.9582</td>
<td>1.3729</td>
</tr>
<tr>
<td>B</td>
<td>lognor</td>
<td>Large</td>
<td>High</td>
<td>$20 \times 20$</td>
<td>0.0222</td>
<td>0.0554</td>
<td>0.8645</td>
<td>0.7452</td>
</tr>
<tr>
<td>B</td>
<td>lognor</td>
<td>Large</td>
<td>Low</td>
<td>$30 \times 30$</td>
<td>0.0103</td>
<td>0.0728</td>
<td>0.9065</td>
<td>1.1352</td>
</tr>
</tbody>
</table>
Figure 6 Average absolute bias for the 4 factor loadings for each level of each of the 5 experimental factors.

dependence is small, but the bias of the factor loadings is much greater for factor loading structure A when the range of spatial dependence is large.

Figures 9 and 10 illustrate the main effects and two-way interactions for the MSE averaged over the 4 factor loadings. The main effect with the greatest impact on MSE is the range, followed by reliability and sample size. The significant interactions illustrated in Figure 10 involve range, reliability, and sample size. Any combination of large range, low reliability, and small sample size results in particularly large MSE, i.e., estimation difficulty. This can also be seen in Table 2, where the two largest MSE values occur for combinations of large spatial dependency range, low reliability, and small field. Also, for factor loading structure A, the MSE is somewhat smaller for the normal factors than
Figure 7 Two-way interaction plots for average absolute bias for the 4 factor loadings.
Figure 8 Comparison of bias for each of the four factor loading estimates at each of the two levels of range. Separate plots are given for the data generated with factor loading structure A (top) and factor loading structure B (bottom). In each plot, “S” and a dotted line is used to indicate the average bias when using the small range values, and “L” and a dashed line is used to indicate the average bias when using the large range values.
Figure 9 Average MSE for the 4 factor loadings for each level of each of the 5 experimental factors.
Figure 10 Two-way interaction plots for average MSE for the 4 factor loadings.
for the lognormal factors.

Figures 11 and 12 are for the coverage probability averaged over the 4 factor loadings. The range and sample size have the greatest impact on coverage. As shown in Figure 12, a small sample size has a negative effect only with large range of spatial dependency. Four of the five lowest coverage probabilities in Table 2 were associated with large range and small sample size. But, the coverage probability is rather reasonable even for these 5 cases, and is very close to the nominal level for all other cases.

The median of $\text{ASE}_{W_1}$ defined in (5.1) is covered by Figures 13 and 14. The prediction accuracy is influenced most by reliability, and to a lesser extent, by factor loading structure and sample size. Note from Table 2 that the median of $\text{ASE}_{W_1}$ is always much less than 1 for the high reliability cases and much less than 3 for the low reliability cases. This indicates that the limited-information BLUP (4.5) is a much better predictor of the factor $W_1(s)$ than is $Z_3(s)$.

6 Analysis of precision farming data

Since the time that the Global Positioning System (GPS) became available for public use nearly ten years ago, there has been an increased interest in precision farming. The GPS allows the agronomist or soil scientist to ascertain a particular location within a field to a high level of accuracy. This precise measure of location in a field, combined with soil characteristics, yield measurements, and other data associated with the location can then be used to extract valuable information about the complex process of plant growth. This information collected at multiple locations within a field can be used to formulate yield-maximization strategies such as variable-rate fertilizer, herbicide or pesticide application. Though such data are now readily available, there remains a need for spatial modeling tools that can assist in explaining the relationships among soil characteristics, environmental conditions, and yield. We present here an application of the generalized shifted-factor model and the augmented observation model-fitting approach for assessing soil quality from multivariate soil characteristic data.

The data to be analyzed here were drawn from part of a privately owned farm
Figure 11 Average coverage probability for the 4 factor loadings for each level of each of the 5 experimental factors.
Figure 12 Two-way interaction plots for average coverage probability for the 4 factor loadings.
Figure 13 Median $A_{SE_{W_i}}$ for each level of each of the 5 factors.
Figure 14 Two-way interaction plots for median ASE_{W_1}. 
in southeastern Boone County, Iowa (Colvin, et al., 1997). In May of 1997, several soil characteristics (including log(Ca), log(Cu), log(Mg), log(Mn), and log(Zn)) were measured in parts per million at each of 215 sites. All sites come from an approximately 350 m × 350 m portion of the field. The sites are located along 8 equally-spaced east-west transects. Along each transect there are 28 sites, spaced approximately 12.2 m apart. The soil samples consisted of 6 cores, 1 inch in diameter and collected to a depth of 8 inches. Samples were air dried, ground, extracted, and analyzed for the various nutrients using inductively-coupled plasma (ICP) techniques (Karlen and Colvin, 1998). Figure 15 is a map of the sampling locations used in our analyses.

The researcher may hypothesize that the five soil characteristics of interest (log(Ca), log(Cu), log(Mg), log(Mn), and log(Zn)) can be modeled as being dependent on a single spatially correlated "soil quality" process. That is, each observed characteristic is a function of the latent factor, plus a general error process which is independent of the factor process and other error processes. Although no theoretical reason exists for lagged relationships between the observed variables and the latent variable, our methodology can be used to evaluate potential shifts. Consider a model

\[
\begin{align*}
\log[Ca](s) &= \gamma_1 + \lambda_{110}W_1(s) + \lambda_{11n}W_1(s + \text{"n"}) + \lambda_{11e}W_1(s + \text{"e"}) + e_1(s) \\
\log[Cu](s) &= \gamma_2 + \lambda_{210}W_1(s) + \lambda_{21n}W_1(s + \text{"n"}) + \lambda_{21e}W_1(s + \text{"e"}) + e_2(s) \\
\log[Mg](s) &= \gamma_3 + \lambda_{310}W_1(s) + \lambda_{31n}W_1(s + \text{"n"}) + \lambda_{31e}W_1(s + \text{"e"}) + e_3(s) \\
\log[Mn](s) &= \gamma_4 + \lambda_{410}W_1(s) + \lambda_{41n}W_1(s + \text{"n"}) + \lambda_{41e}W_1(s + \text{"e"}) + e_4(s) \\
\log[Zn](s) &= W_1(s) + e_5(s),
\end{align*}
\] (6.1)

where \((s + \text{"n"})\) refers to the site 48 m to the north (the nearest north neighbor) and \((s + \text{"e"})\) refers to the site 12.2 m to the east (the nearest east neighbor).

For these data, the range of spatial dependence was estimated to be approximately 150 m, and \(\tau_A\) was set to 150 m in the calculation of \(\hat{\Gamma}\). The augmented observation vector \(Z^*(s)\) used to fit the model included the elements of \(Z^*(s)\), the four nearest neighbors for each process, and all other lagged versions of \(\log[Zn](s)\) that were prescribed by the simple expansion approach given in Section 2. For these data, virtually the same
parameter estimates were obtained using the OLS weight ($I_{na}$) and the normality-based weight ($((V[\Sigma(\hat{\theta})])^{-1})$. The factor loading estimates, standard errors, and $p$-values for model (6.1) using the OLS weight (intercepts $\gamma_1$'s are suppressed) are

\[
\begin{align*}
\log[Ca](s) &= 1.80 \ W_1(s) + (-0.25) \ W_1(s + "n") + 0.41 \ W_1(s + "e") + \epsilon_1(s), \\
\text{std. err.} &= 0.81 \quad 0.96 \quad 0.81 \\
p-value &= 0.027 \quad 0.792 \quad 0.615
\end{align*}
\]

\[
\begin{align*}
\log[Cu](s) &= 0.98 \ W_1(s) + (-0.19) \ W_1(s + "n") + 0.21 \ W_1(s + "e") + \epsilon_2(s), \\
\text{std. err.} &= 0.34 \quad 0.46 \quad 0.32 \\
p-value &= 0.004 \quad 0.678 \quad 0.507
\end{align*}
\]

\[
\begin{align*}
\log[Mg](s) &= 0.82 \ W_1(s) + (-0.08) \ W_1(s + "n") + 0.18 \ W_1(s + "e") + \epsilon_3(s), \\
\text{std. err.} &= 0.41 \quad 0.37 \quad 0.36 \\
p-value &= 0.046 \quad 0.818 \quad 0.624
\end{align*}
\]

\[
\begin{align*}
\log[Mn](s) &= 0.69 \ W_1(s) + (-0.15) \ W_1(s + "n") + 0.07 \ W_1(s + "e") + \epsilon_4(s), \\
\text{std. err.} &= 0.48 \quad 0.37 \quad 0.47 \\
p-value &= 0.153 \quad 0.679 \quad 0.882
\end{align*}
\]

\[
\begin{align*}
\log[Zn](s) &= \ W_1(s) + \epsilon_5(s).
\end{align*}
\]

Note that all of the shifted-factor terms in the model are nonsignificant. Thus, the model with no shifted factors was fitted to obtain the final model.
The values for the factor process can be predicted using the limited-information BLUP approach (4.5). Figure 16 gives image plots of each of the five observed variables as well as the estimated factor process. The predicted factor is a noise-free smooth process which incorporates features of each of the five observed variables. Because our final model for these data involves only one factor process and no shifted factors, each observed variable estimates a linear function of the factor at that location. As seen in Figure 16, the LIBLUP of the factor delineates the areas of high and low soil richness much more clearly than any observed variable.

The generalized shifted-factor model, the augmented observation (AO) model-fitting approach, and the associated methodology provide the applied researcher with powerful tools for exploring and modeling spatially correlated multivariate data. The procedures proposed in this paper are useful for a very broad range of problems and data, and provide accurate and informative analysis.
Figure 16 Image plots (black = high, white = low) for the five observed variables and the estimated factor process.

Appendix

This appendix presents a proof of Theorem 1. First, a central limit theorem for univariate spatial processes is given as a lemma.

Lemma (Central limit theorem for univariate random fields) Let $X(s)$ be a zero-mean second-order stationary random field with $s \in D$, where $D$ is an $n_1 \times n_2$ subset of a rectangular grid. Assume

(i) there exists a finite $b$ such that $E\{|X(s)|^3\} \leq b$,

(ii) for $D_1 \subset D$ and $D_2 \subset D$,
\[
\min_{s_1 \in D_1, s_2 \in D_2} \|s_1 - s_2\| > \tau \implies \{X(s), s \in D_1\} \text{ is independent of } \{X(s), s \in D_2\}.
\]

Then, as \(n_1 \to \infty, n_2 \to \infty,\)

\[
\frac{1}{\sqrt{n_1 n_2}} \sum_{s \in D} X(s) \xrightarrow{\mathcal{L}} N(0, \omega^2),
\]

where \(\omega^2 = \sum_{h \in \mathcal{M}} C_X(h),\)

\[
\mathcal{M} \equiv \{(i, j) : \| (\alpha_c, j \alpha_r) \| \leq \tau; i, j \in (\ldots, -1, 0, 1, \ldots)\}, \tag{A.1}
\]

\(\alpha_c\) is the distance between columns of the grid, and \(\alpha_r\) is the distance between rows of the grid.

**Proof:** Without loss of generality, assume that \(n_1 = n_2 = n\) and that the \(n^2\) observational locations in \(D\) are on a square grid. Let \(l = \lfloor n^{1/8} \rfloor, q = \lfloor \frac{n}{l^2} \rfloor, \) and \(m = \lfloor \tau + 1 \rfloor,\) where \(\lfloor y \rfloor\) denotes the integer part of \(y.\) Partition the locations in the first \(lq\) rows and the first \(lq\) columns into a \(q \times q\) grid of \(l \times l\) blocks, and denote these blocks \(F_{cd}, c = 1, \ldots, q, d = 1, \ldots, q.\) Let the region of \(D\) containing the locations in the final \(n - ql\) rows and the final \(n - ql\) columns be denoted \(E.\) Define \(A_{cd}\) to be the first \(l - m\) rows and the first \(l - m\) columns of \(F_{cd}, c = 1, \ldots, q, d = 1, \ldots, q.\) Let the region of \(F_{cd}\) containing the locations in the final \(m\) rows and the final \(m\) columns be denoted \(B_{cd}, c = 1, \ldots, q, d = 1, \ldots, q.\) Then,

\[
\frac{1}{\sqrt{n_1 n_2}} \sum_{s \in D} X(s) = \frac{1}{n} \sum_{c=1}^{q} \sum_{d=1}^{q} \left( \sum_{s \in A_{cd}} X(s) + \sum_{s \in B_{cd}} X(s) \right) + \frac{1}{n} \sum_{s \in E} X(s).
\]

An illustration of this partitioning of \(D\) is given in Figure 17.

We now show that the effect of ignoring the locations in \(B_{cd}, c = 1, \ldots, q, d = 1, \ldots, q,\) and \(E\) is negligible as \(n \to \infty.\) Let \(\sigma^2\) be the variance of \(X(s)\) and denote \(|G|\) to be the number of locations in the area \(G.\) Note that for any \(s \in B_{cd},\) the number of locations
Figure 17 Partitioning of the $n_1 \times n_2$ grid $D$ in the proof of Lemma.
within $m$ of $s$ is always less than $\frac{3}{4} \pi m^2$, so

$$\text{var} \left\{ \sum_{s \in B_{cd}} X(s) \right\} = |B_{cd}| \sigma^2 + \sum_{s_i \in B_{cd}} \sum_{s_j \in B_{cd}} \text{cov}(X(s_i), X(s_j)) \quad \text{for} \quad 0 < \|s_i - s_j\| < m$$

$$< |B_{cd}| \sigma^2 + 2 \left( \frac{3}{4} \pi m^2 \right) |B_{cd}| \sigma^2$$

$$= \left( \frac{3}{2} \pi m^2 + 1 \right) |B_{cd}| \sigma^2$$

$$= \left( \frac{3}{2} \pi m^2 + 1 \right) (2lm - m^2) \sigma^2.$$  

Note that for any $s \in B_{cd}$, there are less than $5m^2$ locations close enough to covary with locations in other $B_{cd}$'s. Each of these $5m^2$ locations can covary with up to $\frac{\pi}{2} m^2$ locations in other $B_{cd}$'s, and each of these covariances is bounded by $\sigma^2$, so

$$\text{var} \left\{ \sum_{c=1}^{q} \sum_{d=1}^{q} \sum_{s \in B_{cd}} X(s) \right\} < \sum_{c=1}^{q} \sum_{d=1}^{q} \text{var} \left\{ \sum_{s \in B_{cd}} X(s) \right\} + q^2 \left( \frac{5}{2} \pi m^4 \sigma^2 \right)$$

$$= q^2 \left( \frac{3}{2} \pi m^2 + 1 \right) (2lm - m^2) \sigma^2 + q^2 (\frac{5}{2} \pi m^4 \sigma^2)$$

$$= O_p(\eta^{7/4}) O_p(\eta^{1/8}) + O_p(\eta^{7/4})$$

$$= O_p(\eta^{15/8}),$$

and since $E \left\{ \sum_{c=1}^{q} \sum_{d=1}^{q} \sum_{s \in B_{cd}} X(s) \right\} = 0$,

$$\text{var} \left\{ \sum_{c=1}^{q} \sum_{d=1}^{q} \sum_{s \in B_{cd}} X(s) \right\} = O_p(\eta^{15/8})$$

$$\implies \sum_{c=1}^{q} \sum_{d=1}^{q} \sum_{s \in B_{cd}} X(s) = O_p(\eta^{15/16})$$

$$\implies \frac{1}{n} \sum_{c=1}^{q} \sum_{d=1}^{q} \sum_{s \in B_{cd}} X(s) = O_p(n^{-1/16}) = o_p(1). \quad (A.2)$$

The number of locations $s \in E$ is less than $2nl - l^2$. Note that for any $s \in E$, the number of locations within $m$ units of $s$ is always less than $\pi m^2$ and the covariance of
X(s_t) and X(s_{t'}) is bounded by \sigma^2. This implies that
\begin{align*}
\text{var} \left\{ \sum_{s \in B} X(s) \right\} &= |B| \sigma^2 + \sum_{s_i \in B} \sum_{s_j \in B} \text{cov} \{ X(s_i), X(s_j) \} \\
&< |B| \sigma^2 + 2\pi m^2 |E| \sigma^2 \\
&< (2\pi m^2 + 1)(2nl - l^2) \sigma^2 \\
&= O_p(n^{9/8}),
\end{align*}
and since \( E \{ \sum_{s \in B} X(s) \} = 0 \),
\begin{align*}
\text{var} \left\{ \sum_{s \in B} X(s) \right\} &= O_p(n^{9/8}) \\
\implies \sum_{s \in B} X(s) &= O_p(n^{9/16}) \\
\implies \frac{1}{n} \sum_{s \in B} X(s) &= O_p(n^{-7/16}) = o_p(1). \quad (A.3)
\end{align*}

Now define \( Y_{cd} = E_{s \in A_{cd}} X(s) \).
\begin{align*}
\text{var} \{ Y_{cd} \} &= \frac{1}{(l - m)^2} \sum_{s_i \in A_{cd}} \sum_{s_j \in A_{cd}} C_X(s_i - s_j) \\
&= \frac{1}{(l - m)^2} \sum_{h \in M} [(l - m)^2 - r_h] C_X(h) \\
&= \sum_{h \in M} \frac{(l - m)^2 - r_h}{(l - m)^2} C_X(h) \\
&= \omega_n^2
\end{align*}
where \( M \) is defined in equation (A.1) and
\begin{align*}
r_h &= r_{(h_1, h_2)} \\
&= (|h_1| + |h_2|)(l - m) - |h_1||h_2| \\
&= O_p(n^{1/8}).
\end{align*}
As \( n \to \infty \),
\[ \omega_n^2 \overset{p}{\to} \omega^2 = \sum_{h \in M} C_X(h). \]

In order to apply the Liapounov central limit theorem to \( Y_{cd}, c = 1, \ldots, q, d = 1, \ldots, q \),
we verify the following properties of this sequence of random variables:
(1) \( Y_{cd}, c = 1, \ldots, q, d = 1, \ldots, q \) is a triangular array since these are mutually independent random variables and since \( q \to \infty \) as \( n \to \infty \),

(2) \( E\{Y_{cd}\} = 0 \),

(3) \( E\{Y_{cd}^2\} = \omega_n^2 < \infty \),

(4) \( v_n^2 = \sum_{c=1}^q \sum_{d=1}^q \text{var}\{Y_{cd}\} = q^2 \omega_n^2 \).

Then, since

\[
\frac{v_n^{-3} \sum_{c=1}^q \sum_{d=1}^q E\{|Y_{cd}|^3\}}{q^2 \omega_n^3 \sum_{c=1}^q \sum_{d=1}^q \frac{1}{(l-m)^3} E\left\{ \left| \sum_{s \in A_{cd}} X(s) \right|^3 \right\}} \\
\leq \frac{1}{q^2 \omega_n^3} \sum_{c=1}^q \sum_{d=1}^q \frac{1}{(l-m)^3} (l-m)^3 b \\
= \frac{1}{q \omega_n^2} b \to 0,
\]

we can conclude by the Liapounov central limit theorem that

\[
Q_n = \frac{1}{q} \sum_{c=1}^q \sum_{d=1}^q Y_{cd} \overset{L}{\to} N(0, \omega^2).
\]  \hspace{1cm} (A.4)

Using (A.2), (A.3), (A.4), and Slutsky's Theorem, we conclude that

\[
\frac{1}{n} \sum_{s \in E} X(s) = \frac{1}{n} \sum_{c=1}^q \sum_{d=1}^q \sum_{s \in A_{cd}} X(s) + \frac{1}{n} \sum_{c=1}^q \sum_{d=1}^q \sum_{s \in B_{cd}} X(s) + \frac{1}{n} \sum_{s \in E} X(s) \\
= \frac{1}{n} \sum_{c=1}^q \sum_{d=1}^q (l-m) Y_{cd} + o_p(1) + o_p(1) \\
\overset{L}{\longrightarrow} \left( \frac{q(l-m)/n}{N(0, \omega^2)} + o_p(1) \right) \overset{\to}{\longrightarrow} 0 \\
\overset{L}{\longrightarrow} N(0, \omega^2)
\]

We now give the proof of Theorem 1 (multivariate central limit theorem for random fields) which makes use of the result of Lemma.
Proof of Theorem 1: As in the proof of Lemma, we assume (without loss of generality) that \( n_1 = n_2 = n \) and that the \( n^2 \) observational locations in \( D \) are on a square grid. For a non-zero \( t \in \mathbb{R}^p \), consider

\[
t' \left( \frac{1}{n} \sum_{s \in D} X(s) \right) = \frac{1}{n} \sum_{s \in D} T(s)
\]

where \( T(s) = t'X(s) \). Note that \( T(s) \) is a zero-mean second-order stationary process satisfying \((ii)\) in Lemma. The condition \((i)\) is also satisfied, because \( E\{T(s)\} \) is bounded by \( c_\varepsilon b \) for some constant \( c_\varepsilon \) depending on \( t \). Thus, by Lemma,

\[
\frac{1}{n} \sum_{s \in D} T(s) \xrightarrow{\mathcal{L}} N(0, \omega_T^2),
\]

where the limiting variance is defined in Lemma to be

\[
\omega_T^2 = \sum_{h \in \mathcal{M}} C_T(h) = t' \left( \sum_{h \in \mathcal{M}} C_X(h) \right) t = t'\Omega t.
\]

Hence, the result follows from the Cramér-Wold device. \( \Box \)

References


MODELING AND PREDICTION FOR MULTIVARIATE SPATIAL DATA USING SPATIALLY-CORRELATED LATENT VARIABLES

A paper to be submitted to *Biometrika*

William F. Christensen and Yasuo Amemiya

**Abstract**

Factor analysis of multivariate spatial data is considered. A systematic approach for modeling the underlying structure of geo-referenced vector observations is proposed. Statistical inference procedures for model checking and building are discussed. We derive a condition under which a simple and practical inference procedure is valid without specifying the form of distributions and factor covariance functions. The multivariate prediction problem is also discussed, and a procedure combining the latent variable modeling and a measurement-error-free kriging technique is introduced. Simulation results and an example using agricultural data are presented.

**1 Introduction**

When multiple spatially-correlated measurements are collected, the researcher is often interested in some low-dimensional version of the multivariate process. For example, an ecologist may be interested in a small number of underlying factors that drive a high-dimensional process such as abundance of many species (e.g., Ver Hoef, Reiter, and Glenn-Lewin, 1993), or an engineer may be interested in extracting a signal or true image from multi-channel satellite data (e.g., Switzer and Green, 1984). Factor analysis-like...
approaches for multivariate spatial data have been discussed by several authors including Wartenberg (1985a, 1985b), Wackernagel (1988), Grunsky and Agterberg (1991, 1992), Switzer and Green (1984), and Cook et al. (1994). These papers had limited coverage of statistical estimation, inference, prediction, and model fit assessment. Christensen and Amemiya (1999) proposed a general latent variable model for spatial data with possible asymmetric structure, and suggested parameter estimation and inference procedures. In this paper, statistical issues involved in building and using a basic factor analytic model for multivariate spatial data are discussed comprehensively.

Suppose that a \( p \times 1 \) observation \( Z(s) = (Z_1(s), ..., Z_p(s))' \) is taken at \( s \) for \( s \in D \) where \( D \) is a sample region in \( Z^2 \). Assume that the \( p \) observed variables \( Z_i, i = 1, ..., p \), can be expressed as linear functions of a \( k \) \((< p)\) dimensional unobservable factor \( W(s) = (W_1(s), ..., W_k(s))' \) except for \( p \) variable-specific errors \( e_i(s), i = 1, ..., p \). The basic exploratory factor analysis model for such a situation can be expressed in an identifiable form using the so-called errors-in-variables parameterization as

\[
Z(s) = \begin{pmatrix} \gamma \\ 0_k \end{pmatrix} + \begin{pmatrix} B \\ I_k \end{pmatrix} W(s) + e(s),
\]

where \( e(s) = (e_1(s), ..., e_p(s))' \). and

\[
B = \begin{bmatrix} 
\lambda_{11} & \cdots & \lambda_{1k} \\
\lambda_{21} & \cdots & \lambda_{2k} \\
\vdots & \ddots & \vdots \\
\lambda_{p-k,1} & \cdots & \lambda_{p-k,k} 
\end{bmatrix}.
\]

See, e.g., Fuller (1987). Since the inter-relationships among the \( p \) observed variables \( Z_i(s) \) are to be explained by the \( k \) factors \( W_j(s) \), we assume that the \( p + 1 \) processes \( W(s), e_1(s), ..., e_p(s) \) are independent. Our interest is in modeling \( Z(s) \) after removing systematic trend, and thus we assume that \( W(s) \) and \( e_i(s) \) are second-order stationary processes with \( E\{e_i(s)\} = 0 \). But, to develop an exploratory method useful for a broad range of situations, the distributional forms and covariance functions for \( W(s) \) and \( e_i(s) \) are unspecified.
Suppose that the standard random sample normal maximum likelihood method is applied to $Z(s)$ in (1.1) ignoring the spatial dependency and the assumption of unspecified distribution. The validity of the resulting inference procedures for relevant parameters and model fit has not been discussed in the literature. Section 2 shows this validity under a practically meaningful condition, and presents an alternative estimation procedure for cases in which the condition is not met. Another topic that has not been treated in the literature is the development of a proper method for statistically assessing and building a factor analytic model in the spatial setting. For this problem, Section 3 discusses a systematic approach that combines a test for the condition derived in Section 2 and the valid goodness-of-fit test under the condition. Section 4 shows that prediction of the underlying factor $W(s)$ at any $s$ is possible using a measurement-error-free kriging approach under the condition of Section 2. Various estimation, inference, and prediction issues are addressed in simulation studies in Section 5. A simple example from precision agriculture is presented in Section 6.

2 Estimation and inference procedures

Suppose that model (1.1) holds for observations $Z(s)$ taken on an $n_1 \times n_2$ grid $D$. If $Z(s)$ were incorrectly assumed to be independent normal vectors, the normal maximum likelihood method would be the standard approach. Such a method can be easily implemented using the existing software such as SAS PROC CALIS (SAS Institute Inc., 1989). Since the normality-based method uses only the first two moments, the parameter estimators may have reasonable properties, even for nonnormal spatial data. But, the standard errors and chi-square goodness-of-fit test using the independent normal method cannot be expected to be valid. We derive a condition under which such standard errors and goodness-of-fit tests are valid and practically useful.
To express the first two moments of $Z(s)$, let

$$
\begin{align*}
\mu &= \mathbb{E}\{W(s)\}, \\
\Phi &= \text{var}\{W(s)\}, \\
\psi &= \text{var}\{e_i(s)\}, \\
\psi &= (\psi_1, ..., \psi_p)'.
\end{align*}
$$

We assume a general form for $B$ in model (1.1), allowing for possible zero or equality restrictions for factor loadings $\lambda_{ij}$, and write $B = B(\lambda)$ where $\lambda$ contains all unknown $\lambda_{ij}$'s.

The independent normal maximum likelihood method uses the first two sample moments

$$
\begin{align*}
\tilde{Z} &= \frac{1}{n_1n_2} \sum_{s \in D} Z(s) = \begin{pmatrix} \tilde{Z}^{(1)} \\ \tilde{Z}^{(2)} \end{pmatrix}, \\
S &= \frac{1}{n_1n_2 - 1} \sum_{s \in D} (Z(s) - \tilde{Z}) (Z(s) - \tilde{Z})',
\end{align*}
$$

where $\tilde{Z}^{(1)}$ is $(p - k) \times 1$ and $\tilde{Z}^{(2)}$ is $k \times 1$. Given an estimator $\hat{\lambda}$ of $\lambda$, $\mu$ and $\gamma$ are estimated by

$$
\begin{align*}
\hat{\mu} &= \tilde{Z}^{(2)}, \\
\hat{\gamma} &= Z^{(1)} - B(\hat{\lambda}) \tilde{Z}^{(2)}.
\end{align*}
$$

Thus, we concentrate on estimation of

$$
\theta = \begin{pmatrix} \lambda \\ \text{vech } \Phi \\ \psi \end{pmatrix}
$$

based on $S$, where $\text{vech } \Phi$ denotes the $\frac{k(k+1)}{2}$ vector consisting of the distinct elements of $\Phi$. Let

$$
\Sigma(\theta) = \text{var}\{Z(s)\} = \begin{pmatrix} B(\lambda) \\ I_k \end{pmatrix} \Phi ((B(\lambda))', I_k) + \text{diag}(\psi).
$$
Then, the independent normal maximum likelihood estimator $\hat{\theta} = (\hat{\lambda}', (\text{vech } \hat{\Phi})', \hat{\psi})'$ is the value of $\theta$ in the parameter space $\Theta$ that minimizes

$$l(\theta) = n_1 n_2 \left( \log |\Sigma(\theta)| + \text{tr}[S (\Sigma(\theta))^{-1}] - \log |S| - p \right).$$

(2.4)

The standard likelihood ratio test statistic for the goodness of fit of model (1.1) is $l(\hat{\theta})$. Under the independent normal assumption, a large value of $l(\hat{\theta})$ compared to the $\chi^2_k$ distribution indicates a poor fit of model (1.1) with $k$ factors, where

$$d = \frac{p(p + 1)}{2} - \text{dim}(\theta).$$

(2.5)

The following theorem gives the limiting distributions of $\hat{\theta}$ and $l(\hat{\theta})$ for spatially dependent observations with unspecified distribution. Denote the true values of the unknown factor loading parameters $\lambda$ and the error variances $\psi$ by $\lambda_0$ and $\psi_0$, respectively.

**Theorem** Let model (1.1) hold. Assume that $W(s), e_1(s),...,e_p(s)$ are independent second-order stationary processes with $E[e_i(s)] = 0$, and that $W(s)$ satisfies, as $n_1 \to \infty, n_2 \to \infty$,

$$\bar{W} = \frac{1}{n_1 n_2} \sum_{s \in D} W(s) \to \mu_0, \text{ a.s.}$$

$$\frac{1}{n_1 n_2 - 1} \sum_{s \in D} (W(s) - \bar{W})(W(s) - \bar{W})' \to \Phi_0, \text{ a.s.} .$$

For this $\Phi_0$, let

$$\theta_0 = \left( \begin{array}{c} \lambda_0 \\ \text{vech } \Phi_0 \\ \psi_0 \end{array} \right).$$

Assume the identifiability conditions that for any $\epsilon > 0$, there is a $\delta_\epsilon > 0$ satisfying for any $\theta \in \Theta$,

$$||\theta - \theta_0|| > \epsilon \implies ||\text{vech } \Sigma(\theta) - \text{vech } \Sigma(\theta_0)|| > \delta_\epsilon,$$

and that $\frac{\partial \text{vec} \Sigma(\theta)}{\partial \theta'}|_{\theta = \theta_0}$ has full column rank.
If $e_i(s)$ is a spatially-independent process, $i = 1, ..., p$, then as $n_1 \to \infty$, $n_2 \to \infty$,

$$\sqrt{n_1 n_2} (\dot{\lambda} - \lambda_0) \xrightarrow{c} N[0, V_\lambda],$$

(2.7)

$$l(\dot{\theta}) \xrightarrow{c} \chi^2_d,$$

(2.8)

where $d$ is given in (2.5), and $V_\lambda$ is obtained as if $W(s)$ is a spatially-independent normal process and $e_i(s)$ is a normal process, $i = 1, ..., p$. If, in addition, $e_i(s)$ is a normal process, $i = 1, ..., p$, then as $n_1 \to \infty$, $n_2 \to \infty$,

$$\sqrt{n_1 n_2} \left[ \begin{pmatrix} \dot{\lambda} \\ \dot{\psi} \end{pmatrix} - \begin{pmatrix} \lambda_0 \\ \psi_0 \end{pmatrix} \right] \xrightarrow{c} N[0, V_{\lambda \psi}],$$

(2.9)

where $V_{\lambda \psi}$ is obtained as if $W(s)$ is a spatially-independent normal process.

Proof: The proofs of statements (2.7), (2.8), and (2.9) can be obtained by applying Corollary 3 of Anderson and Amemiya (1988), Theorem 1 of Amemiya and Anderson (1990), and Theorem 2R of Amemiya, Fuller, and Pantula (1987) and its proof, respectively.

Note that no distributional form of $W(s)$ and $e(s)$ is assumed for the results (2.7) and (2.8). The condition (2.6) holds for a large class of stationary processes with finite fourth moments. The importance of Theorem is that a special case of model (1.1) with spatially independent $e(s)$ permits the use of simple and valid statistical inference procedures. Note that $V_\lambda$ and $V_{\lambda \psi}$ are functions of only second moments and can be estimated easily using only $\Sigma(\dot{\theta})$. In fact, an estimated covariance matrix obtained by an independent normal maximum likelihood software package can be used for making inference for $\lambda$ and $\psi$. Such procedures using only sample second moments are not only simple but also accurate in finite samples. For $\lambda$ containing relationship coefficients (factor loadings), no distributional form for $W(s)$ and $e(s)$ needs to be specified. To make inference for the error variances $\psi$, the normality of $e(s)$ is required, but the distribution of $W(s)$ can be of any form.
Another significant implication of this theorem is that the result (2.8) provides a proper goodness-of-fit procedure. That is, a test using \( l(\hat{\theta}) \) and the \( \chi^2 \) distribution is an asymptotically correct test for assessing the fit of model (1.1) with spatially independent errors. This test can be carried out without specifying any distributional form, and is readily available in the software packages. In addition, as discussed in the next section, the knowledge of what is being tested allows the use of the test as a part of an overall model checking and building procedure in general situations. The finite sample properties of \( \hat{\theta} \), associated inference procedures, and \( l(\hat{\theta}) \) are assessed by simulation in Section 5.

The independent normal maximum likelihood estimator \( \hat{\theta} \) defined in (2.4) is consistent for \( \theta \) for any distribution of \( W(s) \) and \( e_i(s) \), whether or not the \( e_i(s) \)'s are spatially independent. But, since the observations \( Z(s) \) are spatially dependent and may not have a normal distribution, \( \hat{\theta} \) based only on \( S \) may not be very efficient. Also, if the errors \( e_i(s) \)'s are not spatially independent, then the associated inference procedures are not guaranteed to be valid. Hence, we consider an alternative estimator of \( \theta \) incorporating some sample lagged spatial covariances, which can be used without specifying particular forms of distributions and covariance functions.

Given \( p \times 1 \) observations \( Z(s) \) obtained on an \( n_1 \times n_2 \) grid \( D \), we first create a \( 5p \times 1 \) observation vector \( Z^*(s) \) that contains \( Z_i(s) \), \( Z_i(s + \Delta_N) \), \( Z_i(s - \Delta_N) \), \( Z_i(s + \Delta_E) \), and \( Z_i(s - \Delta_E) \), \( i = 1, \ldots, p \), where \( \Delta_N \) and \( \Delta_E \) are 1-unit lags to the north and east, respectively. Then, under model (1.1) with general \( W(s) \) and \( e_i(s) \), \( \Omega = \text{var}\{Z^*\} \) can be expressed in terms of \( \theta \) in (2.3) and covariance functions of \( W(s) \) and \( e_i(s) \) at various lags. For example, for each \( i \), \( \text{cov}\{e_i(s), e_i(s + h)\} \) for \( h = \Delta_N, \Delta_E, 2\Delta_N, 2\Delta_E, \Delta_N + \Delta_E, \) and \( \Delta_N - \Delta_E \), appear in \( \Omega \). Cross-covariances between two elements of \( W(s) \) also appear. Let \( \eta \) be the vector of all parameters appearing in \( \Omega \) including \( \theta \) and additional covariances, and we write \( \Omega = \Omega(\eta) \). Because of the second-order stationarity of \( Z(s) \), some of the elements of \( \Omega(\eta) \) are redundant (beyond the usual symmetry redundancy). Let \( \omega(\eta) \) denote a \( p_\omega \times 1 \) vector of unique elements of \( \Omega(\eta) \). Each element of \( \omega(\eta) \) can
be estimated by the sample covariance function for $Z(s)$ of the form

$$
\text{cov}\{Z_i(s + h_1), Z_{\nu}(s + h_2)\} = \frac{1}{N_{h_2-h_1} - 1} \sum_{s_i, s_m \in D \atop s_m - s_i = h_2 - h_1} (Z_i(s_i) - \bar{Z}_i)(Z_{\nu}(s_m) - \bar{Z}_{\nu}),
$$

(2.10)

where $N_{h_2-h_1}$ is the number of pairs $(s_i, s_m)$ such that $s_m - s_i = h_2 - h_1$. Let $\hat{\omega}$ denote the $p_{\omega} \times 1$ vector consisting of such sample covariance functions estimating $\omega(\eta)$. Then, our augmented observation (AO) estimator $\hat{\eta}$ of $\eta$ is the least squares estimator minimizing $\|\hat{\omega} - \omega(\eta)\|^2$. The AO estimator of $\theta$ is $\hat{\theta}_A$ given as a part of $\hat{\eta}$.

Note that $\hat{\theta}_A$ uses the information in spatial lag covariances not appearing in $S$ of (2.1). Thus, we might expect $\hat{\theta}_A$ to be more efficient than the independent normal maximum likelihood estimate $\hat{\theta}$, especially when the spatial correlations in $Z(s)$ are strong and/or the distribution of $Z(s)$ differs considerably from normal. But, there is no practical test of model fit associated with the AO approach, and a proper estimated covariance matrix of $\hat{\theta}_A$ cannot be obtained without an additional assumption. Under the assumption that $Z(s)$ is spatially independent beyond a certain known range, Christensen and Amemiya (1999) suggest an estimate of the covariance matrix of a vector containing terms similar to (2.10). Using such an estimator of $\text{var}\{\hat{\omega}\}$ in the so-called sandwich formula, it is possible to obtain an estimator of $\text{var}\{\hat{\theta}_A\}$. A simulation study in Section 5 compares $\hat{\theta}$ and $\hat{\theta}_A$ and their associated inference procedures under various situations.

It turns out that the estimated $\text{var}\{\hat{\omega}\}$ obtained under the spatial range assumption, but still without specific distribution or covariance function forms, tends to be rather variable and singular, even in fairly large samples. Thus, a practical model fit test procedure based on $\hat{\theta}_A$ is not available. This is why the simple inference procedures based on the independent normal maximum likelihood $\hat{\theta}$ and the goodness-of-fit test statistic $l(\hat{\theta})$ are of practical importance. Also, a test suggested in the next section for the condition of spatially independent errors under which such simple procedures are valid should be a part of data analysis and model building.
3 Tests for model fit

As mentioned in Section 1, factor analytic models similar to (1.1) have been proposed in the literature, but no proper method for testing goodness of fit has been proposed. The difficulty in developing such a test is partly related to the fact that any estimator of moments of order higher than two is very variable and unstable when based on spatially dependent data. Here, based on Theorem in the previous section, we propose a practical approach for assessing the fit of model (1.1) with general, spatially dependent \( W(s) \) and \( e_i(s) \), and for selecting an appropriate model of type (1.1).

We start with an interpretation of the \( \chi^2 \) goodness-of-fit test implied by Theorem. That is, if the test using \( l(\hat{\theta}) \) and \( \chi^2 \) is nonsignificant, then model (1.1) with a particular \( k \) (number of factors) and with spatially independent \( e_i(s) \)'s is not rejected. The test statistic \( l(\hat{\theta}) \) will be large when an insufficient number of factors have been included in the model and/or the errors are spatially dependent. As we will show by simulation in Section 5, this test has very large power for detecting the insufficient number of fitted factors but has limited power for detecting the spatial error dependence with the correct number of factors. This is due to the fact that \( \hat{\theta} \) is consistent for \( \theta \) if model (1.1) holds with correct \( k \) even if the \( e_i(s) \)'s are spatially dependent. Then, even if the \( l(\hat{\theta}) \)-test is nonsignificant, we need to further examine the error spatial dependence before comfortably using the standard errors suggested in Theorem. We develop such a test that can be combined with the \( l(\hat{\theta}) \)-test to produce a systematic approach for model checking and building.

It can be shown that if model (1.1) holds with correct \( k \), then \( \hat{\theta} - \theta = O_p(\frac{1}{\sqrt{n_1 n_2}}) \). Hence, \( \hat{\theta} \) can be used to form an observed error contrast that estimates a linear function of \( e(s) \). Write

\[
Z(s) = \begin{pmatrix} Z^{(1)}(s) \\ Z^{(2)}(s) \end{pmatrix},
\]

where \( Z^{(1)}(s) \) is \( (p - k) \times 1 \) and \( Z^{(2)}(s) \) is \( k \times 1 \). Then, using \( \hat{\theta} \), define the observed error
contrasts

\[ \hat{v}(s) = Z^{(1)}(s) - \hat{\gamma} - B(\hat{\lambda}) Z^{(2)}(s), \]  
(3.2)

where \( \hat{\gamma} \) is given in (2.2) and \( \hat{\lambda} \) is the normal maximum likelihood estimator of \( \lambda \). This \((p-k) \times 1\) vector estimates

\[ v(s) = Z^{(1)}(s) - \gamma_0 - B(\lambda_0) Z^{(2)}(s) \]

\[ = (I_{p-k}, -B(\lambda_0)) e(s). \]

If \( e(s) \) is spatially independent, so is \( v(s) \). Under model (1.1), functions of \( e(s) \) that are not linear functions of \( v(s) \) are not estimable. Hence, a test for spatial independence of \( v(s) \) based on \( \hat{v}(s) \) can be used as a reasonable test for the spatial independence of \( e(s) \) under model (1.1).

The standard test for spatial independence of a univariate process is a test based on Moran’s contiguity ratio which is discussed in detail by Cliff and Ord (1981). For a collection of data \( x = (x(s_1), ..., x(s_N))' \) at \( N \) locations, Moran’s ratio is

\[ \frac{N x' G x}{(1_N' G 1_N) (x' x)'}, \]

where \( G = \{g_{ij}\} \) is a given weight matrix, and \( 1_N \) is an \( N \times 1 \) vector of ones. Typically, the elements of the weight matrix \( \{g_{ij}\} \) are chosen based on some function of the distance between \( s_i \) and \( s_j \). The asymptotic normal distribution of this test statistic under the assumption of spatial independence is used to examine the significance level. See, e.g., Cliff and Ord (1981).

To apply the univariate Moran’s spatial dependence test to \( \hat{v}(s) \) in (3.2), we note that \( \hat{v}(s) \) consists of \( (p-k) \) correlated processes. We can estimate the covariance matrix of \( \hat{v}(s) \) under the model using

\[ \text{var}\{\hat{v}(s)\} = \begin{bmatrix} \hat{\psi}_1 & 0 \\ \hat{\psi}_2 & \ddots \\ 0 & \ddots & \hat{\psi}_{p-k} \end{bmatrix} + B(\hat{\lambda}) \begin{bmatrix} \hat{\psi}_{p-k+1} & 0 \\ 0 & \ddots & \hat{\psi}_p \end{bmatrix} (B(\hat{\lambda}))'. \]
Then, the elements of $\hat{v}^*(s) = (\text{Var}\{\hat{v}(s)\})^{-1/2}\hat{v}(s)$ are approximately uncorrelated. We apply the univariate Moran's test (3.3) to each of the $(p - k)$ elements of $\hat{v}^*(s)$. To account for the simultaneity of the tests, we use the Bonferroni approach. That is, our test for the spatial dependence of $e(s)$ is significant at level $\alpha$ if at least one of the $(p - k)$ elements of $\hat{v}^*(s)$ is significant at level $\alpha/(p - k)$. For the weight matrix $G$ in (3.3), we suggest the use of the generic weight $g_{ij} = 1/\|s_i - s_j\|^2$ for each of the $p - k$ elements of $\hat{v}^*(s)$. This Bonferroni test is our joint test for error spatial independence in model (1.1).

To develop a systematic method for selecting an appropriate model, we consider combining the $\chi^2$-test using $l(\hat{\theta})$ and the joint Bonferroni test for error spatial independence. As in the standard factor analysis, we start with a small $k$ (the number of factors) or a $k$ suggested by some theory. First the $\chi^2$-test using $l(\hat{\theta})$ is performed at level $\alpha_1$. If this test is significant, the $k$-factor spatially-independent error (SIE) model does not have an adequate fit. Since this rejection occurs mostly due to an insufficient $k$, the model with $k + 1$ factors should be estimated with the independent normal maximum likelihood approach and the $\chi^2$ test for the new model should be performed. If the model with $k + 1$ factors cannot be fitted by the independent normal maximum likelihood, then we consider the AO estimation of such a model. In the AO fit, the spatially-dependent error (SDE) model can be considered. If the $(k+1)$-factor model cannot be fitted by the AO method, then no factor analysis model fits this data set well. If the $\chi^2$ statistic $l(\hat{\theta})$ for the $k$-factor model is not significant at level $\alpha_1$, we then apply the joint Bonferroni test for error spatial dependence at level $\alpha_2$ (i.e., $\alpha_2/(p - k)$ for each component of $\hat{v}^*(s)$). When this Bonferroni test is also nonsignificant, the $k$-factor SIE model is accepted. If the Bonferroni test for error spatial dependence is significant, then we conclude that the $k$-factor SDE model is the selected model. In this case, the AO approach should be used for making inference about the parameters. Figure 1 graphically illustrates this approach.
Figure 1 Procedure for model selection. Possible outcomes include models with various numbers of factors and with spatially-independent-error (SIE) or spatially-dependent-error (SDE). If a \((k+1)\)-factor model cannot be fitted by the AO method, no factor model fits well.
4 Prediction of latent variables

Frequently, the researcher is interested in obtaining a prediction or map of the latent variable of interest. For example, when measuring ground temperatures using several sources of data, the researcher might wish to use a latent variable model in order to relate the observed variables to the unobservable "actual" temperature process. In such a case, the model parameters can be utilized in order to predict the latent variable at each observed location or at some new locations.

Here we show that a measurement-error-free kriging prediction of \( \mathbf{W}(s) \) in model (1.1) at any \( s_0 \) (\( s_0 \) not necessarily a location on the original observational grid \( \mathbf{D} \)) can be developed if the errors \( e_i(s) \) are spatially uncorrelated (white noise). If the model selection process in Section 3 suggests a model with spatially independent errors, then the white noise condition for \( e_i(s) \) holds. To define the kriging-type predictor of \( \mathbf{W}(s) \), we first define the so-called fixed factor score estimator. Given the normal maximum likelihood estimates (or AO estimates) \( \hat{\gamma}, \hat{\mu}, \hat{\lambda}, \hat{\psi} \), let

\[
\hat{\Lambda} = \begin{pmatrix} \mathbf{B}(\hat{\lambda}) \\ \mathbf{I}_k \end{pmatrix},
\]

and \( \hat{\Psi} = \text{diag}(\hat{\psi}_1, ..., \hat{\psi}_p) \). Then, the fixed factor score estimator is the generalized least squares estimator of \( \mathbf{W}(s) \) given by

\[
\mathbf{W}^{(\text{fixed})}(s) = (\hat{\Lambda}' \hat{\Psi}^{-1} \hat{\Lambda}' \hat{\Psi}^{-1})^{-1} \hat{\Lambda}' \hat{\Psi}^{-1} (\mathbf{Z}(s) - \bar{Z}).
\]

An alternative form of \( \mathbf{W}^{(\text{fixed})}(s) \) without involving \( \hat{\Psi}^{-1} \) is

\[
\mathbf{W}^{(\text{fixed})}(s) = \mathbf{Z}^{(2)}(s) - \hat{U} \hat{\nu}(s),
\]

where

\[
\hat{U} = -\hat{\Psi}_2(\mathbf{B}(\hat{\lambda}))(\hat{\Psi}_1 + \mathbf{B}(\hat{\lambda})\hat{\Psi}_2(\mathbf{B}(\hat{\lambda})))^{-1},
\]

\[
\hat{\Psi}_1 = \text{diag}(\hat{\psi}_1, ..., \hat{\psi}_{p-k}),
\]

\[
\hat{\Psi}_2 = \text{diag}(\hat{\psi}_{p-k+1}, ..., \hat{\psi}_p),
\]
and \( Z^{(2)}(s) \) and \( \tilde{v}(s) \) are defined in (3.1) and (3.2). It can be shown that, except for the parameter estimation errors, \( R(s) = \hat{W}^{(\text{fixed})}(s) - W(s) \) is a linear function of \( e(s) \) and is independent of \( W(s) \). If \( e(s) \) is a white noise process, so is \( R(s) \) (approximately). Also, \( \text{var}\{R(s)\} \) can be estimated by
\[
\text{var}\{R(s)\} = \hat{\Psi}_2 + \hat{\Upsilon}_B(\hat{\lambda}) \hat{\Psi}_2. \tag{4.2}
\]
Hence, treating all \( \hat{W}^{(\text{fixed})}(s), s \in D \), as observations and (4.2) as the white-noise measurement error covariance matrix, we can apply a measurement-error-free kriging approach (see Cressie, 1993, Section 3.2.1) to predict each element of \( W(s) \) at any \( s_0 \). This approach yields a predictor which uses all of the data in the observed multivariate process \( Z(s) \). We will refer to this predictor as the “white-noise-error kriging” (WNEK) predictor \( \hat{W}^{(\text{WNEK})}(s_0) \) of \( W_j(s_0) \). Although the approach for predicting a single factor \( W_j(s_0) \) is given here, a simultaneous prediction of the components of \( W(s_0) \) can be obtained using techniques for co-kriging (see Cressie, 1993, Section 3.2.3) or multivariate spatial prediction (Ver Hoef and Cressie, 1991).

Suppose we model \( \hat{W}^{(\text{fixed})}_j(s) \) using a parametric model for covariance functions (e.g., spherical or exponential) with parameters \( \zeta \) so that \( \text{cov}\{\hat{W}^{(\text{fixed})}_j(s_i), \hat{W}^{(\text{fixed})}_j(s_l)\} = C(s_i - s_l; \zeta) \). Then, the WNEK predictor of \( W_j(s_0) \) from \( \hat{W}^{(\text{fixed})}_j = (\hat{W}^{(\text{fixed})}_j(s_1), \ldots, \hat{W}^{(\text{fixed})}_j(s_N))' \) is of the form
\[
\hat{W}^{(\text{WNEK})}_j(s_0) = a' \hat{W}^{(\text{fixed})}_j. \tag{4.3}
\]
The kriging coefficients \( a \) and Lagrange multiplier \( m \) (ensuring the unbiasedness) are obtained using
\[
\begin{pmatrix}
a \\
m
\end{pmatrix} = C^{-1}_{\hat{W}^{(\text{fixed})}_j} \hat{W}^{(\text{fixed})}_j, \tag{4.4}
\]
where
\[
C_{\hat{y}_j^{(\text{fixed})}} = \begin{bmatrix}
C(s_1 - s_1; \hat{\zeta}) & C(s_1 - s_2; \hat{\zeta}) & \cdots & C(s_1 - s_N; \hat{\zeta}) & 1 \\
C(s_2 - s_1; \hat{\zeta}) & C(s_2 - s_2; \hat{\zeta}) & \cdots & C(s_2 - s_N; \hat{\zeta}) & 1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
C(s_N - s_1; \hat{\zeta}) & C(s_N - s_2; \hat{\zeta}) & \cdots & C(s_N - s_N; \hat{\zeta}) & 1 \\
1 & 1 & 1 & 1 & 0
\end{bmatrix},
\]
(4.5)
and the vector \(c\) is constructed such that
\[
c_i = \begin{cases} 
C(s_0 - s_i; \hat{\zeta}) - I_{(s_0 = s_i)} \sigma^2_{R_j}, & i = 1, \ldots, N, \\
1, & i = N + 1,
\end{cases}
\]
(4.6)
with
\[
I_{(s_0 = s_i)} = \begin{cases} 
1, & s_0 = s_i, \\
0, & s_0 = s_i,
\end{cases}
\]
and \(\sigma^2_{R_j}\) is the \((j, j)\) element of \(\text{var}\{R(s)\}\) in (4.2). Disadvantages of this procedure are that some parametric covariance function model needs to be fitted to \(\hat{W}_j^{(\text{fixed})}(s)\) in the kriging process, and that errors \(e_i(s)\) have to be white noise for practical implementation.

An alternative approach for predicting \(W(s)\) is possible without the white noise or spatially independent error condition. Christensen and Amemiya (1999) describe a limited-information version of the standard best linear unbiased predictor \(W(s)\) which can be implemented based on the augmented observation \(Z^*(s)\) described in Section 2. For each location \(s_0\) on the sampled grid, the prediction of \(W(s_0)\) uses the information in the contemporaneous and neighboring values of \(Z(s_0)\), but not all observations \(Z(s), s \in D\). This "limited-information BLUP" is the estimated BLUP of \(W(s)\) based on \(Z^*(s)\) for each \(s\), where the coefficients needed in constructing the predictor can be obtained using only the AO estimator \(\hat{\eta}\) of \(\eta\) given in Section 2. We denote this predictor by \(\hat{W}^{(\text{LIBLUP})}(s)\). In addition, the fixed factor score estimate of (4.1) can be obtained using \(Z^*(s)\) in place of \(Z(s)\) and the AO estimator \(\hat{\eta}\) in place of the normal maximum likelihood estimator \(\hat{\theta}\). We denote the augmented fixed factor score estimator by \(\hat{W}_{\Lambda}^{(\text{fixed})}(s)\). Note that \(\hat{W}^{(\text{fixed})}(s), \hat{W}_{\Lambda}^{(\text{fixed})}(s), \) and \(\hat{W}^{(\text{LIBLUP})}(s)\) can predict \(W(s)\) only at those \(s\) on the sampled grid \(D\). The measurement-error-free kriging predictor \(\hat{W}^{(\text{WNEK})}(s)\) can be used
to predict $W(s)$ at any $s$, even off the grid, but requires some parametric modeling of the covariance function of $W(s)$. The four predictors are compared by simulation in Section 5.

5 Simulation results

5.1 Inference and model building procedures

In this simulation, we consider the properties associated with the normal maximum likelihood estimator $\hat{\theta}$, including bias, mean squared error (MSE), and coverage probability. We also evaluate the Type I error rate and power of the model-fitting statistics described in Section 3, and assess the usefulness of the model-building procedure given in Section 3. Consider model (1.1) with $p = 8$ for a $20 \times 20$ lattice $D$ with a common row and column distance. For the number of factors, we considered $k = 1$ and $k = 2$. The true factor and error processes were generated as stationary processes with isotropic spherical variograms. The parameters for the spherical variograms were: range $= 6$ units, partial sill $= \frac{7}{3}$, and nugget $= 0$ for each of the factor processes, and range $= \tau_e$, partial sill $= 1$, and nugget $= 0$ for each of the error processes, where $\tau_e = 0$ (spatial independence) or $\tau_e = 2$. For the common distributional form for the factor and errors, we used normal and linearly transformed lognormal. Thus, there are eight configurations depending on the number of factors $k = 1, 2$, the range of error spatial dependence $\tau_e = 0, 2$, and the distributional form, normal or lognormal. For each of the eight configurations, 2000 samples were generated.

For each generated data set, the 1-factor and 2-factor spatially-independent-error models were fitted using the independent normal maximum likelihood. Parameter estimates, their estimated standard errors under the spatially-independent-error model, the $\chi^2$ goodness-of-fit statistic, and the Moran test statistics for $\hat{\nu}^*(s)$ were obtained.

Table 1 summarizes the empirical bias, MSE, and coverage probability for a nominal 95% confidence interval for three groups of parameters ($\Lambda$, $\Phi$, and $\psi$). Each entry is the average value over the parameters in each group. Table 1 only reports on the cases
Table 1 Values of the following quantities averaged over groups of the normal maximum likelihood parameter estimates: bias, MSE, and coverage probability for a nominal 95% confidence interval.

<table>
<thead>
<tr>
<th>Distribution of factor(s) and errors</th>
<th># of factors</th>
<th>Spatially-independent errors?</th>
<th>Parameters</th>
<th>Average bias</th>
<th>Average MSE</th>
<th>Average coverage probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal</td>
<td>1</td>
<td>yes</td>
<td>$\lambda$</td>
<td>0.0021</td>
<td>0.0028</td>
<td>0.9491</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\Phi$</td>
<td>-0.1044</td>
<td>0.2505</td>
<td>0.6015</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\psi$</td>
<td>-0.0029</td>
<td>0.0068</td>
<td>0.9456</td>
</tr>
<tr>
<td>normal</td>
<td>1</td>
<td>no</td>
<td>$\lambda$</td>
<td>0.0044</td>
<td>0.0059</td>
<td>0.8204</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\Phi$</td>
<td>-0.1039</td>
<td>0.2757</td>
<td>0.5765</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\psi$</td>
<td>-0.0095</td>
<td>0.0097</td>
<td>0.8850</td>
</tr>
<tr>
<td>normal</td>
<td>2</td>
<td>yes</td>
<td>$\lambda$</td>
<td>0.0006</td>
<td>0.0029</td>
<td>0.9481</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\Phi$</td>
<td>-0.0740</td>
<td>0.2133</td>
<td>0.6273</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\psi$</td>
<td>-0.0063</td>
<td>0.0096</td>
<td>0.9419</td>
</tr>
<tr>
<td>normal</td>
<td>2</td>
<td>no</td>
<td>$\lambda$</td>
<td>0.0011</td>
<td>0.0055</td>
<td>0.8368</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\Phi$</td>
<td>-0.0702</td>
<td>0.2346</td>
<td>0.6022</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\psi$</td>
<td>-0.0163</td>
<td>0.0136</td>
<td>0.8849</td>
</tr>
<tr>
<td>lognormal</td>
<td>1</td>
<td>yes</td>
<td>$\lambda$</td>
<td>0.0022</td>
<td>0.0031</td>
<td>0.9506</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\Phi$</td>
<td>-0.1014</td>
<td>1.2971</td>
<td>0.3355</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\psi$</td>
<td>-0.0026</td>
<td>0.2575</td>
<td>0.3265</td>
</tr>
<tr>
<td>lognormal</td>
<td>1</td>
<td>no</td>
<td>$\lambda$</td>
<td>0.0043</td>
<td>0.0067</td>
<td>0.8274</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\Phi$</td>
<td>-0.1006</td>
<td>1.3257</td>
<td>0.3290</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\psi$</td>
<td>-0.0087</td>
<td>0.2586</td>
<td>0.3228</td>
</tr>
<tr>
<td>lognormal</td>
<td>2</td>
<td>yes</td>
<td>$\lambda$</td>
<td>0.0010</td>
<td>0.0034</td>
<td>0.9485</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\Phi$</td>
<td>-0.0891</td>
<td>0.9410</td>
<td>0.4488</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\psi$</td>
<td>-0.0052</td>
<td>0.2562</td>
<td>0.3734</td>
</tr>
<tr>
<td>lognormal</td>
<td>2</td>
<td>no</td>
<td>$\lambda$</td>
<td>0.0015</td>
<td>0.0067</td>
<td>0.8348</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\Phi$</td>
<td>-0.0842</td>
<td>0.9623</td>
<td>0.4393</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\psi$</td>
<td>-0.0142</td>
<td>0.2579</td>
<td>0.3639</td>
</tr>
</tbody>
</table>
where the correct number of factors \((k = 1, 2)\) is fitted.

Regardless of the data generation scenario, the independent normal maximum likelihood estimator \(\hat{\lambda}\) has small bias and MSE, and the bias of \(\hat{\psi}\) is small. The MSE’s for the elements of \(\hat{\Phi}\) and \(\hat{\psi}\) are much higher with lognormal distributions than with normal. The coverage probability results are exactly as expected for the theory in Section 2. When error processes are spatially independent, the empirical coverage probability for \(\lambda\) is virtually equal to the nominal 95% level, regardless of distributions and number of factors, as supported by Theorem. Thus, even for a field as small as \(20 \times 20\), the asymptotic confidence interval for \(\lambda\) using the independent normal maximum likelihood is remarkably accurate. But, when the error spatial dependency is present with the range parameter as small as 2, the empirical coverage probability drops to the 82%–84% level. Also as given in Theorem, the coverage probability for \(\psi\) is nearly identical to the nominal level when the errors are spatially independent normal processes. The empirical coverage probability drops to the 88%–89% level with normal spatially dependent errors, and to the 30%–40% level with spatially independent or dependent lognormal errors. Hence, the inference procedures for the parameters using the independent normal maximum likelihood are very accurate and useful, but only for the cases covered in Theorem.

Table 2 gives the proportion of times that the \(\chi^2\) goodness-of-fit test with size \(\alpha = 0.05\) "accepts" (i.e., is nonsignificant for) the 1-factor and 2-factor spatially-independent-error (SIE) models, and the proportion of times that the joint Moran test for \(\hat{\psi}^*(s)\) "accepts" the hypothesis of spatially-independent errors for both the 1-factor and the 2-factor fitted models. When the correct SIE model is fitted, the Type I error rate for the \(\chi^2\) goodness-of-fit test is reasonably close to the nominal (0.05) level (although it can be close to 0.1 for lognormal processes). The power of the goodness-of-fit test was very high (100%) for detecting an insufficient number of factors, but was less than 35% for detecting spatially-correlated errors with the proper number of factors fitted. For this reason, we recommend that the joint Moran test for \(\hat{\psi}^*(s)\) be used to make certain that the \(k\)-factor SIE model should be accepted. When a \(k\)-factor SIE model is fitted to the
Table 2 The proportion of times that the $\chi^2$ goodness-of-fit (GOF) test and the joint Moran test for $\hat{\nu}^*(s)$ are nonsignificant at the $\alpha = 0.05$ level.

<table>
<thead>
<tr>
<th>Distribution of factor(s) and errors</th>
<th># of factors</th>
<th>Spatially-independent errors?</th>
<th>GOF test for SIE (1-factor model)</th>
<th>Moran test “accepts”</th>
<th>GOF test for SIE (2-factor model)</th>
<th>Moran test “accepts”</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal</td>
<td>1</td>
<td>yes</td>
<td>0.9485</td>
<td>0.9510</td>
<td>0.9955</td>
<td>0.9480</td>
</tr>
<tr>
<td>normal</td>
<td>1</td>
<td>no</td>
<td>0.6580</td>
<td>0.0000</td>
<td>0.9555</td>
<td>0.0000</td>
</tr>
<tr>
<td>normal</td>
<td>2</td>
<td>yes</td>
<td>0.0000</td>
<td>0.0025</td>
<td>0.9465</td>
<td>0.9490</td>
</tr>
<tr>
<td>normal</td>
<td>2</td>
<td>no</td>
<td>0.0000</td>
<td>0.0025</td>
<td>0.7355</td>
<td>0.0000</td>
</tr>
<tr>
<td>lognormal</td>
<td>1</td>
<td>yes</td>
<td>0.9010</td>
<td>0.9425</td>
<td>0.9965</td>
<td>0.9445</td>
</tr>
<tr>
<td>lognormal</td>
<td>1</td>
<td>no</td>
<td>0.6650</td>
<td>0.0000</td>
<td>0.9595</td>
<td>0.0000</td>
</tr>
<tr>
<td>lognormal</td>
<td>2</td>
<td>yes</td>
<td>0.0000</td>
<td>0.0095</td>
<td>0.9070</td>
<td>0.9460</td>
</tr>
<tr>
<td>lognormal</td>
<td>2</td>
<td>no</td>
<td>0.0000</td>
<td>0.0100</td>
<td>0.7445</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

SIE data with the true number of factors $\leq k$, the Type I error rate for the joint Moran test for $\hat{\nu}^*(s)$ is very close to the nominal level. If the errors are spatially correlated and/or we have insufficient factors in the fitted model, the joint Moran test for $\hat{\nu}^*(s)$ has power $\geq 99\%$.

The model selection procedure described in Section 3 combining the $\chi^2$ goodness-of-fit and the joint Moran test was applied to the 2000 data sets generated under each scenario considered in the study. For significance level, we used $\alpha_1 = \alpha_2 = 0.025$. Table 3 gives the proportion of data sets that were classified as each of the following models: 1 factor SIE, 1 factor SDE, 2 factors SIE, 2 factors SDE, and “others.” The “others” category includes SIE and SDE models with $> 2$ factors, and the results of no well-fitting factor model. Table 3 gives classification rates, where the proportions in each row sum to 1.

Table 3 shows that data sets generated from SDE models were never erroneously classified as being generated from a SIE model. With the choice of $\alpha_1 = \alpha_2 = 0.025$, the probability of selecting the correct model is reasonably high for all types of the true
Table 3 Data set classification rates by the combined model selection approach with $\alpha_1 = \alpha_2 = 0.025$.

<table>
<thead>
<tr>
<th>Distribution of factor(s) and errors</th>
<th>Spatially-independent factors?</th>
<th>1 factor</th>
<th>1 factor</th>
<th>2 factors</th>
<th>2 factors</th>
<th>&quot;others&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal</td>
<td>yes</td>
<td>0.9460</td>
<td>0.0225</td>
<td>0.0310</td>
<td>0.0000</td>
<td>0.0005</td>
</tr>
<tr>
<td>normal</td>
<td>no</td>
<td>0.0000</td>
<td>0.7540</td>
<td>0.0000</td>
<td>0.2220</td>
<td>0.0240</td>
</tr>
<tr>
<td>normal</td>
<td>yes</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.9440</td>
<td>0.0265</td>
<td>0.0295</td>
</tr>
<tr>
<td>normal</td>
<td>no</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.8195</td>
<td>0.1805</td>
</tr>
<tr>
<td>lognormal</td>
<td>yes</td>
<td>0.8985</td>
<td>0.0315</td>
<td>0.0680</td>
<td>0.0010</td>
<td>0.0010</td>
</tr>
<tr>
<td>lognormal</td>
<td>no</td>
<td>0.0000</td>
<td>0.7380</td>
<td>0.0000</td>
<td>0.2415</td>
<td>0.0205</td>
</tr>
<tr>
<td>lognormal</td>
<td>yes</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.9145</td>
<td>0.0250</td>
<td>0.0605</td>
</tr>
<tr>
<td>lognormal</td>
<td>no</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.8115</td>
<td>0.1885</td>
</tr>
</tbody>
</table>

model. This classification procedure could be considered to be even more accurate if only those models not containing the true model are considered incorrect. Thus, this model selection approach is useful in practice.

5.2 Comparison of parameter estimators and latent variable predictors

In this section, we compare the independent normal maximum likelihood (ML) and augmented observation (AO) model-fitting approaches in terms of parameter estimation and factor value prediction. We used model (1.1) with $p = 5$ and $k = 1$ factor on the domain $D$ of a $20 \times 20$ square lattice. The factor process (always a linearly transformed lognormal) and error processes (either normal or linearly transformed lognormal) have isotropic spherical variograms with 0 nugget. For the factor process, range = 5 units and partial sill = $\frac{5}{3}$. For each of the error processes, range = $\tau_e = 0$ or 2 units, and partial sill = 1. For each of four configurations specified by the error distribution (normal or lognormal) and by the error spatial dependence ($\tau_e = 0$ or 2), 500 samples were generated. From each sample, the ML fit of the SIE model and the AO fit of the SDE model were obtained.

Table 4 reports empirical bias, MSE, and coverage probability of a nominal 95%
Table 4 Bias, MSE, and coverage probability for estimated factor loadings.

<table>
<thead>
<tr>
<th>Distribution of errors</th>
<th>Spatially-independent errors?</th>
<th>Model-fitting approach</th>
<th>Average bias</th>
<th>Average MSE</th>
<th>Average coverage probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal</td>
<td>yes</td>
<td>ML</td>
<td>0.0038</td>
<td>0.0032</td>
<td>0.9515</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AO</td>
<td>0.0027</td>
<td>0.0033</td>
<td>0.9769</td>
</tr>
<tr>
<td>normal</td>
<td>no</td>
<td>ML</td>
<td>0.0042</td>
<td>0.0062</td>
<td>0.8325</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AO</td>
<td>0.0052</td>
<td>0.0086</td>
<td>0.9613</td>
</tr>
<tr>
<td>lognormal</td>
<td>yes</td>
<td>ML</td>
<td>0.0044</td>
<td>0.0029</td>
<td>0.9524</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AO</td>
<td>0.0045</td>
<td>0.0032</td>
<td>0.9839</td>
</tr>
<tr>
<td>lognormal</td>
<td>no</td>
<td>ML</td>
<td>0.0054</td>
<td>0.0056</td>
<td>0.8495</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AO</td>
<td>0.0075</td>
<td>0.0082</td>
<td>0.9703</td>
</tr>
</tbody>
</table>

Confidence interval, averaged over the 4 factor loadings in the model. The average MSE of the ML-SIE estimator is smaller than that of the AO-SDE estimator for all four configurations. Thus, the use of augmented lagged variables does not seem to improve the efficiency of the factor loading estimator. The ML confidence interval is very accurate for SIE data, but has a coverage probability that tends to be smaller for SDE data. The coverage probability of the AO confidence interval is reasonably close to the nominal level for all four cases.

For predicting the factor process \( W_1(s) \), we considered \( \hat{W}_1^{(\text{fixed})}(s) \) and \( \hat{W}_1^{(\text{WNEK})}(s) \) in (4.1) and (4.3) using ML, and \( \hat{W}_{A,1}(s) \) and \( \hat{W}_1^{(\text{LIBLUP})}(s) \) described in Section 4 using AO. For each sample and for any predictor \( \hat{W}_1(s) \), we obtained the average squared error (ASE)

\[
\text{ASE}_{\hat{W}} = \frac{\sum_{i=1}^{N}(\hat{W}(s_i) - W_1(s_i))^2}{N},
\]

where \( N = 400 \) is the sample size. Then, the median of the 500 ASE's over replicate samples was recorded for each predictor. Table 5 summarizes the median ASE's. For comparison, note that naively using \( Z_p(s) \) as a predictor for \( W_1(s) \) yields a median ASE of approximately 1 for these data (since the variance of the \( \epsilon_p(s) \) process is equal to 1).

For spatially-independent errors, \( \hat{W}_1^{(\text{WNEK})}(s) \) and \( \hat{W}_1^{(\text{LIBLUP})}(s) \) have nearly equal
Table 5 Median ASE for four predictors of $W_1(s)$.

<table>
<thead>
<tr>
<th>Distribution of errors</th>
<th>Spatially-independent errors?</th>
<th>ML Approach</th>
<th>AO Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\hat{W}_1^{(\text{fixed})}(s)$</td>
<td>$\hat{W}_1^{(\text{WNEK})}(s)$</td>
</tr>
<tr>
<td>normal</td>
<td>yes</td>
<td>0.1862</td>
<td>0.1481</td>
</tr>
<tr>
<td>normal</td>
<td>no</td>
<td>0.1910</td>
<td>0.1815</td>
</tr>
<tr>
<td>lognormal</td>
<td>yes</td>
<td>0.1627</td>
<td>0.1327</td>
</tr>
<tr>
<td>lognormal</td>
<td>no</td>
<td>0.1672</td>
<td>0.1615</td>
</tr>
</tbody>
</table>

median ASE's. The inability of $\hat{W}_1^{(\text{WNEK})}(s)$ to dominate here may be due to the fact that a covariance function for $\hat{W}_1^{(\text{fixed})}(s)$ must be modeled in order to apply this approach, whereas the AO approach directly estimates required moments. But, $W_1(s_0)$ at $s_0$ outside of the data locations can be predicted only by the WNEK approach. The spatially-dependent errors violate the assumption needed for $\hat{W}_1^{(\text{WNEK})}(s)$, and $\hat{W}_1^{(\text{LIBLUP})}(s)$ incorporating the spatial error dependence information gives a smaller ASE. The fixed factor score estimates used for predicting $W_1(s)$ do not perform as well as their random predictor counterparts.

6 Illustrative example

We now apply the general approach developed in this paper to a data set containing measurements of several soil constituents over an approximately 350 m x 350 m field in southwestern Boone County, Iowa (Colvin, et al., 1997). Measurements of log(Ca), log(Cu), log(Mg), and log(Zn) were observed at 215 of the 224 possible locations on an 8 x 28 grid. The sites are located along 8 east-west transects, spaced 48 m apart. Along each transect there are 28 sites, spaced approximately 12.2 m apart. An agronomist
may wish postulate that this 4-dimensional process can be modeled as

\[
\begin{align*}
\log[Ca](s) &= \gamma_1 + \lambda_{11} \ W_1(s) + e_1(s), \\
\log[Cu](s) &= \gamma_2 + \lambda_{21} \ W_1(s) + e_2(s), \\
\log[Mg](s) &= \gamma_3 + \lambda_{31} \ W_1(s) + e_3(s), \\
\log[Zn](s) &= W_1(s) + e_4(s),
\end{align*}
\]

representing the existence of an underlying “soil richness.”

Our first step is to fit this model with spatially independent errors using independent normal maximum likelihood. The fitted model ignoring the intercepts \( \gamma_i \) is

\[
\begin{align*}
\log[Ca](s) &= 1.93 \ W_1(s) + e_1(s), \\
\text{std. err.} &= 0.128 \\
p-value &= < 0.0001 \\

\log[Cu](s) &= 1.11 \ W_1(s) + e_2(s), \\
\text{std. err.} &= 0.067 \\
p-value &= < 0.0001 \\

\log[Mg](s) &= 1.02 \ W_1(s) + e_3(s), \\
\text{std. err.} &= 0.070 \\
p-value &= < 0.0001 \\

\log[Zn](s) &= W_1(s) + e_4(s),
\end{align*}
\]

and the \( p \)-value for the \( \chi^2 \) goodness-of-fit test in (2.8) is 0.0586. Since this \( p \)-value is greater than \( \alpha_1 = 0.025 \), we accept the 1-factor model but still need to check whether the SIE assumption is appropriate. For each of the 3 elements of \( \hat{v}^*(s) \), the \( p \)-value for the Moran test is less than 0.0001. Thus, the joint Moran test for \( \hat{v}^*(s) \) described in Section 3 rejects SIE, and the 1-factor spatially-dependent-error (SDE) model is selected as our model for the data set. After estimating the range of spatial dependence to be equal to 150 m, we fitted the 1-factor SDE model using the AO model-fitting approach, producing a fitted model (ignoring the intercepts)
\[
\log[\text{Ca}](s) = 2.14 \ W_1(s) + e_1(s),
\]
\[
\text{std. err.} \quad 0.619
\]
\[
p-value \quad 0.0005
\]

\[
\log[\text{Cu}](s) = 1.11 \ W_1(s) + e_2(s),
\]
\[
\text{std. err.} \quad 0.276
\]
\[
p-value \quad 0.0001
\]

\[
\log[\text{Mg}](s) = 0.98 \ W_1(s) + e_3(s),
\]
\[
\text{std. err.} \quad 0.361
\]
\[
p-value \quad 0.0064
\]

\[
\log[\text{Zn}](s) = \ W_1(s) + e_4(s).
\]

Although the factor loading estimates in (6.1) and (6.2) are similar, the standard errors in (6.1) are too small, assuming the SIE model erroneously.

Using the limited-information BLUP predictor \(\hat{W}_1^{(\text{BLUP})}(s)\), the value of an underlying factor \(W_1(s)\) was predicted at each \(s\) in the sample domain. Figure 2 gives image plots of each of the four observed variables as well as the predicted factor process \(\hat{W}_1^{(\text{BLUP})}(s)\).

Note that the predicted factor process is a spatially-continuous process which clearly delineates the areas of high soil nutrient richness. Also, \((\hat{\lambda}_{11}, \hat{\lambda}_{21}, \hat{\lambda}_{31}, 1)' \hat{W}_1^{(\text{BLUP})}(s)\) is a predictor of the error-free version of \((\log[\text{Ca}](s), \log[\text{Cu}](s), \log[\text{Mg}](s), \log[\text{Zn}](s))'\).

**References**


Figure 2 Image plots (black = high, white = low) for the four observed variables and the predicted factor process.


GENERAL CONCLUSION

With the advent of satellite technology and modern computing, we now have the ability to collect and explore large, multivariate, geo-referenced data. Since disciplines such as the atmospheric and environmental sciences, ecology, and precision agriculture are closely tied to the analysis of such data, it has become necessary to develop statistical methodology for modeling and understanding the often complex relationships that exist among multiple variables collected in space.

In this dissertation, we introduced a new factor analytic methodology for the analysis of multivariate spatial data. The generalized shifted-factor model is a general latent variable model which incorporates the concept of shifted factors so that variables exhibiting lagged dependencies or asymmetric covariance relationships can be considered. The model fitting techniques incorporate important spatial correlation information and inferential procedures provide useful model building and model assessing tools. The parameter estimates obtained using these techniques have been shown to have good statistical properties. When each error process has no spatial dependence, simpler approaches for latent variable modeling of spatial data are available, taking advantage of standard statistical software packages and providing excellent estimation and inference properties. Information from the latent variable model can easily be incorporated into a predictor of spatially continuous latent variables.
ACKNOWLEDGMENTS

I would like to thank Tom Colvin and Douglas Karlen of the National Soil Tilth Laboratory, Ames, Iowa, for their kind permission to use the 1997 Baker South data in the second and third papers.

I am indebted to my major professor, Yasuo Amemiya, for demonstrating what it is like to be a selfless educator, putting the student's needs first. Thanks also to the faculty and staff working in the Iowa State University Department of Statistics, and to Alvin Rencher, my M.S. thesis advisor at Brigham Young University, for his continued mentoring and friendship.

I am grateful for the continual emotional, spiritual, and financial support that have been provided by my parents Fred and Martha Christensen, my wife's parents Michael and Maria Moody, and my brothers and sisters.

Thanks most of all to Mary Elizabeth. Her confidence, support, and positive attitude are the reasons we have been successful in this educational venture. Thanks also to Michael, Sarah, and Beth, who patiently sacrificed some "normal-kid" things while Dad finished school.