Statistica Sinica **16**(2006), 981-1020

ON ESTIMATION AND PREDICTION FOR MULTIVARIATE MULTIRESOLUTION TREE-STRUCTURED SPATIAL LINEAR MODELS

Wei Yue and Jun Zhu

University of Wisconsin - Madison

Abstract: Multiresolution tree-structured models are attractive when dealing with large amounts of spatial data in environmental sciences. With the multiresolution tree structure, a change-of-resolution Kalman filter algorithm has been devised to predict spatial processes in a computationally efficient manner (see, e.g., Huang and Cressie (1997) and Huang, Cressie and Gabrosek (2002)). In this article, we extend the multiresolution tree-structured model to account for multiple response variables. Despite the increased model complexity, we derive the theoretical properties of statistical inference and develop direct and fast algorithms for computation. For spatial process prediction, we develop a general theory of optimal projection and generalize the existing change-of-resolution Kalman filter to accommodate singularity. For model parameter estimation, we consider a factorization of the likelihood function to ensure computational efficiency. Moreover, under a fairly mild condition, we derive the distributional properties of both maximum likelihood estimates and restricted maximum likelihood estimates. We evaluate the theory and methods developed here by a simulation study.

Key words and phrases: best linear unbiased predictor, change-of-resolution Kalman filter, factorization of likelihood function.

1. Introduction

For a spatial random process $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in D\}$, where D is a spatial domain of interest in $I\!R^2$, consider a measurement error model $Z(\boldsymbol{s}) = Y(\boldsymbol{s}) + \epsilon(\boldsymbol{s})$, where $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ denotes a latent process representing the underlying truth and $\{\epsilon(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ denotes independent measurement errors. Traditional kriging predicts the latent process $Y(\cdot)$ using the best linear unbiased predictor (BLUP), based on data $Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n)$ at sampling locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$. However, when the data size n becomes large, the kriging methods suffer from slow computation because they involve operations of order $\mathcal{O}(n^3)$. Multiresolution tree-structured models have been developed in recent years to overcome the computational difficulties by imposing a multiresolution tree structure on the latent process $Y(\cdot)$. The multiresolution tree-structured model was first proposed by Chou, Willsky and Nikoukhah (1994). Huang and Cressie (1997), Huang, Cressie and Gabrosek (2002) and Zhu and Yue (2005) further developed the model to accommodate mass balance across resolutions and applied the methodology to process spatial data, such as total column of ozone. The main idea is to partition the spatial domain D into cells $\{D_{j,k} : k = 1, \ldots, N_j, j = 1, \ldots, J\}$ in a nested fashion from coarser resolutions to finer resolutions, where N_j is the number of cells on the jth resolution and J is the total number of resolutions. Associated with each cell is a node and the nodes form a multiresolution tree structure by in-between-resolution directed edges from a parent node in $D_{j,k}$ on the jth resolution to its children nodes in $D_{ch(j,k)}$ on the $(j+1)$th resolution, where the cell $D_{ch(j,k)}$ is nested in the cell $D_{j,k}$. The latent process $Y(\cdot)$ is aggregated within each cell to $y_{j,k} = |D_{j,k}|^{-1} \int_{D_{j,k}} Y(\boldsymbol{s}) d\boldsymbol{s}$ for the kth cell on the jth resolution. Then the corresponding datum is $z_{j,k} = y_{j,k} + \epsilon_{j,k}$ where $\epsilon_{j,k}$ is the corresponding measurement error. Coupled with the multiresolution tree-structured model is a change-of-resolution Kalman filter algorithm for computing the BLUP of y, which consists of a high-to-low-resolution filtering step and a low-to-high-resolution smoothing step. In the high-to-low-resolution filtering step, the optimal predictor of y is computed based on the data on the higher resolutions whereas in the low-to-high-resolution smoothing step, the prediction is based on all the data. Thus the algorithm involves operations of order $\mathcal{O}(n)$ and is more attractive than kriging for processing large amounts of data.

Here we extend the multiresolution tree-structured spatial linear model (MTSLM) in Huang and Cressie (1997), Huang, Cressie and Gabrosek (2002) and Zhu and Yue (2005) to account for multiple response variables. That is, we consider a multivariate version of MTSLM, which we call multivariate multiresolution tree-structured spatial linear model (MMTSLM). There are several challenges in extending the multiresolution tree-structured spatial linear model from univariate to multivariate response variables. One difficulty is in the change-of-resolution Kalman filter algorithm. For univariate multiresolution tree-structured models, the change-of-resolution Kalman filter assumes nonsingularity in the variance matrix involving y to ensure invertibility in the filtering and smoothing steps. For multivariate multiresolution tree-structured models, however, the variance matrix may be singular due to possible linear constraints among the latent variables. Simple adjustment can be made to ensure nonsingularity, but it is in general cumbersome to adjust the multivariate latent processes. Thus an automatic and more elegant procedure is needed. The other difficulty is in the statistical inference of the model parameters. In the literature, a popular parameter estimation is maximum likelihood (ML) using an EM algorithm (Huang, Cressie and Gabrosek (2002) and Zhu and Yue (2005)). While the EM algorithm is numerically stable, its rate of convergence can be very slow. Hence it is important to devise a computationally efficient algorithm for model parameter estimation and statistical inference. In particular, Johannesson and Cressie (2004) developed a fast and statistically efficient parameter estimation method, which utilizes a certain specific model parameterization.

Here we propose novel approaches to address both these two concerns. For spatial process prediction, we develop a general theory of optimal projection and generalize the existing change-of-resolution Kalman filter to accommodate singularity. The results are suitable not only for Gaussian processes, but also general processes with finite second moments. For model parameter estimation, we consider a factorization of the likelihood function to ensure fast computation. Furthermore, we utilize statistical linear model theory to derive the distributional properties of both ML and restricted maximum likelihood (REML) estimates, which, to our knowledge, has not been explored before.

In Section 2, we describe the multivariate multiresolution tree-structured spatial linear model (MMTSLM). We develop general optimal prediction theory and a generalized change-of-resolution Kalman filter algorithm in Section 3. In Section 4, we establish statistical inference via ML and REML and their distributional properties. In Section 5, we illustrate the theory and methods by a simulation study.

2. Multivariate Multiresolution Tree-Structured Spatial Linear Model

2.1. Model specification and assumptions

For the m response variables at each node of a multiresolution tree structure, we use a measurement error model:

$$\boldsymbol{z}_{j,k} = \boldsymbol{y}_{j,k} + \boldsymbol{\epsilon}_{j,k}, \quad k = 1, \ldots, N_j, \quad j = 1, \ldots, J, \tag{1}$$

where $\boldsymbol{z}_{j,k} = (z_{jk1}, \ldots, z_{jkm})$ is an m-dimensional row vector of the response variables, $\boldsymbol{y}_{j,k} = (y_{jk1}, \ldots, y_{jkm})$ is an m-dimensional row vector of the latent processes, and $\boldsymbol{\epsilon}_{j,k} = (\epsilon_{jk1}, \ldots, \epsilon_{jkm})$ is an m-dimensional row vector of the measurement errors that captures exogenous variability independent of $\boldsymbol{y}_{j,k}$, for the kth node on the jth resolution; $k = 1, \ldots, N_j$, $j = 1, \ldots, J$. We further assume that the measurement errors $\{\boldsymbol{\epsilon}'_{j,k}\}$ are independent and follow a multivariate normal distribution:

$$\boldsymbol{\epsilon}'_{j,k} \sim N(\mathbf{0}_m, \boldsymbol{\Phi}_j), \quad k = 1, \ldots, N_j, \quad j = 1, \ldots, J, \tag{2}$$

with mean $\mathbf{0}_m = (0, \ldots, 0)'$ and variance matrix $\boldsymbol{\Phi}_j = \text{diag}\{\phi_{j1}, \ldots, \phi_{jm}\}$ where $\phi_{ji} > 0, i = 1, \ldots, m$. For the latent process, we assume a linear regression mean structure:

$$\boldsymbol{y}_{j,k} = \boldsymbol{x}'_{j,k}\boldsymbol{\beta} + \boldsymbol{u}_{j,k}, \quad k = 1, \ldots, N_j, \quad j = 1, \ldots, J, \tag{3}$$

where $\boldsymbol{x}_{j,k} \in I\!R^p$ is a p-dimensional column vector of covariates for each response variable of $\boldsymbol{z}_{j,k}$, $\boldsymbol{\beta}$ is a $p \times m$ matrix of regression coefficients and $\boldsymbol{u}_{j,k} = (u_{jk1}, \ldots, u_{jkm})$ is an m-dimensional row vector of the residual process. Further, we assume that the tree structure is homogeneous such that within a given resolution, the number of children for each node is the same. We model the residual process $\{\boldsymbol{u}_{j,k}\}$ by a multiresolution tree structure:

$$\begin{aligned}&\boldsymbol{u}'_{1,k} \sim N(\mathbf{0}_m, \boldsymbol{\Sigma}_1), k = 1, \ldots, N_1,\\ &\boldsymbol{u}_{ch(j,k)} = \mathbf{1}_{n_j}\boldsymbol{u}_{j,k} + \boldsymbol{\omega}_{ch(j,k)}, k = 1, \ldots, N_j, j = 1, \ldots, J-1,\\ &\boldsymbol{\omega}'_{j,k} \sim N(\mathbf{0}_m, \boldsymbol{\Sigma}_j), \quad k = 1, \ldots, N_j, \quad j = 2, \ldots, J,\end{aligned} \tag{4}$$

where $\boldsymbol{u}_{ch(j,k)} \equiv [\boldsymbol{u}'_{ch(j,k,1)}, \ldots, \boldsymbol{u}'_{ch(j,k,n_j)}]'$ is an $n_j \times m$ matrix that denotes the n_j children of $\boldsymbol{u}_{j,k}$, $ch(j,k,i)$ is the ith child node of (j,k), $\mathbf{1}_{n_j} \equiv (1, \ldots, 1)'$ and $\boldsymbol{\omega}_{ch(j,k)} \equiv [\boldsymbol{\omega}'_{ch(j,k,1)}, \ldots, \boldsymbol{\omega}'_{ch(j,k,n_j)}]'$ is an $n_j \times m$ matrix that denotes the n_j error terms of $ch(j,k)$ and captures random fluctuations independent of $\boldsymbol{u}_{j,k}$. More specifically, the child nodes of (j,k) are $\{(j+1,(k-1)n_j+1), \ldots, (j+1,(k-1)n_j+n_j)\}$ and hence the ith child node is $ch(j,k,i) \equiv (j+1,(k-1)n_j+i)$ for $i = 1, \ldots, n_j$. Here $\boldsymbol{\Sigma}_1$ is an $m \times m$ variance matrix that captures the covariance among the m residuals in $\boldsymbol{u}_{1,k}$ and $\boldsymbol{\Sigma}_j$ is an $m \times m$ variance matrix that captures the covariance among the m error terms in $\boldsymbol{w}_{j,k}$.

For generality, we assume a flexible correlation structure among the children $\boldsymbol{u}_{ch(j,k)}$. Let $\boldsymbol{u}_1 = [\boldsymbol{u}'_{1,1}, \ldots, \boldsymbol{u}'_{1,N_1}]'$ denote the collection of the residual process on the coarsest resolution. Then we model $\boldsymbol{u}_1$ by an $N_1 \times m$ random matrix that follows a normal distribution (see, e.g., Appendix C, Lauritzen (1996)):

$$\boldsymbol{u}_1 \sim N_{N_1 \times m}(\mathbf{0}_{N_1 \times m}, \boldsymbol{H}_1 \otimes \boldsymbol{\Sigma}_1), \tag{5}$$

where $\mathbf{0}_{N_1 \times m}$ is an $N_1 \times m$ matrix of zeros, $\boldsymbol{H}_1$ is an $N_1 \times N_1$ correlation matrix that captures the correlation among the root nodes and $\otimes$ denotes the Kronecker product. We further model the error term $\boldsymbol{\omega}_{ch(j,k)}$ by an $n_j \times m$ random matrix such that

$$\boldsymbol{\omega}_{ch(j,k)} \sim N_{n_j \times m}(\mathbf{0}_{n_j \times m}, \boldsymbol{H}_{j+1} \otimes \boldsymbol{\Sigma}_{j+1}),\ k=1, \ldots, N_j,\ j=1, \ldots, J-1, \tag{6}$$

where $\boldsymbol{H}_{j+1}$ is an $n_j \times n_j$ correlation matrix that captures the correlation among the child nodes $ch(j,k)$.

2.2. Alternative model specification via vectorization

For notational convenience, we proceed to vectorize the individual scalar nodes in the multiresolution tree structure. Let $(0,1)$ denote an imaginary node on the imaginary 0th resolution, which has the N_1 root nodes as its child nodes. Then $N_0 \equiv 1$, $n_0 \equiv N_1$, and $ch(0,1) \equiv \{(1,1),\dots,(1,N_1)\}$. For the jth resolution, let the vector node $\{j,k\} \equiv \{(j,(k-1)n_{j-1}+1),\dots,(j,(k-1)n_{j-1}+n_{j-1})\}$ denote the group of nodes that share a common parent, where $k=1,\dots,N_{j-1}, j=1,\dots,J$. In fact, the common parent of $\{j,k\}$ is the scalar node $(j-1,k)$ on the $(j-1)$th resolution (i.e $\{j,k\} \equiv ch(j-1,k)$). We also define the parent vector node of $\{j,k\}$, $pa\{j,k\}$, to be the vector node that contains the scalar node $(j-1,k)$ (parent of $\{j,k\}$) on the $(j-1)$th resolution. Since the number of vector nodes on the jth resolution is the number of (scalar) parent nodes on the $(j-1)$th resolution (i.e., N_{j-1}), we have $k=1,\dots,N_{j-1}$ for $\{j,k\}$.

Now, we define a vectorization operator $\rightarrow$ such that for an $n \times m$ matrix $\boldsymbol{A} = [a_{ij}], i=1,\dots,n, j=1,\dots,m$, $\mathrm{vec}(\boldsymbol{A}) = \vec{\boldsymbol{A}} \equiv (a_{11},\dots,a_{1m},\dots,a_{n1},\dots,a_{nm})'$. That is $\boldsymbol{A}$ is vectorized by row to form an nm-dimensional column vector (see, e.g., Chapter 16.2, Harville (1997)). Now let

$$\begin{aligned} &\boldsymbol{Z}_{j,k} \equiv \vec{\boldsymbol{z}}_{ch(j-1,k)}, \quad \boldsymbol{Y}_{j,k} \equiv \vec{\boldsymbol{y}}_{ch(j-1,k)}, \quad \boldsymbol{U}_{j,k} \equiv \vec{\boldsymbol{u}}_{ch(j-1,k)}, \\ &\boldsymbol{W}_{j,k} \equiv \vec{\boldsymbol{\omega}}_{ch(j-1,k)}, \quad \boldsymbol{e}_{j,k} \equiv \vec{\boldsymbol{\epsilon}}_{ch(j-1,k)}, \quad \boldsymbol{X}_{j,k} \equiv \boldsymbol{x}_{ch(j-1,k)} \otimes \boldsymbol{I}_m, \end{aligned} \tag{7}$$

denote the vector of response variables, the vector of the latent processes (original and residual), the vector of the error terms, the vector of the measurement errors, and the matrix of the covariates, all of which correspond to the vector node $\{j,k\}$. By the fact that for matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$,

$$\mathrm{vec}(\boldsymbol{A}+\boldsymbol{B}) = \vec{\boldsymbol{A}} + \vec{\boldsymbol{B}} \ \text{ and } \mathrm{vec}(\boldsymbol{ABC}') = (\boldsymbol{A} \otimes \boldsymbol{C})\vec{\boldsymbol{B}} \tag{8}$$

((B.5), Lauritzen (1996)), the MMTSLM can be rewritten in vector form as follows. The measurement error model is

$$\boldsymbol{Z}_{j,k} = \boldsymbol{Y}_{j,k} + \boldsymbol{e}_{j,k}, \ \boldsymbol{e}_{j,k} \sim N(\boldsymbol{0}_{n_{j-1}m}, \boldsymbol{I}_{n_{j-1}} \otimes \boldsymbol{\Phi}_j), k=1,\dots,N_{j-1}, j=1,\dots,J. \tag{9}$$

The latent process model is

$$\boldsymbol{Y}_{j,k} = \boldsymbol{X}_{j,k}\boldsymbol{B} + \boldsymbol{U}_{j,k}, \tag{10}$$

and the residual process model is

$$\boldsymbol{U}_{1,1} \sim N(\boldsymbol{0}_{n_0 m}, \boldsymbol{H}_1 \otimes \boldsymbol{\Sigma}_1), \tag{11}$$

$$\boldsymbol{U}_{j,k} = \boldsymbol{A}_{j,k}\boldsymbol{U}_{pa\{j,k\}} + \boldsymbol{W}_{j,k}, \ \boldsymbol{W}_{j,k} \sim N(\boldsymbol{0}_{n_{j-1}m}, \boldsymbol{H}_j \otimes \boldsymbol{\Sigma}_j), \quad k=1,\dots,N_{j-1}, \quad j=2,\dots,J, \tag{12}$$

where $\boldsymbol{B} = \vec{\boldsymbol{\beta}}$, $\boldsymbol{A}_{j,k} = \boldsymbol{D}_{j,k} \otimes \boldsymbol{I}_m$ and $\boldsymbol{D}_{j,k}$ is an $n_{j-1} \times n_{j-2}$ matrix consisting of $\boldsymbol{1}_{n_{j-1}}$ in the ith column and $\boldsymbol{0}_{n_{j-1}}$ in the other columns if the scalar parent node $(j-1,k)$ is the ith node within the vector node $pa\{j,k\}$. Here $\boldsymbol{W}_{j,k}$ are mutually independent and are independent of $\boldsymbol{U}_{pa\{j,k\}}$. Furthermore $\boldsymbol{W}_{j,k}$ represents the departure of $\boldsymbol{U}_{j,k}$ from its parent $\boldsymbol{U}_{pa\{j,k\}}$.

2.3. Model properties

For the MMTSLM defined in (9)−(12), we now explore the mean, variance and covariance structure of the variables in the model. For this purpose, we denote the N_{j-1} variables on a given jth resolution by $\boldsymbol{Z}_{j\cdot} \equiv (\boldsymbol{Z}'_{j,1}, \ldots, \boldsymbol{Z}'_{j,N_{j-1}})'$, $\boldsymbol{Y}_{j\cdot} \equiv (\boldsymbol{Y}'_{j,1}, \ldots, \boldsymbol{Y}'_{j,N_{j-1}})'$, $\boldsymbol{U}_{j\cdot} \equiv (\boldsymbol{U}'_{j,1}, \ldots, \boldsymbol{U}'_{j,N_{j-1}})'$ and $\boldsymbol{W}_{j\cdot} \equiv (\boldsymbol{W}'_{j,1}, \ldots, \boldsymbol{W}'_{j,N_{j-1}})'$ for the response variables, the latent processes (original and residual), and the error terms, respectively. Also let $\boldsymbol{X}_{j\cdot} \equiv [\boldsymbol{X}'_{j,1}, \ldots, \boldsymbol{X}'_{j,N_{j-1}}]'$ denote the covariates on the jth resolution.

By (11) and (12), $\boldsymbol{U}_{j+1\cdot} = (\boldsymbol{I}_{N_j} \otimes \boldsymbol{1}_{n_j} \otimes \boldsymbol{I}_m)\boldsymbol{U}_{j\cdot} + \boldsymbol{W}_{j+1\cdot}$, $E(\boldsymbol{U}_{j\cdot}) = \boldsymbol{0}_{N_j m}$, $\mathrm{Var}\,(\boldsymbol{U}_{1\cdot}) = \boldsymbol{H}_1 \otimes \boldsymbol{\Sigma}_1$, and $\mathrm{Var}\,(\boldsymbol{U}_{j+1\cdot}) = (\boldsymbol{I}_{N_j} \otimes \boldsymbol{1}_{n_j} \otimes \boldsymbol{I}_m)\mathrm{Var}\,(\boldsymbol{U}_{j\cdot})(\boldsymbol{I}_{N_j} \otimes \boldsymbol{1}_{n_j} \otimes \boldsymbol{I}_m)' + \mathrm{Var}\,(\boldsymbol{W}_{j+1\cdot})$, where $\mathrm{Var}\,(\boldsymbol{W}_{j+1\cdot}) = \boldsymbol{I}_{N_j} \otimes \boldsymbol{H}_{j+1} \otimes \boldsymbol{\Sigma}_{j+1}$; $j = 1, \ldots, J-1$. A simplification of $\mathrm{Var}\,(\boldsymbol{U}_{j\cdot})$ gives

$$\begin{aligned}\mathrm{Var}\,(\boldsymbol{U}_{j\cdot}) = {} & \boldsymbol{I}_{N_{j-1}} \otimes \boldsymbol{H}_j \otimes \boldsymbol{\Sigma}_j + \boldsymbol{I}_{N_{j-2}} \otimes \boldsymbol{H}_{j-1} \otimes (\boldsymbol{1}_{n_{j-1}}\boldsymbol{1}'_{n_{j-1}}) \otimes \boldsymbol{\Sigma}_{j-1} \\ & + \cdots + \boldsymbol{I}_{N_1} \otimes \boldsymbol{H}_2 \otimes (\boldsymbol{1}_{n_2 \cdots n_{j-1}}\boldsymbol{1}'_{n_2 \cdots n_{j-1}}) \otimes \boldsymbol{\Sigma}_2 \\ & + \boldsymbol{I}_{N_0} \otimes \boldsymbol{H}_1 \otimes (\boldsymbol{1}_{n_1 \cdots n_{j-1}}\boldsymbol{1}'_{n_1 \cdots n_{j-1}}) \otimes \boldsymbol{\Sigma}_1, j = 1, \ldots, J. \qquad (13)\end{aligned}$$

Further, $\mathrm{Cov}\,(\boldsymbol{U}_{j\cdot}, \boldsymbol{U}_{j'\cdot}) = \mathrm{Var}\,(\boldsymbol{U}_{j\cdot})(\boldsymbol{I}_{N_j} \otimes \boldsymbol{1}'_{n_j \cdots n_{j'-1}} \otimes \boldsymbol{I}_m)$, $1 \leq j < j' \leq J$. The mean, variance, and covariance f $\boldsymbol{Y}_{j\cdot}$ and $\boldsymbol{Z}_{j\cdot}$ follow directly.

The MMTSLM presented here is suitable for modeling observations that are available at different resolutions, such as those that are collected from multiple sources. However, in practice, there are usually only observations on one resolution, such as those that are collected from a single source. Here we focus on the single-source case of MMTSLM, even though the MMTSLM is suitable for the multi-source cases. We let $\boldsymbol{\theta} \equiv (\boldsymbol{B}', \boldsymbol{\eta}', \boldsymbol{\zeta}')'$ denote the model parameters for the MMTSLM, with the regression coefficients $\boldsymbol{B}$, the parameters $\boldsymbol{\eta}$ for the among-node correlation matrices $\{\boldsymbol{H}_j : j = 1, \ldots, J\}$, and the parameters $\boldsymbol{\zeta}$ for the within-node variance matrices $\{\boldsymbol{\Sigma}_j : j = 1, \ldots, J\}$. To ensure identifiability, the measurement error variance $\boldsymbol{\Phi}_J$ is assumed to be known for the MMTSLM and can oftentimes be estimated from external data (see, e.g., Zhu and Yue (2005)). Note that when the measurement error variances increase, the predicted values of $\{\boldsymbol{Y}_{j,k}\}$ tend to be smoother; whereas when the measurement error variances decrease, the predicted values of $\{\boldsymbol{Y}_{j,k}\}$ tend to be closer to the original data.

2.4. Mass balance property

The mass-balance property introduced by Huang, Cressie and Gabrosek (2002) and featured in Zhu and Yue (2005) can be readily included in the MMTSLM defined in (1)−(6) as a special case. A multiresolution tree structure is mass-balanced if the average of all the children's values is equal to their parent's value. That is, $n_j^{-1}(\mathbf{1}'_{n_j}\boldsymbol{y}_{ch(j,k)}) = \boldsymbol{y}_{j,k}$, where $\boldsymbol{y}_{ch(j,k)} \equiv [\boldsymbol{y}'_{ch(j,k,1)}, \ldots, \boldsymbol{y}'_{ch(j,k,n_j)}]'$ is an $n_j \times m$ matrix that denotes the children processes of $\boldsymbol{y}_{j,k}$, $j = 1, \ldots, J-1$. The mass-balance property ensures that the latent process is consistent when aggregated across resolutions, which is a condition that most physical processes satisfy. The following conditions are sufficient for an MMTSLM to have mass balance:

$$n_j^{-1}\left(\mathbf{1}'_{n_j}\boldsymbol{u}_{ch(j,k)}\right) = \boldsymbol{u}_{j,k} \quad \text{and} \quad n_j^{-1}\left(\mathbf{1}'_{n_j}\boldsymbol{x}_{ch(j,k)}\right) = \boldsymbol{x}'_{j,k}, \qquad k = 1, \ldots, N_j, \qquad j = 1, \ldots, J-1, \tag{14}$$

where the rows of the matrix $\boldsymbol{x}_{ch(j,k)} \equiv [\boldsymbol{x}_{ch(j,k,1)}, \ldots, \boldsymbol{x}_{ch(j,k,n_j)}]'$ correspond to the children covariates of $\boldsymbol{x}_{j,k}$. It follows from (4) and (14) that $\mathbf{1}'_{n_j}\boldsymbol{\omega}_{ch(j,k)} = \mathbf{0}'_m$. If we assume that $\boldsymbol{H}_j$ is compound symmetric, then we can obtain $\boldsymbol{H}_1 = \boldsymbol{I}_{N_1}$ and $\boldsymbol{H}_{j+1} = (n_j/(n_j-1))(\boldsymbol{I}_{n_j} - \mathbf{1}_{n_j}\mathbf{1}'_{n_j}/n_j)$, $j = 1, \ldots, J-1$, where $\boldsymbol{I}_{N_1}$ is the $N_1 \times N_1$ identity matrix. Further, we can simplify (13) to

$$\begin{aligned}
&\mathrm{Var}\,(\boldsymbol{U}_{j\cdot}) \\
&= \frac{n_{j-1}}{n_{j-1}-1}\boldsymbol{I}_{N_j} \otimes \boldsymbol{\Sigma}_j + \boldsymbol{I}_{N_{j-1}} \otimes (\mathbf{1}_{n_{j-1}}\mathbf{1}'_{n_{j-1}}) \otimes (\frac{n_{j-2}}{n_{j-2}-1}\boldsymbol{\Sigma}_{j-1} - \frac{1}{n_{j-1}-1}\boldsymbol{\Sigma}_j) \\
&\quad + \cdots + \boldsymbol{I}_{N_2} \otimes (\mathbf{1}_{n_2\cdots n_{j-1}}\mathbf{1}'_{n_2\cdots n_{j-1}}) \otimes (\frac{n_1}{n_1-1}\boldsymbol{\Sigma}_2 - \frac{1}{n_2-1}\boldsymbol{\Sigma}_3) \\
&\quad + \boldsymbol{I}_{N_1} \otimes (\mathbf{1}_{n_1\cdots n_{j-1}}\mathbf{1}'_{n_1\cdots n_{j-1}}) \otimes (\boldsymbol{\Sigma}_1 - \frac{1}{n_1-1}\boldsymbol{\Sigma}_2), j = 1, \ldots, J,
\end{aligned} \tag{15}$$

and $\mathrm{Cov}\,(\boldsymbol{U}_{j\cdot}, \boldsymbol{U}_{j'\cdot}) = \mathrm{Var}\,(\boldsymbol{U}_{j\cdot})(\boldsymbol{I}_{N_j} \otimes \mathbf{1}'_{n_j\cdots n_{j'-1}} \otimes \boldsymbol{I}_m)$, $1 \le j < j' \le J$. When the response variable is univariate with $m = 1$, (15) reduces to (9) of Zhu and Yue (2005).

3. Generalized Change-of-Resolution Kalman Filter

In this section, we consider optimal prediction of the latent process in the MMTSLM (9)−(12). First we consider prediction of the residual process $\boldsymbol{U}_{j,k}$. It is well-known that the best linear unbiased predictor (BLUP) is the conditional mean $E(\boldsymbol{U}_{j,k}|\boldsymbol{Z})$ of the latent process $\boldsymbol{U}_{j,k}$ given the observations $\boldsymbol{Z}$. For normal distributions, the BLUP is also the best unbiased predictor (see, e.g., Harvey (1989, Sec. 3.2.3)). Computing $E(\boldsymbol{U}_{j,k}|\boldsymbol{Z})$ usually involves operations of order $\mathcal{O}(n^3)$ and can be computationally inefficient for large n. For a univariate response variable, a change-of-resolution Kalman-filter algorithm has been developed to obtain the BLUP. It exploits the multiresolution tree structure and involves operations of only order $\mathcal{O}(n)$ (see, e.g., Chou, Willsky and Nikoukhah

(1994), Huang, Cressie and Gabrosek (2002) and Zhu and Yue (2005)). However, the existing BLUP theory and change-of-resolution Kalman-filter algorithm for univariate MTSLM assume that any variance matrix involved is nonsingular and thus invertible. For MMTSLM, however, a variance matrix can be singular, and for a normal distribution it is said to be singular normal (see, e.g., Searl (1997, Chap. 2.7)). An example would be when the m latent variables in $\boldsymbol{U}_{j,k}$ are subject to linear constraints. One approach to deal with the singular normal problem is to reduce the dimension of the residual process by transforming the residual process to a new variable that has a nonsingular variance (see, e.g., Appendix D, Luettgen (1993)). The dimension-reduction approach can be cumbersome in the case of multiple response variables, because different problems may require different ways of reducing the dimension. Thus it is unclear how to extend the BLUP theory and the change-of-resolution Kalman-filter algorithm from univariate MTSLM to MMTSLM. Here we develop a general theory of optimal prediction, based on which we derive a generalized change-of-resolution Kalman-filter algorithm. Our approach bears similarity to Luettgen (1993) and Luettgen and Willsky (1995), who also allow for singular variance matrices, but our approach is more general because we allow for multiple response variables, flexible mean and variance structures, and missing observations. Further, our derivation of the change-of-resolution Kalman-filter algorithm is based on general theory of optimal prediction and does not assume normal distributions, which can be of independent interest in the Kalman filter literature. Related work includes Jørgensen, Lundbye-Christensen, Song and Sun (1999), which considered optimal prediction theory for longitudinal data.

3.1. General optimal prediction theory

Following the notation in Chapter 2 of Brockwell and Davis (1991), we consider the space $L^2(\Omega, \mathcal{F}, P)$, which is the collection of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ with finite second moments. Here we abbreviate $L^2(\Omega, \mathcal{F}, P)$ to L^2. For $y_1, y_2, y \in L^2$, define an inner product $\langle y_1, y_2 \rangle \equiv E(y_1 y_2)$ and a norm (or distance) as $\|y\| \equiv \langle y, y^{1/2} \rangle = \sqrt{E(y^2)}$. Equipped with this inner product, L^2 is a real Hilbert space (Brockwell and Davis (1991, Example 2.2.2)). For $z_1, \ldots, z_n \in L^2$, define a closed subspace of L^2 as $\overline{sp}\{1, z_1, \ldots, z_n\} \equiv \{\mu + \beta_1 z_1 + \cdots + \beta_n z_n : \ \mu \in IR, \ \beta_i \in IR, \ i = 1, \ldots, n\}$. For $y \in L^2$, the optimal linear predictor of y in term of $\{z_i \in L^2 : \ i = 1, \ldots, n\}$ is defined as the element in $\overline{sp}\{1, z_1, \ldots, z_n\}$ that has the smallest distance from y. Theorem 2.3.3 of Brockwell and Davis (1991) establishes the existence and uniqueness of the optimal linear predictor. Now, we extend the definition of optimal linear predictor to multivariate random vector $\boldsymbol{Y} = (y_1, \ldots, y_m)'$. We define the space L_m^2 as the collection of m-dimensional random vectors whose elements belong to

L^2. For $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \boldsymbol{Y} \in L^2_m$, define an inner product as $\langle \boldsymbol{Y}_1, \boldsymbol{Y}_2 \rangle \equiv E(\boldsymbol{Y}_1' \boldsymbol{Y}_2)$ and a norm (or distance) as $\|\boldsymbol{Y}\| \equiv \langle \boldsymbol{Y}, \boldsymbol{Y}^{12} \rangle = \sqrt{E(\boldsymbol{Y}'\boldsymbol{Y})}$. It follows that L^2_m is also a real Hilbert space. For $\boldsymbol{Z} = (z_1, \ldots, z_n)' \in L^2_n$, define a closed subspace $\overline{sp}\{\boldsymbol{Z}\}^m$ of L^2_m as the collection of m-dimensional random vectors whose elements belong to $\overline{sp}\{1, z_1, \ldots, z_n\}$. Then the optimal linear predictor of $\boldsymbol{Y}$ given $\boldsymbol{Z}$ is defined as the element in $\overline{sp}\{\boldsymbol{Z}\}^m$ that has the smallest distance from $\boldsymbol{Y}$. Adopting notation from Jørgensen, Lundbye-Christensen, Song and Sun (1999), we denote the optimal linear predictor of $\boldsymbol{Y}$ given $\boldsymbol{Z}$ and the corresponding mean-squared prediction error (MSPE) as $\boldsymbol{Y}|\boldsymbol{Z} \sim [\boldsymbol{m}_{Y|Z}, \boldsymbol{C}_{Y|Z}]$, where $\boldsymbol{m}_{Y|Z}$ denotes the optimal linear predictor of $\boldsymbol{Y}$ given $\boldsymbol{Z}$ and $\boldsymbol{C}_{Y|Z} \equiv E[(\boldsymbol{Y} - \boldsymbol{m}_{Y|Z})(\boldsymbol{Y} - \boldsymbol{m}_{Y|Z})'] = \text{Var}\,(\boldsymbol{Y} - \boldsymbol{m}_{Y|Z})$ denotes the corresponding MSPE. For ease of notation, we sometimes write $\boldsymbol{m}(\boldsymbol{Y}|\boldsymbol{Z}) \equiv \boldsymbol{m}_{Y|Z}$ and $\boldsymbol{C}(\boldsymbol{Y}|\boldsymbol{Z}) \equiv \boldsymbol{C}_{Y|Z}$.

For $\boldsymbol{Y} \in L^2_m$, we use $\boldsymbol{Y} \sim [\boldsymbol{\mu}_{\boldsymbol{Y}}, \boldsymbol{\Sigma}_{\boldsymbol{YY}}]$ to denote that the mean of $\boldsymbol{Y}$ is $\boldsymbol{\mu}_{\boldsymbol{Y}}$ and the variance is $\boldsymbol{\Sigma}_{\boldsymbol{YY}}$. For $\binom{\boldsymbol{Y}}{\boldsymbol{Z}} \sim \left[\binom{\boldsymbol{\mu}_Y}{\boldsymbol{\mu}_Z}, \begin{pmatrix} \boldsymbol{\Sigma}_{YY} & \boldsymbol{\Sigma}_{YZ} \\ \boldsymbol{\Sigma}_{ZY} & \boldsymbol{\Sigma}_{ZZ} \end{pmatrix}\right]$, we establish the following results about optimal linear prediction.

Theorem 1. *For $\boldsymbol{Y} = (y_1, \ldots, y_m)' \in L^2_m$ and $\boldsymbol{Z} = (z_1, \ldots, z_n)' \in L^2_n$, the optimal linear predictor of $\boldsymbol{Y}$ given $\boldsymbol{Z}$ exists and is unique with $E(\boldsymbol{m}_{Y|Z}) = \boldsymbol{\mu}_Y$ and* $\text{Cov}\,(\boldsymbol{Z}, \boldsymbol{Y} - \boldsymbol{m}_{Y|Z}) = \boldsymbol{0}_{n \times m}$.

Theorem 2. *For $\boldsymbol{Y} = (y_1, \ldots, y_m)' \in L^2_m$ and $\boldsymbol{Z} = (z_1, \ldots, z_n)' \in L^2_n$, the optimal linear predictor $\boldsymbol{Y}|\boldsymbol{Z} \sim [\boldsymbol{m}_{Y|Z}, \boldsymbol{C}_{Y|Z}]$ is given by $\boldsymbol{m}_{Y|Z} = \boldsymbol{\mu}_Y + \boldsymbol{\Sigma}_{YZ}\boldsymbol{\Sigma}^+_{ZZ}(\boldsymbol{Z} - \boldsymbol{\mu}_Z)$ and $\boldsymbol{C}_{Y|Z} = \boldsymbol{\Sigma}_{YY} - \boldsymbol{\Sigma}_{YZ}\boldsymbol{\Sigma}^+_{ZZ}\boldsymbol{\Sigma}_{ZY}$ where $\boldsymbol{\Sigma}^+_{ZZ}$ is the unique Moore-Penrose pseudo inverse of $\boldsymbol{\Sigma}_{ZZ}$. If $\boldsymbol{Y}$ and $\boldsymbol{Z}$ have normal distributions, then $\boldsymbol{m}_{Y|Z} = E(\boldsymbol{Y}|\boldsymbol{Z})$ and $\boldsymbol{C}_{Y|Z} = \text{Var}\,(\boldsymbol{Y}|\boldsymbol{Z})$.*

The proof of Theorems 1 and 2 are given in Appendix I. Theorem 1 establishes the existence and uniqueness of the optimal linear predictor whereas Theorem 2 gives the explicit forms of the optimal linear predictor and the corresponding MSPE. Although we restrict our attention to random variables with finite second moments, the variance matrix does not need to be nonsingular. Further, the results apply to but are not restricted to the case of normal distributions. Finally the theorems provide us with an elegant way of deriving a generalized change-of-resolution Kalman filter algorithm when the variance matrices are not necessarily nonsingular.

3.2. Generalized change-of-resolution Kalman filter

Using Theorems 1 and 2, we derive here a generalized change-of-resolution Kalman filter algorithm based on the MMTSLM (9)$-$(12), which provides an efficient way of computing the optimal linear predictor $\{\hat{\boldsymbol{U}}_{j,k} : k = 1, \ldots, N_{j-1}, j = 1, \ldots, J\}$ and involves two steps: a high-to-low-resolution filtering step followed by a low-to-high-resolution smoothing step. In the filtering step, the algorithm moves from finer resolutions to coarser resolutions, recursively computing the optimal predictor of the latent process based on the data on relevant higher resolutions. Once the coarsest resolution is reached, the algorithm goes back from coarser resolutions to finer resolutions, recursively computing the optimal predictor of $\boldsymbol{U}_{j,k}$ on each resolution based on all the data. In the final step of the recursion, the optimal prediction of $\{\boldsymbol{U}_{j,k}\}$, given all the data, is achieved.

Denote $\{j',k'\} \prec \{j,k\}$ if $\{j',k'\}$ is a descendant vector node of $\{j,k\}$ and let $\gamma_{j,k} \equiv \mathcal{I}\{\boldsymbol{Z}_{j,k}$ is observed$\}$ denote whether all the observations at $\{j,k\}$ are observed. In the high-to-low-resolution filtering step, we start with the finest resolution J and compute, for $k = 1, \ldots, N_{J-1}$;

$$\hat{\boldsymbol{U}}_{J,k|J,k} \equiv \boldsymbol{m}(\boldsymbol{U}_{J,k}|\boldsymbol{Z}_{de\{J,k\}}) = \gamma_{J,k}\boldsymbol{V}_{J,k}(\boldsymbol{V}_{J,k}+\boldsymbol{I}_{n_{J-1}}\otimes\boldsymbol{\Phi}_J)^{-1}(\boldsymbol{Z}_{J,k}-\boldsymbol{X}_{J,k}\boldsymbol{B}), \quad (16)$$

$$\hat{\boldsymbol{V}}_{J,k|J,k} \equiv \boldsymbol{C}(\boldsymbol{U}_{J,k}|\boldsymbol{Z}_{de\{J,k\}}) = \boldsymbol{V}_{J,k} - \gamma_{J,k}\boldsymbol{V}_{J,k}(\boldsymbol{V}_{J,k}+\boldsymbol{I}_{n_{J-1}}\otimes\boldsymbol{\Phi}_J)^{-1}\boldsymbol{V}_{J,k}, \quad (17)$$

where $\boldsymbol{Z}_{de\{J,k\}} \equiv \{\gamma_{J,k}\boldsymbol{Z}_{J,k}\}$, and $\boldsymbol{V}_{J,k} \equiv \mathrm{Var}\,(\boldsymbol{U}_{J,k}) = (\boldsymbol{1}_{n_{J-1}}\boldsymbol{1}'_{n_{J-1}}) \otimes (\sum_{j'=1}^{J-1}\boldsymbol{\Sigma}_{j'}) + \boldsymbol{H}_J \otimes \boldsymbol{\Sigma}_J$ can be obtained from $\mathrm{Var}\,(\boldsymbol{U}_{J\cdot})$ in (13). As we move from the resolution $j = J-1$ to the coarsest resolution $j = 1$, we compute, for a given vector node $\{j,k\}$,

$$\hat{\boldsymbol{U}}_{j,k|ch\{j,k,i\}} \equiv \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}_{de\{ch\{j,k,i\}\}}) = \boldsymbol{B}_{ch\{j,k,i\}}\hat{\boldsymbol{U}}_{ch\{j,k,i\}|ch\{j,k,i\}}, \quad (18)$$

$$\begin{aligned}\hat{\boldsymbol{V}}_{j,k|ch\{j,k,i\}} &\equiv \boldsymbol{C}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}_{de\{ch\{j,k,i\}\}}) \\ &= \boldsymbol{B}_{ch\{j,k,i\}}\hat{\boldsymbol{V}}_{ch\{j,k,i\}|ch\{j,k,i\}}\boldsymbol{B}'_{ch\{j,k,i\}} + \boldsymbol{R}_{ch\{j,k,i\}}, \quad (19)\end{aligned}$$

where $\boldsymbol{Z}_{de\{j,k\}} \equiv \{\boldsymbol{Z}_{j',k'} : \gamma_{j',k'} = 1, \{j',k'\} \prec \{j,k\}\}$ denotes the descendants of $\boldsymbol{Z}_{j,k}$ including $\boldsymbol{Z}_{j,k}$, $ch\{j,k,i\}$ denotes the ith vector child node of $\{j,k\}$, $\boldsymbol{B}_{ch\{j,k,i\}} \equiv \boldsymbol{V}_{j,k}\boldsymbol{A}'_{ch\{j,k,i\}}\boldsymbol{V}^{+}_{ch\{j,k,i\}}$, $\boldsymbol{R}_{ch\{j,k,i\}} \equiv \boldsymbol{V}_{j,k} - \boldsymbol{V}_{j,k}\boldsymbol{A}'_{ch\{j,k,i\}}\boldsymbol{V}^{+}_{ch\{j,k,i\}}\boldsymbol{A}_{ch\{j,k,i\}}\boldsymbol{V}_{j,k}$, $i = 1, \ldots, n_{j-1}$, and $+$ denotes the Moore-Penrose pseudo inverse. Here $\boldsymbol{V}_{j,k} \equiv \mathrm{Var}\,(\boldsymbol{U}_{j,k}) = (\boldsymbol{1}_{n_{j-1}}\boldsymbol{1}'_{n_{j-1}}) \otimes (\sum_{j'=1}^{j-1}\boldsymbol{\Sigma}_{j'}) + \boldsymbol{H}_j \otimes \boldsymbol{\Sigma}_j$ and $\boldsymbol{V}_{ch\{j,k,i\}} \equiv \mathrm{Var}\,(\boldsymbol{U}_{ch\{j,k,i\}}) = (\boldsymbol{1}_{n_j}\boldsymbol{1}'_{n_j}) \otimes (\sum_{j'=1}^{j}\boldsymbol{\Sigma}_{j'}) + \boldsymbol{H}_{j+1} \otimes \boldsymbol{\Sigma}_{j+1}$ can be obtained from $\mathrm{Var}\,(\boldsymbol{U}_{j\cdot})$ and $\mathrm{Var}\,(\boldsymbol{U}_{j+1\cdot})$ as defined in (13). Further,

$$\hat{\boldsymbol{U}}^{*}_{j,k|j,k} \equiv \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}^{*}_{de\{j,k\}}) = \hat{\boldsymbol{V}}^{*}_{j,k|j,k}\left(\sum_{i=1}^{n_{j-1}}\hat{\boldsymbol{V}}^{+}_{j,k|ch\{j,k,i\}}\hat{\boldsymbol{U}}_{j,k|ch\{j,k,i\}}\right), \quad (20)$$

$$\hat{\boldsymbol{V}}^{*}_{j,k|j,k} \equiv \boldsymbol{C}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}^{*}_{de\{j,k\}}) = \left\{\boldsymbol{V}^{+}_{j,k} + \sum_{i=1}^{n_{j-1}}(\hat{\boldsymbol{V}}^{+}_{j,k|ch\{j,k,i\}} - \boldsymbol{V}^{+}_{j,k})\right\}^{+}, \quad (21)$$

$$\begin{aligned}\hat{\boldsymbol{U}}_{j,k|j,k} &\equiv \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}_{de\{j,k\}})\\ &= \hat{\boldsymbol{V}}_{j,k|j,k}\left\{\gamma_{j,k}(\boldsymbol{I}_{n_{j-1}}\otimes\boldsymbol{\Phi}_j^{-1})(\boldsymbol{Z}_{j,k}-\boldsymbol{X}_{j,k}\boldsymbol{B})+(\hat{\boldsymbol{V}}^*_{j,k|j,k})^+\hat{\boldsymbol{U}}^*_{j,k|j,k}\right\}, \quad (22)\\ \hat{\boldsymbol{V}}_{j,k|j,k} &\equiv \boldsymbol{C}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}_{de\{j,k\}})\\ &= \hat{\boldsymbol{V}}^*_{j,k|j,k}-\gamma_{j,k}\hat{\boldsymbol{V}}^*_{j,k|j,k}(\hat{\boldsymbol{V}}^*_{j,k|j,k}+\boldsymbol{I}_{n_{j-1}}\otimes\boldsymbol{\Phi}_j)^{-1}\hat{\boldsymbol{V}}^*_{j,k|j,k}, \quad (23)\end{aligned}$$

where $\boldsymbol{Z}^*_{de\{j,k\}} \equiv \{\boldsymbol{Z}_{j',k'} : \gamma_{j',k'} = 1, \{j',k'\} \prec \{j,k\}, \{j',k'\} \neq \{j,k\}\}$ denotes the descendants of $\boldsymbol{Z}_{j,k}$ not including $\boldsymbol{Z}_{j,k}$. At the end of the filtering step, the root vector node is reached and hence the BLUP for $\{1,1\}$ is

$$\hat{\boldsymbol{U}}_{1,1} \equiv \boldsymbol{m}(\boldsymbol{U}_{1,1}|\boldsymbol{Z}) = \hat{\boldsymbol{U}}_{1,1|1,1}, \qquad \hat{\boldsymbol{V}}_{1,1} \equiv \boldsymbol{C}(\boldsymbol{U}_{1,1}|\boldsymbol{Z}) = \hat{\boldsymbol{V}}_{1,1|1,1}, \tag{24}$$

where $\boldsymbol{Z} \equiv \{\boldsymbol{Z}_{j,k} : \gamma_{j,k} = 1, k = 1,\ldots,N_{j-1}, j = 1,\ldots,J\}$ consists of all the observations.

In the low-to-high-resolution smoothing step, we move from the coarsest resolution $j = 2$ to the finest resolution $j = J$ and compute for a given node $\{j,k\}$, where $k = 1,\ldots,N_{j-1}$,

$$\hat{\boldsymbol{U}}_{j,k} \equiv \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}) = \hat{\boldsymbol{U}}_{j,k|j,k} + \boldsymbol{J}_{j,k}(\hat{\boldsymbol{U}}_{pa\{j,k\}} - \hat{\boldsymbol{U}}_{pa\{j,k\}|j,k}), \tag{25}$$

$$\hat{\boldsymbol{V}}_{j,k} \equiv \boldsymbol{C}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}) = \hat{\boldsymbol{V}}_{j,k|j,k} + \boldsymbol{J}_{j,k}(\hat{\boldsymbol{V}}_{pa\{j,k\}} - \hat{\boldsymbol{V}}_{pa\{j,k\}|j,k})\boldsymbol{J}'_{j,k}, \tag{26}$$

where $\boldsymbol{J}_{j,k} \equiv \hat{\boldsymbol{V}}_{j,k|j,k}\boldsymbol{B}'_{j,k}\hat{\boldsymbol{V}}^+_{pa\{j,k\}|j,k}$ and $\boldsymbol{B}_{j,k} \equiv \boldsymbol{V}_{pa\{j,k\}}\boldsymbol{A}'_{j,k}\boldsymbol{V}^+_{j,k}$.

In Appendix II, we prove (16)−(26) using Theorems 1 and 2.

3.3. Optimal Prediction of $\{\boldsymbol{Y}_{j,k}\}$

The optimal prediction of the latent processes $\{\boldsymbol{Y}_{j,k}\}$ is achieved in two steps. First we assume that the model parameters $\boldsymbol{\theta} = (\boldsymbol{B}', \boldsymbol{\eta}', \boldsymbol{\zeta}')'$ are known and combine the regression mean and the predicted residual process $\{\hat{\boldsymbol{U}}_{j,k}\}$: $\boldsymbol{m}_{Y_{j,k}|Z;\ \theta} = \boldsymbol{X}_{j,k}\boldsymbol{B} + \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{Z};\ \boldsymbol{\theta})$. Then we plug $\hat{\boldsymbol{B}}$ into the formula $\boldsymbol{m}_{Y_{j,k}|Z;\ \theta}$ to obtain the predictor

$$\hat{\boldsymbol{Y}}_{j,k} = \boldsymbol{m}_{Y_{j,k}|Z;\ \hat{B},\eta,\zeta} = \boldsymbol{X}_{j,k}\hat{\boldsymbol{B}} + \boldsymbol{m}_{U_{j,k}|Z;\ \hat{B},\eta,\zeta}, \tag{27}$$

where $\hat{\boldsymbol{B}} \equiv [\boldsymbol{X}'\mathrm{Var}\,(\boldsymbol{Z})^{-1}\boldsymbol{X}]^{-1}[\boldsymbol{X}'\mathrm{Var}\,(\boldsymbol{Z})^{-1}\boldsymbol{Z}]$ is the generalized least squares (GLS) estimate of $\boldsymbol{B}$. By arguments similar to Harville (1985), the predictor $\hat{\boldsymbol{Y}}_{j,k}$ is the BLUP. The MSPE of the elements in $\hat{\boldsymbol{Y}}_{j,k}$ can be obtained from the diagonal elements of the matrix

$$\boldsymbol{C}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}) + (\boldsymbol{X}_{j,k} - \boldsymbol{V}_{j,k,\cdot,\cdot}\boldsymbol{D}\boldsymbol{X})\ \mathrm{Var}\,(\hat{\boldsymbol{B}})(\boldsymbol{X}_{j,k} - \boldsymbol{V}_{j,k,\cdot,\cdot}\boldsymbol{D}\boldsymbol{X})', \tag{28}$$

where $\boldsymbol{V}_{j,k,\cdot,\cdot} \equiv \mathrm{Cov}\,(\boldsymbol{U}_{j,k}, \boldsymbol{Z})$, $\boldsymbol{D} \equiv (\mathrm{Var}\,(\boldsymbol{Z}))^{-1}$, and $\boldsymbol{X} \equiv [\boldsymbol{X}'_{j,k} : \gamma_{j,k} = 1]'$. Finally we plug the estimates of $\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\zeta}}$ into (27)−(28) to obtain the empirical BLUP and the corresponding empirical MSPE. To compute the MSPE's (28) efficiently, we propose an algorithm based on the generalized change-of-resolution Kalman filter as follows. First, we compute $\boldsymbol{C}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}) = \hat{\boldsymbol{V}}_{j,k}$ by the generalized change-of-resolution Kalman filter. Then we obtain Var$(\hat{\boldsymbol{B}})$ using (34), to be shown in Section 4. To compute $(\boldsymbol{X}_{j,k} - \boldsymbol{V}_{j,k,\cdot,\cdot}\boldsymbol{D}\boldsymbol{X})$, it suffices to compute $\boldsymbol{V}_{j,k,\cdot,\cdot}\boldsymbol{D}\boldsymbol{X}$ efficiently. We treat, for the moment, the covariates in $\boldsymbol{X}$ as observations and process them using the generalized change-of-resolution Kalman filter. More specifically, let $\boldsymbol{X} = [\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(pm)}]$, where $\boldsymbol{X}^{(i)}$ is the ith column of $\boldsymbol{X}$. For $i = 1, \ldots, pm$, we assume

$$\begin{pmatrix}\boldsymbol{U}_{j,k}\\ \boldsymbol{X}^{(i)}\end{pmatrix} \sim \left[\begin{pmatrix}\boldsymbol{0}_{n_{j-1}m}\\ \boldsymbol{0}_N\end{pmatrix}, \begin{pmatrix}\boldsymbol{V}_{j,k} & \boldsymbol{V}_{j,k,\cdot,\cdot}\\ \boldsymbol{V}'_{j,k,\cdot,\cdot} & \boldsymbol{V}+\boldsymbol{\Phi}\end{pmatrix}\right],$$

where $N \equiv m\sum_{j,k}\gamma_{j,k}n_{j-1}$ is the dimension of $\boldsymbol{Z}$, $\boldsymbol{V} \equiv \text{Var}(\boldsymbol{U})$, $\boldsymbol{\Phi} \equiv \text{Var}(\boldsymbol{e})$, $\boldsymbol{U} \equiv (\boldsymbol{U}'_{j,k} : \gamma_{j,k} = 1)'$, $\boldsymbol{e} \equiv (\boldsymbol{e}'_{j,k} : \gamma_{j,k} = 1)'$. Since $\boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{X}^{(i)}) = \boldsymbol{V}_{j,k,\cdot,\cdot}\boldsymbol{D}\boldsymbol{X}^{(i)}$, we can use the generalized change-of-resolution Kalman filter to compute $\boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{X}^{(i)})$ as we do with $\boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{Z})$. Thus $\boldsymbol{V}_{j,k,\cdot,\cdot}\boldsymbol{D}\boldsymbol{X} = [\boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{X}^{(1)}), \ldots, \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{X}^{(pm)})]$ and the operations remain of order $\mathcal{O}(n)$. A similar approach can be taken to compute the covariance of the BLUPs using an operation of order $\mathcal{O}(n)$ (see Yue and Zhu (2005) for details).

4. Model Parameter Estimation and Inference

Here we consider both maximum likelihood (ML) and restricted maximum likelihood (REML) estimation of the parameters in the MMTSLM (9)−(12). Let $\boldsymbol{Z} \equiv (\boldsymbol{Z}'_{j,k} : \gamma_{j,k} = 1)'$, $\boldsymbol{X} \equiv [\boldsymbol{X}'_{j,k} : \gamma_{j,k} = 1]'$, $\boldsymbol{U} \equiv (\boldsymbol{U}'_{j,k} : \gamma_{j,k} = 1)'$, and $\boldsymbol{e} \equiv (\boldsymbol{e}'_{j,k} : \gamma_{j,k} = 1)'$ denote the vectorized observations, covariates, underlying residual process, and measurement errors, respectively, and let $\boldsymbol{V} \equiv \text{Var}(\boldsymbol{U})$, $\boldsymbol{\Phi} \equiv \text{Var}(\boldsymbol{e})$. Then $\boldsymbol{Z} \sim N(\boldsymbol{X}\boldsymbol{B}, \boldsymbol{V}+\boldsymbol{\Phi})$, with log-likelihood function

$$\log\mathcal{L}(\boldsymbol{\theta}) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{V}+\boldsymbol{\Phi}| - \frac{1}{2}(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{B})'(\boldsymbol{V}+\boldsymbol{\Phi})^{-1}(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{B}), \quad (29)$$

where $\boldsymbol{\theta} = (\boldsymbol{B}', \boldsymbol{\eta}', \boldsymbol{\zeta}')'$ is the vector of model parameters, $N = m\sum_{j,k}\gamma_{j,k}n_{j-1}$ is the dimension of $\boldsymbol{Z}$, and $\boldsymbol{V}+\boldsymbol{\Phi}$ is invertible by Lemma 3 (ii) in Appendix II. The restricted log-likelihood function of $\boldsymbol{Z}$ is

$$\begin{aligned}\log\mathcal{L}_*(\boldsymbol{\eta}, \boldsymbol{\zeta}) = &-\frac{N-pm}{2}\log(2\pi) + \frac{1}{2}\log|\boldsymbol{X}'\boldsymbol{X}| - \frac{1}{2}\log|\boldsymbol{V}+\boldsymbol{\Phi}|\\ &-\frac{1}{2}\log|\boldsymbol{X}'(\boldsymbol{V}+\boldsymbol{\Phi})^{-1}\boldsymbol{X}|\\ &-\frac{1}{2}(\boldsymbol{Z}-\boldsymbol{X}\hat{\boldsymbol{B}})'(\boldsymbol{V}+\boldsymbol{\Phi})^{-1}(\boldsymbol{Z}-\boldsymbol{X}\hat{\boldsymbol{B}}), \quad (30)\end{aligned}$$

where $\hat{\boldsymbol{B}} = [\boldsymbol{X}'(\boldsymbol{V}+\boldsymbol{\Phi})^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}'(\boldsymbol{V}+\boldsymbol{\Phi})^{-1}\boldsymbol{Z}$ is the ML estimates (MLE) of $\boldsymbol{B}$ (see, e.g., Smyth and Verbyla (1996)). Recall that REML estimates (REMLE) of the variance parameters $(\boldsymbol{\eta}', \boldsymbol{\zeta}')'$ use a marginal likelihood function that does not depend on the mean parameters $\boldsymbol{B}$. Moreover, REMLEs and MLEs are asymptotically equivalent under mild conditions (see, e.g., Richardson and Welsh (1994)).

Direct computation of both the MLEs and the REMLEs may not be feasible for a large data size. In Huang, Cressie and Gabrosek (2002) and Zhu and Yue (2005), statistical inference is based on ML only, and the MLEs are obtained using an EM algorithm where the latent process is treated as observable but missing. While the EM algorithm is numerically stable, it often requires a large number of iterations before convergence is achieved. Here we propose a direct algorithm, which is of Newton-Raphson type and involves factorization of the likelihood function according to an ordering of the nodes in the multiresolution tree structure. Further, we consider the distributional properties of MLEs and REMLEs, which have not been addressed before.

4.1. Factorization and fast evaluation of the likelihood function

We order the $N_Z \equiv \sum \gamma_{j,k}$ vector nodes on the multiresolution tree structure and let $\boldsymbol{Z} = (\boldsymbol{Z}_1', \ldots, \boldsymbol{Z}_{N_Z}')'$ denote the response variables, where $\boldsymbol{Z}_i$ is the vectorized observation corresponding to the ith vector node according to a particular ordering. Then the likelihood function can be factorized to

$$\mathcal{L}(\boldsymbol{\theta}) = f(\boldsymbol{Z}_1|\boldsymbol{\theta}) \prod_{l=2}^{N_Z} f(\boldsymbol{Z}_l|\boldsymbol{Z}_{l-1}, \ldots, \boldsymbol{Z}_1; \boldsymbol{\theta}),$$

where $f(\boldsymbol{Z}_l|\boldsymbol{Z}_{l-1}, \ldots, \boldsymbol{Z}_1; \boldsymbol{\theta})$ is the conditional probability density function of $\boldsymbol{Z}_l$ given $\boldsymbol{Z}_{l-1}, \ldots, \boldsymbol{Z}_1$, and $\boldsymbol{\theta}$, $l = 2, \ldots, N_Z$. For normal distribution, to evaluate $\mathcal{L}$, it suffices to determine the conditional mean $E(\boldsymbol{Z}_l|\boldsymbol{Z}_{l-1}, \ldots, \boldsymbol{Z}_1; \boldsymbol{\theta})$ and conditional variance $\mathrm{Var}\,(\boldsymbol{Z}_l|\boldsymbol{Z}_{l-1}, \ldots, \boldsymbol{Z}_1; \boldsymbol{\theta})$, $l = 2, \ldots, N_Z$. More specifically, we define a function $s : \{j, k\} \mapsto s(j, k)$ where $s(j, k)$ is the order of vector node $\{j, k\}$ and $\boldsymbol{Z}^s_{j,k} \equiv \{\boldsymbol{Z}_{j',k'} : \gamma_{j',k'} = 1, s(j', k') < s(j, k)\}$. There are different ways of ordering the vector nodes so that an algorithm similar to the generalized change-of-resolution Kalman filter in Section 3 can be devised. Here we use an ordering developed by Luettgen (1993). Using the corresponding change-of-resolution Kalman filter algorithm (see Yue and Zhu (2005) for details), we can obtain the BLUP $\hat{U}^s_{j,k} \equiv \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}^s_{j,k})$ and the corresponding MSPE $\hat{\boldsymbol{V}}^s_{j,k} \equiv \boldsymbol{C}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}^s_{j,k})$, $k = 1, \ldots, N_{j-1}$, $j = 1, \ldots, J$. Thus the conditional mean and conditional variance are

$$\hat{\boldsymbol{Z}}_{j,k} \equiv E(\boldsymbol{Z}_{j,k}|\boldsymbol{Z}^s_{j,k}) = \boldsymbol{m}(\boldsymbol{Z}_{j,k}|\boldsymbol{Z}^s_{j,k}) = \boldsymbol{X}_{j,k}\boldsymbol{B} + \hat{\boldsymbol{U}}^s_{j,k},$$
$$\boldsymbol{\Lambda}_{j,k} \equiv \mathrm{Var}\,(\boldsymbol{Z}_{j,k}|\boldsymbol{Z}^s_{j,k}) = \boldsymbol{C}(\boldsymbol{Z}_{j,k}|\boldsymbol{Z}^s_{j,k}) = \hat{\boldsymbol{V}}^s_{j,k} + \boldsymbol{I}_{n_{j-1}} \otimes \boldsymbol{\Phi}_j,$$

and the log-likelihood function can be factorized into

$$\log \mathcal{L}(\boldsymbol{\theta}) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\sum_{\gamma_{j,k}=1}\log|\boldsymbol{\Lambda}_{j,k}| \\ -\frac{1}{2}\sum_{\gamma_{j,k}=1}(\boldsymbol{Z}_{j,k} - \boldsymbol{X}_{j,k}\boldsymbol{B} - \hat{\boldsymbol{U}}^s_{j,k})'\boldsymbol{\Lambda}^{-1}_{j,k}(\boldsymbol{Z}_{j,k} - \boldsymbol{X}_{j,k}\boldsymbol{B} - \hat{\boldsymbol{U}}^s_{j,k}), \quad (31)$$

where $\boldsymbol{\Lambda}_{j,k}$ is invertible because of Lemma 3 (ii) in Appendix II. To compute the restricted log-likelihood $\log \mathcal{L}_*(\boldsymbol{\eta}, \boldsymbol{\zeta})$, we utilize the property that $\boldsymbol{\Lambda}_{j,k}$ and $\hat{\boldsymbol{U}}^s_{j,k}$ do not depend on $\boldsymbol{B}$. It is straightforward to show that,

$$\log \mathcal{L}_*(\boldsymbol{\eta}, \boldsymbol{\zeta}) = -\frac{1}{2}(N - pm)\log(2\pi) + \frac{1}{2}\log|\sum_{\gamma_{j,k}=1}\boldsymbol{X}'_{j,k}\boldsymbol{X}_{j,k}| - \frac{1}{2}\sum_{\gamma_{j,k}=1}\log|\boldsymbol{\Lambda}_{j,k}| \\ -\frac{1}{2}\log|\sum_{\gamma_{j,k}=1}\boldsymbol{X}'_{j,k}\boldsymbol{\Lambda}^{-1}_{j,k}\boldsymbol{X}_{j,k}| \\ -\frac{1}{2}\sum_{\gamma_{j,k}=1}\Big\{(\boldsymbol{Z}_{j,k} - \boldsymbol{X}_{j,k}\hat{\boldsymbol{B}} - \hat{\boldsymbol{U}}^s_{j,k})'\boldsymbol{\Lambda}^{-1}_{j,k}(\boldsymbol{Z}_{j,k} - \boldsymbol{X}_{j,k}\hat{\boldsymbol{B}} - \hat{\boldsymbol{U}}^s_{j,k})\Big\}, \quad (32)$$

where

$$\hat{\boldsymbol{B}} = \Big[\sum_{\gamma_{j,k}=1}\boldsymbol{X}'_{j,k}\boldsymbol{\Lambda}^{-1}_{j,k}\boldsymbol{X}_{j,k}\Big]^{-1}\Big[\sum_{\gamma_{j,k}=1}\boldsymbol{X}'_{j,k}\boldsymbol{\Lambda}^{-1}_{j,k}\boldsymbol{Z}_{j,k}\Big]. \quad (33)$$

Moreover, we obtain

$$\mathrm{Var}\,(\hat{\boldsymbol{B}}) = \big[\boldsymbol{X}'(\boldsymbol{V} + \boldsymbol{\Phi})^{-1}\boldsymbol{X}\big]^{-1} = \Big[\sum_{\gamma_{j,k}=1}\boldsymbol{X}'_{j,k}\boldsymbol{\Lambda}^{-1}_{j,k}\boldsymbol{X}_{j,k}\Big]^{-1}. \quad (34)$$

The factorization of the likelihood functions ensures a fast computation of the log-likelihood and the restricted log-likelihood function. Thus we can obtain the ML and REML estimators using numerical maximization. The variances of the MLEs are approximated by the inverse of the observed information matrix $\boldsymbol{I}(\boldsymbol{\theta}) \equiv -\partial^2 \log \mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ evaluated at the MLEs. For $\boldsymbol{B}$, we use $\partial^2 \log \mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{B}\partial\boldsymbol{B}' = -\sum_{\gamma_{j,k}=1}\{\boldsymbol{X}'_{j,k}\boldsymbol{\Lambda}^{-1}_{j,k}\boldsymbol{X}_{j,k}\}$ and, for the other elements of $\boldsymbol{I}(\boldsymbol{\theta})$, we use numerical differentiation. Similarly the variances of the REMLEs are approximated by the inverse of $\boldsymbol{I}_*(\boldsymbol{\theta}) \equiv -\partial^2 \log \mathcal{L}_*(\boldsymbol{\eta}, \boldsymbol{\zeta})/\partial(\boldsymbol{\eta}', \boldsymbol{\zeta}')'\partial(\boldsymbol{\eta}, \boldsymbol{\zeta})$ evaluated at the REMLEs.

4.2. Analytical results

For analytical results, we restrict our attention to single-source data without any missing values. Thus the MMTSLM is $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{U} + \boldsymbol{e}$, where $\boldsymbol{Z} = (\boldsymbol{Z}'_{J,1}, \ldots, \boldsymbol{Z}'_{J,N_{J-1}})'$, $\boldsymbol{U} = (\boldsymbol{U}'_{J,1}, \ldots, \boldsymbol{U}'_{J,N_{J-1}})'$, $\boldsymbol{e} = (\boldsymbol{e}'_{J,1}, \ldots, \boldsymbol{e}'_{J,N_{J-1}})'$, $\boldsymbol{X} =$

$[\boldsymbol{X'}_{J,1},\ldots,\boldsymbol{X'}_{J,N_{J-1}}]'$ and $\boldsymbol{B}=\vec{\boldsymbol{\beta}}$ correspond to the response variables, the residual process, the measurement errors, the covariates, and the regression coefficients, respectively. That is, $\boldsymbol{Z}\sim N(\boldsymbol{XB},\boldsymbol{\Omega})$, where $\boldsymbol{\Omega}\equiv\boldsymbol{V}+\boldsymbol{\Phi}$, $\boldsymbol{V}\equiv\mathrm{Var}\,(\boldsymbol{U})$, $\boldsymbol{\Phi}\equiv\mathrm{Var}\,(\boldsymbol{e})=\boldsymbol{I}_{N_J}\otimes\boldsymbol{\Phi}_J$, and $\boldsymbol{\Phi}_J$ is a full rank $m\times m$ diagonal matrix which is assumed to be known or estimated from external data (see, e.g., Zhu and Yue (2005)). Further, we restrict our attention to the case where $\boldsymbol{\Omega}$ can be decomposed as

$$\boldsymbol{\Omega}=\sum_{j=1}^{J}a_j(\boldsymbol{A}_j\otimes\boldsymbol{\Psi}_j), \tag{35}$$

where $a_j\equiv N_J/N_j=n_j\cdots n_{J-1}$, $j=0,\ldots,J-1$, $a_J\equiv 1$, $\boldsymbol{A}_j\equiv[\boldsymbol{I}_{N_j}\otimes(\mathbf{1}_{a_j}\mathbf{1}'_{a_j})]/a_j$ is an $\boldsymbol{N}_J\times\boldsymbol{N}_J$ matrix with $j=0,\ldots,J$ and $\boldsymbol{\Psi}_j$ is an $m\times m$ semi-positive definite matrix with $j=1,\ldots,J$. Define

$$\boldsymbol{D}_j\equiv\sum_{k=j}^{J}a_k\boldsymbol{\Psi}_k, j=1,\ldots,J.$$

Here we assume that $\boldsymbol{\Omega}$ and $\boldsymbol{D}_j$ are invertible. For example, we consider two special cases of (35). One case involves independence among the child nodes. From (13) and $\boldsymbol{H}_j=\boldsymbol{I}_{n_{j-1}}$, $j=1,\ldots,J$, we have $\boldsymbol{\Omega}\equiv\mathrm{Var}\,(\boldsymbol{Z})=\sum_{j=1}^{J}a_j(\boldsymbol{A}_j\otimes\boldsymbol{\Psi}_j)$, where $\boldsymbol{\Psi}_j=\boldsymbol{\Sigma}_j$, $j=1,\ldots,J-1$ and $\boldsymbol{\Psi}_J=\boldsymbol{\Sigma}_J+\boldsymbol{\Phi}_J$. The other case involves mass balance. From (15), we have $\boldsymbol{\Omega}\equiv\mathrm{Var}\,(\boldsymbol{Z})=\sum_{j=1}^{J}a_j(\boldsymbol{A}_j\otimes\boldsymbol{\Psi}_j)$, where $\boldsymbol{\Psi}_1=\boldsymbol{\Sigma}_1-(1/(n_1-1))\boldsymbol{\Sigma}_2$, $\boldsymbol{\Psi}_j=(n_{j-1}/(n_{j-1}-1))\boldsymbol{\Sigma}_j-(1/(n_j-1))\boldsymbol{\Sigma}_{j+1}$, $j=2,\ldots,J-1$, and $\boldsymbol{\Psi}_J=(n_{J-1}/(n_{J-1}-1))\boldsymbol{\Sigma}_J+\boldsymbol{\Phi}_J$.

We define the following $m\times m$ matrices of sum of squares

$$SS_0(\boldsymbol{\beta})\equiv(\boldsymbol{z}-\boldsymbol{x\beta})'\boldsymbol{A}_0(\boldsymbol{z}-\boldsymbol{x\beta}), \tag{36}$$

$$SS_j(\boldsymbol{\beta})\equiv(\boldsymbol{z}-\boldsymbol{x\beta})'[\boldsymbol{A}_j-\boldsymbol{A}_{j-1}](\boldsymbol{z}-\boldsymbol{x\beta}), j=1,\ldots J, \tag{37}$$

$$SST(\boldsymbol{\beta})\equiv(\boldsymbol{z}-\boldsymbol{x\beta})'\boldsymbol{A}_J(\boldsymbol{z}-\boldsymbol{x\beta})=(\boldsymbol{z}-\boldsymbol{x\beta})'(\boldsymbol{z}-\boldsymbol{x\beta})=\sum_{j=0}^{J}SS_j(\boldsymbol{\beta}), \tag{38}$$

where, for ease of presentation, $\boldsymbol{z}=[\boldsymbol{z}'_{J,1},\ldots,\boldsymbol{z}'_{J,N_J}]'$, $\boldsymbol{x}=[\boldsymbol{x}_{J,1},\ldots,\boldsymbol{x}_{J,N_J}]'$, $\boldsymbol{u}=[\boldsymbol{u}'_{J,1},\ldots,\boldsymbol{u}'_{J,N_J}]'$, and $\boldsymbol{\epsilon}=[\boldsymbol{\epsilon}'_{J,1},\ldots,\boldsymbol{\epsilon}'_{J,N_J}]'$ are written in matrix forms based on the scalar nodes corresponding to the response variables, the covariates, the residual process, and the measurement errors, respectively. It follows that $\boldsymbol{z}=\boldsymbol{x\beta}+\boldsymbol{u}+\boldsymbol{\epsilon}$ and $\boldsymbol{Z}=\vec{\boldsymbol{z}}$, $\boldsymbol{U}=\vec{\boldsymbol{u}}$, $\boldsymbol{e}=\vec{\boldsymbol{\epsilon}}$, $\boldsymbol{X}=\boldsymbol{x}\otimes\boldsymbol{I}_m$, and $\boldsymbol{B}=\vec{\boldsymbol{\beta}}$. Hence from (29) and Lemmas 8 and 9 in Appendix III, we have

$$\log\mathcal{L}(\boldsymbol{\theta})=-\frac{N_Jm}{2}\log(2\pi)-\frac{1}{2}\Big[N_1\log|\boldsymbol{D}_1|+\sum_{j=2}^{J}(N_j-N_{j-1})\log|\boldsymbol{D}_j|\Big]$$
$$-\frac{1}{2}\Big[\mathrm{tr}[SS_0(\boldsymbol{\beta})\boldsymbol{D}_1^{-1}]+\sum_{j=1}^{J}\mathrm{tr}[SS_j(\boldsymbol{\beta})\boldsymbol{D}_j^{-1}]\Big]. \tag{39}$$

Similarly, from (30) and Lemmas 8 and 9 in Appendix III, we have

$$\log \mathcal{L}_*(\boldsymbol{\eta},\boldsymbol{\zeta}) = -\frac{(N_J-p)m}{2}\log(2\pi) + \frac{1}{2}\log|\boldsymbol{X}'\boldsymbol{X}| \\ -\frac{1}{2}\Big[N_1\log|\boldsymbol{D}_1| + \sum_{j=2}^{J}(N_j-N_{j-1})\log|\boldsymbol{D}_j|\Big] - \frac{1}{2}\log|\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}| \\ -\frac{1}{2}\Big[\sum_{j=1}^{J}\mathrm{tr}[SS_j(\hat{\boldsymbol{\beta}})\boldsymbol{D}_j^{-1}] + \mathrm{tr}[SS_0(\hat{\boldsymbol{\beta}})\boldsymbol{D}_1^{-1}]\Big], \tag{40}$$

where $\hat{\boldsymbol{\beta}}$ is obtained from $\mathrm{vec}(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{B}} = [\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1}[\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{Z}]$.

To obtain the MLEs and REMLEs, we differentiate the components of (39) and (40) with respect to the parameters. Using Lemmas 10 (ii), 11 and 12 in Appendix III, we differentiate $\log\mathcal{L}(\boldsymbol{\theta})$ in (29) with respect to $\boldsymbol{B}$, and differentiate $\log\mathcal{L}(\boldsymbol{\theta})$ in (39) and $\log\mathcal{L}_*(\boldsymbol{\eta},\boldsymbol{\zeta})$ in (40) with respect to $\boldsymbol{D}_j^{-1}$, $j=1,\ldots,J$, such that

$$\boldsymbol{b} \equiv \frac{\partial \log\mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{B}} = \boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{Z} - \boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}\boldsymbol{B}, \tag{41}$$

$$\boldsymbol{M}_j \equiv \frac{\partial \log\mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{D}_j^{-1}} = \begin{cases} \frac{N_1}{2}\boldsymbol{D}_1' - \frac{1}{2}(SS_1(\boldsymbol{\beta}) + SS_0(\boldsymbol{\beta}))' & \text{if } j=1, \\ \frac{N_j-N_{j-1}}{2}\boldsymbol{D}_j' - \frac{1}{2}SS_j(\boldsymbol{\beta})' & \text{if } j=2,\ldots,J, \end{cases} \tag{42}$$

$$\boldsymbol{M}_j^* \equiv \frac{\partial \log\mathcal{L}_*(\boldsymbol{\eta},\boldsymbol{\zeta})}{\partial \boldsymbol{D}_j^{-1}} \\ = \begin{cases} \frac{N_1}{2}\boldsymbol{D}_1' - \frac{1}{2}\Big[\mathrm{tr}[\boldsymbol{P}(\boldsymbol{A}_1\otimes\boldsymbol{Q}_{hi})]\Big]_{hi} - \frac{1}{2}(SS_1(\hat{\boldsymbol{\beta}}) + SS_0(\hat{\boldsymbol{\beta}}))' & \text{if } j=1, \\ \frac{N_j-N_{j-1}}{2}\boldsymbol{D}_j' - \frac{1}{2}\Big[\mathrm{tr}[\boldsymbol{P}((\boldsymbol{A}_j-\boldsymbol{A}_{j-1})\otimes\boldsymbol{Q}_{hi})]\Big]_{hi} - \frac{1}{2}SS_j(\hat{\boldsymbol{\beta}})' & \text{if } j=2,\ldots,J, \end{cases} \tag{43}$$

where we use $[g(h,i)]_{hi}$ to denote a matrix whose (h,i)th element is $g(h,i)$, $\boldsymbol{P} \equiv \boldsymbol{X}\left[\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}\right]^{-1}\boldsymbol{X}'$, and $\boldsymbol{Q}_{hi}$ is an $m\times m$ matrix with 1 for the (h,i)th element and 0 otherwise. For an element $\theta_i \in (\boldsymbol{\eta}',\boldsymbol{\zeta}')'$, by Lemma 10 (vi) in Appendix III the score functions are

$$\frac{\partial \log\mathcal{L}(\boldsymbol{\theta})}{\partial\theta_i} = \mathrm{tr}\left[\Big(\frac{\partial \log\mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{D}_j^{-1}}\Big)'\Big(\frac{\partial \boldsymbol{D}_j^{-1}}{\partial\theta_i}\Big)\right], \tag{44}$$

$$\frac{\partial \log\mathcal{L}_*(\boldsymbol{\eta},\boldsymbol{\zeta})}{\partial\theta_i} = \mathrm{tr}\left[\Big(\frac{\partial \log\mathcal{L}_*(\boldsymbol{\eta},\boldsymbol{\zeta})}{\partial \boldsymbol{D}_j^{-1}}\Big)'\Big(\frac{\partial \boldsymbol{D}_j^{-1}}{\partial\theta_i}\Big)\right]. \tag{45}$$

Theorem 3. *For the MMTSLM* (9)$-$(12), *under* (35), *the score functions in*

(41), (44) *and* (45) *are unbiased. That is,*

$$E\left(\frac{\partial \log \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{B}}\right) = \mathbf{0}_{pm}, \quad E\left(\frac{\partial \log \mathcal{L}(\boldsymbol{\theta})}{\partial \theta_i}\right) = 0, \quad E\left(\frac{\partial log \mathcal{L}_*(\boldsymbol{\eta}, \boldsymbol{\zeta})}{\partial \theta_i}\right) = 0 \quad \text{for } \theta_i \in (\boldsymbol{\eta}', \boldsymbol{\zeta}')'. \tag{46}$$

It is obvious that $E(\boldsymbol{b}) = \mathbf{0}_{pm}$. Lemma 13 given in Appendix III shows that $E(\boldsymbol{M}_j) = \mathbf{0}_{m\times m}$ and $E(\boldsymbol{M}_j^*) = \mathbf{0}_{m\times m}$ for $j = 1, \ldots, J$. Thus Theorem 3 follows from (44) and (45).

4.3. Special cases

Here we derive the explicit forms of MLEs and REMLEs when $\boldsymbol{x} = \mathbf{1}_{N_J}$, $\boldsymbol{X} = \mathbf{1}_{N_J} \otimes \boldsymbol{I}_m$, $\boldsymbol{\beta} = (\mu_1, \ldots, \mu_m)$ and $\boldsymbol{B} = (\mu_1, \ldots, \mu_m)'$. That is, the MMTSLM only has intercepts in the regression mean. Let $\hat{a}$ denote the MLE of a and $\tilde{a}$ denote the REMLE of a. By (41)−(43) and Lemma 14 in Appendix III, we obtain the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{D}$ along with their expectations and variances,

$$\begin{aligned}
&\hat{\boldsymbol{\beta}} = \tfrac{1}{N_J}\mathbf{1}'_{N_J}\boldsymbol{z}, && E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}, \\
& && \mathrm{Var}\,(\hat{\boldsymbol{\beta}}') = [\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1} = \tfrac{\boldsymbol{D}_1}{N_J}, \\
&\hat{\boldsymbol{D}}_1 = \tfrac{1}{N_1} SS_1(\hat{\boldsymbol{\beta}}), && E(\hat{\boldsymbol{D}}_1) = \tfrac{N_1-1}{N_1}\boldsymbol{D}_1, \\
& && \mathrm{Var}\,(\hat{\boldsymbol{D}}_{1hi}) = \tfrac{N_1-1}{N_1^2}(\boldsymbol{D}_{1hi}^2 + \boldsymbol{D}_{1hh}\boldsymbol{D}_{1ii}), \\
&\hat{\boldsymbol{D}}_j = \tfrac{1}{N_j - Nj-1} SS_j(\hat{\boldsymbol{\beta}}), && E(\hat{\boldsymbol{D}}_j) = \boldsymbol{D}_j, \\
& && \mathrm{Var}\,(\hat{\boldsymbol{D}}_{jhi}) = \tfrac{1}{N_j - Nj-1}(\boldsymbol{D}_{jhi}^2 + \boldsymbol{D}_{jhh}\boldsymbol{D}_{jii}), j = 2, \ldots, J, \\
&\tilde{\boldsymbol{D}}_j = \tfrac{1}{N_j - N_{j-1}} SS_j(\hat{\boldsymbol{\beta}}) && E(\tilde{\boldsymbol{D}}_j) = \boldsymbol{D}_j, \\
& && \mathrm{Var}\,(\tilde{\boldsymbol{D}}_{jhi}) = \tfrac{1}{N_j - Nj-1}(\boldsymbol{D}_{jhi}^2 + \boldsymbol{D}_{jhh}\boldsymbol{D}_{jii}), j = 1, \ldots, J,
\end{aligned} \tag{47}$$

where we note that $\hat{\boldsymbol{D}}_j = \tilde{\boldsymbol{D}}_j$ for $j = 2, \ldots, J$, and $\boldsymbol{D}_{jhi}$ is the (h, i)th element of $\boldsymbol{D}_j, j = 1, \ldots, J, h, i = 1, \ldots, m$. That is, the estimates of $\boldsymbol{\beta}$ and D_j are all unbiased except for the MLEs $\hat{\boldsymbol{D}}_1$ on the coarsest resolution. Furthermore, we obtain the exact distributions of the sums of squares $SS_j(\cdot)$ as follows.

Theorem 4. *For the MMTSLM* (9)−(12) *that has a constant mean for each response variable, under* (35), $SS_0(\hat{\boldsymbol{\beta}}) = \mathbf{0}_{m\times m}$, $SS_j(\hat{\boldsymbol{\beta}}) \sim W_m(N_j - N_{j-1}, \boldsymbol{D}_j)$, $j = 1, \ldots, J$, *where* $W_m(N_j - N_{j-1}, \boldsymbol{D}_j)$ *denotes an* m*-dimensional Wishart distribution with* $N_j - N_{j-1}$ *degrees of freedom and parameter* $\boldsymbol{D}_j$*. Furthermore,* $\{SS_j(\hat{\boldsymbol{\beta}}) : j = 1, \ldots, J\}$ *are mutually independent and are independent of* $\hat{\boldsymbol{\beta}}$*.*

The Wishart distribution is defined as in C.9 of Lauritzen (1996). The proof of Theorem 4 is given in Lemma 15 of Appendix III. The results give the exact distributions of the sums of squares $SS_j(\cdot)$, which are the building blocks for $\boldsymbol{D}_j$. The independence among these sums of squares is also a nice feature, and facilitates the computation of variances in many cases. Even though the results here are specifically for the MMTSLM, the techniques used for derivation could be of interest in the linear model theory literature. For example, Rao and Heckler (1998) showed analytical results for a multivariate one-way random effects model, which can be viewed as a special case of MMTSLM with $J = 2$ resolutions.

4.3.1. Compound symmetry

When the matrices $\boldsymbol{\Psi}_j$, $j = 1, \ldots, J$, are further parameterized, more explicit forms of the MLEs and REMLEs may be available. Here we consider the case where $\boldsymbol{\Psi}_j$ has a compound symmetry structure with diagonal elements ψ_{j1} and off-diagonal elements ψ_{j2}, $j = 1, \ldots, J$. Thus the $m \times m$ matrix $\boldsymbol{D}_j$ also has a compound symmetry structure with diagonal elements d_{j1} and off-diagonal element d_{j2}, where $d_{j1} = \sum_{k=j}^{J} a_k \psi_{k1}$ and $d_{j2} = \sum_{k=j}^{J} a_k \psi_{k2}$, $j = 1, \ldots, J$. Equivalently, $\boldsymbol{D}_j = (d_{j1} - d_{j2})\boldsymbol{I}_m + d_{j2}\mathbf{1}_m\mathbf{1}_m'$, where $j = 1, \ldots, J$. It is easy to verify that $\boldsymbol{D}_j^{-1} = (1/(d_{j1} - d_{j2}))(\boldsymbol{I}_m - (d_{j2}/(d_{j1} + (m-1)d_{j2}))\mathbf{1}_m\mathbf{1}_m')$ has a compound symmetry structure with diagonal elements $d_{j1}^* = 1/d_{j1} - d_{j2} - d_{j2}/(d_{j1} + (m-1)d_{j2})$ and off-diagonal elements $d_{j2}^* = -d_{j2}/(d_{j1} + (m-1)d_{j2})$, $j = 1, \ldots, J$. We have obtained $\partial \log \mathcal{L}(\boldsymbol{\theta})/\partial \boldsymbol{D}_j^{-1}$ and $\partial \log \mathcal{L}_*(\boldsymbol{\eta}, \boldsymbol{\zeta})/\partial \boldsymbol{D}_j^{-1}$ in (42)−(43). Now we compute $\partial \boldsymbol{D}_j^{-1}/\partial \theta_i$. Using (44)−(45), $\partial \boldsymbol{D}_j^{-1}/\partial d_{j1}^* = \boldsymbol{I}_m$, and $\partial \boldsymbol{D}_j^{-1}/\partial d_{j2}^* = \mathbf{1}_m\mathbf{1}_m' - \boldsymbol{I}_m$, we obtain the MLE and REMLE of d_{j1}, d_{j2}:

$$\hat{d}_{j1} = \begin{cases} \frac{1}{N_1 m} \operatorname{tr}[SS_1(\hat{\boldsymbol{\beta}})] & \text{if } j = 1, \\ \frac{1}{(N_j - N_{j-1})m} \operatorname{tr}[SS_j(\hat{\boldsymbol{\beta}})] & \text{if } j = 2, \ldots, J, \end{cases} \tag{48}$$

$$\hat{d}_{j2} = \begin{cases} \frac{1}{N_1 m(m-1)} \operatorname{tr}[SS_1(\hat{\boldsymbol{\beta}})(\mathbf{1}_m\mathbf{1}_m' - \boldsymbol{I}_m)] & \text{if } j = 1, \\ \frac{1}{(N_j - N_{j-1})m(m-1)} \operatorname{tr}[SS_j(\hat{\boldsymbol{\beta}})(\mathbf{1}_m\mathbf{1}_m' - \boldsymbol{I}_m)] & \text{if } j = 2, \ldots, J, \end{cases} \tag{49}$$

$$\tilde{d}_{j1} = \frac{1}{(N_j - N_{j-1})m} \operatorname{tr}[SS_j(\hat{\boldsymbol{\beta}})] \quad \text{if } j = 1, \ldots, J, \tag{50}$$

$$\tilde{d}_{j2} = \frac{1}{(N_j - N_{j-1})m(m-1)} \operatorname{tr}[SS_j(\hat{\boldsymbol{\beta}})(\mathbf{1}_m\mathbf{1}_m' - \boldsymbol{I}_m)] \quad \text{if } j = 1, \ldots, J, \tag{51}$$

where we note that $\hat{d}_{ji} = \tilde{d}_{ji}$ for $j = 2, \ldots, J$, $i = 1, 2$.

Using Lemmas 5 (vii), 15 (iii) and (v), and 16 in Appendix III, for $j =$

$1, \ldots, J$,

$$E(\mathrm{tr}[SS_j(\hat{\boldsymbol{\beta}})]) = (N_j - N_{j-1}) m d_{j1},$$

$$\mathrm{Var}\,(\mathrm{tr}[SS_j(\hat{\boldsymbol{\beta}})]) = 2(N_j - N_{j-1}) m (d_{j1}^2 + (m-1) d_{j2}^2),$$

$$E(\mathrm{tr}[SS_j(\hat{\boldsymbol{\beta}})](\mathbf{1}_m \mathbf{1}_m' - \boldsymbol{I}_m)) = (N_j - N_{j-1}) m (m-1) d_{j2},$$

$$\mathrm{Var}\,(\mathrm{tr}[SS_j(\hat{\boldsymbol{\beta}})](\mathbf{1}_m \mathbf{1}_m' - \boldsymbol{I}_m)] = 2(N_j - N_{j-1}) m (m-1) \times \Big[d_{j1}^2 + 2(m-2) d_{j1} d_{j2} + (m^2 - 3m + 3) d_{j2}^2\Big].$$

Then the expectation and variance of the MLE and REMLE of d_{j1}, d_{j2} are

$$\begin{aligned}
&E(\hat{d}_{11}) = \tfrac{N_1 - 1}{N_1} d_{11}, \quad \mathrm{Var}\,(\hat{d}_{11}) = \tfrac{2(N_1 - 1)}{N_1^2 m} (d_{11}^2 + (m-1) d_{12}^2),\\
&E(\hat{d}_{j1}) = d_{j1},\\
&\mathrm{Var}\,(\hat{d}_{j1}) = \tfrac{2}{(N_j - N_{j-1}) m} (d_{j1}^2 + (m-1) d_{j2}^2), \qquad j = 2, \ldots, J,\\
&E(\hat{d}_{12}) = \tfrac{N_1 - 1}{N_1} d_{12},\\
&\mathrm{Var}\,(\hat{d}_{12}) = \tfrac{2(N_1 - 1)}{N_1^2 m (m-1)} \Big[d_{11}^2 + 2(m-2) d_{11} d_{12} + (m^2 - 3m + 3) d_{12}^2\Big],\\
&E(\hat{d}_{j2}) = d_{j2},\\
&\mathrm{Var}\,(\hat{d}_{j2}) = \tfrac{2}{(N_j - N_{j-1}) m (m-1)} \Big[d_{j1}^2 + 2(m-2) d_{j1} d_{j2} + (m^2 - 3m + 3) d_{j2}^2\Big], \quad j = 2, \ldots, J,\\
&E(\tilde{d}_{j1}) = d_{j1},\\
&\mathrm{Var}\,(\tilde{d}_{j1}) = \tfrac{2}{(N_j - N_{j-1}) m} (d_{j1}^2 + (m-1) d_{j2}^2), \qquad j = 1, \ldots, J,\\
&E(\tilde{d}_{j2}) = d_{j2},\\
&\mathrm{Var}\,(\tilde{d}_{j2}) = \tfrac{2}{(N_j - N_{j-1}) m (m-1)} \Big[d_{j1}^2 + 2(m-2) d_{j1} d_{j2} + (m^2 - 3m + 3) d_{j2}^2\Big], \quad j = 1, \ldots, J.
\end{aligned} \tag{52}$$

We note that the MLEs and REMLEs are all unbiased with the exception of the MLEs on the coarsest resolution $\hat{d}_{11}, \hat{d}_{12}$. In fact, other than the coarsest resolution, the MLEs and REMLEs of the parameters are the same in the case of constant regression mean. Now we reconsider the two special cases: $\boldsymbol{H}_j = \boldsymbol{I}_{n_{j-1}}$, $j = 1, \ldots, J$, which we call the independence case, and $\boldsymbol{H}_1 = \boldsymbol{I}_{N_1}$, $\boldsymbol{H}_{j+1} = (n_j/(n_j - 1))(\boldsymbol{I}_{n_j} - \mathbf{1}_{n_j} \mathbf{1}_{n_j}'/n_j)$, $j = 1, \ldots, J-1$, which we call the mass-balance case.

4.3.2. Independence case

Here $\boldsymbol{\Omega}$ and $\boldsymbol{D}_j$, $j=1,\dots,J$ are in fact invertible because of Lemma 3 (ii) in Appendix II. It is easy to obtain $\boldsymbol{\Sigma}_j = \boldsymbol{\Psi}_j = (\boldsymbol{D}_j - \boldsymbol{D}_{j+1})/a_j$, $j=1,\dots,J-1$, and $\boldsymbol{\Sigma}_J = \boldsymbol{\Psi}_J - \boldsymbol{\Phi}_J = \boldsymbol{D}_J - \boldsymbol{\Phi}_J$. Using (47) and Lemma 15 in Appendix III, we have the MLE and REMLE of $\boldsymbol{\Sigma}_j$ along with their expectations and variances,

$$
\begin{aligned}
&\hat{\boldsymbol{\Sigma}}_1 = \tfrac{1}{a_1}(\hat{\boldsymbol{D}}_1 - \hat{\boldsymbol{D}}_2), && E(\hat{\boldsymbol{\Sigma}}_1) = \boldsymbol{\Sigma}_1 - \tfrac{\boldsymbol{D}_1}{N_J},\\
& && \mathrm{Var}\,(\hat{\boldsymbol{\Sigma}}_{1hi}) = \tfrac{1}{a_1^2}[\mathrm{Var}\,(\hat{\boldsymbol{D}}_{1hi}) + \mathrm{Var}\,(\hat{\boldsymbol{D}}_{2hi})],\\
&\hat{\boldsymbol{\Sigma}}_j = \tfrac{1}{a_j}(\hat{\boldsymbol{D}}_j - \hat{\boldsymbol{D}}_{j+1}), && E(\hat{\boldsymbol{\Sigma}}_j) = \boldsymbol{\Sigma}_j,\\
& && \mathrm{Var}\,(\hat{\boldsymbol{\Sigma}}_{jhi}) = \tfrac{1}{a_j^2}[\mathrm{Var}\,(\hat{\boldsymbol{D}}_{jhi}) + \mathrm{Var}\,(\hat{\boldsymbol{D}}_{j+1hi})],\\
& && j=2,\dots,J-1,\\
&\hat{\boldsymbol{\Sigma}}_J = \hat{\boldsymbol{D}}_J - \boldsymbol{\Phi}_J, && E(\hat{\boldsymbol{\Sigma}}_J) = \boldsymbol{\Sigma}_J,\\
& && \mathrm{Var}\,(\hat{\boldsymbol{\Sigma}}_{Jhi}) = \tfrac{1}{a_J^2}\mathrm{Var}\,(\hat{\boldsymbol{D}}_{Jhi}),\\
&\tilde{\boldsymbol{\Sigma}}_j = \tfrac{1}{a_j}(\tilde{\boldsymbol{D}}_j - \tilde{\boldsymbol{D}}_{j+1}), && E(\tilde{\boldsymbol{\Sigma}}_j) = \boldsymbol{\Sigma}_j,\\
& && \mathrm{Var}\,(\tilde{\boldsymbol{\Sigma}}_{jhi}) = \tfrac{1}{a_j^2}[\mathrm{Var}\,(\tilde{\boldsymbol{D}}_{jhi}) + \mathrm{Var}\,(\tilde{\boldsymbol{D}}_{j+1hi})],\\
& && j=1,\dots,J-1,\\
&\tilde{\boldsymbol{\Sigma}}_J = \tilde{\boldsymbol{D}}_J - \boldsymbol{\Phi}_J, && E(\tilde{\boldsymbol{\Sigma}}_J) = \boldsymbol{\Sigma}_J,\\
& && \mathrm{Var}\,(\tilde{\boldsymbol{\Sigma}}_{Jhi}) = \tfrac{1}{a_J^2}\mathrm{Var}\,(\tilde{\boldsymbol{D}}_{Jhi}),
\end{aligned}
\tag{53}
$$

where we note that $\hat{\boldsymbol{\Sigma}}_j = \tilde{\boldsymbol{\Sigma}}_j$ for $j=2,\dots,J$.

If $\boldsymbol{\Sigma}_j$ has a compound symmetry structure with diagonal element σ_{j1} and off-diagonal element σ_{j2}, $j=1,\dots,J$, and $\boldsymbol{\Phi}_J$ has a compound symmetry structure with diagonal element ϕ_J and off-diagonal element 0, then the $m\times m$ matrices $\boldsymbol{D}_j$ has a compound symmetry structure with diagonal element d_{j1} and off-diagonal element d_{j2}, where $d_{j1} = \sum_{k=j}^{J} a_k\sigma_{k1} + \phi_J$ and $d_{j2} = \sum_{k=j}^{J} a_k\sigma_{k2}$, $j=1,\dots,J$. Hence using (48)$-$(52), we have the MLE and REMLE of σ_{j1}, σ_{j2} along with their expectations and variances,

$$
\begin{aligned}
&\hat{\sigma}_{11} = \tfrac{1}{a_1}(\hat{d}_{11} - \hat{d}_{21}), && E(\hat{\sigma}_{11}) = \sigma_{11} - \tfrac{d_{11}}{N_J},\\
& && \mathrm{Var}\,(\hat{\sigma}_{11}) = \tfrac{1}{a_1^2}[\mathrm{Var}\,(\hat{d}_{11}) + \mathrm{Var}\,(\hat{d}_{21})],\\
&\hat{\sigma}_{j1} = \tfrac{1}{a_j}(\hat{d}_{j1} - \hat{d}_{j+1\,1}), && E(\hat{\sigma}_{j1}) = \sigma_{j1},\\
& && \mathrm{Var}\,(\hat{\sigma}_{j1}) = \tfrac{1}{a_j^2}[\mathrm{Var}\,(\hat{d}_{j1}) + \mathrm{Var}\,(\hat{d}_{j+1\,1})], j=2,\dots,J-1,\\
&\hat{\sigma}_{J1} = \hat{d}_{J1} - \phi_J, && E(\hat{\sigma}_{J1}) = \sigma_{J1}, \quad \mathrm{Var}\,(\hat{\sigma}_{J1}) = \mathrm{Var}\,(\hat{d}_{J1}),
\end{aligned}
$$

$$
\begin{aligned}
&\hat{\sigma}_{12} = \tfrac{1}{a_1}(\hat{d}_{12} - \hat{d}_{22}), && E(\hat{\sigma}_{12}) = \sigma_{12} - \tfrac{d_{12}}{N_J},\\
& && \mathrm{Var}\,(\hat{\sigma}_{12}) = \tfrac{1}{a_1^2}[\mathrm{Var}\,(\hat{d}_{12}) + \mathrm{Var}\,(\hat{d}_{22})],\\
&\hat{\sigma}_{j2} = \tfrac{1}{a_j}(\hat{d}_{j2} - \hat{d}_{j+1\,2}), && E(\hat{\sigma}_{j2}) = \sigma_{j2},\\
& && \mathrm{Var}\,(\hat{\sigma}_{j2}) = \tfrac{1}{a_j^2}[\mathrm{Var}\,(\hat{d}_{j2}) + \mathrm{Var}\,(\hat{d}_{j+1\,2})], j = 2,\ldots,J-1,\\
&\hat{\sigma}_{J2} = \hat{d}_{J2}, && E(\hat{\sigma}_{J2}) = \sigma_{J2}, \quad \mathrm{Var}\,(\hat{\sigma}_{J2}) = \mathrm{Var}\,(\hat{d}_{J2}),\\
&\tilde{\sigma}_{j1} = \tfrac{1}{a_j}(\tilde{d}_{j1} - \tilde{d}_{j+1\,1}), && E(\tilde{\sigma}_{j1}) = \sigma_{j1}, \qquad (54)\\
& && \mathrm{Var}\,(\tilde{\sigma}_{j1}) = \tfrac{1}{a_j^2}[\mathrm{Var}\,(\tilde{d}_{j1}) + \mathrm{Var}\,(\tilde{d}_{j+1\,1})], j = 1,\ldots,J-1,\\
&\tilde{\sigma}_{J1} = \tilde{d}_{J1} - \phi_J, && E(\tilde{\sigma}_{J1}) = \sigma_{J1}, \quad \mathrm{Var}\,(\tilde{\sigma}_{J1}) = \mathrm{Var}\,(\tilde{d}_{J1}),\\
&\tilde{\sigma}_{j2} = \tfrac{1}{a_j}(\tilde{d}_{j2} - \tilde{d}_{j+1\,2}), && E(\tilde{\sigma}_{j2}) = \sigma_{j2},\\
& && \mathrm{Var}\,(\tilde{\sigma}_{j2}) = \tfrac{1}{a_j^2}[\mathrm{Var}\,(\tilde{d}_{j2}) + \mathrm{Var}\,(\tilde{d}_{j+1\,2})], j = 1,\ldots,J-1,\\
&\tilde{\sigma}_{J2} = \tilde{d}_{J2}, && E(\tilde{\sigma}_{J2}) = \sigma_{J2}, \quad \mathrm{Var}\,(\tilde{\sigma}_{J2}) = \mathrm{Var}\,(\tilde{d}_{J2}),
\end{aligned}
$$

where we note that $\hat{\sigma}_{ji} = \tilde{\sigma}_{ji}$ for $j = 2,\ldots,J$, $i = 1,2$.

4.3.3. Mass-balance case

Here $\boldsymbol{\Omega}$ and $\boldsymbol{D}_j, j = 1,\ldots,J$, are in fact invertible because of Lemma 3 (ii) in Appendix II. Moreover

$$
\boldsymbol{D}_j \equiv \sum_{k=j}^{J} a_k \boldsymbol{\Psi}_k = \begin{cases} a_1 \boldsymbol{\Sigma}_1 + \boldsymbol{\Phi}_J & \text{if } j = 1,\\ \frac{a_{j-1}}{n_{j-1}-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Phi}_J & \text{if } j = 2,\ldots,J. \end{cases} \qquad (55)
$$

Hence using (47), we have the MLE and REMLE of $\boldsymbol{\Sigma}_j$ along with their expectations and variances:

$$
\begin{aligned}
&\hat{\boldsymbol{\Sigma}}_1 = \tfrac{1}{a_1}(\hat{\boldsymbol{D}}_1 - \boldsymbol{\Phi}_J), && E(\hat{\boldsymbol{\Sigma}}_1) = \boldsymbol{\Sigma}_1 - \tfrac{\boldsymbol{D}_1}{N_J}, \quad \mathrm{Var}\,(\hat{\boldsymbol{\Sigma}}_{1hi}) = \tfrac{1}{a_1^2}\mathrm{Var}\,(\hat{\boldsymbol{D}}_{1hi}),\\
&\hat{\boldsymbol{\Sigma}}_j = \tfrac{n_{j-1}-1}{a_{j-1}}(\hat{\boldsymbol{D}}_j - \boldsymbol{\Phi}_J), && E(\hat{\boldsymbol{\Sigma}}_j) = \boldsymbol{\Sigma}_j,\\
& && \mathrm{Var}\,(\hat{\boldsymbol{\Sigma}}_{jhi}) = \tfrac{(n_{j-1}-1)^2}{a_{j-1}^2}\mathrm{Var}\,(\hat{\boldsymbol{D}}_{jhi}), j = 2,\ldots,J,\\
&\tilde{\boldsymbol{\Sigma}}_1 = \tfrac{1}{a_1}(\tilde{\boldsymbol{D}}_1 - \boldsymbol{\Phi}_J), && E(\tilde{\boldsymbol{\Sigma}}_1) = \boldsymbol{\Sigma}_1, \quad \mathrm{Var}\,(\tilde{\boldsymbol{\Sigma}}_{1hi}) = \tfrac{1}{a_1^2}\mathrm{Var}\,(\tilde{\boldsymbol{D}}_{1hi}), \qquad (56)\\
&\tilde{\boldsymbol{\Sigma}}_j = \tfrac{n_{j-1}-1}{a_{j-1}}(\tilde{\boldsymbol{D}}_j - \boldsymbol{\Phi}_J), && E(\tilde{\boldsymbol{\Sigma}}_j) = \boldsymbol{\Sigma}_j,\\
& && \mathrm{Var}\,(\tilde{\boldsymbol{\Sigma}}_{jhi}) = \tfrac{(n_{j-1}-1)^2}{a_{j-1}^2}\mathrm{Var}\,(\tilde{\boldsymbol{D}}_{jhi}), j = 2,\ldots,J,
\end{aligned}
$$

where we note that $\hat{\boldsymbol{\Sigma}}_j = \tilde{\boldsymbol{\Sigma}}_j$ for $j = 2,\ldots,J$.

If we assume $\boldsymbol{\Sigma}_j$ has a compound symmetry structure with diagonal element σ_{j1} and off-diagonal element σ_{j2}, $j = 1, \ldots, J$, and $\boldsymbol{\Phi}_J$ has a compound symmetry structure with diagonal element ϕ_J and off-diagonal element 0, then the $m \times m$ matrices $\boldsymbol{D}_j$ has a compound symmetry structure with diagonal element d_{j1} and off-diagonal element d_{j2} where $d_{11} = a_1\sigma_{11} + \phi_J$, $d_{12} = a_1\sigma_{12}$, $d_{j1} = (a_{j-1}/(n_{j-1} - 1))\sigma_{j1} + \phi_J$, and $d_{j2} = (a_{j-1}/(n_{j-1} - 1))\sigma_{j2}$, $j = 2, \ldots, J$.

Hence using (48)−(52), we have the MLE and REMLE of σ_{j1}, σ_{j2} along with their expectations and variances,

$$
\begin{aligned}
&\hat{\sigma}_{11} = \tfrac{1}{a_1}(\hat{d}_{11} - \phi_J), && E(\hat{\sigma}_{11}) = \sigma_{11} - \tfrac{d_{11}}{N_J}, \quad \mathrm{Var}\,(\hat{\sigma}_{11}) = \tfrac{1}{a_1^2}\mathrm{Var}\,(\hat{d}_{11}),\\
&\hat{\sigma}_{j1} = \tfrac{n_{j-1}-1}{a_{j-1}}(\hat{d}_{j1} - \phi_J), && E(\hat{\sigma}_{j1}) = \sigma_{j1},\\
& && \mathrm{Var}\,(\hat{\sigma}_{j1}) = \tfrac{(n_{j-1}-1)^2}{a_{j-1}^2}\mathrm{Var}\,(\hat{d}_{j1}), j = 2, \ldots, J,\\
&\hat{\sigma}_{12} = \tfrac{1}{a_1}\hat{d}_{12}, && E(\hat{\sigma}_{12}) = \sigma_{12} - \tfrac{d_{12}}{N_J}, \quad \mathrm{Var}\,(\hat{\sigma}_{12}) = \tfrac{1}{a_1^2}\mathrm{Var}\,(\hat{d}_{12}),\\
&\hat{\sigma}_{j2} = \tfrac{n_{j-1}-1}{a_{j-1}}\hat{d}_{j2}, && E(\hat{\sigma}_{j2}) = \sigma_{j2},\\
& && \mathrm{Var}\,(\hat{\sigma}_{j2}) = \tfrac{(n_{j-1}-1)^2}{a_{j-1}^2}\mathrm{Var}\,(\hat{d}_{j2}), j = 2, \ldots, J,\\
&\tilde{\sigma}_{11} = \tfrac{1}{a_1}(\tilde{d}_{11} - \phi_J)), && E(\tilde{\sigma}_{11}) = \sigma_{11}, \quad \mathrm{Var}\,(\tilde{\sigma}_{11}) = \tfrac{1}{a_1^2}\mathrm{Var}\,(\tilde{d}_{11}),\\
&\tilde{\sigma}_{j1} = \tfrac{n_{j-1}-1}{a_{j-1}}(\tilde{d}_{j1} - \phi_J), && E(\tilde{\sigma}_{j1}) = \sigma_{j1},\\
& && \mathrm{Var}\,(\tilde{\sigma}_{j1}) = \tfrac{(n_{j-1}-1)^2}{a_{j-1}^2}\mathrm{Var}\,(\tilde{d}_{j1}), j = 2, \ldots, J,\\
&\tilde{\sigma}_{12} = \tfrac{1}{a_1}\tilde{d}_{12}, && E(\tilde{\sigma}_{12}) = \sigma_{12}, \quad \mathrm{Var}\,(\tilde{\sigma}_{12}) = \tfrac{1}{a_1^2}\mathrm{Var}\,(\tilde{d}_{12}),\\
&\tilde{\sigma}_{j2} = \tfrac{n_{j-1}-1}{a_{j-1}}\tilde{d}_{j2}, && E(\tilde{\sigma}_{j2}) = \sigma_{j2},\\
& && \mathrm{Var}\,(\tilde{\sigma}_{j2}) = \tfrac{(n_{j-1}-1)^2}{a_{j-1}^2}\mathrm{Var}\,(\tilde{d}_{j2}), j = 2, \ldots, J,
\end{aligned}
\tag{57}
$$

where we note that $\hat{\sigma}_{ji} = \tilde{\sigma}_{ji}$ for $j = 2, \ldots, J$, $i = 1, 2$.

5. Simulation Study

Here we conduct a Monte Carlo simulation to evaluate the theory and methods concerning the ML and REML estimators in Section 4. For the multiresolution tree structure, we focus on a 4-resolution quad-tree (i.e., $J = 4$, $n_j \equiv 4$, $j = 1, 2, 3$). For the MMTSLM, we consider the case of single-source 3-variable data without missing values, but with mass balance and compound symmetry in the variance structure (i.e., $m = 3$, $\boldsymbol{H}_1 = \boldsymbol{I}, \boldsymbol{H}_j$ are compound symmetric, $j = 2, 3, 4$, and $\boldsymbol{\Sigma}_j$ are compound symmetric, $j = 1, \ldots, 4$). The parameters associated with $\boldsymbol{\Sigma}_j$ are the diagonal entries σ_{j1} and off-diagonal entries σ_{j2}, $j = 1, \ldots, 4$. The value used for the variance of measurement error is set at $\phi_J = 50$.

By varying the number of root nodes on the coarsest resolution (N_1), we vary the size of the data (N). Here we consider $N_1 = 16$, 64, which correspond to data size $N = 1,024$, $4,096$. For each data size, we consider two MMTSLMs, one with constant and the other with regression means for the response variables. In the case of constant means the parameters are $\boldsymbol{\beta} = (\beta_{11}, \beta_{12}, \beta_{13})$, which are the intercepts for the 3 response variables. In the case of regression means, the parameters are $\boldsymbol{\beta} = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \\ \beta_{41} & \beta_{42} & \beta_{43} \end{bmatrix}$. The true parameter values are shown in Tables 1−4.

Table 1. Maximum likelihood estimates (MLE) and restricted maximum likelihood estimates (REML) for a multivariate multiresolution tree-structured spatial linear model (MMTSLM) with a 16-root, 4-resolution, quad-tree structure and with constant means. Reported are the true parameters, relative bias (R-bias), variance, and mean squared error (MSE) based on both theory and 1,000 MLE and REML estimates computed by analytical formulas and numerical maximization.

		Theory			Formula			Maximization		
MLE	Truth	R-Bias	Variance	MSE	R-Bias	Variance	MSE	R-Bias	Variance	MSE
σ_{11}	200.0	-0.06	1605.98	1763.45	-0.07	1440.07	1616.07	-0.24	1128.06	3366.98
σ_{12}	-20.0	-0.06	653.94	655.51	-0.14	652.55	659.94	-0.36	373.24	424.17
σ_{21}	100.0	0.00	148.25	148.25	0.01	152.29	152.66	0.01	280.61	281.32
σ_{22}	-10.0	0.00	60.61	60.61	-0.09	61.02	61.72	-0.04	95.01	95.06
σ_{31}	50.0	0.00	12.42	12.42	0.01	12.93	13.02	0.01	19.43	19.73
σ_{32}	5.0	0.00	7.28	7.28	-0.01	6.85	6.84	-0.04	20.35	20.37
σ_{41}	25.0	0.00	3.40	3.40	0.01	3.58	3.63	0.01	5.60	5.65
σ_{42}	2.5	0.00	1.84	1.84	-0.02	1.77	1.77	-0.01	1.78	1.78
β_{11}	40.0	0.00	12.55	12.55	0.00	12.81	12.80	0.00	12.88	12.87
β_{12}	20.0	0.00	12.55	12.55	-0.02	13.46	13.46	-0.01	13.48	13.48
β_{13}	10.0	0.00	12.55	12.55	-0.01	13.07	13.07	-0.01	13.12	13.12
REML	Truth	R-Bias	Variance	MSE	R-Bias	Variance	MSE	R-Bias	Variance	MSE
σ_{11}	200.0	0.00	1827.25	1827.25	-0.00	1638.47	1637.51	-0.23	1180.49	3289.35
σ_{12}	-20.0	0.00	744.04	744.04	-0.09	742.46	744.58	-0.36	397.03	447.71
σ_{21}	100.0	0.00	148.25	148.25	0.01	152.29	152.66	0.02	271.95	274.32
σ_{22}	-10.0	0.00	60.61	60.61	-0.09	61.02	61.72	-0.05	96.15	96.28
σ_{31}	50.0	0.00	12.42	12.42	0.01	12.93	13.02	0.02	21.40	22.19
σ_{32}	5.0	0.00	7.28	7.28	-0.01	6.85	6.84	-0.05	26.40	26.45
σ_{41}	25.0	0.00	3.40	3.40	0.01	3.58	3.63	0.01	5.63	5.66
σ_{42}	2.5	0.00	1.84	1.84	-0.02	1.77	1.77	-0.03	2.69	2.69
β_{11}	40.0	0.00	12.55	12.55	0.00	12.81	12.80	0.00	12.85	12.84
β_{12}	20.0	0.00	12.55	12.55	-0.01	13.46	13.46	-0.01	13.48	13.47
β_{13}	10.0	0.00	12.55	12.55	-0.01	13.07	13.07	-0.01	13.11	13.11

Table 2. Maximum likelihood estimates (MLE) and restricted maximum likelihood estimates (REML) for a multivariate multiresolution tree-structured spatial linear model (MMTSLM) with a 16-root, 4-resolution, quad-tree structure and with a regression mean. Reported are the true parameters, relative bias (R-bias), variance, and mean squared error (MSE) based on 1,000 MLE and REML estimates computed by numerical maximization.

		MLE			REML		
	Truth	R-Bias	Variance	MSE	R-Bias	Variance	MSE
σ_{11}	200.0	-0.36	987.63	6290.50	-0.37	979.96	6357.72
σ_{12}	-20.0	-0.60	358.16	500.70	-0.58	366.44	502.28
σ_{21}	100.0	0.02	209.56	214.16	0.03	227.42	235.52
σ_{22}	-10.0	-0.05	123.01	123.11	-0.04	107.64	107.72
σ_{31}	50.0	0.06	30.92	39.26	0.07	27.12	40.96
σ_{32}	5.0	-0.03	22.37	22.36	-0.05	38.89	38.91
σ_{41}	25.0	0.02	24.90	25.25	0.02	12.38	12.59
σ_{42}	2.5	-0.04	29.10	29.08	0.04	8.72	8.72
β_{11}	100.0	0.00	20.69	20.72	0.00	20.56	20.58
β_{12}	50.0	0.00	21.20	21.18	0.00	21.11	21.09
β_{13}	25.0	-0.01	21.61	21.70	-0.01	21.38	21.47
β_{21}	40.0	0.00	0.00	0.00	0.00	0.00	0.00
β_{22}	10.0	0.00	0.00	0.00	0.00	0.00	0.00
β_{23}	20.0	0.00	0.00	0.00	0.00	0.00	0.00
β_{31}	20.0	0.00	0.00	0.00	0.00	0.00	0.00
β_{32}	40.0	0.00	0.00	0.00	0.00	0.00	0.00
β_{33}	10.0	0.00	0.00	0.00	0.00	0.00	0.00
β_{41}	10.0	0.00	2.17	2.17	0.00	2.17	2.17
β_{42}	20.0	0.00	1.94	1.94	0.00	1.92	1.92
β_{43}	40.0	0.00	2.10	2.09	0.00	2.07	2.07

For each of the four cases (two sample sizes and two types of regression means), we simulate $S = 1,000$ data sets based on the corresponding MMTSLM evaluated at the true parameter values. For each data set, we compute the ML and REML estimates using numerical maximization. In addition, in the case of constant means, we have the explicit formulas (57) for the ML and REML estimates. Based on $S = 1,000$ ML and REML estimates, we compute an estimate of the mean and variance of the ML and REML estimates. Using these mean and variance estimates, we obtain an estimate of the relative bias (R-bias), variance, and mean squared error (MSE) of the ML and REML estimates. Again in the case of constant means, we have the explicit formulas (57) for the R-bias, variance, and MSE for the ML and REML estimates. Here the empirical relative bias is defined as the average estimates minus the true parameter value and divided by the true parameter value.

For the constant-mean MMTSLM, the results are shown in Tables 1 and 3. First, the results suggest that our analytical results are correct, as the ML and REML estimates using the explicit formulas match well with the empirical R-bias, variance, and MSE. Second, the theoretical and empirical results match quite well with the ML and REML estimates obtained from numerical maximization, except for some of the variance parameters on the coarser resolutions. Thus the maximization procedure works reasonably well and so does the change-of-resolution Kalman filter algorithm we use to evaluate the loglikelihood functions. We suspect that the under-performance of the variance estimates on the coarser resolutions is due to the smaller number of nodes on these resolutions. Finally, as the data size increases, there is a decrease in the R-bias, variance, and MSE, as one would expect.

Table 3. Maximum likelihood estimates (MLE) and restricted maximum likelihood estimates (REML) for a multivariate multiresolution tree-structured spatial linear model (MMTSLM) with a 64-root, 4-resolution, quad-tree structure and with constant means. Reported are the true parameters, relative bias (R-bias), variance, and mean squared error (MSE) based on both theory and 1,000 MLE and REML estimates computed by analytical formulas and numerical maximization.

		Theory			Formula			Maximization		
MLE	Truth	R-Bias	Variance	MSE	R-Bias	Variance	MSE	R-Bias	Variance	MSE
σ_{11}	200.0	-0.02	421.57	431.41	-0.03	420.37	465.73	-0.12	609.82	1176.96
σ_{12}	-20.0	-0.02	171.66	171.76	-0.01	197.42	197.26	-0.16	156.26	166.04
σ_{21}	100.0	0.00	37.06	37.06	0.00	40.20	40.16	0.00	131.07	130.94
σ_{22}	-10.0	0.00	15.15	15.15	0.00	15.15	15.14	0.01	18.33	18.31
σ_{31}	50.0	0.00	3.10	3.10	-0.00	3.46	3.45	0.00	5.24	5.24
σ_{32}	5.0	0.00	1.82	1.82	0.03	1.91	1.92	0.03	1.96	1.98
σ_{41}	25.0	0.00	0.85	0.85	0.00	0.83	0.83	0.00	1.10	1.10
σ_{42}	2.5	0.00	0.46	0.46	-0.01	0.49	0.49	-0.01	0.50	0.50
β_{11}	40.0	0.00	3.14	3.14	0.00	3.45	3.45	0.00	3.48	3.48
β_{12}	20.0	0.00	3.14	3.14	-0.01	3.50	3.52	-0.01	3.52	3.53
β_{13}	10.0	0.00	3.14	3.14	0.00	3.44	3.44	0.00	3.42	3.42
REML	Truth	R-Bias	Variance	MSE	R-Bias	Variance	MSE	R-Bias	Variance	MSE
σ_{11}	200.0	0.00	435.06	435.06	-0.02	433.83	446.98	-0.12	599.62	1208.84
σ_{12}	-20.0	0.00	177.15	177.16	0.01	203.74	203.55	-0.17	155.13	165.99
σ_{21}	100.0	0.00	37.06	37.06	0.00	40.20	40.16	0.00	133.49	133.39
σ_{22}	-10.0	0.00	15.15	15.15	0.00	15.15	15.14	0.01	18.62	18.62
σ_{31}	50.0	0.00	3.10	3.10	-0.00	3.46	3.45	0.00	5.07	5.09
σ_{32}	5.0	0.00	1.82	1.82	0.03	1.91	1.92	0.03	2.03	2.06
σ_{41}	25.0	0.00	0.85	0.85	0.00	0.83	0.83	0.00	1.13	1.13
σ_{42}	2.5	0.00	0.46	0.46	-0.01	0.49	0.49	-0.00	0.50	0.50
β_{11}	40.0	0.00	3.14	3.14	0.00	3.45	3.45	0.00	3.48	3.48
β_{12}	20.0	0.00	3.14	3.14	-0.01	3.50	3.52	-0.01	3.52	3.53
β_{13}	10.0	0.00	3.14	3.14	0.00	3.44	3.44	0.00	3.42	3.42

Table 4. Maximum likelihood estimates (MLE) and restricted maximum likelihood estimates (REML) for a multivariate multiresolution tree-structured spatial linear model (MMTSLM) with a 64-root, 4-resolution, quad-tree structure and with a regression mean. Reported are the true parameters, relative bias (R-bias), variance, and mean squared error (MSE) based on 1,000 MLE and REML estimates computed by numerical maximization.

		MLE			REML		
	Truth	R-Bias	Variance	MSE	R-Bias	Variance	MSE
σ_{11}	200.0	-0.14	614.92	1386.45	-0.13	617.42	1329.76
σ_{12}	-20.0	-0.18	169.45	182.24	-0.18	168.55	180.58
σ_{21}	100.0	0.01	142.16	142.91	0.01	143.88	145.00
σ_{22}	-10.0	0.02	26.42	26.46	0.02	22.01	22.02
σ_{31}	50.0	0.01	6.35	6.80	0.02	4.70	5.38
σ_{32}	5.0	0.03	5.51	5.53	0.02	2.39	2.40
σ_{41}	25.0	0.00	1.12	1.13	0.00	1.04	1.04
σ_{42}	2.5	0.00	0.56	0.56	0.01	0.56	0.56
β_{11}	100.0	0.00	4.84	4.88	0.00	4.86	4.89
β_{12}	50.0	0.00	4.82	4.82	0.00	4.78	4.77
β_{13}	25.0	0.00	4.75	4.74	0.00	4.75	4.74
β_{21}	40.0	0.00	0.00	0.00	0.00	0.00	0.00
β_{22}	10.0	0.00	0.00	0.00	0.00	0.00	0.00
β_{23}	20.0	0.00	0.00	0.00	0.00	0.00	0.00
β_{31}	20.0	0.00	0.00	0.00	0.00	0.00	0.00
β_{32}	40.0	0.00	0.00	0.00	0.00	0.00	0.00
β_{33}	10.0	0.00	0.00	0.00	0.00	0.00	0.00
β_{41}	10.0	0.00	0.33	0.33	0.00	0.33	0.33
β_{42}	20.0	0.00	0.30	0.30	0.00	0.31	0.30
β_{43}	40.0	0.00	0.33	0.33	0.00	0.33	0.33

For the regression-mean MMTSLM, the results are shown in Tables 2 and 4. Here we do not have theoretical results to compare to, but we can still evaluate the performance of the ML and REML estimates. Overall the ML and REML estimates have small R-bias, except for the coarsest resolution. Again as the data size increases, there is a decrease in the R-bias, variance, and MSE. Our experience suggests that the bias in estimating σ_{12} is a consequence of a relatively large bias in the estimate of σ_{11}. Finally, there seems very little difference between the MLE and REML estimates for the two sample sizes under consideration.

Acknowledgement

Funding has been provided for this research by the USDA Cooperative State Research, Education and Extension Service (CSREES) Hatch project WIS04676.

We thank two referees, an associate editor, and the Editor for helpful comments that improved the presentation of the paper.

Appendix I. General Optimal Prediction Theory

For $\binom{\boldsymbol{Y}}{\boldsymbol{Z}} \sim \left[\binom{\boldsymbol{\mu}_Y}{\boldsymbol{\mu}_Z}, \begin{pmatrix} \boldsymbol{\Sigma}_{YY} & \boldsymbol{\Sigma}_{YZ} \\ \boldsymbol{\Sigma}_{ZY} & \boldsymbol{\Sigma}_{ZZ} \end{pmatrix}\right]$, we obtain $\boldsymbol{m}_{Y|Z}$ by minimizing $\|\boldsymbol{Y}-\boldsymbol{Z}_0\|$ for all $\boldsymbol{Z}_0 \in \overline{sp}\{\boldsymbol{Z}\}^m$, where $\boldsymbol{Z}_0 = \boldsymbol{\mu} + \boldsymbol{\beta Z}$ with $\boldsymbol{\mu} \in IR^m$, $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m]'$, and $\boldsymbol{\beta}_i \in IR^n$, $i = 1, \ldots, m$. It is straightforward to show that $f(\boldsymbol{\mu}, \boldsymbol{\beta}) \equiv \|\boldsymbol{Y}-\boldsymbol{Z}_0\|^2 = \boldsymbol{\mu}_Y'\boldsymbol{\mu}_Y - 2\boldsymbol{\mu}'\boldsymbol{\mu}_Y + \boldsymbol{\mu}'\boldsymbol{\mu} + 2\boldsymbol{\mu}'\boldsymbol{\beta\mu}_Z + \mathrm{tr}\{-2\boldsymbol{\beta\mu}_Z\boldsymbol{\mu}_Y' + \boldsymbol{\beta\mu}_Z\boldsymbol{\mu}_Z'\boldsymbol{\beta}' + \boldsymbol{\Sigma}_{YY} - 2\boldsymbol{\beta\Sigma}_{ZY} + \boldsymbol{\beta\Sigma}_{ZZ}\boldsymbol{\beta}'\}$. We minimize $f(\boldsymbol{\mu}, \boldsymbol{\beta})$ by taking the first-order partial derivatives with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\beta}$ using Lemma 1 Harville (1997, Chap. 15).

Lemma 1. *For an m-dimensional column vector $\boldsymbol{\mu}$, an $m \times n$ matrix $\boldsymbol{\beta}$, and a conformable matrix $\boldsymbol{A}$ that does not depend on $\boldsymbol{\mu}$ and $\boldsymbol{\beta}$:*

(i) $\dfrac{\partial \boldsymbol{\mu}'\boldsymbol{A}}{\partial \boldsymbol{\mu}} = \boldsymbol{A}$; (ii) $\dfrac{\partial \boldsymbol{\mu}'\boldsymbol{A\mu}}{\partial \boldsymbol{\mu}} = (\boldsymbol{A} + \boldsymbol{A}')\boldsymbol{\mu}$; (iii) $\dfrac{\partial\, \mathrm{tr}(\boldsymbol{A\beta}')}{\partial \boldsymbol{\beta}} = \dfrac{\partial\, \mathrm{tr}(\boldsymbol{\beta A}')}{\partial \boldsymbol{\beta}} = \boldsymbol{A}$;
(iv) $\dfrac{\partial\, \mathrm{tr}(\boldsymbol{\beta A \beta}')}{\partial \boldsymbol{\beta}} = \boldsymbol{\beta}(\boldsymbol{A} + \boldsymbol{A}')$.

Proof of Theorem 1. By Lemma 1, we have

$$\frac{\partial f(\boldsymbol{\mu}, \boldsymbol{\beta})}{\partial \boldsymbol{\mu}} = 2(\boldsymbol{\mu} - \boldsymbol{\mu}_Y + \boldsymbol{\beta\mu}_Z),$$
$$\frac{\partial f(\boldsymbol{\mu}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 2(\boldsymbol{\mu} - \boldsymbol{\mu}_Y + \boldsymbol{\beta\mu}_Z)\boldsymbol{\mu}_Z' - 2\boldsymbol{\Sigma}_{YZ} + 2\boldsymbol{\beta\Sigma}_{ZZ}.$$

Setting the partial derivatives to zero, we obtain the normal equations and their equivalence:

$$\boldsymbol{\mu} = \boldsymbol{\mu}_Y - \boldsymbol{\beta\mu}_Z, \qquad \boldsymbol{\Sigma}_{ZZ}\boldsymbol{\beta}' = \boldsymbol{\Sigma}_{ZY}, \tag{58}$$
$$\boldsymbol{\mu} = \boldsymbol{\mu}_Y - \boldsymbol{\beta\mu}_Z, \qquad \boldsymbol{\Sigma}_{ZZ}\boldsymbol{\beta}_i = \boldsymbol{\Sigma}_{Zy_i}, i = 1, \ldots, m. \tag{59}$$

For any optimal linear predictor $\boldsymbol{m}_{Y|Z} = \hat{\boldsymbol{\mu}} + \hat{\boldsymbol{\beta}}\boldsymbol{Z}$, $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\beta}}$ must satisfy the normal equations (58). Then $E(\boldsymbol{m}_{Y|Z}) = E(\hat{\boldsymbol{\mu}} + \hat{\boldsymbol{\beta}}\boldsymbol{Z}) = E(\boldsymbol{\mu}_Y - \hat{\boldsymbol{\beta}}\boldsymbol{\mu}_Z + \hat{\boldsymbol{\beta}}\boldsymbol{Z}) = \boldsymbol{\mu}_Y + \hat{\boldsymbol{\beta}}E(\boldsymbol{Z} - \boldsymbol{\mu}_Z) = \boldsymbol{\mu}_Y$ and $\mathrm{Cov}\,(\boldsymbol{Z},\ \boldsymbol{Y} - \boldsymbol{m}_{Y|Z}) = \mathrm{Cov}\,(\boldsymbol{Z},\ \boldsymbol{Y} - \boldsymbol{\mu}_Y + \hat{\boldsymbol{\beta}}\boldsymbol{\mu}_Z - \hat{\boldsymbol{\beta}}\boldsymbol{Z}) = \boldsymbol{\Sigma}_{ZY} - \boldsymbol{\Sigma}_{ZZ}\hat{\boldsymbol{\beta}}' = \boldsymbol{0}_{n\times m}$ where the last equality holds because of (58).

Suppose there exists another optimal linear predictor $\tilde{\boldsymbol{m}}_{Y|Z}$ of $\boldsymbol{Y}$ given $\boldsymbol{Z}$, then let $\boldsymbol{Y} = \boldsymbol{m}_{Y|Z} + \boldsymbol{e}$ and $\boldsymbol{Y} = \tilde{\boldsymbol{m}}_{Y|Z} + \tilde{\boldsymbol{e}}$. We have $E(\boldsymbol{m}_{Y|Z}) = E(\tilde{\boldsymbol{m}}_{Y|Z}) = \boldsymbol{\mu}_Y$ and $\mathrm{Cov}\,(\boldsymbol{m}_{Y|Z} - \tilde{\boldsymbol{m}}_{Y|Z}, \boldsymbol{e} - \tilde{\boldsymbol{e}}) = \boldsymbol{0}_{m\times m}$ because $\boldsymbol{m}_{Y|Z} - \tilde{\boldsymbol{m}}_{Y|Z} \in \overline{sp}\{\boldsymbol{Z}\}^m$ and $\mathrm{Cov}\,(\boldsymbol{Z}, \boldsymbol{e} - \tilde{\boldsymbol{e}}) = \boldsymbol{0}_{n\times m}$. Then $\boldsymbol{0}_{m\times m} = \mathrm{Var}\,(\boldsymbol{Y} - \boldsymbol{Y}) = \mathrm{Var}\,[(\boldsymbol{m}_{Y|Z} - \tilde{\boldsymbol{m}}_{Y|Z}) +$

$(\boldsymbol{e}-\tilde{\boldsymbol{e}})] = \mathrm{Var}\,(\boldsymbol{m}_{Y|Z}-\tilde{\boldsymbol{m}}_{Y|Z}) + \mathrm{Var}\,(\boldsymbol{e}-\tilde{\boldsymbol{e}})$. Comparing the diagonal elements of both sides, since the variance of any random variable in L^2 is non-negative, we obtain $\mathrm{Var}\,(m_{y_i|Z}-\tilde{m}_{y_i|Z}) = 0$ for $i = 1,\ldots,m$ which implies that $m_{y_i|Z} = \tilde{m}_{y_i|Z}$, or, $\boldsymbol{m}_{Y|Z} = \tilde{\boldsymbol{m}}_{Y|Z}$. Hence the optimal linear predictor of $\boldsymbol{Y}$ given $\boldsymbol{Z}$ is unique.

The following lemma establishes the consistency of the the normal equations (59).

Lemma 2. *The normal equations $\boldsymbol{\Sigma}_{ZZ}\boldsymbol{\beta}_i = \boldsymbol{\Sigma}_{Zy_i}$ are consistent, where $\boldsymbol{\Sigma}_{Zy_i} \equiv \mathrm{Cov}\,(\boldsymbol{Z}, y_i)$ for $i = 1,\ldots,m$.*

Proof. To show that the normal equations are consistent, it suffices to show that for any $\boldsymbol{\alpha} \in I\!R^n$, if $\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{ZZ} = \boldsymbol{0}_n$, then $\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{Zy_i} = 0$. Suppose for $\boldsymbol{\alpha} \in I\!R^n$, $\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{ZZ} = \boldsymbol{0}_n$, then $\mathrm{Var}\,(\boldsymbol{\alpha}'\boldsymbol{Z}) = \boldsymbol{\alpha}'\boldsymbol{\Sigma}_{ZZ}\boldsymbol{\alpha} = 0$, which implies that $\boldsymbol{\alpha}'\boldsymbol{Z} = \boldsymbol{\alpha}'\boldsymbol{\mu}_Z$. Hence $\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{Zy_i} = \boldsymbol{\alpha}'\,\mathrm{Cov}\,(\boldsymbol{Z}, y_i) = \mathrm{Cov}\,(\boldsymbol{\alpha}'\boldsymbol{Z}, y_i) = \mathrm{Cov}\,(\boldsymbol{\alpha}'\boldsymbol{\mu}_Z, y_i) = 0$.

Now recall (Searle (1997, Chap. 1.3)) that the Moore-Penrose pseudo inverse of a matrix $\boldsymbol{\Sigma}$ is the unique matrix $\boldsymbol{\Sigma}^+$ which satisfies the following four conditions:

$$\boldsymbol{\Sigma}\boldsymbol{\Sigma}^+\boldsymbol{\Sigma} = \boldsymbol{\Sigma},\ \boldsymbol{\Sigma}^+\boldsymbol{\Sigma}\boldsymbol{\Sigma}^+ = \boldsymbol{\Sigma}^+,\ (\boldsymbol{\Sigma}^+\boldsymbol{\Sigma})' = \boldsymbol{\Sigma}^+\boldsymbol{\Sigma} \quad \text{and} \quad (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^+)' = \boldsymbol{\Sigma}\boldsymbol{\Sigma}^+. \tag{60}$$

Furthermore, by transposing both sides of the four conditions in (60), we obtain

$$(\boldsymbol{\Sigma}')^+ = (\boldsymbol{\Sigma}^+)'. \tag{61}$$

Proof of Theorem 2. Since the normal equations (59) are consistent, from Theorem 1 of Chapter 1.6 of Searl (1997), one of the solutions of the normal equations and its equivalence is

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \boldsymbol{\mu}_Y - \hat{\boldsymbol{\beta}}'\boldsymbol{\mu}_Z, & \hat{\boldsymbol{\beta}}_i &= \boldsymbol{\Sigma}_{ZZ}^+\boldsymbol{\Sigma}_{Zy_i}, \\ \hat{\boldsymbol{\mu}} &= \boldsymbol{\mu}_Y - \hat{\boldsymbol{\beta}}'\boldsymbol{\mu}_Z, & \hat{\boldsymbol{\beta}} &= (\boldsymbol{\Sigma}_{ZZ}^+\boldsymbol{\Sigma}_{ZY})'. \end{aligned} \tag{62}$$

Then we have $\boldsymbol{m}_{Y|Z} = \hat{\boldsymbol{\mu}} + \hat{\boldsymbol{\beta}}\boldsymbol{Z} = \boldsymbol{\mu}_Y + \hat{\boldsymbol{\beta}}(\boldsymbol{Z}-\boldsymbol{\mu}_Z)$ and

$$\boldsymbol{C}_{Y|Z} = E[(\boldsymbol{Y}-\boldsymbol{m}_{Y|Z})(\boldsymbol{Y}-\boldsymbol{m}_{Y|Z})'] = \mathrm{Var}\,(\boldsymbol{Y}-\hat{\boldsymbol{\beta}}\boldsymbol{Z}) = \Sigma_{YY} - \boldsymbol{\Sigma}_{YZ}\hat{\boldsymbol{\beta}}',$$

due to (58).

Since $\boldsymbol{\Sigma}_{ZZ}$ is symmetric, using (61) and the uniqueness of Moore-Penrose pseudo inverse, we have $(\boldsymbol{\Sigma}_{ZZ}^+)' = \boldsymbol{\Sigma}_{ZZ}^+$, i.e., $\boldsymbol{\Sigma}_{ZZ}^+$ is symmetric. Then we obtain $\boldsymbol{m}_{Y|Z} = \boldsymbol{\mu}_Y + \boldsymbol{\Sigma}_{YZ}\boldsymbol{\Sigma}_{ZZ}^+(\boldsymbol{Z}-\boldsymbol{\mu}_Z)$ and $\boldsymbol{C}_{Y|Z} = \Sigma_{YY} - \boldsymbol{\Sigma}_{YZ}\boldsymbol{\Sigma}_{ZZ}^+\boldsymbol{\Sigma}_{ZY}$. When $\boldsymbol{Y}$

and $\boldsymbol{Z}$ are normally distributed, $\boldsymbol{m}_{Y|Z}$ and $\boldsymbol{C}_{Y|Z}$ are the conditional mean and conditional variance respectively (Lauritzen (1996, Proposition C.5)).

Appendix II. Generalized Change-of-Resolution Kalman Filter

First, we recall and introduce some notation. Write $\{j',k'\} \prec \{j,k\}$ if $\{j',k'\}$ is a descendant vector node of $\{j,k\}$. Here a node is assumed to be a descendant of itself. Further,

$$\begin{aligned}
\gamma_{j,k} &\equiv \mathcal{I}\{\boldsymbol{Z}_{j,k} \ \text{ is } \ \text{observed}\} = \begin{cases} 1; \text{if } \boldsymbol{Z}_{j,k} \text{ is observed,} \\ 0; \text{otherwise,} \end{cases} \\
\boldsymbol{Z} &\equiv \{\boldsymbol{Z}_{j,k} : \gamma_{j,k} = 1\}, \\
\boldsymbol{Z}_{de\{j,k\}} &\equiv \Big\{\boldsymbol{Z}_{j',k'} : \gamma_{j',k'} = 1, \{j',k'\} \prec \{j,k\}\Big\}, \\
\boldsymbol{Z}^*_{de\{j,k\}} &\equiv \Big\{\boldsymbol{Z}_{j',k'} : \gamma_{j',k'} = 1, \{j',k'\} \prec \{j,k\}, \{j',k'\} \neq \{j,k\}\Big\}, \\
\boldsymbol{Z}^c_{de\{j,k\}} &\equiv \boldsymbol{Z} \setminus \boldsymbol{Z}_{de\{j,k\}}, \\
\boldsymbol{U}_{j,k}|\boldsymbol{Z}_{de\{j',k'\}} &\sim [\hat{\boldsymbol{U}}_{j,k|j',k'}, \hat{\boldsymbol{V}}_{j,k|j',k'}], \\
\boldsymbol{U}_{j,k}|\boldsymbol{Z}^*_{de\{j',k'\}} &\sim [\hat{\boldsymbol{U}}^*_{j,k|j',k'}, \hat{\boldsymbol{V}}^*_{j,k|j',k'}], \\
\boldsymbol{U}_{j,k}|\boldsymbol{Z} &\sim [\hat{\boldsymbol{U}}_{j,k}, \hat{\boldsymbol{V}}_{j,k}], \\
\boldsymbol{V}_{j,k} &\equiv \mathrm{Var}\,(\boldsymbol{U}_{j,k}), \\
\boldsymbol{V}_{j,k,j',k'} &\equiv \mathrm{Cov}\,(\boldsymbol{U}_{j,k}, \boldsymbol{U}_{j',k'}), \\
\boldsymbol{B}_{j,k} &\equiv \boldsymbol{V}_{pa\{j,k\}}\boldsymbol{A}'_{j,k}\boldsymbol{V}^+_{j,k}, \\
\boldsymbol{R}_{j,k} &\equiv \boldsymbol{V}_{pa\{j,k\}} - \boldsymbol{V}_{pa\{j,k\}}\boldsymbol{A}'_{j,k}\boldsymbol{V}^+_{j,k}\boldsymbol{A}_{j,k}\boldsymbol{V}_{pa\{j,k\}}, \\
\boldsymbol{J}_{j,k} &\equiv \hat{\boldsymbol{V}}_{j,k|j,k}\boldsymbol{B}'_{j,k}\hat{\boldsymbol{V}}^+_{pa\{j,k\}|j,k}.
\end{aligned}$$

Before deriving the generalized change-of-resolution Kalman filter algorithm, we present some useful results about matrix operations and Moore-Penrose pseudo inverse in the following lemma.

Lemma 3. *For matrices* $\boldsymbol{A}$ *and* $\boldsymbol{B}$*:*

(i) *If* $\boldsymbol{A}$ *is an* $n \times n$ *symmetric positive semi-definite matrix, then there exists an* $n \times m$ *matrix* $\boldsymbol{L}$ *with full column rank such that* $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}'$*, where* $m = \mathrm{Rank}(\boldsymbol{A})$.

(ii) *If* $\boldsymbol{A}$ *is an* $n \times n$ *positive semi-definite matrix and* $\boldsymbol{B}$ *is an* $n \times n$ *positive definite matrix, then* $\boldsymbol{A} + \boldsymbol{B}$ *is invertible.*

(iii) *If* $\boldsymbol{A}$ *is an* $m \times n$ *matrix,* $\boldsymbol{B}$ *is an* $n \times m$ *matrix, and* $\boldsymbol{I}_n + \boldsymbol{B}\boldsymbol{A}$ *is invertible, then* $(\boldsymbol{I}_m + \boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{I}_m - \boldsymbol{A}(\boldsymbol{I}_n + \boldsymbol{B}\boldsymbol{A})^{-1}\boldsymbol{B}$.

(iv) *If* $\boldsymbol{A}$ *is an* $n \times m$ *matrix with full column rank and* $\boldsymbol{B}$ *is an* $m \times n$ *matrix with full row rank, then* $(\boldsymbol{A}\boldsymbol{B})^+ = \boldsymbol{B}^+\boldsymbol{A}^+$.

(v) *If* $\boldsymbol{A}$ *is an* $n \times m$ *matrix, then* $\boldsymbol{A}^{+}\boldsymbol{A} = \boldsymbol{P}_{A'}$ *and* $\boldsymbol{A}\boldsymbol{A}^{+} = \boldsymbol{P}_{A}$*, where* $\boldsymbol{P}_{A'} \equiv \boldsymbol{A}'(\boldsymbol{A}\boldsymbol{A}')^{-}\boldsymbol{A}$ *and* $\boldsymbol{P}_{A} \equiv \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-}\boldsymbol{A}'$ *are the projections matrices corresponding to* $\boldsymbol{A}'$ *and* $\boldsymbol{A}$*, where* $(\boldsymbol{A}'\boldsymbol{A})^{-}$ *denotes the generalized inverse of* $\boldsymbol{A}'\boldsymbol{A}$*. Moreover* $\boldsymbol{P}_A\boldsymbol{A} = \boldsymbol{A}$, $\boldsymbol{A}'\boldsymbol{P}_A = \boldsymbol{A}'$, $\boldsymbol{A}^{+}\boldsymbol{A}\boldsymbol{A}' = \boldsymbol{A}'\boldsymbol{A}\boldsymbol{A}^{+} = \boldsymbol{A}'$ *and* $(\boldsymbol{A}')^{+} = (\boldsymbol{A}^{+})'$*. For matrix* $\boldsymbol{B}$*, if* $\mathrm{Col}(\boldsymbol{A}) = \mathrm{Col}(\boldsymbol{B})$*, then* $\boldsymbol{P}_A = \boldsymbol{P}_B$ *where* $\mathrm{Col}(\boldsymbol{A})$ *denotes the column space of* $\boldsymbol{A}$*.*

(vi) *If* $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{M}\boldsymbol{U} + \boldsymbol{e}$, $\boldsymbol{U} \sim [\boldsymbol{0}, \boldsymbol{V}]$, $\boldsymbol{e} \sim [\boldsymbol{0}, \boldsymbol{P}]$*, where* $\boldsymbol{X}$, $\boldsymbol{B}$ *and* $\boldsymbol{M}$ *are deterministic matrices,* $\boldsymbol{U}$ *and* $\boldsymbol{e}$ *are uncorrelated , and* $\boldsymbol{P}$ *is invertible, then* $\boldsymbol{C}_{U|Z} = \boldsymbol{L}(\boldsymbol{I} + \boldsymbol{L}'\boldsymbol{M}'\boldsymbol{P}^{-1}\boldsymbol{M}\boldsymbol{L})^{-1}\boldsymbol{L}'$, $\boldsymbol{m}_{U|Z} = \boldsymbol{C}_{U|Z}\boldsymbol{M}'\boldsymbol{P}^{-1}(\boldsymbol{Z} - \boldsymbol{X}\boldsymbol{B})$, $\boldsymbol{C}^{+}_{U|Z}\boldsymbol{m}_{U|Z} = (\boldsymbol{L}')^{+}\boldsymbol{L}'\boldsymbol{M}'\boldsymbol{P}^{-1}(\boldsymbol{Z} - \boldsymbol{X}\boldsymbol{B})$*, and* $\boldsymbol{C}_{U|Z}\boldsymbol{C}^{+}_{U|Z}\boldsymbol{m}_{U|Z} = \boldsymbol{m}_{U|Z}$ *where* $\boldsymbol{V} = \boldsymbol{L}\boldsymbol{L}'$ *and* $\boldsymbol{L}$ *has full column rank.*

Proof. See details in Yue and Zhu (2005).

In the high-to-low-resolution filtering step, we start with the finest resolution J.

Proof of (16)−(17). For a leaf node $\{J, k\}$, $k = 1, \ldots, N_{J-1}$, if $\gamma_{J,k} = 1$, we have

$$\begin{pmatrix} \boldsymbol{U}_{J,k} \\ \boldsymbol{Z}_{J,k} \end{pmatrix} \sim \left[\begin{pmatrix} \boldsymbol{0}_{n_{J-1}m} \\ \boldsymbol{X}_{J,k}\boldsymbol{B} \end{pmatrix}, \begin{pmatrix} \boldsymbol{V}_{J,k} & \boldsymbol{V}_{J,k} \\ \boldsymbol{V}_{J,k} & \boldsymbol{V}_{J,k} + \boldsymbol{I}_{n_{J-1}} \otimes \boldsymbol{\Phi}_J \end{pmatrix} \right].$$

From Theorem 2, we have optimal linear predictor $\boldsymbol{U}_{J,k}|\boldsymbol{Z}_{J,k} \sim [\boldsymbol{V}_{J,k}(\boldsymbol{V}_{J,k} + \boldsymbol{I}_{n_{J-1}} \otimes \boldsymbol{\Phi}_J)^{-1}(\boldsymbol{Z}_{J,k} - \boldsymbol{X}_{J,k}\boldsymbol{B}), \boldsymbol{V}_{J,k} - \boldsymbol{V}_{J,k}(\boldsymbol{V}_{J,k} + \boldsymbol{I}_{n_{J-1}} \otimes \boldsymbol{\Phi}_J)^{-1}\boldsymbol{V}_{J,k}]$, where $(\boldsymbol{V}_{J,k} + \boldsymbol{I}_{n_{J-1}} \otimes \boldsymbol{\Phi}_J)$ is invertible because of Lemma 3 (ii). If $\gamma_{J,k} = 0$, we have $\hat{\boldsymbol{U}}_{J,k} = 0$ and $\hat{\boldsymbol{V}}_{J,k|J,k} = \boldsymbol{V}_{J,k}$. Hence for a leaf node $\{J, k\}$,

$$\begin{aligned} \hat{\boldsymbol{U}}_{J,k|J,k} &= \gamma_{J,k}\boldsymbol{V}_{J,k}(\boldsymbol{V}_{J,k} + \boldsymbol{I}_{n_{J-1}} \otimes \boldsymbol{\Phi}_J)^{-1}(\boldsymbol{Z}_{J,k} - \boldsymbol{X}_{J,k}\boldsymbol{B}), \\ \hat{\boldsymbol{V}}_{J,k|J,k} &= \boldsymbol{V}_{J,k} - \gamma_{J,k}\boldsymbol{V}_{J,k}(\boldsymbol{V}_{J,k} + \boldsymbol{I}_{n_{J-1}} \otimes \boldsymbol{\Phi}_J)^{-1}\boldsymbol{V}_{J,k}. \end{aligned}$$

Now we move from the resolution $j = J - 1$ to the coarsest resolution $j = 1$.

Proof of (18)−(19). From (12), we have

$$\begin{pmatrix} \boldsymbol{U}_{pa\{j,k\}} \\ \boldsymbol{U}_{j,k} \end{pmatrix} \sim \left[\begin{pmatrix} \boldsymbol{0}_{n_{j-2}m} \\ \boldsymbol{0}_{n_{j-1}}m \end{pmatrix}, \begin{pmatrix} \boldsymbol{V}_{pa\{j,k\}} & \boldsymbol{V}_{pa\{j,k\}}\boldsymbol{A}'_{j,k} \\ \boldsymbol{A}_{j,k}\boldsymbol{V}_{pa\{j,k\}} & \boldsymbol{V}_{j,k} \end{pmatrix} \right].$$

Then $\boldsymbol{U}_{pa\{j,k\}}|\boldsymbol{U}_{j,k} \sim [\boldsymbol{B}_{j,k}\boldsymbol{U}_{j,k}, \boldsymbol{R}_{j,k}]$. Hence we have $\boldsymbol{U}_{pa\{j,k\}} = \boldsymbol{B}_{j,k}\boldsymbol{U}_{j,k} + \boldsymbol{\xi}_{j,k}$ where $\boldsymbol{\xi}_{j,k} = \boldsymbol{U}_{pa\{j,k\}} - \boldsymbol{B}_{j,k}\boldsymbol{U}_{j,k}$ and $\mathrm{Var}\,(\boldsymbol{\xi}_{j,k}) = \boldsymbol{R}_{j,k}$. From Theorem 1, for any vector node $\{j, k\}$, $\boldsymbol{\xi}_{j,k}$ is uncorrelated with $\boldsymbol{U}_{j,k}$ and is also uncorrelated with $\{\boldsymbol{U}_{j',k'} : \{j', k'\} \prec j, k\} \cup \boldsymbol{Z}_{de\{j,k\}}$, whose elements can be written as $\boldsymbol{U}_{j,k}$ plus error terms and measurement errors that are independent of $\boldsymbol{\xi}_{j,k}$. Similarly, we have $\boldsymbol{U}_{j,k} = \boldsymbol{B}_{ch\{j,k,i\}}\boldsymbol{U}_{ch\{j,k,i\}} + \boldsymbol{\xi}_{ch\{j,k,i\}}$ and $\boldsymbol{\xi}_{ch\{j,k,i\}}$ is uncorrelated with $\boldsymbol{Z}_{de\{ch\{j,k,i\}\}}$. Then

$$\hat{\boldsymbol{U}}_{j,k|ch\{j,k,i\}} = \boldsymbol{B}_{ch\{j,k,i\}}\hat{\boldsymbol{U}}_{ch\{j,k,i\}|ch\{j,k,i\}},$$
$$\hat{\boldsymbol{V}}_{j,k|ch\{j,k,i\}} = \boldsymbol{B}_{ch\{j,k,i\}}\hat{\boldsymbol{V}}_{ch\{j,k,i\}|ch\{j,k,i\}}\boldsymbol{B}'_{ch\{j,k,i\}} + \boldsymbol{R}_{ch\{j,k,i\}},$$

because $\boldsymbol{\xi}_{ch\{j,k,i\}}$ is uncorrelated with $\boldsymbol{Z}_{de\{ch\{j,k,i\}\}}$.

Next we compute $\hat{\boldsymbol{U}}^*_{j,k|j,k}$ and $\hat{\boldsymbol{V}}^*_{j,k|j,k}$.

Proof of (20)−(21). From (9)−(12), $\boldsymbol{Z}_{de\{ch\{j,k,i\}\}} = \boldsymbol{X}_{de\{ch\{j,k,i\}\}}\boldsymbol{B} + \boldsymbol{M}_{de\{ch\{j,k,i\}\}}\boldsymbol{U}_{j,k} + \boldsymbol{e}_{de\{ch\{j,k,i\}\}}$ where $\boldsymbol{X}_{de\{ch\{j,k,i\}\}}$ depends on $\{\boldsymbol{X}_{j',k'} : \{j',k'\} \in T_{j,k}\}$, $\boldsymbol{M}_{de\{ch\{j,k,i\}\}}$ is a deterministic matrix depending on $\{\boldsymbol{A}_{j',k'} : \{j',k'\} \in T_{j,k}\}$, $\boldsymbol{e}_{de\{ch\{j,k,i\}\}}$ is a random vector depending on $\{\boldsymbol{W}_{j',k'} : \{j',k'\} \in T_{j,k}\}$ and $\{\boldsymbol{\epsilon}_{j',k'} : \{j',k'\} \in T_{j,k}\}$, and $T_{j,k} \equiv \{\{j',k'\} : \gamma_{j',k'} = 1, \{j',k'\} \prec \{j,k\}\}$.

We have $\boldsymbol{Z}^*_{de\{j,k\}} = \boldsymbol{X}^*_{de\{j,k\}}\boldsymbol{B} + \boldsymbol{M}^*_{de\{j,k\}}\boldsymbol{U}_{j,k} + \boldsymbol{e}^*_{de\{j,k\}}$, where $\boldsymbol{Z}^*_{de\{j,k\}} = (\boldsymbol{Z}'_{de\{ch\{j,k,1\}\}}, \ldots, \boldsymbol{Z}'_{de\{ch\{j,k,n_{j-1}\}\}})'$, $\boldsymbol{X}^*_{de\{j,k\}} = [\boldsymbol{X}'_{de\{ch\{j,k,1\}\}}, \ldots, \boldsymbol{X}'_{de\{ch\{j,k,n_{j-1}\}\}}]'$, $\boldsymbol{M}^*_{de\{j,k\}} = [\boldsymbol{M}'_{de\{ch\{j,k,1\}\}}, \ldots, \boldsymbol{M}'_{de\{ch\{j,k,n_{j-1}\}\}}]'$ and $\boldsymbol{e}^*_{de\{j,k\}} = (\boldsymbol{e}'_{de\{ch\{j,k,1\}\}}, \ldots, \boldsymbol{e}'_{de\{ch\{j,k,n_{j-1}\}\}})'$. Define $\boldsymbol{P}_{de\{ch\{j,k,i\}\}} \equiv \mathrm{Var}\,(\boldsymbol{e}_{de\{ch\{j,k,i\}\}})$ and $\boldsymbol{P}^*_{de\{j,k\}} \equiv \mathrm{Var}\,(\boldsymbol{e}^*_{de\{j,k\}})$. Then

$$\boldsymbol{P}^*_{de\{j,k\}} = \begin{pmatrix} \boldsymbol{P}_{de\{ch\{j,k,1\}\}} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \ldots & \boldsymbol{P}_{de\{ch\{j,k,n_{j-1}\}\}} \end{pmatrix}.$$

Suppose $\boldsymbol{V}_{j,k} = \boldsymbol{L}_{j,k}\boldsymbol{L}'_{j,k}$ where $\boldsymbol{L}_{j,k}$ is a matrix with full column rank. From Lemma 3 (iv), we have

$$\begin{aligned}
\hat{\boldsymbol{V}}^{*+}_{j,k|j,k} &= (\boldsymbol{L}'_{j,k})^+\Big[\boldsymbol{I} + \boldsymbol{L}'_{j,k}\boldsymbol{M}^{*'}_{de\{j,k\}}\boldsymbol{P}^{*-1}_{de\{j,k\}}\boldsymbol{M}^*_{de\{j,k\}}\boldsymbol{L}_{j,k}\Big]\boldsymbol{L}^+_{j,k} \\
&= (\boldsymbol{L}'_{j,k})^+\Big[\boldsymbol{I} + \boldsymbol{L}'_{j,k}(\sum_{i=1}^{n_{j-1}}\boldsymbol{M}^{*'}_{de\{ch\{j,k,i\}\}}\boldsymbol{P}^{*-1}_{de\{ch\{j,k,i\}\}}\boldsymbol{M}^*_{de\{ch\{j,k,i\}\}})\boldsymbol{L}_{j,k}\Big]\boldsymbol{L}^+_{j,k} \\
&= (\boldsymbol{L}'_{j,k})^+\boldsymbol{L}^+_{j,k} + \sum_{i=1}^{n_{j-1}}\Big[\hat{\boldsymbol{V}}^+_{j,k|ch\{j,k,i\}} - (\boldsymbol{L}'_{j,k})^+\boldsymbol{L}^+_{j,k}\Big] \\
&= \boldsymbol{V}^+_{j,k} + \sum_{i=1}^{n_{j-1}}\Big[\hat{\boldsymbol{V}}^+_{j,k|ch\{j,k,i\}} - \boldsymbol{V}^+_{j,k}\Big],
\end{aligned}$$

where the third equality holds because of Lemma 3 (vi). Hence

$$\hat{\boldsymbol{V}}^*_{j,k|j,k} = \Big\{\boldsymbol{V}^+_{j,k} + \Big[\sum_{i=1}^{n_{j-1}}\hat{\boldsymbol{V}}^+_{j,k|ch\{j,k,i\}} - \boldsymbol{V}^+_{j,k}\Big]\Big\}^+.$$

From Lemma 3 (vi),

$$\begin{aligned}
&\hat{\boldsymbol{V}}^{*+}_{j,k|j,k}\hat{\boldsymbol{U}}^{*}_{j,k|j,k}\\
&= (\boldsymbol{L}'_{j,k})^{+}\boldsymbol{L}'_{j,k}\boldsymbol{M}^{*'}_{de\{j,k\}}\boldsymbol{P}^{*-1}_{de\{j,k\}}(\boldsymbol{Z}^{*}_{de\{j,k\}} - \boldsymbol{X}^{*}_{de\{j,k\}}\boldsymbol{B})\\
&= (\boldsymbol{L}'_{j,k})^{+}\boldsymbol{L}'_{j,k}\Big[\sum_{i=1}^{n_{j-1}}\boldsymbol{M}^{*'}_{de\{ch\{j,k,i\}\}}\boldsymbol{P}^{*-1}_{de\{ch\{j,k,i\}\}}(\boldsymbol{Z}_{de\{ch\{j,k,i\}\}}-\boldsymbol{X}_{de\{ch\{j,k,i\}\}}\boldsymbol{B})\Big]\\
&= \sum_{i=1}^{n_{j-1}}\hat{\boldsymbol{V}}_{j,k|ch\{j,k,i\}}\hat{\boldsymbol{U}}_{j,k|ch\{j,k,i\}}.
\end{aligned}$$

Then from Lemma 3 (vi),

$$\hat{\boldsymbol{U}}^{*}_{j,k|j,k} = \hat{\boldsymbol{V}}^{*}_{j,k|j,k}\hat{\boldsymbol{V}}^{*+}_{j,k|j,k}\hat{\boldsymbol{U}}^{*}_{j,k|j,k} = \hat{\boldsymbol{V}}^{*}_{j,k|j,k}\left\{\sum_{i=1}^{n_{j-1}}\hat{\boldsymbol{V}}_{j,k|ch\{j,k,i\}}\hat{\boldsymbol{U}}_{j,k|ch\{j,k,i\}}\right\}.$$

The final step in each update is to compute $\hat{\boldsymbol{U}}_{j,k|j,k}$ and $\hat{\boldsymbol{V}}_{j,k|j,k}$.

Proof of (22)−(23). If $\gamma_{j,k} = 0$, then $\boldsymbol{U}_{j,k|j,k} = \boldsymbol{U}^{*}_{j,k|j,k}$ and $\hat{\boldsymbol{V}}_{j,k|j,k} = \hat{\boldsymbol{V}}^{*}_{j,k|j,k}$. If $\gamma_{j,k} = 1$, we define $\boldsymbol{X}_{de\{j,k\}} \equiv [\boldsymbol{X}'_{j,k}, \boldsymbol{X}^{*'}_{de\{j,k\}}]'$, $\boldsymbol{M}_{de\{j,k\}} \equiv [\boldsymbol{I}_{n_{j-1}m}, \boldsymbol{M}^{*'}_{de\{j,k\}}]'$, $\boldsymbol{e}_{de\{j,k\}} \equiv (\boldsymbol{e}'_{j,k}, \boldsymbol{e}^{*'}_{de\{j,k\}})'$, and $\boldsymbol{P}_{de\{j,k\}} = \mathrm{diag}(\boldsymbol{I}_{n_{j-1}} \otimes \boldsymbol{\Phi}_{j,k}, \boldsymbol{P}^{*}_{de\{j,k\}})$. Then $\boldsymbol{Z}_{de\{j,k\}} = (\boldsymbol{Z}'_{j,k}, \boldsymbol{Z}^{*'}_{de\{j,k\}})' = \boldsymbol{X}_{de\{j,k\}}\boldsymbol{B} + \boldsymbol{M}_{de\{j,k\}}\boldsymbol{U}_{j,k} + \boldsymbol{e}_{de\{j,k\}}$. Hence from Lemma 3 (vi),

$$\begin{aligned}
\hat{\boldsymbol{V}}^{+}_{j,k|j,k} &= (\boldsymbol{L}'_{j,k})^{+}(\boldsymbol{I} + \boldsymbol{L}'_{j,k}\boldsymbol{M}'_{j,k}\boldsymbol{P}^{-1}_{j,k}\boldsymbol{L}_{j,k})\boldsymbol{L}^{+}_{j,k}\\
&= (\boldsymbol{L}'_{j,k})^{+}\big[\boldsymbol{I} + \boldsymbol{L}'_{j,k}\boldsymbol{M}^{*'}_{j,k}\boldsymbol{P}^{*-1}_{j,k}\boldsymbol{L}_{j,k} + \boldsymbol{L}'_{j,k}(\boldsymbol{I}_{n_{j-1}} \otimes \boldsymbol{\Phi}^{-1}_{j,k})\boldsymbol{L}_{j,k}\big]\boldsymbol{L}^{+}_{j,k}\\
&= \hat{\boldsymbol{V}}^{*}_{j,k|j,k} + (\boldsymbol{L}'_{j,k})^{+}\boldsymbol{L}'_{j,k}(\boldsymbol{I}_{n_{j-1}} \otimes \boldsymbol{\Phi}^{-1}_{j,k})\boldsymbol{L}_{j,k}\boldsymbol{L}^{+}_{j,k}.
\end{aligned}$$

Suppose $\hat{\boldsymbol{V}}^{*}_{j,k|j,k} = \boldsymbol{Q}_{j,k}\boldsymbol{Q}'_{j,k}$ where $\boldsymbol{Q}_{j,k}$ has full column rank. Then

$$\begin{aligned}
&\Big\{\hat{\boldsymbol{V}}^{*}_{j,k|j,k} - \hat{\boldsymbol{V}}^{*}_{j,k|j,k}(\hat{\boldsymbol{V}}^{*}_{j,k|j,k} + \boldsymbol{I}_{n_{j-1}} \otimes \boldsymbol{\Phi}_{j})^{-1}\hat{\boldsymbol{V}}^{*}_{j,k|j,k}\Big\}^{+}\\
&= \{\boldsymbol{Q}_{j,k}[\boldsymbol{I}_{n_{j-1}m} - \boldsymbol{Q}'_{j,k}(\boldsymbol{Q}_{j,k}\boldsymbol{Q}'_{j,k} + \boldsymbol{I}_{n_{j-1}} \otimes \boldsymbol{\Phi}_{j})^{-1}\boldsymbol{Q}_{j,k}]\boldsymbol{Q}'_{j,k}\}^{+}\\
&= \Big\{\boldsymbol{Q}_{j,k}[\boldsymbol{I}_{n_{j-1}m} + \boldsymbol{Q}'_{j,k}(\boldsymbol{I}_{n_{j-1}} \otimes \boldsymbol{\Phi}^{-1}_{j})\boldsymbol{Q}_{j,k}]^{-1}\boldsymbol{Q}'_{j,k}\Big\}^{+}\\
&= (\boldsymbol{Q}'_{j,k})^{+}[\boldsymbol{I}_{n_{j-1}m} + \boldsymbol{Q}'_{j,k}(\boldsymbol{I}_{n_{j-1}} \otimes \boldsymbol{\Phi}^{-1}_{j})\boldsymbol{Q}_{j,k}]\boldsymbol{Q}^{+}_{j,k}\\
&= (\boldsymbol{Q}'_{j,k})^{+}\boldsymbol{Q}^{+}_{j,k} + (\boldsymbol{Q}'_{j,k})^{+}\boldsymbol{Q}'_{j,k}(\boldsymbol{I}_{n_{j-1}} \otimes \boldsymbol{\Phi}^{-1}_{j})\boldsymbol{Q}_{j,k}\boldsymbol{Q}^{+}_{j,k}
\end{aligned}$$

$$
\begin{aligned}
&= \hat{\boldsymbol{V}}^{*+}_{j,k|j,k} + \boldsymbol{P}_{Q_{j,k}}(\boldsymbol{I}_{n_{j-1}} \otimes \boldsymbol{\Phi}_j^{-1})\boldsymbol{P}_{Q_{j,k}} \\
&= \hat{\boldsymbol{V}}^{*+}_{j,k|j,k} + \boldsymbol{P}_{L_{j,k}}(\boldsymbol{I}_{n_{j-1}} \otimes \boldsymbol{\Phi}_j^{-1})\boldsymbol{P}_{L_{j,k}} \\
&= \hat{\boldsymbol{V}}^{*+}_{j,k|j,k} + (\boldsymbol{L}'_{j,k})^{+}\boldsymbol{L}'_{j,k}(\boldsymbol{I}_{n_{j-1}} \otimes \boldsymbol{\Phi}_{j,k}^{-1})\boldsymbol{L}_{j,k}\boldsymbol{L}^{+}_{j,k} \\
&= \hat{\boldsymbol{V}}^{+}_{j,k|j,k},
\end{aligned}
$$

where the second, third and fifth equality hold because of Lemma 3 (iii), (iv) and (v) respectively. Since $\hat{\boldsymbol{V}}^{*}_{j,k|j,k} = \boldsymbol{L}_{j,k}(\boldsymbol{I} + \boldsymbol{L}'_{j,k}\boldsymbol{M}^{*\prime}_{j,k}\boldsymbol{P}^{*-1}_{j,k}\boldsymbol{L}_{j,k})^{-1}\boldsymbol{L}'_{j,k}$, we have $\mathrm{Col}(\hat{\boldsymbol{V}}^{*}_{j,k|j,k}) \subseteq \mathrm{Col}(\boldsymbol{L}_{j,k})$. Since $\mathrm{Rank}(\hat{\boldsymbol{V}}^{*}_{j,k|j,k}) = \mathrm{Rank}(\boldsymbol{L}_{j,k})$, we have $\mathrm{Col}(\hat{\boldsymbol{V}}^{*}_{j,k|j,k}) = \mathrm{Col}(\boldsymbol{L}_{j,k})$. Similarly, we have $\mathrm{Col}(\hat{\boldsymbol{V}}^{*}_{j,k|j,k}) = \mathrm{Col}(\boldsymbol{Q}_{j,k})$. From Lemma 3 (v), we have $\boldsymbol{P}_{Q_{j,k}} = \boldsymbol{P}_{\hat{V}^{*}_{j,k|j,k}} = \boldsymbol{P}_{L_{j,k}}$. Hence

$$
\hat{\boldsymbol{V}}_{j,k|j,k} = \hat{\boldsymbol{V}}^{*}_{j,k|j,k} - \hat{\boldsymbol{V}}^{*}_{j,k|j,k}(\hat{\boldsymbol{V}}^{*}_{j,k|j,k} + \boldsymbol{I}_{n_{j-1}} \otimes \boldsymbol{\Phi}_j)^{-1}\hat{\boldsymbol{V}}^{*}_{j,k|j,k}.
$$

From Lemma 3 (vi),

$$
\begin{aligned}
\hat{\boldsymbol{V}}^{+}_{j,k|j,k}\hat{\boldsymbol{U}}_{j,k|j,k} &= (\boldsymbol{L}'_{j,k})^{+}\boldsymbol{L}'_{j,k}\boldsymbol{M}'_{de\{j,k\}}\boldsymbol{P}^{-1}_{de\{j,k\}}(\boldsymbol{Z}_{de\{j,k\}} - \boldsymbol{X}_{de\{j,k\}}\boldsymbol{B}) \\
&= (\boldsymbol{L}'_{j,k})^{+}\boldsymbol{L}'_{j,k}\Big[\boldsymbol{M}^{*\prime}_{de\{j,k,\}}\boldsymbol{P}^{*-1}_{de\{j,k\}}(\boldsymbol{Z}^{*}_{de\{j,k\}} - \boldsymbol{X}^{*}_{de\{j,k,\}}\boldsymbol{B}) \\
&\quad + (\boldsymbol{I}_{n_{j-1}} \otimes \boldsymbol{\Phi}^{-1}_{j,k})(\boldsymbol{Z}_{j,k} - \boldsymbol{X}_{j,k}\boldsymbol{B})\Big] \\
&= \hat{\boldsymbol{V}}^{*+}_{j,k|j,k}\hat{\boldsymbol{U}}^{*}_{j,k|j,k} + (\boldsymbol{L}'_{j,k})^{+}\boldsymbol{L}'_{j,k}(\boldsymbol{I}_{n_{j-1}} \otimes \boldsymbol{\Phi}^{-1}_{j,k})(\boldsymbol{Z}_{j,k} - \boldsymbol{X}_{j,k}\boldsymbol{B}),
\end{aligned}
$$

where the last equality holds because of Lemma 3 (vi). Then from Lemma 3 (vi),

$$
\begin{aligned}
\hat{\boldsymbol{U}}_{j,k|j,k} &= \hat{\boldsymbol{V}}_{j,k|j,k}\hat{\boldsymbol{V}}^{+}_{j,k|j,k}\hat{\boldsymbol{U}}_{j,k|j,k} \\
&= \hat{\boldsymbol{V}}_{j,k|j,k}\hat{\boldsymbol{V}}^{*+}_{j,k|j,k}\hat{\boldsymbol{U}}^{*}_{j,k|j,k} + \hat{\boldsymbol{V}}_{j,k|j,k}(\boldsymbol{L}'_{j,k})^{+}\boldsymbol{L}'_{j,k}(\boldsymbol{I}_{n_{j-1}} \otimes \boldsymbol{\Phi}^{-1}_{j,k})(\boldsymbol{Z}_{j,k} - \boldsymbol{X}_{j,k}\boldsymbol{B}) \\
&= \hat{\boldsymbol{V}}_{j,k|j,k}\left\{(\boldsymbol{I}_{n_{j-1}} \otimes \boldsymbol{\Phi}^{-1}_{j})(\boldsymbol{Z}_{j,k} - \boldsymbol{X}_{j,k}\boldsymbol{B}) + (\hat{\boldsymbol{V}}^{*}_{j,k|j,k})^{+}\hat{\boldsymbol{U}}^{*}_{j,k|j,k}\right\},
\end{aligned}
$$

where the last equality holds because from Lemma 3 (vi),

$$
\begin{aligned}
&\hat{\boldsymbol{V}}_{j,k|j,k}(\boldsymbol{L}'_{j,k})^{+}\boldsymbol{L}'_{j,k} \\
&= (\boldsymbol{L}'_{j,k})^{+}\Big[\boldsymbol{I}_{n_{j-1}m} + \boldsymbol{L}'_{j,k}\boldsymbol{M}'_{de\{j,k\}}\boldsymbol{P}^{-1}_{de\{j,k\}}\boldsymbol{M}_{de\{j,k\}}\boldsymbol{L}_{j,k}\Big]^{-1}\boldsymbol{L}'_{j,k}(\boldsymbol{L}'_{j,k})^{+}\boldsymbol{L}'_{j,k} \\
&= (\boldsymbol{L}'_{j,k})^{+}\Big[\boldsymbol{I}_{n_{j-1}m} + \boldsymbol{L}'_{j,k}\boldsymbol{M}'_{de\{j,k\}}\boldsymbol{P}^{-1}_{de\{j,k\}}\boldsymbol{M}_{de\{j,k\}}\boldsymbol{L}_{j,k}\Big]^{-1}\boldsymbol{L}'_{j,k} \\
&= \hat{\boldsymbol{V}}_{j,k|j,k}.
\end{aligned}
$$

At the end of the filtering step, the root nodes are reached and hence the BLUP's for $\{1,1\}$ are

$$
\hat{\boldsymbol{U}}_{1,1} \equiv \boldsymbol{m}_{U_{1,1}|Z} = \hat{\boldsymbol{U}}_{1,1|1,1}, \qquad \hat{\boldsymbol{V}}_{1,1} \equiv \boldsymbol{C}_{U_{1,1}|Z} = \hat{\boldsymbol{V}}_{1,1|1,1},
$$

where $\boldsymbol{Z} \equiv \{\boldsymbol{Z}_{j,k} : \gamma_{j,k} = 1, k = 1, \ldots, N_{j-1}, j = 1, \ldots, J\}$ consists of all the observations.

In the low-to-high-resolution smoothing step, we move from the coarsest resolution $j = 2$ to the finest resolution $j = J$ and compute for a given node $\{j,k\}$, where $k = 1, \ldots, N_{j-1}$. We start with the following Lemma.

Lemma 4. $\boldsymbol{m}(\tilde{\boldsymbol{U}}_{j,k|j,k}|\boldsymbol{\nu}_{j,k|j,k}) = \boldsymbol{J}_{j,k}\boldsymbol{m}(\tilde{\boldsymbol{U}}_{pa\{j,k\}|j,k}|\boldsymbol{\nu}_{j,k|j,k})$, *where* $\boldsymbol{\nu}_{j,k|j,k} \equiv \boldsymbol{Z}^c_{de\{j,k\}} - \boldsymbol{m}(\boldsymbol{Z}^c_{de\{j,k\}}|\boldsymbol{Z}_{de\{j,k\}})$ *and* $\boldsymbol{J}_{j,k} \equiv \hat{\boldsymbol{V}}_{j,k|j,k}\boldsymbol{B}'_{j,k}\hat{\boldsymbol{V}}^+_{pa\{j,k\}|j,k}$.

Proof. See details in Yue and Zhu (2005).

Proof of (25)−(26). $\boldsymbol{\nu}_{j,k|j,k} \equiv \boldsymbol{Z}^c_{de\{j,k\}} - \boldsymbol{m}(\boldsymbol{Z}^c_{de\{j,k\}}|\boldsymbol{Z}_{de\{j,k\}})$ is the information provided by the observations $\boldsymbol{Z}^c_{de\{j,k\}}$ given $\boldsymbol{Z}_{de\{j,k\}}$. From Theorem 1, $\boldsymbol{\nu}_{j,k|j,k}$ and $\boldsymbol{Z}_{de\{j,k\}}$ are uncorrelated. Since $\overline{sp}\{\boldsymbol{Z}\}^{mn_{j-1}} = \overline{sp}\{\boldsymbol{Z}_{de\{j,k\}} \cup \boldsymbol{Z}^c_{de\{j,k\}}\}^{mn_{j-1}} = \overline{sp}\{\boldsymbol{Z}_{de\{j,k\}} \cup \boldsymbol{\nu}_{de\{j,k\}}\}^{mn_{j-1}}$ and $\boldsymbol{\nu}_{j,k|j,k}$ and $\boldsymbol{Z}_{de\{j,k\}}$ are uncorrelated, we have $\hat{\boldsymbol{U}}_{j,k} = \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}) = \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}_{de\{j,k\}}) + \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{\nu}_{j,k|j,k}) = \hat{\boldsymbol{U}}_{j,k|j,k} + \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{\nu}_{j,k|j,k})$. Define $\tilde{\boldsymbol{U}}_{j,k|j,k} \equiv \boldsymbol{U}_{j,k} - \hat{\boldsymbol{U}}_{j,k|j,k}$. Then $\tilde{\boldsymbol{U}}_{j,k|j,k}$ is uncorrelated with $\hat{\boldsymbol{U}}_{j,k|j,k}$, and $\boldsymbol{U}_{j,k} = \hat{\boldsymbol{U}}_{j,k|j,k} + \tilde{\boldsymbol{U}}_{j,k|j,k}$. Hence

$$
\begin{aligned}
\hat{\boldsymbol{U}}_{j,k} &= \hat{\boldsymbol{U}}_{j,k|j,k} + \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{\nu}_{j,k|j,k}) \\
&= \hat{\boldsymbol{U}}_{j,k|j,k} + \boldsymbol{m}(\hat{\boldsymbol{U}}_{j,k} + \tilde{\boldsymbol{U}}_{j,k}|\boldsymbol{\nu}_{j,k|j,k}) \\
&= \hat{\boldsymbol{U}}_{j,k|j,k} + \boldsymbol{m}(\hat{\boldsymbol{U}}_{j,k|j,k}|\boldsymbol{\nu}_{j,k|j,k}) + \boldsymbol{m}(\tilde{\boldsymbol{U}}_{j,k|j,k}|\boldsymbol{\nu}_{j,k|j,k}) \\
&= \hat{\boldsymbol{U}}_{j,k|j,k} + \boldsymbol{m}(\tilde{\boldsymbol{U}}_{j,k|j,k}|\boldsymbol{\nu}_{j,k|j,k}),
\end{aligned} \tag{63}
$$

where in the third equality, $\boldsymbol{m}(\hat{\boldsymbol{U}}_{j,k|j,k}|\boldsymbol{\nu}_{j,k|j,k}) = \boldsymbol{0}_{mn_{j-1}}$ because $\hat{\boldsymbol{U}}_{j,k|j,k}$ and $\boldsymbol{\nu}_{j,k|j,k}$ are uncorrelated. Similarly,

$$
\hat{\boldsymbol{U}}_{pa\{j,k\}} = \hat{\boldsymbol{U}}_{pa\{j,k\}|j,k} + \boldsymbol{m}(\tilde{\boldsymbol{U}}_{pa\{j,k\}|j,k}|\boldsymbol{\nu}_{j,k|j,k}) \tag{64}
$$

where $\tilde{\boldsymbol{U}}_{pa\{j,k\}|j,k} \equiv \boldsymbol{U}_{pa\{j,k\}} - \hat{\boldsymbol{U}}_{pa\{j,k\}|j,k}$.

From Lemma 4, we have $\boldsymbol{m}(\tilde{\boldsymbol{U}}_{j,k|j,k}|\boldsymbol{\nu}_{j,k|j,k}) = \boldsymbol{J}_{j,k}\boldsymbol{m}(\tilde{\boldsymbol{U}}_{pa\{j,k\}|j,k}|\boldsymbol{\nu}_{j,k|j,k})$. Combining (63) and (64), we have $\hat{\boldsymbol{U}}_{j,k} = \hat{\boldsymbol{U}}_{j,k|j,k} + \boldsymbol{J}_{j,k}[\hat{\boldsymbol{U}}_{pa\{j,k\}} - \hat{\boldsymbol{U}}_{pa\{j,k\}|j,k}]$. Then $\tilde{\boldsymbol{U}}_{j,k} \equiv \boldsymbol{U}_{j,k} - \hat{\boldsymbol{U}}_{j,k} = \boldsymbol{U}_{j,k} - \hat{\boldsymbol{U}}_{j,k|j,k} - \boldsymbol{J}_{j,k}[\hat{\boldsymbol{U}}_{pa\{j,k\}} - \hat{\boldsymbol{U}}_{pa\{j,k\}|j,k}]$. Hence

$$
\tilde{\boldsymbol{U}}_{j,k} + \boldsymbol{J}_{j,k}\hat{\boldsymbol{U}}_{pa\{j,k\}} = \tilde{\boldsymbol{U}}_{j,k|j,k} + \boldsymbol{J}_{j,k}\hat{\boldsymbol{U}}_{pa\{j,k\}|j,k}. \tag{65}
$$

Now we compute the variances of both sides of (65):

$$
\begin{aligned}
\mathrm{Var}\,(\tilde{\boldsymbol{U}}_{j,k} + \boldsymbol{J}_{j,k}\hat{\boldsymbol{U}}_{pa\{j,k\}}) &= \mathrm{Var}\,(\tilde{\boldsymbol{U}}_{j,k}) + \mathrm{Var}\,(\boldsymbol{J}_{j,k}\hat{\boldsymbol{U}}_{pa\{j,k\}}) \\
&= \mathrm{Var}\,(\boldsymbol{U}_{j,k} - \hat{\boldsymbol{U}}_{j,k}) + \boldsymbol{J}_{j,k}\mathrm{Var}\,(\hat{\boldsymbol{U}}_{pa\{j,k\}})\boldsymbol{J}'_{j,k} \\
&= \hat{\boldsymbol{V}}_{j,k} + \boldsymbol{J}_{j,k}(\boldsymbol{V}_{pa\{j,k\}} - \hat{\boldsymbol{V}}_{pa\{j,k\}})\boldsymbol{J}'_{j,k},
\end{aligned} \tag{66}
$$

where the first equality holds because $\tilde{\boldsymbol{U}}_{j,k}$ is uncorrelated with $\boldsymbol{Z}$ and $\hat{\boldsymbol{U}}_{pa\{j,k\}} \in \overline{sp}\{\boldsymbol{Z}\}^{mn_{j-2}}$, and the last equality holds because $\mathrm{Var}\,(\boldsymbol{U}_{j,k} - \hat{\boldsymbol{U}}_{j,k}) = \boldsymbol{C}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}) = \hat{\boldsymbol{V}}_{j,k}$ and $\boldsymbol{V}_{pa\{j,k\}} = \mathrm{Var}\,(\boldsymbol{U}_{pa\{j,k\}}) = \mathrm{Var}\,(\hat{\boldsymbol{U}}_{pa\{j,k\}} + \tilde{\boldsymbol{U}}_{pa\{j,k\}}) = \mathrm{Var}\,(\hat{\boldsymbol{U}}_{pa\{j,k\}}) + \mathrm{Var}\,(\tilde{\boldsymbol{U}}_{pa\{j,k\}}) = \mathrm{Var}\,(\hat{\boldsymbol{U}}_{pa\{j,k\}}) + \mathrm{Var}\,(\boldsymbol{U}_{pa\{j,k\}} - \hat{\boldsymbol{U}}_{pa\{j,k\}}) = \mathrm{Var}\,(\hat{\boldsymbol{U}}_{pa\{j,k\}}) + \hat{\boldsymbol{V}}_{pa\{j,k\}}$. Similarly,

$$\begin{aligned} \mathrm{Var}\,(\tilde{\boldsymbol{U}}_{j,k|j,k} + \boldsymbol{J}_{j,k}\hat{\boldsymbol{U}}_{pa\{j,k\}|j,k}) &= \mathrm{Var}\,(\tilde{\boldsymbol{U}}_{j,k|j,k}) + \boldsymbol{J}_{j,k}\mathrm{Var}\,(\hat{\boldsymbol{U}}_{pa\{j,k\}|j,k})\boldsymbol{J}'_{j,k} \\ &= \hat{\boldsymbol{V}}_{j,k|j,k} + \boldsymbol{J}_{j,k}(\boldsymbol{V}_{pa\{j,k\}} - \hat{\boldsymbol{V}}_{pa\{j,k\}|j,k})\boldsymbol{J}'_{j,k}, \quad (67) \end{aligned}$$

where the first equality holds because $\tilde{\boldsymbol{U}}_{j,k|j,k}$ is uncorrelated with $\boldsymbol{Z}_{de\{j,k\}}$ and $\hat{\boldsymbol{U}}_{pa\{j,k\}|j,k} \in \overline{sp}\{\boldsymbol{Z}_{de\{j,k\}}\}^{mn_{j-2}}$. From formula (66) and (67), we obtain $\hat{\boldsymbol{V}}_{j,k} = \hat{\boldsymbol{V}}_{j,k|j,k} + \boldsymbol{J}_{j,k}(\hat{\boldsymbol{V}}_{pa\{j,k\}} - \hat{\boldsymbol{V}}_{pa\{j,k\}|j,k})\boldsymbol{J}'_{j,k}$.

For single-source data, the change-of-resolution Kalman-filter algorithm remains the same, except that $\gamma_{j,k} = 0$ whenever $j < J$.

Appendix III. ML and REML Estimators

Lemma 5 gives some useful matrix results. Lemmas 6 and 8 give the inverse and the determinant of the matrix $\boldsymbol{\Omega}$, whereas Lemma 7 gives auxiliary results about $\boldsymbol{A}_j$ and $\boldsymbol{\Omega}$ and Lemma 9 provides a useful decomposition of $(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{B})'\boldsymbol{\Omega}^{-1}(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{B})$. To establish Theorem 3, we use Lemmas 10, 11 and 12, which give the differentiation of $(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{B})'\boldsymbol{\Omega}^{-1}(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{B})$ and $\log|\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}|$ with respect to $\boldsymbol{D}_j^{-1}$. Lemma 13 gives the expectation of the sums of squares $SS_j(\cdot)$. Finally, we consider explicit formulas for the MLEs and REMLEs under the assumption of a constant regression mean. Lemma 14 gives an intermediate step and Lemma 15 establishes the distributional properties of the sums of squares $SS_j(\cdot)$. In deriving the results assuming a compound symmetry structure for the covariance matrix, we use the auxiliary Lemma 16. To save space, we present only the proof of Lemma 15, deferring all the other proofs to Yue and Zhu (2005).

Lemma 5. *For matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$,*

(i) $E(\boldsymbol{Z}'\boldsymbol{A}\boldsymbol{Z}) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Omega}) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}$, *where $\boldsymbol{A}$ is an deterministic square matrix and* $\boldsymbol{Z} \sim N(\boldsymbol{\mu}, \boldsymbol{\Omega})$.

(ii) $\mathrm{tr}(\boldsymbol{A}\boldsymbol{Q}_{hi}) = a_{ih}$, *where $\boldsymbol{A}$ is an $m \times n$ matrix, $\boldsymbol{Q}_{hi}$ is an $n \times m$ matrix whose (h,i)th element is one and zero otherwise, and a_{ih} is the (i,h)th element of matrix $\boldsymbol{A}$.*

(iii) $\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{B}\boldsymbol{A})$, *where $\boldsymbol{A}$ is an $m \times n$ matrix and $\boldsymbol{B}$ is an $n \times m$ matrix.*

(iv) $\mathrm{tr}(\boldsymbol{A} \otimes \boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A})\,\mathrm{tr}(\boldsymbol{B})$, $(\boldsymbol{A} \otimes \boldsymbol{B})' = \boldsymbol{A}' \otimes \boldsymbol{B}'$, *and* $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{A}\boldsymbol{C}) \otimes (\boldsymbol{B}\boldsymbol{D})$.

(v) $\vec{z}'(\boldsymbol{A} \otimes \boldsymbol{B})\vec{z} = \langle \vec{z}, (\boldsymbol{A} \otimes \boldsymbol{B})\vec{z} \rangle = \mathrm{tr}[\boldsymbol{z}'\boldsymbol{A}\boldsymbol{z}\boldsymbol{B}]$, *where* $\boldsymbol{z}$ *is an* $m \times n$ *matrix,* $\boldsymbol{A}$ *is an* $m \times m$ *matrix and* $\boldsymbol{B}$ *is an* $n \times n$ *matrix.*

(vi) *For* $m \times m$ *matrices* $\boldsymbol{A}$ *and* $\boldsymbol{B}$, $n \times n$ *matrices* $\boldsymbol{C}$ *and* $\boldsymbol{D}$, $|\boldsymbol{A} \otimes \boldsymbol{C} + \boldsymbol{B} \otimes \boldsymbol{D}| = |\boldsymbol{C} \otimes \boldsymbol{A} + \boldsymbol{D} \otimes \boldsymbol{B}|$, *where* $|\boldsymbol{A}|$ *denotes the determinant of matrix* $\boldsymbol{A}$. *Moreover* $|\boldsymbol{A} \otimes \boldsymbol{I}_n + \boldsymbol{B} \otimes (\mathbf{1}_n \mathbf{1}_n')| = |\boldsymbol{A}|^{n-1}|\boldsymbol{A} + n\boldsymbol{B}|$ *and* $|\boldsymbol{A} \otimes \boldsymbol{C}| = |\boldsymbol{A}|^n |\boldsymbol{C}|^m$.

(vii) *For an* $n \times n$ *symmetric matrix* $\boldsymbol{A}$, *if* $\boldsymbol{z} \sim W_m(n, \boldsymbol{\Sigma})$, *then* $\mathrm{Var}\,(\mathrm{tr}[\boldsymbol{A}\boldsymbol{z}]) = 2n\,\mathrm{tr}[\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma}]$, *where* $W_m(n, \boldsymbol{\Sigma})$ *denotes a Wishart distribution with* n *degrees of freedom and parameter* $\boldsymbol{\Sigma}$ (*Definition C.9 of* Lauritzen (1996)).

Lemma 6. $\boldsymbol{\Omega}^{-1} = \sum_{j=1}^{J} -\boldsymbol{A}_j \otimes \boldsymbol{C}_j = \sum_{j=1}^{J} (\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{D}_j^{-1} + \boldsymbol{A}_0 \otimes \boldsymbol{D}_1^{-1}$, *where* $\boldsymbol{C}_J \equiv -\boldsymbol{D}_J^{-1}$ *and* $\boldsymbol{C}_j \equiv \boldsymbol{D}_{j+1}^{-1} - \boldsymbol{D}_j^{-1}, j = 1, \ldots, J-1$. *Moreover,* $\sum_{k=j}^{J} \boldsymbol{C}_k = -\boldsymbol{D}_j^{-1}, j = 1, \ldots, J$, *and* $\boldsymbol{\Psi}_J = \boldsymbol{D}_J$ *and* $\boldsymbol{\Psi}_j = (\boldsymbol{D}_j - \boldsymbol{D}_{j+1})/a_j$, $j = 1, \ldots, J-1$.

Lemma 7. *Properties of* $\{\boldsymbol{A}_j : j = 0, 1, \ldots, J\}$ *and* $\boldsymbol{\Omega}$ *are the following.*

(i) $\boldsymbol{A}_j \boldsymbol{A}_k = \boldsymbol{A}_{\min(j,k)}$ *and* $\boldsymbol{A}_j^2 = \boldsymbol{A}_j = \boldsymbol{A}_j'$, *where* $j, k = 0, 1, \ldots, J$.

(ii) $(\boldsymbol{A}_j - \boldsymbol{A}_k)^2 = \boldsymbol{A}_j - \boldsymbol{A}_k, 0 \le k \le j \le J$.

(iii) $\boldsymbol{A}_j \mathbf{1}_{N_J} = \mathbf{1}_{N_J}$, $j = 0, \ldots, J$.

(iv) $\mathrm{Rank}(\boldsymbol{A}_j) = \mathrm{tr}(\boldsymbol{A}_j) = N_j, j = 0, 1, \ldots, J$.

(v) $\mathrm{Rank}(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) = \mathrm{tr}(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) = N_j - N_{j-1}$, $j = 1, \ldots, J$.

(vi) $\boldsymbol{\Omega}[(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{C}] = (\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes (\boldsymbol{D}_j \boldsymbol{C})$, *where* $\boldsymbol{C}$ *is any* $m \times m$ *matrix.*

(vii) $\boldsymbol{\Omega}(\mathbf{1}_{N_J} \otimes \boldsymbol{I}_m) = \mathbf{1}_{N_J} \otimes \boldsymbol{D}_1$ *and* $(\mathbf{1}_{N_J} \otimes \boldsymbol{I}_m)'\boldsymbol{\Omega}(\mathbf{1}_{N_J} \otimes \boldsymbol{I}_m) = N_J \boldsymbol{D}_1$.

(viii) $\boldsymbol{\Omega}^{-1}(\mathbf{1}_{N_J} \otimes \boldsymbol{I}_m) = \mathbf{1}_{N_J} \otimes \boldsymbol{D}_1^{-1}$ *and* $(\mathbf{1}_{N_J} \otimes \boldsymbol{I}_m)'\boldsymbol{\Omega}^{-1}(\mathbf{1}_{N_J} \otimes \boldsymbol{I}_m) = N_J \boldsymbol{D}_1^{-1}$.

Lemma 8. $|\boldsymbol{\Omega}| = |\boldsymbol{D}_J|^{N_J - N_{J-1}} |\boldsymbol{D}_{J-1}|^{N_{J-1} - N_{J-2}} \cdots |\boldsymbol{D}_2|^{N_2 - N_1} |\boldsymbol{D}_1|^{N_1}$.

Lemma 9. $(\boldsymbol{Z} - \boldsymbol{X}\boldsymbol{B})'\boldsymbol{\Omega}^{-1}(\boldsymbol{Z} - \boldsymbol{X}\boldsymbol{B}) = \sum_{j=1}^{J} \mathrm{tr}[SS_j(\boldsymbol{\beta})\boldsymbol{D}_j^{-1}] + \mathrm{tr}[SS_0(\boldsymbol{\beta})\boldsymbol{D}_1^{-1}]$.

Lemma 10.

(i) $\partial\,\mathrm{tr}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B})/\partial \boldsymbol{X} = \boldsymbol{A}'\boldsymbol{B}$, *where* $\boldsymbol{A}$, $\boldsymbol{B}$, *and* $\boldsymbol{X}$ *are conformable matrices.*

(ii) $\partial \log|\boldsymbol{X}|/\partial \boldsymbol{X} = (\boldsymbol{X}^{-1})'$, *where* $\boldsymbol{X}$ *is an* $m \times m$ *invertible matrix.*

(iii) $\partial \boldsymbol{F}\boldsymbol{G}/\partial x_i = \boldsymbol{F}(\partial \boldsymbol{G}/\partial x_i) + (\partial \boldsymbol{F}/\partial x_i)\boldsymbol{G}$ *and* $\partial\,\mathrm{tr}(\boldsymbol{F}\boldsymbol{G})/\partial x_i = \mathrm{tr}[\boldsymbol{F}(\partial \boldsymbol{G}/\partial x_i)] + \mathrm{tr}[(\partial \boldsymbol{F}/\partial x_i)\boldsymbol{G}]$, *where* $\boldsymbol{F}$ *and* $\boldsymbol{G}$ *are conformable matrices that depend on* $\boldsymbol{x} = (x_1, \ldots, x_n)'$.

(iv) $\partial \boldsymbol{F}^{-1}/\partial x_i = -\boldsymbol{F}^{-1}(\partial \boldsymbol{F}/\partial x_i)\boldsymbol{F}^{-1}$, *where* $\boldsymbol{F}$ *is an invertible matrix that depends on* $\boldsymbol{x} = (x_1, \ldots, x_n)'$.

(v) $\partial \log|\boldsymbol{A}\boldsymbol{F}^{-1}\boldsymbol{B}|/\partial x_i = -\,\mathrm{tr}[\boldsymbol{F}^{-1}\boldsymbol{B}(\boldsymbol{A}\boldsymbol{F}^{-1}\boldsymbol{B})^{-1}\boldsymbol{A}\boldsymbol{F}^{-1}(\partial \boldsymbol{F}/\partial x_i)]$, *where* $\boldsymbol{A}$, $\boldsymbol{B}$ *and* $\boldsymbol{F}$ *are conformable matrices and only* $\boldsymbol{F}$ *depends on* $\boldsymbol{x} = (x_1, \ldots, x_n)'$.

(vi) $\partial f/\partial x_i = \mathrm{tr}[(\partial g/\partial \boldsymbol{H})'(\partial \boldsymbol{H}/\partial x_i)]$, *where* $f(\boldsymbol{x}) = g(\boldsymbol{H}(\boldsymbol{x}))$, g *is a function that depends on the elements of* $\boldsymbol{H}$ *and* $\boldsymbol{H}$ *is a matrix that depends on* $\boldsymbol{x} = (x_1, \ldots, x_n)'$.

Lemma 11. *For* $SS(\boldsymbol{B}) \equiv (\boldsymbol{Z} - \boldsymbol{X}\boldsymbol{B})'\boldsymbol{\Omega}^{-1}(\boldsymbol{Z} - \boldsymbol{X}\boldsymbol{B})$,

$$\frac{\partial SS(\hat{\boldsymbol{B}})}{\partial \boldsymbol{D}_1^{-1}} = [SS_1(\hat{\boldsymbol{\beta}}) + SS_0(\hat{\boldsymbol{\beta}})]' \text{ and } \frac{\partial SS(\hat{\boldsymbol{B}})}{\partial \boldsymbol{D}_j^{-1}} = [SS_j(\hat{\boldsymbol{\beta}})]', j = 2, \ldots, J,$$

where $\hat{\boldsymbol{B}} = [\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1}[\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{Z}]$.

Lemma 12. *Let* $\boldsymbol{D}_{jhi}$ *denote the* (h,i)*th element of* $\boldsymbol{D}_j^{-1}$,

$$\frac{\partial \log |\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}|}{\partial \boldsymbol{D}_{jhi}^{-1}} = \begin{cases} \operatorname{tr}[\boldsymbol{P}(\boldsymbol{A}_j \otimes \boldsymbol{Q}_{hi})] & \text{if } j = 1, \\ \operatorname{tr}\Big[\boldsymbol{P}[(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{Q}_{hi}]\Big] & \text{if } j = 2, \ldots, J, \end{cases}$$

where $\boldsymbol{P} \equiv \boldsymbol{X}[\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}'$ *and* $\boldsymbol{Q}_{hi}$ *is an* $m \times m$ *matrix with* 1 *for the* (h,i)*th element and* 0 *otherwise.*

Lemma 13.

$$E(SS_j(\boldsymbol{\beta})) = \begin{cases} \boldsymbol{D}_1 & \text{if } j = 0, \\ (N_j - N_{j-1})\boldsymbol{D}_j & \text{if } j = 1, \ldots, J, \end{cases}$$

$$E(SS_j(\hat{\boldsymbol{\beta}})) = \begin{cases} N_0\boldsymbol{D}_1 - \Big[\operatorname{tr}[\boldsymbol{P}(\boldsymbol{A}_0 \otimes \boldsymbol{Q}_{hi})]\Big]_{ih} & \text{if } j=0, \\ (N_j - N_{j-1})\boldsymbol{D}_j - \Big[\operatorname{tr}[\boldsymbol{P}((\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{Q}_{hi})]\Big]_{ih} & \text{if } j=1, \ldots, J. \end{cases}$$

Lemma 14. $\operatorname{tr}\Big[\boldsymbol{P}[(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{Q}_{hi}]\Big] = 0, j = 1, \ldots, J,$ *and* $\operatorname{tr}\Big[\boldsymbol{P}[\boldsymbol{A}_0 \otimes \boldsymbol{Q}_{hi}]\Big] = \boldsymbol{D}_{1ih}$.

Lemma 15. *With* $\boldsymbol{X} = \boldsymbol{1}_{N_J} \otimes \boldsymbol{I}_m$,

(i) $\hat{\boldsymbol{B}} = (\boldsymbol{1}'_{N_J}\boldsymbol{z})'/N_J$ *and* $\hat{\boldsymbol{\beta}} = (\boldsymbol{1}'_{N_J}\boldsymbol{z})N_J$ *do not depend on* $\{\boldsymbol{\Psi}_j : j = 1, \ldots, J\}$.

(ii) $SS_0(\hat{\boldsymbol{\beta}}) = \boldsymbol{0}_{m\times m}$.

(iii) $SS_j(\hat{\boldsymbol{\beta}}) \sim W_m(N_j - N_{j-1}, \boldsymbol{D}_j), j = 1, \ldots, J$, *where* $W_m(N_j - N_{j-1}, \boldsymbol{D}_j)$ *is the* m*-dimensional Wishart distribution with* $N_j - N_{j-1}$ *degrees of freedom and parameter* $\boldsymbol{D}_j$.

(iv) $\{SS_j(\hat{\boldsymbol{\beta}}) : j = 1, \ldots, J\}$ *are mutually independent and independent of* $\hat{\boldsymbol{\beta}}$.

(v) $E[SS_j(\hat{\boldsymbol{\beta}})] = (N_j - N_{j-1})\boldsymbol{D}_j$, $\operatorname{Var}(SS_{jhi}(\hat{\boldsymbol{\beta}})) = (N_j - N_{j-1})(\boldsymbol{D}_{jhi}^2 + \boldsymbol{D}_{jhh}\boldsymbol{D}_{jii})$, *and* $\operatorname{Cov}(SS_{jhi}(\hat{\boldsymbol{\beta}}), SS_{jh'i'}\hat{\boldsymbol{\beta}})) = (N_j - N_{j-1})(\boldsymbol{D}_{jhi'}\boldsymbol{D}_{jh'i} + \boldsymbol{D}_{jhh'}\boldsymbol{D}_{jii'})$, *where* $j = 1, \ldots, J$, SS_{jhi} *is the* (h,i)*th element of* $SS_j(\hat{\boldsymbol{\beta}})$ *and* $\boldsymbol{D}_{jhi}$ *is the* (h,i)*th element of* $\boldsymbol{D}_j$.

Proof.

(i) $\hat{\boldsymbol{B}} = [\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{Z} = [N_J\boldsymbol{D}_1^{-1}]^{-1}[\boldsymbol{1}_{N_J} \otimes \boldsymbol{D}_1^{-1}]'\boldsymbol{Z} = (1/N_J)(\boldsymbol{I}_1 \otimes \boldsymbol{D}_1)(\boldsymbol{1}'_{N_J} \otimes \boldsymbol{D}_1^{-1})\boldsymbol{Z} = (1/N_J)(\boldsymbol{1}'_{N_J} \otimes \boldsymbol{I}_m)\boldsymbol{Z} = \operatorname{vec}((1/N_J)\boldsymbol{1}'_{N_J}\boldsymbol{z})$, where the first equality holds because of Lemma 7 (viii). Since $\hat{\boldsymbol{B}}$ and $\hat{\boldsymbol{\beta}}$ only depend on $\boldsymbol{z}$, they do not depend on $\{\boldsymbol{\Psi}_j : j = 1, \ldots, J\}$.

(ii) Since $\boldsymbol{x}\hat{\boldsymbol{\beta}} = \mathbf{1}_{N_J}((1/N_J)\mathbf{1}'_{N_J})\boldsymbol{z} = \boldsymbol{A}_0\boldsymbol{z}$, we have $SS_0(\hat{\boldsymbol{\beta}}) = (\boldsymbol{z} - \boldsymbol{x}\hat{\boldsymbol{\beta}})'\boldsymbol{A}_0(\boldsymbol{z} - \boldsymbol{x}\hat{\boldsymbol{\beta}}) = \boldsymbol{z}'(\boldsymbol{I}_{N_J} - \boldsymbol{A}_0)'\boldsymbol{A}_0(\boldsymbol{I}_{N_J} - \boldsymbol{A}_0)\boldsymbol{z} = \mathbf{0}_{m\times m}$, where the last equality holds because $\boldsymbol{A}_0(\boldsymbol{I}_{N_J} - \boldsymbol{A}_0) = \boldsymbol{A}_0 - \boldsymbol{A}_0^2 = \mathbf{0}_{N_J\times N_J}$.

(iii) For $j = 1,\ldots,J$, $SS_j(\hat{\boldsymbol{\beta}}) = (\boldsymbol{z} - \boldsymbol{x}\hat{\boldsymbol{\beta}})'(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})(\boldsymbol{z} - \boldsymbol{x}\hat{\boldsymbol{\beta}}) = \boldsymbol{z}'(\boldsymbol{I}_{N_J} - \boldsymbol{A}_0)'(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})(\boldsymbol{I}_{N_J} - \boldsymbol{A}_0)\boldsymbol{z} = \boldsymbol{z}'(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\boldsymbol{z}$, where the last equality holds because of Lemma 7 (i). $\text{vec}((\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\boldsymbol{z}) = [(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\otimes\boldsymbol{I}_m]\boldsymbol{Z} \sim N([(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\otimes\boldsymbol{I}_m]\boldsymbol{X}\boldsymbol{B}, [(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\otimes\boldsymbol{I}_m]'\boldsymbol{\Omega}[(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\otimes\boldsymbol{I}_m]) \sim N(\mathbf{0}_{N_Jm}, (\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\otimes\boldsymbol{D}_j)$, where the first equality holds because of (8), the first $\sim$ holds because of $\boldsymbol{Z} \sim N(\boldsymbol{X}\boldsymbol{B}, \boldsymbol{\Omega})$, and the second $\sim$ holds because we can use Lemma 7 (iii) and (vi) to obtain $[(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\otimes\boldsymbol{I}_m]\boldsymbol{X}\boldsymbol{B} = [(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\otimes\boldsymbol{I}_m](\mathbf{1}_{N_J}\otimes\boldsymbol{I}_m)\boldsymbol{B} = \left\{\left[(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\mathbf{1}_{N_J}\right]\otimes\boldsymbol{I}_m\right\}\boldsymbol{B} = \{[\mathbf{1}_{N_J} - \mathbf{1}_{N_J}]\otimes\boldsymbol{I}_m\}\boldsymbol{B} = \mathbf{0}_{N_Jm}$ and $[(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\otimes\boldsymbol{I}_m]'\boldsymbol{\Omega}[(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\otimes\boldsymbol{I}_m] = [(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\otimes\boldsymbol{I}_m][(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\otimes\boldsymbol{D}_j] = [(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})^2]\otimes(\boldsymbol{I}_m\boldsymbol{D}_j) = (\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\otimes\boldsymbol{D}_j$. Then $(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\boldsymbol{z} \sim N_{N_J\times m}(\mathbf{0}_{N_J\times m}, (\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\otimes\boldsymbol{D}_j)$ (Appendix C, Lauritzen (1996)). Since $\boldsymbol{A}_j - \boldsymbol{A}_{j-1}$ is an idempotent matrix which is a generalized inverse of itself, from Proposition C.13 of Lauritzen (1996) and Lemma 7 (v), we have $SS_j(\hat{\boldsymbol{\beta}}) = \boldsymbol{z}'(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\boldsymbol{z} = [(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\boldsymbol{z}]'(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})^-[(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\boldsymbol{z}] \sim W_m(N_j - N_{j-1}, \boldsymbol{D}_j)$, where $(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})^-$ denotes a generalized inverse of $(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})$.

(iv) In the proof of part (iii), we showed that for $j = 1,\ldots,J$, $SS_j(\hat{\boldsymbol{\beta}})$ is a function of $(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\boldsymbol{z}$. Hence it suffices to show that $\{(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\boldsymbol{z} : j = 1,\ldots,J\}$ are mutually independent and independent of $\hat{\boldsymbol{\beta}}$. For $1 \le j < k \le J$,

$$\begin{aligned}
&\text{Cov}\left(\text{vec}((\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\boldsymbol{z}),\quad \text{vec}((\boldsymbol{A}_k - \boldsymbol{A}_{k-1})\boldsymbol{z})\right)\\
&= \text{Cov}\left([(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\otimes\boldsymbol{I}_m]\boldsymbol{Z},\quad [(\boldsymbol{A}_k - \boldsymbol{A}_{k-1})\otimes\boldsymbol{I}_m]\boldsymbol{Z}\right)\\
&= [(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\otimes\boldsymbol{I}_m]'\boldsymbol{\Omega}[(\boldsymbol{A}_k - \boldsymbol{A}_{k-1})\otimes\boldsymbol{I}_m]\\
&= [(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\otimes\boldsymbol{I}_m][(\boldsymbol{A}_k - \boldsymbol{A}_{k-1})\otimes\boldsymbol{D}_k]\\
&= [(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})(\boldsymbol{A}_k - \boldsymbol{A}_{k-1})]\otimes[\boldsymbol{I}_m\boldsymbol{D}_k]\\
&= \mathbf{0}_{N_J\times N_J}\otimes\boldsymbol{D}_k\\
&= \mathbf{0}_{N_Jm\times N_Jm},
\end{aligned}$$

where the first equality holds because of (8), the third equality holds because of Lemma 7 (vi), and the fourth equality holds because $(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})(\boldsymbol{A}_k - \boldsymbol{A}_{k-1}) = \mathbf{0}_{N_J\times N_J}$. Hence $\{SS_j(\hat{\boldsymbol{\beta}}) : j = 1,\ldots,J\}$ are mutually independent. Since $\hat{\boldsymbol{\beta}} = \mathbf{1}'_{N_J}\boldsymbol{z}/N_J$, to show $SS_j(\hat{\boldsymbol{\beta}})$ and $\hat{\boldsymbol{\beta}}$ are independent, it suffices to show that $(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\boldsymbol{z}$ and $\mathbf{1}'_{N_J}\boldsymbol{z}$ are independent.

$$\begin{aligned}
&\mathrm{Cov}\left(\mathrm{vec}((\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\boldsymbol{z}), \quad \mathrm{vec}(\mathbf{1}'_{N_J}\boldsymbol{z})\right)\\
&= \mathrm{Cov}\left([(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{I}_m]\boldsymbol{Z}, \quad [\mathbf{1}'_{N_J} \otimes \boldsymbol{I}_m]\boldsymbol{Z})\right)\\
&= [(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{I}_m]'\boldsymbol{\Omega}[\mathbf{1}_{N_J} \otimes \boldsymbol{I}_m]\\
&= [(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{D}_j][\mathbf{1}_{N_J} \otimes \boldsymbol{I}_m]\\
&= [(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\mathbf{1}_{N_J}] \otimes [\boldsymbol{D}_j\boldsymbol{I}_m]\\
&= \mathbf{0}_{N_J \times N_J} \otimes [\boldsymbol{D}_j\boldsymbol{I}_m]\\
&= \mathbf{0}_{N_J m \times N_J m},
\end{aligned}$$

where the first equality holds because of (8), the third equality holds because of Lemma 7 (vi), and the fifth equality holds because of Lemma 7 (iii).

(v) Use the formulas for the Wishart distribution from Appendix C.2 of Lauritzen (1996).

Lemma 16. *For* $j = 1, \ldots, J$ *and* $\boldsymbol{D}_j = (d_{j1} - d_{j2})\boldsymbol{I}_m + d_{j2}\mathbf{1}_m\mathbf{1}'_m$,

$$\begin{aligned}
\mathrm{tr}[\boldsymbol{D}_j^2] &= m(d_{j1}^2 + (m-1)d_{j2}^2)\\
\mathrm{tr}\left[[\boldsymbol{D}_j(\mathbf{1}_m\mathbf{1}'_m - \boldsymbol{I}_m)]^2\right] &= m(m-1)\left[d_{j1}^2 + 2(m-2)d_{j1}d_{j2} + (m^2 - 3m + 3)d_{j2}^2\right].
\end{aligned}$$

References

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Method*. Springer, New York.

Chou, K. C., Willsky, A. S. and Nikoukhah, R. (1994). Multiscale systems, Kalman filters, and Riccati equation. *IEEE Trans. Automat. Control* **39**, 479-492.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University, Cambridge.

Harville, D. A. (1985). Decomposition of prediction error. *J. Amer. Statist. Assoc.* **80**, 132-138.

Harville, D. A. (1997). *Matrix Algebra from A Statistician's Perspective*. Springer, New York.

Huang, H.-C., Cressie, N. and Gabrosek, J. (2002). Fast, resolution-consistent spatial prediction of global processes from satellite data. *J. Comput. Graph. Statist.* **11**, 1-26.

Huang, H.-C. and Cressie, N. (1997). Multiscale spatial modeling. In 1997 *Proceedings of the Section on Statistics and the Environment*, 49-54. American Statistical Association, Alexandria, VA.

Johannesson, G. and Cressie, N. (2004). Variance-covariance modelling and estimation for multi-resolution spatial models. In *geoEnv IV – Geostatistics for Environmental Applications* (Edited by X. Sanchez-Vila, C. J and J. J. Gomez-Hernandez), 319-330. Kluwer, Dordrecht.

Jørgensen, B., Lundbye-Christensen, S., Song, S. and Sun, L. (1999). A state space model for multivariate longitudinal count data. *Biometrika* **86**, 169-181.

Lauritzen, S. L. (1996). *Graphical Models*. Clarendon, Oxford.

Luettgen, M. R. (1993). Image processing with multiscale stochastic models. Ph.D. Thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science.

Luettgen, M. R. and Willsky, A. S. (1995). Multiscale smoothing error models. *IEEE Trans. Automat. Control* **40**, 173-175.

Rao, P. S. R. S. and Heckler, C. E. (1998). Multivariate one-way random effects model. *Amer. J. Math. Management Sci.* **18**, 109-128.

Richardson, A. M. and Welsh, A. H. (1994). Asymptotic properties of restricted maximum likelihood (REML) estimates for hierarchical mixed linear models. *Austral. J. Statist.* **36**, 31-43.

Searle, S. R. (1997). *Linear Models*. Wiley, New York.

Smyth, G. K. and Verbyla, A. P. (1996). A conditional likelihood approach to residual maximum likelihood estimation in generalized linear models. *J. Roy. Statist. Soc. Ser. B* **58**, 565-572.

Yue, W. and Zhu, J. (2005). On estimation and prediction for multivariate multiresolution tree-structured spatial linear models. Technical Report 1102R, Department of Statistics, University of Wisconsin, Madison, WI.

Zhu, J. and Yue, W. (2005). A multiresolution tree-structured spatial linear model. *J. Comput. Graph. Statist.* **14**, 168-184.

Department of Statistics, University of Wisconsin - Madison, 1300 University Avenue, Madison, WI 53706, U.S.A.

E-mail: wyue@stat.wisc.edu

Department of Statistics, University of Wisconsin - Madison, 1300 University Avenue, Madison, WI 53706, U.S.A.

E-mail: jzhu@stat.wisc.edu

(Received January 2005; accepted July 2005)