On Estimation and Prediction for Multivariate Multiresolution Tree-Structured Spatial Linear Models

Technical Report 1102 R

Wei Yue and Jun Zhu

Department of Statistics

University of Wisconsin - Madison

1300 University Avenue

Madison, WI 53706

Email: jzhu@stat.wisc.edu

June 9, 2005

Abstract

Multiresolution tree-structured models prove to be attractive for dealing with large amounts of spatial data. 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, Huang et al., 2002). In this article, we extend the multiresolution tree-structured model to account for multiple response variables simultaneously. 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 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. For illustration, we analyze simulated data as well as a real data set regarding major crop distributions.

1. Introduction

For a spatial random process $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in D\}$, where D is a spatial domain of interest in $\mathbb{R}^2$, 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 it involves operations of order $\mathcal{O}(n^3)$. Multiresolution tree-structured models have been developed in recent years to overcome the computational difficulties kriging faces by imposing a multiresolution tree structure on the latent process $Y(\cdot)$. The multiresolution tree-structured model was first proposed by Chou et al. (1994). Huang and Cressie (1997), Huang et al. (2002), Zhu and Yue (2004) further developed the model to accommodate mass balance across resolutions and apply the methodology to process spatial data. 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 j^{th} 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 j^{th} resolution to its child 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 k^{th} cell on the j^{th} 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 et al. (2002), and Zhu and Yue (2004) to account for multiple response variables simultaneously. That is, we consider a multivariate version of MTSLM, which we call multivariate multiresolution tree-structured spatial linear model (MMTSLM). The motivating example is the distribution of major crops in United States and some parts of Canada. The data set is part of a larger data base of major crops across the world (Leff et al., 2004). Several features of the data challenge both modeling and computation. One feature is the large size of the data, which are at the continental scale and consist of thousands of grid cells for United States alone. Another feature is the multiple response variables corresponding to different crop types. Here we focus on modeling the major crop fractions and mapping the latent processes representing the true crop fractions.

There are several challenges in extending the multiresolution tree-structured spatial linear model from univariate to multivariate response variables, due to the increased model complexity. 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 way of parameter estimation is maximum likelihood (ML) using an EM algorithm. 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.

Here we propose novel approaches to address both these two concerns. For spatial process prediction, we develop 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. We use computer simulation to verify the the analytical results and compare the finite-sample properties of MLE and

REML when there is no analytical results.

In Section 2, we describe the multivariate multiresolution tree-structured spatial linear model (MMT-SLM). 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 the distributional properties. In Section 5, we illustrate the theory and methods by analyzing simulated data as well as the major crop data set. A conclusion is given in Section 6.

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:

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

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 k^{th} node on the j^{th} 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:

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

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:

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

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:

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

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 i^{th} 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 i^{th} 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:

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

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 0^{th} resolution, which has the N_1 root nodes as its child nodes. Then $N_0 \equiv 1$, $n_0 \equiv N_1$, and $ch(0,1) \equiv \{(1,1), \ldots, (1, N_1)\}$. For the j^{th}

resolution, let the vector node $\{j,k\} \equiv \{(j,(k-1)n_{j-1}+1),\ldots,(j,(k-1)n_{j-1}+n_{j-1})\}$ denote the group of nodes that share a common parent, where $k=1,\ldots,N_{j-1}$ $j=1,\ldots,J$. 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 j^{th} resolution is the number of (scalar) parent nodes on the $(j-1)^{th}$ resolution (i.e. N_{j-1}), we have $k=1,\ldots,N_{j-1}$ for $\{j,k\}$.

Now, we define a vectorization operator $\rightarrow$ such that for an $n\times m$ matrix $\boldsymbol{A}=[a_{ij}],\ i=1,\ldots,n,\ j=1,\ldots,m$, $\mathrm{vec}(\boldsymbol{A})=\vec{\boldsymbol{A}} \equiv (a_{11},\ldots,a_{1m},\ldots,a_{n1},\ldots,a_{nm})'$. 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

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

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 $\boldsymbol{C}$,

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

((B.5), Appendix B, Lauritzen, 1996), the MMTSLM can now be rewritten in the vector form as:

$$\begin{aligned}
\boldsymbol{Z}_{j,k} &= \boldsymbol{Y}_{j,k} + \boldsymbol{e}_{j,k},\ \boldsymbol{e}_{j,k} \sim N(\boldsymbol{0}_{n_{j-1}m}, \boldsymbol{I}_{n_{j-1}} \otimes \boldsymbol{\Phi}_j);\ k=1,\ldots,N_{j-1},\ j=1,\ldots,J, \qquad (9)\\
\boldsymbol{Y}_{j,k} &= \boldsymbol{X}_{j,k}\boldsymbol{B} + \boldsymbol{U}_{j,k}, \qquad (10)\\
\boldsymbol{U}_{1,1} &\sim N(\boldsymbol{0}_{n_0 m}, \boldsymbol{H}_1 \otimes \boldsymbol{\Sigma}_1), \qquad (11)\\
\boldsymbol{U}_{j,k} &= \boldsymbol{A}_{j,k}\boldsymbol{U}_{pa\{j,k\}} + \boldsymbol{W}_{j,k},\ \boldsymbol{W}_{j,k} \sim N(\boldsymbol{0}_{n_{j-1}m}, \boldsymbol{H}_j \otimes \boldsymbol{\Sigma}_j);\ k=1,\ldots,N_{j-1},\ j=2,\ldots,J, \qquad (12)
\end{aligned}$$

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 i^{th} column and $\boldsymbol{0}_{n_{j-1}}$ in the other columns if the scalar parent node $(j-1,k)$ is the i^{th} node within the vector node $pa\{j,k\}$. Here $\boldsymbol{W}_{j,k}$ are mutually independent and are independent of $\boldsymbol{U}_{pa\{j,k\}}$.

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 j^{th} 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. Also let $\boldsymbol{X}_{j\cdot} \equiv [\boldsymbol{X}'_{j,1}, \ldots, \boldsymbol{X}'_{j,N_{j-1}}]'$ denote the covariates on the j^{th} resolution.

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

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

Further, $\text{cov}(\boldsymbol{U}_{j\cdot}, \boldsymbol{U}_{j'\cdot}) = \text{var}(\boldsymbol{U}_{j\cdot})(\boldsymbol{I}_{N_j} \otimes \mathbf{1}'_{n_j \cdots n_{j'-1}} \otimes \boldsymbol{I}_m)$; $1 \leq j < j' \leq J$. A derivation of $\text{var}(\boldsymbol{U}_{j\cdot})$ and $\text{cov}(\boldsymbol{U}_{j\cdot}, \boldsymbol{U}_{j'\cdot})$ is given in Appendix A.

Next, by (10), $E(\boldsymbol{Y}_{j\cdot}) = \boldsymbol{X}_{j\cdot}\boldsymbol{B}$, $\text{var}(\boldsymbol{Y}_{j\cdot}) = \text{var}(\boldsymbol{U}_{j\cdot})$, and $\text{cov}(\boldsymbol{Y}_{j\cdot}, \boldsymbol{Y}_{j'\cdot}) = \text{cov}(\boldsymbol{U}_{j\cdot}, \boldsymbol{U}_{j'\cdot}) = \text{var}(\boldsymbol{U}_{j\cdot})(\boldsymbol{I}_{N_j} \otimes \mathbf{1}'_{n_j \cdots n_{j'}-1} \otimes \boldsymbol{I}_m)$; $1 \leq j < j' \leq J$. Hence the latent process $\{\boldsymbol{Y}_{j,k}\}$ features a linear regression mean. Moreover, the variance-covariance structure of $\{\boldsymbol{Y}_{j,k}\}$ is identical to that of $\{\boldsymbol{U}_{j,k}\}$ shown in (13). Finally, by (9), $E(\boldsymbol{Z}_{j\cdot}) = \boldsymbol{X}_{j\cdot}\boldsymbol{B}$, $\text{var}(\boldsymbol{Z}_{j\cdot}) = \text{var}(\boldsymbol{U}_{j\cdot}) + \boldsymbol{I}_{N_j} \otimes \boldsymbol{\Phi}_j$; $j = 1, \ldots, J$. Also $\text{cov}(\boldsymbol{Z}_{j\cdot}, \boldsymbol{Z}_{j'\cdot}) = \text{cov}(\boldsymbol{Y}_{j\cdot}, \boldsymbol{Y}_{j'\cdot}) = \text{cov}(\boldsymbol{U}_{j\cdot}, \boldsymbol{U}_{j'\cdot}) = \text{var}(\boldsymbol{U}_{j\cdot})(\boldsymbol{I}_{N_j} \otimes \mathbf{1}'_{n_j \cdots n_{j'-1}} \otimes \boldsymbol{I}_m)$; $1 \leq j < j' \leq J$.

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 will 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, 2004). Note that when the measurement error variances are larger, the predicted values of $\{\boldsymbol{Y}_{j,k}\}$ tend to be smoother; whereas when the measurement error variances are smaller, 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 et al. (2002) and featured in Zhu and Yue (2004) 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, $n_j^{-1}\left(\mathbf{1}'_{n_j}\boldsymbol{y}_{ch(j,k)}\right) = \boldsymbol{y}_{j,k}$, 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 following conditions are sufficient for an MMTSLM to have mass balance:

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

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 $\boldsymbol{H}_{j+1} = \frac{n_j}{n_j - 1}(\boldsymbol{I}_{n_j} - \frac{1}{n_j}\mathbf{1}_{n_j}\mathbf{1}'_{n_j})$; $j = 1, \ldots, J-1$ where $\boldsymbol{I}_{N_1}$ is the $N_1 \times N_1$ identity matrix. Further, we can simplify (13) to:

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

and $\operatorname{cov}(\boldsymbol{U}_{j\cdot}, \boldsymbol{U}_{j'\cdot}) = \operatorname{var}(\boldsymbol{U}_{j\cdot})(\boldsymbol{I}_{N_j} \otimes \mathbf{1}'_{n_j \cdots n_{j'-1}} \otimes \boldsymbol{I}_m)$; $1 \le j < j' \le J$. When the response variable is univariate with $m = 1$, (15) reduces to (9) of Zhu and Yue (2004).

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., Section 3.2.3, Harvey, 1989). Computing $E(\boldsymbol{U}_{j,k}|\boldsymbol{Z})$ usually involves operations of order $\mathcal{O}(n^3)$ and can be computationally inefficient for a large n. For a univariate response variable, a change-of-resolution Kalman-filter algorithm has been developed to obtain the BLUP, which exploits the multiresolution tree structure and involves operations of only order $\mathcal{O}(n)$ (see, e.g., Chou et al., 1994; Huang et al., 2002; Zhu and Yue, 2004). 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., Chapter 2.7, Searle, 1997). 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 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), which also allows 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 et al. (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 $\|y\| \equiv \langle y,\ y\rangle^{\frac{1}{2}} = \sqrt{E(y^2)}$. Equipped with this inner product, L^2 is a real Hilbert space (Example 2.2.2, Brockwell and Davis, 1991). For $z_1, \ldots, z_n \in L^2$, define a closed subspace of L^2 as $\overline{sp}\{1, z_1, \ldots, z_n\} \equiv \{\mu + \beta_1 z_1 + \ldots + \beta_n z_n : \ \mu \in I\!R, \ \beta_i \in I\!R, \ i = 1, \ldots, n\}$. 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 $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \boldsymbol{Y} \in L_m^2$, 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 $\|\boldsymbol{Y}\| \equiv \langle \boldsymbol{Y}, \ \boldsymbol{Y} \rangle^{\frac{1}{2}} = \sqrt{E(\boldsymbol{Y}'\boldsymbol{Y})}$. It follows that L_m^2 is also a real Hilbert space. For $\boldsymbol{Z} = (z_1, \ldots, z_n)' \in L_n^2$, define a closed subspace $\overline{sp}\{\boldsymbol{Z}\}^m$ of L_m^2 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 et al. (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 $\boldsymbol{C}_{Y|Z} \equiv E[(\boldsymbol{Y} - \boldsymbol{m}_{Y|Z})(\boldsymbol{Y} - \boldsymbol{m}_{Y|Z})'] = \mathrm{var}(\boldsymbol{Y} - \boldsymbol{m}_{Y|Z})$ 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 $\boldsymbol{Y} \in L_m^2$, 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 the optimal linear prediction.

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

THEOREM 2. *For $\boldsymbol{Y} = (y_1, \ldots, y_m)' \in L_m^2$ and $\boldsymbol{Z} = (z_1, \ldots, z_n)' \in L_n^2$, 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 $\boldsymbol{C}_{Y|Z} = \mathrm{var}(\boldsymbol{Y}|\boldsymbol{Z})$.*

The proof of Theorems 1 and 2 are given in Appendix B1. 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 where 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 the 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 $\gamma_{j,k} \equiv \mathcal{I}\{\boldsymbol{Z}_{j,k} \text{ is observed}\}$ 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 $k = 1, \ldots, N_{J-1}$:

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

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

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

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

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

where $\boldsymbol{Z}_{de\{j,k\}} \equiv \{\boldsymbol{Z}_{j',k'} : \gamma_{j',k'} = 1, \{j', k'\} \prec \{j, k\}\}$, $ch\{j, k, i\}$ denotes the i^{th} 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 $\boldsymbol{V}_{j,k} \equiv \text{var}(\boldsymbol{U}_{j,k}) = (\mathbf{1}_{n_{j-1}}\mathbf{1}'_{n_{j-1}}) \otimes (\sum_{j'=1}^{j-1} \boldsymbol{\Sigma}_{j'}) + \boldsymbol{H}_j \otimes \boldsymbol{\Sigma}_j$ and $\boldsymbol{V}_{ch\{j,k,i\}} \equiv \text{var}(\boldsymbol{U}_{ch\{j,k,i\}}) = (\mathbf{1}_{n_j}\mathbf{1}'_{n_j}) \otimes (\sum_{j'=1}^{j} \boldsymbol{\Sigma}_{j'}) + \boldsymbol{H}_{j+1} \otimes \boldsymbol{\Sigma}_{j+1}$ can be obtained from $\text{var}(\boldsymbol{U}_{j\cdot})$ and $\text{var}(\boldsymbol{U}_{j+1\cdot})$ as defined in (13). Further,

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

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

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

$$\hat{\boldsymbol{V}}_{j,k|j,k} \equiv \boldsymbol{C}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}_{de\{j,k\}}) = \hat{\boldsymbol{V}}^*_{j,k|j,k} - \gamma_{j,k}\hat{\boldsymbol{V}}^*_{j,k|j,k}(\hat{\boldsymbol{V}}^*_{j,k|j,k} + \boldsymbol{I}_{n_{j-1}} \otimes \boldsymbol{\Phi}_j)^{-1}\hat{\boldsymbol{V}}^*_{j,k|j,k}, \tag{23}$$

where $\boldsymbol{Z}^*_{de\{j,k\}} \equiv \{\boldsymbol{Z}_{j',k'} : \gamma_{j',k'} = 1, \{j',k'\} \prec \{j,k\}, \{j',k'\} \neq \{j,k\}\}$. At the end of the filtering step, the root vector node is reached and hence the BLUP for $\{1,1\}$ is:

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

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 B2, 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 $\hat{\boldsymbol{B}} \equiv [\boldsymbol{X}'\text{var}(\boldsymbol{Z})^{-1}\boldsymbol{X}]^{-1}[\boldsymbol{X}'\text{var}(\boldsymbol{Z})^{-1}\boldsymbol{Z}]$ 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 (see Appendix B3). The MSPE of the elements in $\hat{\boldsymbol{Y}}_{j,k}$ can be obtained from the diagonal elements of the matrix:

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

where $\boldsymbol{V}_{j,k,\cdot,\cdot} \equiv \text{cov}(\boldsymbol{U}_{j,k}, \boldsymbol{Z})$, $\boldsymbol{D} \equiv (\text{var}(\boldsymbol{Z}))^{-1}$, 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 $\text{var}(\hat{\boldsymbol{B}})$ 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 i^{th} 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 to be 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 Appendix B3).

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 and let $\boldsymbol{V} \equiv \text{var}(\boldsymbol{U}), \boldsymbol{\Phi} \equiv \text{var}(\boldsymbol{e})$. Then

$\boldsymbol{Z} \sim N(\boldsymbol{XB}, \boldsymbol{V} + \boldsymbol{\Phi})$ with log-likelihood function:

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

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 B2. The restricted log-likelihood function of $\boldsymbol{Z}$ is:

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

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 et al. (2002) and Zhu and Yue (2004), 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 stable numerically, it often requires a large number of iterations before convergence is achieved. Here we propose a direct algorithm, which is a 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 has 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 i^{th} 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 $\mathcal{L}(\boldsymbol{\theta})$, 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 $\boldsymbol{Z}^s_{j,k}\equiv\{\boldsymbol{Z}_{j',k'}:\gamma_{j',k'}=1, s(j',k')<s(j,k)\}$ (Luettgen, 1993). Using an algorithm similar to the generalized change-of-resolution Kalman filter in Section 3 (see Appendix D for details), we can obtain the BLUP $\hat{\boldsymbol{U}}^s_{j,k}\equiv\boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}^s_{j,k})$ 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:

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

and the log-likelihood function can be factorized into:

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

where $\boldsymbol{\Lambda}_{j,k}$ is invertible because Lemma 3 (ii) in Appendix B2. 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,

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

where

$$\hat{\boldsymbol{B}} = \left[\sum_{\gamma_{j,k}=1}\boldsymbol{X}'_{j,k}\boldsymbol{\Lambda}^{-1}_{j,k}\boldsymbol{X}_{j,k}\right]^{-1}\left[\sum_{\gamma_{j,k}=1}\boldsymbol{X}'_{j,k}\boldsymbol{\Lambda}^{-1}_{j,k}\boldsymbol{Z}_{j,k}\right]. \tag{33}$$

Moreover, we can obtain

$$\mathrm{var}(\hat{\boldsymbol{B}})=\left[\boldsymbol{X}'(\boldsymbol{V}+\boldsymbol{\Phi})^{-1}\boldsymbol{X}\right]^{-1}=\left[\sum_{\gamma_{j,k}=1}\boldsymbol{X}'_{j,k}\boldsymbol{\Lambda}^{-1}_{j,k}\boldsymbol{X}_{j,k}\right]^{-1} \tag{34}$$

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 $\boldsymbol{I}(\boldsymbol{\theta}) \equiv -\frac{\partial^2 \log \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$ evaluated at the MLEs. For $\boldsymbol{B}$, we use $\frac{\partial^2 \log \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{B} \partial \boldsymbol{B}'} = -\sum_{\gamma_{j,k}=1} \left\{ \boldsymbol{X}'_{j,k} \boldsymbol{\Lambda}^{-1}_{j,k} \boldsymbol{X}_{j,k} \right\}$ 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 $\boldsymbol{I}_*(\boldsymbol{\theta}) \equiv -\frac{\partial^2 \log \mathcal{L}_*(\boldsymbol{\eta},\boldsymbol{\zeta})}{\partial(\boldsymbol{\eta}',\boldsymbol{\zeta}')' \partial(\boldsymbol{\eta},\boldsymbol{\zeta})}$ 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 $\boldsymbol{B} = \vec{\boldsymbol{\beta}}$. That is, $\boldsymbol{Z} \sim N(\boldsymbol{X}\boldsymbol{B}, \boldsymbol{\Omega})$, 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, 2004). Further, we restrict our attention to the case where $\boldsymbol{\Omega}$ can be decomposed into:

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

where $a_j \equiv N_J/N_j = n_j \cdots n_{J-1}; j = 0, \ldots, J-1$, $a_J \equiv 1$, $\boldsymbol{A}_j \equiv \frac{1}{a_j}[\boldsymbol{I}_{N_j} \otimes (\mathbf{1}_{a_j} \mathbf{1}'_{a_j})]$ 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

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

Here we assume that $\boldsymbol{\Omega}$ and $\boldsymbol{D}_j$ are invertible. 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 $\boldsymbol{\Psi}_1 = \boldsymbol{\Sigma}_1 - \frac{1}{n_1 - 1} \boldsymbol{\Sigma}_2$, $\boldsymbol{\Psi}_j = \frac{n_{j-1}}{n_{j-1}-1} \boldsymbol{\Sigma}_j - \frac{1}{n_j - 1} \boldsymbol{\Sigma}_{j+1}; j = 2, \ldots, J-1$, and $\boldsymbol{\Psi}_J = \frac{n_{J-1}}{n_{J-1}-1} \boldsymbol{\Sigma}_J + \boldsymbol{\Phi}_J$.

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

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

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

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

where $\boldsymbol{z} = [\boldsymbol{z}'_{J,1},\ldots,\boldsymbol{z}'_{J,N_J}]'$, $\boldsymbol{x} = [\boldsymbol{x}_{J,1},\ldots,\boldsymbol{x}_{J,N_J}]'$, $\boldsymbol{u} = [\boldsymbol{u}'_{J,1},\ldots,\boldsymbol{u}'_{J,N_J}]'$, and $\boldsymbol{\epsilon} = [\boldsymbol{\epsilon}'_{J,1},\ldots,\boldsymbol{\epsilon}'_{J,N_J}]'$ are written in matrix forms based on the scalar nodes, for ease of presentation. It is obvious 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 C, we have

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

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

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

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 C, 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

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

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

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

where we use $\Big[g(h,i)\Big]_{hi}$ 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 C, the score functions are:

$$\frac{\partial \log \mathcal{L}(\boldsymbol{\theta})}{\partial \theta_i} = \operatorname{tr}\Big[\Big(\frac{\partial \log \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{D}_j^{-1}}\Big)'\Big(\frac{\partial \boldsymbol{D}_j^{-1}}{\partial \theta_i}\Big)\Big] \quad (44)$$

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

Theorem 3. *For the MMTSLM (9)–(12), under (35), the score functions in (41), (44), and (45) are unbiased. That is,*

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

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

4.3 *Special Cases*

Here we derive the explicit forms of MLEs and REMLEs when $\boldsymbol{x} = \boldsymbol{1}_{N_J}$, $\boldsymbol{X} = \boldsymbol{1}_{N_J} \otimes \boldsymbol{I}_m$, $\boldsymbol{\beta} = (\mu_1, \dots, \mu_m)$ and $\boldsymbol{B} = (\mu_1, \dots, \mu_m)'$. 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 C, we obtain

$$\begin{array}{lll}
\hat{\boldsymbol{\beta}} = \frac{1}{N_J}\boldsymbol{1}_{N_J}'\boldsymbol{z}, & E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}, & \operatorname{var}(\hat{\boldsymbol{\beta}}') = [\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1} = \frac{\boldsymbol{D}_1}{N_J} \\
\hat{\boldsymbol{D}}_1 = \frac{1}{N_1}SS_1(\hat{\boldsymbol{\beta}}), & E(\hat{\boldsymbol{D}}_1) = \frac{N_1-1}{N_1}\boldsymbol{D}_1, & \operatorname{var}(\hat{\boldsymbol{D}}_{1hi}) = \frac{N_1-1}{N_1^2}(\boldsymbol{D}_{1hi}^2 + \boldsymbol{D}_{1hh}\boldsymbol{D}_{1ii}) \\
\hat{\boldsymbol{D}}_j = \frac{1}{N_j - Nj-1}SS_j(\hat{\boldsymbol{\beta}}), & E(\hat{\boldsymbol{D}}_j) = \boldsymbol{D}_j & \operatorname{var}(\hat{\boldsymbol{D}}_{jhi}) = \frac{1}{N_j-Nj-1}(\boldsymbol{D}_{jhi}^2 + \boldsymbol{D}_{jhh}\boldsymbol{D}_{jii}); \quad j = 2, \dots, J \\
\tilde{\boldsymbol{D}}_j = \frac{1}{N_j - N_{j-1}}SS_j(\hat{\boldsymbol{\beta}}) & E(\tilde{\boldsymbol{D}}_j) = \boldsymbol{D}_j, & \operatorname{var}(\tilde{\boldsymbol{D}}_{jhi}) = \frac{1}{N_j-Nj-1}(\boldsymbol{D}_{jhi}^2 + \boldsymbol{D}_{jhh}\boldsymbol{D}_{jii}); \quad j = 1, \dots, J
\end{array} \quad (47)$$

where we note that $\hat{\boldsymbol{D}}_j = \tilde{\boldsymbol{D}}_j$ for $j = 2, \dots, J$ and $\boldsymbol{D}_{jhi}$ is the $(h,i)^{th}$ element of $\boldsymbol{D}_j; j = 1, \dots, J, h, i = 1, \dots, m$. That is, the MLEs and REMLEs of $\boldsymbol{\beta}$ and D_j are all unbiased, except the MLEs on the coarsest resolution. Furthermore, we could 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),

$$SS_0(\hat{\boldsymbol{\beta}}) = \mathbf{0}_{m\times m},$$
$$SS_j(\hat{\boldsymbol{\beta}}) \sim W_m(N_j - N_{j-1}, \boldsymbol{D}_j); \quad j = 1, \ldots, J,$$

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 C. 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, which 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.

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 $\boldsymbol{D}_j^{-1} = \frac{1}{d_{j1}-d_{j2}}(\boldsymbol{I}_m - \frac{d_{j2}}{d_{j1}+(m-1)d_{j2}}\mathbf{1}_m\mathbf{1}_m')$ has a compound symmetry structure with diagonal elements $d_{j1}^* = \frac{1}{d_{j1}-d_{j2}} - \frac{d_{j2}}{d_{j1}+(m-1)d_{j2}}$ and off-diagonal elements $d_{j2}^* = \frac{-d_{j2}}{d_{j1}+(m-1)d_{j2}}$; $j = 1, \ldots, J$. We have obtained $\frac{\partial \log \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{D}_j^{-1}}$ and $\frac{\partial \log \mathcal{L}_*(\boldsymbol{\eta},\boldsymbol{\zeta})}{\partial \boldsymbol{D}_j^{-1}}$ in (42)–(43) . Now we compute $\frac{\partial \boldsymbol{D}_j^{-1}}{\partial \theta_i}$. Using (44)–(45), $\frac{\partial \boldsymbol{D}_j^{-1}}{\partial d_{j1}^*} = \boldsymbol{I}_m$, and $\frac{\partial \boldsymbol{D}_j^{-1}}{\partial d_{j2}^*} = \mathbf{1}_m\mathbf{1}_m' - \boldsymbol{I}_m$, we obtain,

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

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

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

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

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

Using Lemmas 5 (vii), 15 (iii) and (v), and 16 in Appendix C, for $j = 1, \dots, J$,

$$\begin{aligned} E(\mathrm{tr}[SS_j(\hat{\boldsymbol{\beta}})]) &= (N_j - N_{j-1})m d_{j1} \\ \mathrm{var}(\mathrm{tr}[SS_j(\hat{\boldsymbol{\beta}})]) &= 2(N_j - N_{j-1})m(d_{j1}^2 + (m-1)d_{j2}^2) \\ E(\mathrm{tr}[SS_j(\hat{\boldsymbol{\beta}})](\mathbf{1}_m\mathbf{1}_m' - \boldsymbol{I}_m)) &= (N_j - N_{j-1})m(m-1)d_{j2} \\ \mathrm{var}(\mathrm{tr}[SS_j(\hat{\boldsymbol{\beta}})](\mathbf{1}_m\mathbf{1}_m' - \boldsymbol{I}_m)] &= 2(N_j - N_{j-1})m(m-1)\left[d_{j1}^2 + 2(m-2)d_{j1}d_{j2} + (m^2 - 3m + 3)d_{j2}^2\right] \end{aligned}$$

Then

$$\begin{array}{ll} E(\hat{d}_{11}) = \frac{N_1 - 1}{N_1} d_{11}, & \mathrm{var}(\hat{d}_{11}) = \frac{2(N_1-1)}{N_1^2 m}(d_{11}^2 + (m-1)d_{12}^2) \\ E(\hat{d}_{j1}) = d_{j1}, & \mathrm{var}(\hat{d}_{j1}) = \frac{2}{(N_j - N_{j-1})m}(d_{j1}^2 + (m-1)d_{j2}^2) \quad j = 2, \dots, J \\ E(\hat{d}_{12}) = \frac{N_1 - 1}{N_1} d_{12}, & \mathrm{var}(\hat{d}_{12}) = \frac{2(N_1-1)}{N_1^2 m(m-1)}\left[d_{11}^2 + 2(m-2)d_{11}d_{12} + (m^2 - 3m + 3)d_{12}^2\right] \\ E(\hat{d}_{j2}) = d_{j2}, & \mathrm{var}(\hat{d}_{j2}) = \frac{2}{(N_j - N_{j-1})m(m-1)}\left[d_{j1}^2 + 2(m-2)d_{j1}d_{j2} + (m^2 - 3m + 3)d_{j2}^2\right] \quad j = 2, \dots, J \\ E(\tilde{d}_{j1}) = d_{j1}, & \mathrm{var}(\tilde{d}_{j1}) = \frac{2}{(N_j - N_{j-1})m}(d_{j1}^2 + (m-1)d_{j2}^2) \quad j = 1, \dots, J \\ E(\tilde{d}_{j2}) = d_{j2}, & \mathrm{var}(\tilde{d}_{j2}) = \frac{2}{(N_j - N_{j-1})m(m-1)}\left[d_{j1}^2 + 2(m-2)d_{j1}d_{j2} + (m^2 - 3m + 3)d_{j2}^2\right] \quad j = 1, \dots, J \end{array} \quad (52)$$

We note that the MLEs and REMLEs are all unbiased with the exception of the MLEs on the coarsest resolution. 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: $\boldsymbol{H}_j = \boldsymbol{I}_{n_{j-1}}$; $j = 1, \dots, J$, which we call the independence case, and $\boldsymbol{H}_1 = \boldsymbol{I}_{N_1}, \boldsymbol{H}_{j+1} = \frac{n_j}{n_j - 1}(\boldsymbol{I}_{n_j} - \frac{1}{n_j}\mathbf{1}_{n_j}\mathbf{1}_{n_j}')$; $j = 1, \dots, J-1$, 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 B2. It is easy to obtain $\boldsymbol{\Sigma}_j = \boldsymbol{\Psi}_j = \frac{1}{a_j}(\boldsymbol{D}_j - \boldsymbol{D}_{j+1})$; $j = 1, \dots, J-1$ and

$\boldsymbol{\Sigma}_J = \boldsymbol{\Psi}_J - \boldsymbol{\Phi}_J = \boldsymbol{D}_J - \boldsymbol{\Phi}_J$. Using (47) and Lemma 15 in Appendix C, we have

$$\begin{array}{llll}
\hat{\boldsymbol{\Sigma}}_1 = \frac{1}{a_1}(\hat{\boldsymbol{D}}_1 - \hat{\boldsymbol{D}}_2), & E(\hat{\boldsymbol{\Sigma}}_1) = \boldsymbol{\Sigma}_1 - \frac{\boldsymbol{D}_1}{N_J}, & \mathrm{var}(\hat{\boldsymbol{\Sigma}}_{1hi}) = \frac{1}{a_1^2}[\mathrm{var}(\hat{\boldsymbol{D}}_{1hi}) + \mathrm{var}(\hat{\boldsymbol{D}}_{2hi})] & \\
\hat{\boldsymbol{\Sigma}}_j = \frac{1}{a_j}(\hat{\boldsymbol{D}}_j - \hat{\boldsymbol{D}}_{j+1}), & E(\hat{\boldsymbol{\Sigma}}_j) = \boldsymbol{\Sigma}_j, & \mathrm{var}(\hat{\boldsymbol{\Sigma}}_{jhi}) = \frac{1}{a_j^2}[\mathrm{var}(\hat{\boldsymbol{D}}_{jhi}) + \mathrm{var}(\hat{\boldsymbol{D}}_{j+1hi})]; & j = 2, \ldots, J-1 \\
\hat{\boldsymbol{\Sigma}}_J = \hat{\boldsymbol{D}}_J - \boldsymbol{\Phi}_J, & E(\hat{\boldsymbol{\Sigma}}_J) = \boldsymbol{\Sigma}_J, & \mathrm{var}(\hat{\boldsymbol{\Sigma}}_{Jhi}) = \frac{1}{a_J^2}\mathrm{var}(\hat{\boldsymbol{D}}_{Jhi}) & \\
\tilde{\boldsymbol{\Sigma}}_j = \frac{1}{a_j}(\tilde{\boldsymbol{D}}_j - \tilde{\boldsymbol{D}}_{j+1}), & E(\tilde{\boldsymbol{\Sigma}}_j) = \boldsymbol{\Sigma}_j, & \mathrm{var}(\tilde{\boldsymbol{\Sigma}}_{jhi}) = \frac{1}{a_j^2}[\mathrm{var}(\tilde{\boldsymbol{D}}_{jhi}) + \mathrm{var}(\tilde{\boldsymbol{D}}_{j+1hi})]; & j = 1, \ldots, J-1 \\
\tilde{\boldsymbol{\Sigma}}_J = \tilde{\boldsymbol{D}}_J - \boldsymbol{\Phi}_J, & E(\tilde{\boldsymbol{\Sigma}}_J) = \boldsymbol{\Sigma}_J, & \mathrm{var}(\tilde{\boldsymbol{\Sigma}}_{Jhi}) = \frac{1}{a_J^2}\mathrm{var}(\tilde{\boldsymbol{D}}_{Jhi}) &
\end{array} \quad (53)$$

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

If $\boldsymbol{\Sigma}_j$ 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 $d_{j1} = \sum_{k=j}^{J} a_k\sigma_{k1} + \phi_J$ and $d_{j2} = \sum_{k=j}^{J} a_k\sigma_{k2}$; $j = 1, \ldots, J$. Hence using (52), we have

$$\begin{array}{llll}
\hat{\sigma}_{11} = \frac{1}{a_1}(\hat{d}_{11} - \hat{d}_{21}), & E(\hat{\sigma}_{11}) = \sigma_{11} - \frac{d_{11}}{N_J}, & \mathrm{var}(\hat{\sigma}_{11}) = \frac{1}{a_1^2}[\mathrm{var}(\hat{d}_{11}) + \mathrm{var}(\hat{d}_{21})] & \\
\hat{\sigma}_{j1} = \frac{1}{a_j}(\hat{d}_{j1} - \hat{d}_{j+1\,1}), & E(\hat{\sigma}_{j1}) = \sigma_{j1}, & \mathrm{var}(\hat{\sigma}_{j1}) = \frac{1}{a_j^2}[\mathrm{var}(\hat{d}_{j1}) + \mathrm{var}(\hat{d}_{j+1\,1})] & j = 2, \ldots, J-1 \\
\hat{\sigma}_{J1} = \hat{d}_{J1} - \phi_J, & E(\hat{\sigma}_{J1}) = \sigma_{J1}, & \mathrm{var}(\hat{\sigma}_{J1}) = \mathrm{var}(\hat{d}_{J1}) & \\
\hat{\sigma}_{12} = \frac{1}{a_1}(\hat{d}_{12} - \hat{d}_{22}), & E(\hat{\sigma}_{12}) = \sigma_{12} - \frac{d_{12}}{N_J}, & \mathrm{var}(\hat{\sigma}_{12}) = \frac{1}{a_1^2}[\mathrm{var}(\hat{d}_{12}) + \mathrm{var}(\hat{d}_{22})] & \\
\hat{\sigma}_{j2} = \frac{1}{a_j}(\hat{d}_{j2} - \hat{d}_{j+1\,2}), & E(\hat{\sigma}_{j2}) = \sigma_{j2}, & \mathrm{var}(\hat{\sigma}_{j2}) = \frac{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}, & \mathrm{var}(\hat{\sigma}_{J2}) = \mathrm{var}(\hat{d}_{J2}) & \\
\tilde{\sigma}_{j1} = \frac{1}{a_j}(\tilde{d}_{j1} - \tilde{d}_{j+1\,1}), & E(\tilde{\sigma}_{j1}) = \sigma_{j1}, & \mathrm{var}(\tilde{\sigma}_{j1}) = \frac{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}, & \mathrm{var}(\tilde{\sigma}_{J1}) = \mathrm{var}(\tilde{d}_{J1}) & \\
\tilde{\sigma}_{j2} = \frac{1}{a_j}(\tilde{d}_{j2} - \tilde{d}_{j+1\,2}), & E(\tilde{\sigma}_{j2}) = \sigma_{j2}, & \mathrm{var}(\tilde{\sigma}_{j2}) = \frac{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}, & \mathrm{var}(\tilde{\sigma}_{J2}) = \mathrm{var}(\tilde{d}_{J2}) &
\end{array} \quad (54)$$

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 $\mathbf{\Omega}$ and $\boldsymbol{D}_j; j = 1, \ldots, J$ are in fact invertible because of Lemma 3 (ii) in Appendix B2. Moreover

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

Hence using (47), we have

$$\begin{array}{lll} \hat{\mathbf{\Sigma}}_1 = \frac{1}{a_1}(\hat{\boldsymbol{D}}_1 - \mathbf{\Phi}_J), & E(\hat{\mathbf{\Sigma}}_1) = \mathbf{\Sigma}_1 - \frac{\boldsymbol{D}_1}{N_J}, & \text{var}(\hat{\mathbf{\Sigma}}_{1hi}) = \frac{1}{a_1^2}\text{var}(\hat{\boldsymbol{D}}_{1hi}) \\ \hat{\mathbf{\Sigma}}_j = \frac{n_{j-1}-1}{a_{j-1}}(\hat{\boldsymbol{D}}_j - \mathbf{\Phi}_J), & E(\hat{\mathbf{\Sigma}}_j) = \mathbf{\Sigma}_j, & \text{var}(\hat{\mathbf{\Sigma}}_{jhi}) = \frac{(n_{j-1}-1)^2}{a_{j-1}^2}\text{var}(\hat{\boldsymbol{D}}_{jhi}); \quad j = 2, \ldots, J \\ \tilde{\mathbf{\Sigma}}_1 = \frac{1}{a_1}(\tilde{\boldsymbol{D}}_1 - \mathbf{\Phi}_J), & E(\tilde{\mathbf{\Sigma}}_1) = \mathbf{\Sigma}_1, & \text{var}(\tilde{\mathbf{\Sigma}}_{1hi}) = \frac{1}{a_1^2}\text{var}(\tilde{\boldsymbol{D}}_{1hi}) \\ \tilde{\mathbf{\Sigma}}_j = \frac{n_{j-1}-1}{a_{j-1}}(\tilde{\boldsymbol{D}}_j - \mathbf{\Phi}_J), & E(\tilde{\mathbf{\Sigma}}_j) = \mathbf{\Sigma}_j, & \text{var}(\tilde{\mathbf{\Sigma}}_{jhi}) = \frac{(n_{j-1}-1)^2}{a_{j-1}^2}\text{var}(\tilde{\boldsymbol{D}}_{jhi}); \quad j = 2, \ldots, J \end{array} \tag{56}$$

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

If we assume $\mathbf{\Sigma}_j$ has a compound symmetry structure with diagonal element σ_{j1} and off-diagonal element σ_{j2}; $j = 1, \ldots, J$ and $\mathbf{\Phi}_J$ 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 $d_{11} = a_1\sigma_{11} + \phi_J$, $d_{12} = a_1\sigma_{12}$, $d_{j1} = \frac{a_{j-1}}{n_{j-1}-1}\sigma_{j1} + \phi_J$; and $d_{j2} = \frac{a_{j-1}}{n_{j-1}-1}\sigma_{j2}$; $j = 2, \ldots, J$.

Hence using (48)–(52), we have

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

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

5. Data Example

Here we illustrate the methodology developed in Sections 2–4 by a real data set concerning the distribution of major crops in the United States and some parts of Canada (Leff et al. (2004)). The study area ranges from -124.5^o to -67^o in longitude and from 25^o to 49.5^o in latitude (Figure 1). The data describe the fraction of a grid cell occupied by three major crops, namely, pastures, cereals and oilbearing crops. Three-dimensional perspective plots of the three major crop fractions are shown in Figure 2. Furthermore, we consider topography, including elevation, slope, and cosine of aspect, as potential covariates. In particular, we focus on modeling and predicting the latent processes representing the true major crop fractions.

For constructing a multiresolution tree structure, we consider 32 cells on the coarsest resolution (Figure 1) and a 4-resolution quad-tree so that each of the 32 cells on the first resolution $j = 1$ is further divided into 4, 16, 64 subcells for the finer resolutions $j = 2, \ldots, 4$. Within each cell on the finest resolution, the crop fractions of the three major crops (pastures, cereals and oilbearing) form a multivariate datum for that cell. The cells that lie outside the study area are considered to have missing values and thus do not affect the statistical inference. Thus the MMTSLM model has three response variables for the fractions of pastures, cereals and oilbearing crops and three covariates for elevation, slope, and cosine of aspect. The linear coefficients $\boldsymbol{\beta}$ is a 4×3 matrix where the first row corresponds to the intercepts $(\beta_{11}, \beta_{12}, \beta_{13})$, and the other three rows correspond to the slopes for the three covariates, elevation $(\beta_{21}, \beta_{22}, \beta_{23})$, slope $(\beta_{31}, \beta_{32}, \beta_{33})$, and cosine of aspect $(\beta_{41}, \beta_{42}, \beta_{43})$, respectively. On the j^{th} resolution, the error terms in the MMTSLM are assumed to have a compound symmetric variance structure parameterized by (σ_{j1}, ρ_j), where σ_{j1} is the variance and ρ_j is the correlation coefficient of the variance matrix $\boldsymbol{\Sigma}_j$, $j = 1, \ldots 4$. The correlation matrix $\boldsymbol{H}_1$ is an identity matrix and $\boldsymbol{H}_j$, $j = 2, \ldots, 4$ are correlation matrices assuming mass balance. The variance of measurement errors on the finest resolution is set at $\phi_4 = 50$, based on an exponential variogram fitting of the data.

The MLEs and REMLEs of the model parameters are given in Table 1. Also reported are the standard errors of the MLEs, where the standard errors of the estimates of $(\sigma_{j1}, \rho_j), j = 1, \ldots, 4$, are based on the approximate information matrices and the standard errors of the estimates of $\boldsymbol{\beta}$ are based on (34). Given the fitted MMTSLM with the corresponding MLEs, the generalized change-of-resolution

Kalman filter algorithm described in Section 3 is applied to obtain the BLUP of fractions of pastures, cereals and oilbearing crops. Since the difference between the MLEs and the REMLEs is very small, we consider here only the MLEs for prediction. In particular, the change-of-resolution Kalman-filter algorithm is first applied to the detrended data and the estimated trend is then added back to obtain the final prediction of the crop fractions. Figure 2 shows the best predicted values of crop fractions. Also shown are the root MSPEs associated with the predicted values. The prediction surfaces are in general smoother than the raw data surface, but the main features of the original data seem to be retained.

6. Conclusion

Here we have developed a multivariate version of the multiresolution tree-structured model. Despite the increased model complexity, we have derived fast computational algorithms for both prediction of the latent processes and inference of the model parameters. For latent process prediction, we have established optimal prediction theory and a generalized change-of-resolution Kalman filter algorithm that accommodate singular variance matrices in the latent processes. For parameter estimation, we have established distributional properties of both maximum likelihood and restricted maximum likelihood. The analytical results show that there is only difference between the two estimation methods on the coarsest resolution of the multiresolution tree structure for certain cases. We have also conducted Monte Carlo simulations that have verified both the theoretical results and the numerical algorithms (see Appendix E for details). Finally we have illustrated the methodology by a real data set concerning the major crop distributions.

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 Transaction on Automatic Control* **39**, 479–492.

Eaton, M. L. (1983). *Multivariate statistics. A vector space approach*. Wily, New York.

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

Harville, D. A. (1985). Decomposition of prediction error. *Journal of the American Statistical Association* **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. *Journal of Computational and Graphical Statistics* **11**, 1–26.

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

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.

Leff, B., Ramankutty, N. and Foley, J. A. (2004). Geographic distribution of major crops across the world. *Global Biogeochemical Cycles* **18**, GB1009 doi:10.1029/2003GB002108.

Luettgen, M. R. (1993). Image processing with multiscale stochastic models. PhD Thesis, Massachusetts Institue of Techology, Department of Electrical Engineering and Computer Science.

Luettgen, M. R. and Willsky, A. S. (1995). Multiscale smoothing error models. *IEEE Transaction on Automatic Control* **40**, 173–175.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications, Second Edition*. Wiley, New York.

Richardson, A. M. and Welsh, A. H. (1994). Asymptotic properties of restricted maximum likelihood (REML) estimates for hierarchical mixed linear models. *Australian Journal of statistics* **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. *Journal of the Royal Statistical Society, Series B* **58**, 565–572.

Zhu, J. and Yue, W. (2004). A multiresolution tree-structured spatial linear model. *Journal of Computational and Graphical Statistics* page In press.

Figures

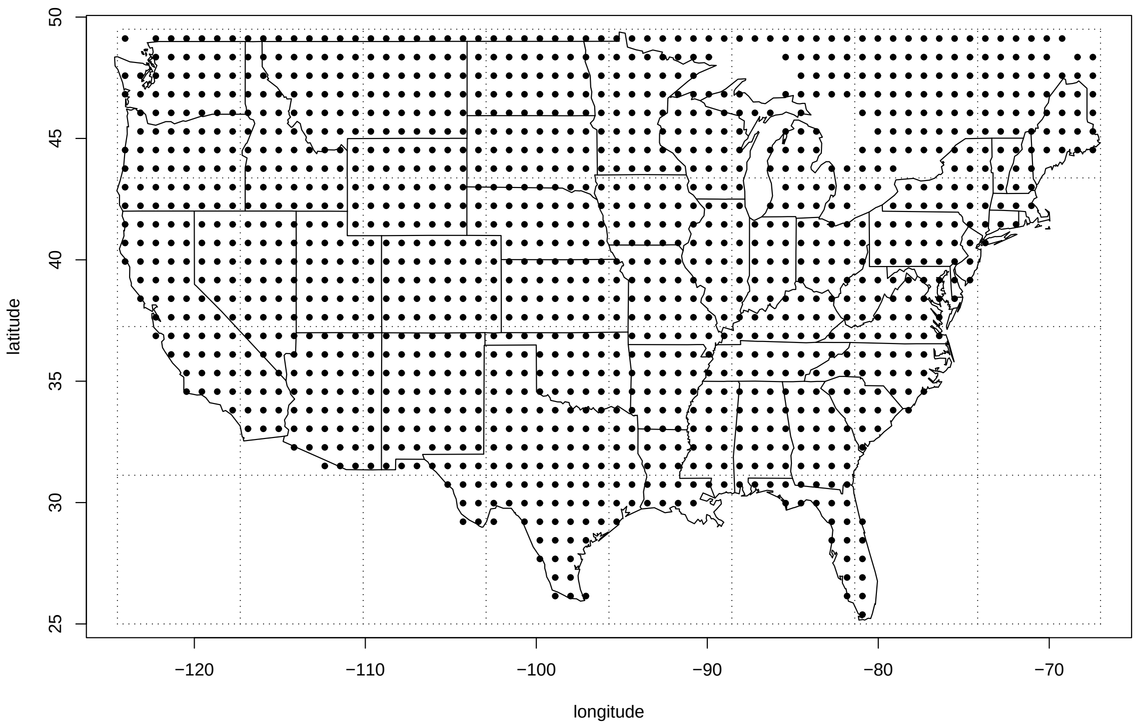

Figure 1: Centroids of the data grid cells in the United States and some parts of Canada.

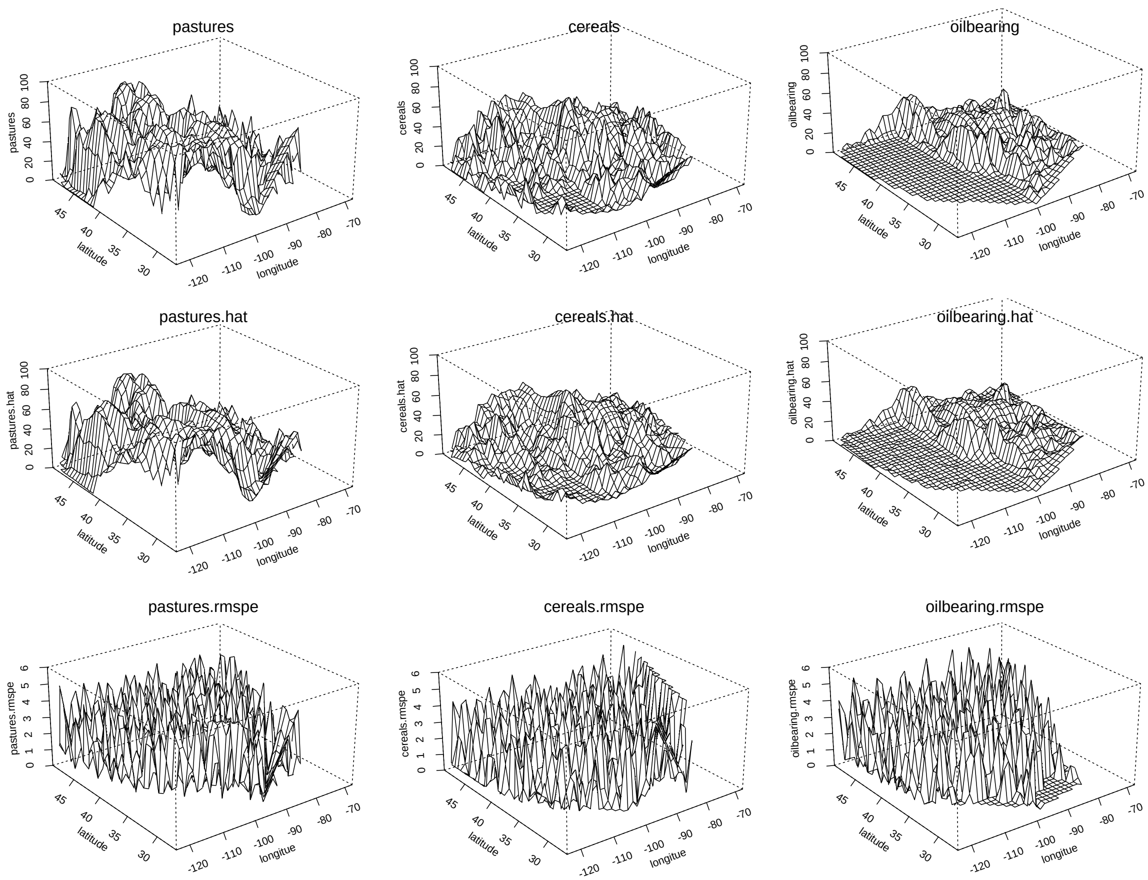

Figure 2: 3D perspective plots of observed fractions (first row); predicted fractions (second row); root mean squared prediction errors (third row) of pastures, cereals, and oilbearing crops.

Tables

	σ_{11}	ρ_1	σ_{21}	ρ_2	σ_{31}	ρ_3	σ_{41}	ρ_4
MLE	131.49	0.0070	94.15	-0.11	77.20	-0.10	33.85	-0.041
StdErr	20.97	0.11	0.72	0.032	0.25	0.049	1.79	0.020
REML	131.24	0.0068	96.11	-0.10	79.14	-0.10	32.57	-0.037
StdErr	0.40	0.11	0.13	0.12	0.13	0.049	0.30	0.038

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	β_{31}	β_{32}	β_{33}	β_{41}	β_{42}	β_{43}
MLE	31.13	29.18	16.07	0.010	-0.0088	-0.0055	-0.000055	-0.000045	-0.000051	1.43	0.084	-0.21
StdErr	1.13	1.13	1.13	0.0011	0.0011	0.0011	0.000018	0.000018	0.000018	0.70	0.70	0.70
REML	31.38	29.15	16.01	0.0099	-0.0088	-0.0054	0.000055	-0.000043	-0.000050	1.42	0.080	-0.20
StdErr	1.13	1.13	1.13	0.0011	0.0011	0.0011	0.000018	0.000018	0.000018	0.70	0.70	0.70

Table 1: Maximum likelihood estimates (MLE) and restricted maximum likelihood estimates (REML) of the model parameters and the corresponding standard errors.

Appendix A: Variance and Covariance of MMTSLM

Proof of (8). Let $\boldsymbol{A}$ be an $m \times n$, $\boldsymbol{B}$ be an $n \times q$, and $\boldsymbol{C}$ be a $p \times q$ matrix. From Appendix B of Lauritzen (1996), if $\boldsymbol{A} = [a_{ij}]$ and $\boldsymbol{C} = [c_{rs}]$, then $(\boldsymbol{A} \otimes \boldsymbol{C})_{ir,js} = a_{ij}b_{rs}$ where $(\boldsymbol{A} \otimes \boldsymbol{C})_{ir,js}$ is the (ir, js) element of $\boldsymbol{A} \otimes \boldsymbol{C}$. Similarly, from Appendix B of Lauritzen (1996), we have

$$\text{vec}(\boldsymbol{ABC}')_{ir} = (\boldsymbol{ABC}')_{ir} = \sum_{vk} (\boldsymbol{A} \otimes \boldsymbol{C})_{ir,vk} \boldsymbol{B}_{vk} = (\boldsymbol{A} \otimes \boldsymbol{C})_{ir,\cdot} \text{vec}(\boldsymbol{B}),$$

where $\text{vec}(\boldsymbol{ABC}')_{ir}$ is the ir-th element of column vector $\text{vec}(\boldsymbol{ABC}')$, $(\boldsymbol{ABC}')_{ir}$ is the (i, r) element of $\boldsymbol{ABC}'$ and $(\boldsymbol{A} \otimes \boldsymbol{C})_{ir,\cdot}$ is the ir-th row of $(\boldsymbol{A} \otimes \boldsymbol{C})$, $\boldsymbol{B}_{vk}$ is the (v, k) element of $\boldsymbol{B}$ and the last equality holds because $\boldsymbol{B}_{vk}$ is the vk-th element of column vector $\text{vec}(\boldsymbol{B})$. Hence $\text{vec}(\boldsymbol{ABC}') = (\boldsymbol{A} \otimes \boldsymbol{C})\text{vec}(\boldsymbol{B})$. □

Proof of (9)–(12). From (1)–(6), we have

$$\begin{aligned}
\boldsymbol{z}_{ch(j-1,k)} &= \boldsymbol{y}_{ch(j-1,k)} + \boldsymbol{\epsilon}_{ch(j-1,k)};\ k = 1, \ldots, N_j,\ j = 1, \ldots, J \\
\boldsymbol{y}_{ch(j-1,k)} &= \boldsymbol{x}_{ch(j-1,k)}\boldsymbol{\beta} + \boldsymbol{u}_{ch(j-1,k)};\ k = 1, \ldots, N_j,\ j = 1, \ldots, J \\
\boldsymbol{u}_{ch(j-1,k)} &= \boldsymbol{D}_{j,k}\boldsymbol{u}_{ch(pa(j-1,k))} + \boldsymbol{\omega}_{ch(j-1,k)};\ k = 1, \ldots, N_j,\ j = 2, \ldots, J
\end{aligned}$$

where $\boldsymbol{D}_{j,k}$ is an $n_{j-1} \times n_{j-2}$ matrix consisting of $\mathbf{1}_{n_{j-1}}$ in the i-th column and $\mathbf{0}_{n_{j-1}}$ in the other columns if the scalar parent node $(j-1, k)$ is the i-th node within the vector node $pa\{j, k\}$. Then using (7) and (8) and $\text{vec}(\boldsymbol{u}_{ch(pa(j-1,k))}) = \boldsymbol{U}_{pa\{j,k\}}$, we obtain (9)–(12). □

Proof of (13). To verify (13), we note that $\text{var}(\boldsymbol{U}_{1\cdot}) = \boldsymbol{H}_1 \otimes \boldsymbol{\Sigma}_1$ and

$$\begin{aligned}
\text{var}(\boldsymbol{U}_{2\cdot}) &= (\boldsymbol{I}_{N_1} \otimes \mathbf{1}_{n_1} \otimes \boldsymbol{I}_m)\text{var}(\boldsymbol{U}_{1\cdot})(\boldsymbol{I}_{N_1} \otimes \mathbf{1}_{n_1} \otimes \boldsymbol{I}_m)' + \text{var}(\boldsymbol{W}_{2\cdot}) \\
&= (\boldsymbol{I}_{N_1} \otimes \mathbf{1}_{n_1} \otimes \boldsymbol{I}_m)(\boldsymbol{H}_1 \otimes \boldsymbol{\Sigma}_1)(\boldsymbol{I}_{N_1} \otimes \mathbf{1}_{n_1} \otimes \boldsymbol{I}_m)' + \boldsymbol{I}_{N_1} \otimes \boldsymbol{H}_2 \otimes \boldsymbol{\Sigma}_2 \\
&= (\boldsymbol{I}_{N_1} \otimes \mathbf{1}_{n_1} \otimes \boldsymbol{I}_m)(\boldsymbol{H}_1 \otimes \boldsymbol{I}_1 \otimes \boldsymbol{\Sigma}_1)(\boldsymbol{I}_{N_1} \otimes \mathbf{1}_{n_1} \otimes \boldsymbol{I}_m)' + \boldsymbol{I}_{N_1} \otimes \boldsymbol{H}_2 \otimes \boldsymbol{\Sigma}_2 \\
&= (\boldsymbol{I}_{N_1}\boldsymbol{H}_1\boldsymbol{I}'_{N_1}) \otimes (\mathbf{1}_{n_1}\boldsymbol{I}_1\mathbf{1}'_{n_1}) \otimes (\boldsymbol{I}_m\boldsymbol{\Sigma}_1\boldsymbol{I}'_m) + \boldsymbol{I}_{N_1} \otimes \boldsymbol{H}_2 \otimes \boldsymbol{\Sigma}_2 \\
&= \boldsymbol{H}_1 \otimes (\mathbf{1}_{n_1}\mathbf{1}'_{n_1}) \otimes \boldsymbol{\Sigma}_1 + \boldsymbol{I}_{N_1} \otimes \boldsymbol{H}_2 \otimes \boldsymbol{\Sigma}_2 \\
&= \boldsymbol{I}_{N_1} \otimes \boldsymbol{H}_2 \otimes \boldsymbol{\Sigma}_2 + \boldsymbol{H}_1 \otimes (\mathbf{1}_{n_1}\mathbf{1}'_{n_1}) \otimes \boldsymbol{\Sigma}_1,
\end{aligned}$$

where $\boldsymbol{I}_1$ in the third equality is an 1×1 identity matrix. Assuming that (13) holds for $\text{var}(\boldsymbol{U}_{j\cdot})$, by induction, we obtain:

$$\begin{aligned}
\text{var}(\boldsymbol{U}_{j+1\cdot}) &= (\boldsymbol{I}_{N_j} \otimes \boldsymbol{1}_{n_j} \otimes \boldsymbol{I}_m)\text{var}(\boldsymbol{U}_{j\cdot})(\boldsymbol{I}_{N_j} \otimes \boldsymbol{1}_{n_j} \otimes \boldsymbol{I}_m)' + \text{var}(\boldsymbol{W}_{j+1\cdot}) \\
&= (\boldsymbol{I}_{N_j} \otimes \boldsymbol{1}_{n_j} \otimes \boldsymbol{I}_m)\Big[\boldsymbol{I}_{N_{j-1}} \otimes \boldsymbol{H}_j \otimes \boldsymbol{\Sigma}_j + \boldsymbol{I}_{N_{j-2}} \otimes \boldsymbol{H}_{j-1} \otimes (\boldsymbol{1}_{n_{j-1}}\boldsymbol{1}'_{n_{j-1}}) \otimes \boldsymbol{\Sigma}_{j-1} \\
&\quad + \ldots + \boldsymbol{I}_{N_1} \otimes \boldsymbol{H}_2 \otimes (\boldsymbol{1}_{n_2\cdots n_{j-1}}\boldsymbol{1}'_{n_2\cdots n_{j-1}}) \otimes \boldsymbol{\Sigma}_2 \\
&\quad + \boldsymbol{I}_{N_0} \otimes \boldsymbol{H}_1 \otimes (\boldsymbol{1}_{n_1\cdots n_{j-1}}\boldsymbol{1}'_{n_1\cdots n_{j-1}}) \otimes \boldsymbol{\Sigma}_1\Big](\boldsymbol{I}_{N_j} \otimes \boldsymbol{1}_{n_j} \otimes \boldsymbol{I}_m)' + \boldsymbol{I}_{N_j} \otimes \boldsymbol{H}_{j+1} \otimes \boldsymbol{\Sigma}_{j+1} \\
&= \boldsymbol{I}_{N_j} \otimes \boldsymbol{H}_{j+1} \otimes \boldsymbol{\Sigma}_{j+1} + \boldsymbol{I}_{N_{j-1}} \otimes \boldsymbol{H}_j \otimes (\boldsymbol{1}_{n_j}\boldsymbol{1}'_{n_j}) \otimes \boldsymbol{\Sigma}_{j-1} \\
&\quad + \ldots + \boldsymbol{I}_{N_1} \otimes \boldsymbol{H}_2 \otimes (\boldsymbol{1}_{n_2\cdots n_j}\boldsymbol{1}'_{n_2\cdots n_j}) \otimes \boldsymbol{\Sigma}_2 \\
&\quad + \boldsymbol{I}_{N_0} \otimes \boldsymbol{H}_1 \otimes (\boldsymbol{1}_{n_1\cdots n_j}\boldsymbol{1}'_{n_1\cdots n_j}) \otimes \boldsymbol{\Sigma}_1,
\end{aligned}$$

where the last equality holds because for $h = 0, \ldots, j-1$,

$$\begin{aligned}
&(\boldsymbol{I}_{N_j} \otimes \boldsymbol{1}_{n_j} \otimes \boldsymbol{I}_m)\left(\boldsymbol{I}_{N_h} \otimes \boldsymbol{H}_{h+1} \otimes (\boldsymbol{1}_{n_{h+1}\cdots n_{j-1}}\boldsymbol{1}'_{n_{h+1}\cdots n_{j-1}}) \otimes \boldsymbol{\Sigma}_{h+1}\right)(\boldsymbol{I}_{N_j} \otimes \boldsymbol{1}_{n_j} \otimes \boldsymbol{I}_m)' \\
= \; &(\boldsymbol{I}_{N_j} \otimes \boldsymbol{1}_{n_j} \otimes \boldsymbol{I}_m)\left([\boldsymbol{I}_{N_h} \otimes \boldsymbol{H}_{h+1} \otimes (\boldsymbol{1}_{n_{h+1}\cdots n_{j-1}}\boldsymbol{1}'_{n_{h+1}\cdots n_{j-1}})] \otimes \boldsymbol{I}_1 \otimes \boldsymbol{\Sigma}_{h+1}\right)(\boldsymbol{I}_{N_j} \otimes \boldsymbol{1}_{n_j} \otimes \boldsymbol{I}_m)' \\
= \; &\left(\boldsymbol{I}_{N_j}[\boldsymbol{I}_{N_h} \otimes \boldsymbol{H}_{h+1} \otimes (\boldsymbol{1}_{n_{h+1}\cdots n_{j-1}}\boldsymbol{1}'_{n_{h+1}\cdots n_{j-1}})]\boldsymbol{I}'_{N_j}\right) \otimes \left(\boldsymbol{1}_{n_j}\boldsymbol{I}_1\boldsymbol{1}'_{n_j}\right) \otimes \left(\boldsymbol{I}_m\boldsymbol{\Sigma}_{h+1}\boldsymbol{I}'_m\right) \\
= \; &\boldsymbol{I}_{N_h} \otimes \boldsymbol{H}_{h+1} \otimes (\boldsymbol{1}_{n_{h+1}\cdots n_{j-1}}\boldsymbol{1}'_{n_{h+1}\cdots n_{j-1}}) \otimes (\boldsymbol{1}_{n_j}\boldsymbol{1}'_{n_j}) \otimes \boldsymbol{\Sigma}_{h+1} \\
= \; &\boldsymbol{I}_{N_h} \otimes \boldsymbol{H}_{h+1} \otimes (\boldsymbol{1}_{n_{h+1}\cdots n_j}\boldsymbol{1}'_{n_{h+1}\cdots n_j}) \otimes \boldsymbol{\Sigma}_{h+1}.
\end{aligned}$$

For the covariance:

$$\begin{aligned}
\text{cov}(\boldsymbol{U}_{j\cdot}, \boldsymbol{U}_{j'\cdot}) &= \text{cov}(\boldsymbol{U}_{j\cdot}, (\boldsymbol{I}_{N_{j'-1}} \otimes \boldsymbol{1}_{n_{j'-1}} \otimes \boldsymbol{I}_m)\boldsymbol{U}_{j'-1\cdot}) \\
&= \text{cov}(\boldsymbol{U}_{j\cdot}, \boldsymbol{U}_{j'-1\cdot})(\boldsymbol{I}_{N_{j'-1}} \otimes \boldsymbol{1}'_{n_{j'-1}} \otimes \boldsymbol{I}_m) \\
&\;\;\vdots \\
&= \text{cov}(\boldsymbol{U}_{j\cdot}, \boldsymbol{U}_{j\cdot})(\boldsymbol{I}_{N_j} \otimes \boldsymbol{1}'_{n_j} \otimes \boldsymbol{I}_m)\cdots(\boldsymbol{I}_{N_{j'-1}} \otimes \boldsymbol{1}'_{n_{j'-1}} \otimes \boldsymbol{I}_m) \\
&= \text{var}(\boldsymbol{U}_{j\cdot})(\boldsymbol{I}_{N_j} \otimes \boldsymbol{1}'_{n_j\cdots n_{j'-1}} \otimes \boldsymbol{I}_m).
\end{aligned}$$

where the last equality holds because for $h = j+1, \ldots, j'-1$,

$$\begin{aligned}
(\boldsymbol{I}_{N_j} \otimes \boldsymbol{1}'_{n_j\cdots n_{h-1}} \otimes \boldsymbol{I}_m)(\boldsymbol{I}_{N_h} \otimes \boldsymbol{1}'_{n_h} \otimes \boldsymbol{I}_m) &= ([\boldsymbol{I}_{N_j} \otimes \boldsymbol{1}'_{n_j\cdots n_{h-1}}] \otimes \boldsymbol{I}_1 \otimes \boldsymbol{I}_m)(\boldsymbol{I}_{N_h} \otimes \boldsymbol{1}'_{n_h} \otimes \boldsymbol{I}_m) \\
&= ([\boldsymbol{I}_{N_j} \otimes \boldsymbol{1}'_{n_j\cdots n_{h-1}}]\boldsymbol{I}_{N_h}) \otimes (\boldsymbol{I}_1\boldsymbol{1}'_{n_h}) \otimes (\boldsymbol{I}_m\boldsymbol{I}_m)
\end{aligned}$$

$$= \boldsymbol{I}_{N_j} \otimes \mathbf{1}'_{n_j \cdots n_{h-1}} \otimes \mathbf{1}'_{n_h} \otimes \boldsymbol{I}_m$$
$$= \boldsymbol{I}_{N_j} \otimes \mathbf{1}'_{n_j \cdots n_h} \otimes \boldsymbol{I}_m . \square$$

Although it is easier to summarize the covariances among $\{\boldsymbol{U}_{j,k}\}$ by the variance-covariance matrix $\text{var}(\boldsymbol{U}_{j\cdot})$ in (13), we now focus on the covariance between any pair of scalar nodes on the same resolution to gain more insight into the spatial dependence structure. While $E(\boldsymbol{u}'_{j,k}) = \mathbf{0}_m$, the usual second-order stationarity or intrinsic stationarity does not hold here because of the special feature of the multiresolution tree structure. However, the model is stationary and isotropic in the following sense. Given two locations (j,k) and (j,k') on the j-th resolution, we denote the first common ancestor as $an(j,k,k')$ and denote its corresponding resolution as j_{an}. When two locations are identical $(k = k')$, then $j_{an} = j$ and $\text{cov}(\boldsymbol{u}'_{j,k}, \boldsymbol{u}'_{j,k'}) = \text{var}(\boldsymbol{u}'_{j,k}) = \sum_{j'=1}^{j} \boldsymbol{\Sigma}_{j'}$. When two locations do not have a common ancestor, then they must be in two subtrees with two different root nodes, in which case we define $j_{an} = 0$. When two locations have a first common ancestor on a resolution $j_{an} = 1, \ldots, j-1$, then it can be shown that $\text{cov}(\boldsymbol{u}'_{j,k}, \boldsymbol{u}'_{j,k'}) = \text{cov}(\boldsymbol{u}'_{an(j,k,k')} + \boldsymbol{\omega}_{j_{an}+1,\ell}, \boldsymbol{u}_{an(j,k,k')} + \boldsymbol{\omega}_{j_{an}+1,\ell'})$, where $(j_{an}+1,\ell)$ and $(j_{an}+1,\ell')$ are the children of $an(j,k,k')$ and the (not shared) ancestors of (j,k) and (j,k') respectively, by the fact that these are the two descendants of $an(j,k,k')$ which are correlated via the matrix $\boldsymbol{H}_{j_{an}+1}$ in (6). Hence $\text{cov}(\boldsymbol{u}'_{j,k}, \boldsymbol{u}'_{j,k'}) = \text{var}(\boldsymbol{u}'_{an(j,k,k')}) + \text{cov}(\boldsymbol{\omega}'_{j_{an}+1,\ell}, \boldsymbol{\omega}'_{j_{an}+1,\ell'}) = \sum_{j'=1}^{j_{an}} \boldsymbol{\Sigma}_{j'} + \boldsymbol{H}_{j_{an}+1\ell\ell'} \boldsymbol{\Sigma}_{j_{an}+1}$, where $\boldsymbol{H}_{j_{an}+1\ell\ell'}$ is the (ℓ,ℓ') element of $\boldsymbol{H}_{j_{an}+1}$. That is,

$$\text{cov}(\boldsymbol{u}'_{j,k}, \boldsymbol{u}'_{j,k'}) = \begin{cases} \sum_{j'=1}^{j_{an}} \boldsymbol{\Sigma}_{j'} + \boldsymbol{H}_{j_{an}+1\ell\ell'} \boldsymbol{\Sigma}_{j_{an}+1} & ; \quad j_{an} = 0, 1, \ldots, j-1, \\ \sum_{j'=1}^{j} \boldsymbol{\Sigma}_{j'} & ; \quad j_{an} = j. \end{cases} \tag{58}$$

Proof of mass balance $\boldsymbol{H}_j$. We assume that $\boldsymbol{u}_{1,k}$ are mutually independent for $k = 1, \ldots, N_1$, then $\boldsymbol{H}_1 = \boldsymbol{I}_{N_1}$. From (4) and (14), we have $\mathbf{1}'_{n_j} \boldsymbol{\omega}_{ch(j,k)} = \mathbf{0}'_m$ for $j = 1, \ldots, J-1$. Using (8), we have $(\mathbf{1}'_{n_j} \otimes \boldsymbol{I}_m) \boldsymbol{W}_{j+1,k} = \text{vec}(\mathbf{1}'_{n_j} \boldsymbol{\omega}_{ch(j,k)}) = \mathbf{0}_m$. Hence $\mathbf{0}_{m \times m} = \text{cov}\left((\mathbf{1}'_{n_j} \otimes \boldsymbol{I}_m) \boldsymbol{W}_{j+1,k}, (\mathbf{1}'_{n_j} \otimes \boldsymbol{I}_m) \boldsymbol{W}_{j+1,k} \right) = (\mathbf{1}'_{n_j} \otimes \boldsymbol{I}_m) \text{var}(\boldsymbol{W}_{j+1,k}) (\mathbf{1}'_{n_j} \otimes \boldsymbol{I}_m)' = (\mathbf{1}'_{n_j} \otimes \boldsymbol{I}_m)(\boldsymbol{H}_{j+1} \otimes \boldsymbol{\Sigma}_{j+1})(\mathbf{1}'_{n_j} \otimes \boldsymbol{I}_m)' = (\mathbf{1}'_{n_j} \boldsymbol{H}_{j+1} \mathbf{1}_{n_j}) \otimes (\boldsymbol{I}_m \boldsymbol{\Sigma}_{j+1} \boldsymbol{I}_m) = (\mathbf{1}'_{n_j} \boldsymbol{H}_{j+1} \mathbf{1}_{n_j}) \otimes \boldsymbol{\Sigma}_{j+1}$. Since $\boldsymbol{\Sigma}_{j+1} \neq \mathbf{0}_{m \times m}$, we have $\mathbf{1}'_{n_j} \boldsymbol{H}_{j+1} \mathbf{1}_{n_j} = 0$. If we assume that $\boldsymbol{H}_{j+1}$ is compound symmetric, such that any two children of $\boldsymbol{\omega}_{j,k}$ have the same correlation coefficient ρ_{j+1}. Then we have $0 = \mathbf{1}'_{n_j} \boldsymbol{H}_{j+1} \mathbf{1}_{n_j} = n_j + (n_j^2 - n_j)\rho_{j+1}$, and hence $\rho_{j+1} = \frac{-1}{n_j - 1}$.$\square$

Appendix B1: 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 $\boldsymbol{Z}_0 = \boldsymbol{\mu} + \boldsymbol{\beta}\boldsymbol{Z}$ with $\boldsymbol{\mu} \in I\!R^m$, $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m]'$, and $\boldsymbol{\beta}_i \in I\!R^n;\ i = 1, \ldots, m$. Let

$$\begin{aligned}
f(\boldsymbol{\mu},\boldsymbol{\beta}) \equiv \|\boldsymbol{Y} - \boldsymbol{Z}_0\|^2 &= E[(\boldsymbol{Y} - \boldsymbol{\mu} - \boldsymbol{\beta}\boldsymbol{Z})'(\boldsymbol{Y} - \boldsymbol{\mu} - \boldsymbol{\beta}\boldsymbol{Z})] \\
&= E\{\mathrm{tr}[(\boldsymbol{Y} - \boldsymbol{\mu} - \boldsymbol{\beta}\boldsymbol{Z})'(\boldsymbol{Y} - \boldsymbol{\mu} - \boldsymbol{\beta}\boldsymbol{Z})]\} \\
&= \mathrm{tr}\{E[(\boldsymbol{Y} - \boldsymbol{\mu} - \boldsymbol{\beta}\boldsymbol{Z})(\boldsymbol{Y} - \boldsymbol{\mu} - \boldsymbol{\beta}\boldsymbol{Z})]'\} \\
&= \mathrm{tr}\{E(\boldsymbol{Y} - \boldsymbol{\mu} - \boldsymbol{\beta}\boldsymbol{Z})E(\boldsymbol{Y} - \boldsymbol{\mu} - \boldsymbol{\beta}\boldsymbol{Z})' + var(\boldsymbol{Y} - \boldsymbol{\mu} - \boldsymbol{\beta}\boldsymbol{Z})\} \\
&= E(\boldsymbol{Y} - \boldsymbol{\mu} - \boldsymbol{\beta}\boldsymbol{Z})'E(\boldsymbol{Y} - \boldsymbol{\mu} - \boldsymbol{\beta}\boldsymbol{Z}) + \mathrm{tr}\{\mathrm{var}(\boldsymbol{Y} - \boldsymbol{\mu} - \boldsymbol{\beta}\boldsymbol{Z})\} \\
&= (\boldsymbol{\mu}_Y - \boldsymbol{\mu} - \boldsymbol{\beta}\boldsymbol{\mu}_Z)'(\boldsymbol{\mu}_Y - \boldsymbol{\mu} - \boldsymbol{\beta}\boldsymbol{\mu}_Z) + \mathrm{tr}\{\boldsymbol{\Sigma}_{YY} - 2\boldsymbol{\beta}\boldsymbol{\Sigma}_{ZY} + \boldsymbol{\beta}\boldsymbol{\Sigma}_{ZZ}\boldsymbol{\beta}'\} \\
&= \boldsymbol{\mu}_Y'\boldsymbol{\mu}_Y - 2\boldsymbol{\mu}_Y'\boldsymbol{\beta}\boldsymbol{\mu}_Z - 2\boldsymbol{\mu}'\boldsymbol{\mu}_Y + \boldsymbol{\mu}'\boldsymbol{\mu} + 2\boldsymbol{\mu}'\boldsymbol{\beta}\boldsymbol{\mu}_Z + \boldsymbol{\mu}_Z'\boldsymbol{\beta}'\boldsymbol{\beta}\boldsymbol{\mu}_Z \\
&\quad + \mathrm{tr}\{\boldsymbol{\Sigma}_{YY} - 2\boldsymbol{\beta}\boldsymbol{\Sigma}_{ZY} + \boldsymbol{\beta}\boldsymbol{\Sigma}_{ZZ}\boldsymbol{\beta}'\} \\
&= \boldsymbol{\mu}_Y'\boldsymbol{\mu}_Y - 2\boldsymbol{\mu}'\boldsymbol{\mu}_Y + \boldsymbol{\mu}'\boldsymbol{\mu} + 2\boldsymbol{\mu}'\boldsymbol{\beta}\boldsymbol{\mu}_Z \\
&\quad + \mathrm{tr}\{-2\boldsymbol{\beta}\boldsymbol{\mu}_Z\boldsymbol{\mu}_Y' + \boldsymbol{\beta}\boldsymbol{\mu}_Z\boldsymbol{\mu}_Z'\boldsymbol{\beta}' + \boldsymbol{\Sigma}_{YY} - 2\boldsymbol{\beta}\boldsymbol{\Sigma}_{ZY} + \boldsymbol{\beta}\boldsymbol{\Sigma}_{ZZ}\boldsymbol{\beta}'\}.
\end{aligned}$$

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 (Chapter 15, Harville, 1997).

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

Proof of Theorem 1. By Lemma 1, we have:

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

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

$$\boldsymbol{\mu} = \boldsymbol{\mu}_Y - \boldsymbol{\beta}\boldsymbol{\mu}_Z, \qquad \boldsymbol{\Sigma}_{ZZ}\boldsymbol{\beta}' = \boldsymbol{\Sigma}_{ZY}, \tag{59}$$

$$\boldsymbol{\mu} = \boldsymbol{\mu}_Y - \boldsymbol{\beta}\boldsymbol{\mu}_Z, \qquad \boldsymbol{\Sigma}_{ZZ}\boldsymbol{\beta}_i = \boldsymbol{\Sigma}_{Zy_i}; \quad i = 1, \ldots, m. \tag{60}$$

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 (59). 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 $\text{cov}(\boldsymbol{Z},\ \boldsymbol{Y} - \boldsymbol{m}_{Y|Z}) = \text{cov}(\boldsymbol{Z},\ \boldsymbol{Y} - \boldsymbol{\mu}_Y + \hat{\boldsymbol{\beta}}\boldsymbol{\mu}_Z - \hat{\boldsymbol{\beta}}\boldsymbol{Z}) = \boldsymbol{\Sigma}_{ZY} - \boldsymbol{\Sigma}_{ZZ}\hat{\boldsymbol{\beta}}' = \mathbf{0}_{n\times m}$ where the last equality holds because of (59).

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 $\text{cov}(\boldsymbol{m}_{Y|Z} - \tilde{\boldsymbol{m}}_{Y|Z}, \boldsymbol{e} - \tilde{\boldsymbol{e}}) = \mathbf{0}_{m\times m}$ because $\boldsymbol{m}_{Y|Z} - \tilde{\boldsymbol{m}}_{Y|Z} \in \overline{sp}\{\boldsymbol{Z}\}^m$ and $\text{cov}(\boldsymbol{Z}, \boldsymbol{e} - \tilde{\boldsymbol{e}}) = \mathbf{0}_{n\times m}$. Then $\mathbf{0}_{m\times m} = \text{var}(\boldsymbol{Y} - \boldsymbol{Y}) = \text{var}[(\boldsymbol{m}_{Y|Z} - \tilde{\boldsymbol{m}}_{Y|Z}) + (\boldsymbol{e} - \tilde{\boldsymbol{e}})] = \text{var}(\boldsymbol{m}_{Y|Z} - \tilde{\boldsymbol{m}}_{Y|Z}) + \text{var}(\boldsymbol{e} - \tilde{\boldsymbol{e}})$. Comparing the diagonal elements of both sides, since the variance of any random variable in L^2 is non-negative, we obtain $\text{var}(m_{y_i|Z} - \tilde{m}_{y_i|Z}) = 0$ 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. □

We obtain the optimal linear predictor using any solution of the normal equations (59) starting with the following lemma.

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

Proof. Consistency of linear equations $\boldsymbol{\Sigma}_{ZZ}\boldsymbol{\beta}_i = \boldsymbol{\Sigma}_{Zy_i}$ means that if any linear relationship exists among the rows of $\boldsymbol{\Sigma}_{ZZ}$, then the linear relationship also exists among the corresponding elements of $\boldsymbol{\Sigma}_{Zy_i}$ (Chapter 1.2, pg 7, Searle, 1997). To show that the normal equations are consistent, it suffices to show that for any $\boldsymbol{\alpha} \in I\!R^n$, if $\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{ZZ} = \mathbf{0}_n$, then $\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{Zy_i} = 0$. Suppose for $\boldsymbol{\alpha} \in I\!R^n$, $\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{ZZ} = \mathbf{0}_n$, then $\text{var}(\boldsymbol{\alpha}'\boldsymbol{Z}) = \boldsymbol{\alpha}'\boldsymbol{\Sigma}_{ZZ}\boldsymbol{\alpha} = 0$, which implies that $\boldsymbol{\alpha}'\boldsymbol{Z} = \boldsymbol{\alpha}'\boldsymbol{\mu}_Z$. Hence $\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{Zy_i} = \boldsymbol{\alpha}'\text{cov}(\boldsymbol{Z}, y_i) = \text{cov}(\boldsymbol{\alpha}'\boldsymbol{Z}, y_i) = \text{cov}(\boldsymbol{\alpha}'\boldsymbol{\mu}_Z, y_i) = 0$. □

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

$$\boldsymbol{\Sigma}\boldsymbol{\Sigma}^+\boldsymbol{\Sigma} = \boldsymbol{\Sigma}, \quad \boldsymbol{\Sigma}^+\boldsymbol{\Sigma}\boldsymbol{\Sigma}^+ = \boldsymbol{\Sigma}^+, \quad (\boldsymbol{\Sigma}^+\boldsymbol{\Sigma})' = \boldsymbol{\Sigma}^+\boldsymbol{\Sigma} \quad \text{and } (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^+)' = \boldsymbol{\Sigma}\boldsymbol{\Sigma}^+. \tag{61}$$

Furthermore, by transposing both sides of the four conditions in (61), we obtain

$$(\boldsymbol{\Sigma}')^+ = (\boldsymbol{\Sigma}^+)'. \tag{62}$$

Proof of Theorem 2. Since the normal equations (60) are consistent, from Theorem 1 of Chapter 1.6 of Searle (1997), one of the solutions of the normal equations and its equivalence are

$$\begin{aligned}\hat{\boldsymbol{\mu}} &= \boldsymbol{\mu}_Y - \hat{\boldsymbol{\beta}}'\boldsymbol{\mu}_Z, \qquad \hat{\boldsymbol{\beta}}_i = \boldsymbol{\Sigma}_{ZZ}^{+}\boldsymbol{\Sigma}_{Zy_i},\\ \hat{\boldsymbol{\mu}} &= \boldsymbol{\mu}_Y - \hat{\boldsymbol{\beta}}'\boldsymbol{\mu}_Z, \qquad \hat{\boldsymbol{\beta}} = (\boldsymbol{\Sigma}_{ZZ}^{+}\boldsymbol{\Sigma}_{ZY})'. \end{aligned} \tag{63}$$

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

$$\begin{aligned}\boldsymbol{C}_{Y|Z} &= E[(\boldsymbol{Y} - \boldsymbol{m}_{Y|Z})(\boldsymbol{Y} - \boldsymbol{m}_{Y|Z})']\\ &= E(\boldsymbol{Y} - \boldsymbol{m}_{Y|Z})E(\boldsymbol{Y} - \boldsymbol{m}_{Y|Z})' + \mathrm{var}(\boldsymbol{Y} - \boldsymbol{m}_{Y|Z})\\ &= \mathrm{var}(\boldsymbol{Y} - \hat{\boldsymbol{\beta}}\boldsymbol{Z})\\ &= \Sigma_{YY} - \boldsymbol{\Sigma}_{YZ}\hat{\boldsymbol{\beta}}' - \hat{\boldsymbol{\beta}}\boldsymbol{\Sigma}_{ZY} + \hat{\boldsymbol{\beta}}\boldsymbol{\Sigma}_{ZZ}\hat{\boldsymbol{\beta}}'\\ &= \Sigma_{YY} - \boldsymbol{\Sigma}_{YZ}\hat{\boldsymbol{\beta}}' - \hat{\boldsymbol{\beta}}\boldsymbol{\Sigma}_{ZY} + \hat{\boldsymbol{\beta}}\boldsymbol{\Sigma}_{ZY}\\ &= \Sigma_{YY} - \boldsymbol{\Sigma}_{YZ}\hat{\boldsymbol{\beta}}',\end{aligned}$$

where the fifth equality holds because of (59).

Since $\boldsymbol{\Sigma}_{ZZ}$ is symmetric, using (62) 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 (Proposition C.5, Lauritzen, 1996). □

Appendix B2: Generalized Change-of-Resolution Kalman Filter

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

$$\begin{aligned}\gamma_{j,k} &\equiv \mathcal{I}\{\boldsymbol{Z}_{j,k} \text{ is observed}\} = \begin{cases} 1; & \text{if } \boldsymbol{Z}_{j,k} \text{ is observed,}\\ 0; & \text{otherwise,}\end{cases}\\ \boldsymbol{Z} &\equiv \{\boldsymbol{Z}_{j,k} : \gamma_{j,k} = 1\},\\ \boldsymbol{Z}_{de\{j,k\}} &\equiv \left\{\boldsymbol{Z}_{j',k'} : \gamma_{j',k'} = 1, \{j', k'\} \prec \{j, k\}\right\},\\ \boldsymbol{Z}^{*}_{de\{j,k\}} &\equiv \left\{\boldsymbol{Z}_{j',k'} : \gamma_{j',k'} = 1, \{j', k'\} \prec \{j, k\}, \{j', k'\} \neq \{j, k\}\right\},\\ \boldsymbol{Z}^{c}_{de\{j,k\}} &\equiv \boldsymbol{Z} \setminus \boldsymbol{Z}_{de\{j,k\}},\end{aligned}$$

$$
\begin{aligned}
\boldsymbol{U}_{j,k}|\boldsymbol{Z}_{de\{j',k'\}} &\sim [\hat{\boldsymbol{U}}_{j,k|j',k'}, \hat{\boldsymbol{V}}_{j,k|j',k'}],\\
\boldsymbol{U}_{j,k}|\boldsymbol{Z}^*_{de\{j',k'\}} &\sim [\hat{\boldsymbol{U}}^*_{j,k|j',k'}, \hat{\boldsymbol{V}}^*_{j,k|j',k'}],\\
\boldsymbol{U}_{j,k}|\boldsymbol{Z} &\sim [\hat{\boldsymbol{U}}_{j,k}, \hat{\boldsymbol{V}}_{j,k}],\\
\boldsymbol{V}_{j,k} &\equiv \mathrm{var}(\boldsymbol{U}_{j,k}),\\
\boldsymbol{V}_{j,k,j',k'} &\equiv \mathrm{cov}(\boldsymbol{U}_{j,k}, \boldsymbol{U}_{j',k'}),\\
\boldsymbol{B}_{j,k} &\equiv \boldsymbol{V}_{pa\{j,k\}}\boldsymbol{A}'_{j,k}\boldsymbol{V}^+_{j,k},\\
\boldsymbol{R}_{j,k} &\equiv \boldsymbol{V}_{pa\{j,k\}} - \boldsymbol{V}_{pa\{j,k\}}\boldsymbol{A}'_{j,k}\boldsymbol{V}^+_{j,k}\boldsymbol{A}_{j,k}\boldsymbol{V}_{pa\{j,k\}},\\
\boldsymbol{J}_{j,k} &\equiv \hat{\boldsymbol{V}}_{j,k|j,k}\boldsymbol{B}'_{j,k}\hat{\boldsymbol{V}}^+_{pa\{j,k\}|j,k}.
\end{aligned}
$$

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

Proof. (i) Theorem 14.3.7 of Harville (1997).

(ii) From (i), there exist matrices $\boldsymbol{L}$ and $\boldsymbol{K}$ such that $\boldsymbol{A} = \boldsymbol{LL}'$ and $\boldsymbol{B} = \boldsymbol{KK}'$. $\boldsymbol{K}$ is invertible because $\text{Rank}(\boldsymbol{K}) = \text{Rank}(\boldsymbol{B})$. Then $\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{K}[(\boldsymbol{K}^{-1}\boldsymbol{L})(\boldsymbol{K}^{-1}\boldsymbol{L})' + \boldsymbol{I}_n]\boldsymbol{K}'$. From Lemma 8 of section 1.6 of Searle (1997), $(\boldsymbol{K}^{-1}\boldsymbol{L})(\boldsymbol{K}^{-1}\boldsymbol{L})' + \boldsymbol{I}_n$ is invertible, hence $\boldsymbol{A} + \boldsymbol{B}$ is invertible.

(iii) $(\boldsymbol{I}_m + \boldsymbol{AB})(\boldsymbol{I}_m - \boldsymbol{A}(\boldsymbol{I}_n + \boldsymbol{BA})^{-1}\boldsymbol{B}) = \boldsymbol{I}_m + \boldsymbol{AB} - \boldsymbol{A}(\boldsymbol{I}_n + \boldsymbol{BA})^{-1}\boldsymbol{B} - \boldsymbol{ABA}(\boldsymbol{I}_n + \boldsymbol{BA})^{-1}\boldsymbol{B} = \boldsymbol{I}_m + \boldsymbol{AB} - \boldsymbol{A}(\boldsymbol{I}_n + \boldsymbol{BA})(\boldsymbol{I}_n + \boldsymbol{BA})^{-1}\boldsymbol{B} = \boldsymbol{I}_m$.

(iv) From formula (1.2) of section 20.1 of Harville (1997), $(\boldsymbol{AB})^+ = \boldsymbol{B}'(\boldsymbol{BB}')^{-1}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$. Formulas (2.1) and (2.2) of section 20.2 of Harville (1997) show that $\boldsymbol{B}^+ = \boldsymbol{B}'(\boldsymbol{BB}')^{-1}$ and $\boldsymbol{A}^+ = (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$. Hence $(\boldsymbol{AB})^+ = \boldsymbol{B}^+\boldsymbol{A}^+$.

(v) From Theorem 20.5.3 of Harville (1997), $(\boldsymbol{A}')^+ = (\boldsymbol{A}^+)'$. From Theorem 20.5.1 of Harville (1997), $\boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{P}_{A'}$ and $\boldsymbol{AA}^+ = \boldsymbol{P}_A$. Theorem 12.3.4 of Harville (1997) shows $\boldsymbol{P}_{A'}\boldsymbol{A}' = \boldsymbol{A}'$ and $\boldsymbol{P}_A\boldsymbol{A} = \boldsymbol{A}$. Hence $\boldsymbol{A}^+\boldsymbol{AA}' = \boldsymbol{P}_{A'}\boldsymbol{A}' = \boldsymbol{A}'$ and $\boldsymbol{A}'\boldsymbol{AA}^+ = \boldsymbol{A}'\boldsymbol{P}_A = (\boldsymbol{P}_A\boldsymbol{A})' = \boldsymbol{A}'$. If $\text{Col}(\boldsymbol{A}) = \text{Col}(\boldsymbol{B})$, Theorem 12.3.1 of Harville (1997) shows that $\boldsymbol{P}_A = \boldsymbol{P}_B$.

(vi) Suppose

$$\begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{Z} \end{pmatrix} \sim \left[\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{XB} \end{pmatrix}, \begin{pmatrix} \boldsymbol{V} & \boldsymbol{VM}' \\ \boldsymbol{M}'\boldsymbol{V} & \boldsymbol{MVM}' + \boldsymbol{P} \end{pmatrix} \right].$$

From Theorem 2,

$$\begin{aligned} \boldsymbol{m}_{U|Z} &= \boldsymbol{VM}'(\boldsymbol{MVM}' + \boldsymbol{P})^{-1}(\boldsymbol{Z} - \boldsymbol{XB}) \\ \boldsymbol{C}_{U|Z} &= \boldsymbol{V} - \boldsymbol{VM}'(\boldsymbol{MVM}' + \boldsymbol{P})^{-1}\boldsymbol{MV}. \end{aligned}$$

where $\boldsymbol{MVM}' + \boldsymbol{P}$ is invertible because of part (ii).

$$\begin{aligned}
\boldsymbol{C}_{U|Z} &= \boldsymbol{V} - \boldsymbol{VM}'(\boldsymbol{MVM}' + \boldsymbol{P})^{-1}\boldsymbol{MV} \\
&= \boldsymbol{L}[\boldsymbol{I} - \boldsymbol{L}'\boldsymbol{M}'(\boldsymbol{MVM}' + \boldsymbol{P})^{-1}\boldsymbol{ML}]\boldsymbol{L}' \\
&= \boldsymbol{L}[\boldsymbol{I} - \boldsymbol{L}'\boldsymbol{M}'\boldsymbol{P}^{-1}(\boldsymbol{MVM}'\boldsymbol{P}^{-1} + \boldsymbol{I})^{-1}\boldsymbol{ML}]\boldsymbol{L}' \\
&= \boldsymbol{L}[\boldsymbol{I} + \boldsymbol{L}'\boldsymbol{M}'\boldsymbol{P}^{-1}\boldsymbol{ML}]^{-1}\boldsymbol{L}', \qquad (64)
\end{aligned}$$

where the last equality holds because of part (iii).

From part (iv), we have $\boldsymbol{C}^{+}_{U|Z} = (\boldsymbol{L}')^{+}[\boldsymbol{I} + \boldsymbol{L}'\boldsymbol{M}'\boldsymbol{P}^{-1}\boldsymbol{ML}]\boldsymbol{L}^{+}$. Then

$$\begin{aligned}
\boldsymbol{C}^{+}_{U|Z}\boldsymbol{m}_{U|Z} &= (\boldsymbol{L}')^{+}[\boldsymbol{I} + \boldsymbol{L}'\boldsymbol{M}'\boldsymbol{P}^{-1}\boldsymbol{ML}]\boldsymbol{L}^{+}[\boldsymbol{VM}'(\boldsymbol{MVM}' + \boldsymbol{P})^{-1}(\boldsymbol{Z} - \boldsymbol{XB})] \\
&= [(\boldsymbol{L}')^{+}\boldsymbol{L}^{+}\boldsymbol{VM}' + (\boldsymbol{L}')^{+}\boldsymbol{L}'\boldsymbol{M}'\boldsymbol{P}^{-1}\boldsymbol{MLL}^{+}\boldsymbol{VM}'][(\boldsymbol{MVM}' + \boldsymbol{P})^{-1}(\boldsymbol{Z} - \boldsymbol{XB})] \\
&= [(\boldsymbol{L}')^{+}\boldsymbol{L}^{+}\boldsymbol{LL}'\boldsymbol{M}' + (\boldsymbol{L}')^{+}\boldsymbol{L}'\boldsymbol{M}'\boldsymbol{P}^{-1}\boldsymbol{MLL}^{+}\boldsymbol{LL}'\boldsymbol{M}'][(\boldsymbol{MVM}' + \boldsymbol{P})^{-1}(\boldsymbol{Z} - \boldsymbol{XB})] \\
&= [(\boldsymbol{L}')^{+}\boldsymbol{L}'\boldsymbol{M}' + (\boldsymbol{L}')^{+}\boldsymbol{L}'\boldsymbol{M}'\boldsymbol{P}^{-1}\boldsymbol{MLL}'\boldsymbol{M}'][(\boldsymbol{MVM}' + \boldsymbol{P})^{-1}(\boldsymbol{Z} - \boldsymbol{XB})] \\
&= [(\boldsymbol{L}')^{+}\boldsymbol{L}'\boldsymbol{M}'\boldsymbol{P}^{-1}(\boldsymbol{P} + \boldsymbol{MVM}')][(\boldsymbol{MVM}' + \boldsymbol{P})^{-1}(\boldsymbol{Z} - \boldsymbol{XB})] \\
&= (\boldsymbol{L}')^{+}\boldsymbol{L}'\boldsymbol{M}'\boldsymbol{P}^{-1}(\boldsymbol{Z} - \boldsymbol{XB}),
\end{aligned}$$

where the fourth equality holds because of part (v).

From (64), $\mathrm{Col}(\boldsymbol{C}_{U|Z}) \subseteq \mathrm{Col}(\boldsymbol{L})$. Since $\mathrm{Rank}(\boldsymbol{C}_{U|Z}) = \mathrm{Rank}(\boldsymbol{L})$, we have $\mathrm{Col}(\boldsymbol{C}_{U|Z}) = \mathrm{Col}(\boldsymbol{L})$. Then

$$\begin{aligned}
\boldsymbol{C}_{U|Z}\boldsymbol{C}^{+}_{U|Z}\boldsymbol{m}_{U|Z} &= \boldsymbol{P}_{C_{U|Z}}\boldsymbol{m}_{U|Z} \\
&= \boldsymbol{P}_{L}\boldsymbol{m}_{U|Z} \\
&= \boldsymbol{LL}^{+}\boldsymbol{VM}'(\boldsymbol{MVM}' + \boldsymbol{P})^{-1}(\boldsymbol{Z} - \boldsymbol{XB}) \\
&= \boldsymbol{LL}^{+}\boldsymbol{LL}'\boldsymbol{M}'(\boldsymbol{MVM}' + \boldsymbol{P})^{-1}(\boldsymbol{Z} - \boldsymbol{XB}) \\
&= \boldsymbol{LL}'\boldsymbol{M}'(\boldsymbol{MVM}' + \boldsymbol{P})^{-1}(\boldsymbol{Z} - \boldsymbol{XB}) \\
&= \boldsymbol{m}_{U|Z},
\end{aligned}$$

where the first, second and third equalities hold because of part (v).

Hence

$$\begin{aligned}\boldsymbol{m}_{U|Z} &= \boldsymbol{C}_{U|Z}(\boldsymbol{L}')^{+}\boldsymbol{L}'\boldsymbol{M}'\boldsymbol{P}^{-1}(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{B}) \\ &= \boldsymbol{C}_{U|Z}\boldsymbol{P}_{L}\boldsymbol{M}'\boldsymbol{P}^{-1}(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{B}) \\ &= \boldsymbol{C}_{U|Z}\boldsymbol{P}_{C_{U|Z}}\boldsymbol{M}'\boldsymbol{P}^{-1}(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{B}) \\ &= \boldsymbol{C}_{U|Z}\boldsymbol{M}'\boldsymbol{P}^{-1}(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{B}),\end{aligned}$$

where the last equality holds because of part (v). □

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

Proof of (16)–(17). For a leaf node $\{J,k\}$, $k=1,\dots,N_{J-1}$, 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 $\boldsymbol{U}_{J,k}|\boldsymbol{Z}_{J,k} \sim \Big[\boldsymbol{V}_{J,k}(\boldsymbol{V}_{J,k}+\boldsymbol{I}_{n_{J-1}}\otimes\boldsymbol{\Phi}_J)^{-1}(\boldsymbol{Z}_{J,k} - \boldsymbol{X}_{J,k}\boldsymbol{B}), \boldsymbol{V}_{J,k}-\boldsymbol{V}_{J,k}(\boldsymbol{V}_{J,k}+\boldsymbol{I}_{n_{J-1}}\otimes\boldsymbol{\Phi}_J)^{-1}\boldsymbol{V}_{J,k}\Big]$, 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 $\{J,k\}$,

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

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 $\Big\{\boldsymbol{U}_{j',k'}:\{j',k'\}\prec j,k\Big\}\cup\boldsymbol{Z}_{de\{j,k\}}$ 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

$$\begin{aligned}
\hat{\boldsymbol{U}}_{j,k|ch\{j,k,i\}} &= \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}_{de\{ch\{j,k,i\}\}}) \\
&= \boldsymbol{m}(\boldsymbol{B}_{ch\{j,k,i\}}\boldsymbol{U}_{ch\{j,k,i\}} + \boldsymbol{\xi}_{ch\{j,k,i\}}|\boldsymbol{Z}_{de\{ch\{j,k,i\}\}}) \\
&= \boldsymbol{B}_{ch\{j,k,i\}}\boldsymbol{m}(\boldsymbol{U}_{ch\{j,k,i\}}|\boldsymbol{Z}_{de\{ch\{j,k,i\}\}}) + \boldsymbol{m}(\boldsymbol{\xi}_{ch\{j,k,i\}}|\boldsymbol{Z}_{de\{ch\{j,k,i\}\}}) \\
&= \boldsymbol{B}_{ch\{j,k,i\}}\boldsymbol{m}(\boldsymbol{U}_{ch\{j,k,i\}}|\boldsymbol{Z}_{de\{ch\{j,k,i\}\}}) \\
&= \boldsymbol{B}_{ch\{j,k,i\}}\hat{\boldsymbol{U}}_{ch\{j,k,i\}|ch\{j,k,i\}},
\end{aligned}$$

where in the third equality, $\boldsymbol{m}(\boldsymbol{\xi}_{ch\{j,k,i\}}|\boldsymbol{Z}_{de\{ch\{j,k,i\}\}}) = 0$ because $\boldsymbol{\xi}_{ch\{j,k,i\}}$ is uncorrelated with $\boldsymbol{Z}_{de\{ch\{j,k,i\}\}}$.

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

where in the fourth equality, $\boldsymbol{C}(\boldsymbol{\xi}_{ch\{j,k,i\}}|\boldsymbol{Z}_{de\{ch\{j,k,i\}\}}) = \text{var}(\boldsymbol{\xi}_{ch\{j,k,i\}}) = \boldsymbol{R}_{j,k}$ because $\boldsymbol{\xi}_{ch\{j,k,i\}}$ is uncorrelated with $\boldsymbol{Z}_{de\{ch\{j,k,i\}\}}$. $\square$

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 $\boldsymbol{P}_{de\{ch\{j,k,i\}\}} \equiv \text{var}(\boldsymbol{e}_{de\{ch\{j,k,i\}\}})$ and $\boldsymbol{P}^*_{de\{j,k\}} \equiv \text{var}(\boldsymbol{e}^*_{de\{j,k\}})$, then

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

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),

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

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),

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

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

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

where the second, third, and fifth equality hold because of Lemma 3 (iii), (iv) and (v) respectively. Since $\hat{\boldsymbol{V}}^*_{j,k|j,k} = \boldsymbol{L}_{j,k}(\boldsymbol{I}+\boldsymbol{L}'_{j,k}\boldsymbol{M}^{*'}_{j,k}\boldsymbol{P}^{*-1}_{j,k}\boldsymbol{L}_{j,k})^{-1}\boldsymbol{L}'_{j,k}$, 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),

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

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

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

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

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

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

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

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. We have the uptree formula

$$\boldsymbol{U}_{pa\{j,k\}}=\boldsymbol{B}_{j,k}\boldsymbol{U}_{j,k}+\boldsymbol{\xi}_{j,k} \tag{65}$$

where $\xi_{j,k}=\boldsymbol{U}_{pa\{j,k\}}-\boldsymbol{B}_{j,k}\boldsymbol{U}_{j,k}$ and $\mathrm{var}(\boldsymbol{\xi}_{j,k})=\boldsymbol{R}_{j,k}$. Moreover, we have the downtree formula

$$\boldsymbol{U}_{j,k}=\boldsymbol{A}_{j,k}\boldsymbol{U}_{pa\{j,k\}}+\boldsymbol{W}_{j,k}. \tag{66}$$

Define $T^c_{j,k}\equiv\Big\{\{j',k'\}:\gamma_{j',k'}=1,\{j',k'\}\text{ is not a descendant of }\{j,k\}\Big\}$ be the collection of all the nodes which are not descendants of node $\{j,k\}$. If node $\{j',k'\}\in T^c_{j,k}$, then there is a path from $pa\{j,k\}$ to $\{j',k'\}$ that starts from $pa\{j,k\}$, moves up to the common ancestor, and moves down to $\{j',k'\}$. Hence $\boldsymbol{U}_{j',k'}$ can be written as a linear transformation of $\boldsymbol{U}_{pa\{j,k\}}$ plus a function depending on $\{\boldsymbol{\xi}_{j'',k''}:\{j'',k''\}\in T^c_{j,k}\}$ and $\{\boldsymbol{W}_{j'',k''}:\{j'',k''\}\in T^c_{j,k}\}$. Now we obtain $\boldsymbol{Z}^c_{de\{j,k\}}=\boldsymbol{F}_{j,k}\boldsymbol{U}_{pa\{j,k\}}+\boldsymbol{\kappa}_{j,k}$ where $\boldsymbol{F}_{j,k}$ is a deterministic matrix which depends on $\{\boldsymbol{A}_{j'',k''},\boldsymbol{B}_{j'',k''}:\{j'',k''\}\in T^c_{j,k}\}$ and $\boldsymbol{\kappa}_{j,k}$ is a random vector which depends on $\{\boldsymbol{\xi}_{j'',k''}, \boldsymbol{W}_{j'',k''}, \boldsymbol{e}_{j'',k''} : \{j'',k''\} \in T^c_{j,k}\}$. It is easy to obtain that $\boldsymbol{\kappa}_{j,k}$ is uncorrelated with $\boldsymbol{Z}_{de\{j,k\}}$.

$$\begin{aligned}
\boldsymbol{\nu}_{j,k|j,k} &\equiv \boldsymbol{Z}^c_{de\{j,k\}} - \boldsymbol{m}(\boldsymbol{Z}^c_{de\{j,k\}}|\boldsymbol{Z}_{de\{j,k\}}) \\
&= \boldsymbol{F}_{j,k}\boldsymbol{U}_{pa\{j,k\}} + \boldsymbol{\kappa}_{j,k} - \boldsymbol{m}(\boldsymbol{F}_{j,k}U_{pa\{j,k\}} + \boldsymbol{\kappa}_{j,k}|\boldsymbol{Z}_{de\{j,k\}}) \\
&= \boldsymbol{F}_{j,k}[\boldsymbol{U}_{pa\{j,k\}} - \boldsymbol{m}(\boldsymbol{U}_{pa\{j,k\}}|\boldsymbol{Z}_{de\{j,k\}})] + \boldsymbol{\kappa}_{j,k} - \boldsymbol{m}(\boldsymbol{\kappa}_{j,k}|\boldsymbol{Z}_{de\{j,k\}}) \\
&= \boldsymbol{F}_{j,k}\tilde{\boldsymbol{U}}_{pa\{j,k\}|j,k} + \boldsymbol{\kappa}_{j,k}
\end{aligned}$$

where $\tilde{\boldsymbol{U}}_{pa\{j,k\}|j,k} \equiv \boldsymbol{U}_{pa\{j,k\}} - \boldsymbol{m}(\boldsymbol{U}_{pa\{j,k\}}|\boldsymbol{Z}_{de\{j,k\}})$ and the last equality holds because $\boldsymbol{m}(\boldsymbol{\kappa}_{j,k}|\boldsymbol{Z}_{de\{j,k\}}) = \boldsymbol{0}$. Then

$$\begin{aligned}
\boldsymbol{m}(\tilde{\boldsymbol{U}}_{j,k|j,k}|\tilde{\boldsymbol{U}}_{pa\{j,k\}}, \boldsymbol{\nu}_{j,k|j,k}) &= \boldsymbol{m}(\tilde{\boldsymbol{U}}_{j,k|j,k}|\tilde{\boldsymbol{U}}_{pa\{j,k\}|j,k}, \boldsymbol{\nu}_{j,k|j,k} - \boldsymbol{F}_{j,k}\tilde{\boldsymbol{U}}_{pa\{j,k\}|j,k}) \\
&= \boldsymbol{m}(\tilde{\boldsymbol{U}}_{j,k|j,k}|\tilde{\boldsymbol{U}}_{pa\{j,k\}|j,k}, \boldsymbol{\kappa}_{j,k}) \\
&= \boldsymbol{m}(\tilde{\boldsymbol{U}}_{j,k|j,k}|\tilde{\boldsymbol{U}}_{pa\{j,k\}|j,k}) \qquad (67)
\end{aligned}$$

where $\tilde{\boldsymbol{U}}_{j,k|j,k} \equiv \boldsymbol{U}_{j,k} - \hat{\boldsymbol{U}}_{j,k|j,k}$ and the last equality holds because $\tilde{\boldsymbol{U}}_{j,k|j,k}$ and $\boldsymbol{\kappa}_{j,k}$ are uncorrelated.

Using the uptree formula (65) and $\hat{\boldsymbol{U}}_{pa\{j,k\}|j,k} = \boldsymbol{B}_{j,k}\hat{\boldsymbol{U}}_{j,k|j,k}$, we obtain $\tilde{\boldsymbol{U}}_{pa\{j,k\}|j,k} \equiv \boldsymbol{U}_{pa\{j,k\}|j,k} - \hat{\boldsymbol{U}}_{pa\{j,k\}|j,k} = \boldsymbol{B}_{j,k}(\boldsymbol{U}_{j,k} - \hat{\boldsymbol{U}}_{j,k|j,k}) + \boldsymbol{\xi}_{j,k} = \boldsymbol{B}_{j,k}\tilde{\boldsymbol{U}}_{j,k|j,k} + \xi_{j,k}$. Moreover, $\text{var}(\tilde{\boldsymbol{U}}_{pa\{j,k\}|j,k}) = \text{var}(\boldsymbol{U}_{pa\{j,k\}} - \hat{\boldsymbol{U}}_{pa\{j,k\}|j,k}) = \boldsymbol{C}(\boldsymbol{U}_{pa\{j,k\}}|\boldsymbol{Z}_{de\{j,k\}}) = \hat{\boldsymbol{V}}_{pa\{j,k\}|j,k}$. Similarly, we obtain $\text{var}(\tilde{\boldsymbol{U}}_{j,k|j,k}) = \hat{\boldsymbol{V}}_{j,k|j,k}$. Then,

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

and

$$\begin{aligned}
\boldsymbol{m}(\tilde{\boldsymbol{U}}_{j,k|j,k}|\tilde{\boldsymbol{U}}_{pa\{j,k\}|j,k}) &= \hat{\boldsymbol{V}}_{j,k|j,k}\boldsymbol{B}'_{j,k}\hat{\boldsymbol{V}}^+_{pa\{j,k\}|j,k}\tilde{\boldsymbol{U}}_{pa\{j,k\}|j,k} \\
&= \boldsymbol{J}_{j,k}\tilde{\boldsymbol{U}}_{pa\{j,k\}|j,k}. \qquad (68)
\end{aligned}$$

Hence

$$\begin{aligned}
\boldsymbol{m}(\tilde{\boldsymbol{U}}_{j,k|j,k}|\boldsymbol{\nu}_{j,k|j,k}) &= \boldsymbol{m}\Big[\boldsymbol{m}(\tilde{\boldsymbol{U}}_{j,k|j,k}|\boldsymbol{\nu}_{j,k|j,k}, \tilde{\boldsymbol{U}}_{pa\{j,k\}|j,k})|\boldsymbol{\nu}_{j,k|j,k}\Big] \\
&= \boldsymbol{m}\Big[\boldsymbol{m}(\tilde{\boldsymbol{U}}_{j,k|j,k}|\tilde{\boldsymbol{U}}_{pa\{j,k\}|j,k})|\boldsymbol{\nu}_{j,k|j,k}\Big] \\
&= \boldsymbol{m}(\boldsymbol{J}_{j,k}\tilde{\boldsymbol{U}}_{pa\{j,k\}|j,k}|\boldsymbol{\nu}_{j,k|j,k}) \\
&= \boldsymbol{J}_{j,k}\boldsymbol{m}(\tilde{\boldsymbol{U}}_{pa\{j,k\}|j,k}|\boldsymbol{\nu}_{j,k|j,k}),
\end{aligned}$$

where the second equality holds because of (67) and the third equality holds because of (68). □

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

$$\begin{aligned}
\hat{\boldsymbol{U}}_{j,k} &= \hat{\boldsymbol{U}}_{j,k|j,k} + \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{\nu}_{j,k|j,k}) \\
&= \hat{\boldsymbol{U}}_{j,k|j,k} + \boldsymbol{m}(\hat{\boldsymbol{U}}_{j,k} + \tilde{\boldsymbol{U}}_{j,k}|\boldsymbol{\nu}_{j,k|j,k}) \\
&= \hat{\boldsymbol{U}}_{j,k|j,k} + \boldsymbol{m}(\hat{\boldsymbol{U}}_{j,k|j,k}|\boldsymbol{\nu}_{j,k|j,k}) + \boldsymbol{m}(\tilde{\boldsymbol{U}}_{j,k|j,k}|\boldsymbol{\nu}_{j,k|j,k}) \\
&= \hat{\boldsymbol{U}}_{j,k|j,k} + \boldsymbol{m}(\tilde{\boldsymbol{U}}_{j,k|j,k}|\boldsymbol{\nu}_{j,k|j,k}) \qquad (69)
\end{aligned}$$

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,

$$\hat{\boldsymbol{U}}_{pa\{j,k\}} = \hat{\boldsymbol{U}}_{pa\{j,k\}|j,k} + \boldsymbol{m}(\tilde{\boldsymbol{U}}_{pa\{j,k\}|j,k}|\boldsymbol{\nu}_{j,k|j,k}) \qquad (70)$$

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 (69) and (70), 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

$$\tilde{\boldsymbol{U}}_{j,k} + \boldsymbol{J}_{j,k}\hat{\boldsymbol{U}}_{pa\{j,k\}} = \tilde{\boldsymbol{U}}_{j,k|j,k} + \boldsymbol{J}_{j,k}\hat{\boldsymbol{U}}_{pa\{j,k\}|j,k}. \qquad (71)$$

Now we compute the variances of both sides of (71):

$$\begin{aligned}
\mathrm{var}(\tilde{\boldsymbol{U}}_{j,k} + \boldsymbol{J}_{j,k}\hat{\boldsymbol{U}}_{pa\{j,k\}}) &= \mathrm{var}(\tilde{\boldsymbol{U}}_{j,k}) + \mathrm{var}(\boldsymbol{J}_{j,k}\hat{\boldsymbol{U}}_{pa\{j,k\}}) \\
&= \mathrm{var}(\boldsymbol{U}_{j,k} - \hat{\boldsymbol{U}}_{j,k}) + \boldsymbol{J}_{j,k}\mathrm{var}(\hat{\boldsymbol{U}}_{pa\{j,k\}})\boldsymbol{J}'_{j,k} \\
&= \hat{\boldsymbol{V}}_{j,k} + \boldsymbol{J}_{j,k}(\boldsymbol{V}_{pa\{j,k\}} - \hat{\boldsymbol{V}}_{pa\{j,k\}})\boldsymbol{J}'_{j,k}, \qquad (72)
\end{aligned}$$

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 $\text{var}(\boldsymbol{U}_{j,k} - \hat{\boldsymbol{U}}_{j,k}) = \boldsymbol{C}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}) = \hat{\boldsymbol{V}}_{j,k}$ and $\boldsymbol{V}_{pa\{j,k\}} = \text{var}(\boldsymbol{U}_{pa\{j,k\}}) = \text{var}(\hat{\boldsymbol{U}}_{pa\{j,k\}} + \tilde{\boldsymbol{U}}_{pa\{j,k\}}) = \text{var}(\hat{\boldsymbol{U}}_{pa\{j,k\}}) + \text{var}(\tilde{\boldsymbol{U}}_{pa\{j,k\}}) = \text{var}(\hat{\boldsymbol{U}}_{pa\{j,k\}}) + \text{var}(\boldsymbol{U}_{pa\{j,k\}} - \hat{\boldsymbol{U}}_{pa\{j,k\}}) = \text{var}(\hat{\boldsymbol{U}}_{pa\{j,k\}}) + \hat{\boldsymbol{V}}_{pa\{j,k\}}$. Similarly,

$$
\begin{aligned}
\text{var}(\tilde{\boldsymbol{U}}_{j,k|j,k} + \boldsymbol{J}_{j,k}\hat{\boldsymbol{U}}_{pa\{j,k\}|j,k}) &= \text{var}(\tilde{\boldsymbol{U}}_{j,k|j,k}) + \boldsymbol{J}_{j,k}\text{var}(\hat{\boldsymbol{U}}_{pa\{j,k\}|j,k})\boldsymbol{J}'_{j,k} \\
&= \hat{\boldsymbol{V}}_{j,k|j,k} + \boldsymbol{J}_{j,k}(\boldsymbol{V}_{pa\{j,k\}} - \hat{\boldsymbol{V}}_{pa\{j,k\}|j,k})\boldsymbol{J}'_{j,k}, \qquad (73)
\end{aligned}
$$

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 (72) and (73), we obtain $\hat{\boldsymbol{V}}_{j,k} = \hat{\boldsymbol{V}}_{j,k|j,k} + \boldsymbol{J}_{j,k}(\hat{\boldsymbol{V}}_{pa\{j,k\}} - \hat{\boldsymbol{V}}_{pa\{j,k\}|j,k})\boldsymbol{J}'_{j,k}$. $\square$

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

Appendix B3: BLUP

Here we show that the best linear unbiased predictor (BLUP) of the underlying true process $\{\boldsymbol{Y}_{j,k}\}$ is $\hat{\boldsymbol{Y}}_{j,k} = \boldsymbol{X}_{j,k}\hat{\boldsymbol{B}} + \hat{\boldsymbol{U}}_{j,k}$, where $\hat{\boldsymbol{U}}_{j,k} = \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{Z})$ is the BLUP of $\{\boldsymbol{U}_{j,k}\}$. Moreover, we derive the mean squared prediction error (MSPE) associated with $\hat{\boldsymbol{Y}}_{j,k}$.

View the MMTSLM (9)–(12) as a mixed effects model such that $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{U} + \boldsymbol{e}$ where $\boldsymbol{Z} \equiv (\boldsymbol{Z}'_{j,k} : \gamma_{j,k} = 1)'$, $\boldsymbol{X} \equiv [\boldsymbol{X}'_{j,k} : \gamma_{j,k} = 1]'$, $\boldsymbol{U} \equiv (\boldsymbol{U}'_{j,k} : \gamma_{j,k} = 1)'$, and $\boldsymbol{e} \equiv (\boldsymbol{e}'_{j,k} : \gamma_{j,k} = 1)'$ are the vectorized observations, covariates, underlying residual process, and measurement errors. Define $\boldsymbol{V} \equiv \text{var}(\boldsymbol{U})$ and $\boldsymbol{\Phi} \equiv \text{var}(\boldsymbol{e})$. Then $\boldsymbol{Z} \sim N(\boldsymbol{X}\boldsymbol{B}, \boldsymbol{V} + \boldsymbol{\Phi})$. Since

$$
\begin{pmatrix} \boldsymbol{U}_{j,k} \\ \boldsymbol{Z} \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{0}_{n_{j-1}m} \\ \boldsymbol{X}\boldsymbol{B} \end{pmatrix}, \begin{pmatrix} \boldsymbol{V}_{j,k} & \boldsymbol{V}_{j,k,\cdot,\cdot} \\ \boldsymbol{V}'_{j,k,\cdot,\cdot} & \boldsymbol{V} + \boldsymbol{\Phi} \end{pmatrix} \right),
$$

where $\boldsymbol{V}_{j,k,\cdot,\cdot} \equiv \text{cov}(\boldsymbol{U}_{j,k}, \boldsymbol{Z}) = \text{cov}(\boldsymbol{U}_{j,k}, \boldsymbol{U}), j = 1, \ldots J, k = 1, \ldots, N_{j-1}$. Then $\boldsymbol{U}_{j,k}|\boldsymbol{Z} \sim [\boldsymbol{V}_{j,k,\cdot,\cdot}\boldsymbol{D}(\boldsymbol{Z} - \boldsymbol{X}\boldsymbol{B}), \boldsymbol{V}_{j,k} - \boldsymbol{V}_{j,k,\cdot,\cdot}\boldsymbol{D}\boldsymbol{V}'_{j,k,\cdot,\cdot}]$, where $\boldsymbol{D} = (\boldsymbol{V} + \boldsymbol{\Phi})^{-1}$.

The linear predictor of $\boldsymbol{Y}_{j,k}$ is:

$$
\begin{aligned}
\boldsymbol{P}_{j,k}(\boldsymbol{B}) &\equiv \boldsymbol{m}(\boldsymbol{Y}_{j,k}|\boldsymbol{Z};\boldsymbol{\theta}) \\
&= \boldsymbol{X}_{j,k}\boldsymbol{B} + \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{Z};\boldsymbol{\theta}) \\
&= \boldsymbol{X}_{j,k}\boldsymbol{B} + \boldsymbol{V}_{j,k,\cdot,\cdot}\boldsymbol{D}(\boldsymbol{Z} - \boldsymbol{X}\boldsymbol{B}) \\
&= (\boldsymbol{X}_{j,k} - \boldsymbol{V}_{j,k,\cdot,\cdot}\boldsymbol{D}\boldsymbol{X})\boldsymbol{B} + \boldsymbol{V}_{j,k,\cdot,\cdot}\boldsymbol{D}\boldsymbol{Z}.
\end{aligned}
$$

If the parameter $\boldsymbol{B}$ is known, then the BLUP of $\boldsymbol{Y}_{j,k}$ is $\boldsymbol{m}(\boldsymbol{Y}_{j,k}|\boldsymbol{Z};\boldsymbol{\theta})$. Since $\boldsymbol{B}$ is usually unknown, it is natural to estimate $\boldsymbol{B}$ using the data. Let $\tilde{\boldsymbol{B}} \equiv \boldsymbol{AZ}$ denote a linear estimator of $\boldsymbol{B}$ such that $\boldsymbol{P}_{j,k}(\tilde{\boldsymbol{B}})$ is an unbiased predictor of $\boldsymbol{Y}_{j,k}$. Because $\boldsymbol{P}_{j,k}(\boldsymbol{B}) - \boldsymbol{P}_{j,k}(\tilde{\boldsymbol{B}}) = (\boldsymbol{X}_{j,k} - \boldsymbol{V}_{j,k,\cdot,\cdot}\boldsymbol{DX})(\boldsymbol{B} - \tilde{\boldsymbol{B}})$ is a linear function of $\boldsymbol{Z}$, and $\boldsymbol{Y}_{j,k} - \boldsymbol{P}_{j,k}(\boldsymbol{B}) = \boldsymbol{Y}_{j,k} - \boldsymbol{m}(\boldsymbol{Y}_{j,k}|\boldsymbol{Z};\boldsymbol{\theta})$ is uncorrelated with $\boldsymbol{Z}$, we have $\mathrm{cov}\Big(\boldsymbol{Y}_{j,k} - \boldsymbol{P}_{j,k}(\boldsymbol{B}), \boldsymbol{P}_{j,k}(\boldsymbol{B}) - \boldsymbol{P}_{j,k}(\tilde{\boldsymbol{B}})\Big) = \boldsymbol{0}_{n_{j-1}m \times n_{j-1}m}$. Hence the matrix $E\Big[(\boldsymbol{Y}_{j,k} - \boldsymbol{P}_{j,k}(\tilde{\boldsymbol{B}})(\boldsymbol{Y}_{j,k} - \boldsymbol{P}_{j,k}(\tilde{\boldsymbol{B}}))'\Big]$ can be decomposed into

$$
\begin{aligned}
& E\Big[(\boldsymbol{Y}_{j,k} - \boldsymbol{P}_{j,k}(\tilde{\boldsymbol{B}})(\boldsymbol{Y}_{j,k} - \boldsymbol{P}_{j,k}(\tilde{\boldsymbol{B}}))'\Big] \\
= \;& \mathrm{var}\Big(\boldsymbol{Y}_{j,k} - \boldsymbol{P}_{j,k}(\tilde{\boldsymbol{B}})\Big) \\
= \;& \mathrm{var}\Big(\boldsymbol{Y}_{j,k} - \boldsymbol{P}_{j,k}(\boldsymbol{B}) + \boldsymbol{P}_{j,k}(\boldsymbol{B}) - \boldsymbol{P}_{j,k}(\tilde{\boldsymbol{B}})\Big) \\
= \;& \mathrm{var}\Big(\boldsymbol{Y}_{j,k} - \boldsymbol{P}_{j,k}(\boldsymbol{B})\Big) + \mathrm{var}\Big(\boldsymbol{P}_{j,k}(\boldsymbol{B}) - \boldsymbol{P}_{j,k}(\tilde{\boldsymbol{B}})\Big) \\
= \;& \mathrm{var}\Big(\boldsymbol{U}_{j,k} - \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{Z})\Big) + (\boldsymbol{X}_{j,k} - \boldsymbol{V}_{j,k,\cdot,\cdot}\boldsymbol{DX})\mathrm{var}(\tilde{\boldsymbol{B}})(\boldsymbol{X}_{j,k} - \boldsymbol{V}_{j,k,\cdot,\cdot}\boldsymbol{DX})' \\
= \;& \boldsymbol{C}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}) + (\boldsymbol{X}_{j,k} - \boldsymbol{V}_{j,k,\cdot,\cdot}\boldsymbol{DX})\mathrm{var}(\tilde{\boldsymbol{B}})(\boldsymbol{X}_{j,k} - \boldsymbol{V}_{j,k,\cdot,\cdot}\boldsymbol{DX})', \qquad (74)
\end{aligned}
$$

where the first equality holds because $\boldsymbol{P}_{j,k}(\tilde{\boldsymbol{B}})$ is unbiased. The diagonal elements of $E[(\boldsymbol{Y}_{j,k} - \boldsymbol{P}_{j,k}(\tilde{\boldsymbol{B}})(\boldsymbol{Y}_{j,k} - \boldsymbol{P}_{j,k}(\tilde{\boldsymbol{B}}))']$ are the MSPE's of the associated scalar nodes. We know that $\hat{\boldsymbol{U}}_{j,k} = \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{Z};\boldsymbol{\theta})$ is the BLUP of $\boldsymbol{U}_{j,k}$. Furthermore, the generalized least square (GLS) estimator $\hat{\boldsymbol{B}}_{\mathrm{GLS}} \equiv (\boldsymbol{X}'\boldsymbol{DX})^{-1}\boldsymbol{X}'\boldsymbol{DZ}$ is the best linear unbiased estimator of $\boldsymbol{B}$ such that $\mathrm{var}(\boldsymbol{c}'\hat{\boldsymbol{B}}_{\mathrm{GLS}}) \leq \mathrm{var}(\boldsymbol{c}'\tilde{\boldsymbol{B}})$ for any vector $\boldsymbol{c}$ and linear estimator $\tilde{\boldsymbol{B}}$ and $\hat{\boldsymbol{B}}_{\mathrm{GLS}} = \hat{\boldsymbol{B}}$. Then, it follows that $\hat{\boldsymbol{Y}}_{j,k} = \boldsymbol{X}_{j,k}\hat{\boldsymbol{B}} + \hat{\boldsymbol{U}}_{j,k}$ has the minimum MSPE $E[(\boldsymbol{Y}_{j,k} - \hat{\boldsymbol{Y}}_{j,k})(\boldsymbol{Y}_{j,k} - \hat{\boldsymbol{Y}}_{j,k})']$ in the sense that the diagonal elements are minimized. Hence $\hat{\boldsymbol{Y}}_{j,k}$ is the BLUP of $\boldsymbol{Y}_{j,k}$.

In addition, we compute the MSPE associated with the BLUP of a linear combination of the latent variables $c_1\boldsymbol{Y}_{j_1,k_1} + \cdots + c_L\boldsymbol{Y}_{j_L,k_L}$ where $c_1, \ldots, c_L$ are constants. The MSPE is

$$
\begin{aligned}
& \boldsymbol{C}(c_1\boldsymbol{Y}_{j_1,k_1} + \cdots + c_L\boldsymbol{Y}_{j_L,k_L}|\boldsymbol{Z}) \\
= \;& \mathrm{var}\Big(\sum_{l=1}^{L} c_l[\boldsymbol{Y}_{j_l,k_l} - \boldsymbol{P}_{j_l,k_l}(\hat{\boldsymbol{B}})]\Big) \\
= \;& \sum_{l=1}^{L} c_l^2 \mathrm{var}(\boldsymbol{Y}_{j_l,k_l} - \boldsymbol{P}_{j_l,k_l}(\hat{\boldsymbol{B}})) + \sum_{h=1}^{L}\sum_{l=1}^{L} c_h c_l \mathrm{cov}(\boldsymbol{Y}_{j_h,k_h} - \boldsymbol{P}_{j_h,k_h}(\hat{\boldsymbol{B}}), \boldsymbol{Y}_{j_l,k_l} - \boldsymbol{P}_{j_l,k_l}(\hat{\boldsymbol{B}})),
\end{aligned}
$$

where $\mathrm{var}(\boldsymbol{Y}_{j_l,k_l} - \boldsymbol{P}_{j_l,k_l}(\hat{\boldsymbol{B}}))$ can be computed using (74) and the algorithm in Section 3.3 which involves operations of order $\mathcal{O}(n)$.

To obtain $\text{cov}(\hat{\boldsymbol{Y}}_{j_h,k_h} - \boldsymbol{Y}_{j_h,k_h}, \hat{\boldsymbol{Y}}_{j_l,k_l} - \boldsymbol{Y}_{j_l,k_l})$, we need an algorithm to compute the covariance of any two given vector nodes $\{j,k\}$ and $\{h,i\}$:

$$
\begin{aligned}
&\text{cov}(\boldsymbol{Y}_{j,k} - \boldsymbol{P}_{j,k}(\hat{\boldsymbol{B}}), \boldsymbol{Y}_{h,i} - \boldsymbol{P}_{h,i}(\hat{\boldsymbol{B}})) \\
&\quad = \text{cov}\Big([\boldsymbol{Y}_{j,k} - \boldsymbol{P}_{j,k}(\boldsymbol{B})] - [\boldsymbol{P}_{j,k}(\hat{\boldsymbol{B}}) - \boldsymbol{P}_{j,k}(\boldsymbol{B})], [\boldsymbol{Y}_{h,i} - \boldsymbol{P}_{h,i}(\boldsymbol{B})] - [\boldsymbol{P}_{h,k}(\hat{\boldsymbol{B}}) - \boldsymbol{P}_{h,i}(\boldsymbol{B})]\Big) \\
&\quad = \text{cov}\Big([\boldsymbol{Y}_{j,k} - \boldsymbol{P}_{j,k}(\boldsymbol{B})], [\boldsymbol{Y}_{h,i} - \boldsymbol{P}_{h,i}(\boldsymbol{B})]\Big) + \text{cov}\Big([\boldsymbol{P}_{j,k}(\hat{\boldsymbol{B}}) - \boldsymbol{P}_{j,k}(\boldsymbol{B})], [\boldsymbol{P}_{h,k}(\hat{\boldsymbol{B}}) - \boldsymbol{P}_{h,i}(\boldsymbol{B})]\Big) \\
&\quad = \text{cov}\Big([\boldsymbol{Y}_{j,k} - \boldsymbol{P}_{j,k}(\boldsymbol{B})], [\boldsymbol{Y}_{h,i} - \boldsymbol{P}_{h,i}(\boldsymbol{B})]\Big) + \text{cov}\Big((\boldsymbol{X}_{j,k} - \boldsymbol{V}_{j,k,\cdot,\cdot}\boldsymbol{D}\boldsymbol{X})(\hat{\boldsymbol{B}} - \boldsymbol{B}), (\boldsymbol{X}_{h,i} - \boldsymbol{V}_{h,i,\cdot,\cdot}\boldsymbol{D}\boldsymbol{X})(\hat{\boldsymbol{B}} - \boldsymbol{B})\Big) \\
&\quad = \text{cov}\Big([\boldsymbol{Y}_{j,k} - \boldsymbol{P}_{j,k}(\boldsymbol{B})], [\boldsymbol{Y}_{h,i} - \boldsymbol{P}_{h,i}(\boldsymbol{B})]\Big) + (\boldsymbol{X}_{j,k} - \boldsymbol{V}_{j,k,\cdot,\cdot}\boldsymbol{D}\boldsymbol{X})\text{var}(\hat{\boldsymbol{B}})(\boldsymbol{X}_{h,i} - \boldsymbol{V}_{h,i,\cdot,\cdot}\boldsymbol{D}\boldsymbol{X})',
\end{aligned}
$$

where the second and third equalities hold because of the argument before (74) and $(\boldsymbol{X}_{j,k} - \boldsymbol{V}_{j,k,\cdot,\cdot}\boldsymbol{D}\boldsymbol{X})\text{var}(\hat{\boldsymbol{B}})(\boldsymbol{X}_{h,i} - \boldsymbol{V}_{h,i,\cdot,\cdot}\boldsymbol{D}\boldsymbol{X})'$ can be obtained using the algorithm in Section 3.3 which involves operations of order $\mathcal{O}(n)$.

Since $\boldsymbol{Y}_{j,k} - \boldsymbol{P}_{j,k}(\boldsymbol{B}) = \boldsymbol{U}_{j,k} - \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{Z})$ and $\boldsymbol{Y}_{h,i} - \boldsymbol{P}_{h,i}(\boldsymbol{B}) = \boldsymbol{U}_{h,i} - \boldsymbol{m}(\boldsymbol{U}_{h,i}|\boldsymbol{Z})$, we have

$$
\begin{aligned}
&\text{cov}\Big([\boldsymbol{Y}_{j,k} - \boldsymbol{P}_{j,k}(\boldsymbol{B})], [\boldsymbol{Y}_{h,i} - \boldsymbol{P}_{h,i}(\boldsymbol{B})]\Big) \\
&\quad = \text{cov}\Big([\boldsymbol{U}_{j,k} - \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{Z})], [\boldsymbol{U}_{h,i} - \boldsymbol{m}(\boldsymbol{U}_{h,i}|\boldsymbol{Z})]\Big) \\
&\quad = \boldsymbol{J}_{j,k} \cdots \boldsymbol{J}_{j_1,k_1}\text{var}\Big(\boldsymbol{U}_{an\{j,k,h,i\}} - \boldsymbol{m}(\boldsymbol{U}_{an\{j,k,h,i\}}|\boldsymbol{Z})\Big)\boldsymbol{J}_{h_1,i_1} \cdots \boldsymbol{J}_{h,i} \\
&\quad = \boldsymbol{J}_{j,k} \cdots \boldsymbol{J}_{j_1,k_1}\boldsymbol{C}(\boldsymbol{U}_{an\{j,k,h,i\}}|\boldsymbol{Z})\boldsymbol{J}_{h_1,i_1} \cdots \boldsymbol{J}_{h,i},
\end{aligned}
$$

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 $\boldsymbol{B}_{j,k} \equiv \boldsymbol{V}_{pa\{j,k\}}\boldsymbol{A}'_{j,k}\boldsymbol{V}^{+}_{j,k}$ are defined in the smoothing step of Kalman filter right after (26) and the last equality holds because of formula (2.15) of Huang et al. (2002). Here, $an\{j,k,h,i\}$ is the first common ancestor of $\{j,k\}$ and $\{h,i\}$, $(\{j,k\}, \ldots, \{j_1,k_1\}, an\{j,k,h,i\})$ is the path from $(\{j,k\}$ to $an\{j,k,h,i\}$, and $(\{h,i\}, \ldots, \{h_1,i_1\}, an\{j,k,h,i\})$ is the path from $\{h,i\}$ to $an\{j,k,h,i\}$. Since $\boldsymbol{J}_{j,k}$ and $\boldsymbol{C}(\boldsymbol{U}_{an\{j,k,h,i\}}|\boldsymbol{Z})$ can be obtained as by-products of the change-of-resolution Kalman filter, this only involves operations of order $\mathcal{O}(n)$.

Appendix C: 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.

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) $\vec{\boldsymbol{z}}'(\boldsymbol{A}\otimes\boldsymbol{B})\vec{\boldsymbol{z}} = \langle\vec{\boldsymbol{z}},\ (\boldsymbol{A}\otimes\boldsymbol{B})\vec{\boldsymbol{z}}\rangle = \mathrm{tr}[\boldsymbol{z}'\boldsymbol{A}\boldsymbol{z}\boldsymbol{B}]$, *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* $|\boldsymbol{A}\otimes\boldsymbol{I}_n + \boldsymbol{B}\otimes(\boldsymbol{1}_n\boldsymbol{1}_n')| = |\boldsymbol{A}|^{n-1}|\boldsymbol{A}+n\boldsymbol{B}|$ *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 degree of freedom n and parameter $\boldsymbol{\Sigma}$ (Definition C.9 of Lauritzen, 1996).*

Proof. (i) Theorem 1 of Chapter 2.7 of Searle (1997).

(ii) If $\boldsymbol{A} = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n)$ where $\boldsymbol{a}_j$ is the j^{th} column of $\boldsymbol{A}$, then $\boldsymbol{A}\boldsymbol{Q}_{hi}$ has $\boldsymbol{a}_h$ as the i^{th} column and zero otherwise. Hence $\text{tr}(\boldsymbol{A}\boldsymbol{Q}_{hi}) = a_{ih}$.

(iii) Lemma 5.2.1 of Harville (1997).

(iv) Propositions B.2 of Lauritzen (1996).

(v) Formula in the proof of Proposition C.8 of Lauritzen (1996).

(vi) From Theorem 16.3.2 of Harville (1997), $\boldsymbol{C} \otimes \boldsymbol{A} = \boldsymbol{K}_{nm}[\boldsymbol{A} \otimes \boldsymbol{C}]\boldsymbol{K}_{mn}$, $\boldsymbol{D} \otimes \boldsymbol{B} = \boldsymbol{K}_{nm}[\boldsymbol{B} \otimes \boldsymbol{D}]\boldsymbol{K}_{mn}$ where $\boldsymbol{K}_{mn}$ and $\boldsymbol{K}_{nm}$ are the permutation matrices defined in Chapter 16.3 of Harville (1997) such that $\boldsymbol{K}'_{mn} = \boldsymbol{K}^{-1}_{mn} = \boldsymbol{K}_{nm}$. Then $|\boldsymbol{C} \otimes \boldsymbol{A} + \boldsymbol{D} \otimes \boldsymbol{B}| = |\boldsymbol{K}_{nm}||\boldsymbol{A} \otimes \boldsymbol{C} + \boldsymbol{B} \otimes \boldsymbol{D}||\boldsymbol{K}_{mn}| = |\boldsymbol{A} \otimes \boldsymbol{C} + \boldsymbol{B} \otimes \boldsymbol{D}||\boldsymbol{K}_{nm}\boldsymbol{K}_{mn}| = |\boldsymbol{A} \otimes \boldsymbol{C} + \boldsymbol{B} \otimes \boldsymbol{D}|$. From page 68 of Rao (1973), we get $|\boldsymbol{A} \otimes \boldsymbol{I}_n + \boldsymbol{B} \otimes (\boldsymbol{1}_n\boldsymbol{1}'_n)| = |\boldsymbol{A}|^{n-1}|\boldsymbol{A} + n\boldsymbol{B}|$. From Proposition B.2 of Lauritzen (1996), we have $|\boldsymbol{A} \otimes \boldsymbol{C}| = |\boldsymbol{A}|^n|\boldsymbol{C}|^m$.

(vii) Formula from page 307 of Eaton (1983). □

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 $\boldsymbol{\Psi}_j = \frac{1}{a_j}(\boldsymbol{D}_j - \boldsymbol{D}_{j+1}); j = 1, \ldots, J-1$.*

Proof. By the definition of $\boldsymbol{C}_j$ and $\boldsymbol{D}_j$, it is easy to obtain $\sum_{k=j}^{J} \boldsymbol{C}_k = -\boldsymbol{D}_j^{-1}; j = 1, \ldots, J$ and $\boldsymbol{D}_j - \boldsymbol{D}_{j+1} = a_j\boldsymbol{\Psi}_j; j = 1, \ldots, J-1$.

$$
\begin{aligned}
\boldsymbol{\Omega} \times [\sum_{j=1}^{J} -\boldsymbol{A}_j \otimes \boldsymbol{C}_j] &= [\sum_{j=k}^{J} a_j(\boldsymbol{A}_k \otimes \boldsymbol{\Psi}_k)] \times [\sum_{j=1}^{J} -\boldsymbol{A}_j \otimes \boldsymbol{C}_j] \\
&= -\sum_{j=1}^{J}\sum_{k=1}^{J} a_j(\boldsymbol{A}_j \otimes \boldsymbol{\Psi}_j) \times (\boldsymbol{A}_k \otimes \boldsymbol{C}_k) \\
&= -\sum_{j=1}^{J}\sum_{k=1}^{J} a_j(\boldsymbol{A}_j\boldsymbol{A}_k) \otimes (\boldsymbol{\Psi}_j\boldsymbol{C}_k) \\
&= -\sum_{j=1}^{J}\sum_{k=1}^{J} a_j(\boldsymbol{A}_{\min(j,k)}) \otimes (\boldsymbol{\Psi}_j\boldsymbol{C}_k)
\end{aligned}
$$

$$
\begin{aligned}
&= -\sum_{k=1}^{J}[\sum_{j=1}^{k-1} a_j \boldsymbol{A}_j \otimes (\boldsymbol{\Psi}_j \boldsymbol{C}_k) + \sum_{j=k}^{J} a_j \boldsymbol{A}_k \otimes (\boldsymbol{\Psi}_j \boldsymbol{C}_k)] \\
&= -\sum_{k=1}^{J}[\sum_{j=1}^{k-1} a_j \boldsymbol{A}_j \otimes (\boldsymbol{\Psi}_j \boldsymbol{C}_k) + \boldsymbol{A}_k \otimes (\boldsymbol{D}_k \boldsymbol{C}_k)] \\
&= -\sum_{k=1}^{J}\sum_{j=1}^{k-1} a_j \boldsymbol{A}_j \otimes (\boldsymbol{\Psi}_j \boldsymbol{C}_k) - \sum_{k=1}^{J} \boldsymbol{A}_k \otimes (\boldsymbol{D}_k \boldsymbol{C}_k) \\
&= -\sum_{j=1}^{J-1}\sum_{k=j+1}^{J} a_j \boldsymbol{A}_j \otimes (\boldsymbol{\Psi}_j \boldsymbol{C}_k) - \sum_{k=1}^{J} \boldsymbol{A}_k \otimes (\boldsymbol{D}_k \boldsymbol{C}_k) \\
&= -\sum_{j=1}^{J-1} a_j \boldsymbol{A}_j \otimes (\boldsymbol{\Psi}_j \sum_{k=j+1}^{J} \boldsymbol{C}_k) - \sum_{k=1}^{J} \boldsymbol{A}_k \otimes (\boldsymbol{D}_k \boldsymbol{C}_k) \\
&= \sum_{j=1}^{J-1} a_j \boldsymbol{A}_j \otimes (\boldsymbol{\Psi}_j \boldsymbol{D}_{j+1}^{-1}) - \sum_{j=1}^{J} \boldsymbol{A}_j \otimes (\boldsymbol{D}_j \boldsymbol{C}_j) \\
&= \sum_{j=1}^{J-1} \boldsymbol{A}_j \otimes (a_j \boldsymbol{\Psi}_j \boldsymbol{D}_{j+1}^{-1} - \boldsymbol{D}_j \boldsymbol{C}_j) - \boldsymbol{A}_J \otimes (\boldsymbol{D}_J \boldsymbol{C}_J) \\
&= \sum_{j=1}^{J-1} \boldsymbol{A}_j \otimes (a_j \boldsymbol{\Psi}_j \boldsymbol{D}_{j+1}^{-1} - \boldsymbol{D}_j \boldsymbol{C}_j) + \boldsymbol{I}_{N_J m} \\
&= \boldsymbol{I}_{N_J m}
\end{aligned}
$$

where the fourth equality holds because of Lemma 7 (i), the twelfth equality holds because $-\boldsymbol{A}_J \otimes (\boldsymbol{D}_J \boldsymbol{C}_J) = -\boldsymbol{I}_{N_J} \otimes [\boldsymbol{D}_J(-\boldsymbol{D}_J^{-1})] = \boldsymbol{I}_{N_J m}$, and the last equality holds because $a_j \boldsymbol{\Psi}_j \boldsymbol{D}_{j+1}^{-1} - \boldsymbol{D}_j \boldsymbol{C}_j = (\boldsymbol{D}_j - \boldsymbol{D}_{j+1})\boldsymbol{D}_{j+1}^{-1} - \boldsymbol{D}_j(\boldsymbol{D}_{j+1}^{-1} - \boldsymbol{D}_j^{-1}) = \boldsymbol{0}_{m\times m}$. □

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

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

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

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

Proof. (i) For $j = 0, 1, \ldots, J$, $\boldsymbol{A}_j = \frac{1}{a_j}[\boldsymbol{I}_{N_j} \otimes (\mathbf{1}_{a_j}\mathbf{1}'_{a_j})]$ is a symmetric matrix, hence $\boldsymbol{A}_j = \boldsymbol{A}'_j$. For $0 \leq j \leq k \leq J$,

$$\begin{aligned}
\boldsymbol{A}_j\boldsymbol{A}_k &= \frac{1}{a_j a_k}[\boldsymbol{I}_{N_j} \otimes (\mathbf{1}_{a_j}\mathbf{1}'_{a_j})][\boldsymbol{I}_{N_k} \otimes (\mathbf{1}_{a_k}\mathbf{1}'_{a_k})] \\
&= \frac{1}{a_j a_k}[\boldsymbol{I}_{N_j} \otimes (\mathbf{1}_{a_j/a_k}\mathbf{1}'_{a_j/a_k}) \otimes (\mathbf{1}_{a_k}\mathbf{1}'_{a_k})][\boldsymbol{I}_{N_k} \otimes (\mathbf{1}_{a_k}\mathbf{1}'_{a_k})] \\
&= \frac{1}{a_j a_k}[(\boldsymbol{I}_{N_j} \otimes (\mathbf{1}_{a_j/a_k}\mathbf{1}'_{a_j/a_k}))\boldsymbol{I}_{N_k}] \otimes [(\mathbf{1}_{a_k}\mathbf{1}'_{a_k})(\mathbf{1}_{a_k}\mathbf{1}'_{a_k})] \\
&= \frac{a_k}{a_j a_k}[\boldsymbol{I}_{N_j} \otimes (\mathbf{1}_{a_j/a_k}\mathbf{1}'_{a_j/a_k}) \otimes (\mathbf{1}_{a_k}\mathbf{1}'_{a_k})] \\
&= \frac{1}{a_j}[\boldsymbol{I}_{N_j} \otimes (\mathbf{1}_{a_j}\mathbf{1}'_{a_j})] \\
&= \boldsymbol{A}_j,
\end{aligned}$$

where the third equality holds because of Lemma 5 (iv). If $0 \leq k \leq j \leq J$, then $\boldsymbol{A}_j\boldsymbol{A}_k = (\boldsymbol{A}'_k\boldsymbol{A}'_j)' = (\boldsymbol{A}_k\boldsymbol{A}_j)' = \boldsymbol{A}'_k = \boldsymbol{A}_k$. Hence $\boldsymbol{A}_j\boldsymbol{A}_k = \boldsymbol{A}_{\min(j,k)}$ and if $j = k$, we have $\boldsymbol{A}_j^2 = \boldsymbol{A}_j$.

(ii) For $0 \leq k \leq j \leq J$, use part (i), we have $(\boldsymbol{A}_j - \boldsymbol{A}_k)^2 = \boldsymbol{A}_j^2 - \boldsymbol{A}_j\boldsymbol{A}_k - \boldsymbol{A}_k\boldsymbol{A}_j + \boldsymbol{A}_k^2 = \boldsymbol{A}_j - \boldsymbol{A}_k - \boldsymbol{A}_k + \boldsymbol{A}_k = \boldsymbol{A}_j - \boldsymbol{A}_k$.

(iii) For $j = 0, 1, \ldots, J$,

$\boldsymbol{A}_j\mathbf{1}_{N_J} = \frac{1}{a_j}[\boldsymbol{I}_{N_j} \otimes (\mathbf{1}_{a_j}\mathbf{1}'_{a_j})]\mathbf{1}_{N_J} = \frac{1}{a_j}[\boldsymbol{I}_{N_j} \otimes (\mathbf{1}_{a_j}\mathbf{1}'_{a_j})][\mathbf{1}_{N_j} \otimes \mathbf{1}_{a_j}] = \frac{1}{a_j}[\boldsymbol{I}_{N_j}\mathbf{1}_{N_j}] \otimes [(\mathbf{1}_{a_j}\mathbf{1}'_{a_j})\mathbf{1}_{a_j}] = \mathbf{1}_{N_j} \otimes \mathbf{1}_{a_j} = \mathbf{1}_{N_J}$.

(iv) For $j = 0, 1, \ldots, J$, from part (i), $\boldsymbol{A}_j$ is idempotent. Using Theorem 10.2.1 of Harville (1997) and Lemma 5 (iv), we have $\text{Rank}(\boldsymbol{A}_j) = \text{tr}(\boldsymbol{A}_j) = \frac{1}{a_j}\text{tr}[\boldsymbol{I}_{N_j} \otimes (\mathbf{1}_{a_j}\mathbf{1}'_{a_j})] = \frac{1}{a_j}\text{tr}[\boldsymbol{I}_{N_j}]\text{tr}[\mathbf{1}_{a_j}\mathbf{1}'_{a_j}] = \frac{1}{a_j}N_j a_j = N_j$.

(v) For $j = 1, \ldots, J$, from part (ii), $\boldsymbol{A}_j - \boldsymbol{A}_{j-1}$ is idempotent. Similar to the proof of part (iv), we have $\text{Rank}(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) = \text{tr}(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) = N_j - N_{j-1}$.

(vi)

$$\begin{aligned}
\boldsymbol{\Omega}[(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{C}] &= [\sum_{k=1}^{J} a_j(\boldsymbol{A}_k \otimes \boldsymbol{\Psi}_k)][(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{C}] \\
&= [\sum_{k=1}^{J} a_j \boldsymbol{A}_k(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})] \otimes [\boldsymbol{\Psi}_k \boldsymbol{C}] \\
&= [\sum_{k=1}^{j-1} a_k \boldsymbol{A}_k(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) + \sum_{k=j}^{J} a_k \boldsymbol{A}_k(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})] \otimes (\boldsymbol{\Psi}_k \boldsymbol{C}) \\
&= [\sum_{k=j}^{J} a_k(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})] \otimes (\boldsymbol{\Psi}_k \boldsymbol{C}) \\
&= (\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes [\sum_{k=j}^{J} a_k \boldsymbol{\Psi}_k \boldsymbol{C}] \\
&= (\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes (\boldsymbol{D}_j \boldsymbol{C}),
\end{aligned}$$

where the second equality holds because of Lemma 5 (iv) and the fourth equality holds because of part (i).

(vii) $\boldsymbol{\Omega}(\boldsymbol{1}_{N_J} \otimes \boldsymbol{I}_m) = [\sum_{j=1}^{J} a_j(\boldsymbol{A}_j \otimes \boldsymbol{\Psi}_j)][\boldsymbol{1}_{N_J} \otimes \boldsymbol{I}_m] = \sum_{j=1}^{J} a_j(\boldsymbol{A}_j \boldsymbol{1}_{N_J}) \otimes (\boldsymbol{\Psi}_j \boldsymbol{I}_m) = \boldsymbol{1}_{N_J} \otimes (\sum_{j=1}^{J} a_j \boldsymbol{\Psi}_j) = \boldsymbol{1}_{N_J} \otimes \boldsymbol{D}_1$, where the third equality holds because of part (iii). Moreover $(\boldsymbol{1}_{N_J} \otimes \boldsymbol{I}_m)' \boldsymbol{\Omega}(\boldsymbol{1}_{N_J} \otimes \boldsymbol{I}_m) = (\boldsymbol{1}'_{N_J} \otimes \boldsymbol{I}_m)(\boldsymbol{1}_{N_J} \otimes \boldsymbol{D}_1) = (\boldsymbol{1}'_{N_J} \boldsymbol{1}_{N_J}) \otimes (\boldsymbol{I}_m \boldsymbol{D}_1) = N_J \boldsymbol{D}_1$.

(viii) $\boldsymbol{\Omega}^{-1}(\boldsymbol{1}_{N_J} \otimes \boldsymbol{I}_m) = [\sum_{j=1}^{J}(-\boldsymbol{A}_j \otimes \boldsymbol{C}_j)][\boldsymbol{1}_{N_J} \otimes \boldsymbol{I}_m] = \sum_{j=1}^{J}[-\boldsymbol{A}_j \boldsymbol{1}_{N_J}] \otimes [\boldsymbol{C}_j \boldsymbol{I}_m] = -\boldsymbol{1}_{N_J} \otimes [\sum_{j=1}^{J} \boldsymbol{C}_j] = \boldsymbol{1}_{N_J} \otimes \boldsymbol{D}_1^{-1}$, where the first and last equality holds because of Lemma 6. Moreover $(\boldsymbol{1}_{N_J} \otimes \boldsymbol{I}_m)' \boldsymbol{\Omega}^{-1}(\boldsymbol{1}_{N_J} \otimes \boldsymbol{I}_m) = (\boldsymbol{1}'_{N_J} \otimes \boldsymbol{I}_m)(\boldsymbol{1}_{N_J} \otimes \boldsymbol{D}_1^{-1}) = (\boldsymbol{1}'_{N_J} \boldsymbol{1}_{N_J}) \otimes (\boldsymbol{I}_m \boldsymbol{D}_1^{-1}) = N_J \boldsymbol{D}_1^{-1}$. □

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

Proof. Define $f(i,j) \equiv n_i \cdots n_j = \prod_{h=i}^{j} n_h$ where $0 \le i \le j \le J-1$. Then

$$\begin{aligned}
|\boldsymbol{\Omega}| &= |\sum_{j=1}^{J} a_j(\boldsymbol{A}_j \otimes \boldsymbol{\Psi}_j)| \\
&= |\sum_{j=1}^{J} a_j(\boldsymbol{\Psi}_j \otimes \boldsymbol{A}_j)|
\end{aligned}$$

$$
\begin{aligned}
&= |(\boldsymbol{\Psi}_J \otimes \boldsymbol{I}_{N_{J-1}}) \otimes \boldsymbol{I}_{n_{J-1}} + [\boldsymbol{\Psi}_{J-1} \otimes \boldsymbol{I}_{N_{J-1}} + \boldsymbol{\Psi}_{J-2} \otimes \boldsymbol{I}_{N_{J-2}} \otimes (\mathbf{1}_{f(J-2,J-2)}\mathbf{1}'_{f(J-2,J-2)}) \\
&\qquad + \cdots + \boldsymbol{\Psi}_1 \otimes \boldsymbol{I}_{N_1} \otimes (\mathbf{1}_{f(1,J-2)}\mathbf{1}'_{f(1,J-2)})] \otimes (\mathbf{1}_{n_{J-1}}\mathbf{1}'_{n_{J-1}})| \\
&= |\boldsymbol{\Psi}_J \otimes \boldsymbol{I}_{N_{J-1}}|^{n_{J-1}-1}|\boldsymbol{\Psi}_J \otimes \boldsymbol{I}_{N_{J-1}} + n_{J-1}[\boldsymbol{\Psi}_{J-1} \otimes \boldsymbol{I}_{N_{J-1}} + \boldsymbol{\Psi}_{J-2} \otimes \boldsymbol{I}_{N_{J-2}} \otimes (\mathbf{1}_{f(J-2,J-2)}\mathbf{1}'_{f(J-2,J-2)}) \\
&\qquad + \cdots + \boldsymbol{\Psi}_1 \otimes \boldsymbol{I}_{N_1} \otimes (\mathbf{1}_{f(1,J-2)}\mathbf{1}'_{f(1,J-2)})]| \\
&= |\boldsymbol{D}_J|^{N_J - N_{J-1}}|(\boldsymbol{\Psi}_J + n_{J-1}\boldsymbol{\Psi}_{J-1}) \otimes \boldsymbol{I}_{N_{J-1}} + n_{J-1}[\boldsymbol{\Psi}_{J-2} \otimes \boldsymbol{I}_{N_{J-2}} \otimes (\mathbf{1}_{f(J-2,J-2)}\mathbf{1}'_{f(J-2,J-2)}) \\
&\qquad + \cdots + \boldsymbol{\Psi}_1 \otimes \boldsymbol{I}_{N_1} \otimes (\mathbf{1}_{f(1,J-2)}\mathbf{1}'_{f(1,J-2)})]| \\
&= |\boldsymbol{D}_J|^{N_J - N_{J-1}}|\boldsymbol{D}_{J-1} \otimes \boldsymbol{I}_{N_{J-1}} + n_{J-1}[\boldsymbol{\Psi}_{J-2} \otimes \boldsymbol{I}_{N_{J-2}} \otimes (\mathbf{1}_{f(J-2,J-2)}\mathbf{1}'_{f(J-2,J-2)}) \\
&\qquad + \cdots + \boldsymbol{\Psi}_1 \otimes \boldsymbol{I}_{N_1} \otimes (\mathbf{1}_{f(1,J-2)}\mathbf{1}'_{f(1,J-2)})]| \\
&= |\boldsymbol{D}_J|^{N_J - N_{J-1}}|(\boldsymbol{D}_{J-1} \otimes \boldsymbol{I}_{N_{J-2}}) \otimes \boldsymbol{I}_{n_{J-2}} + n_{J-1}[\boldsymbol{\Psi}_{J-2} \otimes \boldsymbol{I}_{N_{J-2}} + \boldsymbol{\Psi}_{J-3} \otimes \boldsymbol{I}_{N_{J-3}} \otimes (\mathbf{1}_{f(J-3,J-3)}\mathbf{1}'_{f(J-3,J-3)}) \\
&\qquad + \cdots + \boldsymbol{\Psi}_1 \otimes \boldsymbol{I}_{N_1} \otimes (\mathbf{1}_{f(1,J-3)}\mathbf{1}'_{f(1,J-3)})] \otimes (\mathbf{1}_{n_{J-2}}\mathbf{1}'_{n_{J-2}})| \\
&= \vdots \\
&= |\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}.
\end{aligned}
$$

where the second, fourth and fifth equalities hold because of Lemma 5 (vi).

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

Proof.

$$
\begin{aligned}
&(\boldsymbol{Z} - \boldsymbol{XB})'\boldsymbol{\Omega}^{-1}(\boldsymbol{Z} - \boldsymbol{XB}) \\
=\;& (\boldsymbol{Z} - \boldsymbol{XB})'[\sum_{j=1}^{J}(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{D}_j^{-1} + \boldsymbol{A}_0 \otimes \boldsymbol{D}_j^{-1}](\boldsymbol{Z} - \boldsymbol{XB}) \\
=\;& \sum_{j=1}^{J}(\boldsymbol{Z} - \boldsymbol{XB})'[(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{D}_j^{-1}](\boldsymbol{Z} - \boldsymbol{XB}) + (\boldsymbol{Z} - \boldsymbol{XB})'[\boldsymbol{A}_0 \otimes \boldsymbol{D}_1^{-1}](\boldsymbol{Z} - \boldsymbol{XB}) \\
=\;& \sum_{j=1}^{J}\mathrm{tr}[(\boldsymbol{z} - \boldsymbol{x\beta})'(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})(\boldsymbol{z} - \boldsymbol{x\beta})\boldsymbol{D}_j^{-1}] + \mathrm{tr}[(\boldsymbol{z} - \boldsymbol{z\beta})'\boldsymbol{A}_0(\boldsymbol{z} - \boldsymbol{z\beta})\boldsymbol{D}_1^{-1}] \\
=\;& \sum_{j=1}^{J}\mathrm{tr}[SS_j(\boldsymbol{\beta})\boldsymbol{D}_j^{-1}] + \mathrm{tr}[SS_0(\boldsymbol{\beta})\boldsymbol{D}_1^{-1}],
\end{aligned}
$$

where the first equality holds because of Lemma 6, and the third equality holds because of Lemma 5 (v). □

LEMMA 10. *(i)* $\frac{\partial \mathrm{tr}(\boldsymbol{AXB})}{\partial \boldsymbol{X}} = \boldsymbol{A}'\boldsymbol{B}$ *where* $\boldsymbol{A}$, $\boldsymbol{B}$, *and* $\boldsymbol{X}$ *are conformable matrices.*

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

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

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

(v) $\frac{\partial \log|\boldsymbol{AF}^{-1}\boldsymbol{B}|}{\partial x_i} = -\mathrm{tr}[\boldsymbol{F}^{-1}\boldsymbol{B}(\boldsymbol{AF}^{-1}\boldsymbol{B})^{-1}\boldsymbol{AF}^{-1}\frac{\partial \boldsymbol{F}}{\partial x_i}]$ *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) $\frac{\partial f}{\partial x_i} = \mathrm{tr}\Big[(\frac{\partial g}{\partial \boldsymbol{H}})'(\frac{\partial \boldsymbol{H}}{\partial x_i})\Big]$ *where function* $f(\boldsymbol{x}) = g(\boldsymbol{H}(\boldsymbol{x}))$, g *is a function that depends on the elements of* $\boldsymbol{H}$ *and* $\boldsymbol{H}$ *is a matrix that depends on* $\boldsymbol{x} = (x_1, \ldots, x_n)'$.

Proof. See Chapter 15 of Harville (1997). □

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

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

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

Proof. Let $\boldsymbol{D}_{jhi}^{-1}$ denote the the (h, i) element of $\boldsymbol{D}_j^{-1}$, we have

$$\begin{aligned}
& \Big[\frac{\partial}{\partial \boldsymbol{D}_{jhi}^{-1}}(\boldsymbol{Z} - \boldsymbol{X}\hat{\boldsymbol{B}})\Big]'\boldsymbol{\Omega}^{-1}(\boldsymbol{Z} - \boldsymbol{X}\hat{\boldsymbol{B}}) \\
= & -\Big[\frac{\partial}{\partial \boldsymbol{D}_{jhi}^{-1}}\hat{\boldsymbol{B}}\Big]'\boldsymbol{X}'\boldsymbol{\Omega}^{-1}(\boldsymbol{Z} - \boldsymbol{X}\hat{\boldsymbol{B}}) \\
= & -\Big[\frac{\partial}{\partial \boldsymbol{D}_{jhi}^{-1}}[\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\Big]'\boldsymbol{X}'\boldsymbol{\Omega}^{-1}(\boldsymbol{Z} - \boldsymbol{X}\hat{\boldsymbol{B}}) \\
= & -\Big[\frac{\partial}{\partial \boldsymbol{D}_{jhi}^{-1}}\boldsymbol{Z}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}[\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1}\Big]\boldsymbol{X}'\boldsymbol{\Omega}^{-1}(\boldsymbol{Z} - \boldsymbol{X}\hat{\boldsymbol{B}}) \\
= & -\boldsymbol{Z}'\Big[\frac{\partial}{\partial \boldsymbol{D}_{jhi}^{-1}}\boldsymbol{\Omega}^{-1}\Big]\boldsymbol{X}[\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}(\boldsymbol{Z} - \boldsymbol{X}\hat{\boldsymbol{B}}) \\
& -\boldsymbol{Z}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}\Big[\frac{\partial}{\partial \boldsymbol{D}_{jhi}^{-1}}[\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1}\Big]\boldsymbol{X}'\boldsymbol{\Omega}^{-1}(\boldsymbol{Z} - \boldsymbol{X}\hat{\boldsymbol{B}})
\end{aligned}$$

$$
\begin{aligned}
&= -\boldsymbol{Z}'\Big[\frac{\partial}{\partial \boldsymbol{D}_{jhi}^{-1}}\boldsymbol{\Omega}^{-1}\Big](\boldsymbol{X}\hat{\boldsymbol{B}}-\boldsymbol{X}\hat{\boldsymbol{B}}) \\
&\quad -\boldsymbol{Z}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}\Big[\frac{\partial}{\partial \boldsymbol{D}_{jhi}^{-1}}[\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1}\Big]\boldsymbol{X}'\boldsymbol{\Omega}^{-1}(\boldsymbol{Z}-\boldsymbol{X}\hat{\boldsymbol{B}}) \\
&= \boldsymbol{Z}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}[\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1}\Big[\frac{\partial}{\partial \boldsymbol{D}_{jhi}^{-1}}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}\Big][\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}(\boldsymbol{Z}-\boldsymbol{X}\hat{\boldsymbol{B}}) \\
&= \boldsymbol{Z}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}[\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1}\Big[\frac{\partial}{\partial \boldsymbol{D}_{jhi}^{-1}}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}\Big](\hat{\boldsymbol{B}}-\hat{\boldsymbol{B}}) \\
&= 0
\end{aligned}
$$

where the fourth equality holds because of Lemma 10 (iii) and the sixth equality holds because of Lemma 10 (iv). Then from Lemma 10 (iii), we have

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{D}_{jhi}^{-1}}SS(\hat{\boldsymbol{B}}) &= \Big[\frac{\partial}{\partial \boldsymbol{D}_{jhi}^{-1}}(\boldsymbol{Z}-\boldsymbol{X}\hat{\boldsymbol{B}})\Big]'\boldsymbol{\Omega}^{-1}(\boldsymbol{Z}-\boldsymbol{X}\hat{\boldsymbol{B}}) \\
&\quad +(\boldsymbol{Z}-\boldsymbol{X}\hat{\boldsymbol{B}})'\Big[\frac{\partial}{\partial \boldsymbol{D}_{jhi}^{-1}}\boldsymbol{\Omega}^{-1}\Big](\boldsymbol{Z}-\boldsymbol{X}\hat{\boldsymbol{B}}) \\
&\quad +(\boldsymbol{Z}-\boldsymbol{X}\hat{\boldsymbol{B}})'\boldsymbol{\Omega}^{-1}\Big[\frac{\partial}{\partial \boldsymbol{D}_{jhi}^{-1}}(\boldsymbol{Z}-\boldsymbol{X}\hat{\boldsymbol{B}})\Big]. \\
&= (\boldsymbol{Z}-\boldsymbol{X}\hat{\boldsymbol{B}})'\Big[\frac{\partial}{\partial \boldsymbol{D}_{jhi}^{-1}}\boldsymbol{\Omega}^{-1}\Big](\boldsymbol{Z}-\boldsymbol{X}\hat{\boldsymbol{B}}).
\end{aligned}
$$

Since $\boldsymbol{\Omega}^{-1} = \sum_{j=1}^{J}(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{D}_j^{-1} + \boldsymbol{A}_0 \otimes \boldsymbol{D}_1^{-1}$, if we define matrix $\boldsymbol{Q}_{hi}$ to be an $m \times m$ matrix with 1 as the (h,i) element and 0 otherwise, then $\frac{\partial}{\partial \boldsymbol{D}_{1hi}^{-1}}\boldsymbol{\Omega}^{-1} = \frac{\partial}{\partial \boldsymbol{D}_{1hi}^{-1}}\boldsymbol{A}_1 \otimes \boldsymbol{D}_1^{-1} = \boldsymbol{A}_1 \otimes \boldsymbol{Q}_{hi}$, and $\frac{\partial}{\partial \boldsymbol{D}_{jhi}^{-1}}\boldsymbol{\Omega}^{-1} = \frac{\partial}{\partial \boldsymbol{D}_{jhi}^{-1}}(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{D}_j^{-1} = (\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{Q}_{hi}$ for $j = 2, \ldots, J$. Hence

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{D}_{jhi}^{-1}}SS(\hat{\boldsymbol{B}}) &= \begin{cases} (\boldsymbol{Z}-\boldsymbol{X}\hat{\boldsymbol{B}})'(\boldsymbol{A}_1 \otimes \boldsymbol{Q}_{hi})(\boldsymbol{Z}-\boldsymbol{X}\hat{\boldsymbol{B}}) & \text{if } j=1 \\ (\boldsymbol{Z}-\boldsymbol{X}\hat{\boldsymbol{B}})'[(\boldsymbol{A}_j - \boldsymbol{A}_{j+1}) \otimes \boldsymbol{Q}_{hi}](\boldsymbol{Z}-\boldsymbol{X}\hat{\boldsymbol{B}}) & \text{if } = 2,\ldots,J \end{cases} \\
&= \begin{cases} \mathrm{tr}\Big[(\boldsymbol{z}-\boldsymbol{x}\hat{\boldsymbol{\beta}})'\boldsymbol{A}_1(\boldsymbol{z}-\boldsymbol{x}\hat{\boldsymbol{\beta}})\boldsymbol{Q}_{hi}\Big] & \text{if } j=1 \\ \mathrm{tr}\Big[(\boldsymbol{z}-\boldsymbol{x}\hat{\boldsymbol{\beta}})'(\boldsymbol{A}_j - \boldsymbol{A}_{j+1})(\boldsymbol{z}-\boldsymbol{x}\hat{\boldsymbol{\beta}})\boldsymbol{Q}_{hi}\Big] & \text{if } = 2,\ldots,J \end{cases} \\
&= \begin{cases} \mathrm{tr}\Big[[SS_1(\hat{\boldsymbol{\beta}}) + SS_0(\hat{\boldsymbol{\beta}})]\boldsymbol{Q}_{hi}\Big] & \text{if } j=1 \\ \mathrm{tr}\Big[SS_j(\hat{\boldsymbol{\beta}})\boldsymbol{Q}_{hi}\Big] & \text{if } = 2,\ldots,J \end{cases} \\
&= \begin{cases} \text{the } (i,h) \text{ element of } SS_1(\hat{\boldsymbol{\beta}}) + SS_0(\hat{\boldsymbol{\beta}}) & \text{if } j=1 \\ \text{the } (i,h) \text{ element of } SS_j(\hat{\boldsymbol{\beta}}) & \text{if } = 2,\ldots,J \end{cases},
\end{aligned}
$$

where the second and fourth equalities hold because of Lemma 5 (v) and (ii) respectively. □

LEMMA 12.

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

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.*

Proof.

$$\begin{aligned} & \frac{\partial \log|\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}|}{\partial \boldsymbol{D}_{jhi}^{-1}} \\ = \; & -\mathrm{tr}\Big[\boldsymbol{\Omega}^{-1}\boldsymbol{X}[\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\frac{\partial \boldsymbol{\Omega}}{\partial \boldsymbol{D}_{jhi}^{-1}}\Big] \\ = \; & \mathrm{tr}\Big[\boldsymbol{\Omega}^{-1}\boldsymbol{X}[\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Omega}(\frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \boldsymbol{D}_{jhi}^{-1}})\boldsymbol{\Omega}\Big] \\ = \; & \mathrm{tr}\Big[\boldsymbol{P}\frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \boldsymbol{D}_{jhi}^{-1}}\Big], \end{aligned}$$

where the first and second equalities hold because of Lemma 10 (v) and (iv) respectively, and the last equality holds because of Lemma 5 (iii). Since $\boldsymbol{\Omega}^{-1} = \sum_{j=1}^{J}(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{D}_j^{-1} + \boldsymbol{A}_0 \otimes \boldsymbol{D}_1^{-1}$, we have $\frac{\partial}{\partial \boldsymbol{D}_{1hi}^{-1}}\boldsymbol{\Omega}^{-1} = \frac{\partial}{\partial \boldsymbol{D}_{1hi}^{-1}}\boldsymbol{A}_1 \otimes \boldsymbol{D}_1^{-1} = \boldsymbol{A}_1 \otimes \boldsymbol{Q}_{hi}$, and $\frac{\partial}{\partial \boldsymbol{D}_{jhi}^{-1}}\boldsymbol{\Omega}^{-1} = \frac{\partial}{\partial \boldsymbol{D}_{jhi}^{-1}}(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{D}_j^{-1} = (\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{Q}_{hi}$ for $j = 2,\dots,J$. □

LEMMA 13.

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

Proof. If we use $SS_{jih}(\boldsymbol{\beta})$ to denote the (i,h) element of $SS_j(\boldsymbol{\beta})$, then for $j = 1,\ldots,J$,

$$\begin{aligned}
E(SS_{jih}(\boldsymbol{\beta})) &= E\Big[\mathrm{tr}[SS_j(\boldsymbol{\beta})\boldsymbol{Q}_{hi}]\Big] \\
&= E\Big[\mathrm{tr}[(\boldsymbol{z}-\boldsymbol{x\beta})'(\boldsymbol{A}_j-\boldsymbol{A}_{j-1})(\boldsymbol{z}-\boldsymbol{x\beta})\boldsymbol{Q}_{hi}]\Big] \\
&= E\Big[(\boldsymbol{Z}-\boldsymbol{XB})'[(\boldsymbol{A}_j-\boldsymbol{A}_{j-1})\otimes\boldsymbol{Q}_{hi}](\boldsymbol{Z}-\boldsymbol{XB})\Big] \\
&= \mathrm{tr}\Big[\boldsymbol{\Omega}[(\boldsymbol{A}_j-\boldsymbol{A}_{j-1})\otimes\boldsymbol{Q}_{hi}]\Big] \\
&= \mathrm{tr}\Big[(\boldsymbol{A}_j-\boldsymbol{A}_{j-1})\otimes(\boldsymbol{D}_j\boldsymbol{Q}_{hi})\Big] \\
&= \mathrm{tr}[\boldsymbol{A}_j-\boldsymbol{A}_{j-1}]\mathrm{tr}[\boldsymbol{D}_j\boldsymbol{Q}_{hi}] \\
&= (N_j-N_{j-1})\boldsymbol{D}_{jih},
\end{aligned}$$

where first, third, fourth and sixth equalities hold because of Lemma 5 (ii), (v), (i) and (iv) respectively, and the fifth and seventh equalities hold because of Lemma 7 (vi) and (v) respectively. Similarly we obtain $E(SS_{0ih}(\boldsymbol{\beta})) = \boldsymbol{D}_{1ih}$. Hence $E[SS_0(\boldsymbol{\beta})] = \boldsymbol{D}_1$ and $E[SS_j(\hat{\boldsymbol{\beta}})] = (N_j - N_{j_1})\boldsymbol{D}_j; \quad j = 1,\ldots,J$ If we define $\boldsymbol{H} = \boldsymbol{I} - \boldsymbol{P}\boldsymbol{\Omega}^{-1}$, then it is easy to obtain $\boldsymbol{H\Omega H}' = \boldsymbol{\Omega} - \boldsymbol{P}$ and $\boldsymbol{Z} - \boldsymbol{X}\hat{\boldsymbol{B}} = \boldsymbol{HZ}$. Hence for $j = 1,\ldots,J$

$$\begin{aligned}
E[SS_{jih}(\hat{\boldsymbol{\beta}})] &= E\Big[\mathrm{tr}[SS_j(\hat{\boldsymbol{\beta}})\boldsymbol{Q}_{hi}]\Big] \\
&= E\Big[\mathrm{tr}[(\boldsymbol{z}-\boldsymbol{x}\hat{\boldsymbol{\beta}})'(\boldsymbol{A}_j-\boldsymbol{A}_{j-1})(\boldsymbol{z}-\boldsymbol{x}\hat{\boldsymbol{\beta}})\boldsymbol{Q}_{hi}]\Big] \\
&= E\Big[(\boldsymbol{Z}-\boldsymbol{X}\hat{\boldsymbol{B}})'[(\boldsymbol{A}_j-\boldsymbol{A}_{j-1})\otimes\boldsymbol{Q}_{hi}](\boldsymbol{Z}-\boldsymbol{X}\hat{\boldsymbol{B}})\Big] \\
&= E\Big[\boldsymbol{Z}'\boldsymbol{H}'[(\boldsymbol{A}_j-\boldsymbol{A}_{j-1})\otimes\boldsymbol{Q}_{hi}]\boldsymbol{HZ}\Big] \\
&= \mathrm{tr}\Big[\boldsymbol{H}'[(\boldsymbol{A}_j-\boldsymbol{A}_{j-1})\otimes\boldsymbol{Q}_{hi}]\boldsymbol{H\Omega}\Big] + (\boldsymbol{XB})'\boldsymbol{H}[(\boldsymbol{A}_j-\boldsymbol{A}_{j-1})\otimes\boldsymbol{Q}_{hi}]\boldsymbol{HXB} \\
&= \mathrm{tr}\Big[\boldsymbol{H\Omega H}'[(\boldsymbol{A}_j-\boldsymbol{A}_{j-1})\otimes\boldsymbol{Q}_{hi}]\Big] \\
&= \mathrm{tr}\Big[(\boldsymbol{\Omega}-\boldsymbol{P})[(\boldsymbol{A}_j-\boldsymbol{A}_{j-1})\otimes\boldsymbol{Q}_{hi}]\Big] \\
&= (N_j-N_{j-1})\boldsymbol{D}_{jih} - \mathrm{tr}\Big[\boldsymbol{P}[(\boldsymbol{A}_j-\boldsymbol{A}_{j-1})\otimes\boldsymbol{Q}_{hi}]\Big]
\end{aligned}$$

where first, third and fifth equalities hold because of Lemma 5 (ii), (v) and (i) respectively, and the sixth equality holds because of $\boldsymbol{HX} = \boldsymbol{0}_{N_j\times pm}$. Similarly, we obtain $E[SS_{0ih}(\hat{\boldsymbol{\beta}})] = N_0\boldsymbol{D}_1 - \Big[\mathrm{tr}[\boldsymbol{P}(\boldsymbol{A}_0\otimes \boldsymbol{Q}_{hi})]\Big]_{ih}$. □

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* $\operatorname{tr}\Big[\boldsymbol{P}[\boldsymbol{A}_0 \otimes \boldsymbol{Q}_{hi}]\Big] = \boldsymbol{D}_{1ih}$.

Proof. $\boldsymbol{P} = \boldsymbol{X}[\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}' = \boldsymbol{X}[N_J\boldsymbol{D}_1^{-1}]^{-1}\boldsymbol{X}' = \frac{1}{N_J}(\mathbf{1}_{N_J}\otimes\boldsymbol{I}_m)(\boldsymbol{I}_1\otimes\boldsymbol{D}_1)\boldsymbol{X}(\mathbf{1}_{N_J}\otimes\boldsymbol{I}_m)' = \frac{1}{N_J}(\mathbf{1}_{N_J}\boldsymbol{I}_1\mathbf{1}'_{N_J})\otimes(\boldsymbol{I}_m\boldsymbol{D}_1\boldsymbol{I}_m) = \boldsymbol{A}_0 \otimes \boldsymbol{D}_1$ where the first equality holds because of Lemma 7 (viii).

$\operatorname{tr}\Big[\boldsymbol{P}[(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\otimes\boldsymbol{Q}_{hi}]\Big] = \operatorname{tr}\Big[[\boldsymbol{A}_0\otimes\boldsymbol{D}_1][(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\otimes\boldsymbol{Q}_{hi}]\Big] = \operatorname{tr}\Big[[\boldsymbol{A}_0(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})]\otimes[\boldsymbol{D}_1\boldsymbol{Q}_{hi}]\Big] = 0$ where the last equality holds because $\boldsymbol{A}_0(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) = \mathbf{0}_{N_J\times N_J}$ for $j = 1, \ldots, J$.

$\operatorname{tr}\Big[\boldsymbol{P}[\boldsymbol{A}_0 \otimes \boldsymbol{Q}_{hi}]\Big] = \operatorname{tr}\Big[[\boldsymbol{A}_0 \otimes \boldsymbol{D}_1][\boldsymbol{A}_0 \otimes \boldsymbol{Q}_{hi}]\Big] = \operatorname{tr}\Big[[\boldsymbol{A}_0^2] \otimes [\boldsymbol{D}_1\boldsymbol{Q}_{hi}]\Big] = \operatorname{tr}\Big[\boldsymbol{A}_0 \otimes [\boldsymbol{D}_1\boldsymbol{Q}_{hi}]\Big] = \operatorname{tr}(\boldsymbol{A}_0)\operatorname{tr}(\boldsymbol{D}_1\boldsymbol{Q}_{hi}) = \boldsymbol{D}_{1ih}$, where the fourth and the last equality holds because of Lemma 5 (iv) and (ii) respectively. □

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

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

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

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

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

Proof. (i) $\hat{\boldsymbol{B}} = [\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{Z} = [N_J\boldsymbol{D}_1^{-1}]^{-1}[\mathbf{1}_{N_J} \otimes \boldsymbol{D}_1^{-1}]'\boldsymbol{Z} = \frac{1}{N_J}(\boldsymbol{I}_1 \otimes \boldsymbol{D}_1)(\mathbf{1}'_{N_J} \otimes \boldsymbol{D}_1^{-1})\boldsymbol{Z} = \frac{1}{N_J}(\mathbf{1}'_{N_J} \otimes \boldsymbol{I}_m)\boldsymbol{Z} = \operatorname{vec}(\frac{1}{N_J}\mathbf{1}'_{N_J}\boldsymbol{z})$, 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 $\boldsymbol{x}\hat{\boldsymbol{\beta}} = \mathbf{1}_{N_J}(\frac{1}{N_J}\mathbf{1}'_{N_J})\boldsymbol{z} = \boldsymbol{A}_0\boldsymbol{z}$, 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). $\text{vec}((\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\boldsymbol{z}) = [(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{I}_m]\boldsymbol{Z} \sim N([(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{I}_m]\boldsymbol{X}\boldsymbol{B}, [(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{I}_m]'\boldsymbol{\Omega}[(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{I}_m]) \sim N(\boldsymbol{0}_{N_J m}, (\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{D}_j)$ 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 $[(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{I}_m]\boldsymbol{X}\boldsymbol{B} = [(\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{I}_m](\boldsymbol{1}_{N_J} \otimes \boldsymbol{I}_m)\boldsymbol{B} = \{[(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\boldsymbol{1}_{N_J}] \otimes \boldsymbol{I}_m\}\boldsymbol{B} = \{[\boldsymbol{1}_{N_J} - \boldsymbol{1}_{N_J}] \otimes \boldsymbol{I}_m\}\boldsymbol{B} = \boldsymbol{0}_{N_J m}$ 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 $(\boldsymbol{A}_j - \boldsymbol{A}_{j-1})\boldsymbol{z} \sim N_{N_J \times m}(\boldsymbol{0}_{N_J \times m}, (\boldsymbol{A}_j - \boldsymbol{A}_{j-1}) \otimes \boldsymbol{D}_j)$ (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 they are independent with $\hat{\boldsymbol{\beta}}$. For $1 \leq j < k \leq J$,

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

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

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

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) Using the formulas of Wishart distribution from Appendix C.2 of Lauritzen (1996). □

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

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

Proof. $\operatorname{tr}[\boldsymbol{D}_j^2] = \operatorname{tr}\Big[[(d_{j1}-d_{j2})\boldsymbol{I}_m + d_{j2}\mathbf{1}_m\mathbf{1}'_m]^2\Big] = \operatorname{tr}[(d_{j1}-d_{j2})^2\boldsymbol{I}_m + 2(d_{j1}-d_{j2})d_{j2}\mathbf{1}_m\mathbf{1}'_m + md_{j2}^2\mathbf{1}_m\mathbf{1}'_m] = m(d_{j1} - d_{j2})^2 + 2m(d_{j1} - d_{j2})d_{j2} + m^2d_{j2}^2 = m(d_{j1}^2 + (m-1)d_{j2}^2)$.

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

Appendix D: A Fast Algorithm to Compute the Log-Likelihood Function

There are various ways of ordering the vector nodes $\{j,k\}$ on a multiresolution tree structure; $k = 1,\ldots,N_{j-1}, j = 1,\ldots,J$. Here we consider an ordering proposed by Luettgen (1993). Define a function $s : \{j,k\} \mapsto s(j,k)$, where $s(j,k)$ is the order of vector node $\{j,k\}$. We start from the vector node $\{J,1\}$ and let t denote the current vector node to be ordered. The function $s(j,k)$ is defined as:

(a) $t = \{J,1\}$; $h = 1$;

(b) If $t = \{j,k\}$ and all nodes $\{\{j',k'\} : \{j',k'\} \prec \{j,k\}, \{j',k'\} \neq \{j,k\}\}$ are ordered, then:

1. $s(j,k) = h$ and $h = h+1$;
2. If $\{j,k\} = \{1,1\}$, then End.
3. If $\{j,k\}$ is the i^{th} child of $pa\{j,k\}$ and $\{j,k\}$ is not the last child (i.e. $i < n_{j-1}$), then $t = ch\{pa\{j,k\}, i+1\}$. Otherwise $t = pa\{j,k\}$.

(c) If $t = \{j,k\}$ has unordered descendants, then $t = ch\{j,k,1\}$.

(d) Go back to step (b).

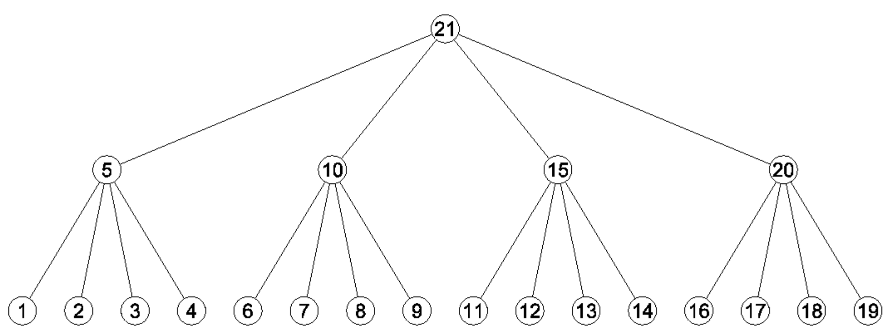

Figure 3: Complete ordering of a quad-tree

Using this order, we proceed to factorize the loglikelihood function and evaluate the individual components based on the corresponding conditional mean and conditional variance. Equivalently, we compute the optimal linear predictor of the residual process $\boldsymbol{U}_{j,k}$ given the collection of the observations on the nodes with a smaller order than that of $\{j,k\}$, by using an algorithm similar to the generalized change-of-resolution Kalman filter in Section 3. In the the high-to-low-resolution filtering step, we start with the finest resolution J and compute the optimal linear predictor of $\boldsymbol{U}_{J,k}$ given $\boldsymbol{Z}_{J,k}$ for $k=1,\ldots,N_{J-1}$:

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

As we move from the resolution $j=J-1$ to the coarsest resolution $j=1$, we compute the optimal linear predictor of $\boldsymbol{U}_{j,k}$ given $\boldsymbol{Z}_{de\{ch\{j,k,i\}\}}$ which are the observations on the subtree of the i^{th} child of $\{j,k\}$. For $k=1,\ldots,N_{j-1}$,

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

where $\boldsymbol{B}_{ch\{j,k,i\}}\equiv\boldsymbol{V}_{j,k}\boldsymbol{A}'_{ch\{j,k,i\}}\boldsymbol{V}^{+}_{ch\{j,k,i\}}$ and $\boldsymbol{R}_{ch\{j,k,i\}}\equiv\boldsymbol{V}_{j,k}-\boldsymbol{V}_{j,k}\boldsymbol{A}'_{ch\{j,k,i\}}\boldsymbol{V}^{+}_{ch\{j,k,i\}}\boldsymbol{A}_{ch\{j,k,i\}}\boldsymbol{V}_{j,k}$; $i=1,\ldots,n_{j-1}$. Further, we compute the optimal linear predictor of $\boldsymbol{U}_{j,k}$ given the combination of observations on the subtrees of the first h children of $\{j,k\}$ where $h=1,\ldots,n_{j-1}$. This depends on the ordering of the child nodes.

$$\begin{aligned}\hat{\boldsymbol{U}}^{h}_{j,k|j,k} &\equiv \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}^{h}_{de\{j,k\}}) = \hat{\boldsymbol{V}}^{h}_{j,k|j,k}\left(\sum_{i=1}^{h}\hat{\boldsymbol{V}}^{+}_{j,k|ch\{j,k,i\}}\hat{\boldsymbol{U}}_{j,k|ch\{j,k,i\}}\right),\\ \hat{\boldsymbol{V}}^{h}_{j,k|j,k} &\equiv \boldsymbol{C}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}^{h}_{de\{j,k\}}) = \left\{\boldsymbol{V}^{+}_{j,k}+\sum_{i=1}^{h}(\hat{\boldsymbol{V}}^{+}_{j,k|ch\{j,k,i\}}-\boldsymbol{V}^{+}_{j,k})\right\}^{+},\end{aligned}$$

where $\boldsymbol{Z}^{h}_{de\{j,k\}}\equiv\bigcup_{i=1}^{h}\boldsymbol{Z}_{de\{ch\{j,k,i\}\}}$ denotes the collection of the observations on the subtrees of the first h children of $\{j,k\}$; $h=1,\ldots,n_{j-1}$. It is obvious that $\boldsymbol{Z}^{n_{j-1}}_{de\{j,k\}}=\boldsymbol{Z}^{*}_{de\{j,k\}}$ for $h=n_{j-1}$. Then, for $\{j,k\}$, we compute the optimal linear predictor of $\boldsymbol{U}_{j,k}$ given all the observations on the subtree of $\{j,k\}$ including $\boldsymbol{Z}_{j,k}$.

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

At the end of the filtering step, the root vector node is reached. For $\{1,1\}$, we compute the optimal linear predictor of $\boldsymbol{U}_{1,1}$ given the combination of observations on the subtrees of its first h children where $h = 0, \dots, n_0$.

$$\hat{\boldsymbol{U}}_{\overline{1,1}} \equiv \boldsymbol{m}(\boldsymbol{U}_{1,1}|\boldsymbol{Z}_{\overline{1,1}}) = \boldsymbol{0}_{n_0 m}, \quad \hat{\boldsymbol{V}}_{\overline{1,1}} \equiv \boldsymbol{C}(\boldsymbol{U}_{1,1}|\boldsymbol{Z}_{\overline{1,1}}) = \boldsymbol{V}_{1,1}, \tag{75}$$

$$\hat{\boldsymbol{U}}^h_{\overline{1,1}} \equiv \boldsymbol{m}(\boldsymbol{U}_{1,1}|\boldsymbol{Z}_{\overline{1,1}}, \boldsymbol{Z}^h_{de\{1,1\}}) = \hat{\boldsymbol{U}}^h_{1,1|1,1}, \ \hat{\boldsymbol{V}}^h_{\overline{1,1}} \equiv \boldsymbol{C}(\boldsymbol{U}_{\overline{1,1}}|\boldsymbol{Z}_{\overline{1,1}}, \boldsymbol{Z}^h_{de\{1,1\}}) = \hat{\boldsymbol{V}}^h_{1,1|1,1}, \ h = 1, \dots, n_0 - 1$$

$$\hat{\boldsymbol{U}}^s_{1,1} \equiv \boldsymbol{m}(\boldsymbol{U}_{1,1}|\boldsymbol{Z}^s_{1,1}) = \hat{\boldsymbol{U}}^{n_0}_{1,1|1,1}, \quad \hat{\boldsymbol{V}}^s_{1,1} \equiv \boldsymbol{C}(\boldsymbol{U}_{1,1}|\boldsymbol{Z}^s_{1,1}) = \hat{\boldsymbol{V}}^{n_0}_{1,1|1,1}, \tag{76}$$

where $\boldsymbol{Z}^s_{1,1} \equiv \boldsymbol{Z} \backslash \boldsymbol{Z}_{1,1}$ denotes the collection of observations on the descendants of $\{1,1\}$ excluding $\{1,1\}$ and $\boldsymbol{Z}_{\overline{1,1}} = \emptyset$ denotes the complement of $\boldsymbol{Z}_{de\{1,1\}}$. The cases $h = 0$ and $h = n_0$ are treated as special cases in (75) and (76) respectively.

In the low-to-high-resolution smoothing step, we move from the resolution $j = 2$ to the finest resolution $j = J$. For a given node $\{j,k\}$ that is not a leaf node ($j = 2, \dots, J-1, k = 1, \dots, N_{j-1}$), if $\{j,k\}$ is the i^{th} child of $pa\{j,k\}$, we compute the optimal linear predictor of $\boldsymbol{U}_{j,k}$ given $\boldsymbol{Z}_{\overline{j,k}}$ and $\boldsymbol{Z}^h_{de\{j,k\}}$ where $h = 0, 1, \dots, n_{j-1}$.

$$\hat{\boldsymbol{U}}_{\overline{j,k}} \equiv \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}_{\overline{j,k}}) = \begin{cases} \hat{\boldsymbol{U}}_{\overline{pa\{j,k\}}} & \text{if } i = 1 \\ \hat{\boldsymbol{U}}^{i-1}_{\overline{pa\{j,k\}}} & \text{if } 1 < i \leq n_{j-2}, \end{cases} \tag{77}$$

$$\hat{\boldsymbol{V}}_{\overline{j,k}} \equiv \boldsymbol{C}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}_{\overline{j,k}}) = \begin{cases} \hat{\boldsymbol{V}}_{\overline{pa\{j,k\}}} + \mathrm{var}(\boldsymbol{W}_{j,k}) & \text{if } i = 1 \\ \hat{\boldsymbol{V}}^{i-1}_{\overline{pa\{j,k\}}} + \mathrm{var}(\boldsymbol{W}_{j,k}) & \text{if } 1 < i \leq n_{j-2}, \end{cases} \tag{78}$$

$$\hat{\boldsymbol{U}}^h_{\overline{j,k}} \equiv \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}_{\overline{j,k}}, \boldsymbol{Z}^h_{de\{j,k\}}) = \hat{\boldsymbol{V}}^h_{\overline{j,k}}[(\hat{\boldsymbol{V}}^h_{j,k|j,k})^+ \hat{\boldsymbol{U}}^h_{j,k|j,k} + \hat{\boldsymbol{V}}^+_{\overline{j,k}} \hat{\boldsymbol{U}}_{\overline{j,k}}], h = 1, \dots, n_{j-1} - 1,$$

$$\hat{\boldsymbol{V}}^h_{\overline{j,k}} \equiv \boldsymbol{C}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}_{\overline{j,k}}, \boldsymbol{Z}^h_{de\{j,k\}}) = [(\hat{\boldsymbol{V}}^h_{j,k|j,k})^+ + \hat{\boldsymbol{V}}_{\overline{j,k}} - \boldsymbol{V}^+_{j,k}]^+, h = 1, \dots, n_{j-1} - 1,$$

$$\hat{\boldsymbol{U}}^s_{j,k} \equiv \boldsymbol{m}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}^s_{j,k}) = \hat{\boldsymbol{U}}^{n_{j-1}}_{\overline{j,k}}, \tag{79}$$

$$\hat{\boldsymbol{V}}^s_{j,k} \equiv \boldsymbol{C}(\boldsymbol{U}_{j,k}|\boldsymbol{Z}^s_{j,k}) = \hat{\boldsymbol{V}}^{n_{j-1}}_{\overline{j,k}}, \tag{80}$$

where $\boldsymbol{Z}^s_{j,k} \equiv \{\boldsymbol{Z}_{j',k'} : \gamma_{j',k'} = 1, s(j',k') < s(j,k)\}$ denotes the collection of the observations on the nodes with smaller order than that of $\{j,k\}$, and $\boldsymbol{Z}_{\overline{j,k}} \equiv \boldsymbol{Z}^s_{j,k} \backslash \boldsymbol{Z}^{n_{j-1}}_{de\{j,k\}}$ denotes $\boldsymbol{Z}^s_{j,k}$ excluding the observations on the descendants of $\{j,k\}$. The cases $h = 0$ and $h = n_{j-1}$ are treated as special cases in (77)–(78) and (79)–(80) respectively. For a leaf node $\{J,k\}$ ($k = 1, \dots, N_{J-1}$), if $\{J,k\}$ is the i^{th} child of $pa\{J,k\}$, we compute the optimal linear predictor of $\boldsymbol{U}_{J,k}$ given all the nodes with small order than that of $\{J, k\}$.

$$
\begin{aligned}
\hat{\boldsymbol{U}}^{s}_{J,k} &\equiv \boldsymbol{m}(\boldsymbol{U}_{J,k} | \boldsymbol{Z}^{s}_{J,k}) = \hat{\boldsymbol{U}}_{\overline{J,k}} = \begin{cases} \hat{\boldsymbol{U}}_{\overline{pa\{J,k\}}} & \text{if } i = 1 \\ \hat{\boldsymbol{U}}^{i-1}_{\overline{pa\{J,k\}}} & \text{if } 1 < i \leq n_{J-2}, \end{cases} \\
\hat{\boldsymbol{V}}^{s}_{J,k} &\equiv \boldsymbol{C}(\boldsymbol{U}_{J,k} | \boldsymbol{Z}^{s}_{J,k}) = \hat{\boldsymbol{V}}_{\overline{J,k}} = \begin{cases} \hat{\boldsymbol{V}}_{\overline{pa\{J,k\}}} + \operatorname{var}(\boldsymbol{W}_{J,k}) & \text{if } i = 1 \\ \hat{\boldsymbol{V}}^{i-1}_{\overline{pa\{J,k\}}} + \operatorname{var}(\boldsymbol{W}_{J,k}) & \text{if } 1 < i \leq n_{J-2}. \end{cases}
\end{aligned}
$$

Appendix E: 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 quadtree (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 $N_1 = 8, 16, 32, 64$, which correspond to data size $N = 512, 1024, 2048, 4096$. 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 2–9. For each of the 8 cases (4 sample sizes and 2 types of regression means), we simulate $S = 1000$ 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 $S = 1000$ 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 then divided by the true parameter value.

For the constant-mean MMTSLM, the results are shown in Tables 2, 4, 6, and 8. 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.

For the regression-mean MMTSLM, the results are shown in Tables 3, 5, 7 and 9. Here we do not have theoretical results to compare to, but we could 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 consequent of a large bias in the estimate of σ_{11}. Finally, there seems very little difference between the MLE and REML estimates for the sample sizes under consideration.

		Theory			Formula			Maximization		
MLE	Truth	R-Bias	Variance	MSE	R-Bias	Variance	MSE	R-Bias	Variance	MSE
σ_{11}	200.0	-0.13	2997.83	3627.72	-0.10	2834.80	3242.64	-0.29	1266.11	4730.67
σ_{12}	-20.0	-0.13	1220.69	1226.94	-0.47	1313.39	1402.09	-0.61	633.52	782.34
σ_{21}	100.0	0.00	296.51	296.51	0.02	308.90	313.69	0.03	417.16	423.14
σ_{22}	-10.0	0.00	121.21	121.21	-0.14	125.17	127.03	-0.05	190.97	191.03
σ_{31}	50.0	0.00	24.83	24.83	0.02	24.36	25.37	0.03	46.15	48.41
σ_{32}	5.0	0.00	14.56	14.56	0.09	12.98	13.16	-0.00	77.85	77.77
σ_{41}	25.0	0.00	6.80	6.80	0.02	6.53	6.72	0.02	25.12	25.45
σ_{42}	2.5	0.00	3.68	3.68	0.06	3.43	3.44	0.06	18.29	18.29
β_{11}	40.0	0.00	25.10	25.10	0.00	26.79	26.77	0.00	26.99	26.97
β_{12}	20.0	0.00	25.10	25.10	0.01	26.41	26.40	0.01	26.65	26.64
β_{13}	10.0	0.00	25.10	25.10	0.04	23.69	23.83	0.04	23.83	23.97
REML	Truth	R-Bias	Variance	MSE	R-Bias	Variance	MSE	R-Bias	Variance	MSE
σ_{11}	200.0	0.00	3915.53	3915.53	0.03	3702.60	3729.40	-0.28	1472.32	4659.92
σ_{12}	-20.0	0.00	1594.37	1594.37	-0.40	1715.45	1777.50	-0.61	745.73	892.70
σ_{21}	100.0	0.00	296.51	296.51	0.02	308.90	313.69	0.03	423.00	431.66
σ_{22}	-10.0	0.00	121.21	121.21	-0.14	125.17	127.03	-0.11	181.90	182.86
σ_{31}	50.0	0.00	24.83	24.83	0.02	24.36	25.37	0.04	35.99	39.26
σ_{32}	5.0	0.00	14.56	14.56	0.09	12.98	13.16	0.07	37.65	37.74
σ_{41}	25.0	0.00	6.80	6.80	0.02	6.53	6.72	0.02	20.23	20.52
σ_{42}	2.5	0.00	3.68	3.68	0.06	3.43	3.44	-0.01	20.75	20.73
β_{11}	40.0	0.00	25.10	25.10	0.00	26.79	26.77	0.00	26.82	26.79
β_{12}	20.0	0.00	25.10	25.10	0.01	26.41	26.40	0.01	26.49	26.47
β_{13}	10.0	0.00	25.10	25.10	0.04	23.69	23.82	0.04	23.77	23.91

Table 2: Maximum likelihood estimates (MLE) and restricted maximum likelihood estimates (REML) for a multivariate multiresolution tree-structured spatial linear model (MMTSLM) with a 8-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 1000 MLE and REML estimates computed by analytical formulas and numerical maximization.

		MLE			REML		
	Truth	R-Bias	Variance	MSE	R-Bias	Variance	MSE
σ_{11}	200.0	-0.32	1394.77	5563.52	-0.32	1322.59	5499.65
σ_{12}	-20.0	-0.63	689.84	846.48	-0.62	641.97	795.37
σ_{21}	100.0	0.02	1238.44	1243.61	0.04	1236.82	1249.05
σ_{22}	-10.0	-0.27	1020.55	1026.97	-0.31	1020.04	1028.48
σ_{31}	50.0	0.09	44.87	64.67	0.11	40.42	69.71
σ_{32}	5.0	-0.01	65.52	65.45	0.02	52.59	52.54
σ_{41}	25.0	0.00	20.11	20.10	0.01	19.90	19.95
σ_{42}	2.5	0.02	21.76	21.74	0.04	20.61	20.59
β_{11}	100.0	0.00	45.05	45.01	0.00	42.34	42.30
β_{12}	50.0	-0.01	44.92	45.05	-0.01	41.47	41.57
β_{13}	25.0	-0.01	44.47	44.48	-0.01	41.38	41.36
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	0.01	5.01	5.01	0.01	4.34	4.35
β_{42}	20.0	0.00	5.21	5.21	0.00	4.23	4.23
β_{43}	40.0	0.00	5.03	5.03	0.00	4.14	4.14

Table 3: Maximum likelihood estimates (MLE) and restricted maximum likelihood estimates (REML) for a multivariate multiresolution tree-structured spatial linear model (MMTSLM) with a 8-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 1000 MLE and REML estimates computed by 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 4: 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 1000 MLE and REML estimates computed by analytical formulas and 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

Table 5: 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 1000 MLE and REML estimates computed by numerical maximization.

		Theory			Formula			Maximization		
MLE	Truth	R-Bias	Variance	MSE	R-Bias	Variance	MSE	R-Bias	Variance	MSE
σ_{11}	200.0	-0.03	829.76	869.13	-0.05	878.08	962.18	-0.18	1098.73	2344.40
σ_{12}	-20.0	-0.03	337.87	338.26	-0.09	352.96	356.01	-0.28	236.57	268.70
σ_{21}	100.0	0.00	74.13	74.13	0.00	77.85	77.85	0.01	189.87	190.32
σ_{22}	-10.0	0.00	30.30	30.30	0.02	31.42	31.43	0.05	43.03	43.22
σ_{31}	50.0	0.00	6.21	6.21	0.00	6.55	6.55	0.01	11.89	11.96
σ_{32}	5.0	0.00	3.64	3.64	0.02	3.62	3.63	0.04	5.64	5.67
σ_{41}	25.0	0.00	1.70	1.70	0.01	1.89	1.92	0.01	2.56	2.60
σ_{42}	2.5	0.00	0.92	0.92	-0.01	0.99	0.99	-0.01	1.12	1.12
β_{11}	40.0	0.00	6.27	6.27	0.00	6.42	6.42	0.00	6.41	6.41
β_{12}	20.0	0.00	6.27	6.27	-0.00	7.18	7.18	-0.00	7.21	7.21
β_{13}	10.0	0.00	6.27	6.27	0.01	6.88	6.88	0.01	6.92	6.92
REML	Truth	R-Bias	Variance	MSE	R-Bias	Variance	MSE	R-Bias	Variance	MSE
σ_{11}	200.0	0.00	884.15	884.15	-0.02	935.64	943.94	-0.18	1079.71	2311.98
σ_{12}	-20.0	0.00	360.02	360.02	-0.06	376.10	377.31	-0.28	256.29	288.06
σ_{21}	100.0	0.00	74.13	74.13	0.00	77.85	77.85	0.01	189.11	189.75
σ_{22}	-10.0	0.00	30.30	30.30	0.020	31.42	31.43	0.06	53.55	53.90
σ_{31}	50.0	0.00	6.21	6.21	0.00	6.55	6.55	0.01	11.22	11.38
σ_{32}	5.0	0.00	3.64	3.64	0.02	3.62	3.63	0.01	12.78	12.77
σ_{41}	25.0	0.00	1.70	1.70	0.01	1.89	1.92	0.01	3.66	3.68
σ_{42}	2.5	0.00	0.92	0.92	-0.01	0.99	0.99	-0.01	1.16	1.16
β_{11}	40.0	0.00	6.27	6.27	0.00	6.42	6.42	0.00	6.41	6.40
β_{12}	20.0	0.00	6.27	6.27	-0.00	7.18	7.18	-0.00	7.21	7.20
β_{13}	10.0	0.00	6.27	6.27	0.01	6.86	6.88	0.01	6.89	6.89

Table 6: Maximum likelihood estimates (MLE) and restricted maximum likelihood estimates (REML) for a multivariate multiresolution tree-structured spatial linear model (MMTSLM) with a 32-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 1000 MLE and REML estimates computed by analytical formulas and numerical maximization.

		MLE			REML		
	Truth	R-Bias	Variance	MSE	R-Bias	Variance	MSE
σ_{11}	200.0	-0.24	1148.70	3380.20	-0.23	1194.32	3345.25
σ_{12}	-20.0	-0.39	256.58	318.31	-0.39	250.61	311.95
σ_{21}	100.0	0.02	186.79	189.23	0.02	191.80	196.03
σ_{22}	-10.0	0.04	53.39	53.50	0.04	45.34	45.45
σ_{31}	50.0	0.03	10.80	12.76	0.03	10.53	13.40
σ_{32}	5.0	0.01	17.02	17.01	0.04	5.05	5.10
σ_{41}	25.0	0.00	3.35	3.36	0.00	1.79	1.80
σ_{42}	2.5	-0.01	1.70	1.70	0.00	1.17	1.17
β_{11}	100.0	0.00	10.67	10.66	0.00	10.62	10.61
β_{12}	50.0	0.00	9.55	9.54	0.00	9.57	9.56
β_{13}	25.0	0.00	10.50	10.50	0.00	10.49	10.50
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	0.71	0.71	0.00	0.71	0.71
β_{42}	20.0	0.00	0.66	0.66	0.00	0.65	0.65
β_{43}	40.0	0.00	0.66	0.66	0.00	0.66	0.66

Table 7: Maximum likelihood estimates (MLE) and restricted maximum likelihood estimates (REML) for a multivariate multiresolution tree-structured spatial linear model (MMTSLM) with a 32-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 1000 MLE and REML estimates computed by 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 8: 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 1000 MLE and REML estimates computed by analytical formulas and 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

Table 9: 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 1000 MLE and REML estimates computed by numerical maximization.

Miscellaneous

LEMMA 17. *With* $\boldsymbol{X} = \mathbf{1}_{N_J} \otimes \boldsymbol{I}_m$ *and* $\boldsymbol{D}_j; j = 1, \ldots, J$ *have compound symmetry variance structure, we have* $\mathrm{var}(\hat{\boldsymbol{\beta}}_{ji}) = var(\hat{\boldsymbol{\beta}}_{jk})$ *where* $\hat{\boldsymbol{\beta}}_{ji}$ *is the* (j, i) *element of the* $p \times m$ *matrix* $\hat{\boldsymbol{\beta}}$*,* $j = 1, \ldots, p$*,* $i,\ k = 1, \ldots, m$*.*

Proof. We have $\boldsymbol{z} = \boldsymbol{x\beta} + \boldsymbol{u} + \boldsymbol{\epsilon}$. Use $\boldsymbol{z}^{(ik)}$ to denote the matrix created by switching the i^{th} and k^{th} columns of $\boldsymbol{z}$, then $\boldsymbol{z}^{(ik)} = \boldsymbol{x\beta}^{(ik)} + \boldsymbol{u}^{(ik)} + \boldsymbol{\epsilon}^{(ik)}$. From formula (34), we have $\mathrm{var}(\mathrm{vec}(\hat{\boldsymbol{\beta}})) = [\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1}$. Since compound symmetric $\boldsymbol{D}_j, j = 1, \ldots, J$ don't depend on the order of the response variables, we have $\boldsymbol{\Omega} \equiv \mathrm{var}(\mathrm{vec}(\boldsymbol{z})) = \mathrm{var}(\mathrm{vec}(\boldsymbol{z}^{(ik)}))$ and $\mathrm{var}(\mathrm{vec}(\hat{\boldsymbol{\beta}}^{(ik)})) = [\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1}$. Hence $\mathrm{var}(\hat{\boldsymbol{\beta}}_{ji}) = \mathrm{var}(\mathrm{vec}(\hat{\boldsymbol{\beta}})_{jm+i}) = \mathrm{var}(\mathrm{vec}(\hat{\boldsymbol{\beta}}^{(ik)})_{jm+i}) = \mathrm{var}(\hat{\boldsymbol{\beta}}^{(ik)}_{ji}) = \mathrm{var}(\hat{\boldsymbol{\beta}}_{jk})$, where $\mathrm{vec}(\hat{\boldsymbol{\beta}})_{jm+i}$ is the $jm + i^{th}$ element of $\mathrm{vec}(\hat{\boldsymbol{\beta}})$. □

For the cell sizes of 4-resolution tree, at the finest resolution, the average cell size is 6278 square km (80*80). The size of each of the 32 subregion is 401843 (640*640). There are 1361 observations available on the finest resolution. Since we have to treat the whole vector node as missing if it contains missing scalar nodes, only 1248 observations are used to fit the MMTSLM.