MODELING AND PREDICTING SPATIO-TEMPORAL DYNAMICS OF $PM_{2.5}$ CONCENTRATIONS THROUGH TIME-EVOLVING COVARIANCE MODELS

Ghulam A. Qadir[1] and Ying Sun[2]

[1] *Heidelberg Institute for Theoretical Studies*

[2] *King Abdullah University of Science and Technology*

Supplementary Material

This file includes the proof of Theorem 1, a detailed explanation of RCL estimation, an examination of model properties, additional insights from the simulation study, and a discussion on the selection of polynomial orders.

S1 Proof of Theorem 1

The proof of the theorem is based on considering spatio-temporal process as a multivariate spatial process with multivariate Matérn covariance model Apanasovich, Genton, and Sun (2012) and providing a valid reparameterization of a particular case of multivariate Matérn model to include temporal components. Let us consider a stationary multivariate process

$\mathbf{Y}(\mathbf{s}) = \{Y_1(\mathbf{s}), \ldots, Y_p(\mathbf{s})\}^{\mathrm{T}},\ \mathbf{s} \in \mathbb{R}^d$, with mutlivariate Matérn covariance model:

$$\mathrm{Cov}\{Y_i(\mathbf{s}), Y_j(\mathbf{s}+\mathbf{h})\} = \mathrm{C}_{ij}(\mathbf{h}) = \rho_{ij}\sigma_i\sigma_j \mathrm{M}(\mathbf{h} \mid \alpha_{ij}, \nu_{ij}), \tag{S1.1}$$

where the validity conditions on the model parameters $\rho_{ij}, \sigma_i, \alpha_{ij}$ and $\nu_{ij},\ i,j = 1, \ldots, p,\ p \geq 1$, are provided in Theorem 1 of Apanasovich et al. (2012). In particular, we consider the following model version derived from Corollary 1(b) of Apanasovich et al. (2012):

$$\mathrm{C}_{ij}(\mathbf{h}) = \frac{\beta_{ij}\sigma_i\sigma_j\alpha_{ij}^d\Gamma(\frac{\nu_i+\nu_j}{2})}{\alpha_{ii}^{d/2}\alpha_{jj}^{d/2}\sqrt{\Gamma(\nu_i)\Gamma(\nu_j)}}\mathrm{M}\{\mathbf{h} \mid \alpha_{ij}, (\nu_i+\nu_j)/2\}, \nu_i, \sigma_i > 0,\ i = 1, \ldots, p, \tag{S1.2}$$

which is valid if: (1) $(\beta_{ij})_{i,j=1}^p$ forms a nonnegative definite matrix and (2) $(-\alpha_{ij}^{-2})_{i,j}^p$ form a conditional nonnegative definite matrix. Now, let $\beta_{ij} = 1,\ i,j = 1, \ldots, p$, and $\sigma_i = \sigma > 0,\ i = 1, \ldots, p$, in (S1.2), we get:

$$\mathrm{C}_{ij}(\mathbf{h}) = \sigma^2 \frac{\alpha_{ij}^d\Gamma(\frac{\nu_i+\nu_j}{2})}{\alpha_{ii}^{d/2}\alpha_{jj}^{d/2}\sqrt{\Gamma(\nu_i)\Gamma(\nu_j)}}\mathrm{M}\{\mathbf{h} \mid \alpha_{ij}, (\nu_i+\nu_j)/2\}, \tag{S1.3}$$

which is valid if $(-\alpha_{ij}^{-2})_{i,j}^p$ forms a conditional nonnegative definite matrix.

Now, let us consider a spatio-temporal process $Y(\mathbf{s}, t), \mathbf{s} \in \mathbb{R}^d,\ t \in \mathbb{R}$ such that $Y(\mathbf{s}, t_i) = Y_i(\mathbf{s})$ for any arbitrary time-point t_i. Also, let $\zeta(t_i, t_j)$ be any positive valued function of time-pairs t_i, t_j. Corresponding adaptation of notations in (S1.3), i.e. $\mathrm{C}_{ij}(\mathbf{h}) = \mathrm{C}(\mathbf{h}, t_i, t_j), \alpha_{ij} = \zeta(t_i, t_j), \nu_i =$

$\nu_s(t_i)$, leads to the following covariance function:

$$\begin{aligned}\mathrm{C}(\mathbf{h}, t_i, t_j) = \sigma^2 \frac{\zeta(t_i,t_j)^d \Gamma\{\frac{\nu_s(t_i)+\nu_s(t_j)}{2}\}}{\zeta(t_i,t_i)^{d/2}\zeta(t_j,t_j)^{d/2}\sqrt{\Gamma\{\nu_s(t_i)\Gamma(\nu_s(t_j)\}}} \times \\ \mathrm{M}\{\mathbf{h} \mid \zeta(t_i,t_j), \frac{\nu_s(t_i)+\nu_s(t_j)}{2}\},\end{aligned} \tag{S1.4}$$

which is valid if $\nu_s(t) > 0,\ t \in \mathbb{R}$, and $-1/\zeta(t_i,t_j)^2$ forms a conditionally nonnegative definite matrix for all $t_i, t_j \in \mathbb{R}$.

Now, since $-\frac{1}{\zeta(t_i,t_j)^2}$ needs to form conditionally nonnegative definite matrix for all $t_i, t_j \in \mathbb{R}$, it equivalently means $\frac{1}{\zeta(t_i,t_j)^2}$ needs to form conditionally negative definite (cnd) matrix for all $t_i, t_j \in \mathbb{R}$. Therefore, we can use positive Bernstein functions $\psi(w) > 0, w \geq 0$, to parameterize $\frac{1}{\zeta(t_i,t_j)^2}$. We let $\frac{1}{\zeta(t_i,t_j)^2} = \{\frac{\psi(|t_i-t_j|^2)}{\overline{\alpha_s}^2} + \frac{1/\alpha_s^2(t_i)+1/\alpha_s^2(t_j)}{2} - \frac{\psi(0)}{\overline{\alpha_s}^2}\},\ \overline{\alpha_s} > 0, \alpha_s(t) > 0, t \in \mathbb{R}$. To prove that the aforementioned parameterization is a valid parameterization, we need to show that $\{\frac{\psi(|t_i-t_j|^2)}{\overline{\alpha_s}^2} + \frac{1/\alpha_s^2(t_i)+1/\alpha_s^2(t_j)}{2} - \frac{\psi(0)}{\overline{\alpha_s}^2}\},\ \overline{\alpha_s} > 0, \alpha_s(t) > 0, t \in \mathbb{R}$ is conditionally negative definite.

As per (Bhatia and Jain, 2015, .S2), there is a one-to-one relation between Bernstein functions and cnd functions, i.e.,"A function $\psi(\cdot)$ on $(0,\infty)$ is a Bernstein function if and only if the function $f(w) = \psi(\|w\|^2)$ is continuous and cnd on $\mathbb{R}^d$ for every $d \geq 1$. Therefore, $\psi(|t_i - t_j|^2)$ is a cnd function.

Now, to show the conditional negative definiteness of $\frac{1/\alpha_s^2(t_i)+1/\alpha_s^2(t_j)}{2}$, let $x_i \in \mathbb{C}$, such that $\sum_i x_i = 0$, then,

$$\sum_i \sum_j x_i \frac{1/\alpha_s^2(t_i) + 1/\alpha_s^2(t_j)}{2} x_j^*$$

$$= \frac{1}{2} \sum_i x_i \frac{1}{\alpha_s^2(t_i)} \sum_j x_j^* + \frac{1}{2} \sum_j x_j^* \frac{1}{\alpha_s^2(t_j)} \sum_i x_i = 0$$

Therefore, $\frac{1/\alpha_s^2(t_i)+1/\alpha_s^2(t_j)}{2}$ always forms conditionally negative definite matrix. Additionally, when $\sum_i x_i = 0$, $\sum_i \sum_j x_i \psi(0) x_j^* = \psi(0) \sum_i x_i \sum_j x_j^* = 0$. Now combining all the three term, we get

$$\sum_i \sum_j x_i \frac{1}{\zeta(t_i, t_j)^2} x_j^* = \sum_i \sum_j x_i \{ \frac{\psi(|t_i - t_j|^2)}{\overline{\alpha_s}^2} + \frac{1/\alpha_s^2(t_i) + 1/\alpha_s^2(t_j)}{2} - \frac{\psi(0)}{\overline{\alpha_s}^2} \} x_j^*$$

$$= \Bigg[\sum_i \sum_j x_i \frac{\psi(|t_i - t_j|^2)}{\overline{\alpha_s}^2} x_j^* + \sum_i \sum_j x_i \frac{1/\alpha_s^2(t_i) + 1/\alpha_s^2(t_j)}{2} x_j^*$$

$$- \sum_i \sum_j x_i \frac{\psi(0)}{\overline{\alpha_s}^2} x_j^* \Bigg] \leq 0.$$

Therefore, $\frac{1}{\zeta(t_i,t_j)^2} = \{ \frac{\psi(|t_i - t_j|^2)}{\overline{\alpha_s}^2} + \frac{1/\alpha_s^2(t_i)+1/\alpha_s^2(t_j)}{2} - \frac{\psi(0)}{\overline{\alpha_s}^2} \}$, $\overline{\alpha_s} > 0, \alpha_s(t) > 0, t \in \mathbb{R}$ is a valid parametrization. Consequently, letting $\frac{1}{\zeta(t_i,t_j)^2} = \{ \frac{\psi(|t_i - t_j|^2)}{\overline{\alpha_s}^2} + \frac{1/\alpha_s^2(t_i)+1/\alpha_s^2(t_j)}{2} - \frac{\psi(0)}{\overline{\alpha_s}^2} \}$, $\overline{\alpha_s} > 0, \alpha_s(t) > 0, t \in \mathbb{R}$ in (S1.4) proves Theorem 1. Note that, if we replace the time-varying functions $\alpha_s(t) > 0$ and $\nu_s(t) > 0$ with space-time varying functions $\alpha_s(\mathbf{s}, t) > 0$ and $\nu_s(\mathbf{s}, t) > 0$, respectively, the parameterization for $\frac{1}{\zeta(t_i,t_j)^2}$ would still be valid and the resulting space-time covariance would be nonstationary both in space and time.

S2 Random Composite Likelihood Estimation

Let $X(\mathbf{s},t), \mathbf{s} \in \mathbb{R}^d, t \in \mathbb{R}$ be a zero mean Gaussian spatio-temporal process, and $\boldsymbol{X}_{\mathcal{S},\mathcal{T}}$ denote the vector of the process X, observed at the set of locations $\mathcal{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_{ns}\} \subset \mathbb{R}^d$, $ns \geq 1$, and the set of time-points $\mathcal{T} = \{t_1, \ldots, t_{nt}\} \subset \mathbb{R}$, $nt \geq 1$, i.e., $\boldsymbol{X}_{\mathcal{S},\mathcal{T}} = \{X(\mathbf{s},t); \mathbf{s} \in \mathcal{S}, t \in \mathcal{T}\}$. The total number of data points is denoted as $N = ns \cdot nt$. The log-likelihood function for $\boldsymbol{X}_{\mathcal{S},\mathcal{T}}$ is given as: $\ell(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S},\mathcal{T}}) = -\{\log\det\Sigma(\boldsymbol{\theta}) + \boldsymbol{X}_{\mathcal{S},\mathcal{T}}^{\mathrm{T}}\Sigma(\boldsymbol{\theta})^{-1}\boldsymbol{X}_{\mathcal{S},\mathcal{T}} + N.\log 2\pi\}/2$, where $\Sigma(\boldsymbol{\theta})$ is the $N \times N$ covariance matrix for $\boldsymbol{X}_{\mathcal{S},\mathcal{T}}$, defined through a spatio-temporal covariance function which depends on the set of parameters $\boldsymbol{\theta}$. The maximum likelihood estimation of $\boldsymbol{\theta}$ requires computing: $\hat{\boldsymbol{\theta}}_{ML} = \mathrm{argmax}_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S}\mathcal{T}})$, generally done through numerical optimization routines which involve iterative evaluation of $\ell(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S}\mathcal{T}})$. The optimization becomes computationally challenging in case both or either of ns and nt are large, as $\Sigma(\boldsymbol{\theta})$ then becomes a large covariance matrix and the iterative evaluation of $\ell(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S},\mathcal{T}})$ becomes time-prohibitive. Additionally, storing an extremely large sized covariance matrix $\Sigma(\boldsymbol{\theta})$ can exhaust the available memory of the machine, thus making the optimization impracticable. A widely used approximate solution to curtail this computational issue is to adopt composite likelihood methods (Vecchia, 1988; Stein et al., 2004; Varin et al., 2011; Eidsvik et al., 2014), in which the optimization is carried out over the product of component likelihoods.

In this work, we too implement the estimation by the means of composite likelihood where the collection of component likelihoods is chosen randomly. Specifically, we randomly create equisized subsets $\mathcal{S}_{ij} \subset \mathcal{S},\ i = 1, \ldots, R_s,\ j = 1, \ldots, M_S$, and $\mathcal{T}_{ij} \subset \mathcal{T},\ i = 1, \ldots, R_t,\ j = 1, \ldots, M_t$, such that for each $i = 1, \ldots, R_s$: $\cup_{j=1}^{M_s} \mathcal{S}_{ij} = \mathcal{S},\ \mathcal{S}_{ik} \cap \mathcal{S}_{il} = \phi, \forall\ k \neq l$, and for each $i = 1, \ldots, R_t$: $\cup_{j=1}^{M_t} \mathcal{T}_{ij} = \mathcal{T},\ \mathcal{T}_{ik} \cap \mathcal{T}_{il} = \phi, \forall\ k \neq l$. Here, M_s and M_t govern the size of subsets of $\mathcal{S}$ and $\mathcal{T}$, respectively, whereas R_s and R_t denote the number of randomly created mutually exclusive and exhaustive partitions of $\mathcal{S}$ and $\mathcal{T}$, respectively. Based on those subsets, we define the following random composite log-likelihood (RCL) function:

$$\ell_{RC}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S},\mathcal{T}}) = \frac{\sum_{i=1}^{R_s} \sum_{j=1}^{M_s} \ell(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S}_{ij},\mathcal{T}})}{2} + \frac{\sum_{i=1}^{R_t} \sum_{j=1}^{M_t} \ell(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S},\mathcal{T}_{ij}})}{2}, \tag{S2.1}$$

and the RCL estimate of $\boldsymbol{\theta}$ is then obtained as $\hat{\boldsymbol{\theta}}_{RCL} = \mathrm{argmax}_{\boldsymbol{\theta}}\, \ell_{RC}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S}\mathcal{T}})$. For large spatio-temporal datasets, computation and optimization of $\ell_{RC}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S}\mathcal{T}})$ is relatively more feasible than that of $\ell(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S}\mathcal{T}})$ as the former includes smaller-sized covariance matrices because the component log-likelihoods are based only on the subset of the data. Additionally, $\ell_{RC}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S}\mathcal{T}})$ can also easily utilize the parallel architecture of modern machines to simultaneously compute the component log-likelihoods, which would lead to further computational speed up. The functions $\ell_{RC}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S}\mathcal{T}})$ and $\ell(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S}\mathcal{T}})$ become increasingly similar for smaller values M_s and M_t, therefore, smaller M_s and M_t leads to more accurate but slower estimation. Note that if $M_s = M_t = 1$, then $\ell_{RC}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S}\mathcal{T}}) = \ell(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S}\mathcal{T}})$ as $M_s = 1 \implies R_s = 1$ and $M_t = 1 \implies R_t = 1$. Therefore, the values of M_s, M_t, R_s and R_t should be chosen by considering the trade-off between accuracy and speed. For our simulation study, we specify $M_s = 20, R_s = 15, M_t = 19$ and $R_t = 1$, whereas, for our data application, we specify $M_s = 69, R_s = 2, M_t = 52$ and $R_t = 1$.

In the following subsection, we now provide an exposition on the properties of $\ell_{RC}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S},\mathcal{T}})$ wherein, we prove that the random composite likelihood score function is always an unbiased estimating function for $\boldsymbol{\theta}$, i.e., $\mathbb{E}\{\frac{\partial \ell_{RC}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S},\mathcal{T}})}{\partial \theta_r}\} = 0$. In addition, we have also included the evaluation for the Hessian of $\ell_{RC}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S},\mathcal{T}})$ and the variance of $\hat{\boldsymbol{\theta}}_{RCL}$.

S2.1 Score and Hessian for RCL

Let $\Sigma_{\mathcal{S}_{ij},\mathcal{T}}$ and $\Sigma_{\mathcal{S},\mathcal{T}_{ij}}$ denote the covariance matrices (that depends on the parameters $\boldsymbol{\theta}$) for $\boldsymbol{X}_{\mathcal{S}_{ij},\mathcal{T}}$ and $\boldsymbol{X}_{\mathcal{S},\mathcal{T}_{ij}}$, respectively.

Under zero-mean Gaussianity, we have (ignoring the scalar terms that do not contain $\boldsymbol{\theta}$):

$$\ell_{RC}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S},\mathcal{T}}) = \Big[\sum_{i=1}^{R_s} \sum_{j=1}^{M_s} \{ -\frac{1}{2} \log(|\Sigma_{\mathcal{S}_{ij},\mathcal{T}}|) - \frac{1}{2} \boldsymbol{X}^T_{\mathcal{S}_{ij},\mathcal{T}} \Sigma^{-1}_{\mathcal{S}_{ij},\mathcal{T}} \boldsymbol{X}_{\mathcal{S}_{ij},\mathcal{T}} \} + \sum_{i=1}^{R_t} \sum_{j=1}^{M_t} \{ -\frac{1}{2} \log(|\Sigma_{\mathcal{S},\mathcal{T}_{ij}}|) - \frac{1}{2} \boldsymbol{X}^T_{\mathcal{S},\mathcal{T}_{ij}} \Sigma^{-1}_{\mathcal{S},\mathcal{T}_{ij}} \boldsymbol{X}_{\mathcal{S},\mathcal{T}_{ij}} \} \Big] \times \frac{1}{2} \tag{S2.2}$$

Let θ_r denote the r^{th} entry of the parameter vector $\boldsymbol{\theta}$, then we differentiate (S2.2) with respect to θ_r to obtain the score function:

$$\frac{\partial \ell_{RC}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S},\mathcal{T}})}{\partial \theta_r} = \frac{1}{2} \times \Big[\sum_{i=1}^{R_s} \sum_{j=1}^{M_s} \{ -\frac{1}{2} \frac{\partial \log(|\Sigma_{\mathcal{S}_{ij},\mathcal{T}}|)}{\partial \theta_r} - \frac{1}{2} \frac{\partial \boldsymbol{X}^T_{\mathcal{S}_{ij},\mathcal{T}} \Sigma^{-1}_{\mathcal{S}_{ij},\mathcal{T}} \boldsymbol{X}_{\mathcal{S}_{ij},\mathcal{T}}}{\partial \theta_r} \} + \sum_{i=1}^{R_t} \sum_{j=1}^{M_t} \{ -\frac{1}{2} \frac{\partial \log(|\Sigma_{\mathcal{S},\mathcal{T}_{ij}}|)}{\partial \theta_r} - \frac{1}{2} \frac{\partial \boldsymbol{X}^T_{\mathcal{S},\mathcal{T}_{ij}} \Sigma^{-1}_{\mathcal{S},\mathcal{T}_{ij}} \boldsymbol{X}_{\mathcal{S},\mathcal{T}_{ij}}}{\partial \theta_r} \} \Big] \tag{S2.3}$$

In what follows, we will make use of the following formulas : (a) $\frac{\partial \log(|\Sigma|)}{\partial \theta_r} = \text{trace}(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_r})$, (b) $\frac{\partial \mathbf{Y}^T \Sigma^{-1} \mathbf{Y}}{\partial \theta_r} = -\mathbf{Y}^T \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_r} \Sigma^{-1} \mathbf{Y}$ and (c) $\mathbb{E}(\mathbf{Y}^T B \mathbf{Y})$ $= \text{trace}(B \Sigma_Y)$, where Σ_Y is the covariance matrix for $\mathbf{Y}$. Using (a) and (b) in (S2.3), we get:

$$\begin{aligned}\frac{\partial \ell_{RC}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S},\mathcal{T}})}{\partial \theta_r} = \frac{1}{2} \times \Big[&\sum_{i=1}^{R_s} \sum_{j=1}^{M_s} \{-\frac{1}{2}\text{trace}(\Sigma_{\mathcal{S}_{ij},\mathcal{T}}^{-1} \frac{\partial \Sigma_{\mathcal{S}_{ij},\mathcal{T}}}{\partial \theta_r}) \\ &+\frac{1}{2}\boldsymbol{X}_{\mathcal{S}_{ij},\mathcal{T}}^T \Sigma_{\mathcal{S}_{ij},\mathcal{T}}^{-1} \frac{\partial \Sigma_{\mathcal{S}_{ij},\mathcal{T}}}{\partial \theta_r} \Sigma_{\mathcal{S}_{ij},\mathcal{T}}^{-1} \boldsymbol{X}_{\mathcal{S}_{ij},\mathcal{T}}\} + \\ &\sum_{i=1}^{R_t} \sum_{j=1}^{M_t} \{-\frac{1}{2}\text{trace}(\Sigma_{\mathcal{S},\mathcal{T}_{ij}}^{-1} \frac{\partial \Sigma_{\mathcal{S},\mathcal{T}_{ij}}}{\partial \theta_r}) + \\ &\frac{1}{2}\boldsymbol{X}_{\mathcal{S},\mathcal{T}_{ij}}^T \Sigma_{\mathcal{S},\mathcal{T}_{ij}}^{-1} \frac{\partial \Sigma_{\mathcal{S},\mathcal{T}_{ij}}}{\partial \theta_r} \Sigma_{\mathcal{S},\mathcal{T}_{ij}}^{-1} \boldsymbol{X}_{\mathcal{S},\mathcal{T}_{ij}}\}\Big]\end{aligned} \tag{S2.4}$$

Now using the formula (c) and taking expectation over both sides in (S2.4), we get:

$$\begin{aligned}\mathbb{E}\Big\{\frac{\partial \ell_{RC}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S},\mathcal{T}})}{\partial \theta_r}\Big\} = \frac{1}{2} \times \Big[&\sum_{i=1}^{R_s} \sum_{j=1}^{M_s} \{-\frac{1}{2}\text{trace}(\Sigma_{\mathcal{S}_{ij},\mathcal{T}}^{-1} \frac{\partial \Sigma_{\mathcal{S}_{ij},\mathcal{T}}}{\partial \theta_r}) \\ &+\frac{1}{2}\text{trace}(\Sigma_{\mathcal{S}_{ij},\mathcal{T}}^{-1} \frac{\partial \Sigma_{\mathcal{S}_{ij},\mathcal{T}}}{\partial \theta_r} \Sigma_{\mathcal{S}_{ij},\mathcal{T}}^{-1} \Sigma_{\mathcal{S}_{ij},\mathcal{T}})\} \\ &+\sum_{i=1}^{R_t} \sum_{j=1}^{M_t} \{-\frac{1}{2}\text{trace}(\Sigma_{\mathcal{S},\mathcal{T}_{ij}}^{-1} \frac{\partial \Sigma_{\mathcal{S},\mathcal{T}_{ij}}}{\partial \theta_r}) + \\ &\frac{1}{2}\text{trace}(\Sigma_{\mathcal{S},\mathcal{T}_{ij}}^{-1} \frac{\partial \Sigma_{\mathcal{S},\mathcal{T}_{ij}}}{\partial \theta_r} \Sigma_{\mathcal{S},\mathcal{T}_{ij}}^{-1} \Sigma_{\mathcal{S},\mathcal{T}_{ij}})\}\Big]\end{aligned} \tag{S2.5}$$

$$\mathbb{E}\Big\{\frac{\partial \ell_{RC}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S},\mathcal{T}})}{\partial \theta_r}\Big\} = \frac{1}{2} \times \Big[\sum_{i=1}^{R_s} \sum_{j=1}^{M_s} \{-\frac{1}{2}\text{trace}(\Sigma_{\mathcal{S}_{ij},\mathcal{T}}^{-1} \frac{\partial \Sigma_{\mathcal{S}_{ij},\mathcal{T}}}{\partial \theta_r}) + \frac{1}{2}\text{trace}(\Sigma_{\mathcal{S}_{ij},\mathcal{T}}^{-1} \frac{\partial \Sigma_{\mathcal{S}_{ij},\mathcal{T}}}{\partial \theta_r})\} + \sum_{i=1}^{R_t} \sum_{j=1}^{M_t} \{-\frac{1}{2}\text{trace}(\Sigma_{\mathcal{S},\mathcal{T}_{ij}}^{-1} \frac{\partial \Sigma_{\mathcal{S},\mathcal{T}_{ij}}}{\partial \theta_r}) + \frac{1}{2}\text{trace}(\Sigma_{\mathcal{S},\mathcal{T}_{ij}}^{-1} \frac{\partial \Sigma_{\mathcal{S},\mathcal{T}_{ij}}}{\partial \theta_r})\}\Big] = 0 \quad \text{(S2.6)}$$

Therefore, the random composite score is always an unbiased estimating function for $\boldsymbol{\theta}$.

Now, let us consider the second derivative of (S2.2) by using the formulas (d): $d\text{trace}(AB) = \text{trace}(dA.B) + \text{trace}(A.dB)$ and (e) $\frac{\partial \Sigma^{-1}}{\partial \theta_r} = -\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_r} \Sigma^{-1}$:

$$\begin{aligned}\frac{\partial^2 \ell_{RC}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S},\mathcal{T}})}{\partial\theta_r\partial\theta_s} = \frac{1}{2} \times \Bigg[&\sum_{i=1}^{R_s}\sum_{j=1}^{M_s}\{-\frac{1}{2}\text{trace}(\Sigma^{-1}_{\mathcal{S}_{ij},\mathcal{T}}\frac{\partial^2\Sigma_{\mathcal{S}_{ij},\mathcal{T}}}{\partial\theta_r\partial\theta_s}) \\ &+\frac{1}{2}\text{trace}(\Sigma^{-1}_{\mathcal{S}_{ij},\mathcal{T}}\frac{\partial\Sigma_{\mathcal{S}_{ij},\mathcal{T}}}{\partial\theta_s}\Sigma^{-1}_{\mathcal{S}_{ij},\mathcal{T}}\frac{\partial\Sigma_{\mathcal{S}_{ij},\mathcal{T}}}{\partial\theta_r}) \\ &+\frac{1}{2}\boldsymbol{X}^T_{\mathcal{S}_{ij},\mathcal{T}}\Sigma^{-1}_{\mathcal{S}_{ij},\mathcal{T}}\frac{\partial^2\Sigma_{\mathcal{S}_{ij},\mathcal{T}}}{\partial\theta_r\partial\theta_s}\Sigma^{-1}_{\mathcal{S}_{ij},\mathcal{T}}\boldsymbol{X}_{\mathcal{S}_{ij},\mathcal{T}} - \\ &\boldsymbol{X}^T_{\mathcal{S}_{ij},\mathcal{T}}\Sigma^{-1}_{\mathcal{S}_{ij},\mathcal{T}}\frac{\partial\Sigma_{\mathcal{S}_{ij},\mathcal{T}}}{\partial\theta_s}\Sigma^{-1}_{\mathcal{S}_{ij},\mathcal{T}}\frac{\partial\Sigma_{\mathcal{S}_{ij},\mathcal{T}}}{\partial\theta_r}\Sigma^{-1}_{\mathcal{S}_{ij},\mathcal{T}}\boldsymbol{X}_{\mathcal{S}_{ij},\mathcal{T}}\} \\ &+\sum_{i=1}^{R_t}\sum_{j=1}^{M_t}\{-\frac{1}{2}\text{trace}(\Sigma^{-1}_{\mathcal{S},\mathcal{T}_{ij}}\frac{\partial^2\Sigma_{\mathcal{S},\mathcal{T}_{ij}}}{\partial\theta_r\partial\theta_s}) + \\ &\frac{1}{2}\text{trace}(\Sigma^{-1}_{\mathcal{S},\mathcal{T}_{ij}}\frac{\partial\Sigma_{\mathcal{S},\mathcal{T}_{ij}}}{\partial\theta_s}\Sigma^{-1}_{\mathcal{S},\mathcal{T}_{ij}}\frac{\partial\Sigma_{\mathcal{S},\mathcal{T}_{ij}}}{\partial\theta_r}) \\ &+\frac{1}{2}\boldsymbol{X}^T_{\mathcal{S},\mathcal{T}_{ij}}\Sigma^{-1}_{\mathcal{S},\mathcal{T}_{ij}}\frac{\partial^2\Sigma_{\mathcal{S},\mathcal{T}_{ij}}}{\partial\theta_r\partial\theta_s}\Sigma^{-1}_{\mathcal{S},\mathcal{T}_{ij}}\boldsymbol{X}_{\mathcal{S},\mathcal{T}_{ij}} - \\ &\boldsymbol{X}^T_{\mathcal{S},\mathcal{T}_{ij}}\Sigma^{-1}_{\mathcal{S},\mathcal{T}_{ij}}\frac{\partial\Sigma_{\mathcal{S},\mathcal{T}_{ij}}}{\partial\theta_s}\Sigma^{-1}_{\mathcal{S},\mathcal{T}_{ij}}\frac{\partial\Sigma_{\mathcal{S},\mathcal{T}_{ij}}}{\partial\theta_r}\Sigma^{-1}_{\mathcal{S},\mathcal{T}_{ij}}\boldsymbol{X}_{\mathcal{S},\mathcal{T}_{ij}}\}\Bigg]\end{aligned} \tag{S2.7}$$

Now taking expectation on both sides of (S2.7), we get:

$$\begin{aligned}\mathbb{E}\Big\{\frac{\partial^2 \ell_{RC}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S},\mathcal{T}})}{\partial\theta_r\partial\theta_s}\Big\} = \frac{1}{2} \times \Bigg[&\sum_{i=1}^{R_s}\sum_{j=1}^{M_s}\{-\frac{1}{2}\text{trace}(\Sigma^{-1}_{\mathcal{S}_{ij},\mathcal{T}}\frac{\partial^2\Sigma_{\mathcal{S}_{ij},\mathcal{T}}}{\partial\theta_r\partial\theta_s}) + \\ &\frac{1}{2}\text{trace}(\Sigma^{-1}_{\mathcal{S}_{ij},\mathcal{T}}\frac{\partial\Sigma_{\mathcal{S}_{ij},\mathcal{T}}}{\partial\theta_s}\Sigma^{-1}_{\mathcal{S}_{ij},\mathcal{T}}\frac{\partial\Sigma_{\mathcal{S}_{ij},\mathcal{T}}}{\partial\theta_r}) \\ &+\frac{1}{2}\text{trace}(\Sigma^{-1}_{\mathcal{S}_{ij},\mathcal{T}}\frac{\partial^2\Sigma_{\mathcal{S}_{ij},\mathcal{T}}}{\partial\theta_r\partial\theta_s}) - \text{trace}(\Sigma^{-1}_{\mathcal{S}_{ij},\mathcal{T}}\frac{\partial\Sigma_{\mathcal{S}_{ij},\mathcal{T}}}{\partial\theta_s}\Sigma^{-1}_{\mathcal{S}_{ij},\mathcal{T}}\frac{\partial\Sigma_{\mathcal{S}_{ij},\mathcal{T}}}{\partial\theta_r})\} \\ &+\sum_{i=1}^{R_t}\sum_{j=1}^{M_t}\{-\frac{1}{2}\text{trace}(\Sigma^{-1}_{\mathcal{S},\mathcal{T}_{ij}}\frac{\partial^2\Sigma_{\mathcal{S},\mathcal{T}_{ij}}}{\partial\theta_r\partial\theta_s}) + \\ &\frac{1}{2}\text{trace}(\Sigma^{-1}_{\mathcal{S},\mathcal{T}_{ij}}\frac{\partial\Sigma_{\mathcal{S},\mathcal{T}_{ij}}}{\partial\theta_s}\Sigma^{-1}_{\mathcal{S},\mathcal{T}_{ij}}\frac{\partial\Sigma_{\mathcal{S},\mathcal{T}_{ij}}}{\partial\theta_r}) \\ &+\frac{1}{2}\text{trace}(\Sigma^{-1}_{\mathcal{S},\mathcal{T}_{ij}}\frac{\partial^2\Sigma_{\mathcal{S},\mathcal{T}_{ij}}}{\partial\theta_r\partial\theta_s}) - \text{trace}(\Sigma^{-1}_{\mathcal{S},\mathcal{T}_{ij}}\frac{\partial\Sigma_{\mathcal{S},\mathcal{T}_{ij}}}{\partial\theta_s}\Sigma^{-1}_{\mathcal{S},\mathcal{T}_{ij}}\frac{\partial\Sigma_{\mathcal{S},\mathcal{T}_{ij}}}{\partial\theta_r})\}\Bigg]\end{aligned} \tag{S2.8}$$

$$\mathbb{E}\Big\{\frac{\partial^2 \ell_{RC}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S},\mathcal{T}})}{\partial\theta_r\partial\theta_s}\Big\} = \tag{S2.9}$$

$$\frac{1}{2} \times \Big[\sum_{i=1}^{R_s}\sum_{j=1}^{M_s}\{-\frac{1}{2}\text{trace}(\Sigma^{-1}_{\mathcal{S}_{ij},\mathcal{T}}\frac{\partial\Sigma_{\mathcal{S}_{ij},\mathcal{T}}}{\partial\theta_s}\Sigma^{-1}_{\mathcal{S}_{ij},\mathcal{T}}\frac{\partial\Sigma_{\mathcal{S}_{ij},\mathcal{T}}}{\partial\theta_r})\} + \sum_{i=1}^{R_t}\sum_{j=1}^{M_t}\{-\frac{1}{2}\text{trace}(\Sigma^{-1}_{\mathcal{S},\mathcal{T}_{ij}}\frac{\partial\Sigma_{\mathcal{S},\mathcal{T}_{ij}}}{\partial\theta_s}\Sigma^{-1}_{\mathcal{S},\mathcal{T}_{ij}}\frac{\partial\Sigma_{\mathcal{S},\mathcal{T}_{ij}}}{\partial\theta_r})\}\Big]$$

Therefore, the negative expected Hessian $H(\boldsymbol{\theta})$ is given as:

$$H(\boldsymbol{\theta}) = -\mathbb{E}\Big\{\frac{\partial^2 \ell_{RC}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S},\mathcal{T}})}{\partial\theta_r\partial\theta_s}\Big\}$$

$$= \frac{1}{4} \times \Big[\sum_{i=1}^{R_s}\sum_{j=1}^{M_s}\{\text{trace}(\Sigma^{-1}_{\mathcal{S}_{ij},\mathcal{T}}\frac{\partial\Sigma_{\mathcal{S}_{ij},\mathcal{T}}}{\partial\theta_s}\Sigma^{-1}_{\mathcal{S}_{ij},\mathcal{T}}\frac{\partial\Sigma_{\mathcal{S}_{ij},\mathcal{T}}}{\partial\theta_r})\} + \sum_{i=1}^{R_t}\sum_{j=1}^{M_t}\{\text{trace}(\Sigma^{-1}_{\mathcal{S},\mathcal{T}_{ij}}\frac{\partial\Sigma_{\mathcal{S},\mathcal{T}_{ij}}}{\partial\theta_s}\Sigma^{-1}_{\mathcal{S},\mathcal{T}_{ij}}\frac{\partial\Sigma_{\mathcal{S},\mathcal{T}_{ij}}}{\partial\theta_r})\}\Big] \tag{S2.10}$$

Typically, for the asymptotically normal estimators which result from unbiased estimating functions, the associated asymptotic covariance for the estimator has a sandwich form (Godambe, 1960; Eidsvik et al., 2014) under the expanding asymptotics paradigm, and therefore: $\hat{\boldsymbol{\theta}}_{RCL} \sim N(\boldsymbol{\theta}, G^{-1})$,

$$G(\boldsymbol{\theta}) = H(\boldsymbol{\theta})J^{-1}(\boldsymbol{\theta})H(\boldsymbol{\theta})$$

where $J(\boldsymbol{\theta}) = \text{var}(\frac{\partial\ell_{RC}(\boldsymbol{\theta}|\boldsymbol{X}_{\mathcal{S},\mathcal{T}})}{\partial\theta_r})$

Let us now compute the variance of the score function:

We rewrite (S2.4) by absorbing non-random terms into a constant C, and denoting $L^r_{\mathcal{S}_{ij},\mathcal{T}} = \Sigma^{-1}_{\mathcal{S}_{ij},\mathcal{T}} \frac{\partial \Sigma_{\mathcal{S}_{ij},\mathcal{T}}}{\partial \theta_r} \Sigma^{-1}_{\mathcal{S}_{ij},\mathcal{T}}$, and $L^r_{\mathcal{S},\mathcal{T}_{ij}} = \Sigma^{-1}_{\mathcal{S},\mathcal{T}_{ij}} \frac{\partial \Sigma_{\mathcal{S},\mathcal{T}_{ij}}}{\partial \theta_r} \Sigma^{-1}_{\mathcal{S},\mathcal{T}_{ij}}$, we get:

$$\begin{aligned}\frac{\partial \ell_{RC}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S},\mathcal{T}})}{\partial \theta_r} = \frac{1}{2} \times \Big[& \sum_{i=1}^{R_s} \sum_{j=1}^{M_s} \{ \frac{1}{2} \boldsymbol{X}^T_{\mathcal{S}_{ij},\mathcal{T}} L^r_{\mathcal{S}_{ij},\mathcal{T}} \boldsymbol{X}_{\mathcal{S}_{ij},\mathcal{T}} \} \\ & + \sum_{i=1}^{R_t} \sum_{j=1}^{M_t} \{ \frac{1}{2} \boldsymbol{X}^T_{\mathcal{S},\mathcal{T}_{ij}} L^r_{\mathcal{S},\mathcal{T}_{ij}} \boldsymbol{X}_{\mathcal{S},\mathcal{T}_{ij}} \} \Big] + C \end{aligned} \tag{S2.11}$$

Now, we take the variance on both sides of (S2.11) by using the formulas: (f): $\mathrm{var}(Y^T B Y) = 2\mathrm{trace}(B \Sigma_Y B \Sigma_Y)$ and (g): $\mathrm{cov}(Y^T B_r Y, Y^T B_s Y) = 2\mathrm{trace}(B_r \Sigma_Y B_s \Sigma_Y)$, we get:

$$\begin{aligned}\mathrm{var}(\frac{\partial \ell_{RC}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S},\mathcal{T}})}{\partial \theta_r}) = \frac{1}{16} \times \Big[& \mathrm{var}[\sum_{i=1}^{R_s} \sum_{j=1}^{M_s} \{ \boldsymbol{X}^T_{\mathcal{S}_{ij},\mathcal{T}} L^r_{\mathcal{S}_{ij},\mathcal{T}} \boldsymbol{X}_{\mathcal{S}_{ij},\mathcal{T}} \}] \\ & + \mathrm{var}[\sum_{i=1}^{R_t} \sum_{j=1}^{M_t} \{ \boldsymbol{X}^T_{\mathcal{S},\mathcal{T}_{ij}} L^r_{\mathcal{S},\mathcal{T}_{ij}} \boldsymbol{X}_{\mathcal{S},\mathcal{T}_{ij}} \}] + \\ \mathrm{cov}(\sum_{i=1}^{R_s} \sum_{j=1}^{M_s} \{ \boldsymbol{X}^T_{\mathcal{S}_{ij},\mathcal{T}} L^r_{\mathcal{S}_{ij},\mathcal{T}} \boldsymbol{X}_{\mathcal{S}_{ij},\mathcal{T}} \}, & \sum_{k=1}^{R_t} \sum_{l=1}^{M_t} \{ \boldsymbol{X}^T_{\mathcal{S},\mathcal{T}_{kl}} L^r_{\mathcal{S},\mathcal{T}_{kl}} \boldsymbol{X}_{\mathcal{S},\mathcal{T}_{kl}} \}) \Big] \end{aligned} \tag{S2.12}$$

Let us first simplify $\mathrm{var}[\sum_{i=1}^{R_s} \sum_{j=1}^{M_s} \{ \boldsymbol{X}^T_{\mathcal{S}_{ij},\mathcal{T}} L^r_{\mathcal{S}_{ij},\mathcal{T}} \boldsymbol{X}_{\mathcal{S}_{ij},\mathcal{T}} \}]$:

$$\operatorname{var}[\sum_{i=1}^{R_s}\sum_{j=1}^{M_s}\{\boldsymbol{X}^T_{\mathcal{S}_{ij},\mathcal{T}}L^r_{\mathcal{S}_{ij},\mathcal{T}}\boldsymbol{X}_{\mathcal{S}_{ij},\mathcal{T}}\}] = \tag{S2.13}$$
$$\sum_{i=1}^{R_s}\operatorname{var}(\sum_{j=1}^{M_s}\boldsymbol{X}^T_{\mathcal{S}_{ij},\mathcal{T}}L^r_{\mathcal{S}_{ij},\mathcal{T}}\boldsymbol{X}_{\mathcal{S}_{ij},\mathcal{T}})$$
$$+\sum_{l\neq m=1}^{R_s}\operatorname{cov}(\sum_{j=1}^{M_s}\boldsymbol{X}^T_{\mathcal{S}_{lj},\mathcal{T}}L^r_{\mathcal{S}_{lj},\mathcal{T}}\boldsymbol{X}_{\mathcal{S}_{lj},\mathcal{T}},\sum_{n=1}^{M_s}\boldsymbol{X}^T_{\mathcal{S}_{mn},\mathcal{T}}L^r_{\mathcal{S}_{mn},\mathcal{T}}\boldsymbol{X}_{\mathcal{S}_{mn},\mathcal{T}})$$

$$\operatorname{var}[\sum_{i=1}^{R_s}\sum_{j=1}^{M_s}\{\boldsymbol{X}^T_{\mathcal{S}_{ij},\mathcal{T}}L^r_{\mathcal{S}_{ij},\mathcal{T}}\boldsymbol{X}_{\mathcal{S}_{ij},\mathcal{T}}\}] = \tag{S2.14}$$
$$\sum_{i=1}^{R_s}\sum_{j=1}^{M_s}\operatorname{var}(\boldsymbol{X}^T_{\mathcal{S}_{ij},\mathcal{T}}L^r_{\mathcal{S}_{ij},\mathcal{T}}\boldsymbol{X}_{\mathcal{S}_{ij},\mathcal{T}})$$
$$+\sum_{i=1}^{R_s}\sum_{j\neq j'}^{M_s}\operatorname{cov}(\boldsymbol{X}^T_{\mathcal{S}_{ij},\mathcal{T}}L^r_{\mathcal{S}_{ij},\mathcal{T}}\boldsymbol{X}_{\mathcal{S}_{ij},\mathcal{T}},\boldsymbol{X}^T_{\mathcal{S}_{ij'},\mathcal{T}}L^r_{\mathcal{S}_{ij'},\mathcal{T}}\boldsymbol{X}_{\mathcal{S}_{ij'},\mathcal{T}})$$
$$+\sum_{l\neq m=1}^{R_s}\sum_{j=1}^{M_s}\sum_{n=1}^{M_s}\operatorname{cov}(\boldsymbol{X}^T_{\mathcal{S}_{lj},\mathcal{T}}L^r_{\mathcal{S}_{lj},\mathcal{T}}\boldsymbol{X}_{\mathcal{S}_{lj},\mathcal{T}},\boldsymbol{X}^T_{\mathcal{S}_{mn},\mathcal{T}}L^r_{\mathcal{S}_{mn},\mathcal{T}}\boldsymbol{X}_{\mathcal{S}_{mn},\mathcal{T}})$$

$$\operatorname{var}[\sum_{i=1}^{R_s}\sum_{j=1}^{M_s}\{\boldsymbol{X}^T_{\mathcal{S}_{ij},\mathcal{T}}L^r_{\mathcal{S}_{ij},\mathcal{T}}\boldsymbol{X}_{\mathcal{S}_{ij},\mathcal{T}}\}] = \tag{S2.15}$$
$$2\sum_{i=1}^{R_s}\sum_{j=1}^{M_s}\operatorname{trace}(L^r_{\mathcal{S}_{ij},\mathcal{T}}\Sigma_{\mathcal{S}_{ij},\mathcal{T}}L^r_{\mathcal{S}_{ij},\mathcal{T}})$$
$$+2\sum_{i=1}^{R_s}\sum_{j\neq j'}^{M_s}\operatorname{trace}(\overline{L^r_{\mathcal{S}_{ij},\mathcal{T}}}\Sigma_{\mathcal{S}_{ij},\mathcal{S}_{ij'},\mathcal{T}}\underline{L^r_{\mathcal{S}_{ij'},\mathcal{T}}}\Sigma_{\mathcal{S}_{ij},\mathcal{S}_{ij'},\mathcal{T}})$$
$$+2\sum_{l\neq m=1}^{R_s}\sum_{j=1}^{M_s}\sum_{n=1}^{M_s}\operatorname{trace}(\overline{L^r_{\mathcal{S}_{lj},\mathcal{T}}}\Sigma_{\mathcal{S}_{ij},\mathcal{S}_{mn},\mathcal{T}}\underline{L^r_{\mathcal{S}_{mn},\mathcal{T}}}\Sigma_{\mathcal{S}_{ij},\mathcal{S}_{mn},\mathcal{T}}),$$

where $\Sigma_{\mathcal{S}_{ij},\mathcal{S}_{mn},\mathcal{T}}$ is the covariance matrix for $(\boldsymbol{X}^T_{\mathcal{S}_{ij},\mathcal{T}}, \boldsymbol{X}^T_{\mathcal{S}_{ij'},\mathcal{T}})^T$,

$$\overline{L^r_{\mathcal{S}_{ij},\mathcal{T}}} = \begin{bmatrix} L^r_{\mathcal{S}_{ij},\mathcal{T}} & 0 \\ 0 & 0 \end{bmatrix}, \quad \underline{L^r_{\mathcal{S}_{ij'},\mathcal{T}}} = \begin{bmatrix} 0 & 0 \\ 0 & L^r_{\mathcal{S}_{ij'},\mathcal{T}} \end{bmatrix}$$

Similarly, we get:

$$\begin{aligned} \text{var}[\sum_{i=1}^{R_t}\sum_{j=1}^{M_t}\{\boldsymbol{X}^T_{\mathcal{S},\mathcal{T}_{ij}} L^r_{\mathcal{S},\mathcal{T}_{ij}} \boldsymbol{X}_{\mathcal{S},\mathcal{T}_{ij}}\}] = \\ 2\sum_{i=1}^{R_t}\sum_{j=1}^{M_t} \text{trace}(L^r_{\mathcal{S},\mathcal{T}_{ij}} \Sigma_{\mathcal{S},\mathcal{T}_{ij}} L^r_{\mathcal{S},\mathcal{T}_{ij}}) \\ +2\sum_{i=1}^{R_t}\sum_{j\neq j'}^{M_t} \text{trace}(\overline{L^r_{\mathcal{S},\mathcal{T}_{ij}}} \Sigma_{\mathcal{S},\mathcal{T}_{ij},\mathcal{T}_{ij'}} \underline{L^r_{\mathcal{S},\mathcal{T}_{ij}}} \Sigma_{\mathcal{S},\mathcal{T}_{ij},\mathcal{T}_{ij'}}) \\ +2\sum_{l\neq m=1}^{R_t}\sum_{j=1}^{M_t}\sum_{n=1}^{M_t} \text{trace}(\overline{L^r_{\mathcal{S},\mathcal{T}_{lj}}} \Sigma_{\mathcal{S},\mathcal{T}_{ij},\mathcal{T}_{mn}} \underline{L^r_{\mathcal{S},\mathcal{T}_{mn}}} \Sigma_{\mathcal{S},\mathcal{S}_{ij},\mathcal{T}_{mn}}) \end{aligned} \tag{S2.16}$$

and

$$\begin{aligned} \sum_{i=1}^{R_s}\sum_{j=1}^{M_s}\sum_{k=1}^{R_t}\sum_{l=1}^{M_t} \text{cov}(\boldsymbol{X}^T_{\mathcal{S}_{ij},\mathcal{T}} L^r_{\mathcal{S}_{ij},\mathcal{T}} \boldsymbol{X}_{\mathcal{S}_{ij},\mathcal{T}}, \boldsymbol{X}^T_{\mathcal{S},\mathcal{T}_{kl}} L^r_{\mathcal{S},\mathcal{T}_{kl}} \boldsymbol{X}_{\mathcal{S},\mathcal{T}_{kl}}) = \\ 2\sum_{i=1}^{R_s}\sum_{j=1}^{M_s}\sum_{k=1}^{R_t}\sum_{l=1}^{M_t} \text{trace}(\overline{L^r_{\mathcal{S}_{ij},\mathcal{T}}} \Sigma_{\mathcal{S}_{ij},\mathcal{T},\mathcal{S},\mathcal{T}_{kl}} \underline{L^r_{\mathcal{S},\mathcal{T}_{kl}}} \Sigma_{\mathcal{S}_{ij},\mathcal{T},\mathcal{S},\mathcal{T}_{kl}}), \end{aligned} \tag{S2.17}$$

where $\Sigma_{\mathcal{S}_{ij},\mathcal{T},\mathcal{S},\mathcal{T}_{kl}}$ is the covariance matrix of $(\boldsymbol{X}^T_{\mathcal{S}_{ij},\mathcal{T}}, \boldsymbol{X}^T_{\mathcal{S},\mathcal{T}_{kl}})^T$

Therefore, by using (S2.15),(S2.16) and (S2.17), we can obtain the diagonal entries of $J(\boldsymbol{\theta})$:

$$
\begin{aligned}
\operatorname{var}(\frac{\partial \ell_{RC}(\boldsymbol{\theta} \mid \boldsymbol{X}_{\mathcal{S},\mathcal{T}})}{\partial \theta_r}) = \frac{1}{8} \times \Bigg[& \sum_{i=1}^{R_s}\sum_{j=1}^{M_s} \mathrm{trace}(L^r_{\mathcal{S}_{ij},\mathcal{T}} \Sigma_{\mathcal{S}_{ij},\mathcal{T}} L^r_{\mathcal{S}_{ij},\mathcal{T}}) \\
& + \sum_{i=1}^{R_s}\sum_{j\neq j'}^{M_s} \mathrm{trace}(\overline{L^r_{\mathcal{S}_{ij},\mathcal{T}}} \Sigma_{\mathcal{S}_{ij},\mathcal{S}_{ij'},\mathcal{T}} \underline{L^r_{\mathcal{S}_{ij'},\mathcal{T}}} \Sigma_{\mathcal{S}_{ij},\mathcal{S}_{ij'},\mathcal{T}}) \\
& + \sum_{l\neq m=1}^{R_s}\sum_{j=1}^{M_s}\sum_{n=1}^{M_s} \mathrm{trace}(\overline{L^r_{\mathcal{S}_{lj},\mathcal{T}}} \Sigma_{\mathcal{S}_{ij},\mathcal{S}_{mn},\mathcal{T}} \underline{L^r_{\mathcal{S}_{mn},\mathcal{T}}} \Sigma_{\mathcal{S}_{ij},\mathcal{S}_{mn},\mathcal{T}}) \\
& + \sum_{i=1}^{R_t}\sum_{j=1}^{M_t} \mathrm{trace}(L^r_{\mathcal{S},\mathcal{T}_{ij}} \Sigma_{\mathcal{S},\mathcal{T}_{ij}} L^r_{\mathcal{S},\mathcal{T}_{ij}}) \\
& + \sum_{i=1}^{R_t}\sum_{j\neq j'}^{M_t} \mathrm{trace}(\overline{L^r_{\mathcal{S},\mathcal{T}_{ij}}} \Sigma_{\mathcal{S},\mathcal{T}_{ij},\mathcal{T}_{ij'}} \underline{L^r_{\mathcal{S},\mathcal{T}_{ij}}} \Sigma_{\mathcal{S},\mathcal{T}_{ij},\mathcal{T}_{ij'}}) \\
& + \sum_{l\neq m=1}^{R_t}\sum_{j=1}^{M_t}\sum_{n=1}^{M_t} \mathrm{trace}(\overline{L^r_{\mathcal{S},\mathcal{T}_{lj}}} \Sigma_{\mathcal{S},\mathcal{T}_{ij},\mathcal{T}_{mn}} \underline{L^r_{\mathcal{S},\mathcal{T}_{mn}}} \Sigma_{\mathcal{S},\mathcal{S}_{ij},\mathcal{T}_{mn}}) \\
& + \sum_{i=1}^{R_s}\sum_{j=1}^{M_s}\sum_{k=1}^{R_t}\sum_{l=1}^{M_t} \mathrm{trace}(\overline{L^r_{\mathcal{S}_{ij},\mathcal{T}}} \Sigma_{\mathcal{S}_{ij},\mathcal{T},\mathcal{S},\mathcal{T}_{kl}} \underline{L^r_{\mathcal{S},\mathcal{T}_{kl}}} \Sigma_{\mathcal{S}_{ij},\mathcal{T},\mathcal{S},\mathcal{T}_{kl}}) \Bigg]. \quad \text{(S2.18)}
\end{aligned}
$$

Similarly, we can obtain the off-diagonal entries of $J(\boldsymbol{\theta})$ and plug it in the formula of the $G(\boldsymbol{\theta})$ to obtain the variance of the parameter estimates $\hat{\boldsymbol{\theta}}_{RCL}$.

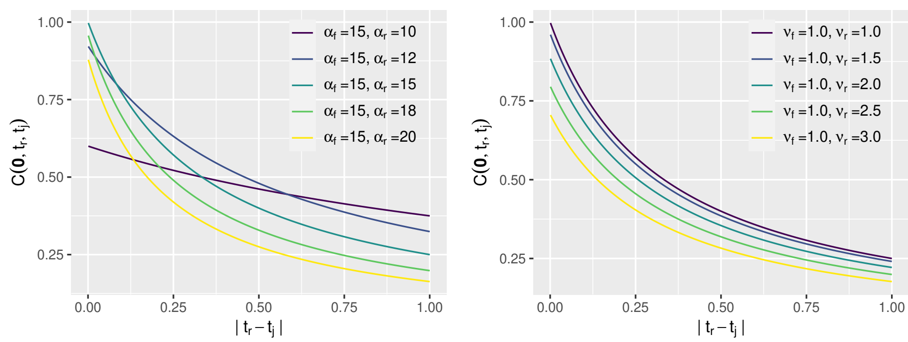
S3 Temporal Nonstationary Effects of $\alpha_s(t)$ and $\nu_s(t)$

Firstly, let us fix: (i) $\nu_s(t) = \nu_f,\ t \in \mathbb{R}$, (ii) $\alpha_s(t_r) = \alpha_r$; $\alpha_s(t_j) = \alpha_f,\ t_j \neq t_r \in \mathbb{R}$, for any arbitrary reference time point t_r, and (iii) $\overline{\alpha}_s = \alpha_f$ in Equation 3.4 of the main manuscript, then the temporal covariance of the reference time-point t_r with any other time point t_j is given as:

$$\mathrm{C}(\mathbf{0}, t_r, t_j) = \frac{\sigma^2 (\frac{\alpha_f^2}{\alpha_r^2})^{d/4}}{\{\psi(|t_r - t_j|^2) - \psi(0) + \frac{1}{2}(1 + \frac{\alpha_f^2}{\alpha_r^2})\}^{d/2}}, \text{ for all } t_j \in \mathbb{R}. \quad \text{(S3.1)}$$

Note that it is reasonable here to fix $\overline{\alpha}_s = \alpha_f$ even under the aforementioned choice of constraint for $\overline{\alpha}_s$, as $\overline{\alpha}_s = \sum_{t_i \in \mathrm{T}} \alpha_s(t_i)/\mathrm{T} \approx \alpha_f$, provided that T includes a large number of training points and α_r is not extremely different from α_f. The function in (S3.1) expresses covariance of a reference time point t_r with any other time point $t_j \in \mathbb{R}$ as a function of $\mathbb{L}_1$ distance between them, i.e., $|t_r - t_j|$, and the term α_f/α_r counter-balances the scale of the covariance and the rate of covariance decay with increasing distance $|t_r - t_j|$. For instance, if $\alpha_r < \alpha_f$, then for non-zero temporal lags, the scale of covariance is increased and the rate of covariance decay is decreased through the term α_f/α_r in the numerator and denominator of (S3.1), respectively. Therefore, the function $\alpha_s(t)$, not only denotes the spatial scale of purely spatial Matérn covariance at time t, but also governs the scaling and rate of temporal covariance decay away from the time point t. Next, we fix (i) $\alpha_s(t) = \alpha_f,\ t \in \mathbb{R}$, (ii) $\overline{\alpha}_s = \alpha_f$, and (iii) $\nu(t_r) = \nu_r; \nu(t_j) = \nu_f, t_j \neq t_r \in \mathbb{R}$ in Equation 3.4 of the main manuscript, we get:

$$\mathrm{C}(\mathbf{0}, t_r, t_j) = \frac{\sigma^2 \Gamma(\frac{\nu_r + \nu_f}{2})}{\sqrt{\Gamma(\nu_r)\Gamma(\nu_f)} \{\psi(|t_r - t_j|^2) - \psi(0) + 1\}^{d/2}}, \text{ for all } t_j \in \mathbb{R}, \quad \text{(S3.2)}$$



(a) (b)

Figure 1: (a) The purely temporal covariance function in (S2.5) as a function of $|t_r - t_j|$ for different choices of α_f and α_r. (b) The purely temporal covariance function in (S2.6) as a function of $|t_r - t_j|$ for different choices of ν_f and ν_r. For both (a) and (b), we have fixed $\sigma = 1$ and $\psi(w) = (3w^{0.5} + 1)$.

where the terms ν_r and ν_f control the scale of the temporal covariance such that the covariance is scaled down if $\nu_r \neq \nu_f$ and the magnitude of this downscaling is directly proportional to the difference between ν_f and ν_r. Hence, the function $\nu_s(t)$ also plays a twofold role where on one hand it controls the smoothness of the purely spatial Matérn covariance at time t, on the other hand, it regulates the scaling of temporal covariance at non-zero temporal lags.

These effect of $\alpha_s(t)$ and $\nu_s(t)$ on the purely temporal covariance function is also illustrated with examples in Figure 1. For $\sigma = 1$ and $\psi(w) = (3w^{0.5} + 1)$, Figure 1(a) and Figure 1(b) show the temporal covariance function in (S3.1) for different combinations of (α_f, α_r) and the temporal covariance function in (S3.2), for different combinations of (ν_f, ν_r), respectively. In particular, the illustrated combinations are $(\alpha_f, \alpha_r) \in \{(15, 10), (15, 12), (15, 15), (15, 18), (15, 20)\}$ and $(\nu_f, \nu_r) \in \{(1, 1), (1, 1.5), (1, 2), (1, 2.5), (1, 3)\}$, where we consider $(\alpha_f, \alpha_r) = (15, 15)$ and $(\nu_f, \nu_r) = (1, 1)$ as the base cases for studying their effects. As shown in Figure 1(a), for the cases when $\alpha_f > \alpha_r$, i.e., $(\alpha_f, \alpha_r) \in \{(15, 10), (15, 12)\}$, the rate of covariance decay is decreased and the scaling is increased compared to the base case, whereas for the cases $\alpha_f < \alpha_r$, i.e., $(\alpha_f, \alpha_r) \in \{(15, 18), (15, 20)\}$, the rate of covariance decay is increased and the scaling is decreased. Moreover, the effect is stronger when the difference between α_f and α_r is higher. Similarly, relative to the base case, the scale of the covariance is clearly decreased in Figure 1(b) for $(\nu_f, \nu_r) \in \{(1, 1.5), (1, 2), (1, 2.5), (1, 3)\}$ and the decrease is the highest when ν_r is the farthest from ν_f, i.e., $(\nu_f, \nu_r) = (1, 3)$.

S4 Extended Discussion from the Simulation Study

S4.1 Temporal nonstationarity of the data generating models in the simulation study

In terms of the purely temporal covariance of the data generating model as shown in Figure 2, the specified $\alpha_s(t)$ and $\nu_s(t)$ impart nonstationarity in Case 1, Case 2 and Case 3, and stationarity in Case 4. In particular, the temporal covariance becomes stronger at the middle of $\mathcal{D}_t$ for Case 1, and at the higher end of $\mathcal{D}_t$ for Case 2 and Case 3.

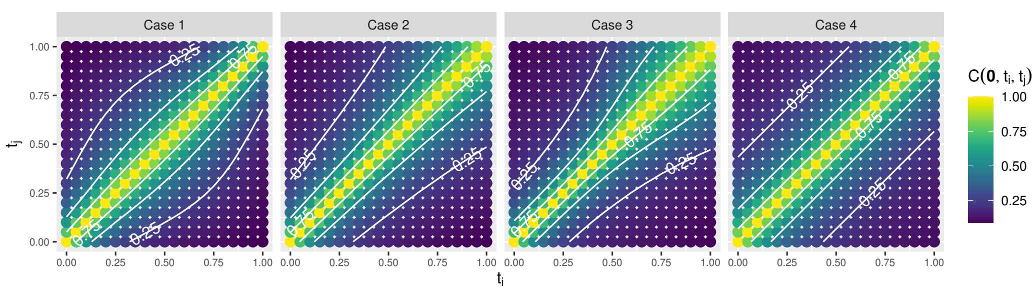


Figure 2: The purely temporal covariance of the true data generating model as a function of time pairs (t_i, t_j), $t_i, t_j \in \mathcal{D}_t$, for all the four cases. The white lines represent the contours at levels: 0.75, 0.5 and 0.25. The purely temporal covariance in Case 1, Case 2 and Case 3 is nonstationary, as the covariance decay is slower in the middle of $\mathcal{D}_t$ for Case 1, and the covariance decay is slower at the higher end of $\mathcal{D}_t$ for Case 2 and Case 3. The purely temporal covariance in Case 4 is stationary, as the structure remain constant throughout t.

S4.2 Additional Figures from the Simulation Study

An example realization of Z for all the four cases from the simulation study presented in the main manuscript can be found in Figures 3–6.

S4.3 RCL parameter estimates from the simulation study

Table 1 reports the average and standard deviation of parameter estimates over the 100 simulation runs, for all the three candidate models under the four simulation cases. The parameter estimates of σ, γ and β under the Gneit.M model are close to their respective true values in all the four cases, however, since the Gneit.M model is misspecified for the time-varying part $\alpha_s(t)$ and $\nu_s(t)$ of the true data generating model, the respective constant estimates are not comparable to the true functions in Cases 1–3. Albeit, for Case 4 where the true $\alpha_s(t)$ and $\nu_s(t)$ are constant, the corresponding estimates under the Gneit.M model are almost equal to their true values. Among the three candidate models, Sep.M is the most extreme misspecification of the true process in all the four cases, and consequently its parameter estimates exhibit the strongest disagreement with their respective true values in all the four cases. All the parameter estimates from the candidate Tvar.M shown in Table 1 are nearly equal to their corresponding true values, in all the four cases. Note that the average estimate and standard

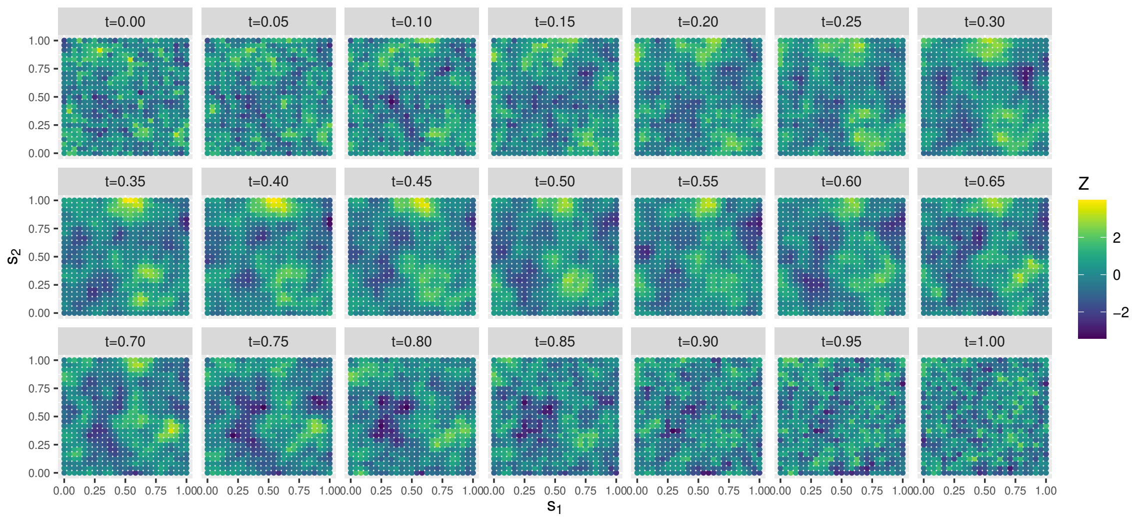


Figure 3: An example of a simulated realization of Z from Case 1.

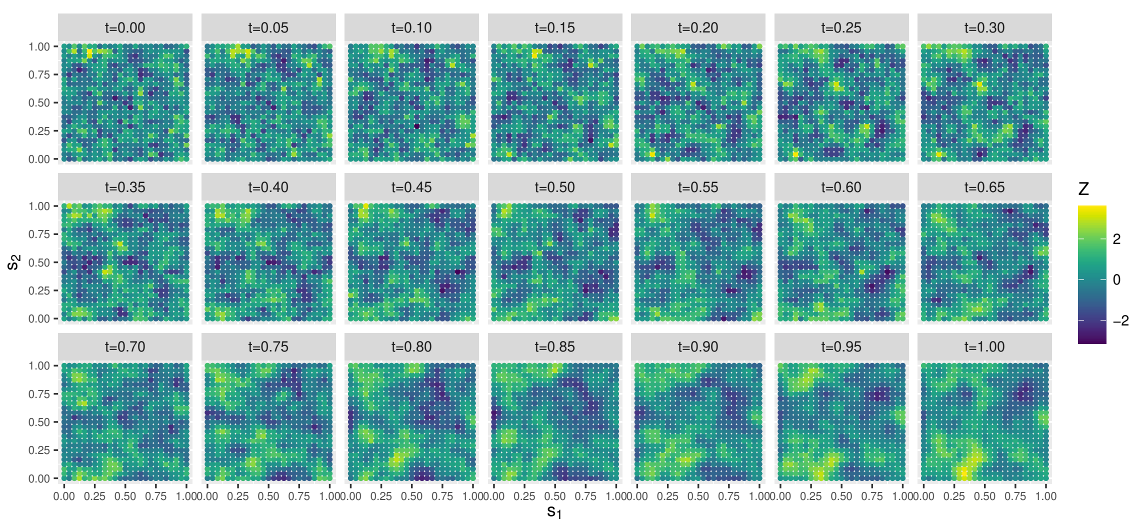


Figure 4: An example of a simulated realization of Z from Case 2.

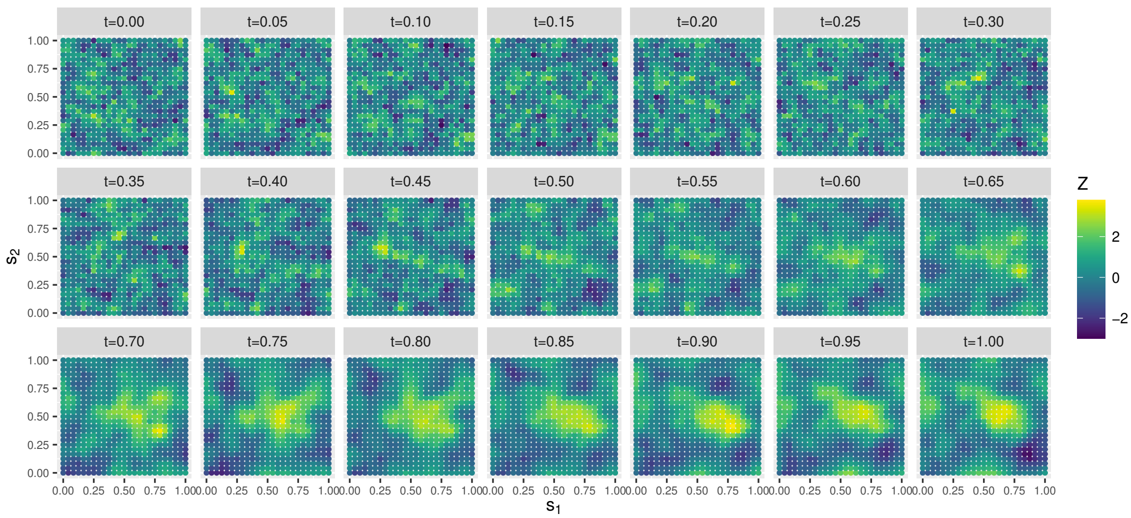


Figure 5: An example of a simulated realization of Z from Case 3.

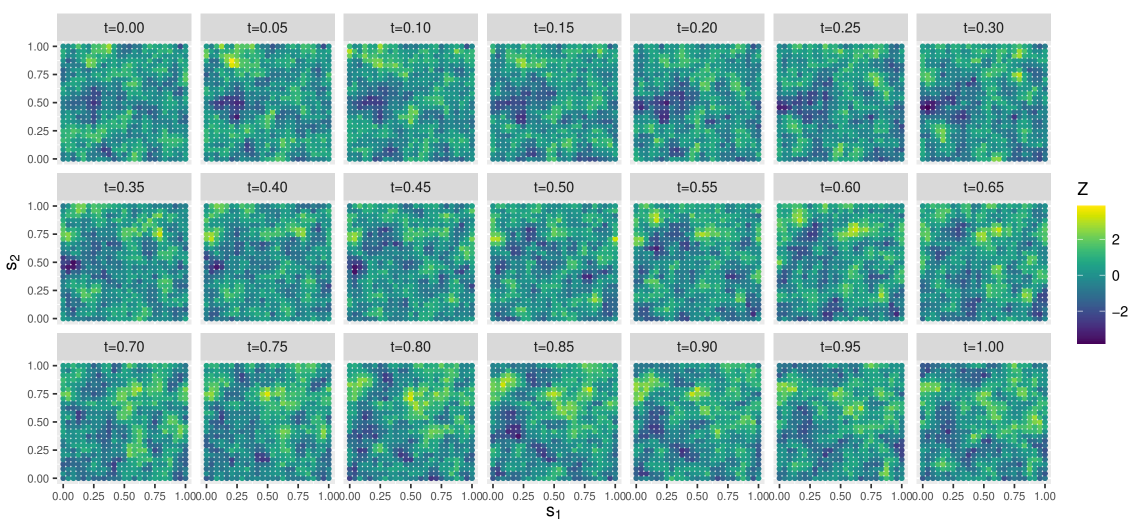


Figure 6: An example of a simulated realization of Z from Case 4.

deviation entries for $\alpha_s(t)$ and $\nu_s(t)$ are left blank under the Tvar.M since the comparison of true and estimated functional parameters $\alpha_s(t)$ and $\nu_s(t)$ under Tvar.M is shown in Figure 3 of the main manuscript. The entry for the average estimate and standard deviation of β under Sep.M is left blank because $\beta = 0$ for separable model.

S5 Polynomial Order Selection

In our study, the time-varying functions $\alpha_s(t)$ and $\nu_s(t)$ are specified as $\exp(p^{\alpha}_{n.\alpha}(t))$ and $\exp(p^{\alpha}_{n.\nu}(t))$, respectively, where $p^{\alpha}_{k}(t)$ and $p^{\nu}_{k}(t)$ are both k-degree polynomials of t. The selection of an appropriate value for $n.\alpha$ and $n.\nu$ is crucial for this polynomial-based specification. In this section, we provide detailed guidelines to ensure an empirically informed, sufficiently flexible, and computationally feasible choice of values for $n.\alpha$ and $n.\nu$.

We begin by estimating the spatial scale $\hat{\alpha}_i$ and smoothness parameter $\hat{\nu}_i$ of the Matérn covariance function independently for each time point $t_i \in \mathrm{T}$. Since our proposed model results in a purely Matérn spatial covariance for each time point, the estimated parameters $\hat{\alpha}_i$ and $\hat{\nu}_i$ should align with the time-varying functions $\alpha_s(t_i)$ and $\nu_s(t_i)$, $t_i \in \mathrm{T}$. Therefore, to determine suitable polynomial orders, we fit regression models $\log(\hat{\alpha}i) \sim p^{\alpha}_{n.\alpha}(t_i)$ and $\log(\hat{\nu}i) \sim p^{\nu}_{n.\nu}(t_i)$ for increasing values of $n.\alpha \geq 0$ and

$n.\nu \geq 0$. Additionally, to achieve balanced flexibility in both time-varying spatial smoothness and spatial scale, we impose the constraint $n.\alpha = n.\nu$ on our choice of polynomial orders in the simulation study and the data application. By jointly analyzing the fitted polynomials through the adjusted R^2 values, visual inspection of the fits, and computational complexity of the resulting time-varying model, we make an informed decision regarding the appropriate orders.

In our simulation study, we fit polynomials of orders $n.\alpha = n.\nu = 0, 1, 2, 3, 4, 5$, for the four cases. For the data application, we fit a polynomial of order $n.\alpha = n.\nu = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$. The selection of the optimal order is based on the adjusted R^2 values, aiming to strike a balance between model flexibility and computational feasibility. We choose the lowest possible order such that the next higher order provides a negligible or no increase in the adjusted R^2 value.

The plots for the fitted polynomials in the four simulation cases are shown in Figures 7–10, and the equivalent plot for the data application is shown in Figure 11. The corresponding adjusted R^2 values for the simulation cases and the data application are reported in Table 2 and Table 3, respectively.

For Case 1 and Case 2, as shown in Figure 7 and Figure 8, the quadratic polynomial visually captures the variation of $\hat{\alpha}_i$ and $\hat{\nu}_i$ well. The adjusted R^2 values also support the use of a quadratic polynomial, as higher orders do not significantly increase the adjusted R^2.

In Case 3, the quadratic polynomial seems insufficient to capture the variation of $\hat{\alpha}_i$ and $\hat{\nu}_i$, as evident from the fitted polynomials in Figure 9. The cubic polynomial appears to be a better fit. The adjusted R^2 values in Table 2 also support the use of a cubic polynomial, as higher orders do not lead to a noticeable increase in adjusted R^2.

For Case 4, we opt for the quadratic polynomial despite the highest adjusted R^2 values in Table 2 being attributed to the fifth-order polynomial. Our decision is informed by the exploratory plots illustrated in Figure 10. The range of $\hat{\alpha}_i$ and $\hat{\nu}_i$ is quite narrow and lacks any discernible pattern that would justify the use of a polynomial of a higher degree. Considering these empirical observations, a quadratic polynomial seems to be a judicious selection.

Similarly, for the data application, a fifth-order polynomial strikes a balance between model flexibility and computational feasibility, as observed from Figure 11 and Table 3.

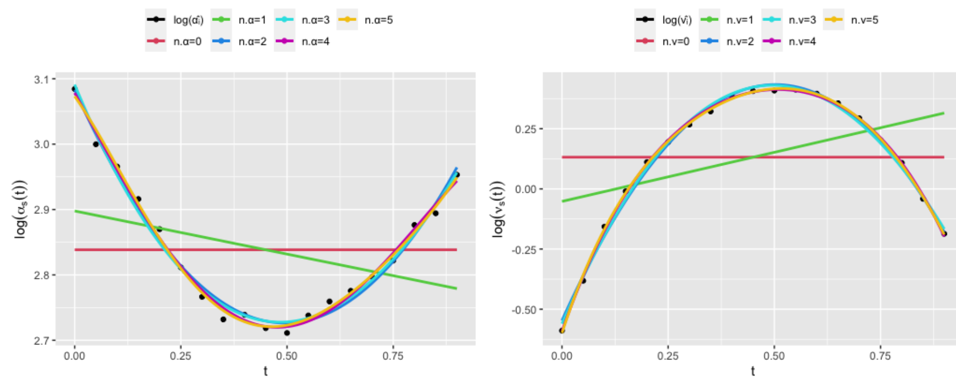


Figure 7: Exploratory plots for the selection on polynomial orders $n.\alpha$ and $n.\nu$ in Case 1. The estimates $\log(\hat{\alpha}_i)$ (left) and $\log(\hat{\nu}_i)$ are based on the average over 100 simulation runs. Colored curves depict polynomial fits of varying degrees.

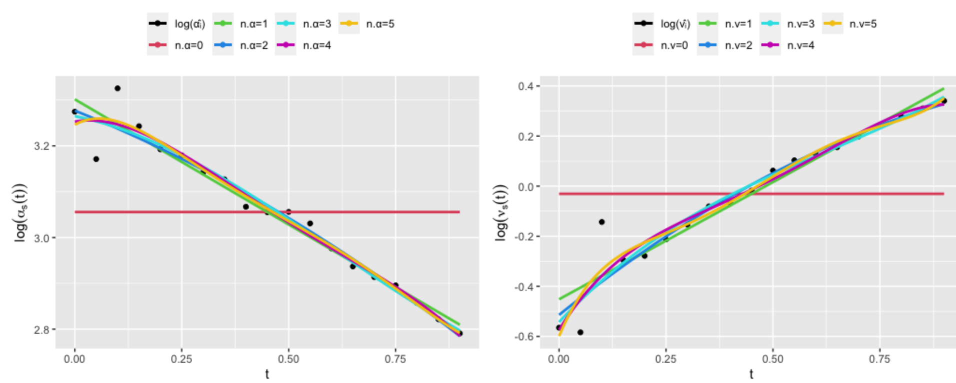


Figure 8: Exploratory plots for the selection on polynomial orders $n.\alpha$ and $n.\nu$ in Case 2. The estimates $\log(\hat{\alpha}_i)$ (left) and $\log(\hat{\nu}_i)$ are based on the average over 100 simulation runs. Colored curves depict polynomial fits of varying degrees.

Table 1: Average and standard deviation of parameter estimates over the 100 simulation runs.

Cases	Parameter/ Function	True value/ True specification	Mean (std. dev.) of the parameter estimates Tvar.M	Gneit.M	Sep.M
Case 1	σ	1	0.99 (0.04)	0.99 (0.04)	0.96 (0.04)
	γ	0.60	0.60 (0.02)	0.61 (0.02)	0.61 (0.02)
	β	0.80	0.75 (0.15)	0.79 (0.12)	–
	δ	0.10	0.17 (0.17)	0.18 (0.17)	1.05 (0.15)
	$\alpha_s(t)$	$20 + 15\sin(\frac{\pi t}{20})$	–	15.16 (2.06)	25.90 (1.99)
	$\nu_s(t)$	$0.5 + \sin(\frac{\pi t}{20})$	–	0.94 (0.09)	1.40 (0.11)
Case 2	σ	1	0.99 (0.03)	0.99 (0.03)	0.96 (0.03)
	γ	0.60	0.60 (0.02)	0.60 (0.01)	0.60 (0.01)
	β	0.80	0.79 (0.14)	0.86 (0.09)	–
	δ	0.10	0.16 (0.17)	0.11 (0.08)	1.03 (0.11)
	$\alpha_s(t)$	$25 - 10t$	–	18.12 (1.87)	32.64 (2.82)
	$\nu_s(t)$	$0.5 + t$	–	0.76 (0.07)	1.23 (0.13)
Case 3	σ	1	1.00 (0.03)	0.98 (0.04)	0.96 (0.04)
	γ	0.60	0.60 (0.02)	0.59 (0.01)	0.59 (0.01)
	β	0.80	0.83 (0.13)	0.90 (0.09)	–
	δ	0.10	0.10 (0.10)	0.07 (0.09)	1.03 (0.13)
	$\alpha_s(t)$	$20 - 10\frac{\exp(10t-5)}{1+\exp(10t-5)}$	–	11.35 (1.62)	18.52 (2.18)
	$\nu_s(t)$	$0.5 + \frac{\exp(10t-5)}{1+\exp(10t-5)}$	–	0.56 (0.06)	0.73 (0.08)
Case 4	σ	1	0.99 (0.03)	1.00 (0.03)	0.97 (0.03)
	γ	0.60	0.60 (0.01)	0.60 (0.01)	0.60 (0.01)
	β	0.80	0.75 (0.11)	0.78 (0.08)	–
	δ	0.10	0.17 (0.15)	0.14 (0.12)	1.00 (0.11)
	$\alpha_s(t)$	20	–	20.51 (2.56)	38.65 (2.76)
	$\nu_s(t)$	1	–	1.03 (0.10)	1.83 (0.17)

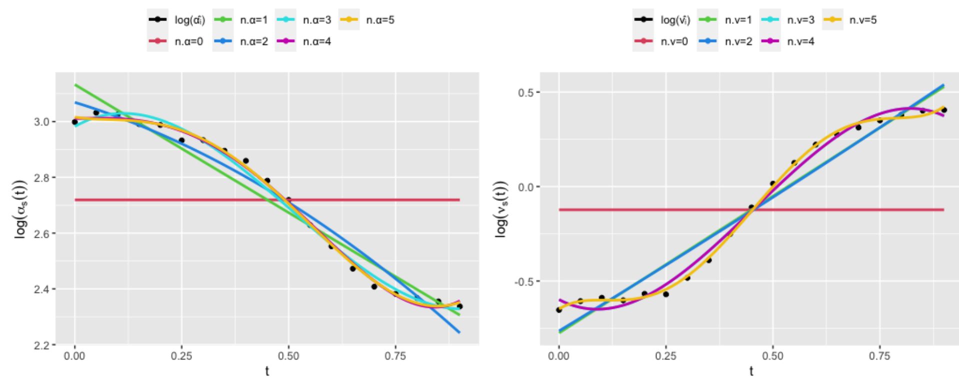


Figure 9: Exploratory plots for the selection on polynomial orders $n.\alpha$ and $n.\nu$ in Case 3. The estimates $\log(\hat{\alpha}_i)$ (left) and $\log(\hat{\nu}_i)$ are based on the average over 100 simulation runs. Colored curves depict polynomial fits of varying degrees.

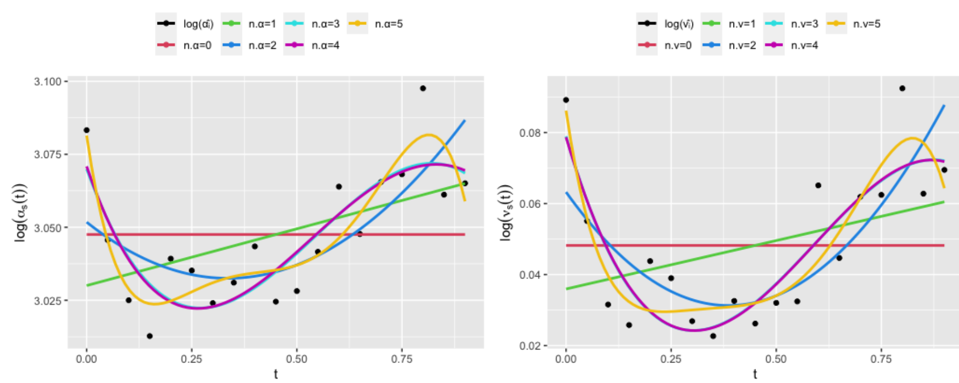


Figure 10: Exploratory plots for the selection on polynomial orders $n.\alpha$ and $n.\nu$ in Case 4. The estimates $\log(\hat{\alpha}_i)$ (left) and $\log(\hat{\nu}_i)$ are based on the average over 100 simulation runs. Colored curves depict polynomial fits of varying degrees.

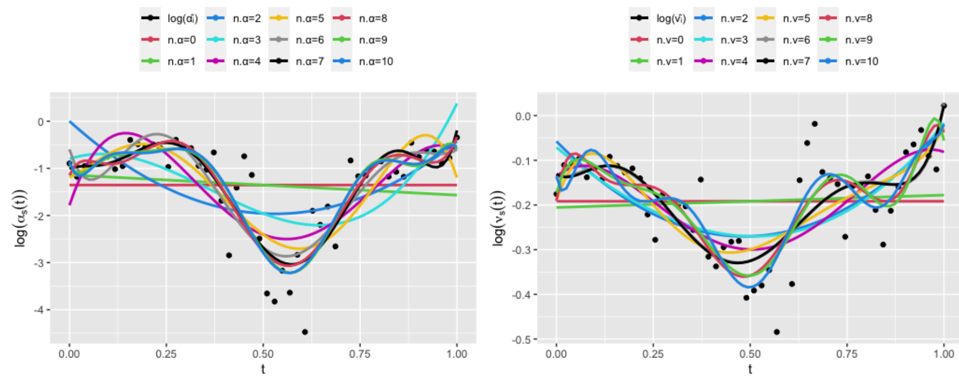


Figure 11: Exploratory plots for the selection on polynomial orders $n.\alpha$ and $n.\nu$ in the data application. The estimates $\log(\hat{\alpha}_i)$ (left) and $\log(\hat{\nu}_i)$ are based on the average over 100 training sets. Colored curves depict polynomial fits of varying degrees.

Table 2: Adjusted R^2 values for polynomial fits of varying degrees ($n.\alpha$ and $n.\nu$) over averaged independent temporal estimates for $\log(\hat{\alpha}_i)$ and $\log(\hat{\nu}_i)$, for the four cases of the simulation study. The first entry in each cell represents the adjusted R^2 for the polynomial fit on independent temporal estimates for $\log(\hat{\alpha}_i)$, while the second entry represents the adjusted R^2 for the polynomial fit on independent temporal estimates for $\log(\hat{\nu}_i)$. The final choices considered in the simulation study are shown in bold.

Case	$n.\alpha = n.\nu = 1$	$n.\alpha = n.\nu = 2$	$n.\alpha = n.\nu = 3$	$n.\alpha = n.\nu = 4$	$n.\alpha = n.\nu = 5$
Case 1	0.068, 0.106	**0.982, 0.994**	0.982, 0.994	0.987, 0.998	0.988, 0.999
Case 2	0.950, 0.920	**0.955, 0.932**	0.954, 0.931	0.953, 0.932	0.950, 0.932
Case 3	0.942, 0.944	0.957, 0.941	**0.990, 0.990**	0.995, 0.990	0.995, 0.998
Case 4	0.194, 0.077	**0.451, 0.548**	0.640, 0.702	0.615, 0.680	0.732, 0.736

Table 3: Adjusted R^2 values for polynomial fits of varying degrees ($n.\alpha$ and $n.\nu$) over averaged independent temporal estimates for $\log(\hat{\alpha}_i)$ and $\log(\hat{\nu}_i)$, for the data application. The first entry in second column represents the adjusted R^2 for the polynomial fit on independent temporal estimates for $\log(\hat{\alpha}_i)$, while the second entry represents the adjusted R^2 for the polynomial fit on independent temporal estimates for $\log(\hat{\nu}_i)$. The final choice considered in the data application is shown in bold.

$n.\alpha = n.\nu$	Adjusted R^2
1	$0.000, 0.000$
2	$0.280, 0.400$
3	$0.388, 0.391$
4	$0.545, 0.441$
5	$\mathbf{0.598, 0.473}$
6	$0.658, 0.511$
7	$0.688, 0.500$
8	$0.686, 0.574$
9	$0.701, 0.574$
10	$0.695, 0.619$

Bibliography

Apanasovich, T. V., M. G. Genton, and Y. Sun (2012). A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components. *Journal of the American Statistical Association 107*, 180–193.

Bhatia, R. and T. Jain (2015). On some positive definite functions. *Positivity 19*(4), 903–910.

Eidsvik, J., B. A. Shaby, B. J. Reich, M. Wheeler, and J. Niemi (2014). Estimation and prediction in spatial models with block composite likelihoods. *Journal of Computational and Graphical Statistics 23*(2), 295–315.

Godambe, V. P. (1960, 12). An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist. 31*(4), 1208–1211.

Stein, M. L., Z. Chi, and L. J. Welty (2004). Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 66*(2), 275–296.

Varin, C., N. Reid, and D. Firth (2011). An overview of composite likelihood methods. *Statistica Sinica*, 5–42.

Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society. Series B (Methodological) 50*(2), 297–312.