Supplementary Material to *Parameter Estimation of Partial Differential Equation Models*

S.1 Calculation of the Penalty Matrix $\mathbf{R}(\boldsymbol{\theta})$

We have that $\mathbf{R}(\boldsymbol{\theta})$ is a $K \times K$ matrix which has $(j, \ell)^{th}$ entry $\int f_j(\mathbf{x}; \boldsymbol{\theta}) f_\ell(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x}$. Using the notation of matrix integration, we write $\mathbf{R}(\boldsymbol{\theta}) = \int \mathbf{f}(\mathbf{x}; \boldsymbol{\theta}) \mathbf{f}^{\mathrm{T}}(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x}$, where $\mathbf{f}(\mathbf{x}; \boldsymbol{\theta}) = \{f_1(\mathbf{x}; \boldsymbol{\theta}), ..., f_K(\mathbf{x}; \boldsymbol{\theta})\}^{\mathrm{T}}$. In our empirical work and simulations based on the PDE model (2), the penalty matrix $\mathbf{R}(\boldsymbol{\theta})$ is the summation of 10 matrix integrals of the same structure, defined as

$$
\begin{aligned}
\mathbf{R}(\boldsymbol{\theta}) &= \int_z \int_t \mathbf{f}(t, z; \boldsymbol{\theta}) \mathbf{f}^{\mathrm{T}}(t, z; \boldsymbol{\theta}) dt dz \\
&= \theta_D^2 \int \frac{\partial^2 \mathbf{b}}{\partial z^2} \frac{\partial^2 \mathbf{b}^{\mathrm{T}}}{\partial z^2} d\mathbf{x} + \theta_S^2 \int \frac{\partial \mathbf{b}}{\partial z} \frac{\partial \mathbf{b}^{\mathrm{T}}}{\partial z} d\mathbf{x} + \theta_A^2 \int \mathbf{b}\mathbf{b}^{\mathrm{T}} d\mathbf{x} \\
&\quad + \theta_D \theta_S \int \left(\frac{\partial^2 \mathbf{b}}{\partial z^2} \frac{\partial \mathbf{b}^{\mathrm{T}}}{\partial z} + \frac{\partial \mathbf{b}}{\partial z} \frac{\partial^2 \mathbf{b}^{\mathrm{T}}}{\partial z^2} \right) d\mathbf{x} + \theta_D \theta_A \int \left(\frac{\partial^2 \mathbf{b}}{\partial z^2} \mathbf{b}^{\mathrm{T}} + \mathbf{b} \frac{\partial^2 \mathbf{b}^{\mathrm{T}}}{\partial z^2} \right) d\mathbf{x} \\
&\quad + \theta_S \theta_A \int \left(\frac{\partial \mathbf{b}}{\partial z} \mathbf{b}^{\mathrm{T}} + \mathbf{b} \frac{\partial \mathbf{b}^{\mathrm{T}}}{\partial z} \right) d\mathbf{x} - \theta_D \int \left(\frac{\partial^2 \mathbf{b}}{\partial z^2} \frac{\partial \mathbf{b}^{\mathrm{T}}}{\partial t} + \frac{\partial \mathbf{b}}{\partial t} \frac{\partial^2 \mathbf{b}^{\mathrm{T}}}{\partial z^2} \right) d\mathbf{x} \\
&\quad - \theta_S \int \left(\frac{\partial \mathbf{b}}{\partial z} \frac{\partial \mathbf{b}^{\mathrm{T}}}{\partial t} + \frac{\partial \mathbf{b}}{\partial t} \frac{\partial \mathbf{b}^{\mathrm{T}}}{\partial z} \right) d\mathbf{x} - \theta_A \int \left(\frac{\partial \mathbf{b}}{\partial t} \mathbf{b}^{\mathrm{T}} + \mathbf{b} \frac{\partial \mathbf{b}^{\mathrm{T}}}{\partial t} \right) d\mathbf{x} + \int \frac{\partial \mathbf{b}}{\partial t} \frac{\partial \mathbf{b}^{\mathrm{T}}}{\partial t} d\mathbf{x} \\
&\triangleq \sum_{\ell=1}^{L} r_\ell(\boldsymbol{\theta}) \mathcal{B}_\ell, \qquad \text{(S.1)}
\end{aligned}
$$

where $L = 10$, $\mathcal{B}_\ell$ are known constant matrices, and $r_\ell(\boldsymbol{\theta})$ are known functions of $\boldsymbol{\theta}$.

We compute $\mathcal{B}_\ell$ for $\ell = 1, ..., 10$ following the same rule. In general, we can use the composite Simpson's rule repeatedly to evaluate the integrals. For a univariate function $\phi(x)$ and an even integer Q, the composite Simpson's rule approximates the integral as

$$
\begin{aligned}
\int_a^b \phi(x) dx &\approx (h/3) \left\{ \phi(x_0) + 2 \textstyle\sum_{q=1}^{Q/2-1} \phi(x_{2q}) + 4 \sum_{q=1}^{Q/2} \phi(x_{2q-1}) + \phi(x_Q) \right\} \\
&= (h/3) \textstyle\sum_{q=0}^{Q} w_q \phi(x_q),
\end{aligned}
$$

where $h = (b - a)/Q$, $x_q = a + qh$, for $q = 0, 1, ..., Q$, are quadrature points, and $(w_0, w_1, w_2, ..., w_{Q-2}, w_{Q-1}, w_Q) = (1, 4, 2, 4, ..., 2, 4, 2, 1)$ assigns weights to quadrature points.

In order to calculate, for example $\mathcal{B}_3$, let Q_1 denote the number of quadrature knots in the time domain, $\mathbf{s}_1 = (t_1, ..., t_{Q_1})$ the vector of knots, and $\mathbf{w}_1$ the vector of weights. Similarly, Q_2, $\mathbf{s}_2 = (z_1, ..., z_{Q_2})$ and $\mathbf{w}_2$ are the number of quadrature knots, knot vector, and weight vector in the range domain. Then the $(i,j)^{th}$ entry $\mathcal{B}_{3,ij}$ is

$$\mathcal{B}_{3,ij} = \int\int b_i(t,z)b_j(t,z)dtdz \approx (h/3)^2 \sum_{k=1}^{Q_2}\sum_{\ell=1}^{Q_1} w_{1,\ell}w_{2,k}b_i(t_\ell,z_k)b_j(t_\ell,z_k). \tag{S.2}$$

Define $\mathbf{W}$ as a diagonal matrix with diagonal elements $\mathbf{w}_1 \otimes \mathbf{w}_2$. Denote the quadrature points by $\mathbf{Z} = \{(t_1,z_1),...,(t_1 z_{Q_2}),...,(t_{Q_1}z_1),...,(t_{Q_1}z_{Q_2})\}^{\mathrm{T}}$, and $\mathbf{B}(\mathbf{Z})$ the matrix of basis function evaluated at the quadrature points. Then the approximation of matrix $\mathcal{B}_3$ can be expressed neatly as $\mathcal{B}_3 \approx \mathbf{B}^{\mathrm{T}}(\mathbf{Z})\mathbf{W}\mathbf{B}(\mathbf{Z})$.

S.2 Implementation of the Variance Estimator

We propose the variance estimate as

$$\widehat{\mathbf{\Sigma}}_{n,\mathrm{prop}} = \mathbf{\Lambda}_n^{-1}(\widehat{\boldsymbol{\theta}})\mathcal{C}(\widehat{\boldsymbol{\theta}})\{\mathbf{\Lambda}_n^{-1}(\widehat{\boldsymbol{\theta}})\}^{\mathrm{T}}, \tag{S.3}$$

where $\mathcal{C}(\widehat{\boldsymbol{\theta}})$ is a matrix whose $(j,k)^{th}$ element is $\widehat{\mathcal{C}}_{jk} = n\widetilde{\lambda}^4\widehat{\sigma}_\epsilon^2\widehat{\boldsymbol{\beta}}^{\mathrm{T}}(\widehat{\boldsymbol{\theta}})\widehat{\mathcal{W}}_j\mathbf{G}_n^{-1}(\widehat{\boldsymbol{\theta}})\mathbf{S}_n\mathbf{G}_n^{-1}(\widehat{\boldsymbol{\theta}})\widehat{\mathcal{W}}_k\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}})$ and $\mathbf{\Lambda}_n(\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^n \partial\Psi_i(\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}^{\mathrm{T}}$. Here $\widehat{\sigma}_\epsilon^2$ is the estimated variance of $\epsilon(\mathbf{x}_i)$ and can be calculated by first fitting a standard spline regression and then forming the residual variance. Also, $\widehat{\mathcal{W}}_j = \widehat{\mathcal{V}}_j + \widehat{\mathcal{V}}_j^{\mathrm{T}}$, where $\widehat{\mathcal{V}}_j = \mathbf{R}(\widehat{\boldsymbol{\theta}})\mathbf{G}_n^{-1}(\widehat{\boldsymbol{\theta}})\mathbf{R}_{j\theta}(\widehat{\boldsymbol{\theta}})$.

The above estimator requires analytic expression of $\partial\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ and $\partial\Psi_i(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. These quantities could be obtained using the implicit function theorem, which is introduced as follows. Dependence on $\boldsymbol{\theta}$ is dropped where appropriate.

To find the first-order derivative of $\widehat{\boldsymbol{\beta}}$ with respect to $\boldsymbol{\theta}$, take total derivative with respect to $\boldsymbol{\theta}$ on both sides of the identity $\partial J(\boldsymbol{\beta}|\boldsymbol{\theta})/\partial\boldsymbol{\beta}|_{\widehat{\boldsymbol{\beta}}} = 0$, we get

$$\frac{d}{d\boldsymbol{\theta}}\left(\left.\frac{\partial J}{\partial\boldsymbol{\beta}}\right|_{\widehat{\boldsymbol{\beta}}}\right) = \left.\frac{\partial^2 J}{\partial\boldsymbol{\theta}^{\mathrm{T}}\partial\boldsymbol{\beta}}\right|_{\widehat{\boldsymbol{\beta}}} + \left.\frac{\partial^2 J}{\partial\boldsymbol{\beta}^{\mathrm{T}}\partial\boldsymbol{\beta}}\right|_{\widehat{\boldsymbol{\beta}}}\frac{\partial\widehat{\boldsymbol{\beta}}}{\partial\boldsymbol{\theta}} = 0.$$

Assuming that $\partial^2 J/\partial\boldsymbol{\beta}^{\mathrm{T}}\partial\boldsymbol{\beta}|_{\widehat{\boldsymbol{\beta}}}$ is non-singular, which is true for our model, we obtain the analytic expression of the first-order derivative of $\widehat{\boldsymbol{\beta}}$ as,

$$\frac{\partial\widehat{\boldsymbol{\beta}}}{\partial\boldsymbol{\theta}} = -\left(\left.\frac{\partial^2 J}{\partial\boldsymbol{\beta}^{\mathrm{T}}\partial\boldsymbol{\beta}}\right|_{\widehat{\boldsymbol{\beta}}}\right)^{-1}\left(\left.\frac{\partial^2 J}{\partial\boldsymbol{\theta}^{\mathrm{T}}\partial\boldsymbol{\beta}}\right|_{\widehat{\boldsymbol{\beta}}}\right). \tag{S.4}$$

It is easily seen from (7) that

$$\frac{\partial^2 J}{\partial\boldsymbol{\beta}^{\mathrm{T}}\partial\boldsymbol{\beta}} = 2\{\mathbf{B}^{\mathrm{T}}\mathbf{B} + \lambda\mathbf{R}(\boldsymbol{\theta})\}, \tag{S.5}$$

and that

$$\frac{\partial^2 J}{\partial\boldsymbol{\theta}^{\mathrm{T}}\partial\boldsymbol{\beta}} = 2\lambda\frac{\partial}{\partial\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\boldsymbol{\beta}\}. \tag{S.6}$$

Substitute the above into (S.4) and we have

$$\frac{\partial \widehat{\boldsymbol{\beta}}}{\partial \boldsymbol{\theta}} = -\lambda\{\mathbf{B}^{\mathrm{T}}\mathbf{B} + \lambda \mathbf{R}(\boldsymbol{\theta})\}^{-1} \left[\frac{\partial}{\partial \boldsymbol{\theta}} \{\mathbf{R}(\boldsymbol{\theta})\boldsymbol{\beta}\} \Big|_{\widehat{\boldsymbol{\beta}}} \right]. \tag{S.7}$$

The first-order derivative of $\Psi_i(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ is, for $i = 1, ..., n$,

$$\frac{\partial \Psi_i}{\partial \boldsymbol{\theta}} = \textstyle\sum_{k=1}^{K} b_k(\mathbf{x}_i)\{Y_i - \mathbf{b}^{\mathrm{T}}(\mathbf{x}_i)\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})\} \dfrac{\partial^2 \widehat{\boldsymbol{\beta}}_k}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}}} - \left(\dfrac{\partial \widehat{\boldsymbol{\beta}}}{\partial \boldsymbol{\theta}} \right)^{\mathrm{T}} \mathbf{b}(\mathbf{x}_i)\mathbf{b}^{\mathrm{T}}(\mathbf{x}_i) \left(\dfrac{\partial \widehat{\boldsymbol{\beta}}}{\partial \boldsymbol{\theta}} \right).$$

To find the second-order derivative of $\widehat{\boldsymbol{\beta}}_k$ with respect to $\boldsymbol{\theta}$, take the second-order total derivative with respect to $\boldsymbol{\theta}$ on both sides of the identity $\partial J / \partial \boldsymbol{\beta}_k|_{\widehat{\boldsymbol{\beta}}_k} = 0$, we get, for $k = 1, ..., K$,

$$\begin{aligned}
\frac{d^2}{d\boldsymbol{\theta}^{\mathrm{T}} d\boldsymbol{\theta}} \left(\frac{\partial J}{\partial \boldsymbol{\beta}_k} \Big|_{\widehat{\boldsymbol{\beta}}_k} \right) &= \frac{d}{d\boldsymbol{\theta}^{\mathrm{T}}} \left\{ \frac{d}{\boldsymbol{\theta}} \left(\frac{\partial J}{\partial \boldsymbol{\beta}_k} \Big|_{\widehat{\boldsymbol{\beta}}_k} \right) \right\} \\
&= \frac{\partial^3 J}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}} \partial \boldsymbol{\beta}_k} \Big|_{\widehat{\boldsymbol{\beta}}_k} + \frac{\partial^3 J}{\partial \boldsymbol{\theta} \partial \boldsymbol{\beta}_k^2} \Big|_{\widehat{\boldsymbol{\beta}}_k} \frac{\partial \widehat{\boldsymbol{\beta}}_k}{\partial \boldsymbol{\theta}^{\mathrm{T}}} + \frac{\partial^2 J}{\partial \boldsymbol{\beta}_k^2} \Big|_{\widehat{\boldsymbol{\beta}}_k} \frac{\partial^2 \widehat{\boldsymbol{\beta}}_k}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}}} \\
&\quad + \frac{\partial^3 J}{\partial \boldsymbol{\beta}_k^3} \Big|_{\widehat{\boldsymbol{\beta}}_k} \frac{\partial \widehat{\boldsymbol{\beta}}_k}{\partial \boldsymbol{\theta}} \frac{\partial \widehat{\boldsymbol{\beta}}_k}{\partial \boldsymbol{\theta}^{\mathrm{T}}} = \mathbf{0}.
\end{aligned}$$

Obviously for our method, we have that $\partial^3 J / \partial \boldsymbol{\beta}_k^3 \equiv 0$, so the last term in the above result disappears. Assuming that $\partial^2 J / \partial \boldsymbol{\beta}_k^2|_{\widehat{\boldsymbol{\beta}}_k} \neq 0$, then the analytic expression for the second-order derivative of $\widehat{\boldsymbol{\beta}}_k$ is obtained as,

$$\frac{\partial^2 \widehat{\boldsymbol{\beta}}_k}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}}} = -\left(\frac{\partial^2 J}{\partial \boldsymbol{\beta}_k^2} \Big|_{\widehat{\boldsymbol{\beta}}_k} \right)^{-1} \left(\frac{\partial^3 J}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}} \partial \boldsymbol{\beta}_k} \Big|_{\widehat{\boldsymbol{\beta}}_k} + \frac{\partial^3 J}{\partial \boldsymbol{\theta} \partial \boldsymbol{\beta}_k^2} \Big|_{\widehat{\boldsymbol{\beta}}_k} \frac{\partial \widehat{\boldsymbol{\beta}}_k}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \right). \tag{S.8}$$

To complete the calculation, we need to know the following quantities,

$$\frac{\partial \mathbf{R}(\boldsymbol{\theta})\boldsymbol{\beta}}{\partial \boldsymbol{\theta}}, \ \frac{\partial^3 J}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}} \partial \boldsymbol{\beta}_k}, \text{ and } \frac{\partial^3 J}{\partial \boldsymbol{\theta} \partial \boldsymbol{\beta}_k^2},$$

all of which involve derivatives of $\mathbf{R}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. The ensuing derivation depends on the particular PDE model of interest.

S.2.1 Implementation with PDE Example (2)

We explain in this section the calculation of above model dependent quantities, in the context of the PDE model (2). We know

$$\mathbf{f}(\mathbf{x}; \boldsymbol{\theta}) = \frac{\partial \mathbf{b}(\mathbf{x})}{\partial t} - \theta_D \frac{\partial^2 \mathbf{b}(\mathbf{x})}{\partial z^2} - \theta_S \frac{\partial \mathbf{b}(\mathbf{x})}{\partial z} - \theta_A \mathbf{b}(\mathbf{x}),$$

where $\mathbf{b}(\mathbf{x}) = \{b_1(\mathbf{x}), ..., b_K(\mathbf{x})\}^{\mathrm{T}}$ is the vector of basis functions. The matrix $\mathbf{R}(\boldsymbol{\theta})$ is shown in (S.1). In this example, the coefficients of matrices $\mathcal{B}_\ell$'s are

$$\begin{aligned}
&r_1(\boldsymbol{\theta}) = \theta_D^2, \quad r_2(\boldsymbol{\theta}) = \theta_S^2, \quad r_3(\boldsymbol{\theta}) = \theta_A^2, \quad r_4(\boldsymbol{\theta}) = \theta_D\theta_S, \quad r_5(\boldsymbol{\theta}) = \theta_D\theta_A, \\
&r_6(\boldsymbol{\theta}) = \theta_S\theta_A, \quad r_7(\boldsymbol{\theta}) = -\theta_D, \quad r_8(\boldsymbol{\theta}) = -\theta_S, \quad r_9(\boldsymbol{\theta}) = -\theta_A, \quad r_{10}(\boldsymbol{\theta}) = 1.
\end{aligned}$$

Then we have

$$\begin{aligned}
\frac{\partial \mathbf{R}(\boldsymbol{\theta})\boldsymbol{\beta}}{\partial \boldsymbol{\theta}} &= \frac{\partial}{\partial \boldsymbol{\theta}} \sum_{\ell=1}^{L} r_\ell(\boldsymbol{\theta})\mathcal{B}_\ell\boldsymbol{\beta} \\
&= \sum_{\ell=1}^{L} \mathcal{B}_\ell\boldsymbol{\beta}\frac{\partial r_\ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}}.
\end{aligned}$$

Notice that $(\partial^2 J/\partial\boldsymbol{\theta}\partial\boldsymbol{\beta}_k)^{\mathrm{T}}$ is the k^{th} row of $\partial^2 J/\partial\boldsymbol{\theta}^{\mathrm{T}}\partial\boldsymbol{\beta}$ given in (S.6). Let $\widetilde{\mathbf{b}}_{\ell,k}$ be the k^{th} row of $\mathcal{B}_\ell$, then $\mathcal{B}_\ell = (\widetilde{\mathbf{b}}_{\ell,1}^{\mathrm{T}}, ..., \widetilde{\mathbf{b}}_{\ell,K}^{\mathrm{T}})^{\mathrm{T}}$. Then, we could write

$$\frac{\partial^2 J}{\partial\boldsymbol{\theta}\partial\boldsymbol{\beta}_k} = 2\lambda \sum_{\ell=1}^{L} \widetilde{\mathbf{b}}_{\ell,k}\boldsymbol{\beta}\frac{\partial r_\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}.$$

Then

$$\frac{\partial^3 J}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}\partial\boldsymbol{\beta}_k} = 2\lambda \sum_{\ell=1}^{L} \widetilde{\mathbf{b}}_{\ell,k}\boldsymbol{\beta}\frac{\partial^2 r_\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}. \tag{S.9}$$

In this simulated example

$$\frac{\partial^2 r_1(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \frac{\partial^2 r_2(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \frac{\partial^2 r_3(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix},$$

and $\frac{\partial^2 r_\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}} \equiv \mathbf{0}$, for $\ell = 4, ..., 10$. Notice that $\partial^2 J/\partial\boldsymbol{\beta}_k^2$ is the k^{th} diagonal element of $\partial^2 J/\partial\boldsymbol{\beta}^{\mathrm{T}}\partial\boldsymbol{\beta}$ given in (S.5), then

$$\frac{\partial^3 J}{\partial\boldsymbol{\theta}\partial\boldsymbol{\beta}_k^2} = 2\lambda \sum_{\ell=1}^{L} \mathcal{B}_\ell(k,k)\frac{\partial r_\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}, \tag{S.10}$$

where $\mathcal{B}_\ell(k,k)$ is the k^{th} diagonal element of the matrix $\mathcal{B}_\ell$.

Finally, substituting $\partial^2 J/\partial\boldsymbol{\beta}_k^2$, (S.9) and (S.10) into (S.8) results in the expression of $\partial^2\widehat{\boldsymbol{\beta}}_k/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}$. The matrices $\mathcal{B}_1, ..., \mathcal{B}_{10}$ are calculated using Simpson's rule, see to **Supplemental Material** Appendix S.1 for detailed calculation.