Published in final edited form as:

Biometrics. 2017 September ; 73(3): 949–959. doi:10.1111/biom.12646.

Estimating Varying Coefficients for Partial Differential Equation Models

Xinyu Zhang¹, Jiguo Cao², and Raymond J. Carroll³

¹Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

²Department of Statistics & Actuarial Science, Simon Fraser University, Burnaby, BC, V5A1S6, Canada

³Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX, 77843-3143, USA, and School of Mathematical and Physical Sciences, University of Technology, Sydney, PO Box 123, Broadway, NSW, 2007, Australia

Summary

Partial differential equations (PDEs) are used to model complex dynamical systems in multiple dimensions, and their parameters often have important scientific interpretations. In some applications, PDE parameters are not constant but can change depending on the values of covariates, a feature that we call *varying coefficients*. We propose a parameter cascading method to estimate varying coefficients in PDE models from noisy data. Our estimates of the varying coefficients are shown to be consistent and asymptotically normally distributed. The performance of our method is evaluated by a simulation study and by an empirical study estimating three varying coefficients in a PDE model arising from LIDAR data.

Keywords

B-splines; Dynamical models; Inverse problems; LIDAR data; Parameter cascading; System identification

1. Introduction

Partial differential equations (PDEs) are used to describe a wide variety of phenomena such as heat, sound, fluid flow and many others. PDEs model these complicated multidimensional dynamical systems by linking some multivariate functions and their partial derivatives together. PDEs are generally much more complicated than ordinary differential equations (ODEs), as ODEs only deal with functions of a single variable. Most PDEs cannot be solved analytically, and many numerical methods have been developed to solve PDEs numerically, for example, the finite element method and finite difference method (Jost, 2002). Parameters in PDEs have important scientific interpretations, and their values determine the behaviour

Supplementary Materials: They include proofs for the asymptotic theory, Figures referenced in Sections 5 and 6, and some additional simulation studies. They are available with this paper at the Biometrics website on Wiley Online Library, as are the computer codes for our simulation studies.

of the dynamical system. However, the values of these parameters are often unknown. There is a need then to estimate these parameters from the measurements or observations of the multidimensional dynamical systems in the presence of measurement errors.

The methods on estimating ODE parameters in the presence of noise have been developed quickly in recent years. Popular methods include the nonlinear least squares approach (Bard (1974), Biegler et al. (1986), Sauer (2006)), the two-step method (Ramsay and Silverman (2005), Chen and Wu (2008), Brunel (2008)), Bayesian methods (Gelman et al. (1996), Huang et al. (2006)), and the parameter cascading method (Ramsay et al. (2007), Cao et al. (2008), Qi and Zhao (2010)). Gugushvili and Klaassen (2012) provided a rigorous analysis of the two-step method. Chen and Wu (2008), Liang et al. (2010), and Cao et al. (2012) proposed the two-step method, the multi-stage smoothing-based approach, and the parameter cascading method, respectively, for estimating time-varying parameters. Xue et al. (2010) considered both numerical error and measurement error and proposed a sieve method for estimating constant and time-varying coefficients in nonlinear ODEs. Cao et al. (2011) proposed a robust estimating method for estimating ODE parameters. Lu et al. (2011) and Wu et al. (2014) proposed multi-stage methods to estimate parameters for high dimensional ODEs. Miao et al. (2014) proposed a generalized ODE model for discrete data. Vujacic et al. (2016) proposed one unifying framework based on generalized Tikhonov regularization and extremum estimation for some existing ODE estimation methods.

The statistical literature on estimating PDE parameters is relatively sparse. Three main methods have been proposed to estimate PDE parameters from noisy data. The first method uses the framework of nonlinear regression by minimizing the distance of the noisy data to the numerical solutions of PDEs for given parameter values (Vallette et al., 1997). Because PDE solutions are very sensitive to parameter values, the nonlinear regression method often has convergence problems, and is computationally costly. This nonlinear regression method also requires the estimation of initial conditions of the PDEs. Müller and Timmer (2002) proposed an extended multiple shooting method to improve the optimization convergence.

The second method is a two-step method (Bär et al., 1999; Voss et al., 1999; Parlitz and Merkwirth, 2000; Coca and Billings, 2000). In the first step, all the state variables and derivatives are estimated from the noisy data by using the multivariate polynomials or nonparametric regression methods such as smoothing splines (Ramsay and Silverman, 2005) or local polynomial regression (Fan and Gijbels, 1996). In the second step, the PDE parameters are estimated by minimizing the distance between the smooth fit and the PDE equations. Both steps have available computer programs to implement, and the computation is also very fast. However, this method depends heavily on the estimation accuracy of derivatives, but derivatives are more difficult to estimate because there are no direct observed data for them. The two-step method also requires that all state variables in the PDEs are observed or measured, which are often hard to achieve in real applications. Müller and Timmer (2004) compared the two-step method with the extended multiple shooting method in their parameter estimates with respect to different noise levels and data resolutions. They concluded that the extended multiple shooting method obtained more accurate parameter estimates than the two-step method, and the superiority of the former method increased with the complexity of the PDE model. The two-step method failed completely when the noise

level is higher than a critical level, due to the difficulty of estimating derivatives. Müller and Timmer (2004) also pointed out that the computational cost of the multiple shooting method is a factor of 1.5×10^3 higher than the two-step method. Dattner and Klaassen (2015) proposed a direct integral approach for estimating ODE parameters and avoided estimating derivatives.

The third method is parameter cascading, which has been shown to obtain more accurate PDE parameter estimates than the two-stage method (Xun et al., 2013). Xun et al. (2013) also established the asymptotic behaviour of the parameter estimates. Frasso et al. (2015) also proposed frequentist and Bayesian approaches for estimating parameters in linear PDEs. Recently, Chkrebtii et al. (2013) considered the uncertainty coming from discretization of system states defined by ODEs or PDEs, and proposed a Bayesian solution to quantify this uncertainty.

The work referenced above assumes that the PDE parameters are constants. In this article, we consider the estimation of PDE parameters which change as functions of covariates, and hence the parameters are varying coefficients. These varying coefficients are often very flexible and easy to interpret, as we show in Section 6. We consider a multidimensional dynamic process $g(\mathbf{x}, w)$, where $\mathbf{x} = (x_1, \dots, x_p)^T \in \mathbb{R}^p$ is a multi-dimensional argument and w is a covariate determining the varying coefficients. Suppose this dynamic process can be modeled with a varying coefficient PDE model

$$\mathcal{F} \left\{ \mathbf{x}, g(\mathbf{x}, w), \frac{\partial g(\mathbf{x}, w)}{\partial x_1}, \dots, \frac{\partial g(\mathbf{x}, w)}{\partial x_p}, \frac{\partial^2 g(\mathbf{x}, w)}{\partial x_1 \partial x_1}, \frac{\partial^2 g(\mathbf{x}, w)}{\partial x_1 \partial x_2}, \dots; \boldsymbol{\theta}(w) \right\} = 0, \quad (1)$$

where $\boldsymbol{\theta}(w) = \{\theta_1(w), \dots, \theta_m(w)\}^T$ is the varying coefficient vector depending on the scalar w . In practice, we do not observe $g(\mathbf{x}, w)$ but instead its surrogate $Y(\mathbf{x}, w)$. We assume that $g(\mathbf{x}, w)$ are observed over a meshgrid of size n , with measurement errors, and that w is observed at $a = 1, \dots, A$ locations. Thus, for $i = 1, \dots, n$ and $a = 1, \dots, A$, we observe the data $(Y_{ai}, \mathbf{x}_{ai}, w_a)$ satisfying

$$Y_{ai} = g(\mathbf{x}_{ai}, w_a) + \varepsilon_{ai}, \quad (2)$$

where the ε_{ai} 's are independent and identically distributed measurement errors and ε_{ai} is assumed here to follow a distribution with mean zero and variance σ^2 . Our goal is to estimate the varying coefficient vector $\boldsymbol{\theta}(w)$ in the PDE model (1) from noisy data, and to quantify the uncertainty of the estimates.

We propose a parameter cascading method to estimate the varying coefficients in PDE models from noisy data (2) in two nested levels of optimization. Without knowing the boundary conditions of the PDEs, it is impossible to obtain the numerical solutions to PDEs. To address this problem, we approximate the PDE solutions using flexible nonparametric functions, which are expressed as linear combinations of basis functions. For given values of

the varying coefficients in PDEs, we estimate the basis coefficients by minimizing the tradeoff between fitting to the data and fidelity to the PDE models in the inner optimization level. Therefore, the estimated basis coefficients, as well as the estimated PDE solution, can be treated as a function of PDE varying coefficients.

Without strong parametric assumptions, we also represent the varying coefficients using flexible nonparametric functions. In the outer optimization level, we estimate the varying coefficients by minimizing the penalized squared errors of the estimated PDE solution to the noisy data in (2). The smoothness of the varying coefficients is controlled by adding a roughness penalty to the squared fitting errors.

The rest of the article is organized as follows. Section 2 introduces our method for estimating varying coefficients in PDEs. Section 3 demonstrates the limiting distribution of the estimates of the varying coefficients. Smoothing parameter selection is described in Section 4. Simulation studies are presented in Section 5 to evaluate the finite performance of our method. Our method is illustrated using LIDAR data in Section 6. Section 7 provides some concluding remarks. Conditions and proofs for the asymptotic theory are provided in the Supplementary Materials, as are the computer codes for our simulation studies.

2. Method for Estimating Varying Coefficients in PDEs

We estimate the varying coefficient vector $\boldsymbol{\alpha}(w)$ in the varying-coefficient PDE model (1) using the parameter cascading method (Cao and Ramsay, 2007, 2010) in two nested levels of optimization. In the inner level of optimization, we estimate the $(p+1)$ -dimensional dynamical process $g(\mathbf{x}, w)$ for any given value of the varying coefficient vector $\boldsymbol{\alpha}(w)$. Hence, the estimated $\hat{g}(\mathbf{x}, w)$ can be viewed as a function of $\boldsymbol{\alpha}(w)$, which is denoted as $\hat{g}\{\mathbf{x}, w, \boldsymbol{\alpha}(w)\}$. In the outer level of optimization, we estimate the varying coefficient vector $\boldsymbol{\alpha}(w)$ by plugging in $\hat{g}\{\mathbf{x}, w, \boldsymbol{\alpha}(w)\}$. The inner level of optimization of $g(\mathbf{x}, w)$ for any given $\boldsymbol{\alpha}(w)$ is embedded inside the outer level of optimization of $\boldsymbol{\alpha}(w)$. The two levels of optimization are implemented iteratively until convergence. The algorithm for the parameter cascading method is also outlined at the end of Section 2.2.

2.1 Inner Level of Optimization: Estimation of $g(\mathbf{x}, w)$

Here we describe the estimation of $g(\mathbf{x}, w)$ in (2) for a given value of $\boldsymbol{\alpha}(w)$. The $(p+1)$ -dimensional dynamical process, $g(\mathbf{x}, w)$, is represented by

$g(\mathbf{x}, w) = \sum_{k=1}^K b_k(\mathbf{x})\beta_k(w) = \mathbf{b}^T(\mathbf{x})\boldsymbol{\beta}(w)$, where $\mathbf{b}(\mathbf{x}) = \{b_1(\mathbf{x}), \dots, b_K(\mathbf{x})\}^T$ is a vector of basis functions depending on $\mathbf{x}$, and $\boldsymbol{\beta}(w) = \{\beta_1(w), \dots, \beta_K(w)\}^T$ is the corresponding vector of basis coefficients depending on w . We further estimate each $\beta_k(w)$ as a linear combination of basis functions

$$\beta_k(w) = \sum_{u=1}^{V_k} \phi_{ku}(w)\alpha_{ku} = \boldsymbol{\phi}_k^T(w)\boldsymbol{\alpha}_k, \quad (3)$$

where $\boldsymbol{\phi}_k(w) = \{\phi_{k1}(w), \dots, \phi_{kV_k}(w)\}^T$ is an additional vector of basis functions, and $\mathbf{a}_k = (\mathbf{a}_{k1}, \dots, \mathbf{a}_{kV_k})^T$ is the corresponding vector of basis coefficients. Let $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^T, \dots, \boldsymbol{\alpha}_K^T)^T$ and $\boldsymbol{\Phi}(w) = \text{diag}\{\phi_1^T(w), \dots, \phi_K^T(w)\}$, where $\text{diag}(A_1, \dots, A_K)$ means a block-diagonal matrix with diagonal elements being $A_1, \dots, A_K$. Then, from (3), $\boldsymbol{\beta}(w) = \boldsymbol{\Phi}(w)\boldsymbol{\alpha}$ and

$$g(\mathbf{x}, w) = \mathbf{b}^T(\mathbf{x})\boldsymbol{\Phi}(w)\boldsymbol{\alpha}. \quad (4)$$

For any given varying coefficient vector, $\boldsymbol{\theta}(w)$, the vector of all basis coefficients $\boldsymbol{\alpha}$ is estimated by minimizing

$$\begin{aligned} J\{\boldsymbol{\alpha}|\boldsymbol{\theta}(w)\} = & \sum_{a=1}^A \sum_{i=1}^n \{Y_{ai} - \mathbf{b}^T(\mathbf{x}_{ai})\boldsymbol{\Phi}(w_a)\boldsymbol{\alpha}\}^2 \\ & + \frac{\lambda_1}{A} \sum_{a=1}^A \int \left[\mathcal{F}\{\mathbf{x}, g(\mathbf{x}, w_a), \frac{\partial g(\mathbf{x}, w_a)}{\partial x_1}, \dots, \frac{\partial g(\mathbf{x}, w_a)}{\partial x_p}, \frac{\partial^2 g(\mathbf{x}, w_a)}{\partial x_1 \partial x_1}, \frac{\partial^2 g(\mathbf{x}, w_a)}{\partial x_1 \partial x_2}, \dots; \boldsymbol{\theta}(w_a)\} \right]^2 d\mathbf{x}, \end{aligned} \quad (5)$$

where the first term measures the fit of (4) to the data, and the second term measures the fidelity of (4) to the PDE model defined in (1). The smoothing parameter λ_1 controls the trade-off between fitting to the data and fidelity to the PDE model.

The estimated basis coefficients $\hat{\mathbf{a}}$ can be obtained for any given varying coefficient vector $\boldsymbol{\theta}(w)$ by minimizing (5), so $\hat{\mathbf{a}}$ can be viewed as a function of $\boldsymbol{\theta}(w)$, which is denoted as $\hat{\mathbf{a}}\{\boldsymbol{\theta}(w)\}$. The $(p+1)$ -dimensional dynamical process can then be estimated as

$$\hat{g}\{\mathbf{x}, w, \boldsymbol{\theta}(w)\} = \mathbf{b}^T(\mathbf{x})\boldsymbol{\Phi}(w)\hat{\mathbf{a}}\{\boldsymbol{\theta}(w)\}. \quad (6)$$

2.2 Outer Level of Optimization: Estimation of $\boldsymbol{\theta}(w)$

The PDE parameter $\boldsymbol{\theta}(w)$ is then estimated in the outer level of optimization. For $k = 1, \dots, m$, write each $\theta_k(w)$ as a linear combination of basis functions

$$\theta_k(w) = \sum_{u=1}^{U_k} \varphi_{ku}(w) c_{ku} = \boldsymbol{\varphi}_k^T(w) \mathbf{c}_k, \quad (7)$$

where $\boldsymbol{\varphi}_k(w) = \{\varphi_{k1}(w), \dots, \varphi_{kU_k}(w)\}^T$ is the vector of basis functions, and $\mathbf{c}_k = (c_{k1}, \dots, c_{kU_k})$ is the vector of basis coefficients. Because the estimator $\hat{\mathbf{a}}\{\boldsymbol{\theta}(w)\}$ is already

regularized, we propose to estimate $\boldsymbol{\theta}(w)$ by minimizing the penalized least squares measure of fit

$$H\{\boldsymbol{\theta}(w)\} = \sum_{a=1}^A \sum_{i=1}^n [Y_{ai} - \hat{g}\{\mathbf{x}_{ai}, w_a, \boldsymbol{\theta}(w_a)\}]^2 + \lambda_2 \sum_{k=1}^m \int \left\{ \frac{d^2 \theta_k(w)}{dw^2} \right\}^2 dw, \quad (8)$$

where λ_2 is a smoothing parameter, which controls the roughness of the estimated varying coefficients $\boldsymbol{\theta}(w)$.

Write $\mathbf{c} = (\mathbf{c}_1^T, \dots, \mathbf{c}_m^T)^T$ and $\boldsymbol{\Psi}(w) = \text{diag}\{\varphi_1^T(w), \dots, \varphi_m^T(w)\}$ such that

$$\boldsymbol{\theta}(w) = \boldsymbol{\Psi}(w)\mathbf{c}. \quad (9)$$

Let $\boldsymbol{\Omega}_k$ be a $U_k \times U_k$ penalty matrix with the $(u_1, u_2)^{\text{th}}$ entry

$$\int \frac{d^2 \varphi_{ku_1}(w)}{dw^2} \frac{d^2 \varphi_{ku_2}(w)}{dw^2} dw$$

and $\boldsymbol{\Omega} = \text{diag}(\boldsymbol{\Omega}_1, \dots, \boldsymbol{\Omega}_m)$. Following (8), we estimate $\mathbf{c}$ by minimizing

$$H(\mathbf{c}) = \sum_{a=1}^A \sum_{i=1}^n [Y_{ai} - \hat{g}\{\mathbf{x}_{ai}, w_a, \boldsymbol{\Psi}(w_a)\mathbf{c}\}]^2 + \lambda_2 \mathbf{c}^T \boldsymbol{\Omega} \mathbf{c}. \quad (10)$$

The function $H(\mathbf{c})$ can be optimized by the Newton-Raphson method, and the gradient and Hessian functions can be calculated analytically. Magnus and Neudecker (2007) is an excellent reference book for matrix differential calculus. Since $\hat{g}\{\mathbf{x}_{ai}, w_a, \boldsymbol{\Psi}(w_a)\mathbf{c}\}$ may have no analytical form, it may be an implicit function of $\mathbf{c}$, in which case the implicit function theorem can be applied to obtain the gradient and Hessian functions analytically. Ramsay et al. (2007) provided some detailed expressions for this optimization procedure for ordinary differential equations, which can be directly extended to partial differential equations.

Denote the minimizer of (10) by $\hat{\mathbf{c}} = (\hat{\mathbf{c}}_1^T, \dots, \hat{\mathbf{c}}_m^T)^T$; the estimates of varying coefficients are then given by $\hat{\theta}_k(w) = \varphi_k^T(w) \hat{\mathbf{c}}_k$ and $\hat{\boldsymbol{\theta}}(w) = \{\hat{\theta}_1(w), \dots, \hat{\theta}_m(w)\}^T$.

Below is the algorithm for the parameter cascading method.

Algorithm 1 Parameter Cascading Method

Outer Level of Optimization - Estimate $\boldsymbol{\theta}(w)$ by minimizing $H\{\boldsymbol{\theta}(w)\}$

- 1 Initialize $\boldsymbol{\theta}(w) = \boldsymbol{\theta}^{(0)}(w)$.
- 2 **Inner Level of Optimization** - Estimate $g(\mathbf{x}, w)$ by minimizing $J\{\mathbf{a}|\boldsymbol{\theta}(w)\}$
 - a. Initialize $\mathbf{a} = \mathbf{a}^{(0)}$.

- b. Update $\mathbf{a}^{(i+1)} = \mathbf{a}^{(i)} - \{(\partial^2 J) / (d\mathbf{a}d\mathbf{a}^T)\}^{-1}(dJ/d\mathbf{a})$ until convergence.
- c. Obtain the estimate $\hat{\mathbf{g}}(\mathbf{x}, w, \boldsymbol{\theta}(w)) = \mathbf{b}^T(\mathbf{x})\Phi(w)\hat{\mathbf{a}}(\boldsymbol{\theta}(w))$.
- 3 Update $\mathbf{c}^{(i+1)} = \mathbf{c}^{(i)} - \{(\partial^2 H)/(d\mathbf{c}d\mathbf{c}^T)\}^{-1}(dH/d\mathbf{c})$ until convergence.
- 4 Obtain the estimate for the varying coefficients: $\hat{\boldsymbol{\theta}}_k(w) = \boldsymbol{\varphi}_k^T(w)\hat{\mathbf{c}}_k$ and $\hat{\boldsymbol{\theta}}(w) = \{\hat{\boldsymbol{\theta}}_1(w), \dots, \hat{\boldsymbol{\theta}}_m(w)\}^T$.

2.3 Specialization to Linear PDE Models

The PDE model (1) can be expressed by substituting (4) into model (1) as

$$\mathcal{F}\{\mathbf{x}, \mathbf{b}^T(\mathbf{x})\Phi(w)\boldsymbol{\alpha}, \{\partial \mathbf{b}^T(\mathbf{x})/\partial x_1\}\Phi(w)\boldsymbol{\alpha}, \dots; \boldsymbol{\theta}(w)\} = 0. \quad (11)$$

When the PDE model (11) is linear in $\boldsymbol{\alpha}$, it can be expressed as

$$\mathcal{F}\left\{\mathbf{x}, \mathbf{b}^T(\mathbf{x})\Phi(w)\boldsymbol{\alpha}, \frac{\partial \mathbf{b}^T(\mathbf{x})}{\partial x_1}\Phi(w)\boldsymbol{\alpha}, \dots; \boldsymbol{\theta}(w)\right\} = \mathbf{f}^T\left\{\mathbf{x}, \mathbf{b}^T(\mathbf{x})\Phi(w), \frac{\partial \mathbf{b}^T(\mathbf{x})}{\partial x_1}\Phi(w), \dots; \boldsymbol{\theta}(w)\right\}\boldsymbol{\alpha} = 0, \quad (12)$$

then $\hat{\mathbf{a}}\{\boldsymbol{\theta}(w)\}$ has a closed form and the algorithm can be stated as follows.

By using (12), the inner level of optimization criterion (5) becomes

$$J\{\boldsymbol{\alpha}|\boldsymbol{\theta}(w)\} = \sum_{a=1}^A \sum_{i=1}^n \{Y_{ai} - \mathbf{b}^T(\mathbf{x}_{ai})\Phi(w_a)\boldsymbol{\alpha}\}^2 + \lambda_1 A^{-1} \sum_{a=1}^A \int \boldsymbol{\alpha}^T \mathbf{f}\left\{\mathbf{x}, \mathbf{b}^T(\mathbf{x})\Phi(w_a), \frac{\partial \mathbf{b}^T(\mathbf{x})}{\partial x_1}\Phi(w_a), \dots; \boldsymbol{\theta}(w_a)\right\} \boldsymbol{\alpha} d\mathbf{x}.$$

Let $N = nA$, $\mathbf{B}_a = \{\mathbf{b}(\mathbf{x}_{1a}), \dots, \mathbf{b}(\mathbf{x}_{na})\}^T$ for $a = 1, \dots, A$, $\mathbf{B}^* = \text{diag}(\mathbf{B}_1, \dots, \mathbf{B}_A)$, $\mathbf{D} = \{\Phi^T(w_1), \dots, \Phi^T(w_A)\}^T$, $\mathbf{B} = \mathbf{B}^*\mathbf{D}$ being a $N \times \sum_{k=1}^K V_k$ basis matrix, $\mathbf{Y} = (Y_{11}, Y_{12}, \dots, Y_{An})^T$, and

$$\begin{aligned} \mathbf{R}\{\boldsymbol{\theta}(w), w\} &= A^{-1} \sum_{a=1}^A \int \mathbf{f}\left\{\mathbf{x}, \mathbf{b}^T(\mathbf{x})\Phi(w_a), \frac{\partial \mathbf{b}^T(\mathbf{x})}{\partial x_1}\Phi(w_a), \dots; \boldsymbol{\theta}(w_a)\right\} \\ &\times \mathbf{f}^T\left\{\mathbf{x}, \mathbf{b}^T(\mathbf{x})\Phi(w_a), \frac{\partial \mathbf{b}^T(\mathbf{x})}{\partial x_1}\Phi(w_a), \dots; \boldsymbol{\theta}(w_a)\right\} d\mathbf{x}. \end{aligned}$$

Then, the inner level of optimization criterion (5) can be expressed in the matrix notation

$$J\{\alpha|\theta(w)\}=(Y-B\alpha)^T(Y-B\alpha)+\lambda_1\alpha^T R\{\theta(w), w\}\alpha. \quad (13)$$

By minimizing (13) for given $\theta(w)$, the estimator for α can be obtained in closed form:

$$\hat{\alpha}\{\theta(w)\}=[B^T B+\lambda_1 R\{\theta(w), w\}]^{-1}B^T Y. \quad (14)$$

In addition, because of (9), we can write $\hat{\alpha}\{\theta(w)\} = \hat{\alpha}(c)$ and $R\{\theta(w), w\} = R(c)$ for simplicity. By substituting into (9), (13) becomes

$$J(\alpha|c)=(Y-B\alpha)^T(Y-B\alpha)+\lambda_1\alpha^T R(c)\alpha, \quad (15)$$

and by substituting in (6), (9) and (14), (10) becomes

$$H(c)=\left[Y-B\{B^T B+\lambda_1 R(c)\}^{-1}B^T Y\right]^T \left[Y-B\{B^T B+\lambda_1 R(c)\}^{-1}B^T Y\right]+\lambda_2 c^T \Omega c. \quad (16)$$

The vector of basis coefficients c can be estimated by minimizing $H(c)$, thus we obtain the estimates for the varying coefficients in the PDE model (1).

3. Limit Distribution of Varying Coefficient Estimator for Linear PDE Models

For simplicity, we denote $c_{11}, c_{12}, \dots, c_m U_m$ by $c_1, c_2, \dots, c_Q$. Let $c_0 = (c_{01}, \dots, c_{0Q})^T$ be the true value of c ; see Condition C.2 in the Supplementary Materials for its definition. Let $\tilde{\lambda}_1 = \lambda_1/N$, $S = N^{-1}B^T B$, $G(c) = S + \tilde{\lambda}_1 R(c)$, $T_1 = E(S)$, and $g = \{g(x_{11}, w_1), g(x_{12}, w_1), \dots, g(x_{An}, w_A)\}^T$. For $q = 1, \dots, Q$, we define $\dot{R}_q(c) = R(c)/c_q$, $\dot{R}_q(c) = R(c)G^{-1}(c)\dot{R}_q(c)$, and $\mathcal{W}_q(c) = \mathcal{V}_q(c) + \mathcal{V}_q^T(c)$. Assume that as $N \rightarrow \infty$, $\tilde{\lambda}_1 R(c)$ converges to a matrix, denoted by $\tilde{R}(c)$. Let $T_2(c) = T_1 + \tilde{R}(c)$, $\alpha(c) = N^{-1}G^{-1}(c)B^T g$, and $\mathcal{C}(c) = \{\mathcal{W}_1(c)\alpha(c), \dots, \mathcal{W}_Q(c)\alpha(c)\}^T$. Also, let $\Lambda(c)$ be a $Q \times Q$ matrix with the $(q_1, q_2)^{th}$ element

$$\Lambda_{q_1 q_2}(c) = \alpha^T(c) \dot{R}_{q_1}(c) \Upsilon_2^{-1}(c) \Upsilon_1 \Upsilon_2^{-1}(c) \dot{R}_{q_2}(c) \alpha(c) + \alpha^T(c) R(c) \Upsilon_2^{-1}(c) \frac{\partial \dot{R}_{q_1}(c)}{\partial c_{q_2}} \alpha(c),$$

where $\Delta = \sigma^2 \Lambda^{-1}(c_0) \mathcal{C}(c_0) \Upsilon_2^{-1}(c_0) \Upsilon_1 \Upsilon_2^{-1}(c_0) \mathcal{C}(c_0)^T \Lambda^{-1}(c_0)$.

In the Supplementary Materials, we show that under Conditions (C.1)-(C4)

$$N^{1/2}(\hat{\mathbf{c}} - \mathbf{c}_0) \rightarrow \text{Normal}(\mathbf{0}, \mathbf{\Delta}), \quad (17)$$

as $N \rightarrow \infty$. Let $\mathbf{\Pi}_k$ be a projection matrix such that $\mathbf{c}_k = \mathbf{\Pi}_k \mathbf{c}$. Then, for any w , we have

$$N^{1/2}\{\hat{\theta}_k(w) - \theta_k(w)\} \rightarrow \text{Normal}\{\mathbf{0}, \boldsymbol{\varphi}_k^T(w) \mathbf{\Pi}_k \mathbf{\Delta} \mathbf{\Pi}_k^T \boldsymbol{\varphi}_k(w)\}, \quad k=1, \dots, m \quad (18)$$

as $N \rightarrow \infty$. To estimate σ^2 , we replace $\mathbf{c}_0$ by $\hat{\mathbf{c}}$ and estimate σ^2 by

$$(N-Q)^{-1} \sum_{a=1}^A \sum_{i=1}^n [Y_{ai} - \hat{g}\{\mathbf{x}_{ai}, w_a, \boldsymbol{\Psi}(w_a) \hat{\mathbf{c}}\}]^2,$$

where Q is the dimension of $\mathbf{c}$.

4. Smoothing Parameter Selection

We propose a two-step procedure for the smoothing parameter selection. In the first step we set $\lambda_1^* = N^{2/5}$ and $\lambda_2^* = N^{-4/5}$, both of which satisfy Condition (C.1) in the Supplementary Materials. For any λ_1 and λ_2 , we rewrite $\hat{\mathbf{c}}$ as $\hat{\mathbf{c}}(\lambda_1, \lambda_2)$ and $\mathbf{c}_0$ as $\mathbf{c}_0(\lambda_1, \lambda_2)$. From (17), we have that as $N \rightarrow \infty$, $N^{1/2}\{\hat{\mathbf{c}}(\lambda_1^*, \lambda_2^*) - \mathbf{c}_0\} \rightarrow \text{Normal}\{\mathbf{0}, \mathbf{\Delta}(\lambda_1^*, \lambda_2^*)\}$.

Let $\hat{\mathbf{\Delta}}(\lambda_1^*, \lambda_2^*)$ be an estimator of $\mathbf{\Delta}(\lambda_1^*, \lambda_2^*)$. In the second step, we select (λ_1, λ_2) by minimizing

$$\begin{aligned} \mathcal{D}(\lambda_1, \lambda_2) &= N\{\hat{\mathbf{c}}(\lambda_1, \lambda_2) - \hat{\mathbf{c}}(\lambda_1^*, \lambda_2^*)\}^T \hat{\mathbf{\Delta}}^{-1}(\lambda_1^*, \lambda_2^*) \{\hat{\mathbf{c}}(\lambda_1, \lambda_2) \\ &\quad - \hat{\mathbf{c}}(\lambda_1^*, \lambda_2^*)\} + N^{-1}[\mathbf{Y} - \mathbf{B}\hat{\boldsymbol{\alpha}}\{\hat{\mathbf{c}}(\lambda_1, \lambda_2)\}]^T [\mathbf{Y} - \mathbf{B}\hat{\boldsymbol{\alpha}}\{\hat{\mathbf{c}}(\lambda_1, \lambda_2)\}]. \end{aligned} \quad (19)$$

The first term of $\mathcal{D}(\lambda_1, \lambda_2)$ in (19) is to ensure that the resulting estimator of $\mathbf{c}_0$ is N -consistent.

Let $\mathcal{U}$ be a candidate set of (λ_1, λ_2) , containing $(\lambda_1^*, \lambda_2^*)$. The estimated smoothing parameters are $(\lambda_1, \lambda_2) = \text{argmin}_{(\lambda_1, \lambda_2) \in \mathcal{U}} \mathcal{D}(\lambda_1, \lambda_2)$. Define $\mathcal{U}^*$ be a subset of $\mathcal{U}$ such that $\hat{\mathbf{c}}(\lambda_1, \lambda_2) - \mathbf{c}_0 = O(N^{-1/2})$ for any $(\lambda_1, \lambda_2) \in \mathcal{U}^*$. Obviously, $\mathcal{U}^*$ is not empty because $(\lambda_1^*, \lambda_2^*)$ belongs to $\mathcal{U}^*$. Assuming $\hat{\mathbf{\Delta}}(\lambda_1^*, \lambda_2^*)$ converges to a positive definite matrix as $N \rightarrow \infty$, we can prove that

$$\Pr\{(\hat{\lambda}_1, \hat{\lambda}_2) \in \mathcal{U}^*\} \rightarrow 1, \quad (20)$$

as $N \rightarrow \infty$. See the Supplementary Materials for the proof of (20). This result means that the estimate $\hat{\mathbf{c}}(\hat{\lambda}_1, \hat{\lambda}_2)$ is N -consistent.

5. Simulations

5.1 Data Generating Mechanism

The following varying coefficient PDE model is used to simulate data

$$\frac{\partial g(t, z, w)}{\partial t} - \theta_D(w) \frac{\partial^2 g(t, z, w)}{\partial z^2} - \theta_S(w) \frac{\partial g(t, z, w)}{\partial z} - \theta_A(w) g(t, z, w) = 0 \quad (21)$$

with varying coefficients $\theta_D(w) = 1 - 0.2(w - 1)^2$, $\theta_S(w) = 0.1 + 0.02 \cos(w\pi)$ and $\theta_A(w) = 0.1 + 0.02 \sin(1.2w\pi)$, where $w \in \{0.1, 0.2, \dots, 2.0\}$.

Model (21) is a PDE of parabolic type in one space dimension, and is also called a (one-dimensional) linear reaction-convection-diffusion equation. Here $\theta_D(\cdot)$ is the diffusion rate, $\theta_S(\cdot)$ is the drift rate, and $\theta_A(\cdot)$ is the reaction rate.

For each w , the PDE model (21) is solved numerically by setting the boundary condition as $g(t, 0, w) = 0$ and the initial condition as $g(0, z, w) = \{1 + 0.1 \times (20 - z)2\}^{-1}$ over a meshgrid in the time domain $t \in [1, 20]$ and the range domain $z \in [1, 40]$.

Thus, for each w , our data is on a 20-by-40 meshgrid in the domain $[1, 20] \times [1, 40]$. The resulting sample size N is $20 \times 20 \times 40 = 16000$. Figure 1 displays the numerical solutions of the PDE (21) when $w = 0.3, 1.0$ and 1.7 .

The observed error-prone data (2) is simulated by adding independent and identically distributed Gaussian noise with standard deviation σ to the PDE solutions. We varied σ , as 0.02 and 0.05. These values are the same as the simulation used in Xun et al. (2013).

5.2 Performance of Parameter Cascading

The parameter cascading method was applied to estimate the three varying coefficients in the PDE model (21) from the simulated data. The two smoothing parameters λ_1 and λ_2 were selected by using the method proposed in Section 4. We used B-splines of order 4 to form the basis functions, with 5, 15, and 3 equally spaced knots in t , z and w domains, respectively. The parameter cascading method was also compared with the two-stage method. The two-stage method estimated varying coefficient parameters for the PDE model in two stages. In the first stage, for each w_a , $a = 1, \dots, A$, $g(\mathbf{x}, w_a)$, and its partial derivatives were estimated by multidimensional penalized signal regression (Marx and Eilers, 2005). In the second stage, similar to (8), we estimated the vector of varying coefficient parameters, $\boldsymbol{\theta}(w_a)$, by minimizing

$$\sum_{a=1}^A \int \left[\mathcal{F} \left\{ \mathbf{x}, g(\mathbf{x}, w_a), \frac{\partial g(\mathbf{x}, w_a)}{\partial x_1}, \dots, \frac{\partial g(\mathbf{x}, w_a)}{\partial x_p}, \frac{\partial^2 g(\mathbf{x}, w_a)}{\partial x_1 \partial x_1}, \frac{\partial^2 g(\mathbf{x}, w_a)}{\partial x_1 \partial x_2}, \dots; \boldsymbol{\theta}(w_a) \right\} \right]^2 d\mathbf{x} + \lambda_3 \sum_{k=1}^m \int \left\{ \frac{d^2 \theta_k(w)}{dw^2} \right\}^2 dw,$$

where $\boldsymbol{\theta}(w_a)$ was estimated as a linear combination of basis functions, as defined in (9), and the basis functions for expanding $\boldsymbol{\theta}(w_a)$ were chosen as the same as the parameter cascading method. The simulation was repeated 1000 times.

Figures S1 and S2 of the Supplementary Materials shows the estimated varying coefficients $\theta_D(w)$, $\theta_S(w)$, and $\theta_A(w)$ for the PDE model (21) from the simulated data in the first 10 simulation replications in our simulation study using the parameter cascading method and the two-stage method, respectively. It is seen that our parameter cascading method can estimate all three varying coefficients reasonably well, but the two-stage method's estimated varying coefficients $\theta_D(w)$ are very different from the true curve. To assess the accuracy of the estimated varying coefficients, we use the squared root of the average squared errors (RASEs), namely

$$\text{RASE}\{\hat{\theta}(w)\} = \left[20^{-1} \sum_{w \in \{0.1, 0.2, \dots, 2\}} \{\hat{\theta}(w) - \theta(w)\}^2 \right]^{1/2}. \quad (22)$$

Table 1 summarizes the mean, median, and standard deviation of the RASEs for the estimated varying coefficients $\hat{\theta}_D(w)$, $\hat{\theta}_S(w)$ and $\hat{\theta}_A(w)$ in 1000 simulation replicates. It shows that our parameter cascading method outperforms the two-stage method. Particularly noticeable is that the two-stage method yields about ten times larger mean, median, and standard deviation of the RASE than the parameter cascading method does for $\theta_D(w)$ and $\theta_S(w)$.

For the parameter cascading method, when the standard deviation of Gaussian noises added to simulated data is increased from 0.02 to 0.05, the mean and median of RASEs for $\hat{\theta}_D(w)$ and $\hat{\theta}_S(w)$ increase around 20% and 71%, respectively, while the mean and median of RASEs for $\hat{\theta}_A(w)$ only increase slightly. The standard deviation of RASEs is doubled for $\hat{\theta}_D(w)$ and $\hat{\theta}_S(w)$ and tripled for $\hat{\theta}_A(w)$ when increasing σ from 0.02 to 0.05.

6. Application

6.1 Background

We have access to a small subset of long range infrared light detection and ranging (LIDAR) data. The data set consists of samples collected for 28 aerosol clouds, 14 of them being biological and the other 14 being non-biological. For each sample, there is a transmitted signal that is sent into the aerosol cloud at 19 different laser wavelengths $w_1, \dots, w_{19}$, and at T time points. For each wavelength and time point, received LIDAR data were observed at equally spaced ranges $z = 1, \dots, Z$. The experiment also included background data which were collected before the aerosol cloud was released. The received LIDAR data from aerosol clouds were then background corrected.

An example of the background-corrected received data for the 13th aerosol cloud (non-biological) is given in Figure 2. Such data are well-described by

$$\frac{\partial g(t, z, w)}{\partial t} - \theta_D(w) \frac{\partial^2 g(t, z, w)}{\partial z^2} - \theta_S(w) \frac{\partial g(t, z, w)}{\partial z} - \theta_A(w) g(t, z, w) = 0, \quad (23)$$

where the parameters $\theta_D(w)$, $\theta_S(w)$ and $\theta_A(w)$ describe the diffusion rate, the drift rate and the reaction rate, respectively, and they are varying with w . In order to estimate the PDE model (23) from the real data, we take $T = 20$ time points and $Z = 60$ range values, so that the sample size N is $20 \times 60 \times 19 = 22800$. We use B-spline basis functions of order 4 constructed with 5 inner knots in the time domain, 17 inner knots in the range domain, and 3 inner knots in the wavelength domain.

6.2 Estimation Results

Figure 3 displays the estimates for the three varying coefficients $\theta_D(w)$, $\theta_S(w)$ and $\theta_A(w)$ for the PDE model (23) from the real data shown in Figure 2. It is obvious that the three varying coefficients are not constant over the range of w . Specifically, the diffusion rate, $\theta_D(w)$, slightly decreases with the early laser wavelengths and then levels off. The drift rate, $\theta_S(w)$, becomes the smallest around $w = 10$, and increases with the laser length thereafter. The reaction rate, $\theta_A(w)$, decreases throughout most of the wavelengths. We used the theoretical result (18) to obtain the 95% confidence intervals for the three varying coefficients, which are also displayed in Figure 3.

Figures 4 and 5 display the means and variances of the estimated varying coefficients $\theta_D(w)$, $\theta_S(w)$ and $\theta_A(w)$ for biological and non-biological aerosol clouds. It is seen that for $\theta_D(w)$ and $\theta_A(w)$, the biological clouds have higher mean varying coefficients than the non-biological clouds in the whole range of wavelengths, while for $\theta_S(w)$, the former has lower mean varying coefficients than the latter in the whole range of wavelengths. For all three coefficients, the non-biological clouds yields larger variances than the biological clouds in the whole range of wavelengths.

We further assessed the equality of the variances of the estimated varying coefficients $\theta_D(w)$, $\theta_S(w)$ and $\theta_A(w)$ for biological and non-biological aerosol clouds by using the Brown-Forsythe test (Brown and Forsythe, 1974), which generally has better control of test level than the Levene's test (Levene, 1960). Figure 5 also shows the pointwise p-values for the Brown-Forsythe test of the equality of the variances of the estimated varying coefficients $\hat{\theta}_D(w)$, $\hat{\theta}_S(w)$ and $\hat{\theta}_A(w)$ for biological and non-biological aerosol clouds. Most p-values are smaller than 0.10 and some of them are even smaller than 0.05. It indicates that the variances of the estimated varying coefficients $\hat{\theta}_D(w)$, $\hat{\theta}_S(w)$ and $\hat{\theta}_A(w)$ for biological and non-biological aerosol clouds are significantly different at most wavelengths. At the significance level of 0.10, the diffusion rate $\theta_D(\cdot)$ shows the significant different variance at the wavelengths in [2,19], while the drift rate $\theta_S(\cdot)$ shows the significant different variance at the wavelengths in [2,6] and [13,15]. The reaction rate $\theta_A(\cdot)$ shows the significant different variance at wavelengths [1,18].

Next, we employ functional principal components analysis to investigate the major variations of the estimated varying coefficients $\hat{\theta}_D(w)$, $\hat{\theta}_S(w)$ and $\hat{\theta}_A(w)$ for all 28 aerosol clouds. Figure S3 of the Supplementary Materials shows the first two functional principal components for the three estimated varying coefficients for all 28 aerosol clouds. The first functional principal component accounts for 91.5%, 98.9%, and 97.4% of total variation of the estimated $\hat{\theta}_D(w)$, $\hat{\theta}_S(w)$ and $\hat{\theta}_A(w)$ for all 28 aerosol clouds, respectively. The first functional principal components for all three estimated varying coefficients change with the wavelength, but they are positive in the whole range of wavelengths. This may be interpreted as that more than 91% of total variation of these estimated varying coefficients comes from the differences of their weighted means, where the weights are defined by the first functional principal components. The second functional principal components account for 7.3%, 0.7%, and 2.1% of total variation of the estimated $\hat{\theta}_D(w)$, $\hat{\theta}_S(w)$ and $\hat{\theta}_A(w)$ for all 28 aerosol clouds, respectively. For all three varying coefficients, the second functional principal components for $\hat{\theta}_D(w)$, $\hat{\theta}_S(w)$ and $\hat{\theta}_A(w)$ are negative when the wavelength is less than 10, 14, 10, respectively, and is positive with larger wavelengths. This may be interpreted as that the second largest source of variations of the three estimated varying coefficients is the change of the varying coefficients from small wavelengths to large wavelengths.

7. Concluding Remarks

We have proposed a parameter cascading method to estimate varying coefficients in PDE models from noisy data. The method employs two nested levels of optimization. In the inner level of optimization, the PDE solution is approximated by a flexible nonparametric function. For any given value of PDE varying coefficients, the nonparametric function is estimated by minimizing the penalized squared errors of the approximated PDE solution to the noisy data, and the fitting penalty is defined by the fit of the nonparametric function to PDEs. In the outer level of optimization, we estimate the varying coefficients as a flexible nonparametric function and add a roughness penalty to control the smoothness of the estimated varying coefficients.

We show that our estimates for varying coefficients are consistent and asymptotically normally distributed. Our simulation study shows that we can obtain accurate estimates for the varying coefficients in PDEs from finite samples in the presence of noises with various scales. We demonstrate our method by estimating three varying coefficients in a PDE model from the LIDAR data. We find that the biological clouds have higher average diffusion rate and reaction rate than the non-biological clouds, but the biological clouds have lower drift rate through the whole range of wavelengths. On the other hand, the non-biological clouds have larger variances for all three varying coefficients than the biological clouds.

Our method allows us to use any multidimensional basis functions to represent the dynamical systems. In our simulation and application, we use tensor product of B-spline basis functions. For this basis function system, the number of basis functions increases quickly with the dimensions of the dynamical system. We believe that we can increase the computational efficiency of our method by defining more efficient multidimensional basis functions. For example, in a bivariate dynamical system, the fast bivariate smoother method

of Xiao et al. (2013) or the spatial spline regression method of Sangalli et al. (2013) are promising in this aspect.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors are very grateful for the very constructive comments from the Editor, an Associate Editor and two reviewers. These comments are extremely helpful for us to improve this work. Zhang's work was partially supported by National Natural Science Foundation of China (Grant numbers 11471324, 71522004 and 71631008) and "Chen Jingrun Future Star" Project. Cao's research was supported by a discovery grant (356044-2013) from the Natural Science and Engineering Research Council of Canada (NSERC). Carroll's research was supported by a grant from the National Cancer Institute (U01-CA057030).

References

- Bär M, Hegger R, Kantz H. Fitting partial differential equations to space-time dynamics. *Physical Review E*. 1999; 59:337–342.
- Bard, Y. *Nonlinear Parameter Estimation*. Academic Press; New York, NY, USA: 1974.
- Biegler L, Damiano JJ, Blau GE. Nonlinear parameter estimation: a case study comparison. *AIChE Journal*. 1986; 32:29–45.
- Brown MB, Forsythe AB. Robust tests for the equality of variances. *Journal of the American Statistical Association*. 1974; 69:364–367.
- Brunel NJ. Parameter estimation of ODE's via nonparametric estimators. *Electronic Journal of Statistics*. 2008; 2:1242–1267.
- Cao J, Fussmann G, Ramsay JO. Estimating a predator-prey dynamical model with the parameter cascades method. *Biometrics*. 2008; 64:959–967. [PubMed: 18047526]
- Cao J, Huang JZ, Wu H. Penalized nonlinear least squares estimation of time-varying parameters in ordinary differential equations. *Journal of Computational and Graphical Statistics*. 2012; 21:42–56. [PubMed: 23155351]
- Cao J, Ramsay J. Parameter cascades and profiling in functional data analysis. *Computational Statistics*. 2007; 22:335–351.
- Cao J, Ramsay J. Linear mixed effects modeling by parameter cascading. *Journal of the American Statistical Association*. 2010; 105:365–374.
- Cao J, Wang L, Xu J. Robust estimation for ordinary differential equation models. *Biometrics*. 2011; 67:1305–1313. [PubMed: 21401565]
- Chen J, Wu H. Efficient local estimation for time-varying coefficients in deterministic dynamic models with applications to HIV-1 dynamics. *Journal of the American Statistical Association*. 2008; 103:369–383.
- Chkrebtii OA, Campbell DA, Calderhead B, Girolami MA. Bayesian solution uncertainty quantification for differential equations. Preprint. 2013; arXiv:1306.2365.
- Coca D, Billings S. Direct parameter identification of distributed parameter systems. *International Journal of Systems Science*. 2000; 31:11–17.
- Dattner I, Klaassen CAJ. Optimal rate of direct estimators in systems of ordinary differential equations linear in functions of the parameters. *Electronic Journal of Statistics*. 2015; 9:1939–1973.
- Fan, J., Gijbels, I. *Local Polynomial Modelling and its Applications*. CRC Press; New York: 1996.
- Frasso G, Jaeger J, Lambert P. Parameter estimation and inference in dynamic systems described by linear partial differential equations. *Advances in Statistical Analysis*. 2015; :1–29. DOI: 10.1007/s10182-015-0257-5

- Gelman A, Bois F, Jiang J. Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *Journal of the American Statistical Association*. 1996; 91:1400–1412.
- Gugushvili S, Klaassen CAJ. n -consistent parameter estimation for systems of ordinary differential equations: bypassing numerical integration via smoothing. *Bernoulli*. 2012; 18:1061–1098.
- Huang Y, Liu D, Wu H. Hierarchical bayesian methods for estimation of parameters in a longitudinal HIV dynamic system. *Biometrics*. 2006; 62:413–423. [PubMed: 16918905]
- Jost, J. *Partial Differential Equations*. Springer-Verlag; New York: 2002.
- Levene, H. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Olkin, I., et al., editors. Stanford University Press; 1960.
- Liang H, Miao H, Wu H. Estimation of constant and time-varying dynamic parameters of HIV infection in a nonlinear differential equation model. *Annals of Applied Statistics*. 2010; 4:460–483. [PubMed: 20556240]
- Lu T, Liang H, Li H, Wu H. High dimensional odes coupled with mixed-effects modeling techniques for dynamic gene regulatory network identification. *Journal of the American Statistical Association*. 2011; 106:1242–1258. [PubMed: 23204614]
- Magnus, JR., Neudecker, H. *Matrix differential calculus with applications in statistics and econometrics*. John Wiley & Sons; New York: 2007.
- Marx B, Eilers P. Multidimensional penalized signal regression. *Technometrics*. 2005; 47:13–22.
- Miao H, Wu H, Xue H. Generalized ordinary differential equation models. *Journal of the American Statistical Association*. 2014; 109:1672–1682. [PubMed: 25544787]
- Müller TG, Timmer J. Fitting parameters in partial differential equations from partially observed noisy data. *Physica D*. 2002; 171:1–7.
- Müller TG, Timmer J. Parameter identification techniques for partial differential equations. *International Journal of Bifurcation and Chaos*. 2004; 14:2053–2060.
- Parlitz U, Merkwirth C. Prediction of spatiotemporal time series based on reconstructed local states. *Physical Review Letters*. 2000; 84:1890–1893. [PubMed: 11017653]
- Qi X, Zhao H. Asymptotic efficiency and finite-sample properties of the generalized profiling estimation of parameters in ordinary differential equations. *The Annals of Statistics*. 2010; 38:435–481.
- Ramsay JO, Hooker G, Campbell D, Cao J. Parameter estimation for differential equations: a generalized smoothing approach (with discussion). *Journal of the Royal Statistical Society, Series B*. 2007; 69:741–796.
- Ramsay, JO., Silverman, BW. *Functional Data Analysis*. Springer; New York: 2005.
- Sangalli L, Ramsay J, Ramsay T. Spatial spline regression models. *J R Stat Soc Ser B Stat Methodol*. 2013; 75:681–703.
- Sauer, T. *Numerical Analysis*. Pearson Addison-Wesley; New York: 2006.
- Vallette D, Jacobs G, Gollub J. Oscillations and spatiotemporal chaos of one-dimensional fluid fronts. *Phys Rev E*. 1997; 55:4274–4287.
- Voss HU, Kolodner P, Abel M, Kurths J. Amplitude equations from spatiotemporal binary-fluid convection data. *Physical Review Letters*. 1999; 83:3422–3425.
- Vujacic I, Mahmoudi SM, Wit E. Generalized Tikhonov regularization in estimation of ordinary differential equations models. *Stat*. 2016; 5:132–143.
- Wu H, Lu T, Xue H, Liang H. Sparse additive ODEs for dynamic gene regulatory network modeling. *JASA*. 2014; 109:700–716. [PubMed: 25061254]
- Xiao L, Li Y, Ruppert D. Fast bivariate p-splines: the sandwich smoother. *Journal of the Royal Statistical Society, Series B*. 2013; 75:577–599.
- Xue H, Miao H, Wu H. Sieve estimation of constant and time-varying coefficients in nonlinear ordinary differential equation models by considering both numerical error and measurement error. *Annals of Statistics*. 2010; 38:2351–2387. [PubMed: 21132064]
- Xun X, Cao J, Mallick B, Maity A, Carroll RJ. Parameter estimation of partial differential equation models. *JASA*. 2013; 108:1009–1020.

Varying Coefficients PDE Models

Figure 1. 3-D plots of the surface of the numerical solution of the PDE (21) when varying $w = 0.3, 1, 1.7$.

Figure 2. Snapshots of the background-corrected received data for the 13th aerosol cloud (non-biological).

Figure 3.

Estimated curves of $\theta_D(w)$ (diffusion rate, top), $\theta_S(w)$ (drift rate, center), and $\theta_A(w)$ (reaction rate, bottom) for the PDE model (21) and their 95% confidence intervals for the 13th aerosol cloud. This figure appears in color in the electronic version of this article.

Figure 4. Mean of the estimated varying coefficients $\hat{\theta}_D(w)$ (diffusion rate, left), $\hat{\theta}_S(w)$ (drift rate, center), and $\hat{\theta}_A(w)$ (reaction rate, right) in the PDE model (21) for 14 biological aerosol clouds (solid lines) and 14 non-biological aerosol clouds (dashed lines). This figure appears in color in the electronic version of this article.

Figure 5.

Sample variance of the estimated varying coefficients $\hat{\theta}_D(w)$ (diffusion rate, top left), $\hat{\theta}_S(w)$ (drift rate, top right), and $\hat{\theta}_A(w)$ (reaction rate, lower left) in the PDE model (21) for 14 biological aerosol clouds (solid lines) and 14 non-biological aerosol clouds (dashed lines). Lower right: p-values to assess the hypothesis of equality of the variances from the two groups of clouds by using the Brown-Forsythe test. Solid line is for estimated curve of $\theta_D(w)$ (diffusion rate), dash-dotted line is for estimated curve of $\theta_S(w)$ (drift rate), and dashed line is for estimated curve of $\theta_A(w)$ (reaction rate). The dotted lines represent $p = 0.10$ and $p = 0.05$. This figure appears in color in the electronic version of this article.

Mean, Median, and Standard Deviation (SD) of the square root of the average squared errors (RASEs) defined in (22) for the estimated varying coefficients $\hat{\theta}_D(w)$, $\hat{\theta}_S(w)$ and $\hat{\theta}_A(w)$ in 1000 simulation replicates by using the proposed parameter cascading method (PC) and the two-stage method (TS). Here σ is the standard deviation of ϵ in (2), $\theta_D(\cdot)$ is the diffusion rate, $\theta_S(\cdot)$ is the drift rate, $\theta_A(\cdot)$ is the reaction rate.

Table 1

		$\sigma = 0.05$		$\sigma = 0.02$	
		PC	TS	PC	TS
RMSE of $\hat{\theta}_D(w)$	Mean $\times 10^2$	1.54	22.76	1.29	15.85
	Median $\times 10^2$	1.53	22.55	1.28	15.88
	SD $\times 10^3$	3.53	40.22	1.58	21.29
RMSE of $\hat{\theta}_S(w)$	Mean $\times 10^3$	1.86	14.94	1.09	10.03
	Median $\times 10^3$	1.85	14.50	1.10	9.90
	SD $\times 10^4$	6.08	43.58	2.78	21.09
RMSE of $\hat{\theta}_A(w)$	Mean $\times 10^3$	1.88	3.46	1.87	2.22
	Median $\times 10^3$	1.88	3.38	1.87	2.19
	SD $\times 10^4$	0.30	7.19	0.11	1.85