JUSTC
Article
http://justc.ustc.edu.cn
Received: January 21, 2022; Accepted: February 23, 2022

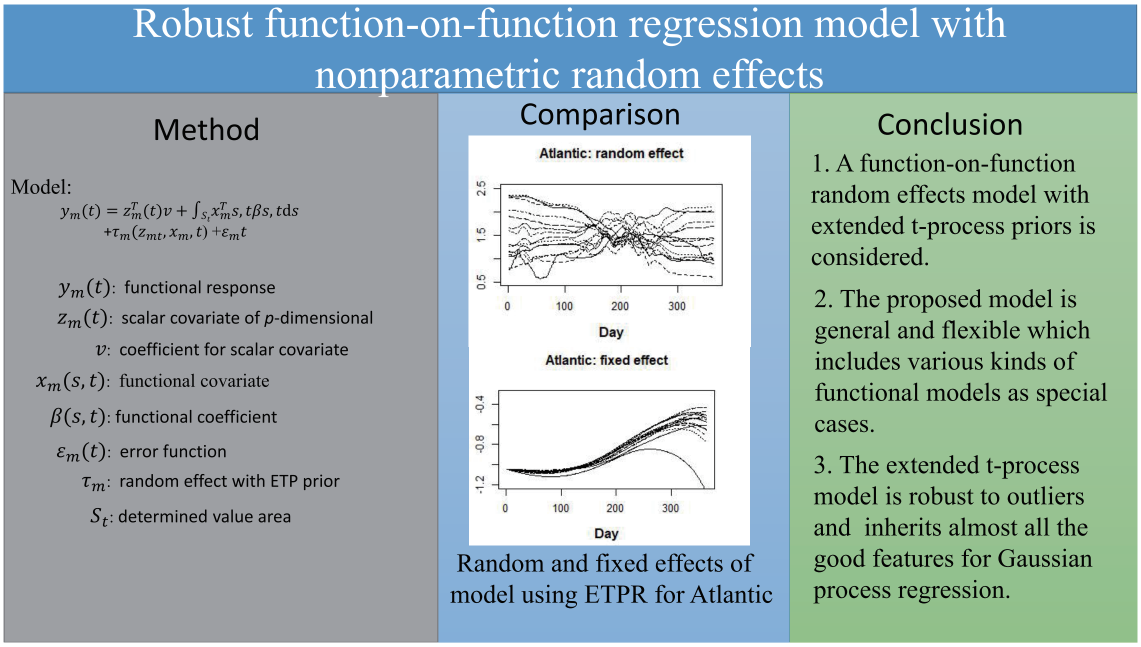
# Robust function-on-function regression model with nonparametric random effects

Shanshan Wang, Hao Ding ✉, and Zhanfeng Wang

*Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei 230026, China*

✉Correspondence: Hao Ding, Email: dinghao@ustc.edu.cn


## Graphical abstract




## Public summary

■ A function-on-function random effects model with extended $t$-process priors is considered.

■ The proposed model is general and flexible which includes various kinds of functional models as special cases.

■ The extended $t$-process model is robust to outliers and inherits almost all the good features for Gaussian process regression.

**Citation:** Wang S S, Ding H, Wang Z F. Robust function-on-function regression model with nonparametric random effects. *JUSTC*, **2022**, 52(4): 5. DOI: 10.52396/JUSTC-2022-0016

JUSTC
Article
http://justc.ustc.edu.cn
Received: January 21, 2022; Accepted: February 23, 2022

# Robust function-on-function regression model with nonparametric random effects

Shanshan Wang, Hao Ding ✉, and Zhanfeng Wang

*Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei 230026, China*

✉Correspondence: Hao Ding, Email: dinghao@ustc.edu.cn



Cite This: *JUSTC*, **2022**, 52(4): 5 (7pp)
Read Online

**Abstract:** Extended $t$-process is robust to outliers and inherits many attractive properties from the Gaussian process. In this paper, we provide a function-on-function nonparametric random-effects model using extended $t$-process priors in which we consider heterogeneity of individual effect, flexible mean function, nonparametric covariance function and robustness. A likelihood-based estimation procedure is constructed to estimate parameters involved in the model. Information consistency for the parameter estimation is provided. Simulation studies and a real data example are further investigated to evaluate the performance of the developed procedures.

**Keywords:** extended $t$-process regression; nonlinear random effects; covariance kernel function; robustness

**CLC number:** O213 **Document code:** A

## 1 Introduction

As the development of science and technology, some data sets are recorded frequently with curves, surfaces and other types, which are usually called functional data that plays an important role in wide fields such as atmospheric science, engineering, medical research, see more details in Ramsay and Silverman[1]. Functional regression models are useful tools in functional data analysis, where one of the most interesting and challenging cases is function-on-function regression, see Ramsay and Silverman[1,2], Yao et al.[3, 4]. In this paper, we consider the following functional model proposed by Wang et al.[5], for $m = 1, \cdots, M$,

$$y_m(t) = z_m^\top(t)\nu + \int_{S_t} x_m^\top(s,t)\beta(s,t)\mathrm{d}s + \tau_m(z_m(t), x_m(\cdot,t)) + \varepsilon_m(t) \tag{1}$$

where $y_m(t)$ is the functional response, $z_m(t)$ is a $p$-vector of functional covariates, $\nu$ is the corresponding parameters, $x_m(s,t)$ is a $q$-dimensional of covariates depends on $s$ and $t$, and $\beta(s,t)$ is a vector of the functional coefficients, $S_t$ is interval for $t$, $\varepsilon_m(t)$ is random error term for the $m$th curve. Model (1) is flexible, and includes some function-on-function models in Gervini[6], Malfait and Ramsay[7], Ramsay and Silverman[2], as special cases. Note that $\tau_m$ is used to model the heterogeneity among the different subjects, which depends on $z_m(t)$, $x_m(\cdot,t)$. Wang et al.[5] considered the above random effects model using Gaussian process priors. More on Gaussian process priors in functional model[8,9].

However, when there exist outliers in the observations, it is not robust to use the model based on Gaussian process priors, see e.g. Wang et al.[10]. Then in order to overcome the influence of outliers, various forms of student t-process have been developed to model a heavy-tailed process, e.g. Yu et al.[11], Zhang and Yeung[12]. Shah et al.[13] pointed out that the t-distribution under addition is not closed to maintain the good properties of Gaussian models. Thus, Wang et al.[10] developed an extended $t$-process regression, which has the following advantages: ① it can maintain the good properties of Gaussian process; ② it has flexible forms, and contains model in Shah et al.[13] as a special case; ③ it is robust. More general discussions on $t$-process can see Refs. [10,14].

In this paper, we consider a functional nonparametric random effects model with extended t-process priors, and propose an estimation procedure. The proposed method has 3 merits. ① It applies the extended $t$-process prior to model the heterogeneity of individual effect in the function-on-function regression model such that the model has robustness; ② A basis expansion smoothing method and a penalized likelihood method are developed to estimate the parameter in the fixed effect and covariance function of random effects, which leads to estimation of the smoothing function and prediction of the random effect; ③ Information consistency of the parameter estimation is obtained.

The remainder of the paper is organized as follows. In Section 2, we present the nonparametric random effects model using extended $t$-process priors, and develop prediction distribution and estimation procedure. In Section 3, we conduct simulation studies and a real data example to evaluate the performance of the proposed method. The conclusions are given in Section 4. All the proofs are given in Appendix.

## 2 Main results

### 2.1 Extended *t*-process

Extended $t$-process proposed by Wang et al.[10] is briefly introduced as follows. Let $f(\cdot)$, a real-valued random function from $\mathcal{X}$ to $R$, satisfy that

$$f \mid r \sim \mathrm{GP}(h, rk), \quad r \sim \mathrm{IG}(v, \omega) \ ,$$

where $\mathrm{GP}(\cdot,\cdot)$ and $\mathrm{IG}(\cdot,\cdot)$ stand for Gaussian process and inverse gamma distribution respectively. Then $f$ follows an extended $t$-process (ETP), and can be denoted by $f \sim \mathrm{ETP}(v, \omega, h, k)$. We call $h(\cdot): \mathcal{X} \to R$ mean function and $k(\cdot,\cdot): \mathcal{X} \times \mathcal{X} \to R$ covariance kernel function. From the definition of ETP, we show that for any points $X = (x_1, \cdots, x_n)^\top$, we have

$$f_n = f(X) = (f(x_1), \cdots, f(x_n))^\top \sim \mathrm{EMTD}(v, \omega, h_n, K_n),$$

meaning that $f_n$ has an extended multivariate $t$-distribution (EMTD) with the following density function,

$$p(z) = |2\pi\omega K_n|^{-1/2} \frac{\Gamma(n/2+v)}{\Gamma(v)} \left(1 + \frac{(z-h_n)^\top K_n^{-1}(z-h_n)}{2\omega}\right)^{-(n/2+v)},$$

where $h_n = (h(x_1), \cdots, h(x_n))^\top, K_n = (k_{ij})_{n\times n}$ and $k_{ij} = k(x_i, x_j)$.

### 2.2 Function-on-function regression model with random effects

In model (1), the random effect $\tau_m$ depicts individual effect. Considering robustness against outliers, an ETP process prior is applied to $\tau_m$. This paper assumes that $\tau_m$ and $\varepsilon_m$ have a joint extended $t$-process,

$$\begin{pmatrix} \tau_m \\ \varepsilon_m \end{pmatrix} \sim \mathrm{ETP}\left(v, \omega, \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} k & 0 \\ 0 & \sigma^2\delta_\varepsilon \end{pmatrix}\right),$$

where $\delta_\varepsilon(t,s) = I(t=s)$ and $I(\cdot)$ is an indicator function.

Note that the random effect $\tau_m$ relies on $\boldsymbol{z_m}(t)$ and $\boldsymbol{x}_m(\cdot,t)$, then following Wang et al.[5], the kernel function $k$ is an expression as

$$\mathrm{Cov}(\tau_m(u_m(t_1)), \tau_m(u_m(t_2))) = k(u_m(t_1), u_m(t_2)) = k_{\boldsymbol{\theta}}(u_m(t_1), u_m(t_2)) =$$
$$\theta_{10} \exp\left\{ -\sum_{i=1}^{p} \frac{\theta_{1i}(z_{mi}(t_1) - z_{mi}(t_2))^2}{2} - \sum_{i=1}^{q} \frac{\theta_{1,p+i} \|x_{mi}(\cdot,t_1) - x_{mi}(\cdot,t_2)\|_\Lambda}{2} \right\} +$$
$$\sum_{i=1}^{p} \theta_{2i} z_{mi}(t_1) z_{mi}(t_2) + \sum_{i=1}^{q} \theta_{2,p+i} \int x_{mi}(s,t_1) x_{mi}(s,t_2)\mathrm{d}s,$$

where $u_m(t) = (z_m^\top(t), x_m^\top(\cdot,t))^\top$, $z_m(t) = (z_{m1}(t), \cdots, z_{mp}(t))^\top$ and $x_m(s,t) = (x_{m1}(s,t), \cdots, x_{mq}(s,t))^\top$. Let $\boldsymbol{\theta} = (\theta_{10}, \theta_{11}, \cdots, \theta_{1Q}, \theta_{21}, \cdots, \theta_{2Q})^\top$ represent a set of hyper-parameters with $Q = p+q$, and $\|g(\cdot)\|_\Lambda$ be a $\Lambda$ norm of function $g$. A choice of $\|\cdot\|_\Lambda$ is the $L_2$ norm of a function, that is, $\|g(\cdot)\|_\Lambda = \int g(s)^2 \mathrm{d}s$ is a $\Lambda$ norm of function $g$.

Let observations $\{y_{mi} = y_m(t_i),\ \ i = 1, \cdots, n, m = 1, \cdots, M\}$, $u_m(t_i) = (z_m^\top(t_i), x_m^\top(\cdot, t_i))^\top$, error term $\varepsilon_{mi} = \varepsilon_m(t_i)$, where $\{t_i\}$ are observed times. Assume that true values of $v$, $\beta$, $\tau_m$ in model (1) are $v_0$, $\beta_0$, $\tau_{0m}$ respectively. From model (1), we further consider the following (true) data model:

$$y_{mi} = z_m^\top(t_i) v_0 + \int_{S_t} x_m^\top(s,t_i)\beta_0(s,t_i)\mathrm{d}s + \tau_{0m}(z_m(t_i), x_m(\cdot,t_i)) + \varepsilon_{mi} \quad (2)$$

This paper aims to develop methods to estimate $v_0$, $\beta_0$, and predict $\tau_{0m}$.

### 2.3 Prediction

Denote $c_m(t) = z_m^\top(t)v + \int_{S_t} x_m^\top(s,t)\beta(s,t)\mathrm{d}s$. From model (1), we have the following results,

$$y_m \mid c_m, \tau_m, \theta, \sigma^2 \sim \mathrm{ETP}(v, w, c_m + \tau_m, \sigma^2\delta_\varepsilon),$$
$$\tau_m \mid \theta \sim \mathrm{ETP}(v, w, 0, k_\theta),$$
$$y_m \mid c_m, \boldsymbol{\theta}, \sigma^2 \sim \mathrm{ETP}(v, w, c_m, k_{\boldsymbol{\theta}} + \sigma^2\delta_\varepsilon).$$

It follows that for the observed data, we have the conditional distributions,

$$\left.\begin{aligned} y_m \mid c_m, \tau_m, \theta, \sigma^2 &\sim \mathrm{EMTD}(v, w, c_m + \tau_m, \sigma^2 I) \\ \tau_m \mid \theta &\sim \mathrm{EMTD}(v, w, \mathbf{0}, K_m) \\ y_m \mid c_m, \theta, \sigma^2 &\sim \mathrm{EMTD}(v, w, c_m, K_m + \sigma^2 I) \end{aligned}\right\} \quad (3)$$

where $y_m = (y_m(t_1), \cdots, y_m(t_n))^\top$ are observations for the $m$th subject at points $\{t_1, \cdots, t_n\}$, similarly, $\tau_m = (\tau_m(u_m(t_1)), \cdots, \tau_m(u_m(t_n)))^\top$, $c_m = (c_m(t_1), \cdots, c_m(t_n))^\top$, $K_m = (k_\theta(u_m(t_i), u_m(t_j)))_{n\times n}$, $\boldsymbol{I}$ is the identity matrix.

Denoted by the data set $\mathcal{D} = \{(y_m(t_j), u_m(t_j)) : j = 1, \cdots, n, m = 1, \cdots, M\}$. Since that

$$\begin{pmatrix} y_m \\ \tau_m \end{pmatrix} \Big| u_m \sim \mathrm{ETMD}\left(v, w, (c_m^\top, \mathbf{0}^\top)^\top, \begin{pmatrix} K_m + \sigma^2 I & K_m \\ K_m & K_m \end{pmatrix}\right),$$

we obtain the posterior distribution of $\boldsymbol{\tau}_m$, that is

$$\tau_m \mid \mathcal{D} \sim \mathrm{EMTD}(v^*, w^*, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m),$$

where $v^* = v + n/2$, $w^* = w + n/2$,

$$\mu_m = K_m(K_m + \sigma^2 I)^{-1}(y_m - c_m),$$
$$\sum\nolimits_m = s_{0m}\{K_m - K_m(K_m + \sigma^2\boldsymbol{I})^{-1}K_m\},$$

and

$$s_{0m} = \frac{2w + (y_m - c_m)^\top (K_m + \sigma^2\boldsymbol{I})^{-1}(y_m - c_m)}{2\omega + n}.$$

For prediction, at a new data point $t^*$, we have

$$\begin{pmatrix} y_m \\ \tau_m(u_m(t^*)) \end{pmatrix} \Big| u_m \sim$$
$$\mathrm{ETMD}\left(v, w, (c_m^\top, 0)^\top, \begin{pmatrix} K_m + \sigma^2 I & k_{mt^*} \\ k_{mt^*}^\top & k(u_m(t^*), u_m(t^*)) \end{pmatrix}\right),$$

where $k_{mt} = (k(u_m(t), u_m(t_1)), \cdots, k(u_m(t), u_m(t_n)))^\top$. It indicates that

$$\tau_m(u_m(t^*)) \mid \mathcal{D} \sim \mathrm{EMTD}(v^*, w^*, \mu_m^*, \Sigma_m^*),$$

where

$$\mu_m^* = k_{mt^*}^\top (K_m + \sigma^2 I)^{-1}(y_m - c_m),$$
$$\sum\nolimits_m^* = s_{0m}\{k(u_m(t^*), u_m(t^*)) - k_{mt^*}^\top (K_m + \sigma^2 I)^{-1} k_{mt^*}\}.$$

Therefore, we can use posterior mean

$$E(y_m(t^*) \mid \mathcal{D}) = \mu_m^* + c_m(t^*)$$

to predict $y_m(t^*)$, denoted by $\hat{y}_m(t^*)$. And using

$$\mathrm{Var}(y_m(t^*) \mid \mathcal{D}) = \Sigma_m^* + s_{0m}\sigma^2$$

estimate variance of $\hat{y}_m(t^*)$.

Moreover, we show $\tau_m \mid \mathcal{D} \sim \mathrm{EMTD}(v^*, w^*, \tilde{\mu}_m, \tilde{\sigma}_m)$, where for data points $u$, $v$,

$$\tilde{\mu}_m(u) = \boldsymbol{k}_{mu}^\top (\boldsymbol{K}_m + \sigma^2\boldsymbol{I})^{-1}(\boldsymbol{y}_m - \boldsymbol{c}_m),$$
$$\tilde{\sigma}_m(u,v) = s_{0m}\{k(u_m(u), u_m(v)) - k_{mu}^\top (K_m + \sigma^2 I)^{-1} k_{mv}\}.$$

Similarly,


DOI: 10.52396/JUSTC-2022-0016
JUSTC, 2022, 52(4): 5

$$\begin{aligned} E(y_m(u)\mid \mathcal{D}) &= \tilde{\mu}_m(u)+c_m(u), \\ \mathrm{Cov}(y_m(u),y_m(v)\mid \mathcal{D}) &= \tilde{\sigma}_m(u,v)+s_{0m}\sigma^2 I(u=v). \end{aligned} \tag{4}$$

It follows that Eq. (4) is an estimation of the covariance function of $\hat{y}_m(\cdot)$.

### 2.4 Parameter estimation

Note that $\beta(s,t)$ in model (1) is a smooth function and can be approximated based on basis functions $\{\phi_k(s),k=1,\cdots,K_s\}$, and $\{\psi_k(s),k=1,\cdots,K_t\}$,

$$\beta(s,t)=\sum_{k=1}^{K_s}\sum_{l=1}^{K_t}\begin{pmatrix} b_{1kl}\\ \vdots \\ b_{qkl}\end{pmatrix}\phi_k(s)\psi_l(t)=\begin{pmatrix}\phi(s)^\top \boldsymbol{B}_1\psi(t)\\ \vdots \\ \phi(s)^\top \boldsymbol{B}_q\psi(t)\end{pmatrix},$$

where $\{b_{ikl}\}$ are coefficients, $\boldsymbol{B}_i=(b_{ikl})_{K_s\times K_t}$, $\phi(s)=(\phi_1(s),\cdots,\phi_{K_s}(s))^\top$,$\psi(t)=(\psi_1(t),\cdots,\psi_{K_t}(t))^\top$. Let $\boldsymbol{\phi}_{xmi}(t)=\int_{S_t}\phi(s)x_{mi}(s,t)\,\mathrm{d}s$,and

$$\begin{aligned} \boldsymbol{\gamma}_m(t) &= (z_m^\top(t),(\psi(t)\otimes\boldsymbol{\phi}_{xm1}(t))^\top,\cdots,(\psi(t)\otimes\boldsymbol{\phi}_{xmq}(t))^\top)^\top, \\ b &= (v^\top,\mathrm{Vec}(\boldsymbol{B}_1)^\top,\cdots,\mathrm{Vec}(\boldsymbol{B}_q)^\top)^\top, \end{aligned}$$

where "⊗" represents the Kronecker product. Hence, $c_m(t)=\boldsymbol{\gamma}_m(t)^\top\boldsymbol{b}$. Let $\boldsymbol{\Gamma}_{mn}=(\boldsymbol{\gamma}_m(t_1),\cdots,\boldsymbol{\gamma}_m(t_n))$, then $c_m=\boldsymbol{\Gamma}_{mn}^\top b$.

Next we estimate $\boldsymbol{\theta}$, $\boldsymbol{b}$ and $\boldsymbol{\sigma}^2$ via using a likelihood method. By Eq. (3), we obtain a likelihood function of $y_m$,

$$f(\mathcal{D}\mid\theta,b,\sigma^2)=\prod_{m=1}^{M}f(y_m\mid\theta,b,\sigma^2)= \prod_{m=1}^{M}|2\pi\omega(K_m+\sigma^2 I)|^{-1/2}\frac{\Gamma(n/2+v)}{\Gamma(v)}\{1+\frac{H_m(\theta,b,\sigma^2)}{2w}\}^{-(n/2+v)},$$

where $H_m(\theta,b,\sigma^2)=(y_m-\boldsymbol{\Gamma}_{mn}^\top b)^\top(K_m+\sigma^2 I)^{-1}(y_m-\boldsymbol{\Gamma}_{mn}^\top b)$. Then we have the following objective function based on the log-likelihood function,

$$l(\theta,b,\sigma^2)=\sum_{m=1}^{M}[\log|K_m+\sigma^2 I|+(n+2v)\log\{2w+H_m(\theta,b,\sigma^2)\}].$$

Due to the smoothness of $\beta(\cdot,\cdot)$, following from Ramsay and Silverman[5], we consider the following penalty functions,

$$\begin{aligned} \mathrm{Pen}_s(\beta(s,t)) &= \int_a^b\int_a^b\|\boldsymbol{L}_s(\beta(s,t))\|^2\mathrm{d}s\mathrm{d}t=\sum_{i=1}^{q}\mathrm{tr}(\boldsymbol{B}_i^\top\boldsymbol{L}_{\phi\phi}\boldsymbol{B}_i\boldsymbol{J}_{\psi\psi}), \\ \mathrm{Pen}_t(\beta(s,t)) &= \int_a^b\int_a^b\|\boldsymbol{L}_t(\beta(s,t))\|^2\mathrm{d}s\mathrm{d}t=\sum_{i=1}^{q}\mathrm{tr}(\boldsymbol{B}_i^\top\boldsymbol{J}_{\phi\phi}\boldsymbol{B}_i\boldsymbol{L}_{\psi\psi}), \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{L}_{\phi\phi} &= \int_a^b[\boldsymbol{L}_s\phi(s)][\boldsymbol{L}_s\phi(s)^\top]\mathrm{d}s,\quad \boldsymbol{J}_{\psi\psi}=\int_a^b\psi(t)\psi(t)^\top\mathrm{d}t, \\ \boldsymbol{L}_{\psi\psi} &= \int_a^b[\boldsymbol{L}_t\psi(t)][\boldsymbol{L}_t\psi(t)^\top]\mathrm{d}t,\quad \boldsymbol{J}_{\phi\phi}=\int_a^b\phi(s)\phi(s)^\top\mathrm{d}s. \end{aligned}$$

Therefore, we develop an objective function,

$$G(\theta,b,\sigma^2)=l(\theta,b,\sigma^2)+\lambda_s\mathrm{Pen}_s(\beta(s,t))+\lambda_t\mathrm{Pen}_t(\beta(s,t)),$$

where $\lambda_s$ and $\lambda_t$ are tuning parameters. Take the derivative of $G(\theta,b,\sigma^2)$ with respect to $\boldsymbol{b}$, we can obtain the estimation equation

$$(n+2v)\sum_{m=1}^{M}\frac{\boldsymbol{\Gamma}_{mn}(K_m+\sigma^2\boldsymbol{I})^{-1}(y_m-\boldsymbol{\Gamma}_{mn}^\top\boldsymbol{b})}{2w+H_m(\boldsymbol{\theta},\boldsymbol{b},\sigma^2)}=\boldsymbol{\Lambda}\boldsymbol{b},$$

where $\boldsymbol{\Lambda}=\mathrm{diag}(\boldsymbol{0}_{p\times p},\lambda_s\boldsymbol{J}_{\psi\psi}\otimes\boldsymbol{L}_{\phi\phi}+\lambda_t\boldsymbol{L}_{\psi\psi}\otimes\boldsymbol{J}_{\phi\phi},\cdots,\lambda_s\boldsymbol{J}_{\psi\psi}\otimes\boldsymbol{L}_{\phi\phi}+\lambda_t\boldsymbol{L}_{\psi\psi}\otimes\boldsymbol{J}_{\phi\phi})$ is a $(p+qK_sK_t)\times(p+qK_sK_t)$ matrix. Similarly, we can get estimation equations with respect to $\theta$ and $\sigma^2$.

From these estimation equations, we construct an estimation procedure as follows.

Step 1 Given an initial estimate of $\boldsymbol{\theta}$;

Step 2 Given $\boldsymbol{\theta}$, we update the estimates of $\boldsymbol{b}$ and $\sigma^2$ via

$$\arg\min_{\boldsymbol{b},\sigma^2}G(\boldsymbol{\theta},\boldsymbol{b},\sigma^2);$$

Step 3 Given $\boldsymbol{b}$ and $\sigma^2$, we update the estimate of $\boldsymbol{\theta}$ via

$$\arg\min_{\theta}l(\theta,b,\sigma^2);$$

Step 4 Repeat Step 2 and Step 3 until convergence.

Similar to Ref. [5], when the absolute value of relative difference of $l(\theta,b,\sigma^2)$ between two successive iterations is less than a given value, the procedure stops.

### 2.5 Information consistency

The common mean structure and its properties have been studied a lot in functional models, see Yao et al.[4], Yuan and Cai[15], Sun et al.[16], and among others. Next we only consider the information consistency. Let $\mathcal{X}=\mathcal{X}_1\times\mathcal{X}_2$, where $\mathcal{X}_1$ and $\mathcal{X}_2$ are spaces covariates $\boldsymbol{z}_m(t)$ and $\boldsymbol{x}_m(\cdot,t)$ belonging to. Let $p_{\sigma_0}(y_m|\tau_{0m},u_m)$ be the density function to generate the data $y_m$ given $u_m$ and $\tau_{0m}$, where $\sigma_0$ is the true value of $\sigma$, $\tau_{0m}$ is the true value of $\tau_m$. Let $p_\theta(\tau)$ be a measurement of the random process $\tau$ on space $\mathcal{F}=\{\tau(\cdot,\cdot):\mathcal{X}\to\mathcal{R}\}$. Let

$$p_{\sigma,\theta}(y_m\mid u_m)=\int_{\mathcal{F}}p_\sigma(y_m\mid\tau,u_m)\,\mathrm{d}p_\theta(\tau),$$

be the density function to generate the data $y_m$ given $u_m$ under model (1). Let $p_{\sigma_0,\hat{\theta}}(y_m|u_m)$ be the estimated density function. Denote

$$D[p_1,p_2]=\int(\log p_1-\log p_2)\mathrm{d}p_1$$

as the Kullback-Leibler divergence between two densities $p_1$ and $p_2$. According to Ref. [6], we only need to show the Kullback-Leibler divergence between two density functions for $y_m|u_m$ from the true and the assumed models tends to zero when $n$ is large enough.

For information consistency of the parameter estimation, we need the following condition.

**Condition (A)**: $\|\tau_{0m}\|_k$ is bounded and

$$E_{u_m}(\log|I+\sigma_0^{-2}K_m|)=o(n),$$

where $\|\tau_{0m}\|_k$ is the reproducing kernel Hilbert space norm of $\tau_{0m}$ associated with $k(\cdot,\cdot;\theta)$, $K_m$ is covariance matrix of $\tau_{0m}$ over $u_m$, $I$ is the $n\times n$ identity matrix.

More details about Condition (A) can see Seeger et al.[17] and Wang et al.[5]. More on reproducing kernel Hilbert space can see Berlinet and Thomas[18].

**Proposition 2.1.** Under the conditions in Lemma A.1 (Appendix) and condition (A), we have

$$\frac{1}{n}E_{u_m}(D[p_{\sigma_0}(y_m|\tau_{0m},u_m),p_{\sigma_0,\hat{\theta}}(y_m|u_m)])\longrightarrow 0,\quad \text{as}\quad n\to\infty,$$

where the expectation is taken over the distribution of $u_m$.

## 3 Numerical results

### 3.1 Simulations

Performance of the proposed method is investigated by nu-


DOI: 10.52396/JUSTC-2022-0016
JUSTC, 2022, 52(4): 5

merical studies. Simulation data are generated by the following model,

$$y_m(t)=z_m^\top(t)\nu+\int_0^1 x_m^\top(s,t)\beta(s,t)\mathrm{d}s+\tau_m(z_m(t),x_m(\cdot,t))+\varepsilon_m(t), \tag{5}$$

where $z_m(\cdot)\sim\mathrm{GP}(h_1,k_1)$, $h_1=h_1(t)=t$, for $t\in(0,1)$, $k_1=k_1(z_m(t_1),z_m(t_2))=g(t_1,t_2)=0.1\exp\{-5(t_1-t_2)^2\}+0.1t_1t_2$, and $x_m(\cdot,\cdot)\sim GP(h_2,k_2)$, $h_2=h_2(t)=t+\cos(s)(s)$, for $t,s\in(0,1)$, $k_2=k_2(x_m(s_1,t),x_m(s_2,t))=g(s_1,s_2)$. Let $\nu=1.0,\theta_{10}=\theta_{12}=\theta_{21}=\theta_{22}=0.1,\theta_{11}=10$, $\sigma^2=0.5$, and $t$ and $s$ take 20 points equally in (0,1). Consider four different combinations of $\tau_m$ and $\beta(s,t)$,

S1: $\tau_m\sim\mathrm{GP}(0,\mathrm{Cov}(\tau_m(u_m(t_1)),\tau_m(u_m(t_2))))$, and $h_2=h_2(t)=t+\cos(s)(s)$, for $s,t\in(0,1)$;

S2: $\tau_m\sim\mathrm{GP}(0,\mathrm{Cov}(\tau_m(u_m(t_1)),\tau_m(u_m(t_2))))$, and $\beta(s,t)=\exp\{-(t^2+s^2)\}/10$, for $s,t\in(0,1)$;

S3: $\tau_m=0$ and $\beta(s,t)=(t^2+\cos(s))/10$, for $s,t\in(0,1)$;

S4: $\tau_m=0$ and $\beta(s,t)=\exp\{-(t^2+s^2)\}/10$, for $s,t\in(0,1)$.

We take sample sizes $M$=10, 20, and 30. All simulations are repeated 500 times.

To show robustness of model (1) with random effect having ETPR, saying ETPR, we also compute model (1) with random effect having GPR, denoted by GPR. Two indices: prediction error (PE),

$$\mathrm{PE}=\frac{1}{nM}\sum_{i=1}^{M}\sum_{k=1}^{n}\left(y(t_i)-\hat f(t_i)\right)^2,$$

and average estimation bias (AB)

$$\mathrm{AB}=\frac{1}{nM}\sum_{i=1}^{M}\sum_{k=1}^{n}\left(\hat f(t_i)-f_0(t_i)\right)^2,$$

are applied to show performance of two methods: ETPR and GPR, where $\hat f(t)=z_m^\top(t)\hat\nu+\int_0^1 x_m^\top(s,t)\hat\beta(s,t)\mathrm{d}s+\hat\tau_m(z_m(t),x_m(\cdot,t))$ is an estimator of the true regression function $f_0(t)=z_m^\top(t)\nu_0+\int_0^1 x_m^\top(s,t)\beta_0(s,t)\mathrm{d}s+\tau_{0m}(z_m(t),x_m(\cdot,t))$. To show robustness of our method, one curve is randomly selected and added with an extra disturbance, $\delta t_3$, where $t_3$ stands for student $t$ distribution with degree of freedom 3. Table 1 presents the values of PE and AB from these two methods. We see that ETPR has smaller PE and AB than GPR, especially with $\delta=1.0$ and small sample sizes. It shows that the proposed method ETPR has more robustness against outliers compared to GPR.

In addition, we also consider one constant disturbance for the abnormal curves with small sample sizes 10 and 20. Tables 2 and 3 present PE and AB of prediction from ETPR method and GPR method for one and two curves disturbed, respectively. We see that ETPR has better performance in prediction compared to GPR.

### 3.2 Real data example

The proposed method is applied to Canadian weather data, which is obtained from the *R* package *fda*. We aim to study fixed effect of temperature on precipitation by common temperature effect of stations in the same region, and random effect of temperature on precipitation by individual effect of each station. Generally, the 35 stations are divided into four regions: Arctic, Atlantic, Pacific and Continental. Obviously, there exists heterogeneity among the stations due to the spatial nature of the weather data. Then we propose the following model:

$$y_{ij}(t)=\nu_0+\int_0^t x_{ij}(s)\beta_i(s,t)\mathrm{d}s+\tau_{ij}(x_{ij}(\cdot,t))+\varepsilon_{ij}(t) \tag{6}$$

where $y_{ij}(t)$ represents precipitation and $x_{ij}(t)$ represents temperature, for time $t$, region $i$ and $j$th station. In this model, we have $z_{ij}(t)=1$ and $x_{ij}(s,t)=x_{ij}(s)$ which effectively simplifies model fit.

**Table 1.** PE and AB of prediction from ETPR method and GPR method, where SDs are presented in parentheses.

| Setup | $\delta$ | Method | $M=10$ | | $M=20$ | | $M=30$ | |
|---|---|---|---|---|---|---|---|---|
| | | | PE | AB | PE | AB | PE | AB |
| S1 | 0.5 | ETPR | 0.420(0.200) | 0.170(0.194) | 0.375(0.073) | 0.124(0.065) | 0.360(0.040) | 0.111(0.029) |
| | | GPP | 0.430(0.212) | 0.188(0.207) | 0.385(0.079) | 0.135(0.072) | 0.37(0.044) | 0.12(0.035) |
| | 1 | ETPR | 0.597(0.718) | 0.3484(0.719) | 0.443(0.26) | 0.192(0.26) | 0.395(0.119) | 0.146(0.117) |
| | | GPP | 0.642(0.851) | 0.3932(0.85) | 0.466(0.296) | 0.216(0.297) | 0.414(0.137) | 0.165(0.135) |
| S2 | 0.5 | ETPR | 0.392(0.183) | 0.143(0.179) | 0.354(0.071) | 0.103(0.063) | 0.341(0.038) | 0.092(0.027) |
| | | GPP | 0.405(0.212) | 0.156(0.21) | 0.355(0.078) | 0.106(0.072) | 0.341(0.042) | 0.092(0.034) |
| | 1 | ETPR | 0.569(0.707) | 0.321(0.712) | 0.42(0.273) | 0.172(0.273) | 0.377(0.119) | 0.129(0.117) |
| | | GPP | 0.608(0.852) | 0.36(0.855) | 0.437(0.295) | 0.187(0.295) | 0.386(0.138) | 0.137(0.137) |
| S3 | 0.5 | ETPR | 0.345(0.229) | 0.094(0.225) | 0.308(0.088) | 0.057(0.083) | 0.293(0.041) | 0.042(0.034) |
| | | GPP | 0.359(0.289) | 0.109(0.283) | 0.319(0.122) | 0.069(0.119) | 0.301(0.061) | 0.051(0.056) |
| | 1 | ETPR | 0.515(0.817) | 0.263(0.814) | 0.38(0.359) | 0.129(0.359) | 0.3275(0.191) | 0.077(0.191) |
| | | GPP | 0.579(1.158) | 0.328(1.15) | 0.425(0.483) | 0.175(0.482) | 0.3568(0.24) | 0.106(0.24) |
| S4 | 0.5 | ETPR | 0.311(0.22) | 0.06(0.218) | 0.276(0.091) | 0.026(0.087) | 0.263(0.04) | 0.013(0.035) |
| | | GPP | 0.326(0.287) | 0.074(0.283) | 0.287(0.123) | 0.036(0.121) | 0.269(0.06) | 0.019(0.057) |
| | 1 | ETPR | 0.481(0.82) | 0.229(0.817) | 0.351(0.363) | 0.1(0.363) | 0.299(0.193) | 0.049(0.194) |
| | | GPP | 0.547(1.154) | 0.295(1.15) | 0.395(0.485) | 0.144(0.485) | 0.326(0.241) | 0.076(0.243) |


DOI: 10.52396/JUSTC-2022-0016
*JUSTC*, **2022**, 52(4): 5

**Table 2.** PE and AB of prediction from ETPR method and GPR method with one curve disturbed by constant 1.0, where SDs are presented in parentheses.

| Setup | Method | $M$ = 10 | | $M$ = 20 | |
|---|---|---|---|---|---|
| | | PE | AB | PE | AB |
| S1 | ETPR | 0.429(0.069) | 0.18(0.042) | 0.388(0.044) | 0.138(0.025) |
| | GPP | 0.457(0.073) | 0.208(0.044) | 0.406(0.046) | 0.155(0.026) |
| S2 | ETPR | 0.401(0.066) | 0.153(0.039) | 0.365(0.042) | 0.114(0.023) |
| | GPP | 0.42(0.069) | 0.173(0.04) | 0.374(0.043) | 0.123(0.023) |
| S3 | ETPR | 0.341(0.053) | 0.09(0.027) | 0.302(0.032) | 0.051(0.01) |
| | GPP | 0.368(0.065) | 0.117(0.044) | 0.312(0.039) | 0.061(0.021) |
| S4 | ETPR | 0.303(0.048) | 0.052(0.024) | 0.268(0.029) | 0.018(0.009) |
| | GPP | 0.33(0.065) | 0.08(0.045) | 0.275(0.037) | 0.025(0.022) |

**Table 3.** PE and AB of prediction from ETPR method and GPR method with two curves disturbed by constant 1.0, where SDs are presented in parentheses.

| Setup | Method | $M$ = 10 | | $M$ = 20 | |
|---|---|---|---|---|---|
| | | PE | AB | PE | AB |
| S1 | ETPR | 0.512(0.082) | 0.263(0.053) | 0.425(0.05) | 0.175(0.029) |
| | GPP | 0.558(0.083) | 0.308(0.055) | 0.453(0.051) | 0.202(0.03) |
| S2 | ETPR | 0.479(0.079) | 0.23(0.049) | 0.399(0.048) | 0.148(0.027) |
| | GPP | 0.517(0.082) | 0.268(0.052) | 0.421(0.05) | 0.169(0.028) |
| S3 | ETPR | 0.423(0.071) | 0.171(0.045) | 0.329(0.037) | 0.078(0.018) |
| | GPP | 0.48(0.081) | 0.227(0.055) | 0.367(0.047) | 0.115(0.032) |
| S4 | ETPR | 0.381(0.068) | 0.129(0.044) | 0.296(0.035) | 0.045(0.017) |
| | GPP | 0.441(0.078) | 0.188(0.053) | 0.333(0.048) | 0.081(0.034) |

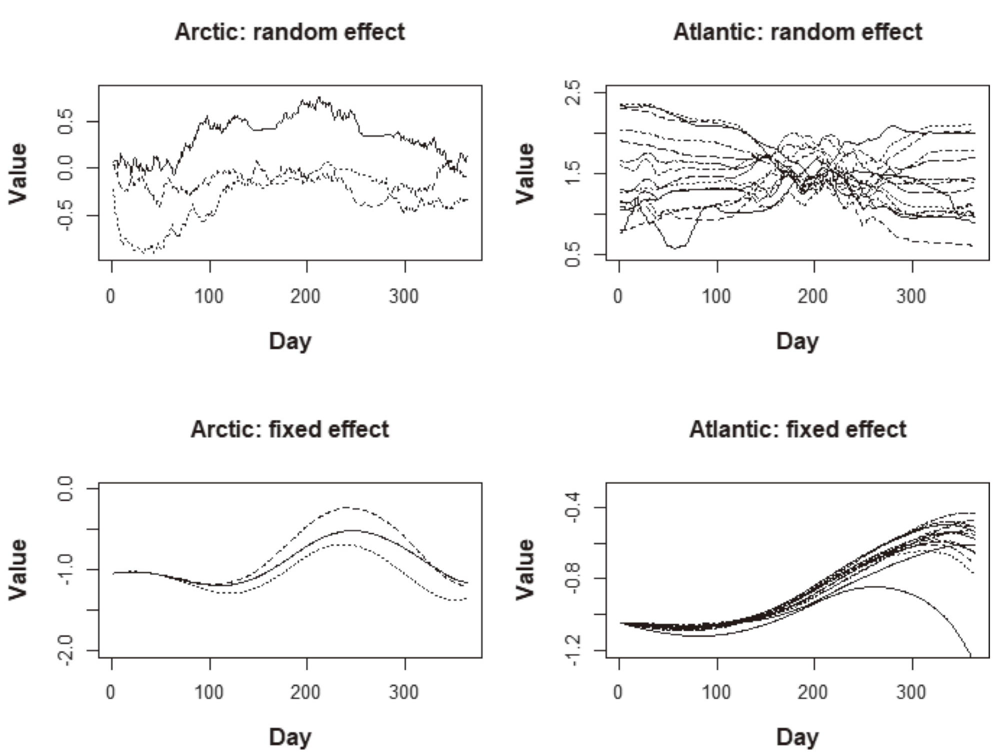



**Fig. 1.** Random and fixed effects of model using ETPR for Arctic and Atlantic.


DOI: 10.52396/JUSTC-2022-0016
*JUSTC*, **2022**, 52(4): 5

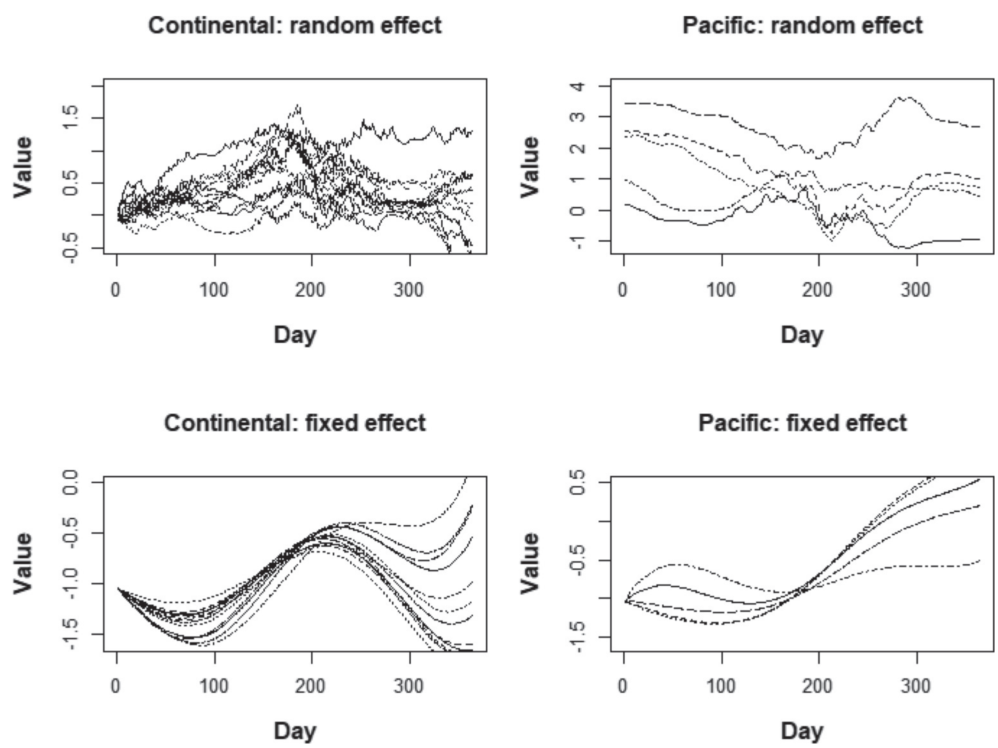



**Fig. 2.** Random and fixed effects of model using ETPR for Continental and Pacific.

Figs. 1 and 2 show random and fixed effects of the 4 regions: Arctic, Atlantic, Pacific and Continental from the proposed method. We see from the random effects that each station in the same region has different temperature effects on the precipitation. To compare performance of prediction from ETPR with GPR, 10-folds cross validation method is used to compute mean squares of prediction errors, 0.310 and 0.314, for ETPR and GPR, respectively. It shows that ETPR has a little better performance in prediction.

## 4 Conclusions

A function-on-function random effects model with extended t-process prior in this paper is developed to analyze functional data which may include outliers. The proposed model is flexible, including various kinds of functional models, such as the function-on-function linear model[2] and the historical functional regression model[7] as special cases. The proposed extended t-process model is not only robust against outliers, but also inherits almost all the nice properties from Gaussian process regression, such as closed form of prediction and convenient computation procedure. The estimation procedure and computing algorithm are developed to estimate the parameters and predict the random effect in the regression model. The functional response considered in this paper has one dimension. In practical application, functional multi-response may consist of several correlated curves. It is interesting that the proposed method is extended to functional data with multi-response, which will be studied in our further work.

## Appendix

**Lemma A.1.** Let $w = v-1$. Under model (1), assume that $y_m$ are independently sampled, the covariance kernel function $k$ is bounded and continuous on the parameter $\theta$, and $\hat{\boldsymbol{\theta}}$ converges to $\theta$ when $n \to \infty$. Then, for a positive constant $c$ and any $\varepsilon > 0$, when $n$ is large enough, we have

$$\frac{1}{n}\left(-\log\{p_{\sigma_0,\hat{\theta}}(y_m|u_m)\} + \log\{p_{\sigma_0}(y_m|\tau_{0m},u_m)\}\right) \leqslant$$
$$\frac{1}{n}\left\{\frac{1}{2}\log|I+\sigma^{-2}K_m| + \frac{q_m^2+2(v-1)}{2(n+2v-2)}\left(\|\tau_{0m}\|_k^2+c\right)+c\right\}+\varepsilon,$$

where $q_m^2 = (y_m - c_{0m} - \tau_{0m})^{\top}(y_m - c_{0m} - \tau_{0m})/\sigma_0^2$, $c_{0m}$ is the true value of $c_m$, $\|\tau_{0m}\|_k$ is the reproducing kernel Hilbert space norm of $\tau_{0m}$ associated with $k(\cdot,\cdot;\theta)$, $K_m$ is covariance matrix of $\tau_{0m}$ over $u_m$, $I$ is the $n\times n$ identity matrix.

**Proof of Lemma A.1.** Assume $r$ is a random variable following inverse gamma distribution $\mathrm{IG}(v,(v-1))$. Conditional on $r$, we have

$$\left(\begin{array}{c}\tau_m\\ \varepsilon_m\end{array}\right)\Big|r \sim \mathrm{GP}\left(\left(\begin{array}{c}0\\0\end{array}\right),\left(\begin{array}{cc}r_m k & 0\\ 0 & r_m\sigma^2\delta_\varepsilon\end{array}\right)\right),$$

where $\mathrm{GP}(h,k)$ stands for Gaussian process with mean function $h$ and covariance function $k$. Then conditional on $r_m$, the extended t-process regression model $y_m = c_m + \tau_m + \varepsilon_m$ becomes Gaussian process regression model

$$y_m = c_m + \tilde{\tau}_m + \tilde{\varepsilon}_m,$$

where $\tilde{\tau}_m = \tau_m|r_m \sim \mathrm{GP}(0, r_m k(\cdot,\cdot;\boldsymbol{\theta}))$, $\tilde{\varepsilon}_m = \varepsilon_m|r_m \sim \mathrm{GP}(0, r\sigma^2\delta_\varepsilon)$, and $\tilde{\tau}_m$ and $\tilde{\varepsilon}_m$ are independent. Denoted the computation of conditional probability density for given $r_m$ by $\tilde{p}$. Let

$$p_G(y_m \mid r_m,u_m) = \int_{\mathcal{F}} p_{\sigma_0}(y_m \mid \tilde{\tau}_m, r_m, u_m)\,\mathrm{d}\tilde{p}_{\theta}(\tilde{\tau}_m),$$
$$p_0(y_m \mid r_m,u_m) = p_{\sigma_0}(y_m \mid \tau_{0m}, r_m, u_m),$$

where $\tilde{p}_{\boldsymbol{\theta}}$ is the induced measure from Gaussian process $\mathrm{GP}(0, r_m k(\cdot,\cdot;\hat{\boldsymbol{\theta}}))$. Note that variable $r$ is independent of $u_m$. We can show that


DOI: 10.52396/JUSTC-2022-0016
*JUSTC*, **2022**, 52(4): 5

$$p_{\sigma_0,\hat\theta}(y_m \mid u_m) = \int p_G(y_m \mid r_m, u_m)\, g(r)\mathrm{d}r,$$
$$p_{\sigma_0}(y_m \mid \tau_{0m}, u_m) = \int p_0(y_m \mid r_m, u_m)\, g(r)\mathrm{d}r.$$

By similar procedures in Seeger et al.[17] and Wang et al.[10], for any given $r$, we have

$$-\log p_G(y_m \mid r_m, u_m) + \log p_0(y_m \mid r_m, u_m) \leqslant \frac{1}{2}\log\left|\boldsymbol{I} + \sigma_0^{-2}K_m\right| + \frac{r_m}{2}\left(\|\tau_{0m}\|_k^2 + c\right) + c + n\varepsilon,$$

then it follows that

$$-\log\int p_G(y_m \mid r, u_m)\, g(r)\mathrm{d}r \leqslant \frac{1}{2}\log\left|\boldsymbol{I} + \sigma_0^{-2}K_m\right| + c + n\varepsilon - \log\int p_0(y_m \mid r, u_m)\exp\left\{-\left(\frac{r}{2}\left(\|\tau_{0m}\|_k^2 + c\right)\right)\right\} g(r)\mathrm{d}r.$$

Let $g^*(r)$ be the density function of $\mathrm{IG}(\nu + n/2, \nu - 1 + q_m^2/2)$. It easily shows that

$$\int p_0(y_m \mid r, u_m)\exp\left\{-\left(\frac{r}{2}\left(\|\tau_{0m}\|_k^2 + c\right)\right)\right\} g(r)\mathrm{d}r = \int p_0(y_m \mid r, u_m)\, g(r)\mathrm{d}r \int \exp\left\{-\left(\frac{r}{2}\left(\|\tau_{0m}\|_k^2 + c\right)\right)\right\} g^*(r)\mathrm{d}r.$$

We have

$$-\log p_{\sigma_0,\hat\theta}(y_m \mid u_m) + \log p_{\sigma_0}(y_m \mid \tau_{0m}, u_m) \leqslant$$
$$\frac{1}{2}\log\left|\boldsymbol{I} + \sigma_0^{-2}K_m\right| + c + n\varepsilon - \log\int \exp\left\{-\left(\frac{r}{2}\left(\|\tau_{0m}\|_k^2 + c\right)\right)\right\} g^*(r)\mathrm{d}r \leqslant$$
$$\frac{1}{2}\log\left|I + \sigma_0^{-2}K_m\right| + c + n\varepsilon + \frac{\|\tau_{0m}\|_k^2 + c}{2}\int r g^*(r)\mathrm{d}r =$$
$$\frac{1}{2}\log\left|I + \sigma_0^{-2}K_m\right| + \frac{q_m^2 + 2(\nu - 1)}{2(n + 2\nu - 2)}\left(\|\tau_{0m}\|_k^2 + c\right) + c + n\varepsilon,$$

which shows that Lemma A.1 holds.

**Proof of Proposition 2.1.** Obviously $q_m^2 = (y_m - c_{0m} - \tau_{0m})^\top \cdot (y_m - c_{0m} - \tau_{0m})/\sigma_0^2 = O(n)$. Under the conditions of Lemma A.1 and condition (A), by Lemma A.1, for a positive constant $c$ and any $\varepsilon > 0$, when $n$ is large enough, we have

$$\frac{1}{n}E_{u_m}\left(D[p_{\sigma_0}(y_m|\tau_{0m}, u_m), p_{\sigma_0,\hat\theta}(y_m|u_m)]\right) =$$
$$E_{u_m}\int \frac{1}{n}\left(-\log p_{\sigma_0,\hat\theta}(y_m|u_m) + \log p_{\sigma_0}(y_m|\tau_{0m}, u_m)\right)\mathrm{d}p_{\sigma_0}(y_m|\tau_{0m}, u_m) \leqslant$$
$$E_{u_m}\int\left(\frac{1}{2n}\log\left|I + \sigma_0^{-2}K_m\right| + \frac{q_m^2 + 2(\nu - 1)}{2n(n + 2\nu - 2)}\left(\|\tau_{0m}\|_k^2 + c\right) + \frac{c}{n} + \varepsilon\right)\mathrm{d}p_{\sigma_0}(y_m|\tau_{0m}, u_m) \longrightarrow 0, \quad \text{as} \quad n \to \infty.$$

Thus, it completes the proof.

## Acknowledgements

We thank the reviewers for their insightful comments and suggestions. This work was supported in part by the National Natural Science Foundation of China (11971457), Anhui Provincial Natural Science Foundation (1908085MA06) and the Fundamental Research Funds for the Central Universities (WK2040000035).

## Conflict of interest

The authors declare that they have no conflict of interest.

## Biographies

**Shanshan Wang** is currently a master student under the supervision of Assoc. Prof. Zhanfeng Wang at the University of Science and Technology of China. Her research mainly focuses on functional data.

**Hao Ding** received his PhD degree from the University of Science and Technology of China (USTC). He is currently a postdoctoral fellow at USTC. His research focuses on robust estimation, functional data analysis.

## References

[1] Wang Z, Noh M, Lee Y, et al. A general robust *t*-process regression model. *Computational Statistics and Data Analysis,* **2021**, *154*: 107093.

[2] Yuan M, Cai T T. A reproducing kernel Hilbert space approach to functional linear regression. *The Annals of Statistics,* **2010**, *38* (6): 3412–3444.

[3] Wang Z, Shi J Q, Lee Y. Extended *t*-process regression models. *Journal of Statistical Planning and Inference,* **2017**, *189*: 38–60.

[4] Seeger M W, Kakade S M, Foster D P. Information consistency of nonparametric Gaussian process methods. *IEEE Transactions on Information Theory,* **2008**, *54*: 2376–2382.

[5] Zhang Y, Yeung D Y. Multi-task learning using generalized *t*-process. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. Cambridge, MA: PMLR, 2010: 964–971.

[6] Yao F, Müller H G, Wang J L. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association,* **2005**, *100*: 577–590.

[7] Wang B, Shi J Q. Generalized gaussian process regression model for non-gaussian functional data. *Journal of the American Statistical Association,* **2014**, *109*: 1123–1133.

[8] Shi J Q, Choi T. Gaussian Process Regression Analysis for Functional Data. Boca Raton, FL: CRC Press, 2011.

[9] Wang Z, Ding H, Chen Z, et al. Nonparametric random effects functional regression model using Gaussian process priors. *Statistica Sinica,* **2021**, *31*: 53–78.

[10] Yu S, Tresp V, Yu K. Robust multi-task learning with *t*-processes. In: Proceedings of the 24th International Conference on Machine Learning. New York: ACM, 2007: 1103–1110.

[11] Berlinet A, Thomas-Agnan C. Reproducing Kernel Hilbert Spaces in Probability and Statistics. Berlin: Springer Science & Business Media, 2011.

[12] Malfait N, Ramsay J O. The historical functional linear model. *Canadian Journal of Statistics,* **2003**, *31*: 115–128.

[13] Sun X, Du P, Wang X, et al. Optimal penalized function-on-function regression under a reproducing kernel Hilbert space framework. *Journal of the American Statistical Association,* **2018**, *113* (524): 1601–1611.

[14] Shah A, Wilson A, Ghahramani Z. Student-*t* processes as alternatives to Gaussian processes. In: Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics. Cambridge, MA: PMLR, 2014: 877–885.

[15] Gervini D. Dynamic retrospective regression for functional data. *Technometrics,* **2015**, *57*: 26–34.

[16] Ramsay J O, Silverman B W. Functional Data Analysis. New York: Springer, 2005.

[17] Ramsay J O, Dalzell C. Some tools for functional data analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* **1991**, *53*: 539–572.

[18] Yao F, Müller H G, Wang J L. Functional linear regression analysis for longitudinal data. *The Annals of Statistics,* **2005**, *33*: 2873–2903.



DOI: 10.52396/JUSTC-2022-0016
*JUSTC*, **2022**, 52(4): 5