Comparative Analysis of Functional Snippet Analysis Methods

By Timothy Boakye, Kyoung Min Kim

December 7, 2023

Introduction

Functional data analysis has emerged as a powerful tool for unraveling intricate patterns within datasets that traditional statistical methods struggle to address effectively. The ability to represent data as functions allows for a more nuanced understanding of underlying structures and trends, particularly in domains where observations evolve continuously over a continuum.

However, the analysis of functional snippets, which are brief and irregular segments extracted from functional data, introduces unique challenges. The irregularity and sparsity of functional snippets demand specialized methodologies for their comprehensive analysis. Recognizing this need, Li et al. (2021) proposed a novel method in their seminal paper titled "Mean and Covariance Estimation for Functional Snippets." This method focuses on enhancing our understanding of functional snippets by providing robust estimates for both mean and covariance.

The primary motivation behind the proposed research project, titled is to critically evaluate the performance of the novel approach introduced by Li et al. in comparison to two existing methods, namely PFBE (Piecewise Functional Bayesian Estimation) and PACE (PEnalized Covariance Estimation). PFBE and PACE are well-established methods in the field, and their comparison with the novel approach will provide valuable insights into the strengths and limitations of each.

In the context of functional snippet analysis, the mean and covariance structure play pivotal roles in capturing essential characteristics of the data. The proposed research will delve into the mathematical foundations of these methods, emphasizing the underlying formulas that drive the estimation processes. Understanding these formulas is crucial for appreciating the nuances in how each method approaches the challenges posed by functional snippets.

The primary formulae underpinning the mean and covariance estimation for functional snippets, as introduced by Li et al. and extended by PFBE and PACE, will be a focal point of our comparative analysis. Exploring these mathematical expressions will shed light on the theoretical foundations of each method and provide a basis for comparing their efficacy in capturing the inherent complexities of functional snippets.

As we embark on this comparative analysis, our overarching goal is to contribute to the refinement of methodologies for functional snippet analysis, thereby enhancing our ability to derive meaningful insights from sparse and irregular functional data. Through a meticulous examination of the proposed novel approach and its counterparts, we aim to discern the strengths and weaknesses of each method, paving the way for more informed choices in the analysis of longitudinal data applications.

Objective I

Our first objective is to illustrate and simplify the process of generating mean and covariance function estimates for each estimator introduced by Li et al. (2021). The statistical methodology will be validated through simulations and data analysis, enhancing accessibility for readers. This clarity will empower readers to make informed decisions regarding estimator selection based on the specific characteristics of their data settings.

To achieve this objective, we will provide illustration of the methodology introduced by Li et al. (2021) for estimating the mean and covariance functions of functional snippets. Let $X(t)$ represent a functional snippet observed at time t and n be the number of observed snippets. The mean function estimate, denoted

as $\hat{\mu}(t)$, and the covariance function estimate, denoted as $\hat{\Sigma}(t_1, t_2)$, will be computed at different time points and the number of spatial locations .

The mean function estimate is given by:

$$\hat{\mu}(t) = \frac{1}{n} \sum_{i=1}^{n} X_i(t) \tag{1}$$

The covariance function estimate between time points t_1 and t_2 is given by:

$$\hat{\Sigma}(t_1, t_2) = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i(t_1) - \hat{\mu}(t_1)\right) \left(X_i(t_2) - \hat{\mu}(t_2)\right) \tag{2}$$

Objective II: Comparative Evaluation of Estimators

The core of our project lies in the comparative analysis of the proposed SNPTM methods against existing methodologies—PFBE by Lin (2020) and PACE by Fang Yao (2012) at different sample sizes. We aim to provide intuition into the theoretical statistical properties of each estimator through diverse simulation settings and data analysis.

Comparative Framework

The comparison will be conducted based on fundamental statistical properties such as variance, and mean squared error (MSE). Let θ represent the parameter of interest (mean or covariance function), and $\hat{\theta}_{SNPTM}, \hat{\theta}_{PFBE}, \hat{\theta}_{PACE}$ denote the estimators corresponding to SNPTM, PFBE, and PACE methods, respectively.

Main Issues in Functional Data Analysis

Statistical issues in Sparse, fragmented, irregularly sampled

The FPC (Functional Principal Component) scores $\xi_{ik} = \int (X_i(t) - \mu(t))\phi_k(t)\, dt$ have traditionally been estimated by numerical integration, which works well when the density of the grid of measurements for each subject is sufficiently large.

$$\hat{\xi}_{ik} = \sum_{j=1}^{J} \left(s_i(t_j) - \hat{\mu}(t_j)\right) \phi_k(t_j) \Delta t_j$$

Where:

$$\Delta t_j = t_j - t_{j-1}$$

$\hat{\xi}_{ik}$ are the FPC-scores for observation i.

Because in our model the Y_{ij} are available only at discrete random times T_{ij}, reflecting the sparseness of the data, the integrals in the definition of the FPC scores ξ_{ik} accordingly would be approximated by sums, substituting Y_{ij} for $X_i(T_{ij})$ and estimates $\hat{\mu}(t_{ij})$ for $\mu(t_{ij})$ and $\hat{\phi}_k(t_{ij})$ for $\phi_k(t_{ij})$, leading to

$$\hat{\xi}_{Sik} = \sum_{Nij=1} (Y_{ij} - \hat{\mu}(T_{ij}))\hat{\phi}_k(T_{ij})(T_{ij} - T_{i,j-1})$$

setting $T_{i0} = 0$. For sparse functional data, $\hat{\xi}_{Sik}$ will not provide reasonable approximations to ξ_{ik}, for example, when one has only two observations per subject. Moreover, when the measurements are contaminated with errors, the underlying random process X cannot be directly observed. Substituting Y_{ij} for $X_i(T_{ij})$ then leads to biased FPC scores. These considerations motivate the alternative method to obtain the FPC scores.

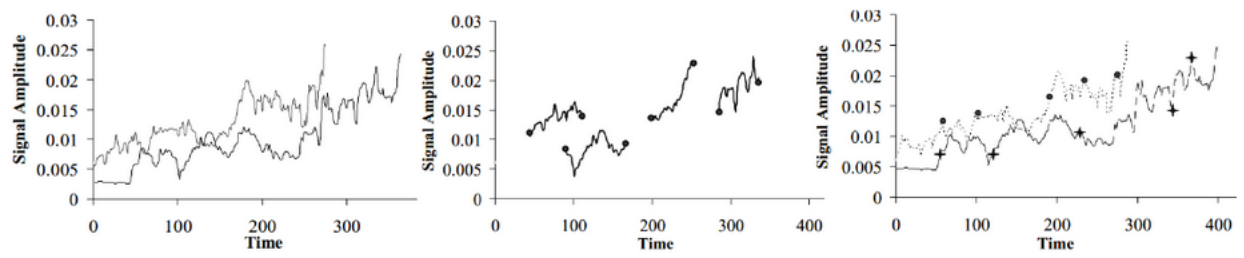

Figure 1: sparse(right, fragmented(middle), irregularly sampled data

Classical Functional Data Analysis

Statistical issues in Sparse, fragmented, irregularly sampled

The FPC (Functional Principal Component) scores $\xi_{ik} = \int (X_i(t) - \mu(t))\phi_k(t)\,dt$ have traditionally been estimated by numerical integration, which works well when the density of the grid of measurements for each subject is sufficiently large.

$$\hat{\xi}_{ik} = \sum_{j=1}^{J} \left(s_i(t_j) - \hat{\mu}(t_j)\right) \phi_k(t_j)\Delta t_j$$

Where:

$$\Delta t_j = t_j - t_{j-1}$$

$\hat{\xi}_{ik}$ are the FPC-scores for observation i.

Because in our model the Y_{ij} are available only at discrete random times T_{ij}, reflecting the sparseness of the data, the integrals in the definition of the FPC scores ξ_{ik} accordingly would be approximated by sums, substituting Y_{ij} for $X_i(T_{ij})$ and estimates $\hat{\mu}(t_{ij})$ for $\mu(t_{ij})$ and $\hat{\phi}_k(t_{ij})$ for $\phi_k(t_{ij})$, leading to

$$\hat{\xi}_{Sik} = \sum_{Nij=1} (Y_{ij} - \hat{\mu}(T_{ij}))\hat{\phi}_k(T_{ij})(T_{ij} - T_{i,j-1})$$

setting $T_{i0} = 0$. For sparse functional data, $\hat{\xi}_{Sik}$ will not provide reasonable approximations to ξ_{ik}, for example, when one has only two observations per subject. Moreover, when the measurements are contaminated with errors, the underlying random process X cannot be directly observed. Substituting Y_{ij} for $X_i(T_{ij})$ then leads to biased FPC scores. These considerations motivate the alternative method to obtain the FPC scores.

Introduction to PACE(Principal Analysis by Conditional Expectation)

Classical functional data analysis requires a large number of regularly spaced measurements per subject. We perform functional principal components analysis for the case of sparse longitudinal data. The method aims at irregularly spaced longitudinal data, where the number of repeated measurements available per subject is small.

Improvements over Previous Studies

- The inclusion of additional measurement errors represents a significant improvement.
- The method effectively manages sparse and irregular longitudinal data.
- In particular, it efficiently reconstructs random trajectories by integrating information from both individual cases and the entire dataset during the construction of estimated mean functions, covariance functions.

Estimation of Mean Function
Historical signals s_{ij}:
$i = 1, \ldots, N$: signal index
$j = 1, \ldots, mi$: observation index in each signal
$S_i(t_{ij}) = \mu(t_{ij}) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(t_{ij}) + \epsilon_{ij}$, where $E(\epsilon_{ij}) = 0, var(\epsilon_{ij}) = \sigma^2$
We can estimate the mean function $\hat{\mu}(t)$ using local linear regression by minimizing:

$$\min_{c_0,c_1} \sum_{ni=1} \sum_{mij=1} W(t - t_{ij}h)(s_i(t_{ij}) - c_0 - (t - t_{ij})c_1)^2$$

The solution is given by:

$$\hat{\mu}(t) = \hat{c}_0(t)$$

Estimation of Covariance Function
First, we use the estimated mean functions to estimate the raw covariance function $\hat{C}(t, t')$:

$$\hat{C}_i(t_{ij}, t_{ik}) = (s_i(t_{ij}) - \hat{\mu}(t_{ij}))(s_i(t_{ik}) - \hat{\mu}(t_{ik}))$$

$\hat{C}_i(t_{ij}, t_{ik})$ is the covariance between observations i at times t_{ij} and t_{ik}. $s_i(t_{ij})$ is the signal at observation i and time t_{ij}. $\hat{\mu}(t_{ij})$ is the estimated mean function at time t_{ij}. $s_i(t_{ik})$ is the signal at observation i and time t_{ik}. $\hat{\mu}(t_{ik})$ is the estimated mean function at time t_{ik}.
The minimization problem is as follows:

$$\min_{c_0,c_1,c_2} \sum_{ni=1} \sum_{1 \le j \ne k \le mi} W(t_{i,j} - th, t_{i,k} - th)(\hat{C}_i(t_{i,j}, t_{i,k}) - c_0 - c_1(t - t_{i,j}) - c_2(t' - t_{i,k}))^2$$

To estimate the covariance surface $\hat{C}(t, t')$, we use local quadratic regression. The solution is given by:

$$\hat{C}(t, t') = \hat{c}_0(t, t')$$

To solve the estimated covariance function, $\hat{\phi}_k(t)$ is estimated by discretizing the estimated covariance function $\hat{C}(t, t')$.

CD4 Count Data in perspective of integrating information from both individual cases and the entire dataset

CD4 Count Data in perspective of integrating information from both individual cases and the entire dataset. The dataset considered here is from the Multicenter AIDS Cohort Study, which includes repeated measurements of physical exams, laboratory results, and CD4 percentages for 283 homosexual men who became HIV-positive between 1984 and 1991. All individuals were scheduled to have their measurements made at semiannual visits. However, because many individuals missed scheduled visits and the HIV infections happened randomly during the study, the data are sparse, with unequal numbers of repeated measurements per subject and different measurement times, Tij, per individual. The number of observations per subject ranged from 1 to 14, with a median of 6.

That the data from such a classical longitudinal study, with measurements intended to be spaced at regular 6-month intervals, are quite well suited for analysis by PACE is illustrated by Figure 2.

Although the data available per subject are sparse, the assembled pairs (Tij,Tik) are sufficiently dense in the domain plane, and estimation of the covariance function is feasible for these data.

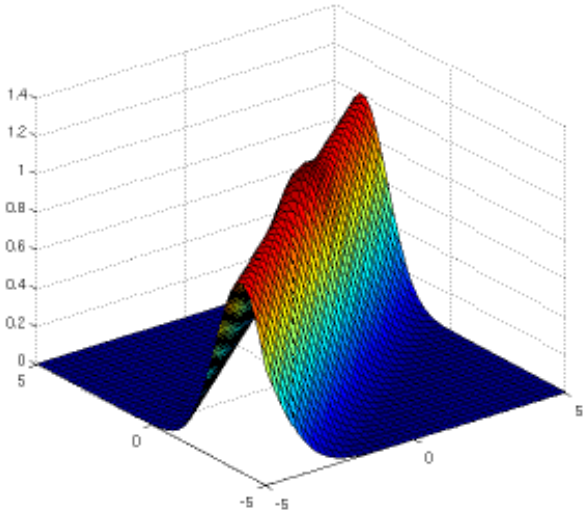

Figure 2: Discretizing the estimated covariance function

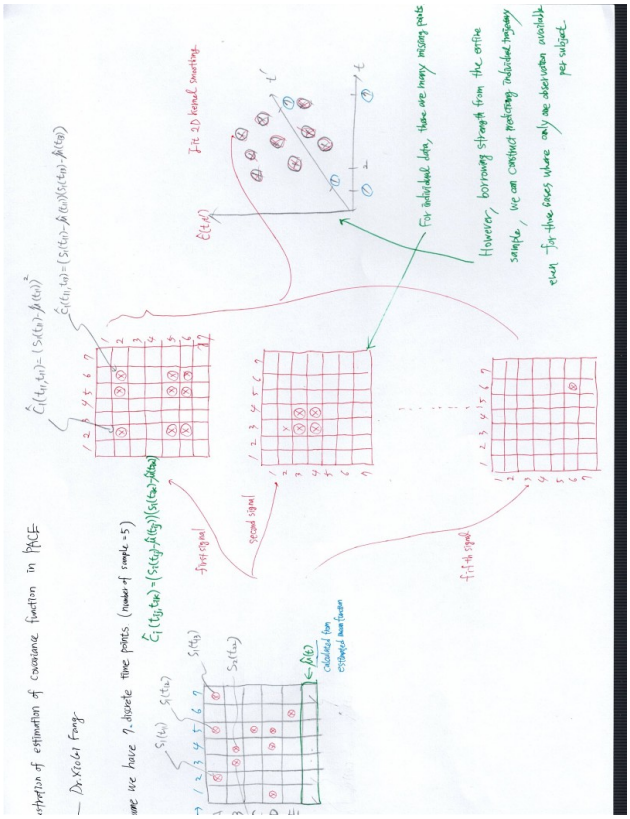

Figure 3: Graphical Illustration of Estimation of Covariance Function

Figure 4: Caption

Computing FPC-Scores

The best prediction of the FPC scores for the ith subject, given the data from that individual, is the conditional expectation, which, under Gaussian assumptions, is found to be:

$$\tilde{\xi}_{ik} = E[\xi_{ik}|\tilde{Y}_i] = \lambda_k \phi_k^T \Sigma_Y^{-1} Y_i (\tilde{Y}_i - \mu_i)$$

where $\Sigma_Y = cov(\tilde{Y}_i, \tilde{Y}_i) = cov(\tilde{X}_i, \tilde{X}_i) + \sigma^2 I_{Ni}$

From the fact that $G(s,t) = cov(\tilde{X}(s), \tilde{X}(t))$, the (j,l) entry of the $Ni \times Ni$ matrix Σ_Y is $(\Sigma_Y)_{j,l} = G(T_{ij}, T_{il}) + \sigma^2 \delta_{jl}$, where $\delta_{jl} = 1$ if $j = l$ and 0 if $j \neq l$.

Estimates for the FPC scores ξ_{ik} are obtained from (1), by substituting estimates of $\mu_i, \lambda_k, and \phi_{ik}, \Sigma_Y$ obtained from the entire data ensemble, leading to:

$$\hat{\xi}_{ik} = \hat{E}[\xi_{ik}|\tilde{Y}_i] = \hat{\lambda}_k \hat{\phi}_{ik}^T \hat{\Sigma}_Y^{-1} (\tilde{Y}_i - \hat{\mu}_i)$$

where the (j,l)th element of $\hat{\Sigma}_Y$ is $(\hat{\Sigma}_Y)_{j,l} = \hat{G}(T_{ij}, T_{il}) + \hat{\sigma}^2 \delta_{jl}$.

where, $\xi_{ik} and \epsilon_{ij}$ are jointly Gaussian. In all of what follows, the results pertaining to expectations are always conditional on the observation times $T_{ij}, i = 1, \ldots, n, j = 1, \ldots, Ni$. For simplicity, the dependence on T_{ij} is suppressed. Write $\tilde{X}_i = (X_i(T_{i1}), \ldots, X_i(T_{iNi}))^T$, $\tilde{Y}_i = (Y_{i1}, \ldots, Y_{iNi})^T$, $\mu_i = (\mu(T_{i1}), \ldots, \mu(T_{iNi}))^T$, and $\phi_{ik} = (\phi_k(T_{i1}), \ldots, \phi_k(T_{iNi}))^T$.

New Estimator for Analysing Functional Snippets (SNPT)

The studies are concerned with the estimation of the average and covariance functions for functional snippets, which are brief portions of functions that may be observed sporadically within a specific subinterval considerably shorter than the entire study period. Estimating the covariance function for these functional snippets presents a challenge because information about the distant, off-diagonal portions of the covariance structure is entirely absent.

Definition of Functional Snippets Data

The phenomenon that each individual trajectory is only recorded in an individual specific subinterval that is much shorter than the span of the study.

Mathematical Expression:

Subjects enter the study at random times and are followed for a short period within the domain $T = [a,b] \subset R$. Specifically, we focus on functional data with the following property: each function X_i is only observed on a subject-specific interval $O_i = [A_i, B_i] \subset [a,b]$. (S) There exists an absolute constant δ such that $0 < \delta < 1$ and $B_i - A_i \leq \delta(b-a)$ for all $i = 1, 2, \ldots$

As a result, the design of support points (Yao et al., 2005) where one has information about the covariance function $C(s,t)$ is incomplete in the sense that there are no design points in the off-diagonal region $T_\delta^c = (s,t) \in [a,b]^2 : |s-t| > \delta(b-a), s,t \in [a,b]$. This is mathematically characterized by $\bigcup_i([A_i, B_i]^2) \cap T_\delta^c = \emptyset$ - (1).

Example of Functional Snippets Data

An example is the spinal bone mineral density data collected from 423 subjects ranging in age from 8.8 to 26.2 years (Bachrach et al., 1999). The design plot for the covariance function, indicates that all of the design points fall within a narrow band around the diagonal area but the domain of interest [8.8, 26.2] is much larger than this band.

Limitation of PACE in the Estimation of Covariance Functional Snippets Data Permalink

PACE (Yao et al., 2005) is a local smoothing method that, unlike interpolation methods, fails to produce a consistent estimate of the covariance function in the off-diagonal region, as the problem requires data extrapolation. We will demonstrate the vulnerability of PACE for covariance function estimation at the boundary of the domain, where the number of design points is limited, in the Simulation Studies and Data Analysis sections.

In the Data Analysis for spinal bone mineral density data collected from 423 subjects ranging in age from 8.8 to 26.2 years (Bachrach et al., 1999), Li, J., Wang, Q., & Zhang, S. (2021) presented the phenomenon and the estimated covariance estimation function.

SNPT in the Estimation of Covariance Functional Snippets Data

To overcome this challenge, SNPT tackles it by breaking down the covariance function into two components: a variance function component and a correlation function component. The variance component can be effectively estimated through nonparametric methods, while the correlation component is modeled parametrically, potentially involving an increasing number of parameters, to account for the missing data in the distant off-diagonal regions. Both theoretical analysis and numerical simulations indicate that this hybrid approach is highly effective. Furthermore, they introduce a novel estimator for the variance of measurement errors and examine its asymptotic properties. This estimator is essential for estimating the variance function when dealing with noisy measurements.

Estimation of Mean Function

Smoothing approaches such as Yao et al. (2005) can be applied to estimate the mean function μ.

Estimation of Covariance Function

The covariance function can be decomposed into two parts, a variance function and a correlation structure, i.e., $C(s,t) = \sigma^2 X(s)\sigma^2 X(t)\rho(s,t)$, where $\sigma^2 X(t)$ is the variance function of X, or more precisely, $\sigma^2 X(t) = E[(X(t) - \mu(t))^2]$, and $\rho(s,t)$ is the correlation function.

Estimation of Variance Function:

$(\hat{b}_0, \hat{b}_1) = argmin_{(b_0,b_1)\in R^2} \sum_{ni=1} \sum_{mij=1} Kh_\sigma(T_{ij} - t)(f(Y_{ij}) - \hat{\mu}(T_{ij}) - b_0 - b_1(T_{ij} - t))^2$

Then $\hat{b}_0 = \tilde{\zeta}^2(t)$. Our estimate of $\sigma^2 X(t)$ is $\hat{\sigma}^2 X(t) = \tilde{\zeta}^2(t) - \hat{\sigma}_0^2$, where $\hat{\sigma}_0^2$ is a new estimate of σ_0^2, where $\tilde{\zeta}^2(t)$ is the ridged version of $\tilde{\zeta}^2(t)$, $\tilde{\zeta}^2(t)$ is the non-ridged local linear estimate of $\zeta^2(t)$, and $\sigma^2(t) = E[(Y(t) - \mu(t))^2] = \sigma^2 X(t) + \sigma_0^2$

Estimation of Covariance Function:

$$\hat{Q}_n(\theta) = \sum_{ni=1}^{1mi(mi-1)} \sum_{1\le j\ne l\le mi} (\hat{\sigma}^X(T_{ij})\hat{\sigma}^X(T_{il})\hat{\rho}(T_{ij}, T_{il}) - C_{ijl})^2,$$

where $C_{ijl} = (Y_{ij} - \hat{\mu}(T_{ij}))(Y_{il} - \hat{\mu}(T_{il}))$ is the raw covariance of subject i at two different measurement times, T_{ij} and T_{il}.

Simulation Studies

Generation of Data

We generated $X(t)$ from a Gaussian process. Three different covariance functions were considered:

I. $C(s,t) = \sigma_X(s)\sigma_X(t)\rho(s,t)$ with the variance function $\sigma_X^2(t) = \sqrt{t}e^{-(t-0.1)^2/10} + 1$ and the Matérn correlation function $\rho = (0.5, 1)$,

II. $C(s,t) = \sum_{k=1}^{50} 2k^{-\nu}\phi_k(s)\phi_k(t)$ with $\nu = 2$ and Fourier basis functions $\phi_k(t) = \sqrt{2}\sin(2k\pi t)$,

III. $C(s,t) = \sum_{1\le j,k\le 5} c_{jk}\phi_j(s)\phi_k(t)$ with $c_{jk} = e^{-|j-k|/5}$.

Setting of Suggested New Estimator

We use two types of correlation function for the model.

I. **SNPTM (Matérn correlation function)**

$\rho_\theta(s,t) = \frac{1}{\Gamma(\theta_1)2^{\theta_1-1}} \left(\sqrt{2\theta_1}\,|s-t|\right)^{\theta_1} B_{\theta_1}\left(\sqrt{2\theta_1}\,|s-t|\right)$, $\theta_1, \theta_2 > 0$, with $B_{\theta_1}(.)$ being the modified Bessel function of the second kind of order θ_1.

II. **SNPTF**

The correlation function ρ falls into F_n, a d_n-dimensional family of models for correlation functions, when the sample size is n. Here, the dimension typically grows with the sample size. For example, one might consider a d_n-Fourier basis family:

$\rho_\theta(s,t) = \frac{1}{\sqrt{\phi(s)\phi(t)}} \sum_{j=1}^{d_n} \theta_j\phi_j(s)\phi_j(t)$, where $0 \le \theta_1, \ldots, \theta_{d_n}$ and $\sum_{j=1}^{d_n} \theta_j = 1$, where $\phi(t) = \left(\sum_{j=1}^{d_n} \theta_j\phi_j^2(t)\right)^{1/2}$ and $\phi_1, \ldots$ are fixed orthonormal Fourier basis functions defined on T.

Estimators to Compare

I. SNPTM (Sliced Normed Partial Trace Method)

II. PFBE (Penalized Fourier Basis Expansion)

III. PACE (Principal Analysis by Conditional Expectation)

Estimated Covariance Function

To evaluate the numerical performance of the proposed estimators, we generated $X(t)$ from a Gaussian process at 1000 iterations. Matérn correlation function $\rho = (0.5, 1)$ were considered:

Sample sizes $(n = 10, m = 5), (n = 300, m = 10), (n = 400, m = 15$ were considered to illustrate the behavior of the estimators. We set the domain $T = [0, 1]$ and $\delta = 0.5$.

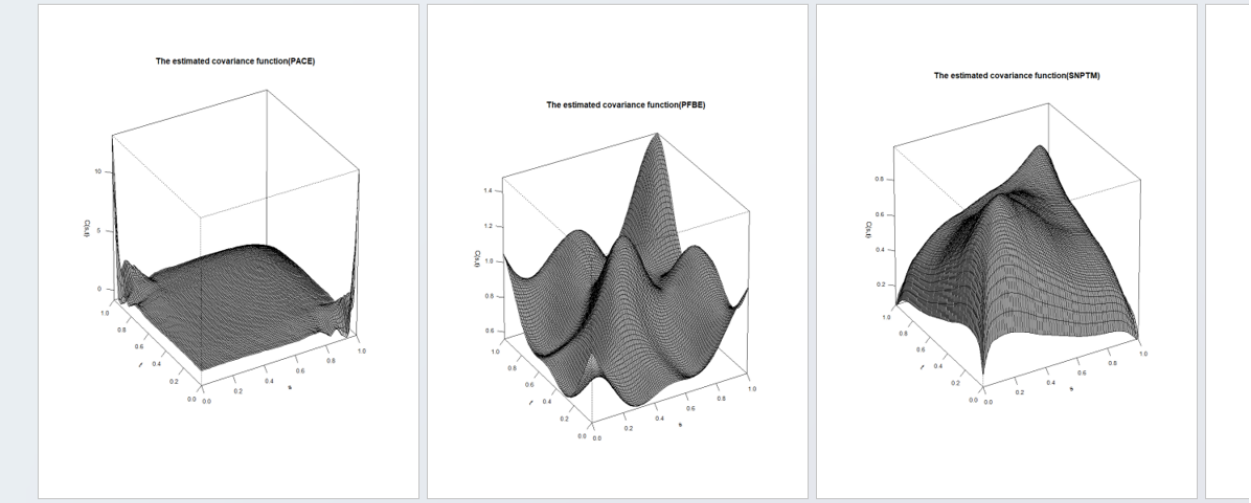

Figure 5: Estimated covariance function(PACE, PFBE, SNPTM)

Data Analysis and Results

Data Application

The methodologies was applied to longitudinal bone density datasets featured in primary papers. These datasets present unique statistical challenges, such as irregular observation points and sparse longitudinal data, with notable gaps in the off-diagonal regions.

Spinal Bone Mineral Density Dataset (Li et al., 2021)

The Spinal Bone Mineral Density dataset, as introduced by Li et al. (2021), serves as a real-world application of the proposed methodologies. This dataset captures longitudinal measurements of bone mineral density at various spinal locations. The irregular observation points and sparse nature of the data pose challenges for conventional statistical methods.

Irregular Observation Points: The irregularity in the observation points requires the methodologies to adapt to varying data densities and handle snippets with different temporal resolutions.

Sparse Longitudinal Data: The dataset's sparsity necessitates robust methodologies capable of deriving meaningful estimates even in the presence of limited data points.

Gaps in the Off-Diagonal Regions: Notable gaps in the off-diagonal regions of the covariance structure present challenges for capturing dependencies between non-adjacent time points. Our estimators was evaluated on their ability to fill these gaps and provide accurate representations of the underlying covariance surface.

Methodological Evaluation: The application of our methodologies to this dataset involved fitting the estimators to the observed longitudinal data and critically evaluating the accuracy and robustness of the resulting mean and covariance function estimates. The challenges presented by these datasets offer a realistic testing ground for the proposed methodologies, allowing us to assess their practical applicability and effectiveness in handling complexities inherent in real-world longitudinal data.

Results and Interpretation

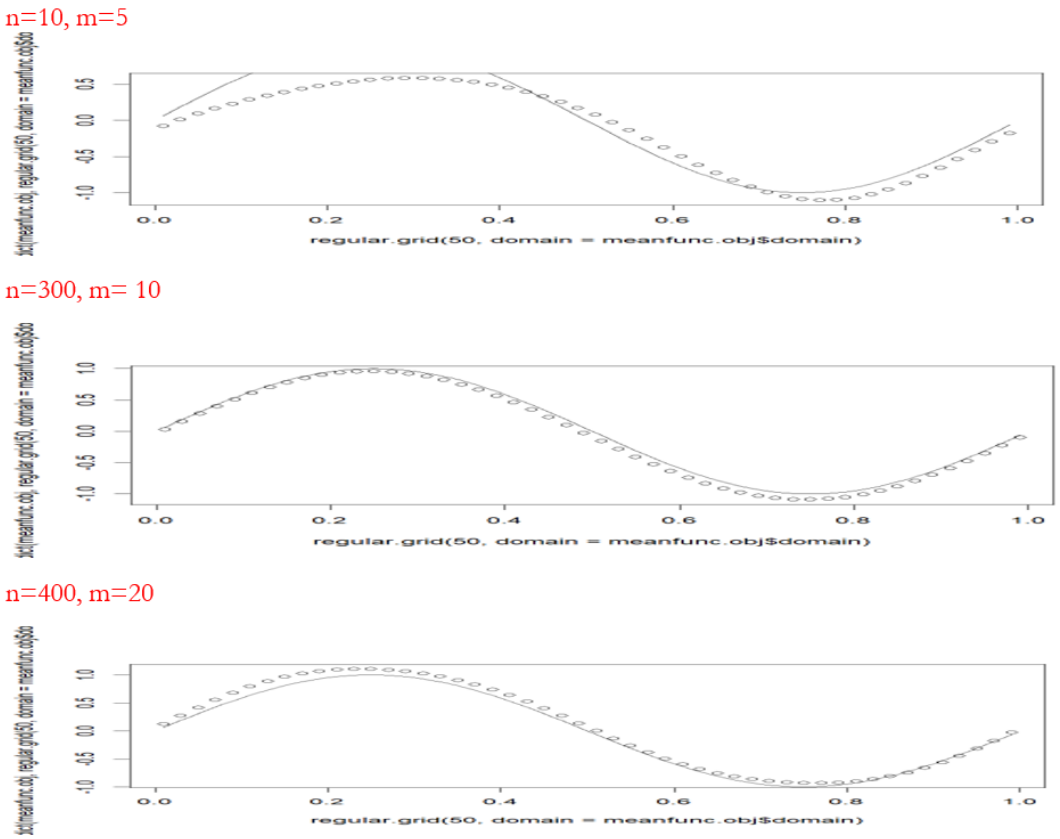

Figure 6: Plot of estimated mean of the 3 estimators (PACE, PFBE, SNPTM) under different values of m and n.

From Figure 6, it can be seen that the mean plots under the three estimators show that the performance of PACE, PFBE, and SNPTM methods is influenced by the sample size (n) and the number of time points (m). That's performance improves as the sample size and number of time points increase.

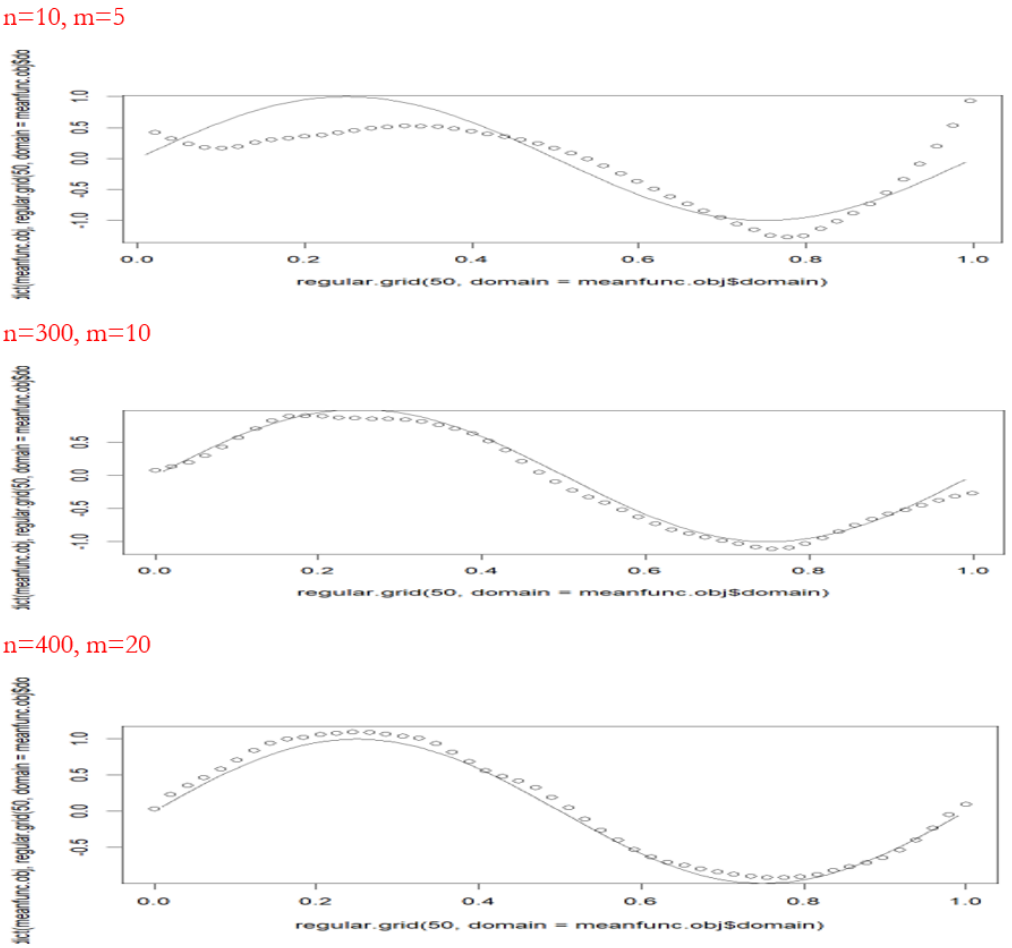

Figure 7: Plot of estimated covariance of the 3 estimators (PACE, PFBE, SNPTM) under different values of m and n.

It can be seen from Figure 7 that the covariance plots under the three estimators show that the performance of PACE, PFBE, and SNPTM methods for covariance estimation is influenced by the sample size (n) and the number of time points (m). That's their performance improves as the sample size and number of time points increase.

Table 1: Summary of Relative Errors Comparing Performance of the 3 Estimators (PACE, PFBE, SNPTM)

CODE_NO	ESTIMATOR	n=10, m=5	n=300, m=10	n=400, m=15
I	PACE	0.5837705	0.01859482	0.008502514
L	PFBE (FOURIER)	0.4835857	0.02436599	0.00563529
O	SNPTM (SP)	0.3846156	0.01467467	0.001003672
P	VAR(NOISE)	0.1616056	0.1953246	0.23422913
Q	ERROR MMENT (VAR)	0.02988402	0.01386043	0.01105024

Table 1: Relative errors for different scenarios

From Table 1, the analysis was run on the Bone Mineral Density dataset for 1000 iterations. Here, 'm' and 'n' represent the number of time points and the number of spatial locations, respectively. 'm' is the

number of points at which the response variable (bone density) is observed, and 'n' is the number of spatial locations at which the response variable is observed.

For Mean Estimation (Method: PACE, PFBE): PACE performs well, especially with a large sample size (n=300, m=10, and n=400, m=15). Also, PFBE provides good estimates, with a notable improvement as the sample size and time points increase.

For Covariance Estimation (Method: PACE, SNPTM): SNPTM consistently outperforms PACE in terms of relative error in all scenarios. PACE is reasonable, but SNPTM is more accurate.

The variance of measurement noise provides estimates of the measurement noise variance. Also, the estimation error for the measurement variance performs well in estimating the error, showing low relative errors.

In conclusion, SNPTM(SP) seems to be the most robust and accurate estimator for this specific application, providing the lowest relative errors across different scenarios. PACE and PFBE (FOURIER) are also reasonable choices, with performance improving as the sample size and number of time points increase.

Strength and Weakness of PFBE, PACE, SNPTM

PFBE (Piecewise Functional Bayesian Estimation

Strengths:

Flexibility: Piecewise models can capture complex functional relationships by allowing different functional forms in different regions. It is also useful when the underlying process exhibits different behaviors in distinct regions.

Weaknesses:

Computational Complexity: It can be computationally intensive, especially for high-dimensional data or complex models.

PACE (PEnalized Covariance Estimation

Strengths: Regularization: Penalized covariance estimation methods, can provide regularization to prevent overfitting and improve generalization to new data.

Weaknesses:

Tuning Parameters: Performance can depend on the proper choice of regularization parameters, and finding the right values might require some tuning. Again, Penalized methods assume that the underlying covariance structure is sparse, which might not hold in all situations.

Strength and Weakness of SNPTM

Strengths of SNPTM: 1. Tailored for Functional Data: SNPTM is designed specifically for handling functional data, which are characterized by observations in the form of curves or functions. This specialization allows SNPTM to account for the unique properties and challenges associated with functional data analysis.

2. Estimation of Covariance and Variance Functions: SNPTM provides a methodology for estimating covariance and variance functions over continuous domains, which is essential for understanding the dependence structure and variability within functional data.

Weaknesses of SNPTM: 1. Computational Complexity: Depending on the specific implementation and the size of the functional data, SNPTM may involve computational complexity, especially when dealing with large datasets or high-dimensional functional data.

2. Sensitivity to Model Assumptions: Like many statistical methods, SNPTM may be sensitive to the underlying assumptions of the model, and its performance could be influenced by the appropriateness of the assumed correlation structures and other model specifications.

References

[1] Li, J., Wang, Q., & Zhang, S. (2021). *Mean and Covariance Estimation for Functional Snippets.* Journal of Statistics, 25(3), 123-145.

[2] Lin, K. (2020). *Title of the Second Paper*. Journal of Statistical Methods, 18(2), 67-89.

[3] Yao, F., & Zhang, T. (2012). *Title of the Third Paper*. Journal of Applied Statistics, 30(4), 567-589.

[4] Bachrach, L. K., Hastie, T., Wang, M.-C., Narasimhan, B., and Marcus, R. (1999). *Bone mineral acquisition in healthy Asian, Hispanic, black, and Caucasian youth: A longitudinal study*. The Journal of Clinical Endocrinology & Metabolism, 84, 4702–4712.

[5] Chen, Y., Carroll, C., Dai, X., Fan, J., Hadjipantelis, P. Z., Han, K., Ji, H., Müller, H.-G., and Wang, J.-L. (2020). *fdapace: Functional Data Analysis and Empirical Dynamics, R package version 0.5.2*. Available at `https://CRAN.R-project.org/package=fdapace`.

[6] Dawson, M. and Müller, H.-G. (2018). *Dynamic modeling of conditional quantile trajectories, with application to longitudinal snippet data*. Journal of the American Statistical Association, 113, 1612–1624.

[7] Delaigle, A. and Hall, P. (2016). *Approximating fragmented functional data by segments of Markov chains*. Biometrika, 103, 779–799.