

Bayesian smoothing of functional data

Jingjing Yang, Hongxiao Zhu, Dennis Cox

Rice University
Virginia Tech

June 5, 2014



- 1 Introduction
- 2 Bayesian hierarchical model
 - Model assumptions
 - Markov chain Monte Carlo (MCMC) algorithm
- 3 Simulation studies
 - Stationary Gaussian data
 - Non-stationary Gaussian data
- 4 Real case studies
 - Spectroscopy data
 - Metabolic data
- 5 Discussion

- 1 Introduction
- 2 Bayesian hierarchical model
- 3 Simulation studies
- 4 Real case studies
- 5 Discussion

Functional data definition (Ramsay & Silverman 2002)

A group of data that are defined by functions, most often smooth functions, which could be any function varying over a continuum.

- Existing smoothing methods: smoothing splines, kernel smoothing with local polynomials, etc.
- We propose a Bayesian hierarchical model (BHM) that simultaneously performs model based smoothing and estimating in the functional data model.

Potential Advantages of BHM

- Borrow strength across all observations to hopefully do a better job of smoothing each one individually.
- Keep local features that may be smoothed out by typical individual curve smoothers
- Give nonparametric estimates for functional mean and covariance
- Give credible intervals of both the smoothed functions and the mean and covariance functions.

- 1 Introduction
- 2 Bayesian hierarchical model
 - Model assumptions
 - Markov chain Monte Carlo (MCMC) algorithm
- 3 Simulation studies
- 4 Real case studies
- 5 Discussion

Model assumptions

Suppose the functional data are noisy observations of n independent trajectories from a Gaussian process, where the i^{th} trajectory has p_i observed values on a grid $\mathbf{t}_i = \{t_{i1}, \dots, t_{ip_i}\}$, $i = 1, \dots, n$.

Signal + Noise model

$$Y_i(t_{ij}) = Z_i(t_{ij}) + \varepsilon_{ij}; \quad t_{ij} \in \mathcal{T}, \quad i = 1, \dots, n, \quad j = 1, \dots, p_i, \quad (1)$$

- $Z_i(\cdot)$: the true Gaussian trajectory i.i.d. $\sim GP(\mu(\cdot), \Sigma(\cdot, \cdot))$ called a signal;
- The noise terms ε_{ij} : i.i.d. $\sim N(0, \sigma_\varepsilon^2)$ and independent of the signals;

Prior distribution

- For the mean, a natural choice is $\mu(\cdot) | \Sigma(\cdot, \cdot) \sim GP(\mu_0(\cdot), \frac{1}{c}\Sigma(\cdot, \cdot))$, for some $c > 0$;
- Another natural choice: $\sigma_\varepsilon^2 \sim \text{Inverse-Gamma}(a_\varepsilon, b_\varepsilon)$, with shape parameter a_ε and scale parameter b_ε .
- The covariance is much harder.

Prior distribution (cont.)

- Based on results in Dawid (1981), Zhu, Dunson, and Strawn (2011) show that if $\Psi(s, t)$ is a covariance function and $\delta > 0$, then the $\text{InvertedWishart}(n + \delta, [\Psi(s_i, s_j) : 1 \leq i, j \leq n])$ defines a family of finite dimensional distributions that satisfy the Kolmogorov consistency conditions.
- Hence, we may define an Inverted Wishart Process (IWP) and use it for a prior on the covariance function:
 $\Sigma(\cdot, \cdot) \sim \text{IWP}(\delta, \Psi(\cdot, \cdot))$.
- We use $\Psi(s, t) = \sigma_s^2 A(|s - t|, \rho)$, where $A(D(\cdot, \rho))$ is a Matérn correlation function with spatial scale parameter ρ (with fixed value given by an empirical estimate) and order 2.5.
- We also put a prior $\sigma_s^2 \sim G(a_s, b_s)$.

Markov chain Monte Carlo (MCMC) algorithm

We assume for now a common grid for the observations:
 $t_1, \dots, t_p$. If we are at step k , we compute updates as follows:

- Step 1: Conditioning on Y_i , $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}^{(k)}$)⁻¹ we can update the signal estimates using standard multivariate normal calculations.
- Step 2: Update the noise variance parameter with
 $(\sigma_\varepsilon^2)^{(k+1)} \sim$
 $IG\left(a_\varepsilon + \frac{\sum_{i=1}^n p_i}{2}, \quad b_\varepsilon + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^p (Y_i(t_j) - Z_i^{(k+1)}(t_j))^2\right).$

Markov chain Monte Carlo (MCMC) algorithm (cont.)

RICE

- Step 3: Conditioning on the current signal estimates we can update parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ (on the common grid) using standard calculations with conjugate normal-inverted Wishart.
- Step 4: Given $\boldsymbol{\Sigma}^{(k+1)}$, we can update the hierarchical parameter $(\sigma_s^2)^{(k+1)}$ from $(\sigma_s^2)^{(k+1)} | \boldsymbol{\Sigma}^{(k+1)} \sim IG\left(a_s + \frac{(\delta+p-1)p}{2}, b_s + \frac{1}{2}\text{trace}(A(\boldsymbol{\Sigma}^{(k+1)})^{-1})\right)$.

Non-common observation grid

- This algorithm can be easily adapted to the setting where the union of all the observation grids is still rather small, which happens sometimes with longitudinal data.
- However, if we have data that are observed on random grids, or if the grid is too fine (so matrices are too large), we propose to update the signals Z_i , mean μ , and covariance Σ on a fixed "working grid" $\boldsymbol{\tau}$, and then use splines to interpolate them on the observation grids $\{t_{ij} : j = 1, \dots, p_i\}, 1 \leq i \leq n$.
- Interpolation of Covariance function:

$$\Sigma(s, t) = b(s)^T B^{-1} \Sigma(\boldsymbol{\tau}, \boldsymbol{\tau}) B^{-T} b(t),$$

where $b(t) = (b_1(t), \dots, b_p(t))$, $b_i(t)$ is i 'th basis spline function, and $B = [b(\boldsymbol{\tau}_1) \cdots b(\boldsymbol{\tau}_p)]$.

Modified MCMC for Spline Approximation

- Step 1: Update $Z_i(\boldsymbol{\tau})$ from $Z_i(\boldsymbol{\tau})|(Y_i(t_i), \mu(\cdot), \Sigma(\cdot, \cdot)) \sim MN[m_{Z_i|Y_i}, V_{Z_i|Y_i}]$, where $m_{Z_i|Y_i} = \mu(\boldsymbol{\tau}) + \Sigma(\boldsymbol{\tau}, t_i)(\Sigma(t_i, t_i) + \sigma_\epsilon^2 I)^{-1}(Y_i(t_i) - \mu(t_i))$, $V_{Z_i|Y_i} = \Sigma(\boldsymbol{\tau}, \boldsymbol{\tau}) - \Sigma(\boldsymbol{\tau}, t_i)(\Sigma(t_i, t_i) + \sigma_\epsilon^2 I)^{-1}\Sigma(t_i, \boldsymbol{\tau})$.
- Step 2: Update σ_ϵ^2 with interpolated $Z_i(t_{ij})$.
- Step 3: Update $\mu(\boldsymbol{\tau})$ and $\Sigma(\boldsymbol{\tau}, \boldsymbol{\tau})$ conditioning on $Z_i(\boldsymbol{\tau}), i = 1, 2, \dots, n$ as before.
- Step 4: Update σ_s^2 conditioning on $\Sigma(\boldsymbol{\tau}, \boldsymbol{\tau})$ as before.

- 1 Introduction
- 2 Bayesian hierarchical model
- 3 **Simulation studies**
 - Stationary Gaussian data
 - Non-stationary Gaussian data
- 4 Real case studies
- 5 Discussion

Where possible we compare with alternative methods:

- Smooth estimates of individual signals using GCV smoothing splines and local linear fits with plug in bandwidths (Fan and Gijbels, 1992)
- With individually smoothed signals, we can estimate μ and Σ
- Smooth covariance estimator PACE from Yao, Mueller, and Wang (2005)
- With PACE, we can also estimate signals Z_i and then μ .
- Computed RIMSE (square root of average mean squared error)

Gaussian data on common grid

For each simulation trial, we generated $n = 100$ stationary functional data curves (true signals) from GP with mean function $\mu(t) = 3 \sin(4t)$, and covariance function $\Sigma(s, t) = \text{Matern}(|s - t|; \nu = 5/2, \rho = 1/2, \sigma^2 = 5)$, denoted by $\mathbf{Z}^*$.

- All curves were obtained on the same equally spaced grid $\{t_{i1} = 0, \dots, t_{ip} = \pi/2\}$ with $p = 40$.
- The noise terms $\{\varepsilon_{ij}; i = 1, \dots, 100, j = 1, \dots, 40\}$ were generated i.i.d. from $N(0, \sigma_\varepsilon = \sqrt{5}/2)$.
- The observed noisy functional data curves were given by $\mathbf{Y} = \mathbf{Z}^* + \boldsymbol{\varepsilon}$.

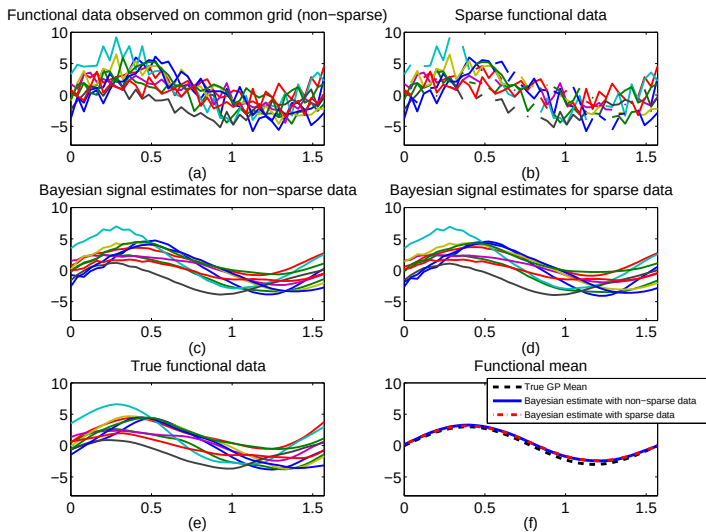
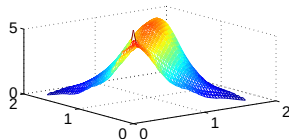
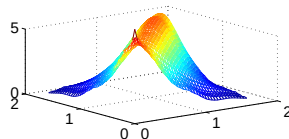


Figure 1 : First 10 stationary functional curves.

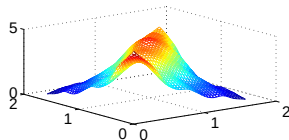
Bayesian covariance estimate (non-sparse)



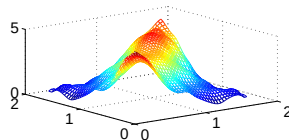
Bayesian covariance estimate (sparse)



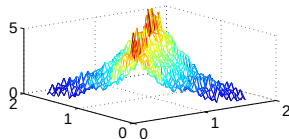
PACE covariance estimate (non-sparse)



PACE covariance estimate (sparse)



Sample covariance estimate (non-sparse)



True covariance

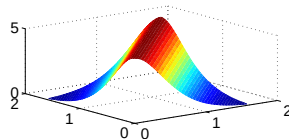


Figure 2 : Covariance estimates (stationary Gaussian data).

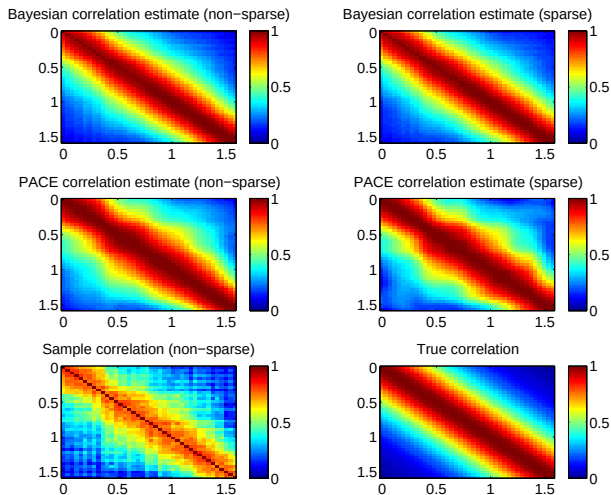


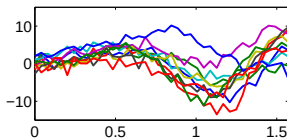
Figure 3 : Heat maps of correlation estimates (stationary Gaussian data).

Non-stationary Gaussian data

We generated non-stationary Gaussian data by imposing a nonlinear transformation on the variable t and a scalar function on the stationary Gaussian process.

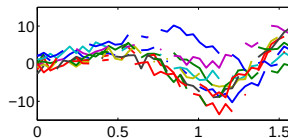
- Use nonlinear transformation function be $\xi(t) = (t)^{2/3}$ and scalar function be $h(t) = t + 1/2$.
- Let $\tilde{X}_i(t)$ denotes a functional trajectory generated from the stationary Gaussian process. Then the non-stationary true trajectory is given by $X_i(t) = h(t)\tilde{X}_i(\xi(t))$.
- For the case with common grid, 100 non-stationary functional trajectories were generated on the same equally spaced grid $t = \{t_{i1} = 0, \dots, t_{ip} = \pi/2\}$.
- Then we obtained the noisy functional data Y by adding on noise terms i.i.d. from $N(0, \sigma_\varepsilon = \sqrt{5}/2)$.

Functional data observed on common grid (non-sparse)



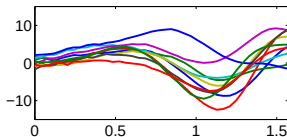
(a)

Sparse functional data



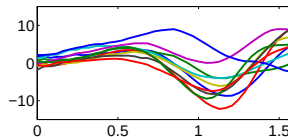
(b)

Bayesian signal estimates for non-sparse data



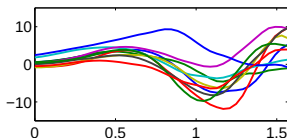
(c)

Bayesian signal estimates for sparse data



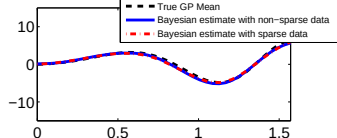
(d)

True functional data



(e)

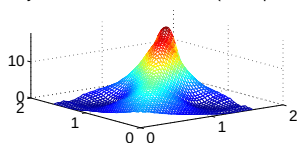
Functional mean



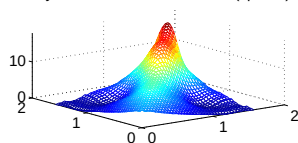
(f)

Figure 4 : First 10 non-stationary functional curves.

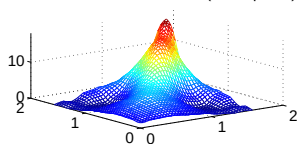
Bayesian covariance estimate (non-sparse)



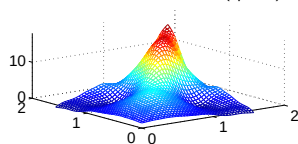
Bayesian covariance estimate (sparse)



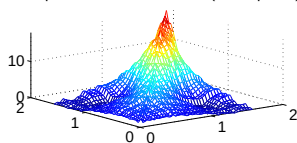
PACE covariance estimate (non-sparse)



PACE covariance estimate (sparse)



Sample covariance estimate (non-sparse)



True covariance

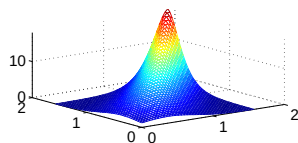


Figure 5 : Covariance estimates (non-stationary Gaussian data).

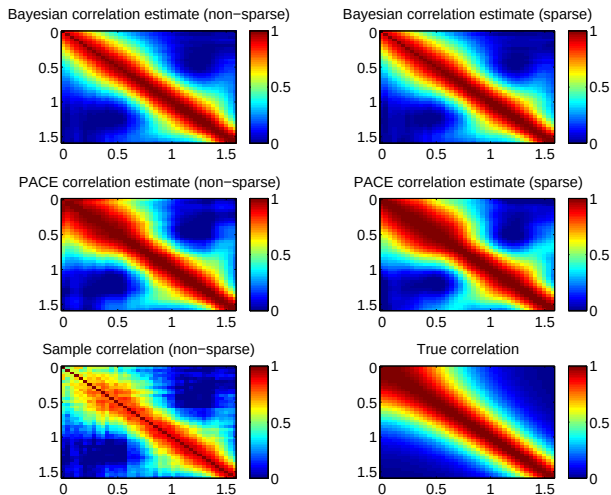


Figure 6 : Heat maps of correlation estimates (non-stationary Gaussian data).

The following results are for 100 simulations of samples of 100 curves, common grid with 40 equally spaced grid points, for stationary and non-stationary examples.

Table 1 : Average RIMSEs (along with standard errors) for signal estimates in comparison with the best possible estimates (conditional mean), which gives a lower-bound of the RIMSE.

Average RIMSEs (Signal)	Splines	Kernel	PACE	Bayesian	LowerBound
Stationary	0.6778 (± 0.0034)	0.6280 (± 0.0015)	0.5374 (± 0.0015)	0.5328 (± 0.0013)	0.5077
Non-Stationary	0.7690 (± 0.0030)	0.7554 (± 0.0016)	0.6207 (± 0.0014)	0.6401 (± 0.0014)	0.5814

Table 2 : Average RIMSEs (along with standard errors) for functional mean estimates in comparison with the best possible least square estimates.

Average RIMSEs (Mean)	Splines	Kernel	PACE	Bayesian	LowerBound
Stationary	0.2785 (± 0.0105)	0.3017 (± 0.0117)	0.2879 (± 0.0116)	0.2689 (± 0.0110)	0.2493 (± 0.0108)
Non-Stationary	0.3603 (± 0.0121)	0.4071 (± 0.0124)	0.4111 (± 0.0129)	0.3563 (± 0.0126)	0.3220 (± 0.0123)

Table 3 : Average RIMSEs (along with standard errors) for the covariance estimates from sample/PACE/Bayesian/sample with Bayesian smoothed data (Sample Smoothed) on the pooled grid.

Average RIMSEs (Covariance)	Sample	PACE	Bayesian	Sample Smoothed
Stationary	1.1503 (± 0.0277)	0.9598 (± 0.0309)	1.1110 (± 0.0489)	0.9407 (± 0.0316)
Non-Stationary	1.8522 (± 0.0467)	1.6813 (± 0.0502)	1.8534 (± 0.0433)	1.6820 (± 0.0498)

Stationary Gaussian data with Random grids

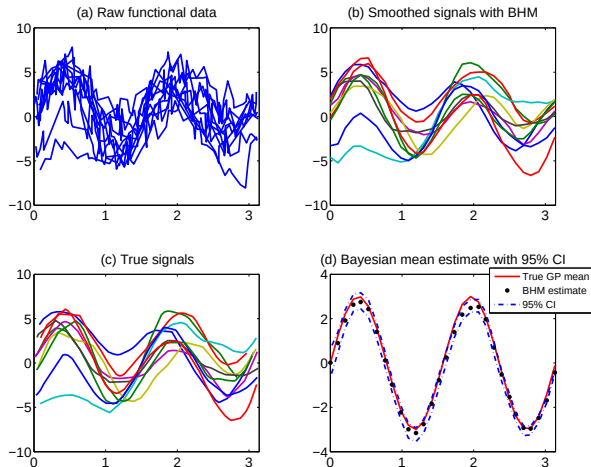


Figure 7 : First 10 functional curves and functional mean.

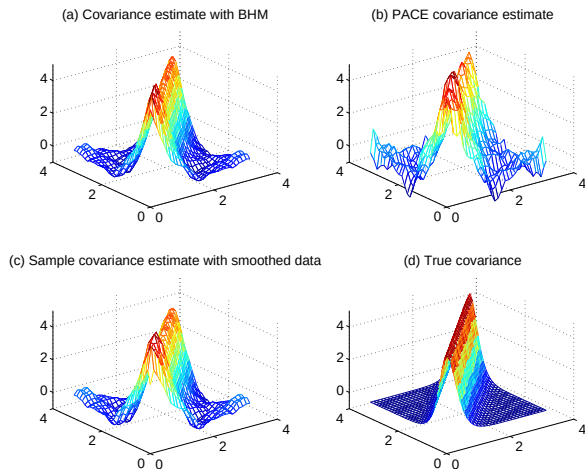


Figure 8 : Covariance estimates and the true covariance.

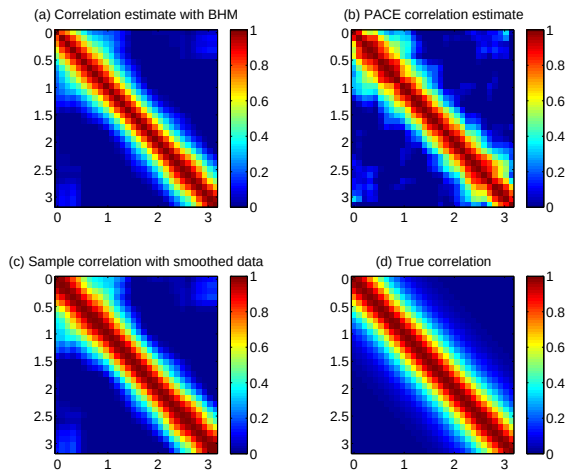


Figure 9 : Heat maps of the correlation.

- 1 Introduction
- 2 Bayesian hierarchical model
- 3 Simulation studies
- 4 Real case studies**
 - Spectroscopy data
 - Metabolic data
- 5 Discussion

Spectroscopy data

- $n=462$ reflectance spectroscopy measurements by one device on a standard
- Device is being investigated for diagnosing cervical cancer in the early stages.
- Each measurement was composed of intensity values (log10 transformed) at wavelengths from 410 to 700 nm.
- Since the functional spectroscopy data are very dense on the emission wavelengths (common grid), we took intensity values at 3nm equally spaced wavelengths from (410, 700) as our observations $\mathbf{Y}$.
- Compared predicted values at the left out observations

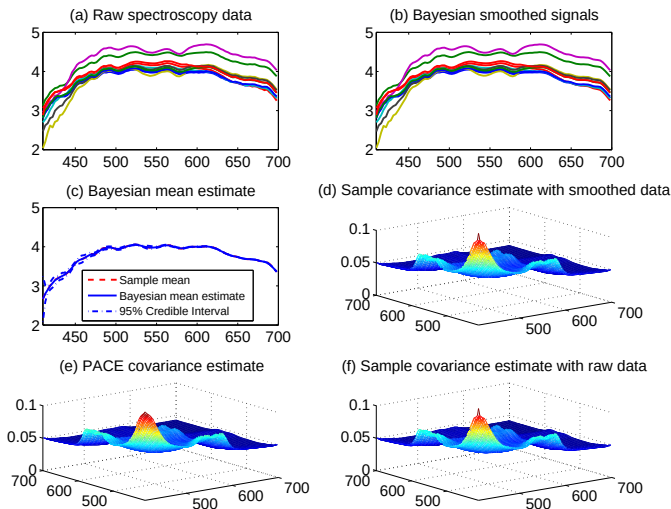


Figure 10 : Ten example raw-spectroscopy-curves and covariance

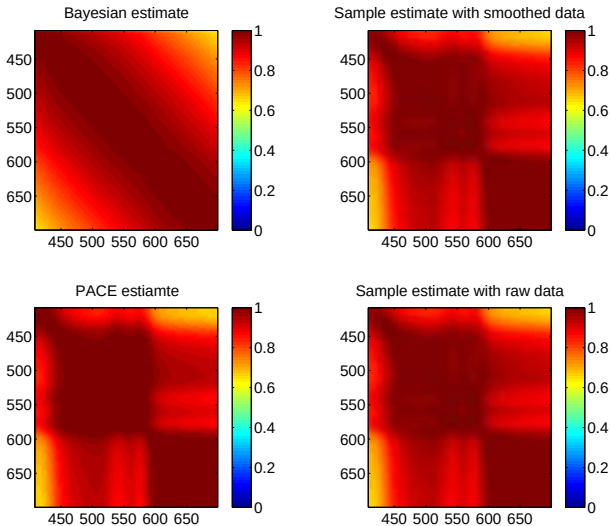


Figure 11 : Heat maps of correlation estimates.

Compare with individual smoothing methods

We used the remaining 2/3 of the data points as a validation data set to compare our Bayesian method with smoothing splines (smoothing parameter decided by GCV), kernel smooth (bandwidth selected by R function “dpill”) and PACE.

Table 4 : RMSEs of the validation subset of the spectroscopy data.

Splines	Kernel	PACE	Bayesian
0.0204	0.0390	0.0254	0.0174

In Table 4, the Bayesian method gave the smallest RMSE, which is a natural result of preserving features in the raw data.

Child Metabolic Data

Study was conducted in the Children's Nutrition Research Center (CNRC), at Baylor College of Medicine.

- measured the energy expenditure (EE) of 106 children over time (minute by minute), which can be treated as a functional evaluations over time.
- The data were collected during 24 hour periods for each child; however, we will only look at the data during the first 405 minutes of child's sleep time.
- We applied our Bayesian smoothing methods on the data of obese and non-obese groups separately. We compared with smoothing splines and kernel smoothing methods.

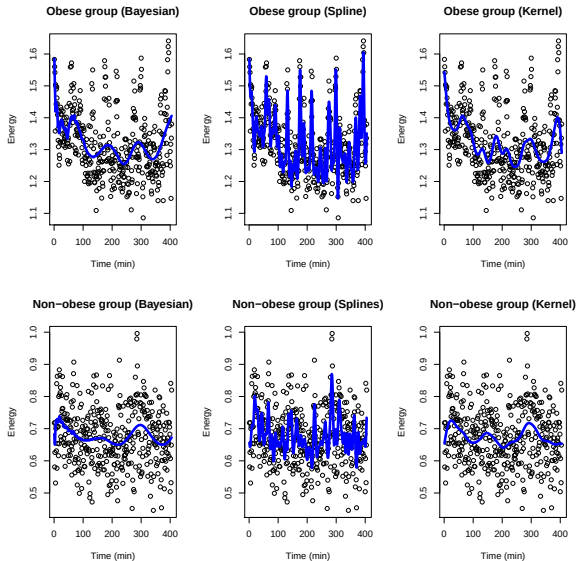


Figure 12 : Smoothed signal estimates.

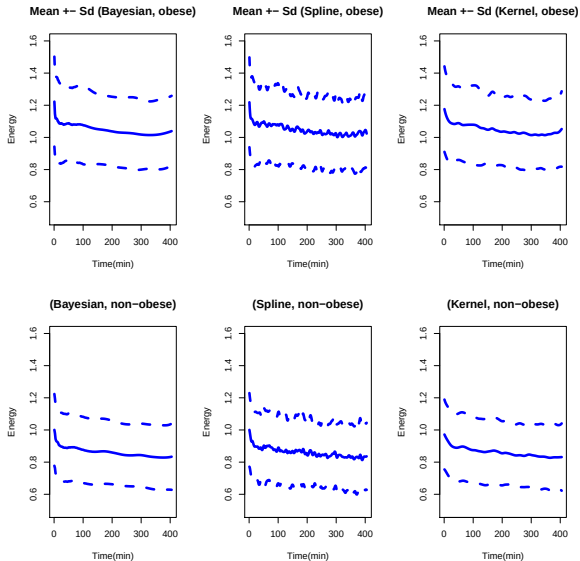


Figure 13 : Sample mean $\pm$ sample standard deviation with

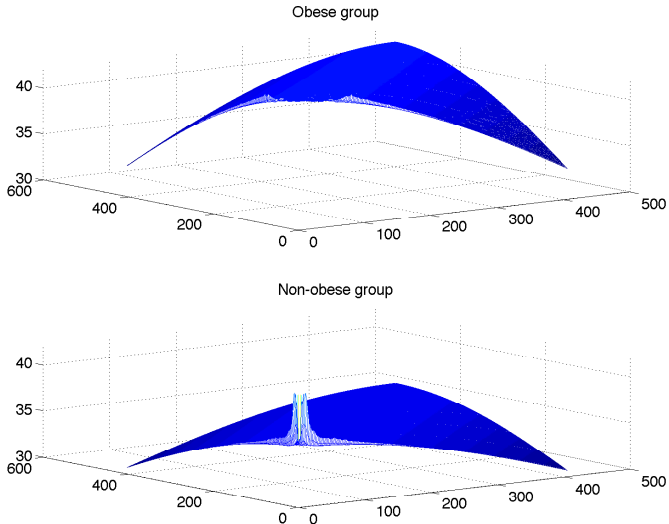


Figure 14 : Bayesian covariance estimates.

- 1 Introduction
- 2 Bayesian hierarchical model
- 3 Simulation studies
- 4 Real case studies
- 5 Discussion**

- Our Bayesian method beats the individual smoothing methods because of borrowing strength across all observations.
- Can also give good estimates along with accuracy assessment (Bayesian credible intervals) for the parameters in the model, i.e. signal mean and covariance.
- In the examples, the PACE method has better covariance estimates.
- How to improve our method? The hierarchical covariance prior with stationary scaling covariance seems too rigid. We would like to develop more flexible priors for the covariance function
- Of course, theoretical properties need to be developed. Besides the usual results (e.g. Bayesian consistency), properties of the Inverted Wishart Covariance Process would be of interest.

Acknowledgement

- This research was partially supported by NIH grant PO1CA082710.
- Thank you for your time.
- And a Special Thanks to Grace Wahba for all she has done for our field.

