Adaptive Bayesian Sum of Trees Model for Covariate Dependent Spectral Analysis

Scott A. Bruce[1]

Department of Statistics
Texas A&M University

Yakun Wang, Department of Statistics, George Mason University
Zeda Li, Paul H. Chook Department of Information System and Statistics, Baruch College, The City University of New York

May 24, 2022

[1]Research is supported by the National Institute Of General Medical Sciences of the NIH under Award Number R01GM140476.

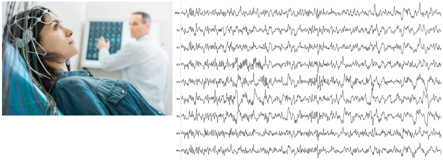
Examples of Biomedical Time Series Data

Clinicians and researchers collect a variety of time series data whose oscillatory patterns are of interest.

Examples of Biomedical Time Series Data

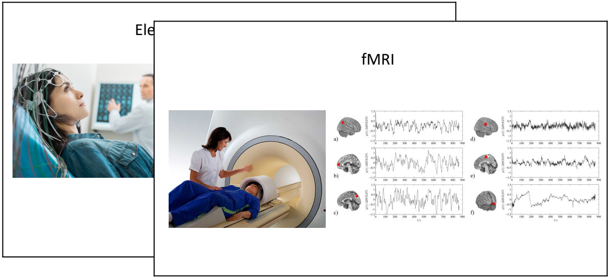
Clinicians and researchers collect a variety of time series data whose oscillatory patterns are of interest.

Electroencephalography (EEG)



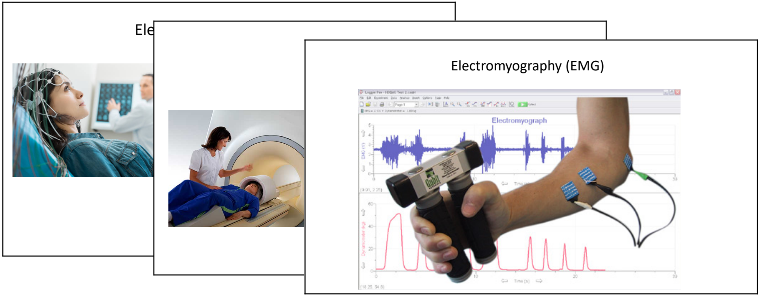
Examples of Biomedical Time Series Data

Clinicians and researchers collect a variety of time series data whose oscillatory patterns are of interest.



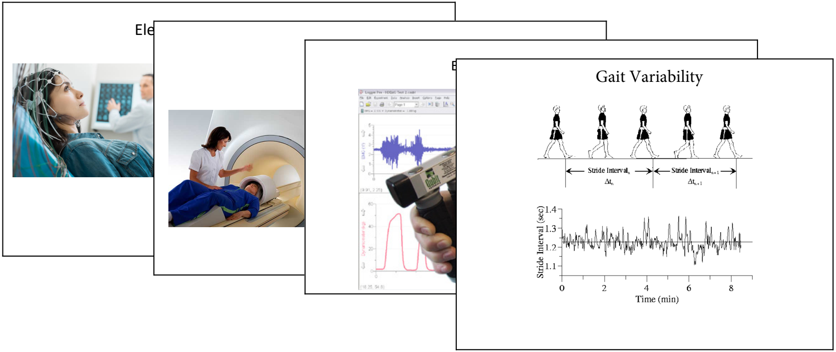
Examples of Biomedical Time Series Data

Clinicians and researchers collect a variety of time series data whose oscillatory patterns are of interest.



Examples of Biomedical Time Series Data

Clinicians and researchers collect a variety of time series data whose oscillatory patterns are of interest.



Gait Maturation Study [Hausdorff et al. (1999)]

Maturation of gait dynamics

- Immature gait in very young children results in unsteady walking patterns and frequent falls.
- Gait is relatively mature by age 3. However, neuromuscular control continues to develop well beyond age 3.
- Researchers are interested in determining if stride-to-stride dynamics continue to become more steady and regular beyond age 3.

Gait Maturation Study [Hausdorff et al. (1999)]

Stride-to-stride time series data

- $N = 50$ healthy children ages 3-14.
- $T = 256$ stride times recorded after removing stride times in the first 60 seconds and last 5 seconds.
- Age, gender, height, weight, leg length, and gait speed are also collected for each child.

Goal: To better understand the maturation of gait dynamics with age in the presence of other related covariates.

Data From Three Subjects

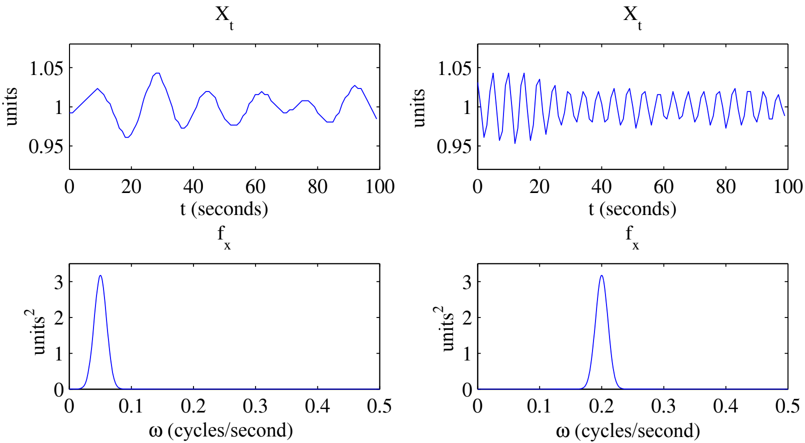
Stationary Time Series

- Consider a zero-mean stationary time series X_t.
- Cramér Representation [Cramér (1942)]:

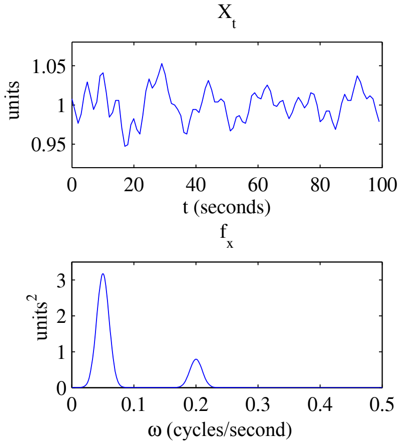
$$X_t = \int_{-1/2}^{1/2} A(\nu) \exp(2\pi i \nu t) dZ(\nu).$$

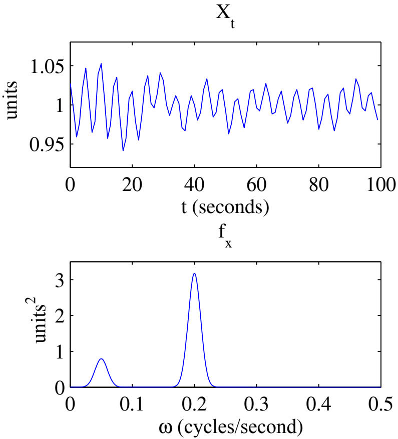
- **Power spectrum**: $f(\nu) = |A(\nu)|^2$.
- The power spectrum represents a decomposition of variance over frequencies.
- $\mathrm{Var}(X_t) = \int_{-1/2}^{1/2} f(\nu) d\nu$.

Two Simulated Examples



Two More Simulated Examples





Periodogram Estimator

- Periodogram from $\mathbf{X} = (X_1, \ldots, X_T)'$:

$$I(\nu_k) = \frac{1}{T} \left| \sum_{t=1}^{T} X_t \exp(-2\pi i \omega_k t) \right|^2.$$

- $\nu_k = k/T$, $k = 1, \ldots, n = \lfloor T/2 \rfloor - 1$.
- Unbiased but noisy estimates of $f(\nu)$.
- Approximately distributed as scaled χ^2 to provide the Whittle likelihood:

$$p(\mathbf{x}|\mathbf{f}) \approx (2\pi)^{-n/2} \prod_{k=1}^{n} \exp\left\{ -\frac{1}{2} [\log f(\nu_k) + I(\nu_k)/f(\nu_k)] \right\}$$

Smoothing

- Periodogram can be smoothed to obtain a consistent estimate.
- One approach - **Bayesian penalized linear spline** [Wahba (1990)]:

$$\log f(\nu) \approx \alpha + \sum_{s=1}^{S} \beta_s \cos(2\pi s \nu)$$

- Priors [Rosen, Wood, and Stoffer (2012)]

$$\begin{aligned} \alpha &\sim N(0, \sigma_\alpha^2) \\ \boldsymbol{\beta} &\sim N(\mathbf{0}, \tau^2 D_S), \text{where } D_S = \text{diag}(\{\sqrt{2}\pi s\}^{-2}) \\ \tau &\sim \text{half-t} \end{aligned}$$

- Sampling via Metropolis-Hastings and Gibbs steps.

Covariate-dependent Power Spectrum

- Covariate-dependent Cramér Representation

$$X_{\ell t} = \int_{-1/2}^{1/2} A(\boldsymbol{\omega}_\ell, \nu) \exp(2\pi i t \nu) dZ_\ell(\nu),$$

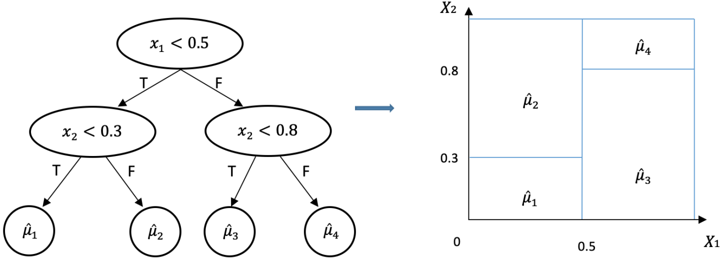
where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_P)'$ is a P-dimensional covariate vector, and $\ell = 1, \ldots, L$ independent subjects.

- **Covariate-dependent power spectrum**:
$f(\boldsymbol{\omega}, \nu) = |A(\boldsymbol{\omega}, \nu)|^2$.

- **Goal**: Develop an adaptive method that can capture both smooth and abrupt changes in power spectra across multiple covariates and provide a tool for variable selection.

One Option: Tree-based Approach

- Regression tree illustration



- Tree-based models provide a flexible and parsimonious approach for partitioning multiple covariates.
- For scalar responses: **Bayesian Additive Regression Tree** (BART) model [Chipman et al. (2010)]

Adaptive Bayesian Sum of Trees Model

- **Idea**: Develop a Bayesian sum-of-trees model for $\log f(\boldsymbol{\omega}, \nu)$

$$\log f(\boldsymbol{\omega}, \nu) \approx \sum_{j=1}^{M} \sum_{b=1}^{B_j} \delta(\boldsymbol{\omega}; U_j, b) \log f_{bj}(\nu),$$

  - M is the number of trees
  - U_j represents the jth tree that has B_j terminal nodes
  - δ is a function that identifies terminal node membership such that $\delta(\boldsymbol{\omega}_\ell; U, b) = 1$ if the ℓth observation falls into the bth terminal node and $\delta(\boldsymbol{\omega}_\ell; U, b) = 0$ otherwise.

- Model specification for $\log f_{bj}(\nu)$ then follows directly from the Bayesian penalized linear spline introduced previously.

Tree Structure Priors

- A regularization prior is applied to encourage each tree to be a weak learner:

$$\Pr(\text{SPLIT}) = \alpha(1+d)^{-\theta}, \qquad \alpha \in (0,1),\ \theta \in [0,\infty),$$

d is the depth of a tree, $\alpha = 0.95$ and $\theta = 2$ as default. [Chipman et al. (2010)]

- Terminal node parameters and trees are assumed to be independent a priori.
- Uniform priors on split variables and cut points.
- Sparsity-inducing Dirichlet prior on split variables can also be used for improved variable selection. [Linero, 2018]

Sampling Scheme

- Backfitting Markov chain Monte Carlo (MCMC) on 'residual' of periodogram

$$\mathbf{R}_{\ell j}(\nu_k) = \log \boldsymbol{I}_\ell(\nu_k) - \sum_{i \neq j} \sum_{b=1}^{B_j} \delta(\boldsymbol{\omega}; U_i, b) \log f_{bi}(\nu)$$

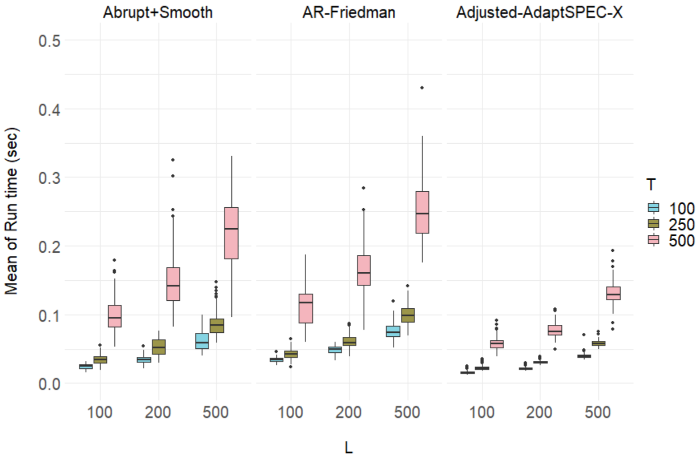
  allows for updating each individual tree structure in turn.

- Reversible jump MCMC
  - Birth: splitting a terminal node into two child nodes
  - Death: dropping two terminal child nodes belonging to the same internal node
  - Change: modifying the variable and cut point associated with an internal node with two terminal child nodes

Overview of Proposed Approach

- Adaptively partition covariate space using tree structures.
- Bayesian penalized spline model for local spectra estimation within each terminal node.
- Bayesian Backfitting MCMC and Reversible jump MCMC techniques to sample from posterior of the trees
- Inference averaged over distribution of trees.

Run times



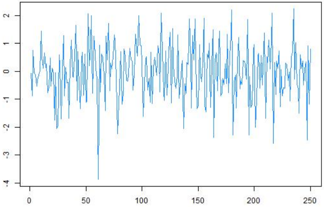
Mean run times for a single tree update over 100 replicates of the three simulation settings with $M = 5$ trees.

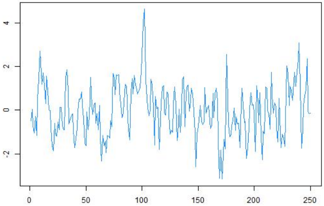
Simulated Abrupt+Smooth Example

- **AR(1)**: $x_{\ell t} = \phi_\ell x_{\ell t-1} + \epsilon_{\ell t}$
- $\phi_\ell =$

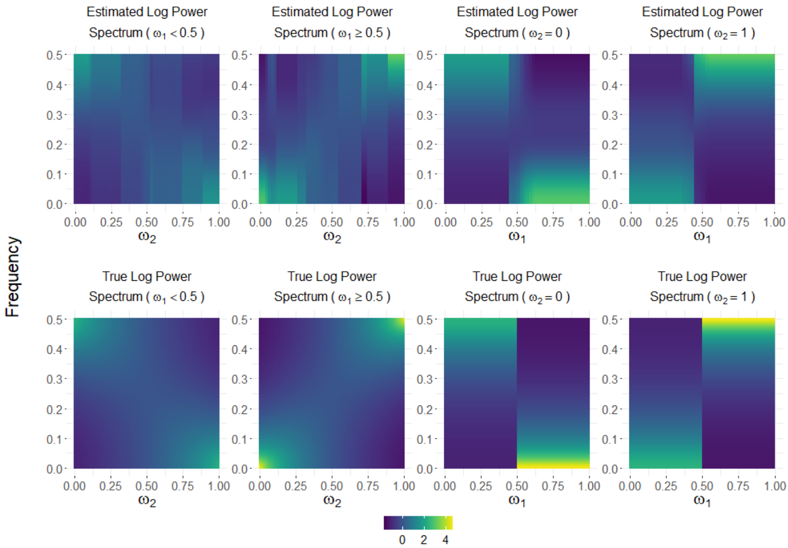
$$\begin{cases} -0.7 + 1.4\omega_2 & \text{for } 0 \leq \omega_1 < 0.5 \\ 0.9 - 1.8\omega_2 & \text{for } 0.5 \leq \omega_1 \leq 1, \end{cases}$$

- $\ell = 1, \cdots, L = 100$ subjects
- $t = 1, \cdots, T = 250$
- $\omega_1, \omega_2 \overset{i.i.d.}{\sim} U(0,1)$
- $\epsilon_{\ell t} \overset{i.i.d.}{\sim} N(0,1)$





Estimation Accuracy for Abrupt+Smooth Simulation

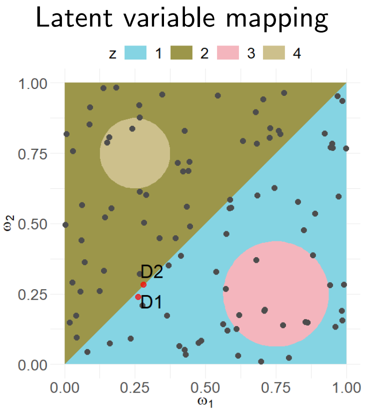


Simulated Latent Variable Example

- **AR(2)**:
  $x_{\ell t} = \phi_{z_\ell 1} x_{\ell t-1} + \phi_{z_\ell 2} x_{\ell t-2} + \epsilon_{\ell t}$

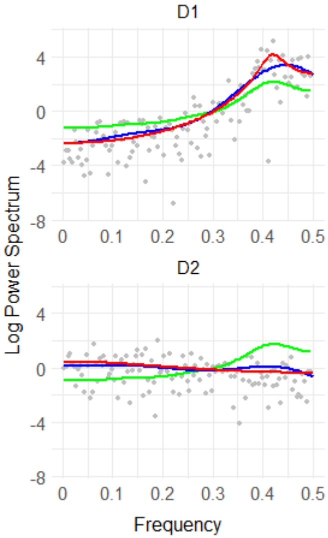
$$(\phi_{z_\ell 1}, \phi_{z_\ell 2}) = \begin{cases} (1.5, -0.75), z_\ell = 1 \\ (-0.8, 0), z_\ell = 2 \\ (-1.5, -0.75), z_\ell = 3 \\ (0.2, 0), z_\ell = 4 \end{cases}$$

- $\ell = 1, \cdots, L = 100$ subjects
- $t = 1, \cdots, T = 250$
- $\omega_1, \omega_2 \overset{i.i.d.}{\sim} U(0,1)$
- $\epsilon_{\ell t} \overset{i.i.d.}{\sim} N(0,1)$



Simulated Latent Variable Example

- Red line: true log power spectra
- Gray points: log periodogram ordinates
- Blue line: estimated log power spectra using the proposed Bayesian sum of trees model
- Green line: estimated log power spectra using the competing smooth model

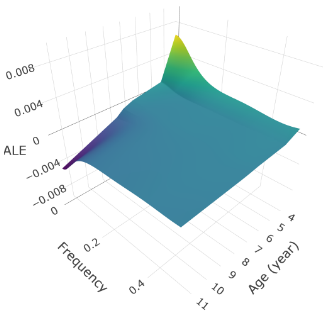


Gait Maturation Analysis

- Data:
  - $N = 50$, $T = 256$ stride-to-stride time series.
  - Ages 3-14 years old.
  - Age, gender and gait speed as covariates.
- Low frequencies (LF)(0.05-0.25 stride^{-1}) represent fluctuations over a longer-term scale (immature gait).
- High frequencies (HF) (0.25-0.5 stride^{-1}) represent fluctuations over a shorter-term scale (mature gait).

Covariate Effects on Power Spectrum

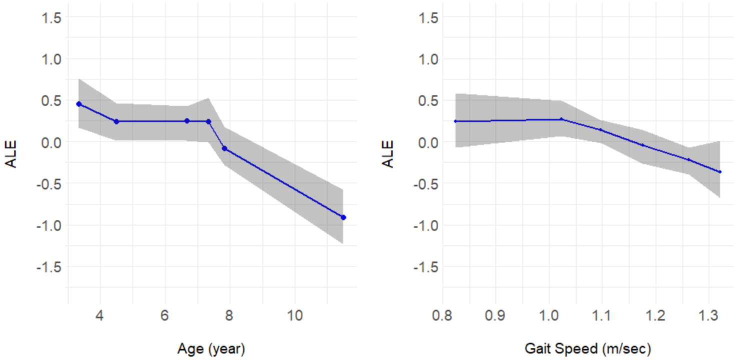
- Accumulated local effects (ALE) is a method for evaluating covariate effects.



1. Power over all frequencies decreases as age increase.
2. Power in low frequencies decreases much more with age relative to higher frequencies.

Covariate Effects on Power Spectrum

- Low-to-high frequency ratio $\widehat{\frac{\text{LF}}{\text{HF}}}(\omega) = \frac{\sum_{\nu_k \in (0.05,0.25)} \hat{f}(\omega,\nu_k)}{\sum_{\nu_k \in [0.25,0.5)} \hat{f}(\omega,\nu_k)}$.



Significant decreases for ages above 7 years and for speeds above 1 m/sec.

Current and Future Work

Current work

- Proposed a nonparametric adaptive Bayesian sum of trees model for covariate-dependent spectral analysis.
- Captures both abrupt and smooth changes.
- Handle complex nonlinear and interaction effects.

Future work

- Extend to time- and covariate-dependent time series.
- Apply alternative partitioning frameworks such as Voronoi tessellations. [Payne et al., 2020]

Current and Future Work

Current work

- Proposed a nonparametric adaptive Bayesian sum of trees model for covariate-dependent spectral analysis.
- Captures both abrupt and smooth changes.
- Handle complex nonlinear and interaction effects.

Future work

- Extend to time- and covariate-dependent time series.
- Apply alternative partitioning frameworks such as Voronoi tessellations. [Payne et al., 2020]

THANK YOU!

Gait Maturation Analysis

- Accumulated local effects (ALE) is a method for evaluating covariate effects
- ALE for $\omega_j = x$ on the power spectrum is

$$f_{j,\text{ALE}}(x, \nu) = \int_{z_{0,j}}^{x} E_{\boldsymbol{\omega}_{\backslash j}|\omega_j} \left[\frac{\delta f(\boldsymbol{\omega}, \nu)}{\delta \omega_j} \middle| \omega_j = z_j \right] dz_j - \text{constant}$$

- $\boldsymbol{\omega} = (\omega_j, \boldsymbol{\omega}_{\backslash j})$ where ω_j denotes the jth covariate and $\boldsymbol{\omega}_{\backslash j}$ denotes all covariates other than the jth covariate
- $\mathbb{Z}_j = \{z_{0,j}, \ldots, z_{H,j}\}$ is a collection of $H+1$ partition points over the effective support of ω_j
- The constant is a value to vertically center the plot