

ADAPTIVE BAYESIAN COVARIATE DEPENDENT SPECTRAL ANALYSIS
OF MULTIPLE TIME SERIES

by

Yakun Wang
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
In Partial fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Statistical Science

Committee:

<u><i>Pramita Bagchi</i></u>	Dr.Pramita Bagchi, Dissertation Director
<u><i>Anand Vidyashankar</i></u>	Dr.Anand Vidyashankar, Committee Member
<u><i>Wanli Qiao</i></u>	Dr.Wanli Qiao, Committee Member
<u><i>Scott A. Bruce</i></u>	Dr.Scott A. Bruce, Committee Member
<u><i>Zeda Li</i></u>	Dr.Zeda Li, Committee Member
<u><i>Jiayang Sun</i></u>	Dr.Jiayang Sun, Department Chair

Date: 05/08/2022 Summer 2022
George Mason University
Fairfax, VA

Adaptive Bayesian Covariate Dependent Spectral Analysis of Multiple Time Series

A dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy at George Mason University

By

Yakun Wang
Master of Science
George Washington University, 2016
Bachelor of Science
Taiyuan University of Technology, 2014

Director: Dr.Pramita Bagchi, Professor
Department of Statistics

Summer 2022
George Mason University
Fairfax, VA

Copyright © 2022 by Yakun Wang
All Rights Reserved

Dedication

I dedicate my dissertation work to my family and friends. A special feeling of gratitude to my loving parents.

Acknowledgments

I wish to thank my committee members who were more than generous with their expertise and precious time. A special thanks to Dr. Scott Bruce, my PhD advisor for his countless hours of reflecting, reading, encouraging, and most of all patience throughout the entire process. Thank you Dr. Pramita Bagchi, Dr. Anand Vidyashankar, Dr. Wanli Qiao, and Dr. Zeda Li for agreeing to serve on my committee.

Table of Contents

	Page
List of Tables	vii
List of Figures	ix
Abstract	xiv
1 Introduction	1
2 Methodological Background	6
2.1 Single Stationary Time Series	6
2.1.1 Bayesian Penalized Spline Estimator	8
2.2 Multiple Stationary Time Series	9
2.2.1 Covariate-dependent Multiple Stationary Time Series	11
2.2.2 Piecewise Stationary Approximation of Covariate Space	12
2.3 Multiple Nonstationary Time Series	13
2.3.1 Piecewise Stationary Approximation of Time and Covariate Space	14
2.4 Tree-Based Modeling of the Power Spectrum	15
2.4.1 Tree-Based Piecewise Partition	15
2.4.2 Local Power Spectrum Estimation	17
2.5 Introduction of BART	19
2.5.1 A sum-of-trees model	20
2.5.2 BART Priors	21
2.5.3 Posterior Distribution	23
3 Adaptive Bayesian Sum of Trees Model for Stationary Time Series Data	26
3.1 Adaptive Bayesian Sum of Trees Model	26
3.1.1 Sum of Trees Model	26
3.1.2 Prior Specification	27
3.1.3 Bayesian Backfitting MCMC	28
3.1.4 Reversible-jump MCMC Sampling	29
3.1.5 Comparison with Bayesian additive regression model	29
3.2 Simulations of Bayesian Sum of Trees Model	30
3.2.1 Settings	30

3.2.2	Results: Estimation Accuracy	32
3.2.3	Comparison with Generalized Additive Model	36
3.2.4	Results: Computation Time	39
3.2.5	Results: Sparse Covariate Effects	42
3.3	Gait Maturation Analysis	44
3.3.1	Estimation Results of Gait Maturation Analysis	44
4	Inverse Regression Framework	49
4.1	Inverse Regression Procedure	50
5	Covariate-dependent Multiple Nonstationary Time Series	53
5.1	Voronoi Tessellation Modeling Structure	53
5.2	Local Power Spectrum Estimation	55
5.3	Sampling Scheme	57
5.4	Simulation Results	58
6	Conclusion and Future Work	62
A	Proof of the Smoothing Parameter τ^2	63
B	Sampling Scheme Details for Bayesian Sum of Trees Model	66
B.1	BIRTH	69
B.2	DEATH	72
B.3	CHANGE	73
C	Sampling Scheme Details for Voronoi Tessellation	75
C.1	BIRTH	78
C.2	DEATH	79
C.3	MOVE	79
C.4	CHANGE	80
D	Additional Simulation and Application Results for Bayesian Sum of Trees Model	81
D.1	Additional Results: Estimation Accuracy	81
D.2	Additional Results: Sparse Covariate Effects	81
D.3	Additional Results: Gait Maturation Analysis	89
D.3.1	Visualization of Tree Structures	89
D.3.2	Credible Intervals for ALE of Power Spectrum	92
D.4	Convergence Diagnostics	94
E	Graphical Posterior Predictive Diagnostics	103
	Bibliography	112

List of Tables

Table	Page
3.1 Mean and standard deviation of MSE over 100 replications for three simulation settings with different lengths (T) and number (L) of time series. Results are presented for the proposed Bayesian sum of trees model (top) with number of trees $M = 50$ and the AdaptSPEC-X model (bottom) with mixture components $C = 50$	35
3.2 Mean and standard deviation of MSE over 100 replications for two simulation settings with different lengths (T) and number (L) of time series. Lower mean MSEs for each setting are bolded for ease of comparison.	38
3.3 Mean and standard deviation of MSE and posterior probability of model inclusion for AR-Friedman setting with $L = 500$ time series of length $T = 250$ over 100 replications with different number of covariates (P) and hyperpriors (Uniform and Dirichlet). Posterior probabilities for model inclusion are reported for important variables, $\omega_1, \dots, \omega_5$, individually and noise variables $\omega_6, \dots, \omega_P$ in aggregate. Results are presented for the proposed Bayesian sum of trees model using $M = 50$ trees.	43
5.1 Mean and standard deviation of MSE for time-varying abrupt-slowly simulation setting with $L = 20$ time series with different length T . Results are presented for the proposed Voronoi tessellation model starting with $M = 5$ centers.	61
D.1 Mean and standard deviation of MSE and posterior probability of model inclusion for Abrupt+Smooth setting with $L = 500$ time series of length $T = 250$ over 100 replications with different number of covariates (P) and hyperpriors (Uniform and Dirichlet). Posterior probabilities for model inclusion are reported for important variables, ω_1, ω_2 , individually and noise variables $\omega_3, \dots, \omega_P$ in aggregate. Results are presented for the proposed Bayesian sum of trees model using $M = 5$ and $M = 50$ trees.	86

D.2	Mean and standard deviation of MSE and posterior probability of model inclusion for AR-Friedman setting with $L = 500$ time series of length $T = 250$ over 100 replications with different number of covariates (P) and hyperpriors (Uniform and Dirichlet). Posterior probabilities for model inclusion are reported for important variables, $\omega_1, \dots, \omega_5$, individually and noise variables $\omega_6, \dots, \omega_P$ in aggregate. Results are presented for the proposed Bayesian sum of trees model using $M = 5$ and $M = 50$ trees	87
D.3	Mean and standard deviation of MSE and posterior probability of model inclusion for Adjusted-AdaptSPEC-X setting with $L = 500$ time series of length $T = 250$ over 100 replications with different number of covariates (P) and hyperpriors (Uniform and Dirichlet). Posterior probabilities for model inclusion are reported for important variables, ω_1, ω_2 , individually and noise variables $\omega_3, \dots, \omega_P$ in aggregate. Results are presented for the proposed Bayesian sum of trees model using $M = 5$ and $M = 50$ trees	88

List of Figures

Figure	Page
1.1 Demeaned stride interval time series for three participants in the gait maturation study ages 4, 7, and 11 years old.	2
2.1 Illustration of transforming tree structure to grid structure blocks	16
2.2 Illustration of two different partitioning methods	16
2.3 Illustration of the sum of regression trees	21
3.1 (a) presents the mapping of covariate values ω_1 and ω_2 to latent variable values z for the Adjusted-AdaptSPEC-X simulation setting. D1 and D2 denote two simulated time series with similar covariate values that are mapped to different latent variable values corresponding to different power spectra. For these two realizations, (b) displays the true log power spectra (red lines), log periodogram ordinates (gray points), and estimated log power spectra using the proposed Bayesian sum of trees model (blue lines) and the AdaptSPEC-X model (green lines).	31
3.2 Estimated and true covariate-dependent conditional log power spectrum for one run of the Abrupt+Smooth simulation setting. The first two columns contain the estimated and true covariate-dependent power spectrum conditional on $\omega_1 < 0.5$ and $\omega_1 \geq 0.5$ respectively; The last two columns display the estimated and true covariate-dependent power spectrum conditional on $\omega_2 = 0$ and $\omega_2 = 1$ respectively.	33
3.3 The distribution of mean run times in seconds for a single tree update over 100 replicates of the three simulations with $M = 5$ trees.	40
3.4 The total number of bottom nodes over iterations for four runs with different L and T for the AR-Friedman simulation setting using 5 trees.	41
3.5 Posterior mean of ALE for age (a) and gait speed (b) effects on the power spectrum and posterior mean of ALE for age (c) and gait speed (d) on LF/HF ratio (blue dotted line) with 95% pointwise credible intervals (shaded gray region).	48

4.1	Scatter plot of estimated age and the true age of 50 participants with LOOCV algorithm. The red reference line shows the line $y = x$	52
5.1	Comparison of Voronoi tessellation structure and tree-based model	54
5.2	Estimated and true covariate-dependent conditional log power spectrum for one run of the time-varying sbrupt-slowly simulation setting conditional on four different covariate spaces respectively. The 1st, 6th, 11th, and 16th observations are selected from each region separately.	59
5.3	Convergence diagnostic plots for the time-varying abrupt-slowly simulation for one replication: (a) average β across time length T and all time series L ; (b) average τ^2 across time length T and all time series L ; (c) number of centers; (d) weights \mathbf{w} for time and two covariates ω_1 and ω_2	60
D.1	The estimated (blue line) and the true (red line) log power spectrum of the eight randomly selected time series for the AR-Friedman simulation. The corresponding log periodogram ordinates are shown with gray dots.	82
D.2	The mapping of the covariates ω_1 and ω_2 to the latent variable z for the Adjusted-AdaptSPEC-X simulation setting with eight specific time series (red dots).	83
D.3	The estimated (blue line) and the true (red line) log power spectrum of the eight time series denoted in Figure D.2. The corresponding log periodogram ordinates are shown with gray dots.	84
D.4	Posterior probabilities of model inclusion under the uniform (left) and Dirichlet (right) priors for important variables (red) and noise variables (black).	85
D.5	Sum of trees visualization for 1st participant in the gait maturation study (age 3.3 years, gait speed 1.04 m/sec, male). The left column is an illustration of the sum of trees structure. Terminal nodes outlined in red correspond to the 1st participant. The right column shows the estimated component of log power spectrum for the 1st participant from each tree.	90
D.6	Sum of trees visualization for 50th participant in the gait maturation study (age 13.6 years, gait speed 1.22 m/sec, male). The left column is an illustration of the sum of trees structure. Terminal nodes outlined in red correspond to the 50th participant. The right column shows the estimated component of log power spectrum for the 50th participant from each tree.	91

D.7	The final log power spectrum estimates for the 1st (age 3.3 years, gait speed 1.04 m/sec, male) and 50th (age 13.6 years, gait speed 1.22 m/sec, male) participants summing over the components from five trees from Figures D.5 and Figures D.6 respectively. Each line corresponds to the addition of another tree component in the sum to see how the components come together to produce the final estimate (dark blue).	92
D.8	95% pointwise credible intervals of ALE for age (a,b) and gait speed (c,d) effects on the power spectrum.	93
D.9	Posterior mean (blue) and 95% pointwise credible intervals (shaded gray regions) of ALE for Gender.	94
D.10	Convergence diagnostic plots for the Abrupt+Smooth simulation for one replication: (a) average mean squared residuals across frequencies and all time series; (b) average estimated log power spectrum across frequencies and all time series; (c) total number of nodes for each of the five trees; (d) total number of bottom nodes across all five trees.	96
D.11	Convergence diagnostics for the AR-Friedman simulation for one replication. Plot (a) contains trace plots of the average mean squared residuals across all time series; Plot (b) shows the average estimated log power spectrum across frequencies and all time series for each iteration; Plot (c) is the trace plots of the total number of nodes for each of the five trees separately (c); Plot (d) is the total number of bottom nodes across all five trees.	97
D.12	Convergence diagnostics for the Adjusted-AdaptSPEC-X simulation for one replication. Plot (a) contains trace plots of the average mean squared residuals across all time series; Plot (b) shows the average estimated log power spectrum across frequencies and all time series for each iteration; Plot (c) is the trace plots of the total number of nodes for each of the five trees separately (c); Plot (d) is the total number of bottom nodes across all five trees.	98
D.13	Convergence diagnostic plots for the Abrupt+Smooth simulation for one replication for 50 trees. (a) average mean squared residuals across frequencies and all time series; (b) average estimated log power spectrum across frequencies and all time series; (c) total number of nodes for each of the five trees; (d) total number of bottom nodes across all 50 trees.	99

D.14	Convergence diagnostics for the AR-Friedman simulation for one replication for 50 trees. Plot (a) contains trace plots of the average mean squared residuals across all time series; Plot (b) shows the average estimated log power spectrum across frequencies and all time series for each iteration; Plot (c) is the trace plots of the total number of nodes for each of the five trees separately (c); Plot (d) is the total number of bottom nodes across all 50 trees.	100
D.15	Convergence diagnostics for the Adjusted-AdaptSPEC-X simulation for one replication for 50 trees. Plot (a) contains trace plots of the average mean squared residuals across all time series; Plot (b) shows the average estimated log power spectrum across frequencies and all time series for each iteration; Plot (c) is the trace plots of the total number of nodes for each of the five trees separately (c); Plot (d) is the total number of bottom nodes across all 50 trees.	101
D.16	Convergence diagnostics for the gait maturation data analysis. Plot (a) contains trace plots of the average mean squared residuals across all time series; Plot (b) shows the average estimated log power spectrum across frequencies and all time series for each iteration; Plot (c) is the trace plots of the total number of nodes for each of the five trees separately (c); Plot (d) is the total number of bottom nodes across all five trees.	102
E.1	Posterior predictive plots for participants 1-8 of gait maturation study. Left: Observed log-periodogram (black line) and posterior predictive log-periodogram draws (light blue). Right: Density plots for observed (black line) and posterior predictive log-periodogram draws (light blue) across frequencies. Age is in years and speed is in meters per second.	105
E.2	Posterior predictive plots for participants 9-16 of gait maturation study. Left: Observed log-periodogram (black line) and posterior predictive log-periodogram draws (light blue). Right: Density plots for observed (black line) and posterior predictive log-periodogram draws (light blue) across frequencies. Age is in years and speed is in meters per second.	106

E.3	Posterior predictive plots for participants 17-24 of gait maturation study. Left: Observed log-periodogram (black line) and posterior predictive log-periodogram draws (light blue). Right: Density plots for observed (black line) and posterior predictive log-periodogram draws (light blue) across frequencies. Age is in years and speed is in meters per second.	107
E.4	Posterior predictive plots for participants 25-32 of gait maturation study. Left: Observed log-periodogram (black line) and posterior predictive log-periodogram draws (light blue). Right: Density plots for observed (black line) and posterior predictive log-periodogram draws (light blue) across frequencies. Age is in years and speed is in meters per second.	108
E.5	Posterior predictive plots for participants 33-40 of gait maturation study. Left: Observed log-periodogram (black line) and posterior predictive log-periodogram draws (light blue). Right: Density plots for observed (black line) and posterior predictive log-periodogram draws (light blue) across frequencies. Age is in years and speed is in meters per second.	109
E.6	Posterior predictive plots for participants 41-48 of gait maturation study. Left: Observed log-periodogram (black line) and posterior predictive log-periodogram draws (light blue). Right: Density plots for observed (black line) and posterior predictive log-periodogram draws (light blue) across frequencies. Age is in years and speed is in meters per second.	110
E.7	Posterior predictive plots for participants 49-50 of gait maturation study. Left: Observed log-periodogram (black line) and posterior predictive log-periodogram draws (light blue). Right: Density plots for observed (black line) and posterior predictive log-periodogram draws (light blue) across frequencies. Age is in years and speed is in meters per second.	111

Abstract

ADAPTIVE BAYESIAN COVARIATE DEPENDENT SPECTRAL ANALYSIS OF MULTIPLE TIME SERIES

Yakun Wang, PhD

George Mason University, 2022

Dissertation Director: Dr.Pramita Bagchi

The frequency-domain properties of time series have been found to contain valuable information in many studies. It is often the case that biomedical time series are collected from multiple participants in conjunction with multiple covariates in order to analyze the association between the characteristics of biological processes and various clinical and behavioral outcomes. In this work, we propose flexible and adaptive nonparametric Bayesian methods to estimate the association between multiple covariates and the power spectrum of multiple time series. For stationary time series data, we introduce a Bayesian sum of trees model to capture complex dependencies and interactions between covariates and the power spectrum. Local power spectra corresponding to terminal nodes within trees are estimated nonparametrically using Bayesian penalized linear splines. The trees are considered to be random and fit using a Bayesian backfitting Markov chain Monte Carlo (MCMC) algorithm that sequentially considers tree modifications via reversible-jump MCMC techniques. For high-dimensional covariates, a sparsity-inducing Dirichlet hyperprior on tree splitting proportions is considered, which provides a sparse estimation of covariate effects and efficient variable selection. For nonstationary time series, Voronoi tessellation is used as the partition model for the partition of both time and covariates spaces. The tessellation

is adaptively updated via the reversible-jump MCMC technique. The Bayesian penalized linear splines model is used to estimate the local power spectra within each disjoint region of the tessellation. Empirical performance is evaluated via simulations to demonstrate the proposed methods' ability to accurately recover complex relationships and interactions. The Bayesian sum of trees model is used to study gait maturation in young children by evaluating age-related changes in power spectra of stride interval time series in the presence of other covariates.

Chapter 1: Introduction

The frequency-domain properties of time series have often been found to contain valuable information. For example, frequency-domain analysis of biomedical time series, such as gait variability, heart rate variability (HRV), and electroencephalography (EEG), provides interpretable information about underlying physiological processes (Hausdorff et al., 1999; Hall et al., 2004; Klimesch, 1999). In many studies, biomedical time series are collected from multiple participants in conjunction with multiple covariates to explore connections between prominent oscillatory patterns in the time series and various clinical and behavioral outcomes. These relationships are often complex and highly interactive. As a result, a flexible, adaptive method that can estimate the association between power spectra and multiple covariates is needed to better understand the complex relationships between physiological processes and important measures of health and functioning.

A prime example and motivating application for this research comes from a study of maturation in gait dynamics in young children (Hausdorff et al., 1999). Immature gait in very young children results in unsteady walking patterns and frequent falls (Shumway-Cook and Williams, 1995). While gait is relatively mature by age 3, neuromuscular control continues to develop well beyond this age (Preis et al., 2008). Accordingly, it is of interest to assess if gait dynamics continue to become more steady and regular beyond age 3, in conjunction with improving neuromuscular control. To assess gait variability and posture control in younger children, stride interval time series consisting of stride times during normal walking were observed in fifty children between the ages of 3 and 14 (Hausdorff et al., 1999). For illustration, Figure 1.1 displays demeaned stride interval time series for three participants age 4, 7, and 11 years old. In addition to age, other covariates were also collected that may influence gait, such as gait speed and gender. In quantifying the

association between age and the power spectra of stride interval time series, we seek to better understand the maturation of gait dynamics and variability associated with developing neuromuscular control with age in the presence of other related covariates.

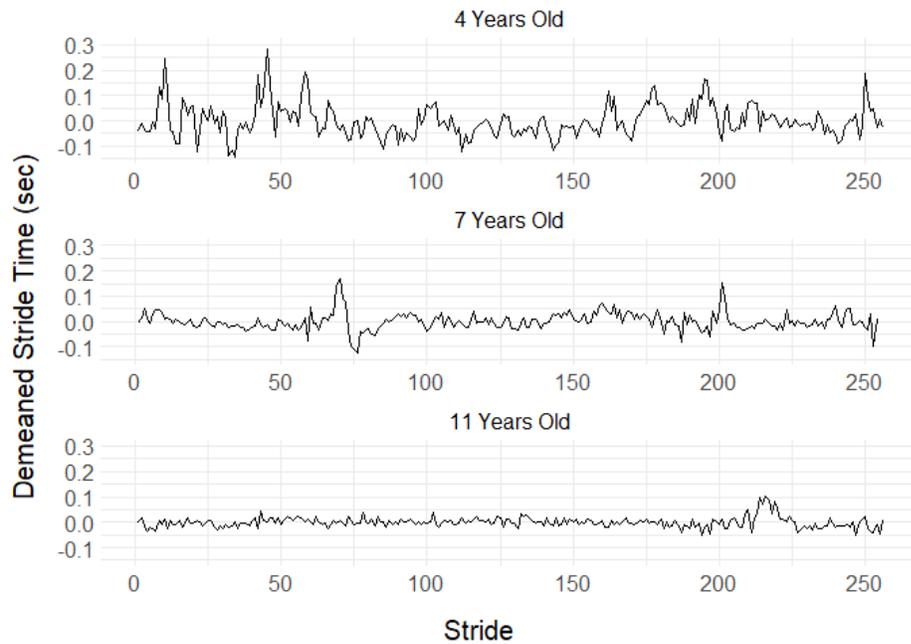


Figure 1.1: Demeaned stride interval time series for three participants in the gait maturation study ages 4, 7, and 11 years old.

In the time series literature, spectral analysis of multiple time series has received much attention in recent years. To quantify the association between a *single* covariate and power spectra, Fiecas and Ombao (2017) and Krafty et al. (2017) propose methods that can capture a smooth covariate effect on the power spectrum. Bruce et al. (2018) proposes an

adaptive Bayesian method that can capture both smooth and abrupt changes in power spectra across a covariate. Li et al. (2021) adapts the method of Bruce et al. (2018) for covariate-dependent spectral analysis of replicated multivariate time series. However, these methods are not readily extendable to incorporate *multiple* covariates, which hinders their applicability to many important studies. Existing methods that can account for multiple covariates are either parametric (Diggle and Al Wasel, 1997) or semi-parametric (Iannaccone and Coles, 2001; Qin et al., 2009; Stoffer et al., 2010; Krafty et al., 2011; Chau and von Sachs, 2016). These approaches characterize covariate effects via design matrices within a linear modeling framework, and thus can not immediately accommodate complex dependencies and interactions among covariates and power spectra. One exception is the approach proposed by Bertolacci et al. (2022), which introduces a mixture modeling approach with covariate-dependent mixture weights to account for complex covariate effects. However, a thin-plate Gaussian process prior is imposed on mixture weights, which is both smooth and stationary, and may not be appropriate for modeling abrupt spectral dynamics. Moreover, this method does not provide a means for variable selection when a large number of covariates are observed. The goal of this work is to introduce a flexible method that can capture both smooth and abrupt changes in power spectra across multiple covariates, without loss of interpretation, and simultaneously provide a tool for variable selection.

For stationary time series data, to capture complex smooth, abrupt, and interaction effects of covariates on power spectra in a parsimonious manner, we propose a tree-based covariate partitioning framework. Tree-based models are not new and have become extremely popular in recent years (Breiman, 2001; Chipman et al., 2010). For example, Chipman et al. (2010) propose a Bayesian additive regression tree (BART) model which can flexibly model complex covariate effects and interactions and demonstrates outstanding predictive performance (Chipman et al., 2013). Accordingly, BART has been widely applied in many different scientific domains for various types of outcomes (Waldmann, 2016; Blattenberger and Fowles, 2017; van der Merwe, 2018) including smooth functional response variables (Starling et al., 2020). In this work, a sum of trees model for the covariate-dependent power spectrum

is introduced to simultaneously partition multiple covariates parsimoniously. A penalized linear spline model is used for local spectrum estimation within terminal nodes of the trees. The framework is formulated in a fully Bayesian setting where the trees are random and fit using an iterative Bayesian backfitting Markov chain Monte Carlo (MCMC) procedure and reversible-jump techniques (Green, 1995) to evaluate various tree modifications.

For nonstationary time series data, a Voronoi tessellation structure (Green and Sibson, 1978; Møller, 1994; Boots et al., 2000) is used to capture both abrupt and smooth change on time and covariates spaces. Similar with the tree-based model, this partition model is a nonparametric method that can recover the relationship between response and covariates. While, instead of a rectangular block of the tree-based model, the Voronoi tessellation structure is more flexible that divides the covariate space into any shape of disjoint regions such that the model is more computational efficient. Hol (2005) and Denison et al. (2002) propose a Bayesian partition model with Voronoi tessellation structure for prediction on continuous covariates for regression and classification. Payne et al. (2020) uses this partition to model conditional density estimation. In the proposed framework, we apply the Bayesian penalized linear spline model within each region for power spectrum estimation.

The proposed methodology expands the scope of covariate dependent power spectra that can be accurately recovered in three meaningful ways. First, the flexible partition modeling framework can recover complex nonlinear relationships and interaction effects without assuming a particular form of the relationship a priori. Second, by averaging over the posterior distributions, the proposed method can recover both smooth and abrupt covariate effects on the power spectrum. While methods that assume completely smooth or completely abrupt changes will likely perform better when these assumptions are valid, the proposed method generally outperforms these methods in the presence of both smooth and abrupt covariate effects without prior knowledge of the nature of such effects. Third, for the sum of trees model, the proposed method can automatically accommodate mixed-type covariates (nominal, ordinal, discrete, continuous) through the underlying tree structures, as well as high-dimensional covariates by placing a sparsity-inducing Dirichlet hyperprior

on the splitting proportions of the regression tree prior (Linero, 2018) for sparse estimation of covariate effects and variable selection.

The rest of the dissertation is organized as follows. Chapter 2 provides the methodological background for the covariate-dependent power spectrum, and the tree-based modeling framework for the stationary time series data. Chapter 3 presents the proposed adaptive Bayesian sum of trees model, the simulation results for various covariate effects (e.g. linear, nonlinear, smooth, abrupt, high-dimensional) and interaction effects, and the application results of the Gait maturation study. Chapter 4 presents the methodology of the inverse regression framework to estimate unknown covariates for interpretable regression and classification. Chapter 5 introduces the Voronoi tessellation modeling structure for covariate-dependent power spectrum of nonstationary time series data and the simulation results. Conclusions and future directions of this work are covered in Chapter 6.

Chapter 2: Methodological Background

To better understand the proposed method for covariate-dependent power spectrum analysis, we first provide some preliminaries for power spectrum analysis of stationary and nonstationary time series data, and the background knowledge of the tree-based model.

2.1 Single Stationary Time Series

The power spectrum represents a decomposition of variance over frequencies (Wei, 2006) and characterizes the contribution of each frequency to the time series. It is the primary tool for frequency-domain time series analysis.

Let $\{X_t\}$ be a weakly stationary zero-mean time series, where $t = 1, 2, \dots$ such that $\mu_t = E(x_t) = \int_{-\infty}^{\infty} x f_t(x) dx$ is constant and does not depend on time t and $\gamma(s, t) = (x_s, x_t) = E[(x_s - \mu_s)(x_t - \mu_t)]$ depends on s and t only through their difference $|s - t|$ (Shumway and Stoffer, 2017, p. 20). The frequency domain characteristics can be obtained from its spectral representation (Cramér, 1942)

$$X_t = \int_{-\frac{1}{2}}^{\frac{1}{2}} A(\nu) \exp(2\pi i t \nu) dZ(\nu), \quad (2.1)$$

where $A(\nu)$ is a complex-valued function that is Hermitian $A(\nu) = \overline{A(-\nu)}$, and $A(\nu) = A(\nu + 2\pi)$. It is square-integrable over $[-1/2, 1/2]$, and has period 1 as a function of frequency. $Z(\nu)$ is a zero-mean orthogonal process that is independent of $A(\nu)$ such that

$$[dZ(\nu), dZ(\nu')] = \begin{cases} 0, & \text{if } \nu \neq \nu'. \\ d\nu, & \text{otherwise.} \end{cases}$$

The power spectral density function $f(\nu)$ is defined as

$$f(\nu) = |A(\nu)|^2 = \sum_{h=-\infty}^{\infty} \gamma(h) \exp(-2\pi i\nu h) \quad -\frac{1}{2} \leq \nu \leq \frac{1}{2}, \quad (2.2)$$

where $\gamma(h) = (X_{t+h}, X_t)$ is the autocovariance function of $\{X_t\}$. Note that we can obtain $\gamma(h)$ by applying the inverse Fourier transform from $f(\nu)$, which can be specified as

$$\gamma(h) = \int_{-\frac{1}{2}}^{\frac{1}{2}} f(\nu) \exp(2\pi i\nu h) d\nu. \quad (2.3)$$

For the estimation of the power spectrum, the periodogram is a standard estimator that has been well studied. Given a realization x_1, x_2, \dots, x_T of a stationary time series $\{X_t\}$, assume T is even without loss of generality. The periodogram at frequency ν is then defined as

$$I(\nu_k) = \frac{1}{T} \left| \sum_{t=1}^T x_t \exp(-2\pi i\nu_k t) \right|^2, \quad (2.4)$$

where the frequencies $\nu_k = k/T$ for $k = 1, \dots, [T/2]-1$ are known as the Fourier frequencies.

It can be shown that the periodogram is an asymptotically unbiased estimator of the power spectrum. However, the variances of periodogram ordinates across frequencies do not converge to 0 as $T \rightarrow \infty$. In order to resolve this dilemma, there are some parametric and non-parametric methods proposed to obtain a consistent estimation of the power spectrum. A parametric approach would be to model the time series using an ARMA(p, q) process, which has a closed-form expression for the power spectrum as a function of the coefficients in this model. The choice of the orders p and q can be determined using standard model selection criteria (e.g. AIC) (Wei, 2006, p. 318). Smoothed estimators have also been proposed which nonparametrically smooth periodogram estimates using a window technique (Shumway and Stoffer, 2017, p. 191).

Likelihood-based power spectrum estimators have also been developed based on the large-sample distribution of the periodogram (Whittle, 1957). Let $n = \lfloor T/2 \rfloor - 1$ and $\nu_k = k/T$ for $k = 1, \dots, n$ be the Fourier frequencies. Assume T is large, the likelihood of $x = (x_1, \dots, x_T)'$ can be approximated by

$$p(x|f) \approx (2\pi)^{-\frac{n}{2}} \prod_{k=1}^n \exp\{-[\log f(\nu_k) + I(\nu_k)/f(\nu_k)]\}. \quad (2.5)$$

This implies that the distribution of the periodogram ordinates can be approximated by a log linear model of the form

$$\log I(\nu_k) = \log f(\nu_k) + \epsilon_k, \quad (2.6)$$

where ϵ_k for $k = 1, \dots, \lfloor T/2 \rfloor - 1$ are asymptotically independent and identically distributed as $\log(\chi_2^2/2)$, and $\epsilon_k \stackrel{i.i.d.}{\sim} \log(\chi_1^2)$ for $k = 0, T/2$.

2.1.1 Bayesian Penalized Spline Estimator

Many estimators of the power spectrum have been developed based on this model (Wahba, 1980; Carter and Kohn, 1997). More recently, Rosen et al. (2012) propose a linear smoothing spline model to estimate the log spectral density within a Bayesian framework. The log spectrum estimator is displayed as

$$\log f(\nu) \approx \alpha + \sum_{s=1}^S \beta_s \cos(2\pi s\nu). \quad (2.7)$$

where the functions $\cos(2\pi s\nu)$ are the Demmler–Reinsch basis functions for periodic even splines observed on an evenly spaced grid (i.e. the Fourier frequencies) (Schwarz and Krivobokova, 2016, Section 3). Only the first $S < N$ basis functions are used to provide a low-rank approximation to the full linear smoothing spline (Eubank, 1999). Let \mathbf{Z}

be a $n \times S$ matrix of the basis functions where $Z_{is} = \cos(2\pi s\nu_i)$. Prior distributions of the vector of coefficients $\boldsymbol{\beta}$ is $\boldsymbol{\beta} = (\beta_1, \dots, \beta_S)' \sim N(0, \tau^2 \mathbf{D}_S)$ where $\mathbf{D}_S = \text{diag}(\{\sqrt{2\pi s}\}^{-2})$ and τ is a smoothing parameter follows uniform prior $p(\tau^2) = 1/\tau^2$. \mathbf{D}_S is formulated such that the distribution of the integrated squared first derivative of the log power spectrum is regulated by the smoothing parameter τ (see Appendix A for more details). The prior distribution of α is $N(0, \sigma_\alpha^2)$. The hyperparameter σ_α^2 is set to be a fixed, large constant. The coefficients α and $\boldsymbol{\beta}$ are also independent a priori. The posterior distribution of parameters α , $\boldsymbol{\beta}$, and τ^2 are then drawn from a two-step sampling scheme as follows

1. Given a realization of the log periodogram, $\log \mathbf{I} = \{\log[I(\nu_1)], \dots, \log[I(\nu_n)]\}'$ and basis functions \mathbf{Z} , α and $\boldsymbol{\beta}$ are sampled jointly via a Metropolis-Hastings (M-H) step from

$$p(\alpha, \boldsymbol{\beta} | \tau^2, \log \mathbf{I}, \mathbf{Z}) \propto \exp \left\{ - \sum_{k=1}^n [\alpha + z'_k \boldsymbol{\beta} + \exp(\log[I(\nu_k)]) - \alpha - z'_k \boldsymbol{\beta}] - \frac{\alpha}{2\sigma_\alpha^2} - \frac{1}{2\tau^2} \boldsymbol{\beta}' \mathbf{D}_S^{-1} \boldsymbol{\beta} \right\}, \quad (2.8)$$

where z'_k is the k th row of \mathbf{Z} .

2. τ^2 is sampled from the inverse gamma distribution with density

$$p(\tau^2 | \boldsymbol{\beta}) \propto (\tau^2)^{-\frac{S}{2}} \exp\left(-\frac{1}{2\tau^2} \boldsymbol{\beta}' \mathbf{D}_S^{-1} \boldsymbol{\beta}\right). \quad (2.9)$$

2.2 Multiple Stationary Time Series

Many studies and methods focus on power spectrum analysis of multiple time series. Assuming the multiple time series are independent realizations of the same underlying process, the Bayesian smoothing spline model covered in the previous section can be easily extended to multiple time series.

Let $X_{\ell t}$ be a collection of stationary time series with length $t = 1, \dots, T$ from $\ell = 1, \dots, L$ independent subjects. The frequency domain characteristics can be obtained from its spectral representation (Cramér, 1942)

$$X_{\ell t} = \int_{-\frac{1}{2}}^{\frac{1}{2}} A(\nu) \exp(2\pi i t \nu) dZ_{\ell}(\nu). \quad (2.10)$$

where Z_{ℓ} and the $A(\nu)$ are defined the same as in Section 2.1. The periodogram at frequency ν is then defined as

$$I(\nu_k) = \frac{1}{T} \left| \sum_{t=1}^T x_{\ell t} \exp(-2\pi i \nu_k t) \right|^2, \quad (2.11)$$

where the frequencies $\nu_k = k/T$ for $k = 1, \dots, \lfloor T/2 \rfloor - 1$ are known as the Fourier frequencies. The overall Whittle likelihood can then be approximated by the product of individual Whittle likelihood

$$p(\mathbf{x}|\mathbf{f}) \approx \prod_{\ell=1}^L (2\pi)^{-\frac{n}{2}} \prod_{k=1}^n \exp\{-[\log f(\nu_k) + I(\nu_k)/f(\nu_k)]\}, \quad (2.12)$$

and the sampling scheme for the model coefficients α , β , and τ^2 of Bayesian penalized spline estimator can be extended as follows:

1. Given a set of realizations of the log periodogram, $\log \mathbf{I}_{\ell} = \{\log[I_{\ell}(\nu_1)], \dots, \log[I_{\ell}(\nu_n)]\}'$ and basis functions \mathbf{Z} , α and β are sampled jointly via a Metropolis-Hastings (M-H) step from

$$p(\alpha, \beta | \tau^2, \log \mathbf{I}, \mathbf{Z}) \propto \exp \left\{ - \sum_{\ell=1}^L \sum_{k=1}^n [\alpha + z'_{k\ell} \beta + \exp(\log[I_{\ell}(\nu_k)])] - \alpha - z'_{k\ell} \beta \right. \\ \left. - \frac{\alpha}{2\sigma_a^2} - \frac{1}{2\tau^2} \beta' \mathbf{D}_S^{-1} \beta \right\}, \quad (2.13)$$

where $n = \lfloor T/2 \rfloor - 1$ and z'_k is the k th row of \mathbf{Z} .

2. τ^2 is sampled from the inverse gamma distribution with density

$$p(\tau^2|\boldsymbol{\beta}) \propto (\tau^2)^{-\frac{S}{2}} \exp\left(-\frac{1}{2\tau^2}\boldsymbol{\beta}'\mathbf{D}_S^{-1}\boldsymbol{\beta}\right). \quad (2.14)$$

However, assuming a common stationary underlying process for multiple time series may not be realistic in practice. Diggle and Al Wasel (1997) note this limitation and introduce random effects models to account for the between-subject variability of the periodogram for multiple time series. Cadonna et al. (2019) construct a Bayesian modeling approach using a mixture of Gaussian distributions to estimate the log-periodogram distribution for a single time series and propose a hierarchical model for multiple time series. These methods can identify groups of time series with similar spectral characteristics and provide an appropriate estimation of the power spectrum in the presence of such clustering effects.

2.2.1 Covariate-dependent Multiple Stationary Time Series

When additional covariates that may be associated with the power spectrum are available, we can directly incorporate covariates under the assumption that the covariates modulate the dependence across multiple time series, i.e. time series with similar covariate values have similar underlying power spectra.

More specifically, we consider modeling a collection of stationary time series $X_{\ell t}$ of length $t = 1, \dots, T$ and p -dimensional covariates $\boldsymbol{\omega}_\ell = \{\omega_{1\ell}, \dots, \omega_{p\ell}\}$ for $\ell = 1, \dots, L$ independent subjects. Consider the following covariate-dependent extension of the Cramér representation

$$X_{\ell t} = \int_{-1/2}^{1/2} A(\boldsymbol{\omega}_\ell, \nu) \exp(2\pi i t \nu) dZ_\ell(\nu). \quad (2.15)$$

where $A(\boldsymbol{\omega}_\ell, \nu)$ is a time-varying complex-valued function of covariate $\boldsymbol{\omega}_\ell$, and frequency $\nu \in \mathbb{R}$ that is Hermitian $A(\boldsymbol{\omega}_\ell, \nu) = \overline{A(\boldsymbol{\omega}_\ell, -\nu)}$, and $A(\boldsymbol{\omega}_\ell, \nu) = A(\boldsymbol{\omega}_\ell, \nu + 2\pi)$. It is square

integrable with respect to frequencies over $[-1/2, 1/2]$, and has period 1 as a function of frequency. \mathbf{Z}_ℓ are zero-mean mutually independent orthogonal processes that are also independent of $A(\boldsymbol{\omega}_\ell, \nu)$. Hence, the covariate-dependent power spectrum is defined as

$$f(\boldsymbol{\omega}, \nu) = |A(\boldsymbol{\omega}, \nu)|^2. \quad (2.16)$$

The power spectrum $f(\boldsymbol{\omega}, \nu)$ can be interpreted as the contribution to the variance at frequency ν , conditional on covariate values $\boldsymbol{\omega}$. We assume that A , and subsequently the spectrum f , are continuous functions of frequency ν , but can have a finite number of discontinuities as functions of covariates in $\boldsymbol{\omega}$. This flexibility allows for modeling abrupt changes over the covariate space. For the power spectrum estimation with the Bayesian penalized spline model, we employ the piecewise stationary approximation method to create a locally stationary partition corresponding to the covariate space which is introduced in the following Section.

2.2.2 Piecewise Stationary Approximation of Covariate Space

A locally stationary time series for covariate space can be approximated by a piecewise stationary process

$$X_{\ell t} \approx \sum_{j=1}^{j=M} X_{\ell t}^{(j)} \delta_j(\boldsymbol{\omega}_\ell), \quad (2.17)$$

where suppose we have M disjoint blocks, $\delta_j(\boldsymbol{\omega}_\ell)$ is a indicator function that $\delta_j(\boldsymbol{\omega}_\ell) = 1$ if the $\boldsymbol{\omega}_\ell \in (\xi_{j-1}, \xi_j]$ and is zero otherwise. $\boldsymbol{\xi} = (\xi_0, \dots, \xi_M)$ form a partition of $[0, 1]$ into approximately stationary segments, and $X_{\ell t}^{(j)}$ are stationary process (Adak, 1998a). Similarly, let $n = \lfloor T/2 \rfloor - 1$ and $\nu_k = k/T$ for $k = 1, \dots, n$ be the Fourier frequencies, and let $I_j(\nu)$ be the local periodogram within the j th segment. The likelihood can then be

approximated by a product of local Whittle likelihood

$$L(f_1, \dots, f_M | \mathbf{x}, \boldsymbol{\xi}) \approx \prod_{\ell=1}^L \prod_{j=1}^M (2\pi)^{-\frac{n}{2}} \prod_{k=1}^n \exp\{-\delta_j[\log f_j(\nu_k) + I_j(\nu_k)/f_j(\nu_k)]\}. \quad (2.18)$$

where the periodogram $I_j(\nu)$ is the same as Equation 2.11.

2.3 Multiple Nonstationary Time Series

A model for the spectrum analysis of nonstationary time series can be defined by allowing the transfer function in the Cramér spectral representation to vary with time (Priestley, 1965). We consider modeling a collection of nonstationary time series $X_{\ell t}$ of length $t = 1, \dots, T$ and p -dimensional covariates $\boldsymbol{\omega}_\ell = \{\omega_{1\ell}, \dots, \omega_{p\ell}\}$ for $\ell = 1, \dots, L$ independent subjects. The covariate-dependent extension of the Cramér representation follows as

$$X_{\ell t} = \int_{-1/2}^{1/2} A(t/T, \boldsymbol{\omega}_\ell, \nu) \exp(2\pi i t \nu) dZ_\ell(\nu). \quad (2.19)$$

where $A(u, \boldsymbol{\omega}_\ell, \nu)$ is a time-varying complex-valued function of covariate $\boldsymbol{\omega}_\ell$, scaled time $u \in [0, 1]$, and frequency $\nu \in \mathbb{R}$ that is Hermitian $A(u, \boldsymbol{\omega}_\ell, \nu) = \overline{A(u, \boldsymbol{\omega}_\ell, -\nu)}$, and $A(u, \boldsymbol{\omega}_\ell, \nu) = A(u, \boldsymbol{\omega}_\ell, \nu + 2\pi)$. It is square integrable with respect to frequencies over $[-1/2, 1/2]$, and has period 1 as a function of frequency. Z_ℓ are zero-mean mutually independent orthogonal processes that are also independent of $A(u, \boldsymbol{\omega}_\ell, \nu)$. Hence, the time-varying covariate-dependent power spectrum is defined as

$$f(u, \boldsymbol{\omega}, \nu) = |A(u, \boldsymbol{\omega}, \nu)|^2. \quad (2.20)$$

The power spectrum $f(u, \boldsymbol{\omega}, \nu)$ can be interpreted as the contribution to the variance at frequency ν , conditional on covariate values $\boldsymbol{\omega}$. We assume that A , and subsequently the spectrum f , are continuous functions of frequency ν , but can have a finite number of

discontinuities as functions of covariates $\boldsymbol{\omega}$ and scaled time u . This flexibility allows for modeling abrupt changes over the covariate and time space.

2.3.1 Piecewise Stationary Approximation of Time and Covariate Space

For nonstationary time series data, a piecewise stationary process can be expressed as

$$X_{\ell t} \approx \sum_{j=1}^{j=M} X_{\ell t}^{(j)} \delta_j(t, \boldsymbol{\omega}_\ell), \quad (2.21)$$

where $\delta_j(t, \boldsymbol{\omega}_\ell)$ is a indicator function that $\delta_j(t, \boldsymbol{\omega}_\ell) = 1$ if both time t and the covariates $\boldsymbol{\omega}_\ell$ falls into the j th block, and $X_{\ell t}^{(j)}$ are stationary process (Adak, 1998a). Let T_j be the number of observations in the j th segment. Also let $n_j = \lfloor T_j/2 \rfloor - 1$ and $\nu_{kj} = k/T_j$ for $k = 1, \dots, n_j$ be the Fourier frequencies for thr j th segment, and let $I_j(\nu)$ be the local periodogram within the j th segment. The likelihood can then be approximated by a product of local Whittle likelihood

$$L(f_1, \dots, f_M | \boldsymbol{x}, \boldsymbol{\xi}) \approx \prod_{\ell=1}^L \prod_{j=1}^M (2\pi)^{-\frac{n_j}{2}} \prod_{k=1}^{n_j} \exp\{-\delta_j[\log f_j(\nu_{kj}) + I_j(\nu_{kj})/f_j(\nu_{kj})]\}. \quad (2.22)$$

where the periodogram $I_j(\nu_{kj})$ at frequency ν_{kj} is defined as

$$I_{j,\ell}(\nu_k) = \frac{1}{T_j} \left| \sum_{t=1}^{T_j} \delta_j(t) x_{\ell t} \exp(-2\pi i \nu_k t) \right|^2, \quad (2.23)$$

where the frequencies $\nu_k = k/T$ for $k = 1, \dots, \lfloor T/2 \rfloor - 1$.

2.4 Tree-Based Modeling of the Power Spectrum

2.4.1 Tree-Based Piecewise Partition

Our method begins by assuming time series with similar covariate values have similar underlying power spectra and can be partitioned by a tree structure accordingly. It should be noted that methods have been proposed to approximate nonstationary time series through piecewise stationary time series, where time series are divided into approximately stationary intervals for time-dependent spectral analysis (Adak, 1998b; Rosen et al., 2012). Our model is different from these approaches as it partitions the covariate space instead of time. More recently, Bruce et al. (2018) introduced a time- and covariate-based piecewise stationary approximation for time- and covariate-dependent spectral analysis using a two-dimensional grid. However, directly extending a two-dimensional grid to higher dimensions to accommodate multiple covariates can easily lead to over-parameterization and undue computational complexity. In particular, a grid-based partition tends to produce a finer partition than is necessary, since partition points for each covariate do not depend on the other covariate. This results in less efficient information sharing across series with similar covariates and less accurate estimation. Tree-based approaches represent a more flexible and parsimonious alternative for partitioning multiple covariates. If the true partition does have a grid structure, a tree-based model can still well-approximate the partition and is thus preferable. Figure 2.1 illustrates how tree structures correspond to a partition of multiple covariates. In this illustration, if we assume that this partition represents the underlying truth of the dependence structure, Figure 2.2 shows the best possible fit using a grid-based partition (a) vs. the true tree structure (b). The grid partition requires six blocks to approximate the true underlying structure. While, if the underlying grid structure in Figure 2.2(b) is true, the tree-based model can still well approximate this structure.

For the tree-based partition, each terminal node is defined through a collection of splitting rules corresponding to the tree structure and represents an region of the covariate space

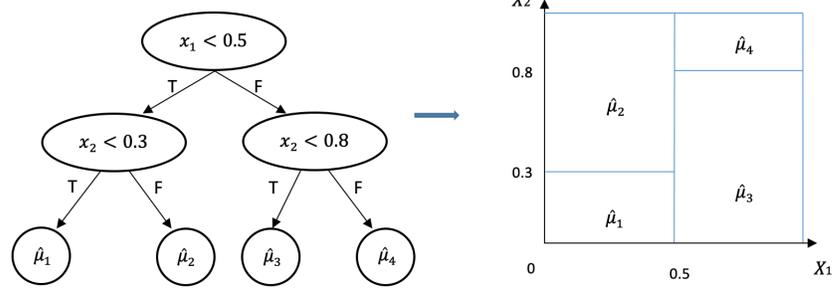


Figure 2.1: Illustration of transforming tree structure to grid structure blocks

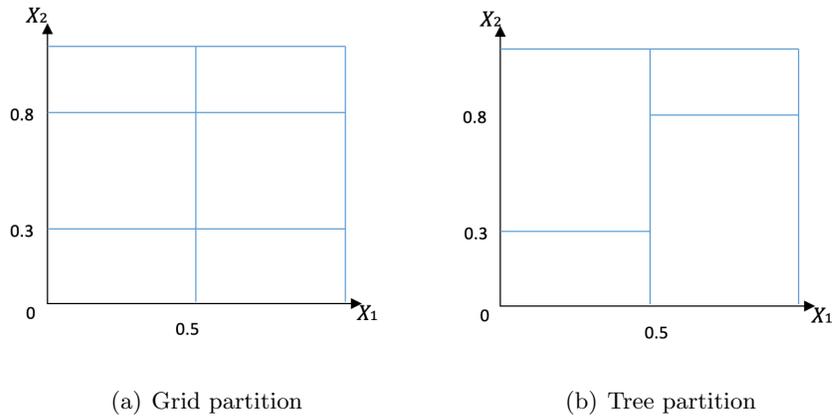


Figure 2.2: Illustration of two different partitioning methods

that shares the power spectrum. Given a tree U with B terminal nodes, the tree-based partition of the covariate-dependent Cramér spectral representation is given by

$$X_{\ell t} \approx \int_{-1/2}^{1/2} \sum_{b=1}^B \delta(\boldsymbol{\omega}_\ell; U, b) A_b(\nu) \exp(2\pi i t \nu) dZ_\ell(\nu), \quad (2.24)$$

where $X_t^{(b)}$ is stationary process with power spectrum $f_b(\nu) = |A_b(\nu)|^2$ for the b th terminal node, and δ is a function that identifies terminal node membership for each observation based on covariates such that $\delta(\boldsymbol{\omega}_\ell; U, b) = 1$ if the ℓ th observation falls into the b th terminal node and $\delta(\boldsymbol{\omega}_\ell; U, b) = 0$ otherwise.

2.4.2 Local Power Spectrum Estimation

We now introduce an estimator for local power spectra within terminal nodes of the tree. Let $N = \lfloor T/2 \rfloor - 1$ and $\nu_k = k/T$ for $k = 1, \dots, N$ be the Fourier frequencies. The periodogram estimator of the power spectrum for the ℓ th time series is $I_\ell(\nu_k) = \frac{1}{T} \left| \sum_{t=1}^T X_{\ell t} \exp(-2\pi i \nu_k t) \right|^2$. The Whittle likelihood (Whittle, 1952), derived from the large sample distribution of the periodogram, can then be used to approximate the overall likelihood for the tree as a product of individual likelihoods, assuming T is sufficiently large,

$$L(\mathbf{I}_1, \dots, \mathbf{I}_L | f_1, \dots, f_B) \approx \prod_{\ell=1}^L \prod_{b=1}^B (2\pi)^{-N/2} \prod_{k=1}^N \exp\{-\delta(\boldsymbol{\omega}_\ell; U, b) [\log f_b(\nu_k) + \exp(\log I_\ell(\nu_k) - \log f_b(\nu_k))]\} \quad (2.25)$$

for $\mathbf{I}_\ell = \{I_\ell(\nu_1), I_\ell(\nu_2), \dots, I_\ell(\nu_N)\}$. Log power spectra within each terminal node $\log f_b(\nu)$ are modeled using a Bayesian penalized linear spline model (Rosen et al., 2012)

$$\log f_b(\nu) \approx \alpha_b + \sum_{s=1}^S \beta_s^{(b)} \cos(2\pi s\nu), \quad (2.26)$$

where the functions $\cos(2\pi s\nu)$ are the Demmler–Reinsch basis functions. In order to achieve good computational efficiency without sacrificing estimation accuracy (Krafty et al., 2017), $S = 7$ basis functions are used for subsequent simulations and real data analyses, which provide good empirical performance. Gaussian priors are assumed such that $\alpha_b \sim N(0, \sigma_\alpha^2)$ where σ_α^2 is a constant value, and $\boldsymbol{\beta}^{(b)} = (\beta_1^{(b)}, \dots, \beta_S^{(b)})' \sim N(0, \tau_b^2 \mathbf{D}_S)$, where $\mathbf{D}_S = \text{diag}(\{\sqrt{2\pi}s\}^{-2})$. τ_b^2 is a smoothing parameter that controls the roughness of the log spectrum. The scaling for the smoothing parameter, $\{\sqrt{2\pi}s\}^{-2}$, provides regularization of the integrated squared first derivative of the log power spectrum (Li and Krafty, 2019). Instead of uniform prior used in Rosen et al. (2012), A half-t prior is placed on τ_b (Gelman, 2006) such that $p(\tau_b) \propto [1 + (\tau_b/A_\tau)^2/\xi_\tau]^{-(\xi_\tau+1)/2}$ for $\tau_b > 0$ where A_τ and ξ_τ are scale and degrees of freedom parameters respectively to complete the Bayesian model specification. A two-step MCMC sampling scheme for α_b , $\boldsymbol{\beta}^{(b)}$, and τ_b following Rosen et al. (2012) is presented below.

1. Let \mathbf{Z}_b be a $N \times S$ matrix of basis functions for b th terminal node such that $\{Z_b\}_{k,s} = \cos(2\pi s\nu_k)$. Given τ_b , basis functions \mathbf{Z}_b , and periodogram ordinates \mathbf{I}_ℓ for $\ell = 1, \dots, L$, α_b and $\boldsymbol{\beta}^{(b)}$ are sampled jointly in a Metropolis-Hastings (M-H) step from

$$p(\alpha_b, \boldsymbol{\beta}^{(b)} | \tau_b^2, \mathbf{I}_\ell, \mathbf{Z}_b) \propto \exp \left\{ - \sum_{\ell=1}^L \sum_{k=1}^N [\alpha_b + \mathbf{z}'_{bk} \boldsymbol{\beta}^{(b)} + \exp(\log I_\ell(\nu_k) - \alpha_b - \mathbf{z}'_{bk} \boldsymbol{\beta}^{(b)})] - \frac{\alpha_b^2}{2\sigma_a^2} - \frac{1}{2\tau_b^2} \boldsymbol{\beta}^{(b)'} \mathbf{D}_S^{-1} \boldsymbol{\beta}^{(b)} \right\}. \quad (2.27)$$

2. By representing the half-t prior as a scale mixture of inverse gamma distributions (Wand et al., 2011), we efficiently obtain draws of τ_b from its full conditional distribution by sampling from

$$(a_b | \tau_b^2) \sim \text{IG}\left(\frac{\xi_\tau + 1}{2}, \frac{\xi_\tau}{\tau_b^2} + \frac{1}{A_\tau^2}\right), \quad (2.28)$$

and

$$(\tau_b^2 | a_b, \boldsymbol{\beta}^{(b)}) \sim \text{IG}\left(\frac{\xi_\tau + S + 1}{2}, \frac{\boldsymbol{\beta}^{(b)'} \boldsymbol{\beta}^{(b)}}{2} + \frac{\xi_\tau}{a_b}\right) \quad (2.29)$$

where a_b is a latent variable, and ξ_τ and A_τ^2 are fixed hyperparameters of the inverse gamma distribution.

2.5 Introduction of BART

Poor mixing of single tree models has been noted in many applications such that the MCMC algorithm becomes stuck in subsets of the covariate space representing local optima and cannot efficiently traverse the entire parameter space (Wu et al., 2007). This can happen when single tree models grow very large in an effort to approximate more complex relationships, thus restricting possible modifications due to low sample size and an abundance of other splits. Chipman et al. (2010) proposed the Bayesian additive regression model which can solve the poor mixing problem by constructing many shallow trees as “weak learners” for estimation. The Bayesian additive regression tree model is a non-parametric regression tree model with a Bayesian MCMC sampling scheme that can recover complex dependencies between a set of predictor variables and a response variable. The basic idea of BART model is to build a sum-of-trees model and impose a prior such that the effect of any individual tree is regularized to be small. It shows an excellent ability to model non-linearity and complex interaction effects (Chipman et al., 2010).

2.5.1 A sum-of-trees model

First, we consider to model an unknown function f with output Y and a p dimensional inputs $\mathbf{x} = (x_1, \dots, x_p)$, which can be expressed as

$$Y = f(\mathbf{x}) + \epsilon, \quad \epsilon \sim N(0, \sigma^2). \quad (2.30)$$

where Y is a scalar response and the function $f(\mathbf{x}) = E(Y|\mathbf{x})$ is the mean of Y given \mathbf{x} . In order to approximate $f(\mathbf{x})$, a sum of M regression trees is used such that $f(\mathbf{x}) \approx \sum_{j=1}^M g_j(\mathbf{x})$ where g_j denotes a regression tree for $j = 1, \dots, M$. Then, the sum-of-trees model can be approximated as

$$Y = \sum_{j=1}^M g_j(\mathbf{x}) + \epsilon, \quad \epsilon \sim N(0, \sigma^2). \quad (2.31)$$

For each g_j , let U_j denote the tree structure which can be defined by the depth of the tree, the collection of internal nodes and terminal nodes, and split rules for all internal nodes. The set of parameters for the estimated value at the terminal nodes is denoted by $\Phi_j = \{\mu_{j1}, \dots, \mu_{jb_j}\}$, where b_j is the number of terminal nodes of j th tree, μ_{jb} for $b = 1, \dots, b_j$ is the parameter to be estimated in b th terminal node. Each node of the tree is grown with the binary splits on a single predictor in the predictor space. Note that for a given observation, it can only be associated with a single terminal node within the j th tree. Then, the observation's predicted value is the sum of the M terminal nodes' values from the M trees individually. We illustrate the process with an example. Suppose there are two predictors in the data to be split X_1 and X_2 , and the tree model is described in Figure 2.3. For the ℓ th observation with $x_{\ell 1} = 103$ and $x_{\ell 2} = 54$, the prediction is $\hat{\mu}_{12} + \hat{\mu}_{21} + \dots + \hat{\mu}_{M2}$.

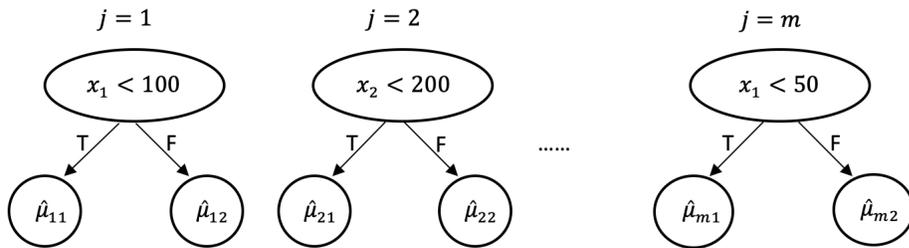


Figure 2.3: Illustration of the sum of regression trees

Unlike other tree-based models, BART is formulated within a Bayesian framework, which can then provide statistical inference on the estimated responses and covariate relationships. By imposing priors on the tree depth, split rules, and node parameters, BART can penalize overly complex tree structures for any single tree and can incorporate the previous experiments' information into the regression model through the priors.

2.5.2 BART Priors

Chipman et al. (2010) introduce three types of priors: priors for the tree structure, priors for the set of terminal parameters' μ and the variance σ^2 of the error term.

For the tree structure, the priors include (1) the probability of a node to be split. We have

$$\Pr(\text{SPLIT}) = \gamma(1 + d)^{-\theta}, \quad \gamma \in (0, 1), \theta \in [0, \infty), \quad (2.32)$$

where $d = 0, 1, \dots$ is the depth of the given node. The prior is a regularization of the tree depth to keep the tree structure to be small. For the hyperparameters in $\Pr(\text{SPLIT})$, Chipman et al. (2010) suggest $\gamma = 0.95$ and $\theta = 2$ which are the most widely used settings (Kapelner and Bleich, 2016; Linero and Yang, 2018a). (2) The probability for selecting a

covariate for splitting. A uniform prior is used such that all covariates have the same prior probability to be selected for splitting; (3) The probability for the possible cutpoints to be selected for a given covariate. A uniform prior is used such that it is invariant for monotone transformations of the corresponding covariate.

For the terminal node parameters, a conjugate normal distribution $N(\mu_\mu, \sigma_\mu^2)$ is proposed for the prior of $\mu_{jb}|U_j$, where μ_{jb} is the parameter in the b th terminal node of j th tree for $j = 1, \dots, M$ and $b = 1, 2, \dots, b_j$. A data-driven method is used to specify the hyperparameters μ_μ and σ_μ^2 . Assume the μ_{jb} 's are independent and identically distributed, for a given observation, the induced prior on the sum of trees $\sum_{j=1}^M g_j$ is $N(M\mu_\mu, M\sigma_\mu^2)$. The values of μ_μ and σ_μ can then be specified in a data-driven fashion by the following two relationships: $M\mu_\mu - k\sqrt{M}\sigma_\mu = y_{\min}$ and $M\mu_\mu + k\sqrt{M}\sigma_\mu = y_{\max}$ where y_{\min} and y_{\max} are the minimum and maximum values of the observed Y values for all observations in the data. $k = 2$ is the common choice to induce a 95% prior probability that $E(Y|x)$ is in the interval (y_{\min}, y_{\max}) . Chipman et al. (2010) suggests to center the prior of μ_{jb} at $\mu_\mu = 0$ and rescale Y so that $y_{\min} = -0.5$ and $y_{\max} = 0.5$. This yields the prior distribution of μ_{jb} as

$$\mu_{jb} \sim N(0, \sigma_\mu^2), \quad \text{where } \sigma_\mu = 0.5/k\sqrt{M}. \quad (2.33)$$

This distribution indicates that when M is large, the deviation of μ_{jb} from 0 will be small, and then the effect of each tree is regularized by shrinking μ_{jb} towards zero.

In terms of the variance σ^2 of the error term, the prior is the inverse chi-square distribution $\sigma^2 \sim \nu\lambda/\chi_\nu^2$. Chipman et al. (2010) suggests the value of ν to be between 3 to 10 to obtain an appropriate shape, and λ to be selected such that there is a high probability that the variance is smaller than the residual variance from the least square linear regression estimate. The probability is suggested as 0.75, 0.90 or 0.99. Cross validation could also be used to select appropriate values.

Based on the above settings, the priors for both the tree structure and the parameters

are specified as

$$p((U_1, \Phi_1), \dots, (U_M, \Phi_M), \sigma) = \left[\prod_j p(U_j, \Phi_j) \right] p(\sigma) = \left[\prod_j p(\Phi_j|U_j)p(U_j) \right] p(\sigma), \quad (2.34)$$

where

$$p(\Phi_j|U_j) = \prod_j p(\mu_{jb}|U_j). \quad (2.35)$$

This expression follows from the assumption trees that are independent of each other and of the error variance σ , and the terminal nodes within each tree are independent of each other also. With these assumptions, the prior can be simplified such that it is a product of the priors for terminal nodes, tree structures, and the error variance.

2.5.3 Posterior Distribution

In order to approximate the target function with a sum-of-trees model, BART employs a Bayesian back-fitting MCMC algorithm which considers modifications to individual trees based on the residual of the fitted response from other trees rather than the observed values directly. This is because for each individual tree, say j th tree, the MCMC sampler of U_j , Φ_j and σ follows

$$(U_j, \Phi_j)|U_{(j)}, \Phi_{(j)}, \sigma, y, \quad (2.36)$$

$$\sigma|U_1, \dots, U_M, \Phi_1, \dots, \Phi_M, y. \quad (2.37)$$

The posterior of σ^2 follows inverse gamma distribution. $U_{(j)}$ is the set of all trees in the sum except U_j , and $\Phi_{(j)}$ is defined in the same way. Observing the relationship between y and $U_{(j)}, \Phi_{(j)}$, it is easily obtained that the conditional distribution $p(U_j, \Phi_j|U_{(j)}, \Phi_{(j)}, \sigma, y)$ depends on $(U_{(j)}, \Phi_{(j)}, y)$ only through

$$R_j = y - \sum_{k \neq j} g(x; U_k, \Phi_k), \quad (2.38)$$

where R_j is the remaining residuals of y fit with a sum of trees model across all trees except for the j th tree. For the posterior distribution

$$p((U_1, \Phi_1), \dots, (U_M, \Phi_M), \sigma | y), \quad (2.39)$$

the backfitting MCMC algorithm samples from this distribution using the two-step sampling scheme in (2.36) and (2.37) which is replaced with $(U_j, \Phi_j) | R_j, \sigma$ and $\sigma | U_1, \dots, U_M, \Phi_1, \dots, \Phi_M, y$. Specifically, the first step can be elaborated to draw U_j and Φ_j separately for each tree. When we calculate the marginal distribution of $p(U_j | R_j, \sigma)$, we can get

$$\begin{aligned} p(U_j | R_j, \sigma) &= \int p(U_j, \Phi_j | R_j, \sigma) d\Phi_j \\ &\propto p(U_j) \int p(R_j | U_j, \Phi_j, \sigma) p(\Phi_j | U_j, \sigma) d\Phi_j. \end{aligned} \quad (2.40)$$

By using a conjugate prior on Φ_j , the integral above turns out to have a closed form up to a norming constant. This is good for posterior sampling such that instead of jointly sampling (U_j, Φ_j) , we can draw them in two successive steps as

$$U_j | R_j, \sigma, \quad (2.41)$$

$$\Phi_j | U_j, R_j, \sigma. \quad (2.42)$$

Next, we would like to clarify the way to build a single tree. In Chipman et al. (2010), four moves are suggested to construct the tree. (1) BIRTH: growing a terminal node to generate two new child nodes. (2) DEATH: pruning the two child nodes from their parent node. The parent node will be a new terminal node. (3) CHANGE: changing the split rule of a selected internal node including which variable to be cut on and the corresponding cutpoint. (4) SWAP: swapping the split rule between a pair of parent and child nodes. Kapelner and Bleich (2016) omit the SWAP move and simplify the CHANGE move to only

consider those internal nodes with two terminal nodes to reduce computational complexity.

To complete the Bayesian estimation, the MCMC procedures are repeated many times until convergence. Then, we will remove the “burn-in” iterations and average over draws from the posterior distribution to obtain our final estimator

$$\hat{f}(x) = \frac{1}{K} \sum_{k=1}^K f_k^*(x), \quad (2.43)$$

where $f_k^*(x) = \sum_{j=1}^M g(x; U_j^*, \Phi_j^*)$ is the estimated value from the k th iteration.

Chapter 3: Adaptive Bayesian Sum of Trees Model for Stationary Time Series Data

In Chapter 2, we extend the simplistic i.i.d. model to a covariate-dependent stationary and nonstationary multiple time series data. In this chapter, we propose an adaptive Bayesian sum of trees method for covariate-dependent multiple stationary time series data that we employ the structure of the Bayesian additive regression tree (BART) model but modify it for power spectrum analysis which is a functional response data. Also, it represents an extension of a method for adaptive partitioning of a single covariate proposed in Bruce et al. (2018). We also describe the Bayesian back-fitting MCMC algorithm and reversible-jump MCMC algorithm needed to sample from the posterior distribution to estimate the covariate-dependent power spectrum.

3.1 Adaptive Bayesian Sum of Trees Model

To adjust the BART model for the power spectrum estimation, we specify the form of terminal parameters according to the coefficients in the Bayesian penalized spline model and employ both the backfitting MCMC sampler and the reversible-jump MCMC (RJMCMC) procedure to sample new tree structures and node parameters. More details are presented in the following sections.

3.1.1 Sum of Trees Model

Let M be the number of trees. A sum of trees model for the log power spectrum is then constructed as

$$\log f(\boldsymbol{\omega}, \nu) \approx \sum_{j=1}^M \sum_{b=1}^{B_j} \delta(\boldsymbol{\omega}; U_j, b) \log f_{bj}(\nu), \quad (3.1)$$

where U_j represents the j th tree that has B_j terminal nodes for $j = 1, \dots, M$. Model specification for local power spectra $\log f_{b_j}(\nu)$ within each tree then follows directly from the specification for the single tree model introduced in Section 2.4.2.

It is important to note that the number of trees and hyperparameters for priors introduced in Section 2.4.2 should be selected to ensure a collection of “weak learners”. More specifically, m should be sufficiently large to avoid overly complex individual tree structures, and estimates of $\log f_{b_j}(\nu)$ should be shrunk toward zero by selecting hyperparameters σ_α^2 and A_τ sufficiently small. Particular selections used in simulation studies and applications are discussed in subsequent corresponding sections, and as recommended in Chipman et al. (2010).

3.1.2 Prior Specification

Let $\Phi_j = \{\log f_{1j}(\nu), \dots, \log f_{B_j j}(\nu)\}$ be the collection of log power spectra across terminal nodes for the j th tree. To complete the Bayesian model specification, priors are imposed on U_j and Φ_j in order to allow the trees to be random and fit from the data. Assuming independence across terminal node parameters and trees a priori, priors can be specified as $p((U_1, \Phi_1), \dots, (U_M, \Phi_M)) = \prod_j p(U_j, \Phi_j) = \prod_j p(\Phi_j|U_j)p(U_j)$, where $p(\Phi_j|U_j) = \prod_b p(\log f_{b_j}(\nu)|U_j)$. The priors for $\Phi_j|U_j$ then correspond to the priors of α_b , $\beta^{(b)}$ and τ_b for the local power spectrum estimator introduced in Section 2.4.2. For the priors on the tree structure U_j , three probabilities need to be considered.

1. The probability of a node to be split is the same as function 2.32. Ročková and Saha (2019) propose a minor modification $\Pr(\text{SPLIT}) \propto \gamma^d$ for some $0 \leq \gamma < 1/2$ to achieve the optimal posterior convergence rate, which can also be adopted within the proposed framework.
2. The probability of selecting the p th covariate for splitting is denoted as s_p for $p = 1, \dots, P$. The proposed model allows for two possible prior specifications: a uniform prior, $s_p = P^{-1}$, such that all covariates have the same probability to be selected,

and a sparsity-inducing Dirichlet prior $(s_1, \dots, s_P) \sim \mathcal{D}\left(\frac{\sigma}{P}, \dots, \frac{\sigma}{P}\right)$ (Linero, 2018).

For the Dirichlet prior, σ determines the degree of sparsity and Linero (2018) offer multiple approaches for modeling this parameter. We set $\sigma = 1$ as suggested by Linero (2018) for computational convenience.

3. The probability of selecting a particular cutpoint for a given covariate is uniform across all cutpoints. For continuous, discrete, and ordinal covariates, cutpoints are selected from a fixed number of evenly spaced points over the range of possible values. For categorical covariates without an intrinsic ordering, a cutpoint represents a particular mapping of categories to the left and right child nodes created by the split. A categorical variable with q categories then has $2^q - 2$ cutpoints that can be selected.

3.1.3 Bayesian Backfitting MCMC

It is important to note that each tree captures particular features of the covariate-dependent power spectrum and depends on the features captured by other trees. While this provides considerable flexibility and adaptive estimation, it presents significant computational challenges in estimating the trees. Following Chipman et al. (2010), we develop a Bayesian backfitting MCMC algorithm for proposing and evaluating modifications to each tree sequentially.

For the j th tree, the posterior distribution $p((U_1, \Phi_1), \dots, (U_M, \Phi_M) | \mathbf{I}_1, \dots, \mathbf{I}_L)$ can be sampled through M successive draws from $p((U_j, \Phi_j) | U_{-j}, \Phi_{-j}, \mathbf{I}_1, \dots, \mathbf{I}_L)$ where U_{-j} is the set of all trees except U_j , and Φ_{-j} is defined similarly. Note the conditional distribution $p((U_j, \Phi_j) | U_{-j}, \Phi_{-j}, \mathbf{I}_1, \dots, \mathbf{I}_L)$ depends on $(U_{-j}, \Phi_{-j}, \mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_L)$ only through $\mathbf{R}_{\ell j} = \log \mathbf{I}_{\ell} - \sum_{i \neq j} \sum_{b=1}^{B_i} \delta(\omega_{\ell}; U_i, b) \log \mathbf{f}_{bi}$, where $\log \mathbf{f}_{bi} = \{\log f_{bi}(\nu_1), \log f_{bi}(\nu_2), \dots, \log f_{bi}(\nu_N)\}$ and $\mathbf{R}_{\ell j}$ is the residual of the log periodogram after removing the fit from the sum of trees across all trees except for the j th tree. Therefore, drawing from the posterior equates to M successive draws from $p((U_j, \Phi_j) | \mathbf{R}_{1j}, \dots, \mathbf{R}_{Lj})$. Hence, the local power spectrum estimator for the j th tree is then fit by replacing $\log \mathbf{I}_{\ell}$ with $\mathbf{R}_{\ell j}$ in Equations (2.25) and (2.27).

3.1.4 Reversible-jump MCMC Sampling

To sample new tree structures, a reversible-jump MCMC (Green, 1995) procedure is developed to jointly propose and evaluate new draws of U and Φ for each tree. The proposed modification takes the form of one of three possible moves: BIRTH, DEATH and CHANGE, as described in Section 2.5.3, with probabilities 0.25, 0.25 and 0.5 respectively. New model parameters for the proposed modification are drawn, and the proposed modification is then accepted or rejected using a Metropolis–Hastings (M-H) step. Each tree is considered in turn for updating within each iteration. Draws using this RJMCMC sampler are averaged over post burn-in draws to obtain the final estimator. Technical details for the RJMCMC sampling scheme are available in Appendix B.

3.1.5 Comparison with Bayesian additive regression model

In this section, we compare our proposed modeling framework with that of the BART model. The first important difference is in the nature of the response variable. In the proposed method, the response variable is the power spectrum, which is a functional parameter, rather than a scalar response as in BART. Additionally, the power spectrum has special properties (periodicity, non-negativity, etc.) that need to be considered in the modeling approach. While BART has been adapted to model generic functional responses (Starling et al., 2020), the proposed method is specifically designed to model the power spectrum and retain these properties using a specially-designed set of basis functions. Second, BART does not use RJMCMC sampling. This is due to the use of conjugate Gaussian priors for terminal node means, which allows these parameters to be integrated out when drawing new tree structures, thus avoiding the need for reversible jumps between continuous spaces of varying dimensions. The proposed model does not have such conjugacy, and thus uses RJMCMC for sampling. Third, for the general regression model in (2.30), there is a need to estimate the variance σ^2 of error term. While, the error term ϵ of the log-linear model in (2.6) follows $\log(\chi_2^2/2)$ such that no parameters need to be estimated.

3.2 Simulations of Bayesian Sum of Trees Model

We consider three simulation settings representing abrupt and smoothly varying dynamics with complex covariate effects and interactions in order to demonstrate strong finite-sample estimation accuracy, as well as the ability to adapt to sparse covariate effects and conduct variable selection.

3.2.1 Settings

1. Abrupt+Smooth: Let $\boldsymbol{\omega} = [\omega_1, \omega_2]'$ where $\omega_1, \omega_2 \stackrel{i.i.d.}{\sim} U(0, 1)$. An AR(1) process for the ℓ th time series is specified as $x_{\ell t} = \phi_{\ell} x_{\ell t-1} + \epsilon_{\ell t}$, $\epsilon_{\ell t} \sim N(0, 1)$, where $\phi_{\ell} = -0.7 + 1.4\omega_2$ when $0 \leq \omega_1 < 0.5$ and $\phi_{\ell} = 0.9 - 1.8\omega_2$, when $0.5 \leq \omega_1 \leq 1$.
2. AR-Friedman: Let $\boldsymbol{\omega} = [\omega_1, \dots, \omega_5]'$ where $\omega_1, \dots, \omega_5 \stackrel{i.i.d.}{\sim} U(0, 1)$. An AR(1) process for the ℓ th time series is specified as $x_{\ell t} = \phi_{\ell} x_{\ell t-1} + \epsilon_{\ell t}$, $\epsilon_{\ell t} \sim N(0, 1)$, where $\phi_{\ell} = 0.5 \sin(\pi\omega_{\ell 1}\omega_{\ell 2}) - (\omega_{\ell 3} - 0.5)^2 + 0.35\text{sign}(\omega_{\ell 4} - 0.5) - 0.15\omega_{\ell 5}$.
3. Adjusted-AdaptSPEC-X: Let $\boldsymbol{\omega} = [\omega_1, \omega_2]'$, where $\omega_1, \omega_2 \stackrel{i.i.d.}{\sim} U(0, 1)$. Each covariate vector $\boldsymbol{\omega}$ is mapped to a latent variable $z_{\ell} \in \{1, 2, 3, 4\}$. Figure D.2 shows the mapping from $\boldsymbol{\omega}$ to z . An AR(2) process is then specified as $x_{\ell t} = \phi_{z_{\ell}1} x_{\ell t-1} + \phi_{z_{\ell}2} x_{\ell t-2} + \epsilon_{\ell t}$, $\epsilon_{\ell t} \sim N(0, 1)$, where the coefficients for the four latent region are defined as follows. If $z_{\ell} = 1$, $(\phi_{z_{\ell}1}, \phi_{z_{\ell}2}) = (1.5, -0.75)$; if $z_{\ell} = 2$, $(\phi_{z_{\ell}1}, \phi_{z_{\ell}2}) = (-0.8, 0)$; if $z_{\ell} = 3$, $(\phi_{z_{\ell}1}, \phi_{z_{\ell}2}) = (-1.5, -0.75)$; if $z_{\ell} = 4$, $(\phi_{z_{\ell}1}, \phi_{z_{\ell}2}) = (0.2, 0)$.

The first setting represents an AR(1) process in which the coefficient varies smoothly across one covariate and abruptly across another. The second setting contains complex linear and nonlinear covariate effects and interactions (Friedman, 1991) adapted for time series data. The third setting represents an abruptly-changing process over two dimensions, similar to that of Bertolacci et al. (2022). Power spectra for covariate-dependent AR(1) processes from the first two settings can be represented as $f(\boldsymbol{\omega}_{\ell}, \nu) = [1 - 2\phi_{\ell} \cos(2\pi\nu) + \phi_{\ell}^2]^{-1}$ where

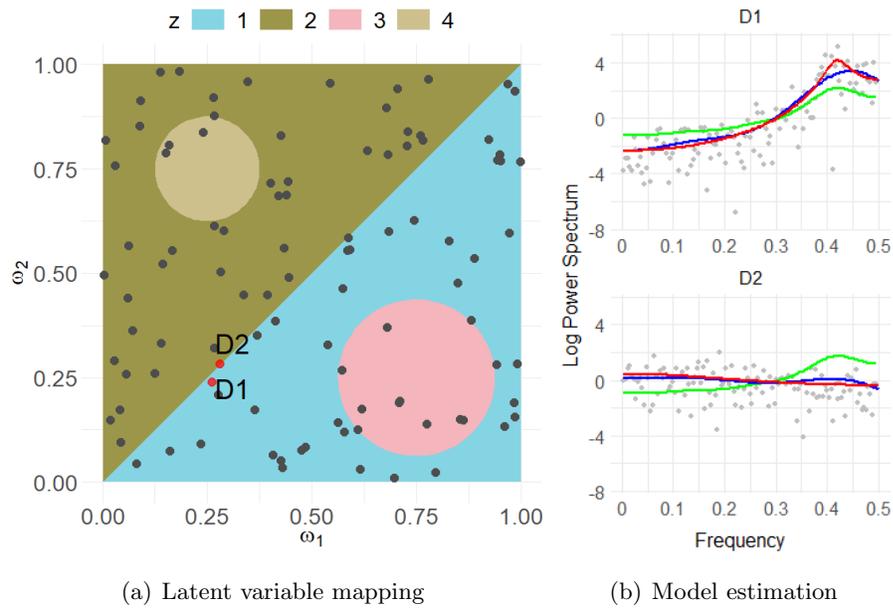


Figure 3.1: (a) presents the mapping of covariate values ω_1 and ω_2 to latent variable values z for the Adjusted-AdaptSPEC-X simulation setting. D1 and D2 denote two simulated time series with similar covariate values that are mapped to different latent variable values corresponding to different power spectra. For these two realizations, (b) displays the true log power spectra (red lines), log periodogram ordinates (gray points), and estimated log power spectra using the proposed Bayesian sum of trees model (blue lines) and the AdaptSPEC-X model (green lines).

$\boldsymbol{\omega}_\ell$ represents the covariate vector for the ℓ th time series and ϕ_ℓ depends on $\boldsymbol{\omega}_\ell$ as defined above. For the third setting, covariate-dependent AR(2) processes can similarly be represented as $f(\boldsymbol{\omega}_\ell, \nu) = [1 + 2(-\phi_{z_\ell 1} + \phi_{z_\ell 1}\phi_{z_\ell 2}) \cos(2\pi\nu) - 2\phi_{z_\ell 2} \cos(4\pi\nu) + \phi_{z_\ell 1}^2 + \phi_{z_\ell 2}^2]^{-1}$.

3.2.2 Results: Estimation Accuracy

Hyperparameters are specified as $\sigma_\alpha^2 = 100$ for the prior variance of α_b in Equation (2.26) and $\xi_\tau = 2$ and $A_\tau = 10$ for the prior on τ_b in Equation (2.28) and Equation (2.29). We use both $M = 5$ trees and $M = 50$ trees for estimation, and different numbers (L) and lengths (T) of time series are considered. The MCMC procedure is run for a total of 10,000 iterations with the first 5,000 discarded as burn-in. In order to assess convergence, trace plots for summary measures of the mean squared residuals, estimated log power spectrum, and tree structures for all settings are available in Appendix D. These diagnostics appear to show convergence after approximately 5,000 iterations across all settings.

Estimates of the covariate-dependent power spectrum for a single run are presented in Figure 3.2 to visually illustrate the ability of the proposed method to capture abrupt and smooth changes in power spectra. The smooth change in the conditional power spectrum over ω_2 is captured by averaging over the posterior distribution of tree structures, which contains many splits across the range of possible values of ω_2 . On the other hand, splits on ω_1 are concentrated around the true abrupt change at $\omega_1 = 0.5$, and the posterior mean estimator of the conditional power spectrum over ω_1 appropriately reflects the abrupt change in the conditional power spectrum.

Given an estimate of the covariate-dependent log power spectrum, $\hat{f}(\boldsymbol{\omega}, \nu)$, the MSE,

$$\text{MSE} = (NL)^{-1} \sum_{\ell=1}^L \sum_{k=1}^N \left[\log \hat{f}(\boldsymbol{\omega}_\ell, \nu_j) - \log f(\boldsymbol{\omega}_\ell, \nu_j) \right]^2, \quad (3.2)$$

will be used to evaluate estimation accuracy. True covariate-dependent log power spectra, $\log f(\boldsymbol{\omega}, \nu)$ for all settings can be fully determined by the representations and coefficients

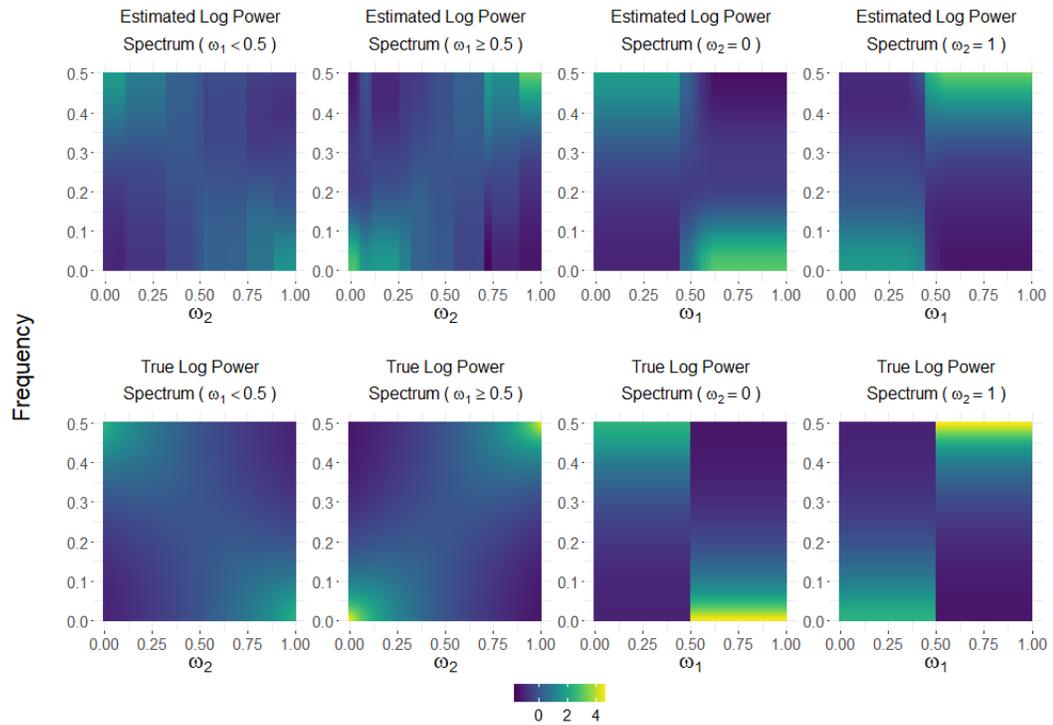


Figure 3.2: Estimated and true covariate-dependent conditional log power spectrum for one run of the Abrupt+Smooth simulation setting. The first two columns contain the estimated and true covariate-dependent power spectrum conditional on $\omega_1 < 0.5$ and $\omega_1 \geq 0.5$ respectively; The last two columns display the estimated and true covariate-dependent power spectrum conditional on $\omega_2 = 0$ and $\omega_2 = 1$ respectively.

presented in Section 3.2.1. The mean and standard deviation of the mean squared error (MSE) is presented in Table 3.1 for 100 replications of all settings. For comparison, AdaptSPEC-X (Bertolacci et al., 2022) is also used to estimate the covariate-dependent power spectrum. It is important to note that AdaptSPEC-X allows for modeling of time- and covariate-varying power spectra with time-varying means, which is a more general setting than what is considered in this work. Accordingly, we implement a simplified version of the AdaptSPEC-X mixture model without the time-varying mean and power spectra components to enable more accurate comparisons. We applied $C = 50$ mixture components to ensure sufficient flexibility in estimating covariate effects on power spectra.

These results show that both methods see improved estimation accuracy as the number (L) and length (T) of time series increase. For the proposed method, using 50 trees provides slight improvements in estimation accuracy compared to using 5 trees when $L > 100$ and $T > 100$ since the additional trees provide more flexibility to capture complexities in the covariate-dependent power spectra. However, the proposed method using either 5 or 50 trees produces smaller MSEs than AdaptSPEC-X for comparable settings. This can be partially attributed to the presence of abrupt changes across one or more covariates in all settings, which are better captured by the proposed tree-based approach. To illustrate this point, Figure 3.1(b) shows the estimated power spectra for two simulated time series with similar covariate values separated by an abrupt change from the Adjusted-AdaptSPEC-X setting. As noted in Bertolacci et al. (2022), the thin-plate Gaussian process prior on the mixture weights, while flexible, is both smooth and stationary. Accordingly, the AdaptSPEC-X power spectrum estimates for these two time series are similar due to smoothing across similar covariate values induced by the thin-plate Gaussian process prior. On the other hand, the proposed Bayesian sum-of-trees model is able to accurately distinguish the abrupt change in the power spectra. Moreover, averaging over the posterior distribution of trees enables the proposed method to recover smooth changes well (Figure 3.2). Taken together, this results in superior estimation accuracy across all settings seen in Table 3.1.

Table 3.1: Mean and standard deviation of MSE over 100 replications for three simulation settings with different lengths (T) and number (L) of time series. Results are presented for the proposed Bayesian sum of trees model (top) with number of trees $M = 50$ and the AdaptSPEC-X model (bottom) with mixture components $C = 50$.

		Abrupt+Smooth	AR-Friedman	Adj. AdaptSPEC-X
		Proposed Bayesian Sum of Trees Model (5 trees)		
L	T			
100	100	0.04(0.01)	0.04(0.01)	0.15(0.02)
100	250	0.02(0.00)	0.02(0.00)	0.08(0.01)
100	500	0.01(0.00)	0.02(0.00)	0.05(0.01)
200	100	0.03(0.01)	0.03(0.00)	0.14(0.02)
200	250	0.02(0.00)	0.02(0.00)	0.07(0.01)
200	500	0.01(0.00)	0.01(0.00)	0.05(0.01)
500	100	0.02(0.00)	0.03(0.00)	0.13(0.01)
500	250	0.01(0.00)	0.02(0.00)	0.07(0.01)
500	500	0.01(0.00)	0.01(0.00)	0.05(0.01)
		Proposed Bayesian Sum of Trees Model (50 trees)		
L	T			
100	100	0.04(0.01)	0.04(0.00)	0.153(0.02)
100	250	0.02(0.00)	0.02(0.00)	0.07(0.02)
100	500	0.01(0.00)	0.01(0.00)	0.05(0.01)
200	100	0.03(0.00)	0.03(0.00)	0.14(0.02)
200	250	0.01(0.00)	0.02(0.00)	0.07(0.01)
200	500	0.01(0.00)	0.01(0.00)	0.05(0.01)
500	100	0.02(0.00)	0.02(0.00)	0.13(0.01)
500	250	0.01(0.00)	0.01(0.00)	0.07(0.01)
500	500	0.01(0.00)	0.01(0.00)	0.05(0.00)
		AdaptSPEC-X Model		
L	T			
100	100	0.06(0.01)	0.05(0.01)	0.63(0.13)
100	250	0.04(0.01)	0.04(0.00)	0.56(0.12)
100	500	0.04(0.01)	0.04(0.01)	0.54(0.12)
200	100	0.05(0.01)	0.05(0.01)	0.52(0.07)
200	250	0.04(0.01)	0.04(0.01)	0.45(0.07)
200	500	0.04(0.01)	0.04(0.01)	0.42(0.06)
500	100	0.05(0.01)	0.04(0.00)	0.46(0.04)
500	250	0.04(0.01)	0.03(0.00)	0.40(0.04)
500	500	0.04(0.01)	0.03(0.00)	0.399(0.04)

3.2.3 Comparison with Generalized Additive Model

When the number of covariates is small, the proposed method can also be compared with a generalized additive model (GAM) approach (Wood, 2017) for estimating the covariate-dependent power spectrum. This is an important benchmark for comparison, since the GAM modeling approach is frequently used in practice, and its smooth additive structure is well-equipped to characterize smooth covariate effects.

We consider two simulation settings representing 1) a combination of abrupt and smoothly varying dynamics, and 2) entirely smoothly varying dynamics in order to compare finite-sample estimation accuracy.

1. Abrupt+Smooth: Let $\boldsymbol{\omega} = (\omega_1, \omega_2)$ where $\omega_1, \omega_2 \stackrel{i.i.d.}{\sim} U(0, 1)$. An AR(1) process for the ℓ th time series is specified as $x_{\ell t} = \phi_{\ell} x_{\ell t-1} + \epsilon_{\ell t}$, $\epsilon_{\ell t} \sim N(0, 1)$, where $\phi_{\ell} = -0.7 + 1.4\omega_2$ when $0 \leq \omega_1 < 0.5$ and $\phi_{\ell} = 0.9 - 1.8\omega_2$, when $0.5 \leq \omega_1 \leq 1$.
2. AR-Friedman-Smooth: Let $\boldsymbol{\omega} = (\omega_1, \omega_2)$ where $\omega_1, \omega_2 \stackrel{i.i.d.}{\sim} U(0, 1)$. An AR(1) process for the ℓ th time series is specified as $x_{\ell t} = \phi_{\ell} x_{\ell t-1} + \epsilon_{\ell t}$, $\epsilon_{\ell t} \sim N(0, 1)$, where $\phi_{\ell} = 0.5 \sin(\pi\omega_{\ell 1}\omega_{\ell 2})$.

The first setting is exactly as appears in the simulation studies Section 3.2.1, and the second setting is a simplification of the AR-Friedman setting in Section 3.2.1 removing all but the first two important covariates, which influence the power spectra in a smooth manner. Since the power spectrum for the covariate-dependent AR(1) processes can be represented as $f(\boldsymbol{\omega}_{\ell}, \nu) = [1 - 2\phi_{\ell} \cos(2\pi\nu) + \phi_{\ell}^2]^{-1}$ where $\boldsymbol{\omega}_{\ell}$ represents the covariate vector for the ℓ th time series and ϕ_{ℓ} is defined as above, plugging in the corresponding expressions for ϕ_{ℓ} leads to the following covariate-dependent power spectra for the above settings.

$$1. f(\boldsymbol{\omega}_{\ell}, \nu) = \begin{cases} [1.49 - 1.96(\omega_2 - \omega_2^2) + (1.4 - 2.8\omega_2) \cos(2\pi\nu)]^{-1} & 0 \leq \omega_1 < 0.5 \\ [1.81 - 3.24(\omega_2 - \omega_2^2) - (1.8 - 3.6\omega_2) \cos(2\pi\nu)]^{-1} & 0.5 \leq \omega_1 \leq 1 \end{cases}$$

$$2. f(\boldsymbol{\omega}_{\ell}, \nu) = [1 - \sin(\pi\omega_1\omega_2) \cos(2\pi\nu) + 0.25 \sin^2(\pi\omega_1\omega_2)]^{-1}$$

From this representation, complex interactions among covariates and frequency, as well as nonlinear effects, are observed, so a general modeling approach using GAMs to approximate covariate-dependent log power spectra must go beyond univariate additive smooths to include multi-dimensional smooths as well. Here is one such possible model for the above simulation settings with two covariates.

$$\mathbb{E}[\log f(\boldsymbol{\omega}_\ell, \nu)] = g_0(\nu) + g_1(\omega_1) + g_2(\omega_2) + g_3(\omega_1, \omega_2) + g_4(\omega_1, \nu) + g_5(\omega_2, \nu) + g_6(\omega_1, \omega_2, \nu) \quad (3.3)$$

where g_0, g_1, \dots, g_6 are functions that vary over the input parameters smoothly. This type of model is well-equipped to capture smooth changes in covariate-dependent log power spectra and can be estimated using the `mgcv` R package (Wood, 2017; R Core Team, 2022).

Given a collection of stationary time series $X_{\ell t}$ of length $t = 1, \dots, T$ and P -dimensional covariate vectors $\boldsymbol{\omega}_\ell = (\omega_{1\ell}, \dots, \omega_{P\ell})'$ for $\ell = 1, \dots, L$ independent subjects, let $N = \lfloor T/2 \rfloor - 1$ and $\nu_k = k/T$ for $k = 1, \dots, N$ be the Fourier frequencies. The periodogram estimator of the power spectrum for the ℓ th time series is $I_\ell(\nu_k) = \frac{1}{T} \left| \sum_{t=1}^T X_{\ell t} \exp(-2\pi i \nu_k t) \right|^2$. The log periodogram ordinates, $\log I_\ell(\nu_k)$, can then be used as the observed response variable values within the GAM modeling framework above to smooth periodogram ordinates over covariate and frequency in order to estimate the covariate-dependent log-power spectrum.

It is important to note that this approach is only feasible when the number of covariates is small, as in these two settings. Directly extending this modeling approach to accommodate many covariates without simplifying assumptions about the impact of covariates on the log power spectrum would require significantly more terms and multi-dimensional smooths. This would significantly increase computational complexity and hamper interpretability.

Given an estimate of the covariate-dependent log power spectrum from either the GAM model or the proposed model, $\hat{f}(\boldsymbol{\omega}, \nu)$, the MSE will be used to evaluate estimation accuracy. For the proposed method, hyperparameters are specified as $\sigma_\alpha^2 = 100$, $\xi_\tau = 2$, and $A_\tau = 10$.

$M = 5$ trees are used for estimation, and different numbers (L) and lengths (T) of time series are considered. The MCMC procedure is run for a total of 10,000 iterations with the first 5,000 discarded as burn-in. Table 3.2 displays the mean and standard deviation for the MSE over 100 replicates using both the proposed method and the GAM modeling approach.

Table 3.2: Mean and standard deviation of MSE over 100 replications for two simulation settings with different lengths (T) and number (L) of time series. Lower mean MSEs for each setting are bolded for ease of comparison.

L	T	Abrupt+Smooth		AR-Friedman-Smooth	
		Proposed (5 trees)	GAM	Proposed (5 trees)	GAM
100	100	0.04(0.01)	0.10(0.01)	0.02(0.00)	0.03(0.01)
100	250	0.02(0.00)	0.06(0.01)	0.01(0.00)	0.01(0.00)
100	500	0.01(0.00)	0.05(0.01)	0.01(0.00)	0.01(0.00)
200	100	0.03(0.01)	0.08(0.01)	0.02(0.00)	0.02(0.00)
200	250	0.02(0.00)	0.06(0.01)	0.01(0.00)	0.01(0.00)
200	500	0.01(0.00)	0.05(0.01)	0.00(0.00)	0.00(0.00)
500	100	0.02(0.00)	0.08(0.01)	0.01(0.00)	0.02(0.00)
500	250	0.01(0.00)	0.06(0.01)	0.01(0.00)	0.01(0.00)
500	500	0.01(0.00)	0.06(0.00)	0.00(0.00)	0.00(0.00)

For the Abrupt+Smooth setting, both methods perform well in estimating the power spectrum under all settings, but the proposed method outperforms the GAM approach. This is expected and can be attributed to the presence of both abrupt (ω_1) and smooth (ω_2) covariate effects, for which our model can approximate reasonably well. The GAM model's smooth approximation leads to biased estimation of abrupt changes, which leads to increased MSEs compared to the proposed sum-of-trees model.

For the AR-Friedman-Smooth setting, again both models perform well in estimating the power spectrum, but performance depends on the number L and length T of time series available for modeling. For smaller L and T values, the proposed method performs slightly better than the GAM model. This could be attributed to a lack of sufficient signal strength in the smaller datasets to accurately estimate the numerous parameters in the GAM model. Also, since the complexity (i.e. tree size) of the model for the proposed method can vary depending on the signal strength, this may offer a slight advantage to the proposed approach. This is not unexpected, as sum of tree style models have recently been shown to offer superior performance in low signal to noise ratio settings (Mentch and Zhou, 2020). However, as L and T increase, the GAM model outperforms the proposed model. This is very much expected, since the smooth GAM model structure is more efficient in learning the covariate-dependent power spectrum when the smoothness assumption is correct.

These results show that methods assuming completely smooth effects, like GAM, can outperform the proposed method when smoothness assumptions are valid and sufficient data are available. However, the proposed method generally outperforms these methods in the presence of both smooth and abrupt covariate effects without prior knowledge of the nature of such effects.

3.2.4 Results: Computation Time

Simulations were carried out on a Windows 10 machine with an 8-core Intel i7 3.6 GHz processor and 64 GB RAM using R version 4.0.3 (R Core Team, 2021). The R code for implementing the proposed model is provided as a zip file in Supporting Information and is described in Web Appendix C. Computationally-intensive aspects of the methodology are written in C++ using `RcppArmadillo` (Eddelbuettel and Sanderson, 2014) for more efficient computation and reduced run times. Replications were run in parallel across six cores. For the simulation settings considered herein, the mean run time for each tree update after burn-in ranges from 0.02 to 0.43 seconds, depending on the number (L) and length (T) of time series, which is shown in Figure 3.3.

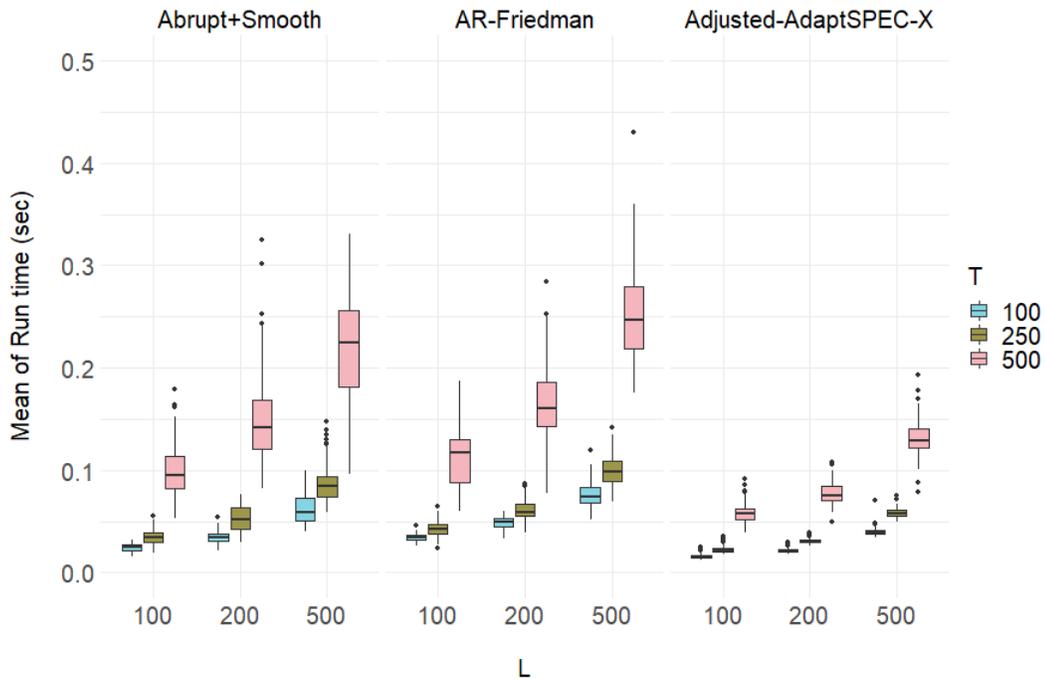


Figure 3.3: The distribution of mean run times in seconds for a single tree update over 100 replicates of the three simulations with $M = 5$ trees.

While run times generally increase as the number and length of time series increase, length increases have a bigger impact on run times relative to increases in the number of time series. Increasing T from 100 to 500 while holding L constant increases mean run times by a multiple of approximately 3.5, while increasing L from 100 to 500 while holding T constant increases mean run times by a multiple of approximately 2.1. This is expected since the number of Fourier frequencies grows with the length of the series and increases both the number of terms being summed in the log Whittle likelihood and the dimension of the cosine basis used to approximate local log power spectra. However, increasing the number of time series does not change the number of Fourier frequencies and only requires adding

more terms in the log Whittle likelihood and posterior distribution for the spline coefficients, which is less computationally expensive. See Equations (2.25)-(2.27) and sampling scheme details in Appendix B for more details.

Tree size also plays an important role in determining run times for tree updates. Larger trees tend to have fewer time series belonging to each terminal node, which reduces the computational burden and run times for evaluating modifications to a single terminal node. This is why the Adjusted-AdaptSPEC-X setting, which requires larger trees to recover the complex covariate effects, has faster mean run times compared to other settings. Additionally, tree size tends to increase as both L and T increase. For illustration, Figure 3.4 shows the total number of bottom nodes for different combinations of L and T for the AR-Friedman simulation.

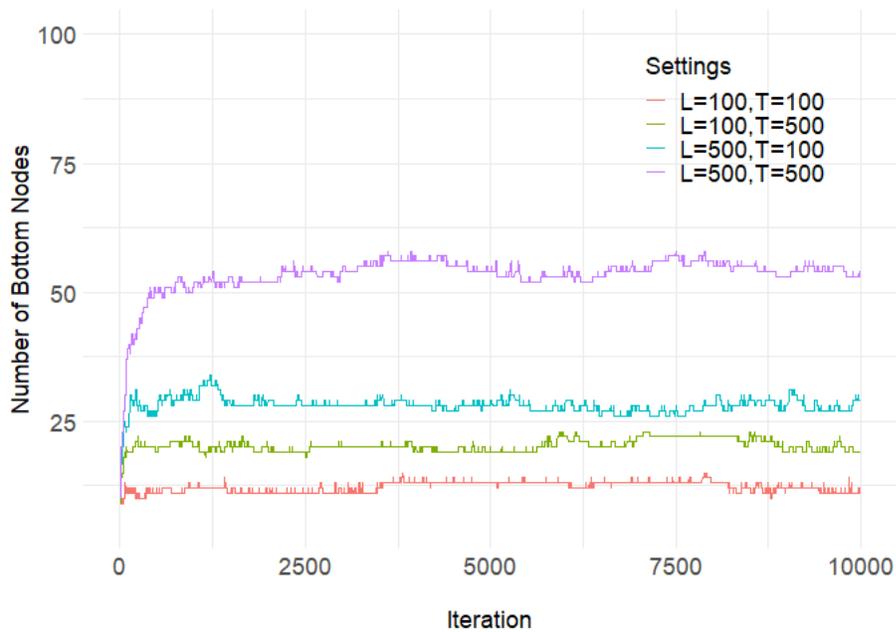


Figure 3.4: The total number of bottom nodes over iterations for four runs with different L and T for the AR-Friedman simulation setting using 5 trees.

The total number of bottom nodes for $L = 500, T = 500$ is over 50 while it is about 13 for $L = 100, T = 100$. This can help explain why the mean run times grow more slowly in L compared to T , as the mean number of time series belonging to each terminal node increases modestly.

3.2.5 Results: Sparse Covariate Effects

To demonstrate the capability of the proposed method in providing efficient variable selection by adapting to sparse covariate effects for high-dimensional covariates, we introduce additional noise covariates into the simulation settings described in Section 3.2.1. Consider augmenting the original covariate vector, $\boldsymbol{\omega}$, with additional noise covariates independently drawn from a standard normal distribution such that $\tilde{\boldsymbol{\omega}} \sim N(\mathbf{0}, \mathbf{I})$ where \mathbf{I} is the identity matrix and $\boldsymbol{\omega}^* = [\boldsymbol{\omega}', \tilde{\boldsymbol{\omega}}']'$ is the augmented covariate vector used for estimating the covariate-dependent power spectrum. We consider augmented covariate vectors of size $P = 100, 200$, and 1000 using $L = 500$ time series of length $T = 250$. Since the original simulation settings only have 2-5 important covariates, this results in highly sparse covariate vectors in which noise covariates constitute 95% to 99.8% of the total covariates. Variable selection efficiency can be investigated by assessing the estimated posterior probability for model inclusion of each covariate, which is the proportion of posterior draws where the covariate appears in at least one split rule for at least one tree. Table 3.3 displays the MSE and posterior probabilities of model inclusion for the AR-Friedman setting using both the uniform and Dirichlet hyperpriors for tree splitting proportions introduced in Section 3.1.2 with $M = 50$ trees. Tables for all simulation settings for the two different hyperpriors using different numbers of trees ($M = 5, 50$) and visualizations of the posterior model inclusion probabilities are available in Appendix D.

Table 3.3: Mean and standard deviation of MSE and posterior probability of model inclusion for AR-Friedman setting with $L = 500$ time series of length $T = 250$ over 100 replications with different number of covariates (P) and hyperpriors (Uniform and Dirichlet). Posterior probabilities for model inclusion are reported for important variables, $\omega_1, \dots, \omega_5$, individually and noise variables $\omega_6, \dots, \omega_P$ in aggregate. Results are presented for the proposed Bayesian sum of trees model using $M = 50$ trees.

P=100	Uniform	Dirichlet
MSE	0.02(0.00)	0.01(0.00)
Noise	0.27(0.35)	0.13(0.33)
ω_1	1.00(0.00)	1.00(0.00)
ω_2	1.00(0.00)	1.00(0.00)
ω_3	1.00(0.00)	1.00(0.00)
ω_4	1.00(0.00)	1.00(0.00)
ω_5	1.00(0.00)	1.00(0.00)
P=200	Uniform	Dirichlet
MSE	0.02(0.00)	0.01(0.00)
Noise	0.21(0.32)	0.08(0.26)
ω_1	1.00(0.00)	1.00(0.00)
ω_2	1.00(0.00)	1.00(0.00)
ω_3	1.00(0.00)	1.00(0.00)
ω_4	1.00(0.00)	1.00(0.00)
ω_5	1.00(0.00)	0.99(0.06)
P=1000	Uniform	Dirichlet
MSE	0.03(0.01)	0.02(0.01)
Noise	0.13(0.29)	0.07(0.23)
ω_1	0.96(0.17)	0.89(0.32)
ω_2	0.94(0.20)	0.83(0.37)
ω_3	0.68(0.44)	0.50(0.48)
ω_4	1.00(0.00)	1.00(0.00)
ω_5	0.63(0.46)	0.47(0.49)

From this table, some important trends should be noted. First, there is a slight improvement in estimation accuracy when using the Dirichlet hyperprior. This can be attributed to the proposed method's ability to recover important variables with high posterior probability

and the Dirichlet hyperprior’s superior ability to adapt to the sparse covariate vectors considered. This can be seen in Table 3.3 by comparing model inclusion posterior probabilities for important and noise variables. Important variables are included in models with high posterior probability ranging from 46.9% to 100% on average. However, the Dirichlet hyperprior shrinks posterior probabilities for noise variables closer to zero for all covariate vector sizes considered by 47%-62% on average compared to the uniform hyperprior. Second, both estimation accuracy and variable selection are negatively impacted by increasing covariate vector size P . This is expected since the additional noise covariates result in sparser covariate vectors, for which resolving important variables is relatively more difficult. Also, additional noise covariates inject additional randomness into the proposed model which leads to implicit regularization (Mentch and Zhou, 2020).

3.3 Gait Maturation Analysis

We now present the analytical results of applying the proposed method to the motivating gait maturation study described in the introduction (Goldberger et al., 2000), and the inverse regression for the estimating of unknown covariates.

3.3.1 Estimation Results of Gait Maturation Analysis

The current analysis considers the effect of age on gait variability to better understand gait maturation in young children in the presence of other factors that may influence gait, such as gender and gait speed. The data contains stride interval time series from 50 healthy children with equal numbers of girls and boys between 3 and 14 years old. The time series consist of $T = 256$ stride times during normal walking after removing the first 60 seconds and last 5 seconds to avoid warm-up and ending effects (Figure 1.1). More details of data processing can be found in Hausdorff et al. (1999). The proposed Bayesian sum of trees model was used to estimate the covariate-dependent power spectrum of stride interval time series using 5 trees and 10,000 total iterations with the first 5,000 iterations discarded as burn-in. See Appendix D for convergence diagnostics and Appendix E for graphical posterior predictive

checks for this application, which demonstrate stable estimation and model adequacy.

Partial dependence (PD) (Friedman, 2001) is the most widely used method for evaluating covariate effects in machine learning models. However, there is an issue with multicollinearity in this dataset, as age and gait speed are significantly correlated ($r = 0.653, p < 0.0001$), which can render PD unreliable due to extrapolation of the response at predictor values far outside the multivariate envelope of the data (Apley and Zhu, 2020). Therefore, we use accumulated local effects (ALE) (Apley and Zhu, 2020) to characterize covariate effects. ALE presents the effect in a small interval of the interested feature, which can mitigate issues with multicollinearity by localizing the estimated effect within the envelope of the data. Let $\boldsymbol{\omega} = (\omega_j, \boldsymbol{\omega}_{\setminus j})$ where ω_j denotes the j th covariate and $\boldsymbol{\omega}_{\setminus j}$ denotes all other covariates. The ALE for $\omega_j = x$ on the power spectrum at frequency ν is defined as

$$f_{j,\text{ALE}}(x, \nu) = \int_{z_{0,j}}^x E_{\boldsymbol{\omega}_{\setminus j} | \omega_j} \left[\frac{\delta f(\boldsymbol{\omega}, \nu)}{\delta \omega_j} \Big|_{\omega_j = z_j} \right] dz_j - \text{constant} \quad (3.4)$$

where $\mathbb{Z}_j = \{z_{0,j}, \dots, z_{H,j}\}$ 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. Let $\hat{f}(z_{h,j}, x_{\setminus j}; \nu)$ be the estimated power spectrum from a single posterior draw for $\omega_j = z_{h,j}$, $h = 1, \dots, H$ and $\boldsymbol{\omega}_{\setminus j} = x_{\setminus j}$ on frequency ν , the uncentered ALE can then be estimated by

$$\hat{g}_{j,\text{ALE}}(x, \nu) = \sum_{h=1}^{h_j(x)} \frac{1}{n_j(h)} \sum_{\{i: x_j^{(i)} \in N_j(h)\}} \left[\hat{f}(z_{h,j}, x_{\setminus j}^{(i)}; \nu) - \hat{f}(z_{h-1,j}, x_{\setminus j}^{(i)}; \nu) \right] \quad (3.5)$$

where $h_j(x)$ is the index for the interval of the partition \mathbb{Z}_j to which the value x belongs, n is the total number of observations and $n_j(h)$ is the number of observations in the h th segment of the partition for ω_j such that $\sum_{h=1}^H n_j(h) = n$. $N_j(h) = (z_{h-1,j}, z_{h,j}]$ represents

the h th interval of the partition for ω_j . Then the estimated centered ALE is

$$\hat{f}_{j,\text{ALE}}(x, \nu) = \hat{g}_{j,\text{ALE}}(x, \nu) - \frac{1}{n} \sum_{i=1}^n \hat{g}_{j,\text{ALE}}(x^{(i)}, \nu). \quad (3.6)$$

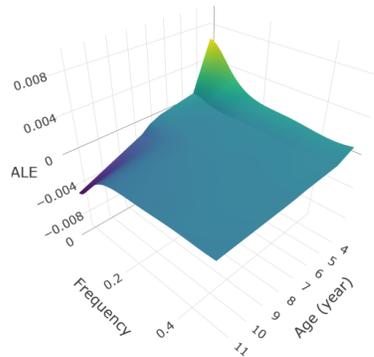
The estimated centered ALE above can be computed for each of the posterior draws from the RJMCMC sampler, and the posterior mean and 95% credible intervals can provide the desired inference on the covariate effects. By partitioning covariates into $H = 5$ intervals containing equal numbers of observations, Figures 3.5(a) and 3.5(b) show the posterior mean of the ALE for age and gait speed on the power spectrum. Corresponding 95% credible intervals are available in Appendix D. Two findings can be concluded from these plots. First, power over all frequencies decreases as age increases. This indicates variability in stride times decreases with age, which is consistent with previous findings (Hausdorff et al., 1999). Second, we can observe that power in low frequencies (LF) (0.05-0.25 stride⁻¹) decreases much more with age relative to higher frequencies (HF) (0.25-0.5 stride⁻¹), which is also expected as low frequency power corresponds to fluctuations over relatively longer time scales and is indicative of less mature neuromuscular control (Hausdorff et al., 1999). It should be noted that Hausdorff et al. (1999) discretizes age as a categorical variable and then considers an ANOVA model to test for age effects. Our model, on the other hand, considers the age-dependent ALE as a continuous surface, which provides a more comprehensive assessment of the association between frequency patterns of gait variability time series and age without subjectively categorizing age into different bins.

Figures 3.5(c) and 3.5(d) present the posterior mean of the ALE of age and gait speed on the LF/HF ratio $\frac{\text{LF}}{\text{HF}}(\omega) = \int_{0.05}^{0.25} f(\omega, \nu) d\nu / \int_{0.25}^{0.5} f(\omega, \nu) d\nu$ along with 95% pointwise credible intervals. This can be computed by replacing the power spectrum $f(\omega, \nu)$ and estimated power spectrum $\hat{f}(z_{h,j}, x_{\setminus j}; \nu)$ in Equations (3.4)-(3.6) with the $\frac{\text{LF}}{\text{HF}}(\omega)$ and its corresponding estimates $\widehat{\frac{\text{LF}}{\text{HF}}}(\omega) = \sum_{\nu_k \in (0.05, 0.25)} \hat{f}(\omega, \nu_k) / \sum_{\nu_k \in (0.25, 0.5)} \hat{f}(\omega, \nu_k)$ where $\widehat{\frac{\text{LF}}{\text{HF}}}(\omega)$ can be

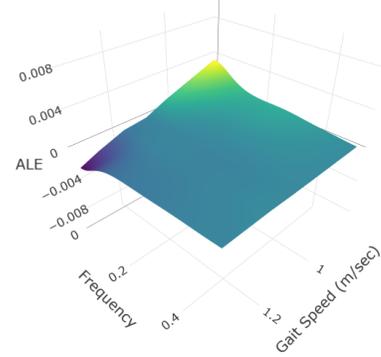
expressed as $\widehat{\frac{LF}{HF}}(z_{h,j}, x_{\setminus j})$ when calculating the ALE for the j th covariate. While $LF/HF(\omega)$ decreases significantly with age, we see relatively larger decreases beyond 7 years of age. Previous analyses of gait variability time series have produced contradictory results. Hausdorff et al. (1999) show that the LF/HF ratio presents a significant decrease in children 7-14 years of age. However, Preis et al. (1997) show that gait maturation occurs more rapidly in children 3-7 years of age and changes slowly after 7 years. Our results are more consistent with analyses suggesting LF/HF ratio decreases beyond 7 years of age.

For gait speed, we observe significantly less low frequency ($< 0.1 \text{ stride}^{-1}$) power (and lower LF/HF ratio) among faster walkers with speeds above 1.1 m/sec. Gender appears to have a much less effect on power spectra (see Appendix D), with only a very small frequency range from 0.05 stride^{-1} to 0.15 stride^{-1} significantly different in power between males and females. There are no significant differences in LF/HF ratio between males and females with a posterior mean ALE of -0.1680 and 95% credible interval (-0.39,0.06).

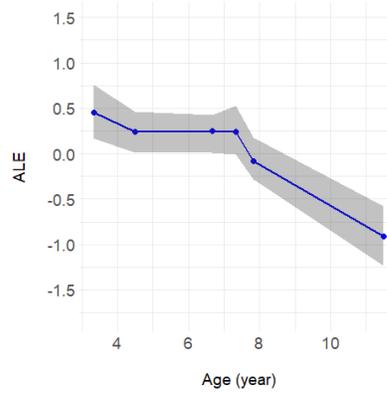
We provide some possible extensions of the proposed method which can be used to accomplish other goals of analyzing gait variability time series. First, other gait maturation studies have focused on classifying subjects into different groups based on the covariates and gait variability time series through support vector machines (Wu and Krishnan, 2009a,b; Wu and Shi, 2011). Second, if disease status is considered as a covariate, then this extension can potentially aide in diagnosing certain diseases, such as Parkinson’s disease, which is another goal of many gait studies (Daliri, 2012; Khorasani and Daliri, 2014). Third, visualization of covariate effects is a major advantage of tree-based models. In Appendix D, a visualization of the tree structures for the gait maturation analysis are presented.



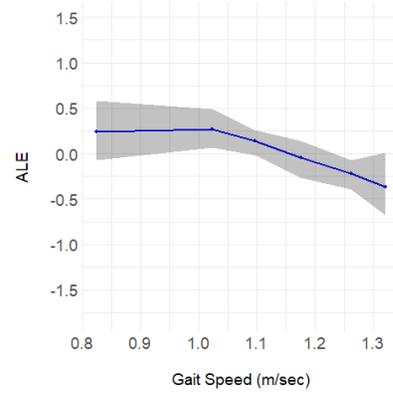
(a)



(b)



(c)



(d)

Figure 3.5: Posterior mean of ALE for age (a) and gait speed (b) effects on the power spectrum and posterior mean of ALE for age (c) and gait speed (d) on LF/HF ratio (blue dotted line) with 95% pointwise credible intervals (shaded gray region).

Chapter 4: Inverse Regression Framework

One of the primary advantages of tree-based methods, like the proposed one, over less transparent machine learning methods, like support vector machines (SVM), is the ability to visualize covariate effects through the tree structures. Appendix D demonstrates this capability for the proposed method in how covariates are associated with the characteristics of the log power spectrum for two different participants in the gait maturation study.

Now suppose instead that we have a realization of a time series and are missing one or more of the covariate values. We can certainly estimate the log power spectra from the individual time series realization, without using the proposed method, via standard techniques such as periodogram smoothing (Shumway and Stoffer, 2017). Can we then use the proposed method’s characterization of the relationship between the log power spectra and covariates to then inversely estimate appropriate values for the unknown covariates? In this way, we can use the proposed method for regression and classification, like SVM, but with the added advantage of interpretability of the covariate effects offered by the proposed method.

To demonstrate a possible framework for such an inverse regression using the proposed method, we use the gait maturation study data to estimate the age of participants via leave-one-out cross-validation (LOOCV). This involves fitting the proposed model using all but a single participant, estimating the age for the single participant left out, and then repeating this process for each of the participants. In this way, we can evaluate the methods ability to serve as a classification/regression algorithm for estimating covariate values.

4.1 Inverse Regression Procedure

Given an estimator of the covariate-dependent log power spectrum $\log f(\boldsymbol{\omega}, \nu)$, from the proposed model, and a new individual time series realization for which we wish to predict the unknown age, the algorithm begins by first estimating the log power spectrum using only the new time series realization. Many methods exist, but we use a modified Daniell smooth of periodogram ordinates such that

$$\tilde{f}(\nu_k) = \frac{1}{2m} \left[\frac{I_\ell(\nu_{k-m})}{2} + I_\ell(\nu_{k-m+1}) + \dots + I_\ell(\nu_k) + \dots + I_\ell(\nu_{k+m-1}) + \frac{I_\ell(\nu_{k+m})}{2} \right] \quad (4.1)$$

where $I_\ell(\nu_k)$ are the periodogram ordinates for the new time series realization, $N = \lfloor T/2 \rfloor - 1$, $\nu_k = k/T$ for $k = 1, \dots, N$ are the Fourier frequencies $T = 256$ is the length of time series. We use a span of $m = 10$ to encourage smoothing, but this parameter can also be selected via cross-validation from a reasonable set of choices.

Second, the age-dependent log power spectrum can be obtained by integrating out other covariates in the model of the covariate-dependent log power spectrum obtained from the RJMCMC draws from the proposed method to construct the partial dependence on age

$$\log f(\omega_{age}, \nu) = E_{\boldsymbol{\omega}_{-age}} [\log f([\omega_{age}, \boldsymbol{\omega}_{-age}], \nu)] = \int \log f([\omega_{age}, \boldsymbol{\omega}_{-age}], \nu) dP(\boldsymbol{\omega}_{-age}) \quad (4.2)$$

which gives an average value of the log power spectrum for a fixed age ω_{age} and the other covariates $\boldsymbol{\omega}_{-age}$ vary over their marginal distribution, $dP(\boldsymbol{\omega}_{-age})$ (Breiman, 2001). This can be approximated using the proposed method as

$$\log \hat{f}(\omega_{age}, \nu) = \frac{1}{L} \sum_{\ell=1}^L \log \hat{f}([\omega_{age}, \boldsymbol{\omega}_{-age, \ell}], \nu) \quad (4.3)$$

where $\boldsymbol{\omega}_{-age, \ell}$ are the covariates for the L time series used to train the model except for

age and $\log \hat{f}([\omega_{age}, \omega_{-age, \ell}], \nu)$ is the posterior mean estimator of the covariate-dependent log power spectrum from the proposed method.

Third, we compute $\log \hat{f}(\omega_{age}, \nu)$ for a reasonable range of possible ages. In what follows, we consider 20 evenly spaced values of ages between 3 and 14 years old, representing the range of ages found in the gait maturation study data. Finally, we estimate the age by finding the age that minimizes the sum of squared difference over frequencies between the individually estimated log power spectrum for the new time series and the age-dependent estimated log power spectra from the proposed method such that

$$\hat{\omega}_{age} = \arg \min_{\omega_{age}} \frac{1}{N} \sum_{k=1}^N \left[\log \hat{f}(\omega_{age}, \nu_k) - \tilde{f}(\nu_k) \right]^2. \quad (4.4)$$

Figure 4.1 shows the comparison between the estimated age using LOOCV from the proposed approach and the true age for each participant. The plot indicates generally good estimation of age with a root mean squared error of 2.408 years, and we can conclude that this is a viable approach for inverse regression using the proposed method to estimate unknown covariate values.

This method can be generalized for estimation of the other covariates such as gait speed and gender very well. If the unknown covariate of interest is categorical rather than continuous, the above framework extends easily. The only modification is with respect to the covariate values for which the estimated log power spectrum is computed. In this case, we only consider covariate values observed in the training dataset as candidates for estimation. The remaining components of the framework extend directly.

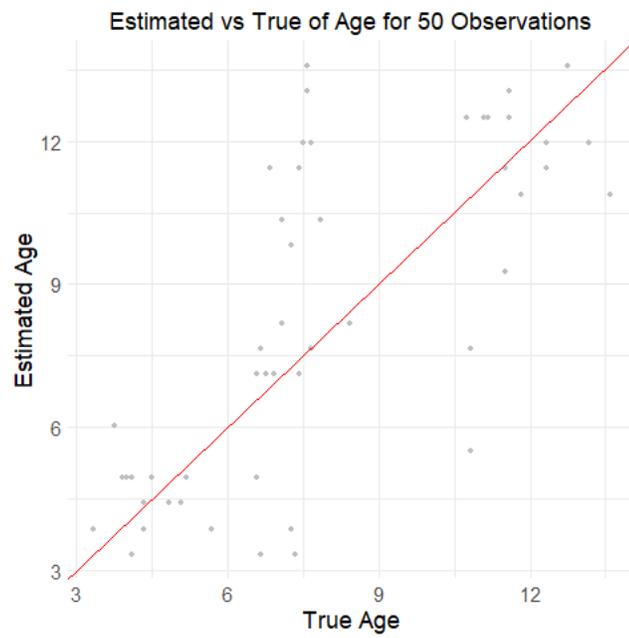


Figure 4.1: Scatter plot of estimated age and the true age of 50 participants with LOOCV algorithm. The red reference line shows the line $y = x$.

Chapter 5: Covariate-dependent Multiple Nonstationary Time Series

In Chapter 3, we introduce the methodology of the Bayesian sum of trees model for stationary time series data and provide the results of simulated data and the Gait maturation study. In this chapter, we propose a Bayesian Voronoi tessellation framework for the analysis of nonstationary time series data.

5.1 Voronoi Tessellation Modeling Structure

The Voronoi tessellation modeling framework is constructed by dividing the covariate space into different disjoint regions. Compared with the tree-based model, Voronoi tessellation can construct irregular shapes of the partition which is more flexible than the tree structure. Figure 5.1 shows the comparison between the Voronoi tessellation and the tree structure. Assuming the true partition is the Voronoi tessellation, it is efficient for the Voronoi tessellation framework to fit the structure by calculating the distance between the observations and the two centers and assigning the observations to the closest center. However, for the tree-based model, we only can repeat the procedure many times around the diagonal line to approximate the partition.

Consider modeling a collection of nonstationary time series $X_{\ell t}$ of length $t = 1, \dots, T$ and p -dimensional covariates $\boldsymbol{\omega}_\ell = \{\omega_{1\ell}, \dots, \omega_{p\ell}\}$ for $\ell = 1, \dots, L$ independent subjects, Voronoi tessellation structure is to construct a partition by dividing the time T and p -dimensional covariates space into M disjoint regions V_1, \dots, V_M with M centers $\mathbf{c}_1, \dots, \mathbf{c}_M$. Each region V_i consists of all the observed $\boldsymbol{\theta}_\ell = (\boldsymbol{\omega}_\ell, t)$ that are closest to center \mathbf{c}_i for $\ell \in L_i$

where L_i is the set of time series within V_i , such that

$$R_i = \{\boldsymbol{\theta} \in \mathbf{D} : \|\boldsymbol{\theta} - \mathbf{c}_i\| < \|\boldsymbol{\theta} - \mathbf{c}_j\| \quad i \neq j\} \quad (5.1)$$

where $\|\boldsymbol{\theta}\| = \sum_{i=1}^{p+1} w_i^2 \theta_i^2$, $\mathbf{w} = \{w_t, w_1, \dots, w_p\}$ is a normalized weighting vector such that $\sum_{k=1}^{p+1} w_k = 1$, and $\mathbf{D} = \mathbb{R}^p \times \mathcal{T}$ where \mathbb{R}^p is the p -dimensional real-valued space and $\mathcal{T} = \{1, 2, \dots, T\}$ is the finite set of time index points.

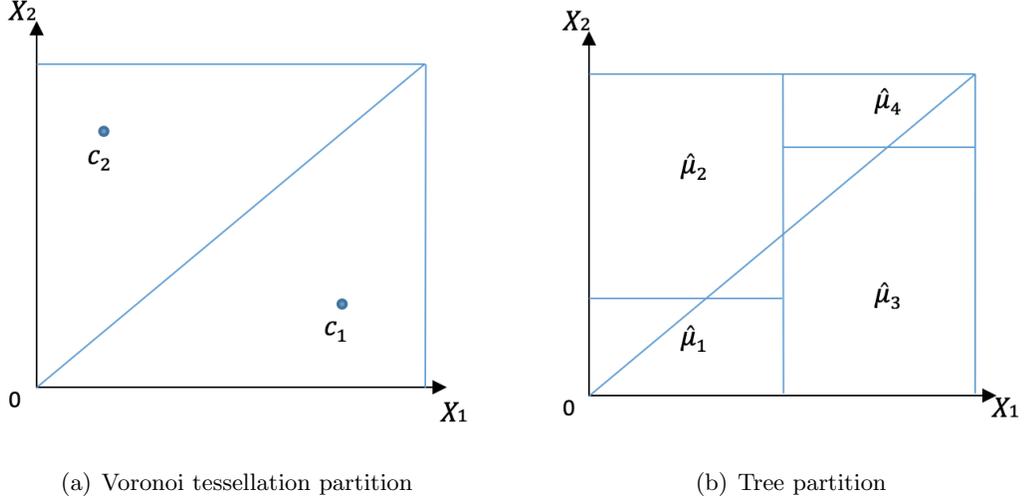


Figure 5.1: Comparison of Voronoi tessellation structure and tree-based model

For the priors of the Voronoi tessellation structure, we follow Payne et al. (2020) and assume that the possible location of the centers are selected from the observed data $\boldsymbol{\theta}$. The prior can be expressed as

$$p(\mathbf{c}, M, \mathbf{w}) = p(\mathbf{c}|M)p(M)p(\mathbf{w}) \quad (5.2)$$

where

$$p(M) = \text{DU}(M|1, \dots, M_{\max}) \quad (5.3)$$

$$p(\mathbf{c}|M) = \text{DU}\left(\mathbf{c}|1, \dots, \binom{n}{M}\right) \quad (5.4)$$

$$p(\mathbf{w}) = \text{Di}(\mathbf{w}|1, \dots, 1) \quad (5.5)$$

that $\text{DU}(x|1, \dots, n)$ is the discrete uniform on $1, \dots, n$ and M_{\max} is the maximum possible number of centers which is a fixed value. $\text{Di}(\mathbf{w}|1, \dots, 1)$ is a Dirichlet prior for vector \mathbf{w} .

5.2 Local Power Spectrum Estimation

Consider we have M centers, the local periodogram for the ℓ th time series of i th cell is

$$I_{i\ell}(\nu) = \frac{1}{T_{i\ell}} \left| \sum_{t \in \Psi_{i\ell}} x_{\ell t} \exp(-2\pi i \nu t) \right|^2 \quad (5.6)$$

$\Psi_{i\ell} = \{t \in \mathcal{T} : \|\boldsymbol{\theta}_\ell - \mathbf{c}_i\| < \|\boldsymbol{\theta}_\ell - \mathbf{c}_j\| \forall i \neq j\}$ where $\boldsymbol{\theta}_\ell = (\mathbf{w}_\ell, t)$. Note that the cardinality of the set $|\Phi_{i\ell}| = T_{i\ell}$. Assuming $T_{i\ell}$ is sufficiently large, the Whittle likelihood is then can be derived as

$$L(\mathbf{I}_1, \dots, \mathbf{I}_L | f_1, \dots, f_M) \approx$$

$$\prod_{i=1}^M \prod_{\ell=1}^L (2\pi)^{-\frac{n_{i\ell}}{2}} \prod_{k=1}^{n_{i\ell}} \exp \left\{ -\delta_{\Phi_{i\ell}} [\log f_i(\nu_{i\ell k}) + I_{i,\ell}(\nu_{i\ell k}) / f_i(\nu_{i\ell k})] \right\} \quad (5.7)$$

where $n_{i\ell} = \lfloor T_{i\ell}/2 \rfloor - 1$ is the number of Fourier frequencies for the ℓ th time series in the i th cell, $\delta_{\Phi_{i\ell}}$ is the indicator function that $\delta_{\Phi_{i\ell}} = 1$ if $\Phi_{i\ell} \neq \emptyset$ and 0 otherwise. Similar with the analysis of covariate-dependent stationary time series data, the local log power spectrum within i th cell $\log f_i(\nu)$ are modeled using a Bayesian penalized linear spline model (Rosen

et al., 2012)

$$\log f_i(\nu) \approx \alpha_i + \sum_{s=1}^S \beta_s^{(i)} \cos(2\pi s\nu), \quad (5.8)$$

where the functions $\cos(2\pi s\nu)$ are the Demmler–Reinsch basis functions. Also, following the previous stationary time series analysis, Gaussian priors are assumed such that $\alpha_i \sim N(0, \sigma_\alpha^2)$ where σ_α^2 is a constant value, and $\boldsymbol{\beta}^{(i)} = (\beta_1^{(i)}, \dots, \beta_S^{(i)})' \sim N(0, \tau_i^2 \mathbf{D}_S)$, where $\mathbf{D}_S = \text{diag}(\{\sqrt{2}\pi s\}^{-2})$. τ_i^2 is a smoothing parameter that controls the roughness of the log spectrum. The scaling for the smoothing parameter, $\{\sqrt{2}\pi s\}^{-2}$, provides regularization of the integrated squared first derivative of the log power spectrum (Li and Krafty, 2019). A half-t prior is placed on τ_i (Gelman, 2006) which is the same as specified in Chapter 3. A two-step MCMC sampling scheme for α_i , $\boldsymbol{\beta}^{(i)}$, and τ_i is presented as

1. Let \mathbf{Z}_i be a $n_{i\ell} \times S$ matrix of basis functions for ℓ th time series of i th center such that $\{Z_i\}_{ks} = \cos(2\pi s\nu_k)$. Given the value of τ_i^2 and the realization of local log periodogram $\mathbf{I}_{i\ell}(\nu)$. The posterior of α_i and $\boldsymbol{\beta}^{(i)}$ is,

$$p(\alpha_i, \boldsymbol{\beta}^{(i)} | \tau_i^2, \mathbf{I}_{i\ell}, \mathbf{Z}_i) \propto \exp \left\{ - \sum_{\ell=1}^L \sum_{k=1}^{n_{i\ell}} [\alpha_i + \mathbf{z}'_{ik} \boldsymbol{\beta}^{(i)} + \exp(\log I_{i\ell}(\nu_{i\ell k}) - \alpha_i - \mathbf{z}'_{ik} \boldsymbol{\beta}^{(i)})] - \frac{\alpha_i^2}{2\sigma_a^2} - \frac{1}{2\tau_i^2} \boldsymbol{\beta}^{(i)'} \mathbf{D}_B^{-1} \boldsymbol{\beta}^{(i)} \right\}. \quad (5.9)$$

2. For the posterior of τ_i^2 , we assume a latent variable a_i for the i th center which has the following prior distribution with τ_i^2

$$(\tau_i^2 | a_i) \sim \text{IG}\left(\frac{\nu_\tau}{2}, \frac{\nu_\tau}{a_i}\right), \quad a_i \sim \text{IG}\left(\frac{1}{2}, \frac{1}{A_\tau^2}\right), \quad (5.10)$$

then the full conditional posterior distributions are

$$(a_i | \tau_i^2) \sim \text{IG}\left(\frac{\nu_\tau + 1}{2}, \frac{\nu_\tau}{\tau_i^2} + \frac{1}{A_\tau^2}\right), \quad (5.11)$$

and

$$(\tau_i^2 | a_i, \boldsymbol{\beta}^{(i)}) \sim \text{IG}\left(\frac{\nu_\tau + S + 1}{2}, \frac{\boldsymbol{\beta}^{(i)'} \boldsymbol{\beta}^{(i)}}{2} + \frac{\nu_\tau}{a_i}\right). \quad (5.12)$$

The MCMC algorithm draws a_i first and then updates τ_i^2 .

5.3 Sampling Scheme

To sample Voronoi tessellation structure, reversible-jump MCMC is developed to jointly propose and evaluate new draws of 1) the number of centers M ; 2) the collection of centers \mathbf{c} ; 3) the vector of weights \mathbf{w} ; and 4) the parameters for Bayesian penalized linear spline model.

The proposed modification takes the form of one of four possible steps: BIRTH, DEATH, MOVE, and CHANGE, with an equal probability of 0.25 for each step. The BIRTH move grows a new center that is randomly selected from the rest of the observations except the current centers, the DEATH move deletes one of the current center randomly, the MOVE step changes the location of one of the current centers, and the CHANGE move updates one of the vector weights \mathbf{w} . New model parameters for the proposed modification are drawn, and the proposed modification is then accepted or rejected using a Metropolis–Hastings (M-H) step. By using this reversible-jump MCMC sampler, we average the posterior probability over post burn-in draws to obtain the final estimator. Technical details for the sampling scheme are available in Appendix C.

5.4 Simulation Results

We consider the following time-varying abrupt-slowly simulation setting representing abrupt and smoothly varying dynamics in order to demonstrate the performance of the Bayesian Voronoi tessellation model.

Let $\boldsymbol{\omega} = [\omega_1, \omega_2]'$ where $\omega_{1\ell} = \ell/L$ for $\ell = 1, \dots, L$, $\omega_{2\ell} = \ell/(L/2)$ for $\ell = 1, \dots, L/2$, and $\omega_{2\ell} = (\ell - L/2)/(L/2)$ for $\ell = (L/2 + 1), \dots, L$. An AR(1) process for the ℓ th time series is specified as $x_{\ell t} = \phi_{\ell} x_{\ell t-1} + \epsilon_{\ell t}$, $\epsilon_{\ell t} \sim N(0, 1)$ where

$$\phi(U) = \begin{cases} \phi_{\ell} = -0.3 + (t/T) * 0.6, & \text{if } \omega_{1\ell} \leq 0.5 \text{ and } \omega_{2\ell} \leq 0.5 \\ \phi_{\ell} = 0.3 - (t/T) * 0.6, & \text{if } \omega_{1\ell} \leq 0.5 \text{ and } \omega_{2\ell} > 0.5 \\ \phi_{\ell} = 0.7 - (t/T) * 1.4, & \text{if } \omega_{1\ell} > 0.5 \text{ and } \omega_{2\ell} \leq 0.5 \\ \phi_{\ell} = -0.7 + (t/T) * 1.4, & \text{if } \omega_{1\ell} > 0.5 \text{ and } \omega_{2\ell} > 0.5 \end{cases} \quad (5.13)$$

for $t = 1, \dots, T$, such that the simulated data changes abruptly on four different covariate regions and smoothly on time space within each region. Figure 5.2 shows the estimated power spectrum (the first row) and the true power spectrum (the second row) for $L = 20$, $T = 1000$ time series data. We start the Voronoi tessellation with $M = 80$, and set $M_{\max} = 200$. The MCMC procedure is run for a total of 10,000 iterations with the first 5,000 discarded as burn-in. In order to assess convergence, trace plots for summary measures are available in Figure 5.3. We can observe that the MCMC procedure converged after 5,000 iterations. To confirm the performance of the model, Table 5.1 presents the MSE over 20 replicates for various time length T for $L = 20$. The accuracy of the proposed model performs increases when we increase the length of the time series.

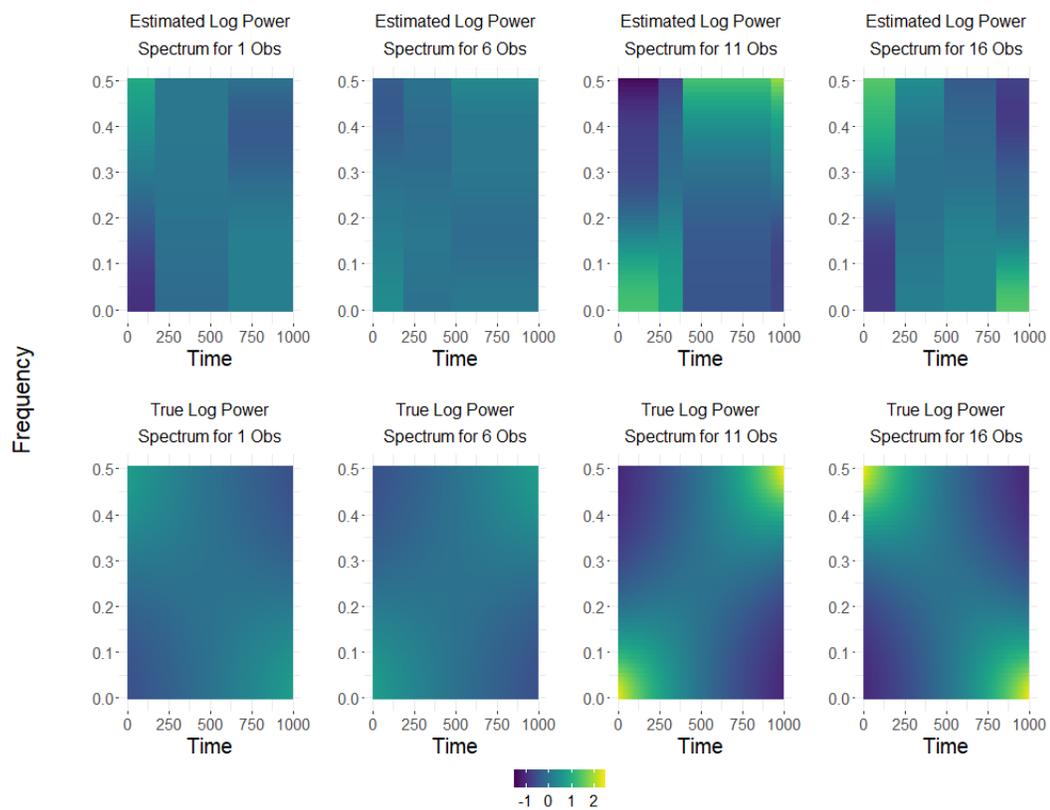


Figure 5.2: Estimated and true covariate-dependent conditional log power spectrum for one run of the time-varying sbrupt-slowly simulation setting conditional on four different covariate spaces respectively. The 1st, 6th, 11th, and 16th observations are selected from each region separately.

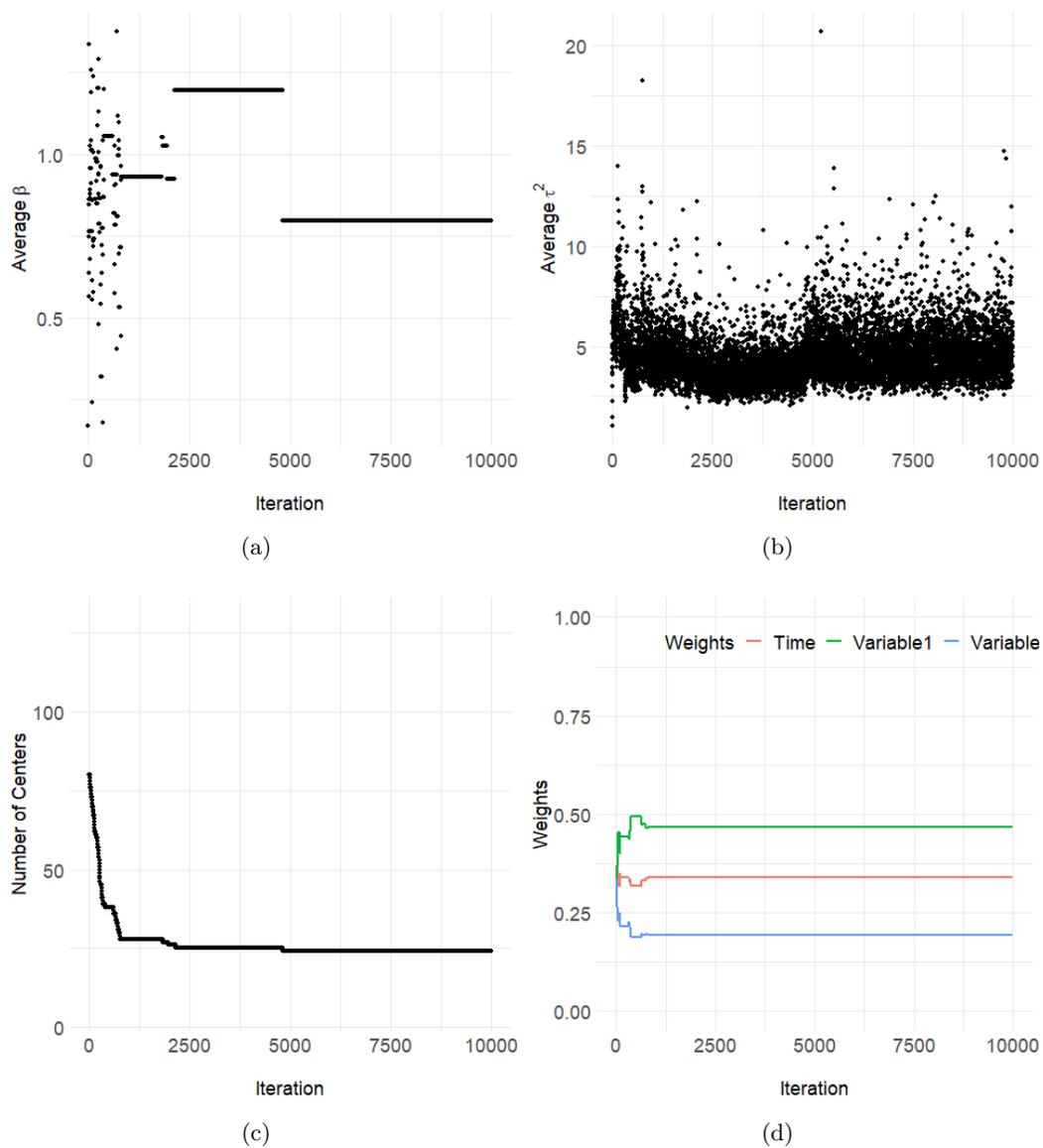


Figure 5.3: Convergence diagnostic plots for the time-varying abrupt-slowly simulation for one replication: (a) average β across time length T and all time series L ; (b) average τ^2 across time length T and all time series L ; (c) number of centers; (d) weights \mathbf{w} for time and two covariates ω_1 and ω_2 .

Table 5.1: Mean and standard deviation of MSE for time-varying abrupt-slowly simulation setting with $L = 20$ time series with different length T . Results are presented for the proposed Voronoi tessellation model starting with $M = 5$ centers.

L=20	Time-varying Abrupt-slowly
250	0.2039(0.0173)
500	0.1651(0.0244)
1000	0.1251(0.0233)

Chapter 6: Conclusion and Future Work

This paper describes two novel adaptive Bayesian covariate-dependent models for the power spectrum estimation of multiple stationary time series and nonstationary time series data. For multiple stationary time series data, we use a Bayesian sum of trees model to characterize covariate effects. This model is flexible and can automatically recover complex nonlinear associations and interactions as well as provide efficient variable selection. For multiple nonstationary time series data, we use the Voronoi tessellation structure instead to create the partition for both time and covariates, which is more flexible than the Bayesian sum of trees model and can recover the time and covariates effects very well.

This work is one of the first approaches to analyzing the power spectrum of multiple time series with multiple covariates in a completely nonparametric manner, but it is not without limitations. For the Bayesian sum of trees model, soft-decision trees (Linero and Yang, 2018b) that better adapt to smooth effects may also be considered for capturing covariate effects in an even more flexible and parsimonious manner. The current method can be extended to conduct the extra spectral variability due to clustering effects (Krafty, 2016) and missingness in time series data. For the Voronoi tessellation modeling framework, we may have the multiple modes problem since the diagnostic plots can stick to a certain value after burn-in iterations, and the performance of the model varies according to different start numbers of centers. Due to the multiple modes, the current collection of moves can't effectively reach and explore these modes due to getting stuck in local modes. To solve the problem, we can develop some new moves that can dramatically change the partition structure and allow for getting to other modes without having to traverse the area of low posterior probability separating the modes.

Appendix A: Proof of the Smoothing Parameter τ^2

The calculation of $\int_{-\frac{1}{2}}^{\frac{1}{2}} (\log' f(\nu))^2 d\nu$ is presented as below. Since

$$\log f(\nu) = \alpha + \sum_{b=1}^B \beta_b \cos(2\pi b\nu), \quad (\text{A.1})$$

we have

$$\log' f(\nu) = 0 + \sum_{b=1}^B [\beta_b (-\sin(2\pi b\nu)) 2\pi b], \quad (\text{A.2})$$

and

$$[\log' f(\nu)]^2 = \sum_{b=1}^B [4\pi^2 b^2 \beta_b^2 \sin^2(2\pi b\nu)] + 2 \sum_{i \neq j} [4\pi^2 ij \beta_i \beta_j \sin(2\pi j\nu) \sin(2\pi i\nu)]. \quad (\text{A.3})$$

So,

$$\begin{aligned} \int_{-\frac{1}{2}}^{\frac{1}{2}} (\log' f(\nu))^2 d\nu &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{b=1}^B [4\pi^2 b^2 \beta_b^2 \sin^2(2\pi b\nu)] d\nu \\ &+ 2 \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{i \neq j} [4\pi^2 ij \beta_i \beta_j \sin(2\pi i\nu) \sin(2\pi j\nu)] d\nu. \end{aligned} \quad (\text{A.4})$$

Now we first deal with the second term of the integration

Since $\sin(x) \sin(y) = \frac{1}{2} [\cos(x-y) + \cos(x+y)]$, we have $\sin(2\pi i\nu) \sin(2\pi j\nu) = \frac{1}{2} [\cos(2\pi\nu(i -$

$j)) + \cos(2\pi\nu(i + j))]$. The integration of the second term can be rewritten as

$$\begin{aligned}
2 \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{i \neq j} [4\pi^2 ij \beta_i \beta_j \sin(2\pi j\nu) \sin(2\pi i\nu)] d\nu &= 4\pi^2 ij \beta_i \beta_j \int_{-\frac{1}{2}}^{\frac{1}{2}} \cos(2\pi\nu(i - j)) \\
&\quad - \cos(2\pi\nu(i + j)) d\nu \\
&= 2\pi^2 ij \beta_i \beta_j \left[\sin[2\pi(i - j)\nu] \times \frac{1}{2\pi(i - j)} \right] \Big|_{-\frac{1}{2}}^{\frac{1}{2}} \\
&\quad - \sin[2\pi(i + j)\nu] \times \frac{1}{2\pi(i + j)} \Big|_{-\frac{1}{2}}^{\frac{1}{2}} \\
&= 0.
\end{aligned} \tag{A.5}$$

Because $i - j$ and $i + j$ are integers and $\sin(c\pi) = 0$ if c is a integer.

So

$$\begin{aligned}
\int_{-\frac{1}{2}}^{\frac{1}{2}} (\log' f(\nu))^2 d\nu &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{b=1}^B [4\pi^2 b^2 \beta_b^2 \sin^2(2\pi b\nu)] d\nu \\
&= \sum_{b=1}^B 4\pi^2 b^2 \beta_b^2 \int_{-\frac{1}{2}}^{\frac{1}{2}} \sin^2(2\pi b\nu) d\nu.
\end{aligned} \tag{A.6}$$

Since $\sin^2(x) = \frac{1}{2}[1 - \cos(2x)]$

$$\begin{aligned}
\int_{-\frac{1}{2}}^{\frac{1}{2}} \sin^2(2\pi b\nu) d\nu &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{2} - \frac{1}{2} \cos(4\pi b\nu) d\nu \\
&= \frac{1}{2} \nu \Big|_{-\frac{1}{2}}^{\frac{1}{2}} - \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \cos(4\pi b\nu) d\nu \\
&= \frac{1}{2} \nu \Big|_{-\frac{1}{2}}^{\frac{1}{2}} - \frac{1}{2} \sin(4\pi b\nu) \frac{1}{4\pi b} \Big|_{-\frac{1}{2}}^{\frac{1}{2}} \\
&= \frac{1}{2} \left(\frac{1}{2} - \left(-\frac{1}{2}\right) \right) - 0 \\
&= \frac{1}{2},
\end{aligned} \tag{A.7}$$

$\frac{1}{2} \sin(4\pi b\nu) \frac{1}{4\pi b} \Big|_{-\frac{1}{2}}^{\frac{1}{2}} = 0$ because $\sin(2\pi b) = 0$ for $b = 0, \pm 1, \pm 2, \dots$. So

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} (\log' f(\nu))^2 d\nu = \sum_{b=1}^B 4\pi^2 b^2 \beta_b^2 \times \frac{1}{2}. \tag{A.8}$$

Let $Z_1, \dots, Z_b \sim N(0, 1)$, where $b = 1, \dots, B$. So $\beta_b = \frac{\tau}{\sqrt{2\pi b}} Z_b$

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} (\log' f(\nu))^2 d\nu = \sum_{b=1}^B 2\pi^2 b^2 \frac{\tau^2}{(\sqrt{2\pi b})^2} Z_b^2 \sim \tau^2 \chi_B^2. \tag{A.9}$$

Hence, it can be concluded that the distribution of $\int_{-\frac{1}{2}}^{\frac{1}{2}} (\log' f(\nu))^2 d\nu$ is controlled by the smoothing parameter τ^2 .

Appendix B: Sampling Scheme Details for Bayesian Sum of Trees Model

The sampling scheme for the proposed Bayesian sum of trees model for the covariate-dependent power spectrum is presented in this section. Suppose we have M trees for the sum of trees model. Let (U_j, Φ_j) and (U_j^*, Φ_j^*) be the current and proposed tree structure and terminal node parameter estimates respectively for the j th tree, and let \mathbf{R}_j denote the residuals of the log periodogram ordinates from the fit corresponding to the sum of all trees except the j th tree for all time series. More details about the types of proposals developed in this work are provided herein. For ease of exposition, the subscript j is dropped in what follows. We implement the reversible jump Markov chain Monte Carlo (MCMC) sampling scheme by using a Metropolis-Hastings algorithm in which the acceptance ratio α is formulated as

$$\alpha = \min\{1, A\},$$

where

$$A = \frac{p(U^*, \Phi^* | \mathbf{R}) \times q(U, \Phi | U^*, \Phi^*)}{p(U, \Phi | \mathbf{R}) \times q(U^*, \Phi^* | U, \Phi)}.$$

More details on the individual components of the acceptance ratio introduced above are provided below.

i. Distribution of $p(U, \Phi|\mathbf{R})$ and $p(U^*, \Phi^*|\mathbf{R})$

The joint posterior distribution $p(U, \Phi|\mathbf{R})$ can be expressed as a product of the following terms

$$\begin{aligned} p(U, \Phi|\mathbf{R}) &= p(\mathbf{R}|U, \Phi) \times p(\Phi|U) \times p(U) \\ &= p(\mathbf{R}|U, \Phi) \times p(\boldsymbol{\beta}, \tau^2|U) \times p(U) \\ &= \underbrace{p(\mathbf{R}|U, \Phi)}_{\text{likelihood}} \times \underbrace{p(\boldsymbol{\beta}|U, \tau^2) \times p(\tau^2|U)}_{\text{prior}} \times p(U), \end{aligned}$$

where the prior of Φ is determined by the joint prior of $(\boldsymbol{\beta}, \tau^2)$. Here, $\boldsymbol{\beta}$ represents the intercept and basis coefficients in Equation (1) of the manuscript. Suppose there are n_b observations in the b th terminal node of j th tree for $b = 1, \dots, B$, the likelihood can be expressed as the product of individual Whittle likelihoods

$$\begin{aligned} p(\mathbf{R}|U, \Phi) &= \prod_{b=1}^B p(\mathbf{R}_{b1}, \dots, \mathbf{R}_{bn_b} | f_{jb}) \approx \\ &\prod_{b=1}^B \prod_{i=1}^{n_b} (2\pi)^{-n/2} \prod_{k=1}^n \exp\{\log f_{jb}(\nu_k) + \exp(\mathbf{R}_{bi}(\nu_k)) / f_{jb}(\nu_k)\}, \quad (\text{B.1}) \end{aligned}$$

where $n = \lfloor T/2 \rfloor - 1$, $\nu_k = k/T$ for $k = 1, \dots, n$ are the Fourier frequencies, T is the length of time series, and \mathbf{R}_{bi} is the residual of the log periodogram ordinates for the i th time series belonging to the b th terminal node.

The prior for $\boldsymbol{\beta}$ is a normal distribution such that $p(\boldsymbol{\beta}|U, \tau^2) \sim N(0, (\sigma_\alpha^2, \tau^2 \mathbf{D}_S))$ as described in Section 3.2 of the manuscript. The prior distribution for τ^2 , $p(\tau^2|U)$, is a half-t distribution. We follow Wand et al. (2011) and express the half-t distribution as a

scale mixture of inverse gamma distributions with latent variable a such that

$$p(\tau^2|a) \sim \text{IG}\left(\frac{\xi_\tau}{2}, \frac{\xi_\tau}{a}\right), \quad p(a) \sim \text{IG}\left(\frac{1}{2}, \frac{1}{A_\tau^2}\right).$$

The prior for the tree structure $p(U)$ is

$$p(U) = \prod_{\eta \in \text{terminals}} [1 - p_{\text{split}}(\eta)] \prod_{\eta \in \text{internals}} p_{\text{split}}(\eta) \prod_{\eta \in \text{internals}} p_{\text{rule}}(\eta),$$

where η is the node of the current tree. The $p_{\text{split}}(\eta) = \gamma(1+d)^{-\theta}$ is the probability of node η to be split into two child nodes with $\gamma \in (0, 1)$, $\theta \in [0, \infty)$ and $d = 0, 1, \dots$ which is the depth of the given node η . The $p_{\text{rule}}(\eta) = \frac{1}{n_{\text{adj}}(\eta)} \times \frac{1}{n_{\text{cutpoint}}(\eta)}$ is the probability of the available variables and the cutpoints to be chosen for node η , where $n_{\text{adj}}(\eta)$ denotes the number of predictors available for the node and $n_{\text{cutpoint}}(\eta)$ is the number of available cutpoints for the selected variable. The posterior distribution $p(U^*, \Phi^* | \mathbf{R})$ is similar with $p(U, \Phi | \mathbf{R})$ by plugging in U^* and Φ^* instead.

ii. Distribution of $q(U^*, \Phi^* | U, \Phi)$ and $q(U, \Phi | U^*, \Phi^*)$

The proposed density $q(U^*, \Phi^* | U, \Phi)$ is defined as

$$\begin{aligned} q(U^*, \Phi^* | U, \Phi) &= q(\Phi^* | U^*, U, \Phi) \times q(U^* | U, \Phi) \\ &= q(\beta^*, \tau^{2*} | U^*, U, \Phi) \times q(U^* | U) \\ &= q(\tau^{2*} | U^*, U, \Phi) \times q(\beta^* | \tau^{2*}, U^*, U, \Phi) \times q(U^* | U), \end{aligned}$$

and similarly, the density $q(U, \Phi|U^*, \Phi^*)$ is

$$\begin{aligned}
q(U, \Phi|U^*, \Phi^*) &= q(\Phi|U^*, \Phi^*, U) \times q(U|U^*, \Phi^*) \\
&= q(\beta, \tau^2|U^*, \Phi^*, U) \times q(U|U^*) \\
&= q(\tau^2|U^*, \Phi^*, U) \times q(\beta|\tau^2, U^*, \Phi^*, U) \times q(U|U^*).
\end{aligned}$$

From part **i** and part **ii**, A can be written as

$$\begin{aligned}
&\underbrace{\frac{p(\mathbf{R}|U^*, \Phi^*)}{p(\mathbf{R}|U, \Phi)}}_{\text{likelihood ratio}} \times \underbrace{\frac{p(\beta^*|U^*, \tau^{2*})p(\tau^{2*}|U^*)}{p(\beta|U, \tau^2)p(\tau^2|U)}}_{\text{prior ratio}} \times \underbrace{\frac{p(U^*)}{p(U)}}_{\text{tree structure ratio}} \times \underbrace{\frac{q(\beta, \tau^2|U^*, \Phi^*, U)}{q(\beta^*, \tau^{2*}|U, \Phi, U^*)}}_{\text{proposed probability ratio}} \times \underbrace{\frac{q(U|U^*)}{q(U^*|U)}}_{\text{transition ratio}}. \\
&\hspace{20em} \text{(B.2)}
\end{aligned}$$

Proposed modifications to the tree structures can take the form of one of three possible moves: BIRTH, DEATH and CHANGE. The BIRTH move grows the tree by splitting a terminal node into two child nodes, the DEATH move prunes the tree by dropping two terminal child nodes belonging to the same internal node, and the CHANGE move modifies the variable and cut point associated with an internal node with two terminal child nodes. Noticing that the prior ratio is the same for all BIRTH, DEATH, and CHANGE moves, the other ratios will be described individually for each of the three types of moves.

B.1 BIRTH

For the BIRTH move, the b th terminal node of the j th tree is selected to be split into two new child nodes. The proposed tree differs from the original tree only through the change from this terminal node to new child nodes. Therefore, the likelihood ratio becomes

$$\frac{p(\mathbf{R}|U^*, \Phi^*)}{p(\mathbf{R}|U, \Phi)} = \frac{p(\mathbf{R}_{b_1}, \dots, \mathbf{R}_{b_{l_i}}|f_{j b_i})p(\mathbf{R}_{b_r}, \dots, \mathbf{R}_{b_{n_r}}|f_{j b_r})}{p(\mathbf{R}_{b_1}, \dots, \mathbf{R}_{b_{n_b}}|f_{j b})},$$

where b_l denotes left child node, b_r is for the right child node, n_l is the number of series corresponding to the proposed left child node, n_r is the number of series corresponding to the proposed right child node, $\mathbf{R}_{b_l,i}$ are residuals of the log periodogram ordinates for the i th series corresponding to the proposed left child node, and $\mathbf{R}_{b_r,i}$ defined similarly for the proposed right child node. The likelihood within the terminal node is as shown in Equation (B.1). The tree structure ratio is expressed as

$$\frac{p(U^*)}{p(U)} = \frac{(1 - p_{\text{split}}(b_l))(1 - p_{\text{split}}(b_r))p_{\text{split}}(b)p_{\text{rule}}(b)}{1 - p_{\text{split}}(b)}.$$

For the transition probability $q(U|U^*)$ and $q(U^*|U)$, they can be expressed as

$$\begin{aligned} q(U^*|U) &= p(\text{GROW}) \times p(\text{selecting the } b\text{th terminal node to grow from}) \\ &\quad \times p(\text{selecting the } q\text{th predictor to split on}) \\ &\quad \times p(\text{selecting the } w\text{th value to split on}) \\ &= p(\text{GROW}) \frac{1}{B} \frac{1}{n_{\text{adj}}(b)} \times \frac{1}{n_{\text{cutpoint}}(b)}, \end{aligned}$$

and

$$\begin{aligned} q(U|U^*) &= p(\text{PRUNE}) \times p(\text{selecting node } \eta \text{ to prune from}) \\ &= p(\text{PRUNE}) \frac{1}{n_{\text{internal}^*}}, \end{aligned}$$

where $p(\text{GROW}) = 0.25$ and $p(\text{PRUNE}) = 0.25$ are the probability of BIRTH and DEATH move to be selected, and n_{internal^*} is the total number of internal nodes that has two terminal child nodes. Thus, we can derive the transition ratio to be

$$\frac{q(U|U^*)}{p(U^*|U)} = \frac{p(\text{PRUNE})Bn_{\text{adj}}(b)n_{\text{cutpoint}}(b)}{q(\text{GROW})n_{\text{internal}^*}}.$$

Then, we follow Rosen et al. (2012) to draw the proposed terminal parameters β^* , τ^{2*} and latent variable a^* . For $\tau^2(a)$, let $\tau_b^2(a_b)$ denote the current $\tau^2(a)$ for the b th terminal node and a uniform distribution is used to generate new parameters $\tau_{b_l}^{2*}(a_{b_l}^*)$ and $\tau_{b_r}^{2*}(a_{b_r}^*)$ for the left and right child. Specifically,

$$\begin{aligned}\tau_{b_l}^{2*} &= \tau_b^2 \times \frac{u_\tau}{1 - u_\tau}, & \tau_{b_r}^{2*} &= \tau_b^2 \times \frac{1 - u_\tau}{u_\tau} \\ a_{b_l}^* &= a_b \times \frac{u_a}{1 - u_a}, & a_{b_r}^* &= a_b \times \frac{1 - u_a}{u_a}\end{aligned}$$

where $u_\tau, u_a \sim U[0, 1]$. For β , an approximated normal distribution is proposed to generate the new parameters $\beta_{b_l}^*$ and $\beta_{b_r}^*$. Specifically, $(\beta_{b^*}^* | \tau_{b^*}^{2*}, U^*, U, \Phi) \sim N(\beta_b^{\max}, \Sigma_b^{\max})$, where

$$\beta_b^{\max} = \operatorname{argmax}_{\beta_{b^*}^*} p(\beta_{b^*}^* | \mathbf{R}_b, \tau_{b^*}^{2*}, U^*), \quad (\text{B.3})$$

and

$$\Sigma_b^{\max} = \left\{ -(\partial^2 \log p(\beta_{b^*}^* | \mathbf{R}_b, \tau_{b^*}^{2*}, U^*)) / (\partial \beta_{b^*}^* \partial \beta_{b^*}^{\prime}) |_{\beta_{b^*}^* = \beta_b^{\max}} \right\}^{-1}, \quad (\text{B.4})$$

where $p(\beta_{b^*}^* | \mathbf{R}_b, \tau_{b^*}^{2*}, U^*)$ is presented in Section 3.2, and b^* represents the left or right child of node b . We then have the proposed probability ratio

$$\begin{aligned}\frac{q(\beta, \tau^2 | U^*, \Phi^*, U)}{q(\beta^*, \tau^{2*} | U, \Phi, U^*)} &= \frac{q(\beta | \tau^2, U^*, \Phi^*, U) q(\tau^2 | U^*, \Phi^*, U) q(a | U^*, \Phi^*, U)}{q(\beta^* | \tau^{2*}, U, \Phi, U^*) q(\tau^{2*} | U^*, \Phi^*, U) q(a^* | U^*, \Phi^*, U)} \\ &= \frac{q(\beta_b)}{q(\beta_{b_l}^*) q(\beta_{b_r}^*) p(u)} \times \left| \frac{\partial(\tau_{b_l}^{2*}, \tau_{b_r}^{2*})}{\partial(\tau_b^2, u_\tau)} \right| \times \left| \frac{\partial(a_{b_l}^*, a_{b_r}^*)}{\partial(a_b, u_a)} \right|,\end{aligned}$$

where $p(u_\tau), p(u_a) = 1$, $q(\beta_b)$, $q(\beta_{b_l}^*)$ and $q(\beta_{b_r}^*)$ are the densities of the approximately normal distribution $N(\beta_b^{\max}, \Sigma_b^{\max})$ with the corresponding current and proposed τ^2 values.

The Jacobian of τ^2 and a are

$$\left| \frac{\partial(\tau_{b_l}^{2*}, \tau_{b_r}^{2*})}{\partial(\tau_b^2, u_\tau)} \right| = \frac{2\tau_b^2}{u_\tau(1-u_\tau)}, \quad \left| \frac{\partial(a_{b_l}^*, a_{b_r}^*)}{\partial(a_b, u_a)} \right| = \frac{2a_b}{u_a(1-u_a)}.$$

B.2 DEATH

The DEATH move is the inverse of the BIRTH move. Suppose b is the selected internal to be pruned by deleting the two child nodes b_l and b_r , the likelihood ratio for the proposed tree and the current tree is hence expressed as

$$\frac{p(\mathbf{R}|U^*, \Phi^*)}{p(\mathbf{R}|U, \Phi)} = \frac{p(\mathbf{R}_{b1}, \dots, \mathbf{R}_{bn_b})|f_{jb}}{p(\mathbf{R}_{b_l,1}, \dots, \mathbf{R}_{b_l,n_l}|f_{jb_l})p(\mathbf{R}_{b_r,1}, \dots, \mathbf{R}_{b_r,n_r}|f_{jb_r})},$$

which is the change from the two child nodes to the internal b node. The ratio of tree structure is presented as

$$\frac{p(U^*)}{p(U)} = \frac{1 - p_{\text{split}}(b)}{(1 - p_{\text{split}}(b_l))(1 - p_{\text{split}}(b_r))p_{\text{split}}(b)p_{\text{rule}}(b)},$$

and the ratio of transition

$$\frac{p(U|U^*)}{p(U^*|U)} = \frac{q(\text{GROW})n_{\text{internal}^*}}{q(\text{PRUNE})(B-1)n_{\text{adj}}(b)n_{\text{cutpoint}}(b)},$$

where $B-1$ is the number of terminal nodes for the proposed prune tree. For the proposed probability ratio, we draw terminal node parameters τ_b^{2*} and a_b^* by taking the inverse of the corresponding BIRTH move

$$\tau_b^{2*} = \sqrt{\tau_{b_l}^2 \tau_{b_r}^2}, \quad a_b^* = \sqrt{a_{b_l} a_{b_r}}.$$

The vector β_b^* is drawn from the approximately normal distribution $N(\beta_b^{\max}, \Sigma_b^{\max})$ given the proposed τ^{2*} value. Hence, the proposed probability ratio is

$$\frac{q(\beta, \tau^2 | U^*, \Phi^*, U)}{q(\beta^*, \tau^{2*} | U, \Phi, U^*)} = \frac{q(\beta_{b_l})q(\beta_{b_r})p(u)}{q(\beta_b^*)} \times \left| \frac{\partial(\tau_b^{2*}, u_\tau)}{\partial(\tau_{b_l}^2, \tau_{b_r}^2)} \right| \times \left| \frac{\partial(a_b^*, u_a)}{\partial(a_{b_l}, a_{b_r})} \right|,$$

where the Jacobian of τ^2 and a are

$$\left| \frac{\partial(\tau_b^{2*}, u_\tau)}{\partial(\tau_{b_l}^2, \tau_{b_r}^2)} \right| = \frac{u_\tau(1-u_\tau)}{2\tau_b^{2*}} = 2(\tau_{b_l} + \tau_{b_r})^2,$$

$$\left| \frac{\partial(a_b^*, u_a)}{\partial(a_{b_l}, a_{b_r})} \right| = \frac{u_a(1-u_a)}{2a_b^*} = 2(\sqrt{a_{b_l}} + \sqrt{a_{b_r}})^2.$$

B.3 CHANGE

A CHANGE move is to change two terminal nodes to a pair of new child nodes by changing the split rule of their parent node. The observations within each terminal node of the proposed tree can be different from the current tree. Thus, the likelihood ratio is

$$\frac{p(\mathbf{R} | U^*, \Phi^*)}{p(\mathbf{R} | U, \Phi)} = \frac{p(\mathbf{R}_{l,1}, \dots, \mathbf{R}_{l,n^{l*}} | f_{j_{b_l^*}})p(\mathbf{R}_{r,1}, \dots, \mathbf{R}_{r,n^{r*}} | f_{j_{b_r^*}})}{p(\mathbf{R}_{l,1}, \dots, \mathbf{R}_{l,n^l} | f_{j_{b_l}})p(\mathbf{R}_{r,1}, \dots, \mathbf{R}_{r,n^r} | f_{j_{b_r}})},$$

where n^{l*} and n^{r*} are the number of observations in the new left and the new right child nodes. The tree structure ratio for the CHANGE move is

$$\frac{p(U^*)}{p(U)} = \frac{(1 - p_{\text{split}}(b_{l^*}))(1 - p_{\text{split}}(b_{r^*}))p_{\text{split}}(b^*)p_{\text{rule}}(b^*)}{(1 - p_{\text{split}}(b_l))(1 - p_{\text{split}}(b_r))p_{\text{split}}(b)p_{\text{rule}}(b)},$$

since the depth of the children nodes does not change, the split probability for each node stays the same. The only change is the number of the available cutpoints of the new variable,

which can be different from the current variable. Therefore, the tree structure ratio is

$$\frac{p(U^*)}{p(U)} = \frac{n_{\text{cutpoint}}(b)}{n_{\text{cutpoint}}(b^*)},$$

where $n_{\text{cutpoint}}(b^*)$ is the number of available cutpoints for the proposed variable. The transition probability from current tree to the proposed tree is

$$\begin{aligned} p(U^*|U) &= p(\text{CHANGE}) \times p(\text{selecting node } b \text{ to change}) \\ &\quad \times p(\text{selecting the new predictor to split on}) \\ &\quad \times p(\text{selecting the new cutpoint to split on}). \end{aligned}$$

The calculation of the $p(U|U^*)$ is similar with $p(U^*|U)$ except that the number of available cutpoints can be different. So the transition ratio is

$$\frac{p(U|U^*)}{p(U^*|U)} = \frac{n_{\text{cutpoint}}(b^*)}{n_{\text{cutpoint}}(b)}.$$

From the results above, the transition ratio and the tree structure ratio are cancelled in the representation of Equation (C.1). So the acceptance ratio for the CHANGE step is only related to the likelihood, prior probability and the proposed probability of terminal node parameters. Two-step Gibbs sampling is used to draw new terminal node parameters. The $\beta_{b^*}^*$ is drawn from the normal approximation $N(\beta_b^{\max}, \Sigma_b^{\max})$ given the current $\tau_{b^*}^2$, and a_{b^*} and $\tau_{b^*}^2$ are draw from their full conditional distributions. The MCMC algorithm draws a_{b^*} first and then updates $\tau_{b^*}^2$.

Appendix C: Sampling Scheme Details for Voronoi Tessellation

The sampling scheme for the proposed Bayesian modeling with Voronoi tessellation structure for the covariate-dependent nonstationary time series analysis is presented in this section. Suppose we have M centers for the Voronoi tessellation structure. Let $\Phi = \{\beta, \tau^2\}$ be the collection of all the parameters for the Bayesian spline model. Here, β represents the intercept and basis coefficients in Equation 5.8 of the manuscript. Let $U = \{M, \mathbf{c}, \mathbf{w}\}$ be the collection of parameters for Voronoi tessellation. We define that (U, Φ) and $((U^*, \Phi^*))$ are the current and proposed Voronoi tessellation structure and parameter estimates respectively, and $\log I$ denotes the log periodogram ordinates for all time series. More details about the types of proposals developed in this work are provided herein. We implement the reversible jump Markov chain Monte Carlo (MCMC) sampling scheme by using a Metropolis-Hastings algorithm in which the acceptance ratio α is formulated as

$$\alpha = \min\{1, A\},$$

where

$$A = \frac{p(U^*, \Phi^* | \log I) \times q(U, \Phi | U^*, \Phi^*)}{p(U, \Phi | \log I) \times q(U^*, \Phi^* | U, \Phi)}.$$

More details on the individual components of the acceptance ratio introduced above are provided below.

i. Distribution of $p(U, \Phi | \log I)$ and $p(U^*, \Phi^* | \log I)$

The joint posterior distribution $p(U, \Phi | \log I)$ can be expressed as a product of the following terms

$$\begin{aligned}
 p(U, \Phi | \log I) &= p(\log I | U, \Phi) \times p(\Phi | U) \times p(U) \\
 &= p(\log I | U, \Phi) \times p(\boldsymbol{\beta}, \tau^2 | U) \times p(U) \\
 &= \underbrace{p(\log I | U, \Phi)}_{\text{likelihood}} \times \underbrace{p(\boldsymbol{\beta} | U, \tau^2) \times p(\tau^2 | U)}_{\text{prior}} \times p(U),
 \end{aligned}$$

where the prior of Φ is determined by the joint prior of $(\boldsymbol{\beta}, \tau^2)$. The likelihood can be expressed as the product of individual Whittle likelihoods is shown in Equation 5.7.

The prior for the Voronoi tessellation $p(U)$ is

$$p(U) = p(\mathbf{c} | M) p(M) p(\boldsymbol{\omega})$$

The prior of the $p(\mathbf{c} | M)$, $p(M)$, and the $p(\boldsymbol{\omega})$ are the same as described in Section 5.1. The distribution of $p(U^*, \Phi^* | \log I)$ is the same as $p(U, \Phi | \log I)$.

ii. Distribution of $q(U^*, \Phi^* | U, \Phi)$ and $q(U, \Phi | U^*, \Phi^*)$

The proposed density $q(U^*, \Phi^* | U, \Phi)$ is defined as

$$\begin{aligned}
 q(U^*, \Phi^* | U, \Phi) &= q(\Phi^* | U^*, U, \Phi) \times q(U^* | U, \Phi) \\
 &= q(\boldsymbol{\beta}^*, \tau^{2*} | U^*, U, \Phi) \times q(U^* | U) \\
 &= q(\tau^{2*} | U^*, U, \Phi) \times q(\boldsymbol{\beta}^* | \tau^{2*}, U^*, U, \Phi) \times q(U^* | U),
 \end{aligned}$$

and similarly, the density $q(U, \Phi|U^*, \Phi^*)$ is

$$\begin{aligned}
q(U, \Phi|U^*, \Phi^*) &= q(\Phi|U^*, \Phi^*, U) \times q(U|U^*, \Phi^*) \\
&= q(\boldsymbol{\beta}, \tau^2|U^*, \Phi^*, U) \times q(U|U^*) \\
&= q(\tau^2|U^*, \Phi^*, U) \times q(\boldsymbol{\beta}|\tau^2, U^*, \Phi^*, U) \times q(U|U^*).
\end{aligned}$$

From part **i** and part **ii**, A can be written as

$$\begin{aligned}
&\underbrace{\frac{p(\log I|U^*, \Phi^*)}{p(\log I|U, \Phi)}}_{\text{likelihood ratio}} \times \underbrace{\frac{p(\boldsymbol{\beta}^*|U^*, \tau^{2*})p(\tau^{2*}|U^*)}{p(\boldsymbol{\beta}|U, \tau^2)p(\tau^2|U)}}_{\text{prior ratio}} \times \underbrace{\frac{p(U^*)}{p(U)}}_{\text{tessellation prior ratio}} \times \underbrace{\frac{q(\boldsymbol{\beta}, \tau^2|U^*, \Phi^*, U)}{q(\boldsymbol{\beta}^*, \tau^{2*}|U, \Phi, U^*)}}_{\text{proposed probability ratio}} \times \underbrace{\frac{q(U|U^*)}{q(U^*|U)}}_{\text{transition ratio}}.
\end{aligned} \tag{C.1}$$

The likelihood ratio is expressed as

$$\frac{p(\log I|U^*, \Phi^*)}{p(\log I|U, \Phi)} = \frac{L(\mathbf{I}_1, \dots, \mathbf{I}_L|f_1, \dots, f_{M^*})}{L(\mathbf{I}_1, \dots, \mathbf{I}_L|f_1, \dots, f_M)} \tag{C.2}$$

The calculation of prior ratio follows that the prior for $\boldsymbol{\beta}$ is a normal distribution such that $p(\boldsymbol{\beta}|U, \tau^2) \sim N(0, (\sigma_a^2, \tau^2 \mathbf{D}_S))$ as described in Section 2.1.1 of the manuscript. The prior distribution for τ^2 , $p(\tau^2|U)$, is a half-t distribution. We follow Wand et al. (2011) and express the half-t distribution as a scale mixture of inverse gamma distributions with latent variable a such that

$$p(\tau^2|a) \sim \text{IG}\left(\frac{\xi_\tau}{2}, \frac{\xi_\tau}{a}\right), \quad p(a) \sim \text{IG}\left(\frac{1}{2}, \frac{1}{A_\tau^2}\right).$$

The Voronoi tessellation structure ratio is

$$\frac{p(U^*)}{p(U)} = \frac{p(\mathbf{c}^*|M^*)}{p(\mathbf{c}|M)} \quad (\text{C.3})$$

since the prior probability of $p(M)$ and $p(\mathbf{w})$ are 1 as described in Section 5.1.

The proposed probability ratio $\frac{q(\boldsymbol{\beta}, \tau^2|U^*, \Phi^*, U)}{q(\boldsymbol{\beta}^*, \tau^{2*}|U, \Phi, U^*)}$ and the transition ratio $\frac{q(U|U^*)}{q(U^*|U)}$ are different for the four proposed BIRTH, DEATH, MOVE, and CHANGE steps which will be described individually in the following sections.

C.1 BIRTH

For the BIRTH step, we randomly select one of the rest of observations as a new center c_{new} . For drawing of $\tau^{2\text{new}}(a^{\text{new}})$ for the new center, we select the most closest center c_c to the new center c_{new} with the $\tau^2(a)$ denoted as $\tau^{2c}(a^c)$. We assume $\tau^{2\text{new}}(a^{\text{new}}) = \tau^{2c}(a^c)$, and the proposed τ^2 is a combination of current and new τ^2 such that $\tau^{2*}(a^*) = \{\tau^2(a), \tau^{2\text{new}}(a^{\text{new}})\}$. Then, an approximated normal distribution is proposed to generate the new parameters $\boldsymbol{\beta}^*$ (Rosen et al., 2012). Specifically, $(\boldsymbol{\beta}^*|\tau^{2*}, U^*, U, \Phi) \sim N(\boldsymbol{\beta}^{\text{max}}, \boldsymbol{\Sigma}^{\text{max}})$, where

$$\boldsymbol{\beta}^{\text{max}} = \operatorname{argmax}_{\boldsymbol{\beta}^*} p(\boldsymbol{\beta}^*|\log I, \tau^{2*}, U^*), \quad (\text{C.4})$$

and

$$\boldsymbol{\Sigma}^{\text{max}} = \left\{ -(\partial^2 \log p(\boldsymbol{\beta}^*|\log I, \tau^{2*}, U^*)) / (\partial \boldsymbol{\beta}^* \partial \boldsymbol{\beta}^{*'}) \Big|_{\boldsymbol{\beta}^* = \boldsymbol{\beta}^{\text{max}}} \right\}^{-1}, \quad (\text{C.5})$$

where $p(\boldsymbol{\beta}^*|\log I, \tau^{2*}, U^*)$ is presented in Section 2.1.1.

We then have the proposed probability ratio

$$\begin{aligned} \frac{q(\boldsymbol{\beta}, \tau^2 | U^*, \Phi^*, U)}{q(\boldsymbol{\beta}^*, \tau^{2*} | U, \Phi, U^*)} &= \frac{q(\boldsymbol{\beta} | \tau^2, U^*, \Phi^*, U)}{q(\boldsymbol{\beta}^* | \tau^{2*}, U, \Phi, U^*)} \\ &= \frac{q(\boldsymbol{\beta})}{q(\boldsymbol{\beta}^*)}, \end{aligned}$$

where $q(\boldsymbol{\beta}^*)$ are the densities of the approximately normal distribution $N(\boldsymbol{\beta}^{\max}, \boldsymbol{\Sigma}^{\max})$ with the corresponding current and proposed τ^2 values.

For the transition ratio $\frac{q(U|U^*)}{q(U^*|U)}$, for most non-boundary cases, $q(U|U^*)$ and $q(U^*|U)$ can be cancelled. When $M = 1$ and we propose a birth, the ratio is $\frac{3}{4}$, and when $M = M_{\max} - 1$, the ratio is $\frac{4}{3}$.

C.2 DEATH

For the DEATH step, we randomly delete one of the current centers which is denoted as c_{delete} . We correspondingly remove the $\tau_{\text{delete}}^2(a_{\text{delete}})$ from the current $\tau^2(a)$ as the proposed $\tau^{2*} a^*$. An approximated normal distribution is proposed to generate the new parameters $\boldsymbol{\beta}^*$ (Rosen et al., 2012) which is exactly the same as in the BIRTH step.

For the transition ratio $\frac{q(U|U^*)}{q(U^*|U)}$, for most non-boundary cases, $q(U|U^*)$ and $q(U^*|U)$ can be cancelled. When $M = 2$ and we propose a death step, the ratio is $\frac{4}{3}$, and when $M = M_{\max}$. the ratio is $\frac{3}{4}$.

C.3 MOVE

For the MOVE step, we randomly delete one of the current centers and remove the corresponding $\tau(a)$. Then we select a new center from the rest of observations denoted as c_{new} . For drawing of $\tau^{2\text{new}}(a^{\text{new}})$ for the new center, the same as BIRTH step, we select the most closest center c_c to the new center c_{new} with the $\tau^2(a)$ denoted as $\tau^{2c}(a^c)$. We assume

$\tau^{2\text{new}}(a^{\text{new}}) = \tau^{2c}(a^c)$. An approximated normal distribution is proposed to generate the new parameters β^* (Rosen et al., 2012) which is exactly the same as in the BIRTH step.

The transition ratio $\frac{q(U|U^*)}{q(U^*|U)}$ is 1 in this situation.

C.4 CHANGE

For the CHANGE step, we randomly select one of the weights denoted as w_{select} and update the proposed value by $w_{\text{select}}^* = w_{\text{select}} + u$ where $u \sim N(0, \sigma)$, and σ is a predefined fixed value. By default, we assume $\sigma = 0.1$. Then, we normalized all the weights to guarantee $\sum_{i=1}^{p+1} w^* = 1$. For the proposed $\tau^*(a^*)$, let $\tau^*(a^*) = \tau(a)$. Then, we update the parameters β^* from the approximated normal distribution (Rosen et al., 2012).

Since the Voronoi tessellation structure does not changed in this step, the transition ratio $\frac{q(U|U^*)}{q(U^*|U)}$ only depends on $\frac{q(\mathbf{w}|\mathbf{w}^*)}{q(\mathbf{w}^*|\mathbf{w})}$, where

$$q(\mathbf{w}^*|\mathbf{w}) = q(u) \tag{C.6}$$

and

$$q(\mathbf{w}|\mathbf{w}^*) = q(v) \tag{C.7}$$

that $v = ((w_{\text{select}} + u)/(1 + u) - w_{\text{select}})/(w_{\text{select}} - 1)$.

Appendix D: Additional Simulation and Application Results for Bayesian Sum of Trees Model

Additional results for the simulations and gait maturation analysis along with convergence diagnostics are shown in the following section. The ordering of the sections follows that of the manuscript in Sections 4 and 5.

D.1 Additional Results: Estimation Accuracy

In this section, set $M = 5$, $L = 100$, and $T = 250$ and use 10,000 total iterations with the first 5,000 discarded as burn-in. For the AR-Friedman simulation setting, where covariates influence the power spectrum in more complicated ways, we randomly selected eight observations for illustration, and Figure D.1 shows the true and estimated log power spectra for selected observations. For the Adjusted-AdaptSPEC-X simulation, two observations are randomly selected within each of the four regions shown in Figure D.2. The corresponding estimation results for the eight observations are shown in Figure D.3. We observe that the proposed model performs very well in capturing the true behavior of the power spectrum in both simulation settings.

D.2 Additional Results: Sparse Covariate Effects

To visualize the effect on variable selection of the sparsity-inducing Dirichlet hyperprior on the splitting proportions of the regression tree prior, effects on high-dimensional data as described in Section 3.2.5, posterior probabilities of model inclusion under the uniform and Dirichlet priors for important variables and noise variables from a single run of the AR-Friedman setting with $L = 500$, $T = 250$, and $M = 50$ is presented in Figure D.4.

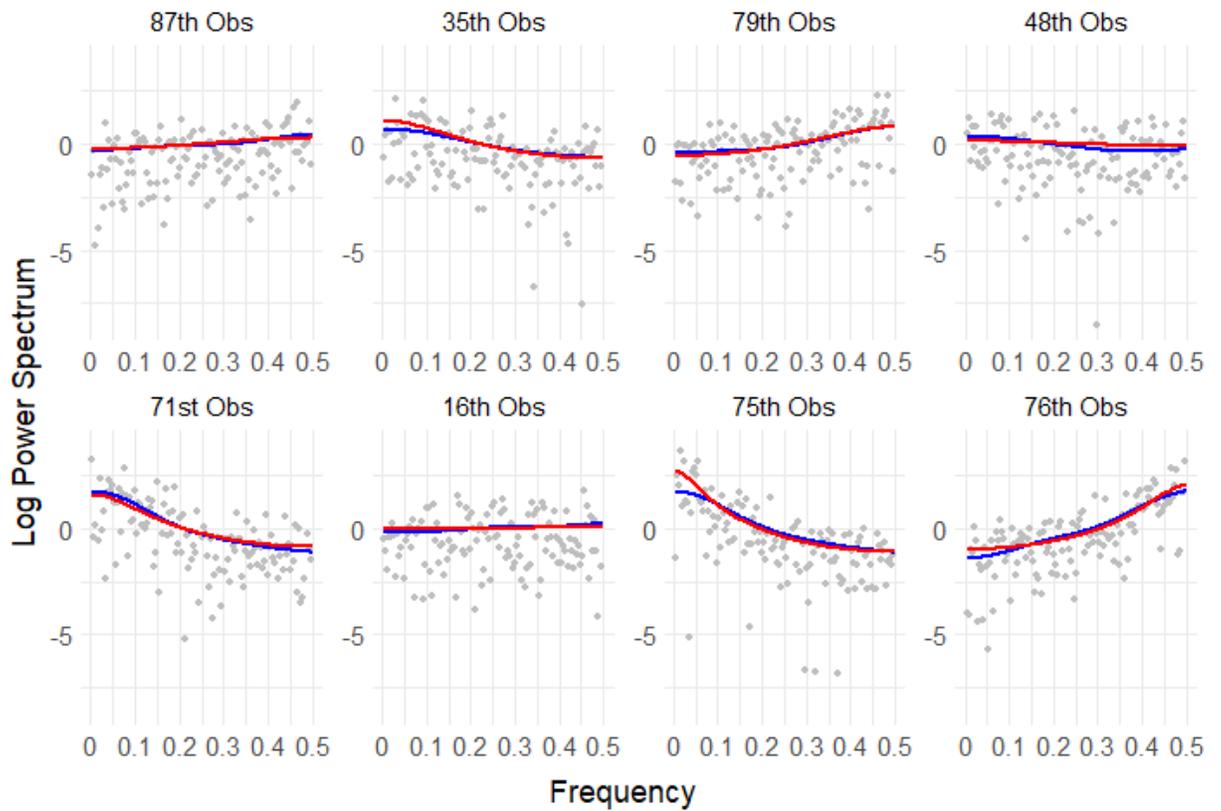


Figure D.1: The estimated (blue line) and the true (red line) log power spectrum of the eight randomly selected time series for the AR-Friedman simulation. The corresponding log periodogram ordinates are shown with gray dots.

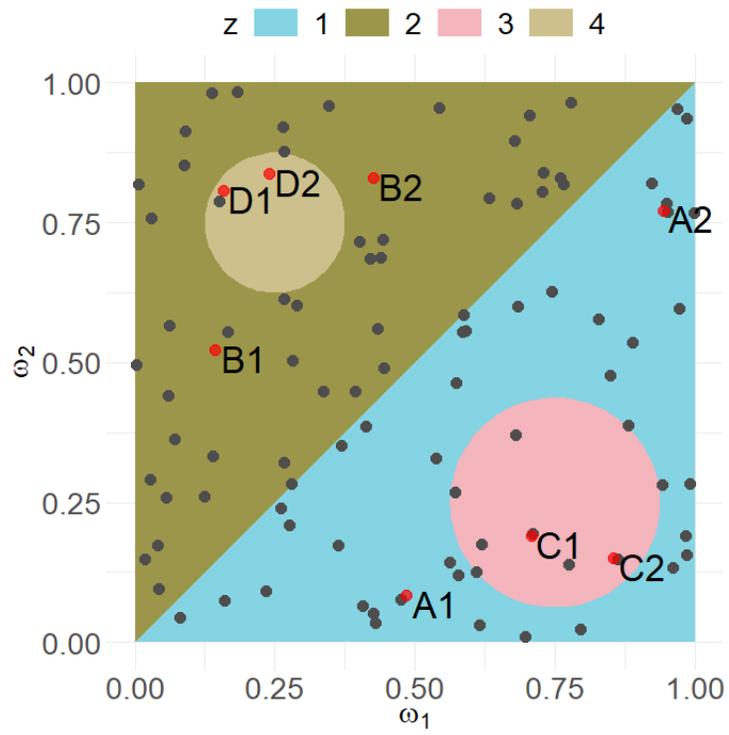


Figure D.2: The mapping of the covariates ω_1 and ω_2 to the latent variable z for the Adjusted-AdaptSPEC-X simulation setting with eight specific time series (red dots).

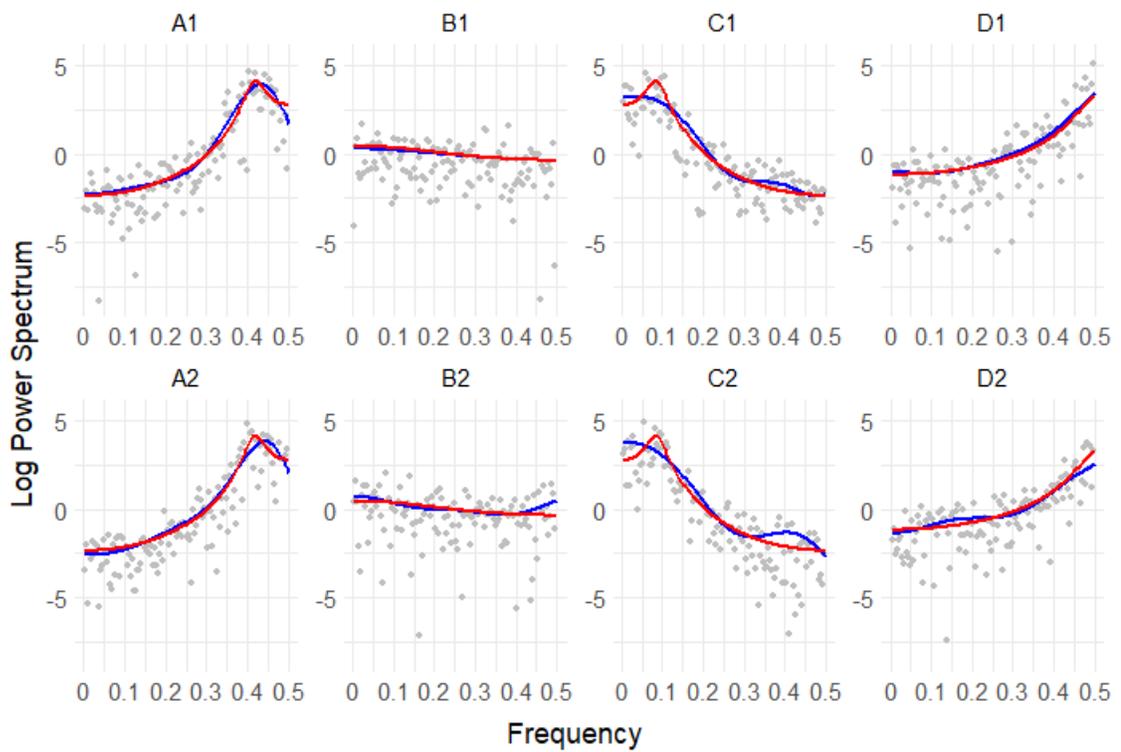


Figure D.3: The estimated (blue line) and the true (red line) log power spectrum of the eight time series denoted in Figure D.2. The corresponding log periodogram ordinates are shown with gray dots.

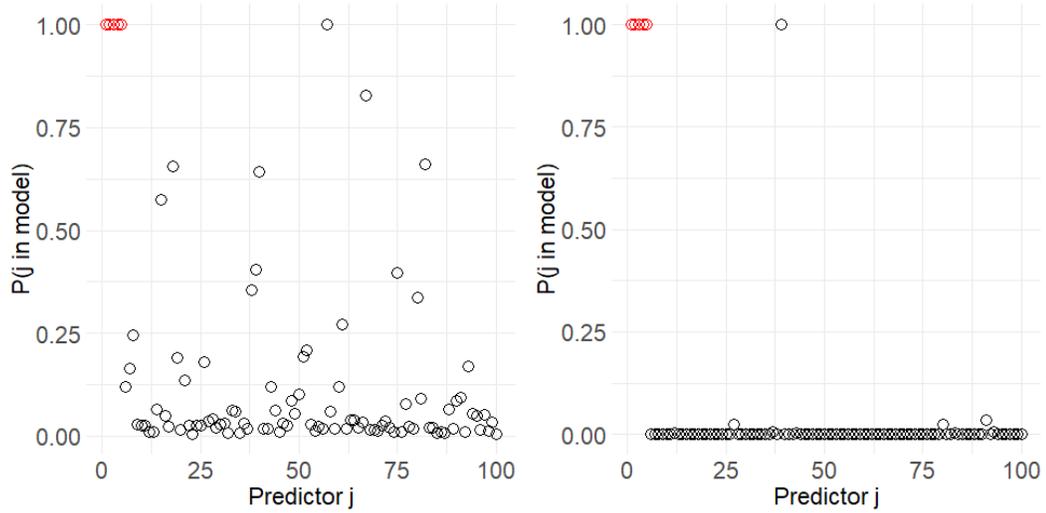


Figure D.4: Posterior probabilities of model inclusion under the uniform (left) and Dirichlet (right) priors for important variables (red) and noise variables (black).

Both forms of the prior correctly estimate the posterior probabilities of model inclusion to be 1 for all five important variables, but the Dirichlet prior accurately achieves a sparser solution, with the posterior probability of inclusion for noise variables being much closer to 0. For sparse, high-dimensional covariates, the Dirichlet prior provides more accurate selection of important covariates.

Tables D.1 - D.3 displays the MSE and posterior probabilities of model inclusion for all three settings from Section 3.2.1 of the manuscript using both the uniform and Dirichlet hyperpriors for tree splitting proportions and using $M = 5$ and $M = 50$ trees. For the Abrupt+Smooth and Adjusted-AdaptSPEC-X settings, the hyperparameter for the prior on tree depth is increased to $\theta = 25$ to encourage shallower trees, since these settings are relatively sparser with only two important variables compared to the AR-Friedman setting with five important variables.

Table D.1: Mean and standard deviation of MSE and posterior probability of model inclusion for Abrupt+Smooth setting with $L = 500$ time series of length $T = 250$ over 100 replications with different number of covariates (P) and hyperpriors (Uniform and Dirichlet). Posterior probabilities for model inclusion are reported for important variables, ω_1, ω_2 , individually and noise variables $\omega_3, \dots, \omega_P$ in aggregate. Results are presented for the proposed Bayesian sum of trees model using $M = 5$ and $M = 50$ trees.

	M=5		M=50	
P=100	Uniform	Dirichlet	Uniform	Dirichlet
MSE	0.0930(0.0220)	0.0645(0.0261)	0.0308(0.0111)	0.0192(0.0060)
Noise	0.4465(0.4547)	0.2903(0.4482)	0.4023(0.4148)	0.1220(0.3086)
ω_1	1.0000(0.0000)	0.9900(0.1000)	1.0000(0.0000)	1.0000(0.0000)
ω_2	1.0000(0.0000)	1.0000(0.0000)	1.0000(0.0000)	1.0000(0.0000)
P=200	Uniform	Dirichlet	Uniform	Dirichlet
MSE	0.1228(0.0174)	0.0836(0.0314)	0.0470(0.0141)	0.0258(0.0154)
Noise	0.3484(0.4427)	0.2235(0.4088)	0.3806(0.4353)	0.1319(0.3175)
ω_1	0.9632(0.1825)	0.8800(0.3266)	0.9998(0.0018)	0.9800(0.1407)
ω_2	0.9825(0.1116)	1.0000(0.0000)	1.0000(0.0000)	1.0000(0.0000)
P=1000	Uniform	Dirichlet	Uniform	Dirichlet
MSE	0.1425(0.0105)	0.1383(0.0212)	0.0648(0.0056)	0.0535(0.0199)
Noise	0.0984(0.2799)	0.0740(0.2603)	0.1413(0.3331)	0.0995(0.2926)
ω_1	0.3521(0.4562)	0.2003(0.4019)	0.5769(0.4766)	0.4615(0.5009)
ω_2	0.7383(0.4206)	0.5900(0.4943)	1.0000(0.0000)	1.0000(0.0000)

Table D.2: Mean and standard deviation of MSE and posterior probability of model inclusion for AR-Friedman setting with $L = 500$ time series of length $T = 250$ over 100 replications with different number of covariates (P) and hyperpriors (Uniform and Dirichlet). Posterior probabilities for model inclusion are reported for important variables, $\omega_1, \dots, \omega_5$, individually and noise variables $\omega_6, \dots, \omega_P$ in aggregate. Results are presented for the proposed Bayesian sum of trees model using $M = 5$ and $M = 50$ trees

	M=5		M=50	
	Uniform	Dirichlet	Uniform	Dirichlet
P=100				
MSE	0.0407(0.0060)	0.0353(0.0067)	0.0162(0.0019)	0.0141(0.0016)
Noise	0.7267(0.3861)	0.5544(0.4897)	0.2681(0.3457)	0.1287(0.3284)
ω_1	1.0000(0.0000)	0.9900(0.1000)	1.0000(0.0000)	1.0000(0.0000)
ω_2	1.0000(0.0000)	0.9800(0.1407)	1.0000(0.0000)	1.0000(0.0000)
ω_3	0.9812(0.1090)	0.9100(0.2876)	1.0000(0.0000)	1.0000(0.0000)
ω_4	1.0000(0.0000)	1.0000(0.0000)	1.0000(0.0000)	1.0000(0.0000)
ω_5	0.9056(0.2589)	0.8903(0.2972)	1.0000(0.0000)	1.0000(0.0000)
P=200				
MSE	0.0451(0.0042)	0.0416(0.0059)	0.0188(0.0030)	0.0143(0.0017)
Noise	0.5807(0.4338)	0.4386(0.4906)	0.2098(0.3224)	0.0787(0.2631)
ω_1	0.9667(0.1491)	0.9170(0.2734)	1.0000(0.0000)	1.0000(0.0000)
ω_2	0.9668(0.1606)	0.9100(0.2876)	1.0000(0.0000)	1.0000(0.0000)
ω_3	0.7878(0.3579)	0.6400(0.4824)	1.0000(0.0000)	1.0000(0.0000)
ω_4	1.0000(0.0000)	1.0000(0.0000)	1.0000(0.0000)	1.0000(0.0000)
ω_5	0.7503(0.3985)	0.6538(0.4755)	0.9999(0.0002)	0.9942(0.0612)
P=1000				
MSE	0.0501(0.0029)	0.0480(0.0042)	0.0285(0.0060)	0.0221(0.0069)
Noise	0.1949(0.3578)	0.1411(0.3459)	0.1259(0.2908)	0.0660(0.2319)
ω_1	0.4736(0.4820)	0.4403(0.4963)	0.9584(0.1692)	0.8859(0.3158)
ω_2	0.3894(0.4499)	0.3800(0.4878)	0.9441(0.2013)	0.8349(0.3713)
ω_3	0.2577(0.4206)	0.2400(0.4292)	0.6782(0.4358)	0.5005(0.4796)
ω_4	0.7976(0.3816)	0.7300(0.4462)	1.0000(0.0000)	1.0000(0.0000)
ω_5	0.2988(0.4386)	0.2652(0.4393)	0.6312(0.4583)	0.4690(0.4859)

Table D.3: Mean and standard deviation of MSE and posterior probability of model inclusion for Adjusted-AdaptSPEC-X setting with $L = 500$ time series of length $T = 250$ over 100 replications with different number of covariates (P) and hyperpriors (Uniform and Dirichlet). Posterior probabilities for model inclusion are reported for important variables, ω_1, ω_2 , individually and noise variables $\omega_3, \dots, \omega_P$ in aggregate. Results are presented for the proposed Bayesian sum of trees model using $M = 5$ and $M = 50$ trees

	M=5		M=50	
P=100	Uniform	Dirichlet	Uniform	Dirichlet
MSE	0.2864(0.0236)	0.2921(0.0273)	0.1501(0.0081)	0.1487(0.0080)
Noise	0.9153(0.2237)	0.8222(0.3792)	0.9526(0.1791)	0.9237(0.2637)
ω_1	1.0000(0.0000)	0.9900(0.1000)	1.0000(0.0000)	1.0000(0.0000)
ω_2	1.0000(0.0000)	1.0000(0.0000)	1.0000(0.0000)	1.0000(0.0000)
P=200	Uniform	Dirichlet	Uniform	Dirichlet
MSE	0.2983(0.0245)	0.2986(0.0280)	0.1511(0.0075)	0.1493(0.0087)
Noise	0.7174(0.3865)	0.6008(0.4872)	0.7853(0.3637)	0.7196(0.4471)
ω_1	0.9908(0.0918)	0.9400(0.2387)	1.0000(0.0000)	1.0000(0.0000)
ω_2	0.9924(0.0553)	0.9400(0.2387)	1.0000(0.0000)	1.0000(0.0000)
P=1000	Uniform	Dirichlet	Uniform	Dirichlet
MSE	0.3125(0.0267)	0.3138(0.0257)	0.1525(0.0080)	0.1535(0.0077)
Noise	0.2313(0.3792)	0.1728(0.3765)	0.2689(0.4103)	0.2285(0.4186)
ω_1	0.5449(0.4807)	0.4300(0.4976)	0.9528(0.2091)	0.9216(0.2702)
ω_2	0.5434(0.4674)	0.4300(0.4976)	0.9259(0.2559)	0.9118(0.2850)

As in the results in Table 3.3 in the manuscript, there is a slight improvement in estimation accuracy when using the Dirichlet hyperprior, as it better regularizes noise variables without losing the ability to recover important variables, and both estimation accuracy and variable selection are negatively impacted by increasing covariate vector size P . Additionally, using more trees results in improvements in MSE, increases in posterior probabilities of model inclusion for important variables, and generally reduced posterior probabilities of model inclusion for noise variables.

D.3 Additional Results: Gait Maturation Analysis

D.3.1 Visualization of Tree Structures

Visualization of covariate effects is one of the major advantages of tree-based models. Complex covariate effects can be translated into a collection of simple rules that translate transparently into actions. For the proposed method, this type of visualization is a bit more difficult due to the sum-of-trees structure of the model and the multiple draws of the tree structures from the posterior distribution using the RJMCMC sampler. However, for a small number of trees, we can visualize a single draw from the posterior by selecting the single draw that maximizes the posterior distribution, known as the maximum a posteriori (MAP) estimate. We demonstrate such a visualization for the gait maturation analysis for two participants (see Figures D.5 and D.6).

From these figures, you can see the tree structures and how they work together to produce the final estimate of the covariate-dependent log power spectrum. For example, the first participant (Figure D.5) is a male age 3.3 years with gait speed of 1.04 m/sec. This participant's male gender, young age, and lower gait speed result in components with relatively higher low frequency power coming from trees 1, 2, and 3. Components from trees 4 and 5 are relatively flat, thus increasing total variability in the power spectrum across all frequencies. Figure D.7 displays the sum of these components for the first participant, which carries relatively more low frequency power compared to high frequency power.

Alternatively, the 50th participant (Figure D.6) is a male age 13.6 years with gait speed of 1.22 m/sec. Compared to the first participant, this participant's older age results in a flatter component of the log spectrum with less overall power from tree 3. This results in less power over all frequencies compared to the younger first participant seen in the sum of these components as the final estimator of the log power spectrum in Figure D.7

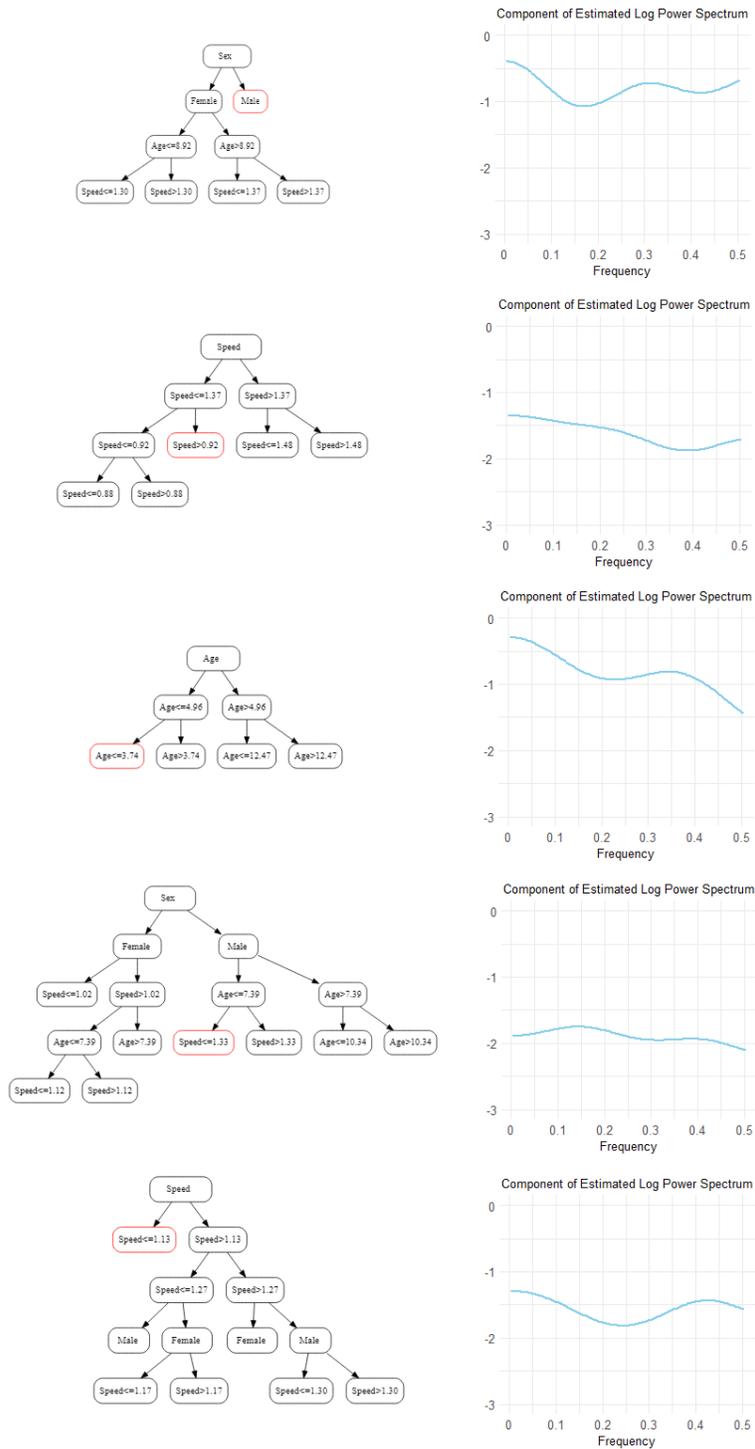


Figure D.5: Sum of trees visualization for 1st participant in the gait maturation study (age 3.3 years, gait speed 1.04 m/sec, male). The left column is an illustration of the sum of trees structure. Terminal nodes outlined in red correspond to the 1st participant. The right column shows the estimated component of log power spectrum for the 1st participant from each tree.

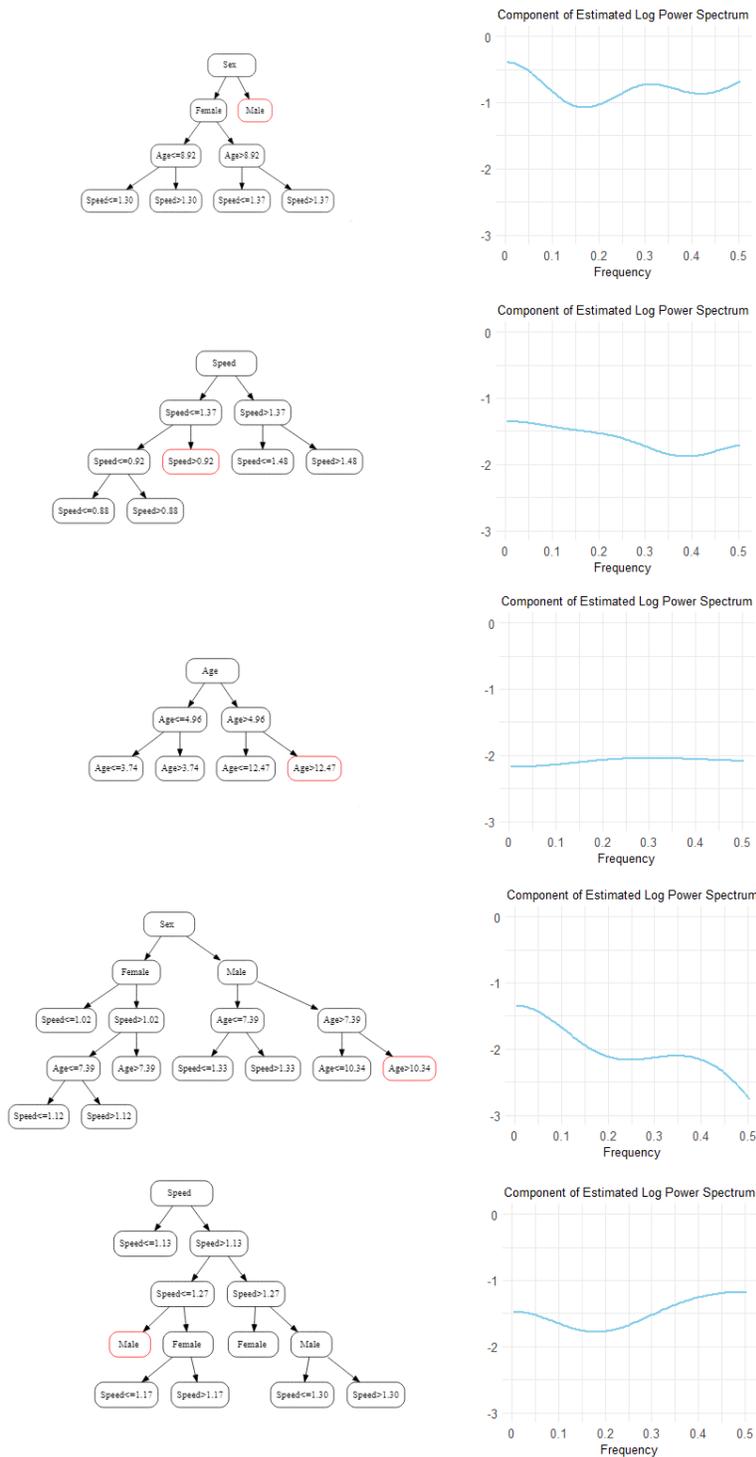


Figure D.6: Sum of trees visualization for 50th participant in the gait maturation study (age 13.6 years, gait speed 1.22 m/sec, male). The left column is an illustration of the sum of trees structure. Terminal nodes outlined in red correspond to the 50th participant. The right column shows the estimated component of log power spectrum for the 50th participant from each tree.

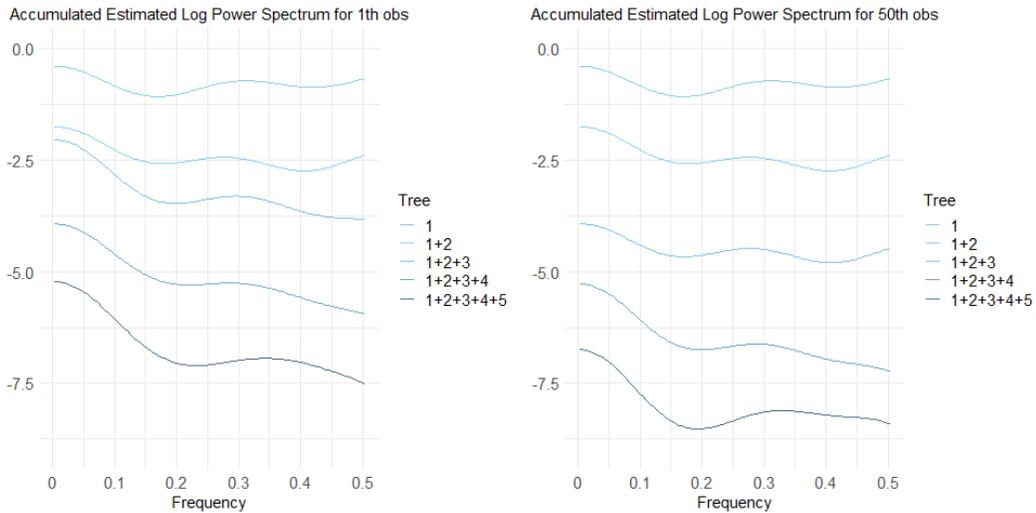


Figure D.7: The final log power spectrum estimates for the 1st (age 3.3 years, gait speed 1.04 m/sec, male) and 50th (age 13.6 years, gait speed 1.22 m/sec, male) participants summing over the components from five trees from Figures D.5 and Figures D.6 respectively. Each line corresponds to the addition of another tree component in the sum to see how the components come together to produce the final estimate (dark blue).

D.3.2 Credible Intervals for ALE of Power Spectrum

95% pointwise credible intervals for (a) and (b) in Figure 3.5 of the manuscript are provided here (see Figure D.8), along with a similar ALE plot for gender (see Figure D.9). These plots help explore covariate effects on power spectra and make inferences about which covariates significantly influence power spectra and which particular frequencies are significantly influenced. For example, age has a significantly non-zero ALE for participants under 4 years of age and above 8 years of age for low frequencies (< 0.1 stride⁻¹). This indicates that participants under 4 years of age have significantly more low frequency power (and higher LF/HF ratio) than other participants, and participants above 8 years of age have significantly less low frequency power (and lower LF/HF ratio) than other participants. For gait

speed, we observe significantly less low frequency power (and lower LF/HF ratio) among faster walkers with speeds above 1.1 m/sec. Gender appears to have a much less effect on power spectra, with only a very small frequency range from 0.05 stride⁻¹ to 0.15 stride⁻¹ significantly different in power between males and females. There are no significant differences in LF/HF ratio between males and females with a posterior mean ALE of -0.1680 and 95% credible interval (-0.3887,0.0551).

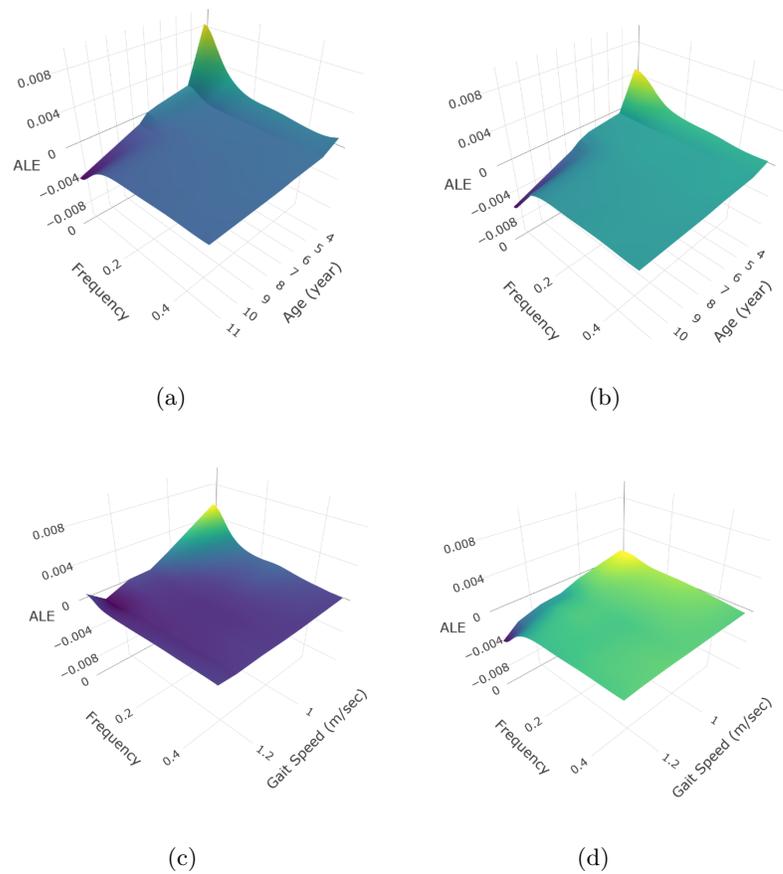


Figure D.8: 95% pointwise credible intervals of ALE for age (a,b) and gait speed (c,d) effects on the power spectrum.

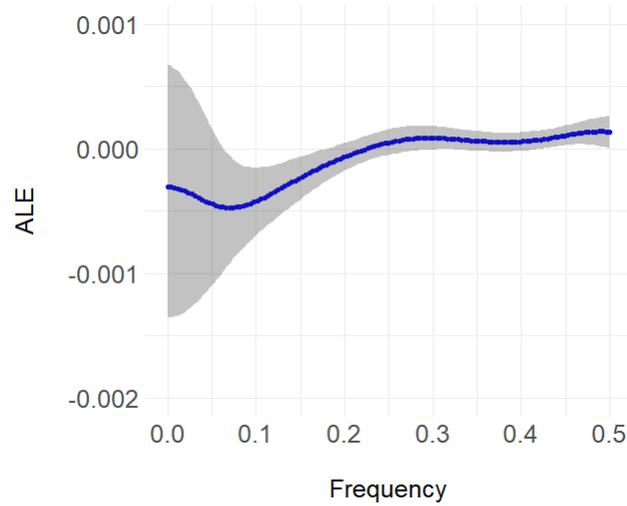


Figure D.9: Posterior mean (blue) and 95% pointwise credible intervals (shaded gray regions) of ALE for Gender.

D.4 Convergence Diagnostics

Convergence diagnostic plots for all three simulation settings using $M = 5$ trees are shown in Figures D.10-D.12, which appear to converge after 5000 burn-in iterations. Convergence diagnostic plots for all three simulation settings using $M = 50$ trees are shown in Figures

D.13-D.15, which also appear to converge after 5000 burn-in iterations. Convergence diagnostics for the gait maturation data analysis are shown in Figure D.16 using $M = 5$ trees. While the sampler appears to converge slower than the simulation settings, the sampler still appears to converge after 5,000 burn-in iterations.

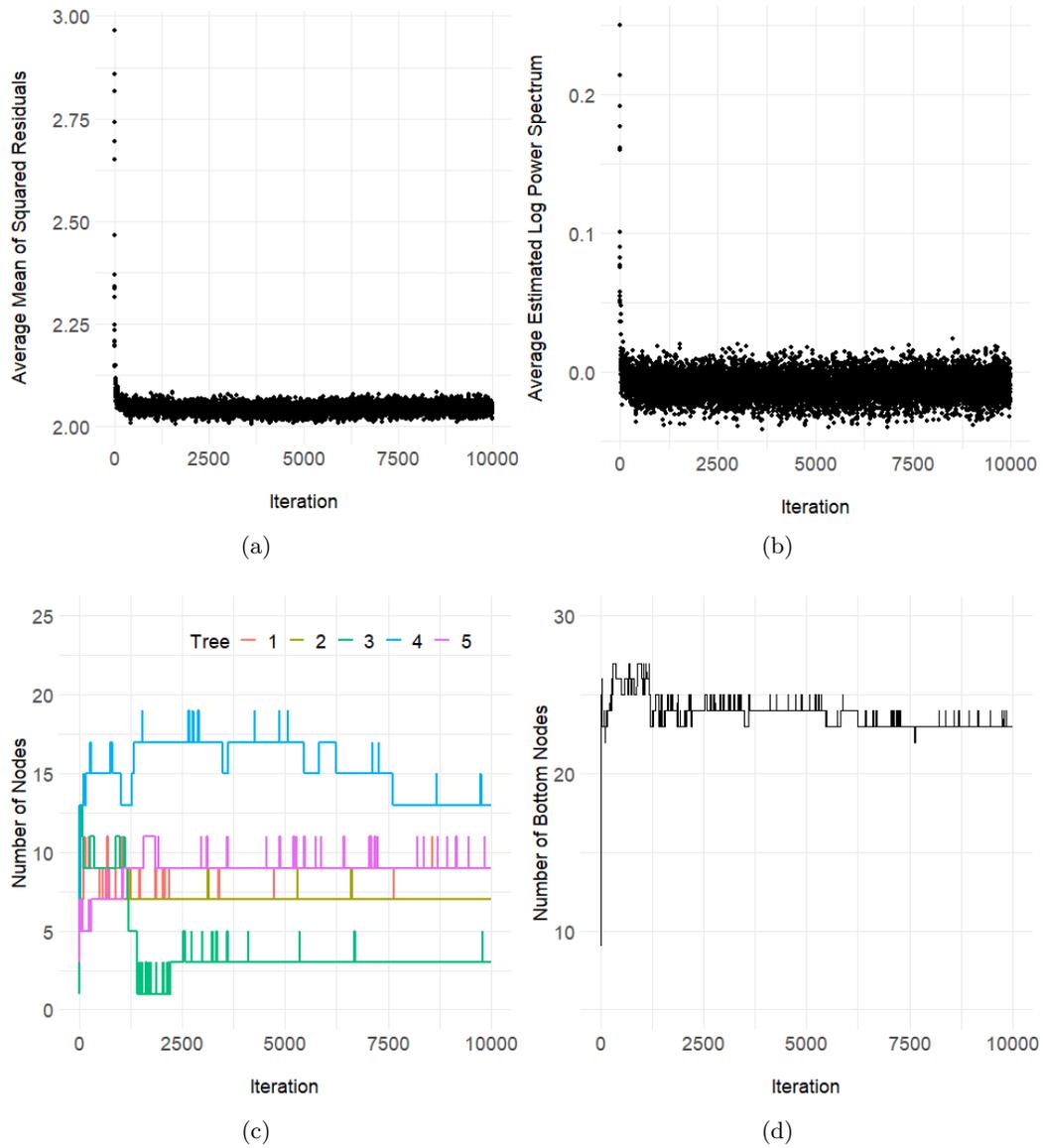
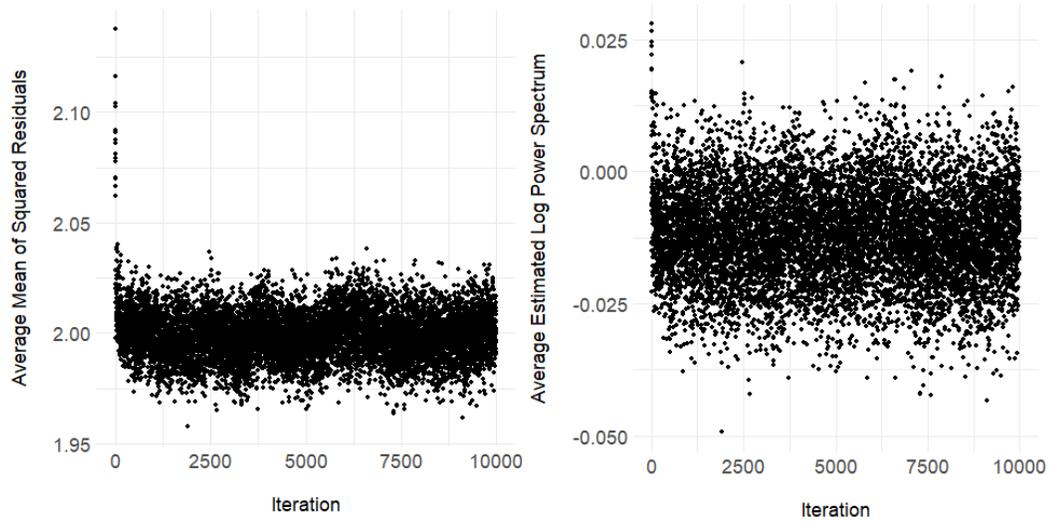
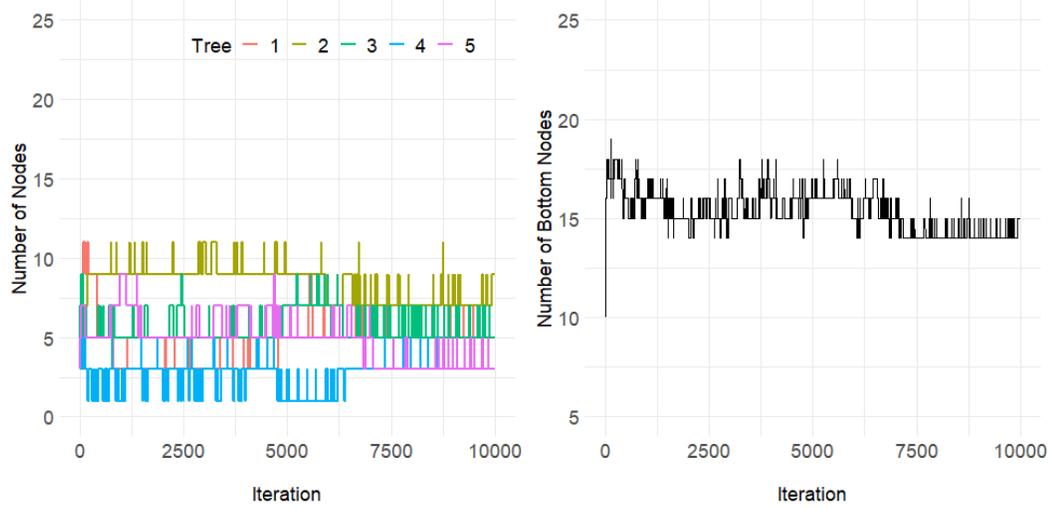


Figure D.10: Convergence diagnostic plots for the Abrupt+Smooth simulation for one replication: (a) average mean squared residuals across frequencies and all time series; (b) average estimated log power spectrum across frequencies and all time series; (c) total number of nodes for each of the five trees; (d) total number of bottom nodes across all five trees.



(a) Average mean of squared residuals (b) Average of the estimated log spectrum



(c) Number of nodes of each tree (d) Number of bottom nodes across all trees

Figure D.11: Convergence diagnostics for the AR-Friedman simulation for one replication. Plot (a) contains trace plots of the average mean squared residuals across all time series; Plot (b) shows the average estimated log power spectrum across frequencies and all time series for each iteration; Plot (c) is the trace plots of the total number of nodes for each of the five trees separately (c); Plot (d) is the total number of bottom nodes across all five trees.

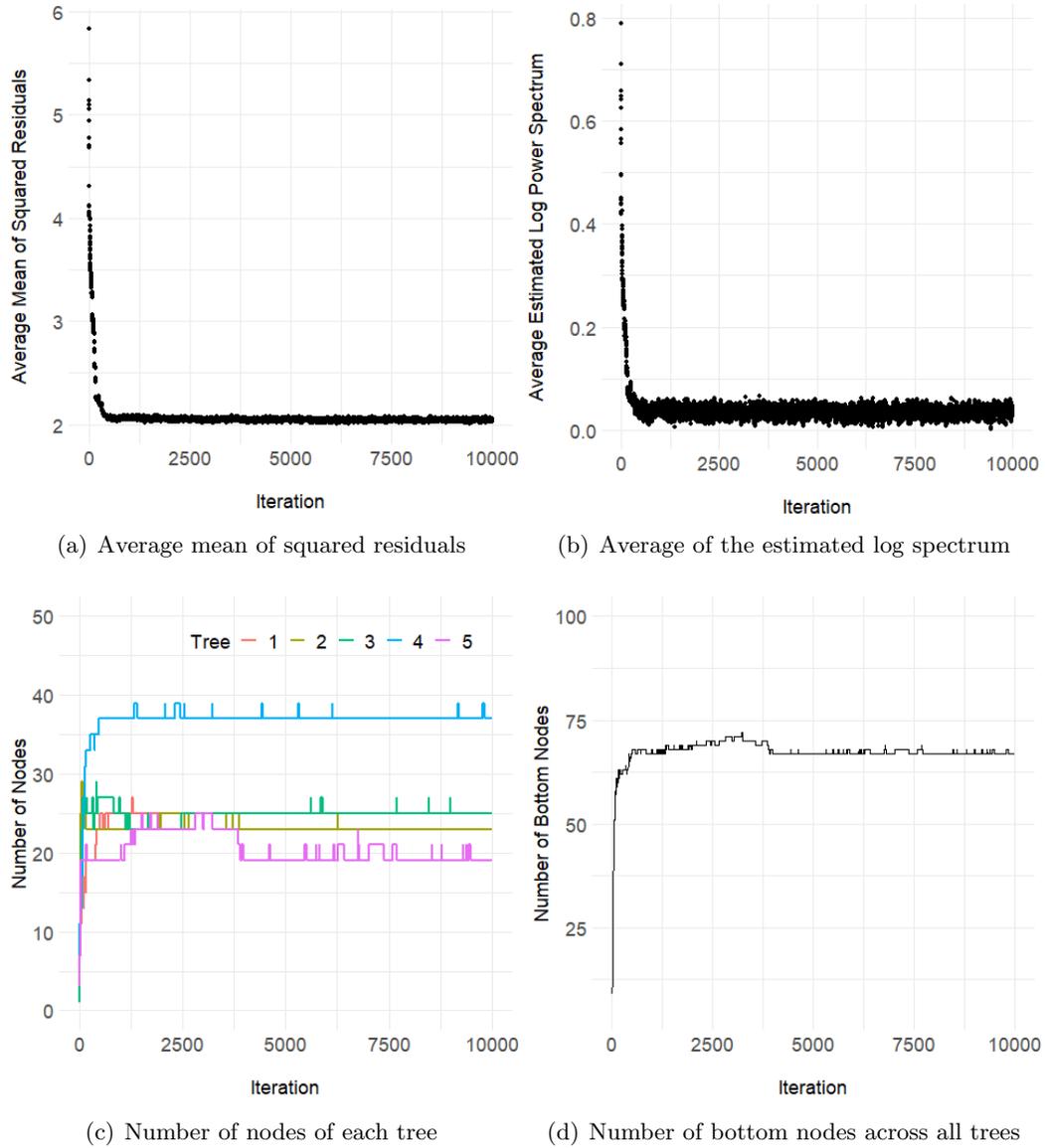


Figure D.12: Convergence diagnostics for the Adjusted-AdaptSPEC-X simulation for one replication. Plot (a) contains trace plots of the average mean squared residuals across all time series; Plot (b) shows the average estimated log power spectrum across frequencies and all time series for each iteration; Plot (c) is the trace plots of the total number of nodes for each of the five trees separately (c); Plot (d) is the total number of bottom nodes across all five trees.

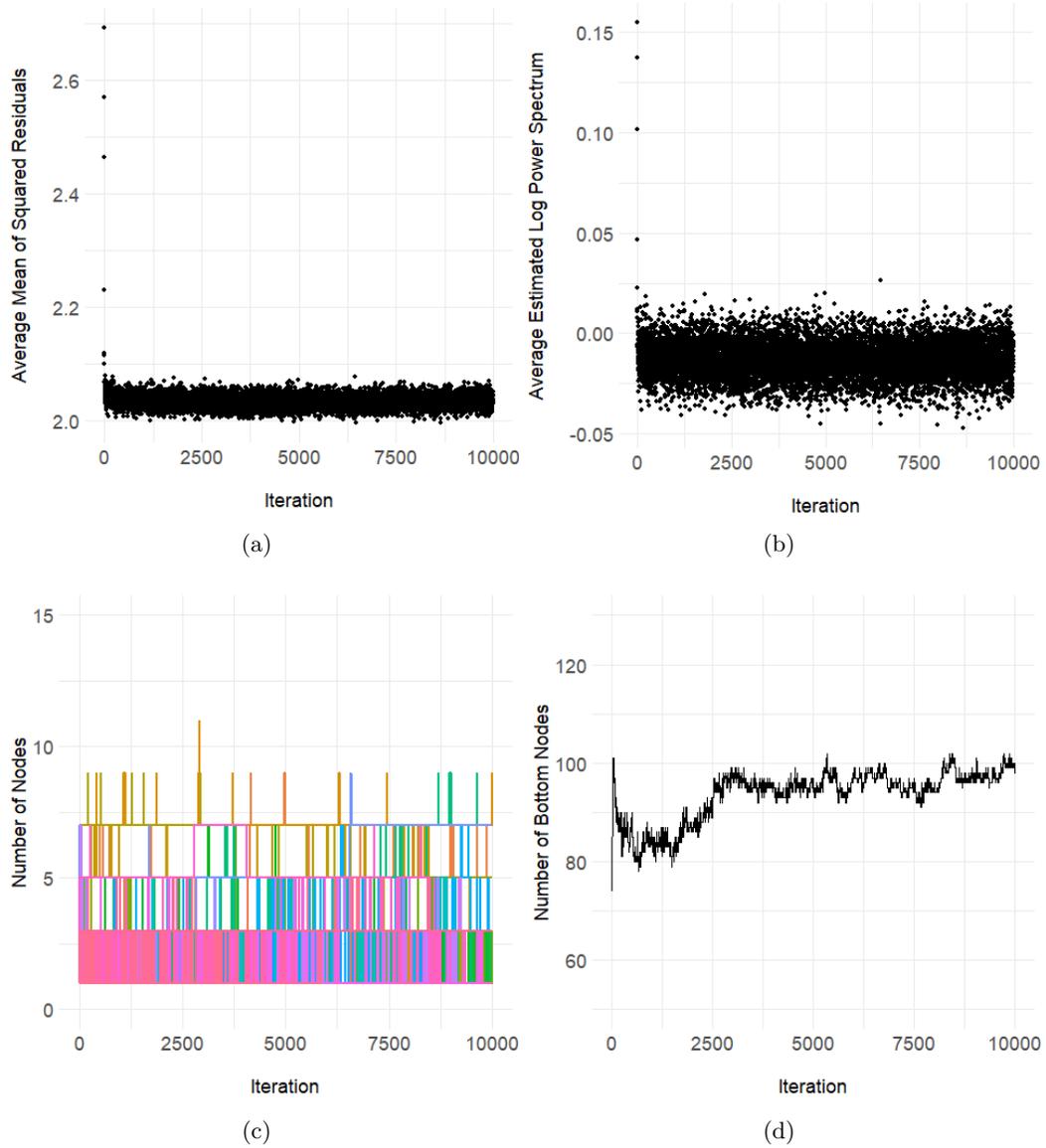


Figure D.13: Convergence diagnostic plots for the Abrupt+Smooth simulation for one replication for 50 trees. (a) average mean squared residuals across frequencies and all time series; (b) average estimated log power spectrum across frequencies and all time series; (c) total number of nodes for each of the five trees; (d) total number of bottom nodes across all 50 trees.

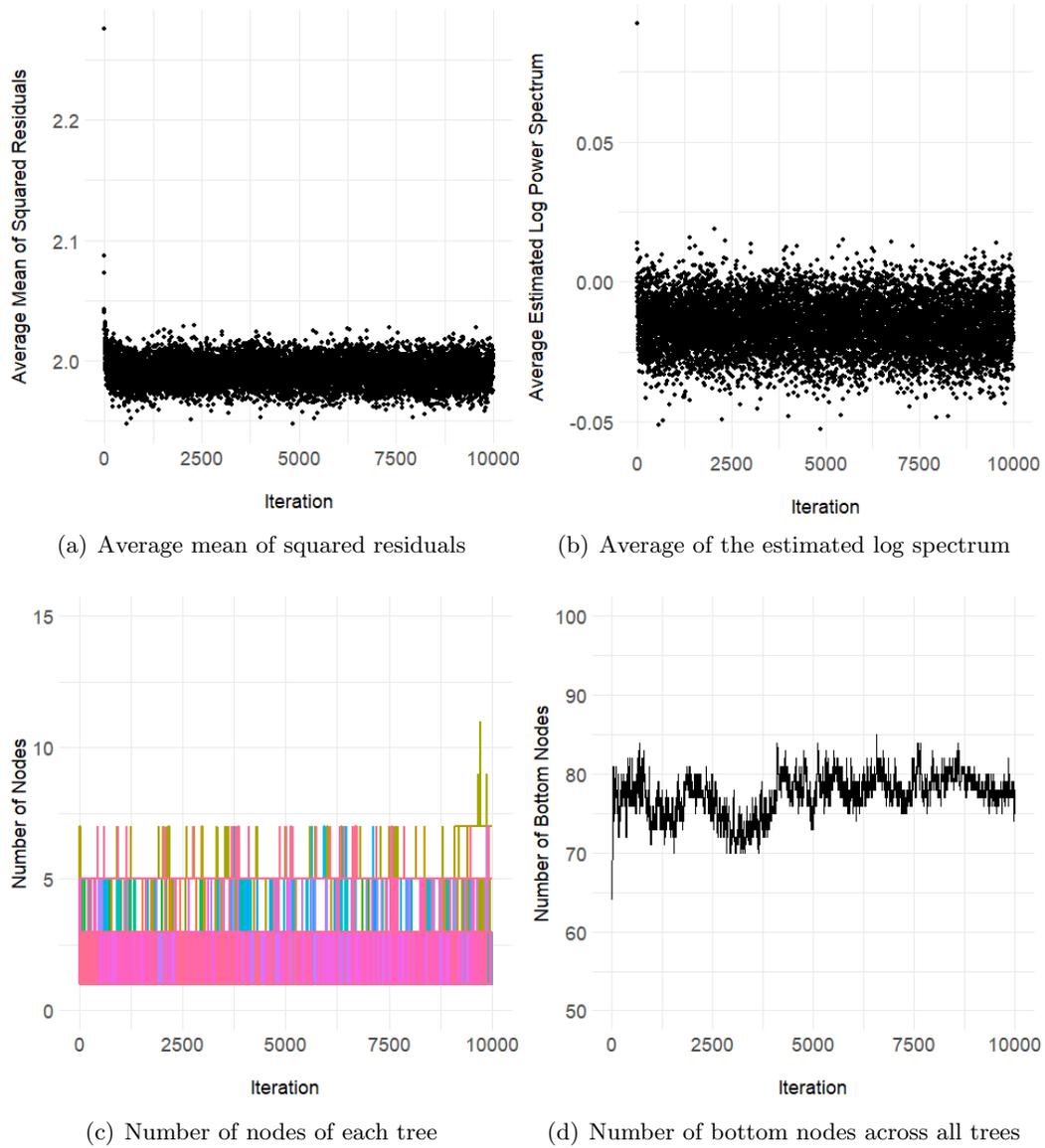


Figure D.14: Convergence diagnostics for the AR-Friedman simulation for one replication for 50 trees. Plot (a) contains trace plots of the average mean squared residuals across all time series; Plot (b) shows the average estimated log power spectrum across frequencies and all time series for each iteration; Plot (c) is the trace plots of the total number of nodes for each of the five trees separately (c); Plot (d) is the total number of bottom nodes across all 50 trees.

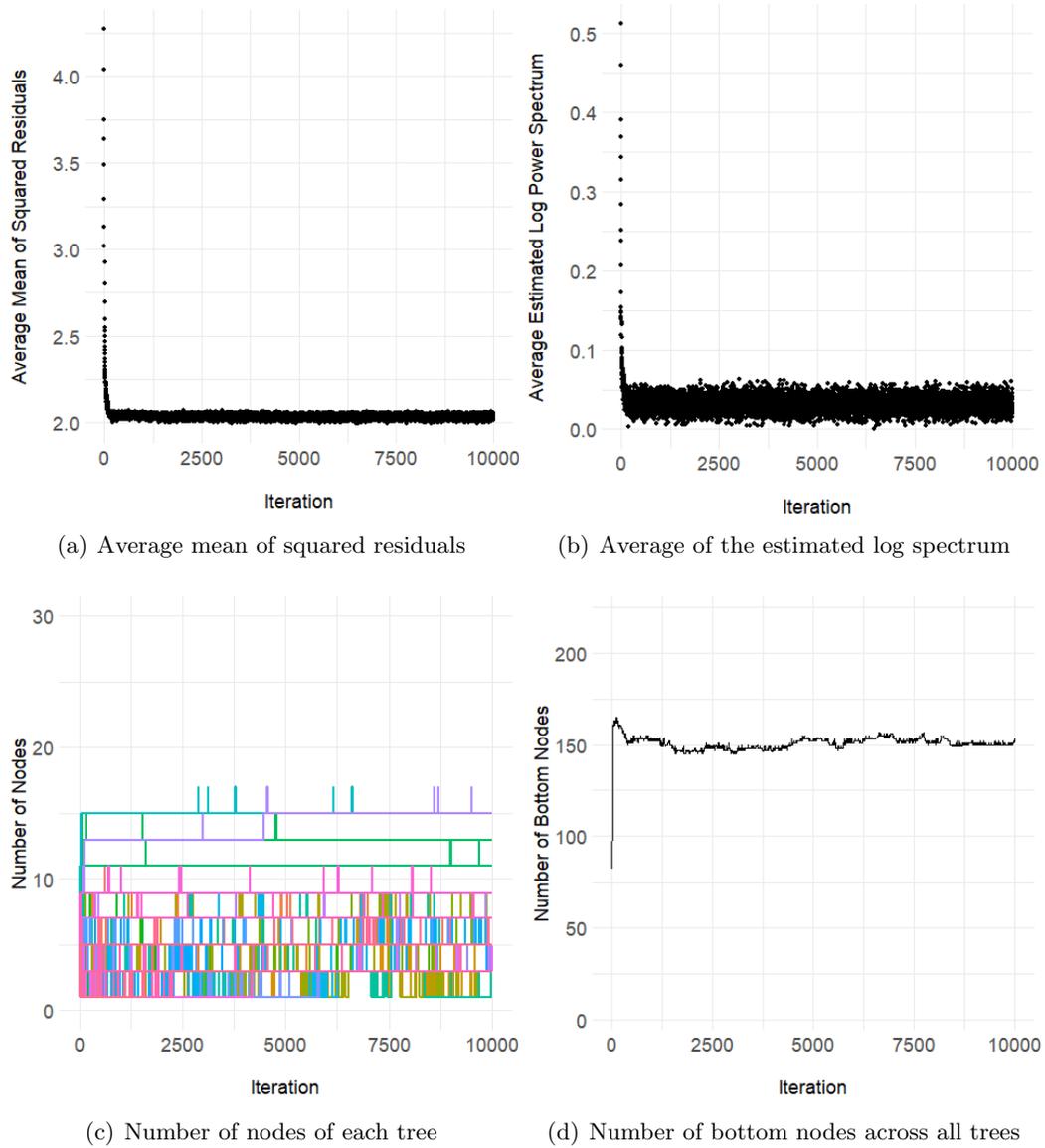
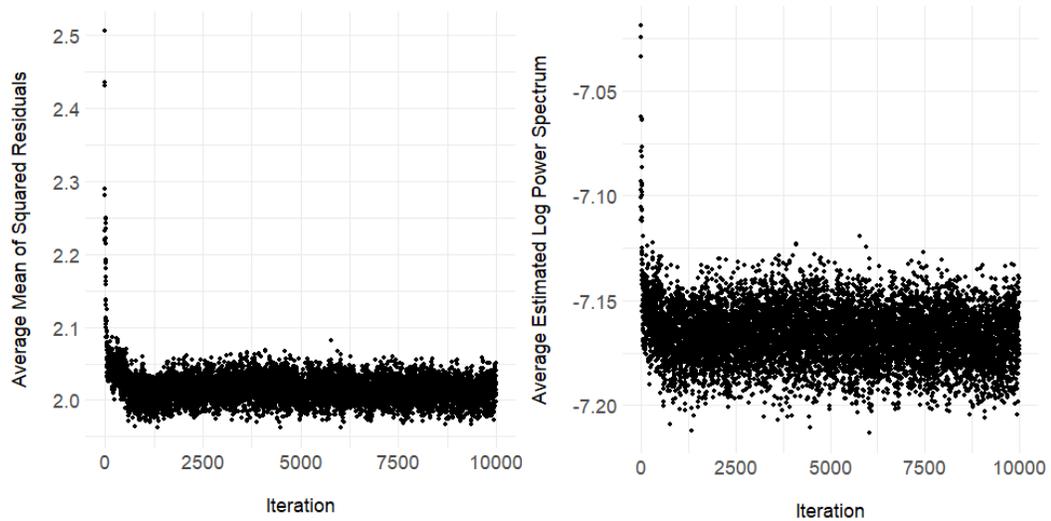
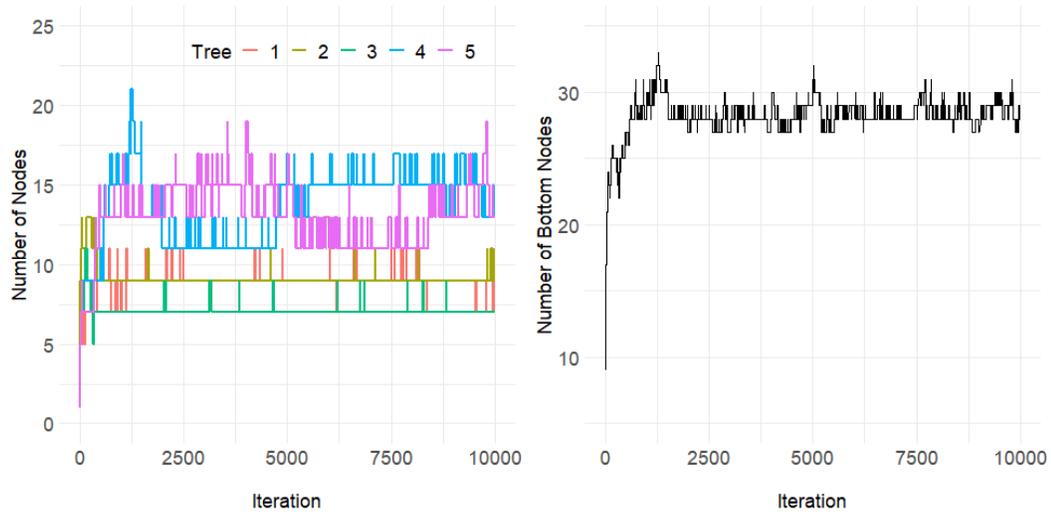


Figure D.15: Convergence diagnostics for the Adjusted-AdaptSPEC-X simulation for one replication for 50 trees. Plot (a) contains trace plots of the average mean squared residuals across all time series; Plot (b) shows the average estimated log power spectrum across frequencies and all time series for each iteration; Plot (c) is the trace plots of the total number of nodes for each of the five trees separately (c); Plot (d) is the total number of bottom nodes across all 50 trees.



(a) Average mean of squared residuals

(b) Average of the estimated log spectrum



(c) Number of nodes of each tree

(d) Number of bottom nodes across all trees

Figure D.16: Convergence diagnostics for the gait maturation data analysis. Plot (a) contains trace plots of the average mean squared residuals across all time series; Plot (b) shows the average estimated log power spectrum across frequencies and all time series for each iteration; Plot (c) is the trace plots of the total number of nodes for each of the five trees separately (c); Plot (d) is the total number of bottom nodes across all five trees.

Appendix E: Graphical Posterior Predictive Diagnostics

We use posterior predictive checks to diagnose proper model fit for the gait maturation analysis. The idea is that if a model is a good fit for the data, then it should be able to generate data similar to that which we observed originally (Gabry et al., 2019). Using the posterior draws of the covariate-dependent log power spectrum from the proposed RJMCMC sampler, in conjunction with the large-sample distribution of the log-periodogram, we can generate draws from the posterior predictive distribution of the log-periodogram for each time series to assess fit to the data.

Using our RJMCMC sampler, we are able to generate draws from the posterior distribution of the covariate-dependent log power spectrum, $p(\log f(\boldsymbol{\omega}, \nu) | \log \mathbf{I}_1, \dots, \log \mathbf{I}_L)$, where $\log \mathbf{I}_\ell = \{\log I_\ell(\nu_1), \log I_\ell(\nu_2), \dots, \log I_\ell(\nu_N)\}$ is the log periodogram for time series $X_{\ell t}$ such that

$$\log I_\ell(\nu_k) = \log \frac{1}{T} \left| \sum_{t=1}^T X_{\ell t} \exp(-2\pi i \nu_k t) \right|^2 \quad (\text{E.1})$$

where $N = \lfloor T/2 \rfloor - 1$, $\nu_k = k/T$ for $k = 1, \dots, N$ are the Fourier frequencies and T is the length of time series. Additionally, the large-sample properties of the log-periodogram provide the following data model (Pawitan and O’Sullivan, 1994)

$$\log I_\ell(\nu_k) = f(\boldsymbol{\omega}_\ell, \nu_k) + \log(\chi_2^2/2) \quad (\text{E.2})$$

for large T . Taken together, we can then generate draws from the posterior predictive distribution

$$p\left(\log \tilde{\mathbf{I}}_1, \dots, \log \tilde{\mathbf{I}}_L | \log \mathbf{I}_1, \dots, \log \mathbf{I}_L\right) = \int p(\log \mathbf{I}_1, \dots, \log \mathbf{I}_L | \log f(\boldsymbol{\omega}, \nu)) p(f(\boldsymbol{\omega}, \nu) | \log \mathbf{I}_1, \dots, \log \mathbf{I}_L) d\boldsymbol{\omega}, \nu \quad (\text{E.3})$$

by obtaining S draws of $\log f(\boldsymbol{\omega}_\ell, \nu_k)$ from its posterior distribution, and then using the data model above to obtain S draws of $\log \tilde{I}_\ell(\nu_k)$ for $\ell = 1, \dots, L$ and $k = 1, \dots, N$.

To demonstrate appropriate model fit for the gait maturation analysis, Figures E.1-E.7 display the observed log periodograms and $S = 250$ draws from the posterior predictive distribution of log periodograms for all 50 participants in the gait maturation study. All observed log-periodograms and densities are well covered by their respective draws from the posterior predictive distribution without any obvious indications of systematic or egregious lack of fit. Thus, we conclude that the model appears to adequately characterize the data from the gait maturation study.

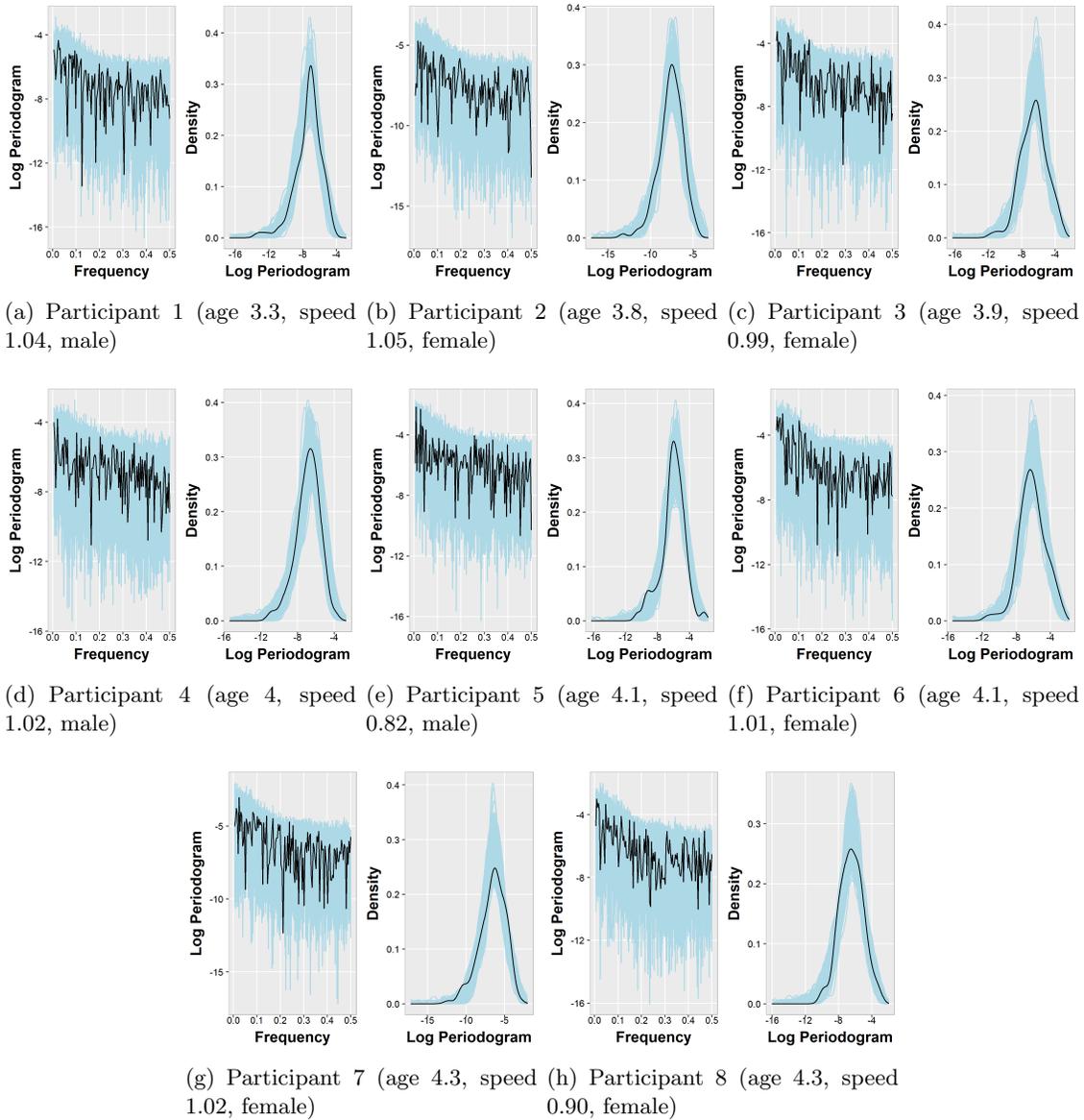


Figure E.1: Posterior predictive plots for participants 1-8 of gait maturation study. Left: Observed log-periodogram (black line) and posterior predictive log-periodogram draws (light blue). Right: Density plots for observed (black line) and posterior predictive log-periodogram draws (light blue) across frequencies. Age is in years and speed is in meters per second.

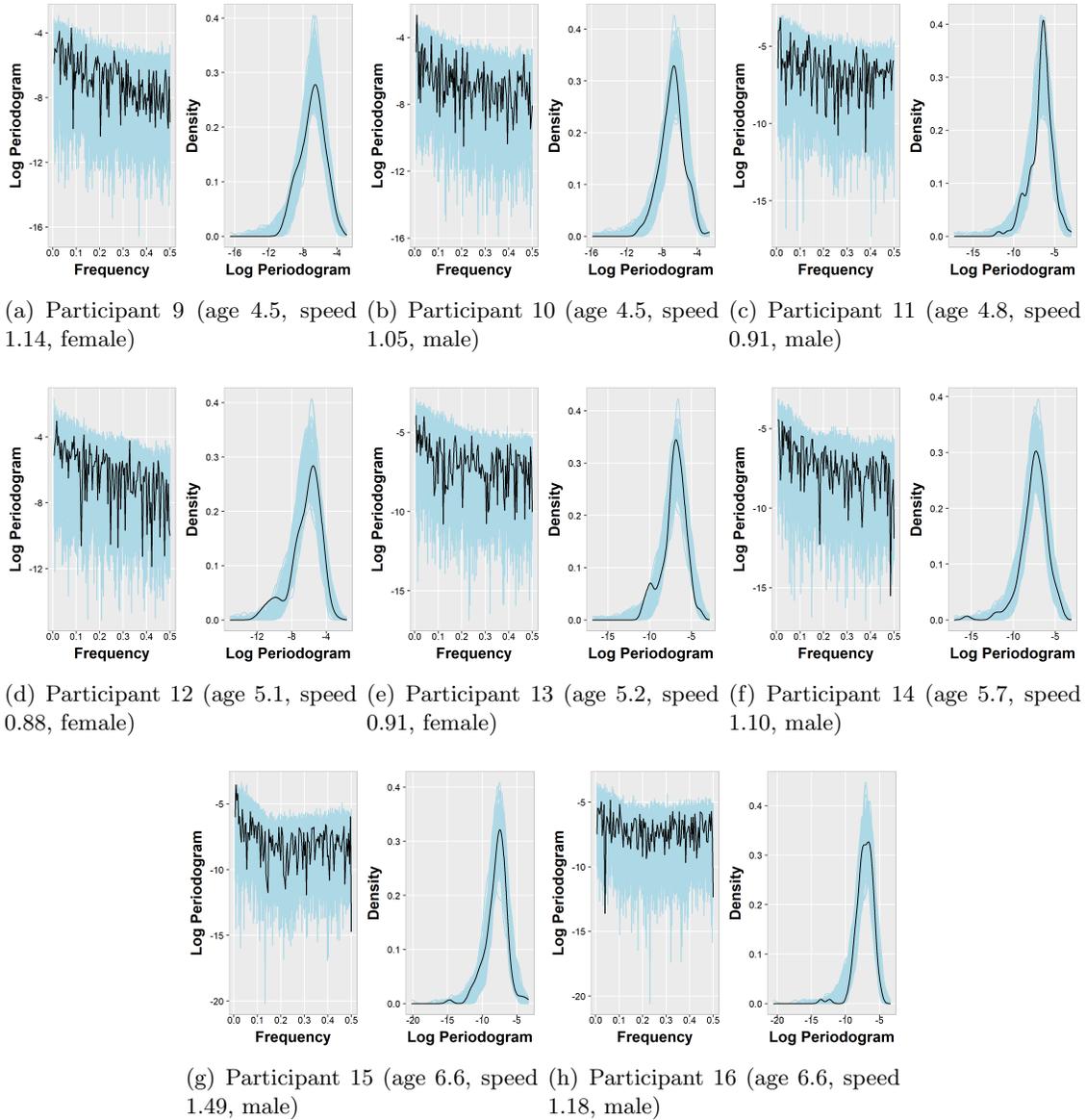


Figure E.2: Posterior predictive plots for participants 9-16 of gait maturation study. Left: Observed log-periodogram (black line) and posterior predictive log-periodogram draws (light blue). Right: Density plots for observed (black line) and posterior predictive log-periodogram draws (light blue) across frequencies. Age is in years and speed is in meters per second.

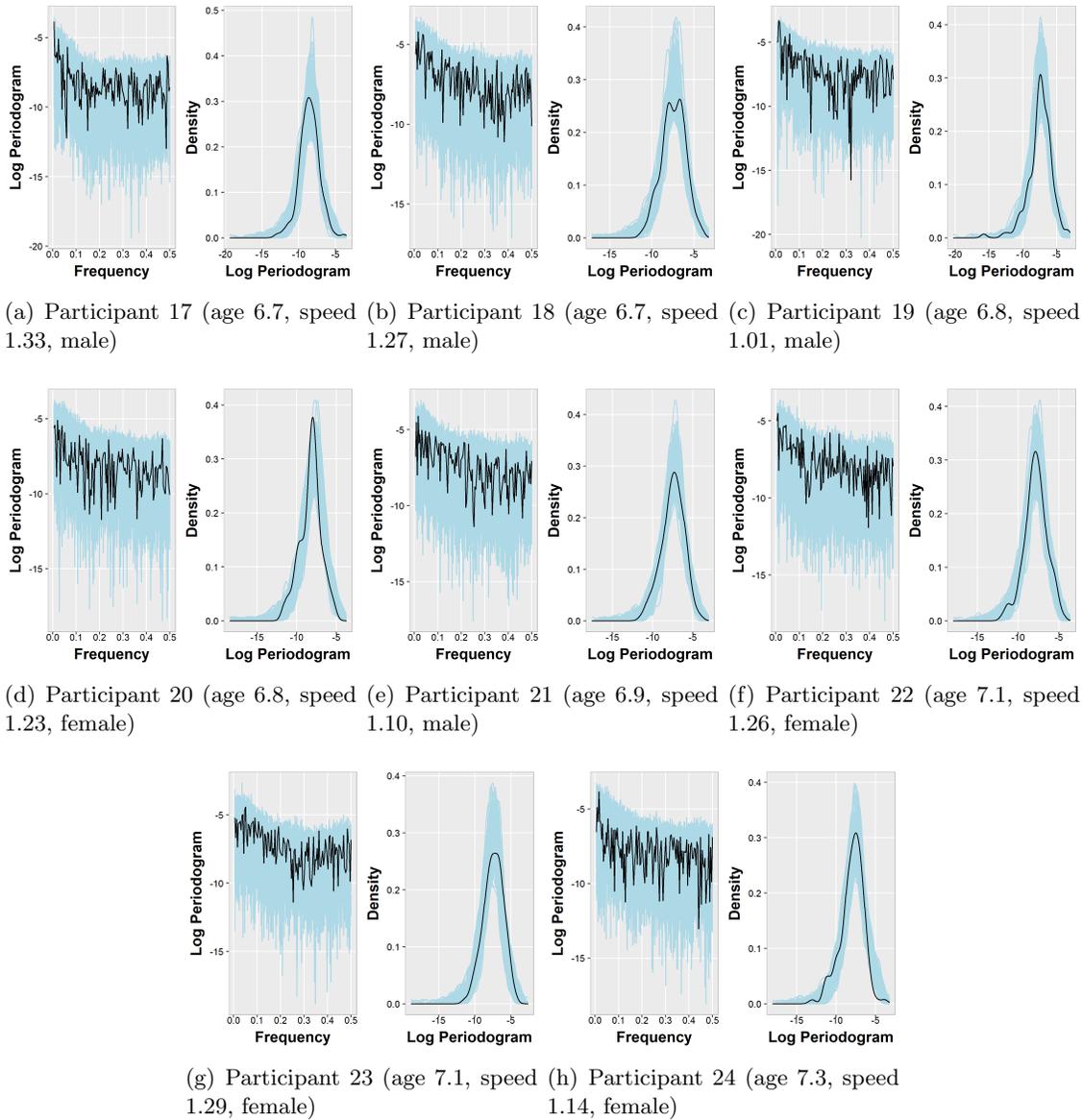


Figure E.3: Posterior predictive plots for participants 17-24 of gait maturation study. Left: Observed log-periodogram (black line) and posterior predictive log-periodogram draws (light blue). Right: Density plots for observed (black line) and posterior predictive log-periodogram draws (light blue) across frequencies. Age is in years and speed is in meters per second.

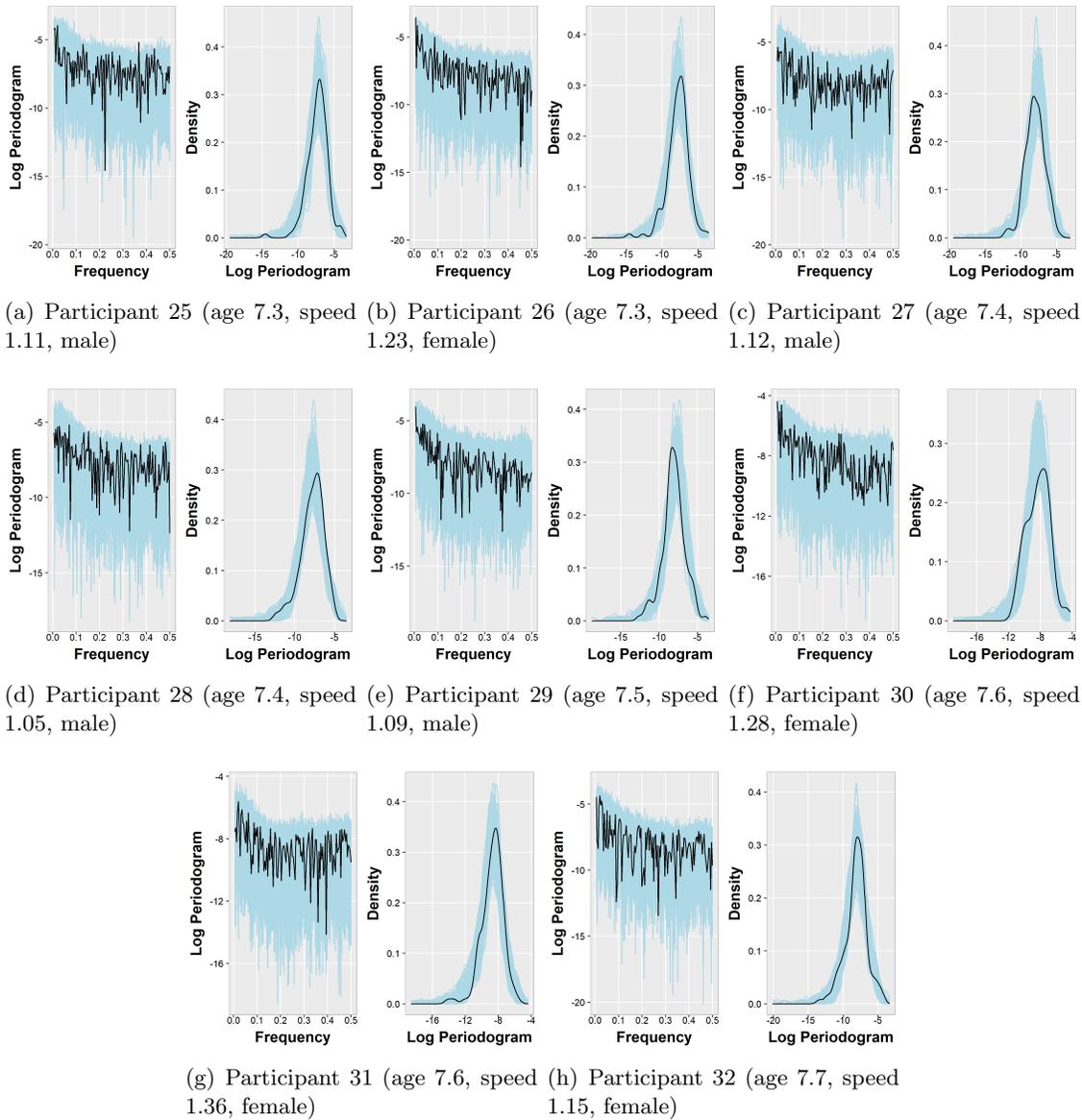


Figure E.4: Posterior predictive plots for participants 25-32 of gait maturation study. Left: Observed log-periodogram (black line) and posterior predictive log-periodogram draws (light blue). Right: Density plots for observed (black line) and posterior predictive log-periodogram draws (light blue) across frequencies. Age is in years and speed is in meters per second.

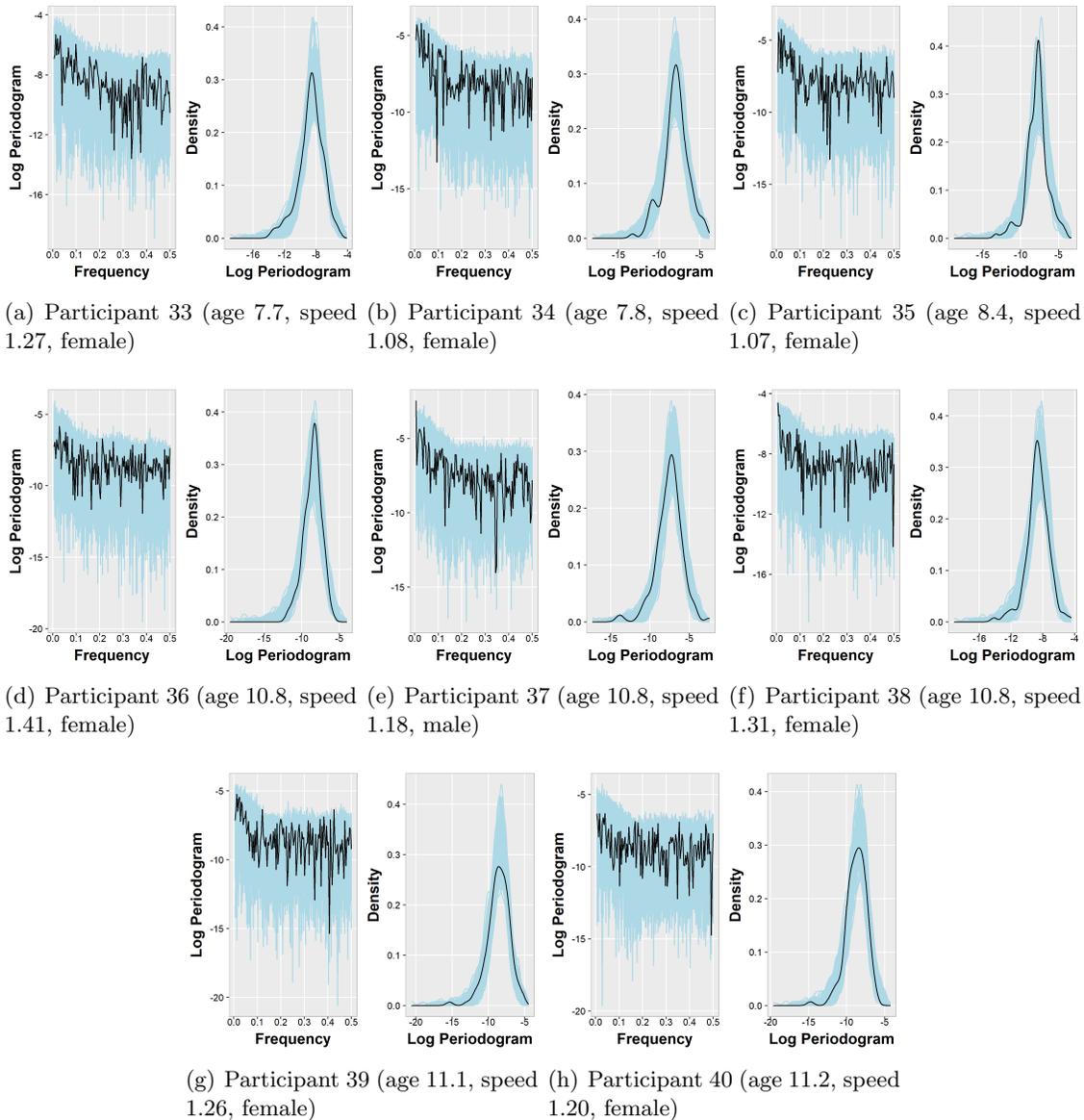


Figure E.5: Posterior predictive plots for participants 33-40 of gait maturation study. Left: Observed log-periodogram (black line) and posterior predictive log-periodogram draws (light blue). Right: Density plots for observed (black line) and posterior predictive log-periodogram draws (light blue) across frequencies. Age is in years and speed is in meters per second.

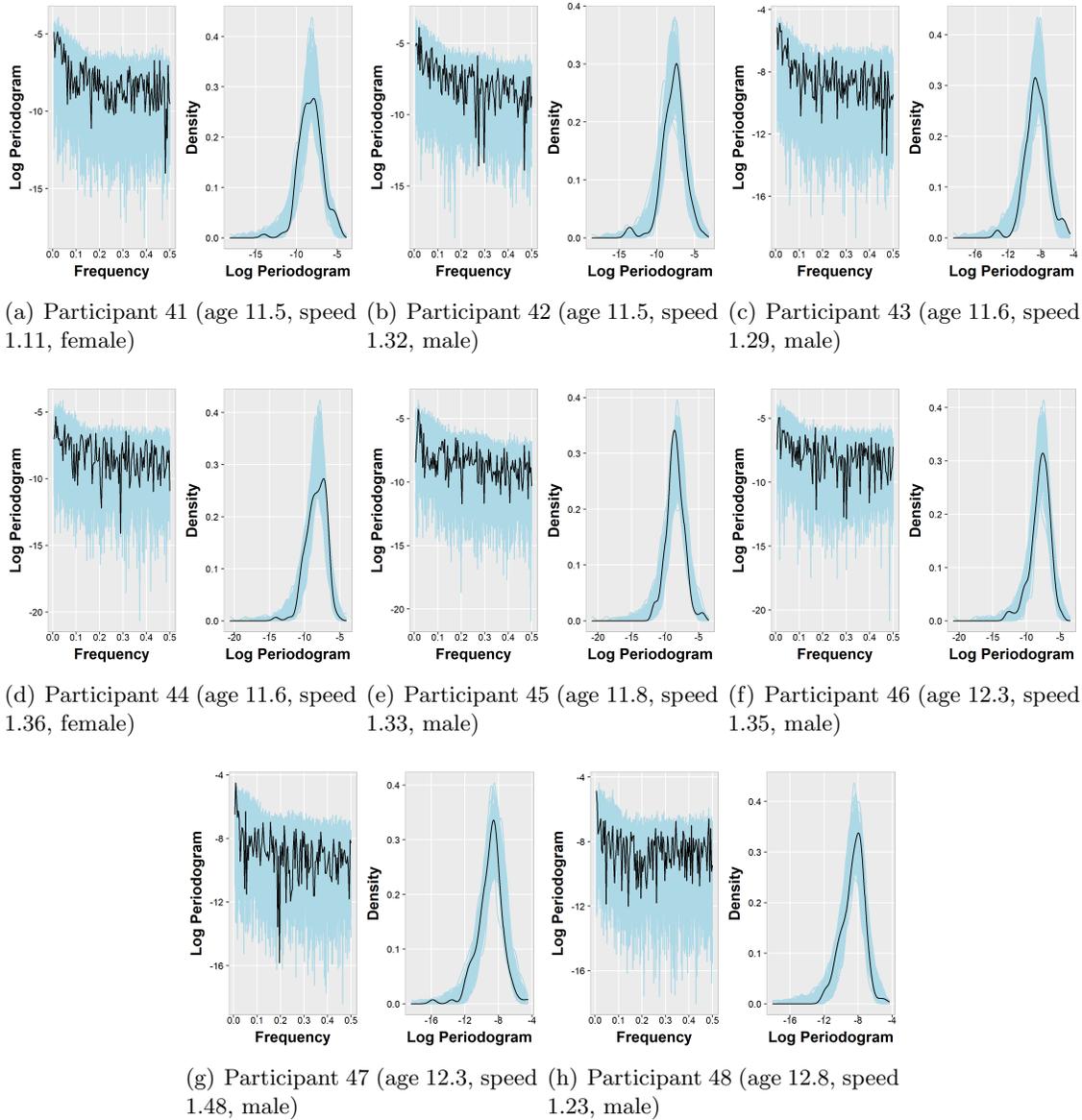


Figure E.6: Posterior predictive plots for participants 41-48 of gait maturation study. Left: Observed log-periodogram (black line) and posterior predictive log-periodogram draws (light blue). Right: Density plots for observed (black line) and posterior predictive log-periodogram draws (light blue) across frequencies. Age is in years and speed is in meters per second.

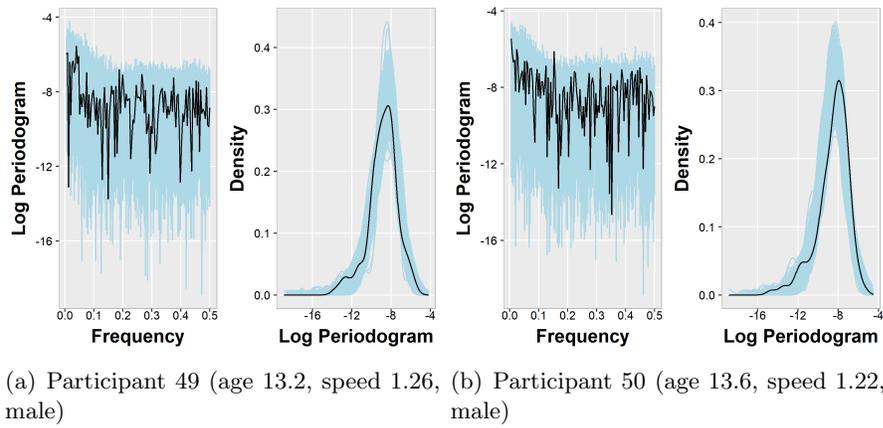


Figure E.7: Posterior predictive plots for participants 49-50 of gait maturation study. Left: Observed log-periodogram (black line) and posterior predictive log-periodogram draws (light blue). Right: Density plots for observed (black line) and posterior predictive log-periodogram draws (light blue) across frequencies. Age is in years and speed is in meters per second.

Bibliography

- (2005), “Bayesian Prediction via Partitioning,” *Journal of computational and graphical statistics*, 14, 811–830.
- Adak, S. (1998a), “Time-Dependent Spectral Analysis of Nonstationary Time Series,” *Journal of the American Statistical Association*, 93, 1488–1501.
- (1998b), “Time-Dependent Spectral Analysis of Nonstationary Time Series,” *Journal of the American Statistical Association*, 93, 1488–1501.
- Apley, D. W. and Zhu, J. (2020), “Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models,” *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 82, 1059–1086.
- Bertolacci, M., Rosen, O., Cripps, E., and Cripps, S. (2022), “AdaptSPEC-X: Covariate dependent spectral modeling of multiple nonstationary time series,” *Journal of Computational and Graphical Statistics*.
- Blattenberger, G. and Fowles, R. (2017), “Treed Avalanche Forecasting: Mitigating Avalanche Danger Utilizing Bayesian Additive Regression Trees,” *Journal of Forecasting*, 36, 165–180.
- Boots, B., Sugihara, K., Chiu, S. N., and Okabe, A. (2000), *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, Hoboken: John Wiley Sons, Incorporated.
- Breiman, L. (2001), “Random Forests,” *Machine Learning*, 45, 5–32.
- Bruce, S. A., Hall, M. H., Buysse, D. J., and Krafty, R. T. (2018), “Conditional Adaptive

- Bayesian Spectral Analysis of Nonstationary Biomedical Time Series,” *Biometrics*, 74, 260–269.
- Cadonna, A., Kottas, A., and Prado, R. (2019), “Bayesian Spectral Modeling for Multiple Time Series,” *Journal of the American Statistical Association*, 114, 1838–1853.
- Carter, C. K. and Kohn, R. (1997), “Semiparametric Bayesian Inference for Time Series with Mixed Spectra,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59, 255–268.
- Chau, J. and von Sachs, R. (2016), “Functional mixed effects wavelet estimation for spectra of replicated time series,” *Electronic Journal of Statistics*, 10, 2461 – 2510.
- Chipman, H., George, E. I., Gramacy, R. B., and McCulloch, R. (2013), “Bayesian Treed Response Surface Models,” *WIREs Data Mining and Knowledge Discovery*, 3, 298–305.
- Chipman, H. A., George, E. I., and McCulloch, R. E. (2010), “BART: Bayesian Additive Regression Trees,” *The Annals of Applied Statistics*, 4, 266–298.
- Cramér, H. (1942), “On Harmonic Analysis in Certain Function Spaces,” *Arkiv för Matematik, Astronomi och Fysik*, 28B, 1–7.
- Daliri, M. R. (2012), “Automatic diagnosis of neuro-degenerative diseases using gait dynamics,” *Measurement*, 45, 1729–1734.
- Denison, D., Adams, N., Holmes, C., and Hand, D. (2002), “Bayesian partition modelling,” *Computational statistics data analysis*, 38, 475–485.
- Diggle, P. J. and Al Wasel, I. (1997), “Spectral analysis of replicated biomedical time series,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46, 31–71.
- Eddelbuettel, D. and Sanderson, C. (2014), “RcppArmadillo: Accelerating R with High-performance C++ Linear Algebra,” *Computational Statistics and Data Analysis*, 71, 1054–1063.

- Eubank, R. (1999), *Nonparametric Regression and Spline Smoothing*, Boca Raton: CRC Press.
- Fiecas, M. and Ombao, H. (2017), “Modeling the Evolution of Dynamic Brain Processes During an Associative Learning Experiment,” *Journal of the American Statistical Association*, 111, 1440–1453.
- Friedman, J. H. (1991), “Multivariate Adaptive Regression Splines,” *The Annals of Statistics*, 19, 1–67.
- (2001), “Greedy Function Approximation: A Gradient Boosting Machine,” *The Annals of Statistics*, 29, 1189–1232.
- Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., and Gelman, A. (2019), “Visualization in Bayesian workflow,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182, 389–402.
- Gelman, A. (2006), “Prior Distributions for Variance Parameters in Hierarchical Models (Comment on Article by Browne and Draper),” *Bayesian Anal.*, 1, 515–534.
- Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., and Stanley, H. E. (2000), “PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals,” *Circulation*, 101, 215–220.
- Green, P. J. (1995), “Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination,” *Biometrika*, 82, 711–732.
- Green, P. J. and Sibson, R. (1978), “Computing Dirichlet Tessellations in the Plane,” *Comput. J.*, 21, 168–173.
- Hall, M., Vasko, R., Buysse, D., Ombao, H., Chen, Q., Cashmere, J. D., Kupfer, D., and Thayer, J. F. (2004), “Acute Stress Affects Heart Rate Variability during Sleep,” *Psychosomatic Medicine*, 66, 56–62.

- Hausdorff, J. M., Zeman, L., Peng, C.-K., and Goldberger, A. L. (1999), “Maturation of Gait Dynamics: Stride-to-Stride Variability and its Temporal Organization in Children,” *Journal of Applied Physiology*, 86, 1040–1047.
- Iannaccone, R. and Coles, S. (2001), “Semiparametric Models and Inference for Biomedical Time Series with Extra-Variation,” *Biostatistics*, 2, 261–276.
- Kapelner, A. and Bleich, J. (2016), “BARTMachine : Machine Learning with Bayesian Additive Regression Trees,” *Journal of Statistical Software*, 70, 1–40.
- Khorasani, A. and Daliri, M. R. (2014), “HMM for classification of Parkinson’s disease based on the raw gait data,” *Journal of medical systems*, 38, 1–6.
- Klimesch, W. (1999), “EEG Alpha and Theta Oscillations Reflect Cognitive and Memory Performance: A Review and Analysis,” *Brain Research Reviews*, 29, 169-195.
- Krafty, R. T. (2016), “Discriminant Analysis of Time Series in the Presence of Within-group Spectral Variability,” *Journal of Time Series Analysis*, 37, 435–450.
- Krafty, R. T., Hall, M., and Guo, W. (2011), “Functional Mixed Effects Spectral Analysis,” *Biometrika*, 98, 583–598.
- Krafty, R. T., Rosen, O., Stoffer, D. S., Buysse, D. J., and Hall, M. H. (2017), “Conditional Spectral Analysis of Replicated Multiple Time Series with Application to Nocturnal Physiology,” *Journal of the American Statistical Association*, 112, 1405–1216.
- Li, Z., Bruce, S. A., Wutzke, C. J., and Long, Y. (2021), “Conditional Adaptive Bayesian Spectral Analysis of Replicated Multivariate Time Series,” *Statistics in Medicine*, 40, 1989–2005.
- Li, Z. and Krafty, R. T. (2019), “Adaptive Bayesian Time-Frequency Analysis of Multivariate Time Series,” *Journal of the American Statistical Association*, 114, 453–465.
- Linero, A. R. (2018), “Bayesian Regression Trees for High-Dimensional Prediction and Variable Selection,” *Journal of the American Statistical Association*, 113, 626–636.

- Linero, A. R. and Yang, Y. (2018a), “Bayesian Regression Tree Ensembles that Adapt to Smoothness and Sparsity,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80, 1087–1110.
- (2018b), “Bayesian regression tree ensembles that adapt to smoothness and sparsity,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80, 1087–1110.
- Mentch, L. and Zhou, S. (2020), “Randomization as regularization: A degrees of freedom explanation for random forest success,” *Journal of Machine Learning Research*, 21, 1–36.
- Møller, J. (1994), *Lectures on random Voronoi tessellations*, Lecture notes in statistics ; 87, New York: Springer-Verlag.
- Pawitan, Y. and O’Sullivan, F. (1994), “Nonparametric spectral density estimation using penalized Whittle likelihood,” *Journal of the American Statistical Association*, 89, 600–610.
- Payne, R. D., Guha, N., Ding, Y., and Mallick, B. K. (2020), “A conditional density estimation partition model using logistic Gaussian processes,” *Biometrika*, 107, 173–190.
- Preis, S., Klemms, A., and Müller, K. (1997), “Gait analysis by measuring ground reaction forces in children: changes to an adaptive gait pattern between the ages of one and five years,” *Developmental Medicine & Child Neurology*, 39, 228–233.
- Preis, S., Klemms, A., and Müller, K. (2008), “Gait Analysis by Measuring Ground Reaction Forces in Children: Changes to an Adaptive Gait Pattern between the Ages of One and Five Years,” *Developmental Medicine and Child Neurology*, 39, 228–233.
- Priestley, M. B. (1965), “Evolutionary spectral and non-stationary processes (with Discussion),” *Journal of the Royal Statistical Society. Series B, Methodological*, 27, 204–.
- Qin, L., Guo, W., and Litt, B. (2009), “A Time-Frequency Functional Model for Locally

- Stationary Time Series Data,” *Journal of Computational and Graphical Statistics*, 18, 675–693.
- R Core Team (2021), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
- (2022), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
- Rosen, O., Wood, S., and Stoffer, D. (2012), “AdaptSPEC: Adaptive Spectral Estimation for Nonstationary Time Series,” *Journal of the American Statistical Association*, 107, 1575–1589.
- Ročková, V. and Saha, E. (2019), “On Theory for BART,” in *Proceedings of Machine Learning Research*, PMLR, vol. 89 of *Proceedings of Machine Learning Research*, pp. 2839–2848.
- Schwarz, K. and Krivobokova, T. (2016), “A Unified Framework for Spline Estimators,” *Biometrika*, 103, 121–131.
- Shumway, R. H. and Stoffer, D. S. (2017), *Time Series Analysis and Its Applications With R Examples*, Springer.
- Shumway-Cook, A. and Williams, M. W. (1995), “Motor Control: Theory and practical applications,” .
- Starling, J. E., Murray, J. S., Carvalho, C. M., Bukowski, R. K., Scott, J. G., et al. (2020), “BART with Targeted Smoothing: An Analysis of Patient-specific Stillbirth Risk,” *Annals of Applied Statistics*, 14, 28–50.
- Stoffer, D. S., Han, S., Qin, L., and Guo, W. (2010), “Smoothing Spline ANOPOW,” *Journal of Statistical Planning and Inference*, 140, 3789–3796.
- van der Merwe, S. (2018), “Time Series Analysis of the Southern Oscillation Index using Bayesian Additive Regression Trees,” *ArXiv*.

- Wahba, G. (1980), “Automatic Smoothing of the Log Periodogram,” *Journal of the American Statistical Association*, 75, 122–132.
- Waldmann, P. (2016), “Genome-Wide Prediction using Bayesian Additive Regression Trees,” *Genetics Selection Evolution (Paris)*, 48, 42–.
- Wand, M. P., Ormerod, J. T., Padoan, S. A., and Frühwirth, R. (2011), “Mean Field Variational Bayes for Elaborate Distributions,” *Bayesian Anal.*, 6, 847–900.
- Wei, W. W. S. (2006), *Time Series Analysis : Univariate and Multivariate Methods*, Boston: Pearson Addison Wesley, 2nd ed.
- Whittle, P. (1952), “The Simultaneous Estimation of a Time Series Harmonic Components and Covariance Structure,” *Trabajos de Estadística*, 3, 43-57.
- (1957), “Curve and Periodogram Smoothing,” *Journal of the Royal Statistical Society. Series B, Methodological*, 19, 38–63.
- Wood, S. (2017), *Generalized Additive Models: An Introduction with R*, Chapman and Hall/CRC, 2nd ed.
- Wu, Y. and Krishnan, S. (2009a), “Computer-aided analysis of gait rhythm fluctuations in amyotrophic lateral sclerosis,” *Medical & biological engineering & computing*, 47, 1165–1171.
- (2009b), “Statistical analysis of gait rhythm in patients with Parkinson’s disease,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 18, 150–158.
- Wu, Y. and Shi, L. (2011), “Analysis of altered gait cycle duration in amyotrophic lateral sclerosis based on nonparametric probability density function estimation,” *Medical engineering & physics*, 33, 347–355.
- Wu, Y., Tjelmeland, H., and West, M. (2007), “Bayesian CART: Prior Specification and Posterior Simulation,” *Journal of Computational and Graphical Statistics*, 16, 44–66.

Curriculum Vitae

RESEARCH INTERESTS

Computational Time Series Analysis; Spectral Analysis; Bayesian Statistical Learning; Applications in Medicine, Public Health

EDUCATION

George Mason University, Fairfax, Virginia, USA **Aug. 2018 - *expected* Aug. 2022**
Ph.D. Statistics

George Washington University, Washington, D.C., USA **Jul. 2014 - Aug. 2016**
M.S. Statistics

Taiyuan University of Technology, Taiyuan, Shanxi, China **Sep. 2010 - Jul. 2014**
B.S. Information and Computing Science

PAPERS

Yakun Wang, Zeda Li, and Scott Bruce. “Adaptive Bayesian Sum of Trees Model for Covariate Dependent Spectral Analysis” *Biometrics*. Invited revisions.

Bryndan Lindsey, **Yakun Wang**, Scott A. Bruce, and Nelson Cortes. (2021) “Eight-Week Individualized Gait Modification Intervention to Reduce Knee Adduction Moment: Preliminary Analysis of a Randomized Controlled Trial”. International Society of Biomechanics Congress, Stockholm, Sweden. Accepted.

PRESENTATIONS

CMStatistics, Virtual, December 2020

International Society for Bayesian Analysis (ISBA), Virtual, June 2021

COMPUTER SKILLS

Statistical Software: R, SAS, MATLAB

Database Software: Oracle

General purpose programming: C++, Python

Applications: L^AT_EX, Microsoft Office Suite

Certification: SAS Base certification, SAS Advanced certification

PROFESSIONAL EXPERIENCE

Data Analyst, **National Science Foundation (NSF)** **May 2021 - Aug.2021**
Alexandria, VA

- Aggregated responses from the Longitudinal Survey of Doctorate Recipients 2015-2019 data and developed multiple summary tables using R.
- Created Sankey plots to visualize career trajectories for recent doctorate recipients and late career doctorates for publication as an InfoBrief on the NSF website.
- Explored retirement trends of late career doctorates across key demographics (e.g. gender, race, and citizenship) using survival analysis.

Data Analyst/Program Manager, **Apr. 2017 - Aug. 2018**
China Transport Telecommunications & Information Center (CTTIC)
Beijing, China

- Developed a data pipeline to extract meaningful insights and summarize historical data using SQL from over 40 million records of transportation vehicles for the Chinese Ministry of Transportation.
- Conducted regular data processing and cleaning to protect data integrity used for downstream analytical purposes.
- Drafted written summary reports for monthly monitoring of improper vehicle registration records and corrections needed.

Consultant, **Georgetown University** **Aug. 2015 - Mar. 2017**
Washington, D.C.

- Analyzed transcripts of conversations with 34 migrant female sex workers and 18 male clients to extract data and investigate the relationship between HIV infections and methamphetamine drug use.
- Constructed summary tables to assess associations between HIV symptoms, living conditions, and demographic information for publication.

RESEARCH EXPERIENCE

Graduate Research Assistant, **George Mason University** **Aug. 2019 - Present**
Fairfax, Virginia

- Developing an adaptive computational Bayesian modeling framework for frequency-domain analysis of large complex biomedical time series data.
- Constructing a Bayesian sum of regression trees model and Voronoi tessellation model to capture complex dependencies and interactions between high-dimensional covariates and power spectra.
- Implementing Bayesian backfitting and reversible-jump Markov chain Monte Carlo algorithms to sample from the high-dimensional posterior distribution for both stationary and nonstationary time series.
- Studying gait maturation in young children by evaluating age-related changes in power spectra of stride interval time series in the presence of other covariates (e.g. sex and gait speed).

Graduate Student Researcher, **George Mason University** **Jun. 2020 - Aug. 2020**
Fairfax, Virginia

- Developed a longitudinal mixed-effects regression model to evaluate the effect of gait modification on peak knee adduction moment in knee osteoarthritis patients.
- Preliminary findings discovered a significant reduction in force across medial compartment of the knee upon completion of 8-week gait retraining program.

Graduate Research Assistant, **George Washington University** **Sep. 2015 - Aug. 2016**
Washington, D.C.

- Explored the robust Bayesian analysis for the priors of multi-variable normal linear regression with ϵ -contaminated class and the density ratio class.
- Conducted the Bayesian hypothesis test for the coefficients of multi-variable linear regression.

Classification of Poisonous Mushrooms, **George Washington University** **Oct. 2015**
Washington, D.C.

- Conducted data cleaning and created exploratory data visualizations by R (corrgram, vcd) with 22 attributes of mushrooms.
- Analyzed how the 22 attributes determine the toxicity of mushrooms using K-means, KNN, logistic regression and hierarchical clustering algorithms.