

AdaptSPEC-X: Covariate Dependent Spectral Modeling of Multiple Nonstationary Time Series

Michael Bertolacci¹, Ori Rosen², Edward Cripps³, and Sally Cripps⁴

¹University of Wollongong

²University of Texas at El Paso

³University of Western Australia

⁴University of Sydney

Abstract

We present the AdaptSPEC-X method for the joint analysis of a panel of possibly nonstationary time series. The approach is Bayesian and uses a covariate-dependent infinite mixture model to incorporate multiple time series, with mixture components parameterized by a time-varying mean and log spectrum. The mixture components are based on AdaptSPEC, a nonparametric model which adaptively divides the time series into an unknown number of segments and estimates the local log spectra by smoothing splines.

AdaptSPEC-X extends AdaptSPEC in three ways. First, through the infinite mixture, it applies to multiple time series linked by covariates. Second, it can handle missing values, a common feature of time series which can cause difficulties for nonparametric spectral methods. Third, it allows for a time-varying mean. Through these extensions, AdaptSPEC-X can estimate time-varying means and spectra at observed and unobserved covariate values, allowing for predictive inference. Estimation

is performed by Markov chain Monte Carlo (MCMC) methods, combining data augmentation, reversible jump, and Riemann manifold Hamiltonian Monte Carlo techniques. We evaluate the methodology using simulated data, and describe applications to Australian rainfall data and measles incidence in the US. Software implementing the method proposed in this paper is available in the R package BayesSpec.

Key words: Locally stationary time series; Measles; Multiple time series; Rainfall; Reversible jump Markov chain Monte Carlo; Whittle likelihood;

1 Introduction

When the available data are multiple time series thought to be realizations of nonstationary random processes, estimation of their time-varying mean and spectrum offers insight into the behavior of the processes, including whether and how they have changed over time. For example, an analysis of the spatial distribution of the frequency domain characteristics of time series of rainfall for sites spanning a wide spatial field can quantify the cyclical variability of the underlying process, while allowing for nonstationarity can suggest ways in which the climate has changed over the observation period.

Joint modeling of multiple time series with similar or identical local spectra improves estimation of the spectra by borrowing strength, and this is the motivation behind the approach taken in this article. This is particularly necessary when the time series are nonstationary and the data are historical, because the prospect of the spectrum changing implies that future observations may not help to estimate the local spectrum in earlier time periods (Priestley, 1965; Dahlhaus, 1997). This is reflected, for example, in theoretical frameworks for estimating evolutionary spectra, such as those of Priestley (1965), Dahlhaus (1997) and Dahlhaus (2012). In those frameworks, the relevant asymptotics are based on infill, requiring new observations of the process at finer time intervals; such infill is not possible for historical time series. Joint modeling of multiple time series with similar or identical local spectra is one way to ameliorate this problem, as estimates for different time series can borrow strength from each other (the same approach is also useful for stationary time series; see, e.g., Diggle and al Wasel, 1997). If additional information is available, such as covariates, it should be incorporated into the model to improve estimation.

This article addresses this problem by presenting methodology for analyzing a panel of possibly nonstationary time series using a covariate-dependent infinite mixture model, with mixture components parameterized by their time-varying mean and spectrum. The mixture components are based on AdaptSPEC (Rosen et al., 2012, henceforth RWS12), which partitions a (centered) time series into an unknown but finite number of segments, estimating the spectral density within each segment by smoothing splines. As part of the proposed method, AdaptSPEC is extended to handle missing values, a common feature of time series which can cause difficulties for nonparametric spectral methods. A second extension is the incorporation of a time-varying mean, which avoids having to de-mean (center) the time series as a preliminary step. The covariates, which are assumed to be time-independent, are incorporated via the mixture using the logistic stick breaking process (LSBP) of Rigon and Durante (2021), where the log odds for each ‘stick break’ are modeled using a thin plate spline Gaussian process over the covariates. The model is formulated in a Bayesian framework, where Markov chain Monte Carlo (MCMC) methods are used for parameter estimation and to deal with missing values. Specifically, as in AdaptSPEC, reversible jump MCMC (RJMCMC) is used to estimate the mixture component parameters, while the LSBP parameters are estimated via the Pólya-Gamma based latent variable expansion of Rigon and Durante (2021) (see also Polson et al., 2013, for the original latent variable expansion in the finite mixture case). The model and sampling scheme are capable of handling large panels, such as that of the measles application which has nearly 200,000 observations. In addition to estimating time-varying spectra for each time series in the panel, the covariate-dependent mixture structure allows inference about the underlying process at unobserved covariate values, enabling predictive inference. For instance, in this work, we use longitude and latitude as covariates when modeling Australian rainfall data, and are able to infer the predictive time-varying spectrum of the rainfall process at unobserved locations.

Many methods have been proposed for the spectral analysis of time series. As the focus of this paper is multiple time series, we provide background in the form of an overview of methods for nonstationary single time series. We then review methods for multiple time series, stationary or otherwise. This excludes methods for multivariate time series, and we refer readers to Li and Krafty (2019) for a review of past and recent work in this active research area.

Approaches to spectral estimation for a single nonstationary time series include fitting para-

metric time series models with time-varying parameters (Kitagawa and Gersch, 1996; Dahlhaus, 1997; West et al., 1999; Yang et al., 2016), smoothing the log periodogram (Ombao et al., 2001; Guo et al., 2003; Qin and Wang, 2008), dividing the time series into locally stationary segments (Adak, 1998; Davis et al., 2006; Rosen et al., 2009, 2012), using short time Fourier transforms (Yang and Zhou, 2020), and through the use of wavelets (Nason et al., 2000). For a recent and extensive overview of methods for single nonstationary time series, see Yang et al. (2016). Most directly relevant to this paper, Rosen et al. (2009) estimate the log of the spectral density using a Bayesian mixture of splines. The time series is partitioned into small sections, and it is assumed that the log spectral density within each partition is given by a mixture of smoothing splines. The mixture weights are assumed to be time-varying. RWS12 introduce the AdaptSPEC method, which avoids the fixed partitions of Rosen et al. (2009). AdaptSPEC partitions the time series into one or more variable length segments in an adaptive manner, modeling the log spectral density within each segment via a smoothing spline. This results in better estimates than those obtained from the method of Rosen et al. (2009). Furthermore, by averaging over the possible locations of the partition points, the method can accommodate both slowly and abruptly varying processes, as well as identify stationary processes. AdaptSPEC forms the basis of our proposed model for multiple time series.

For multiple stationary time series, the seminal work of Diggle and al Wasel (1997) establishes a frequentist framework for replicated stationary time series in which the log spectral density for each replicate varies randomly around a baseline population density. The variation is captured through a mixed-effects model as arising from both random and deterministic (i.e., covariate-dependent) causes, and is estimated through a probability model over the log periodograms of the replicates. Freyermuth et al. (2010) extend this method using tree-structured wavelets to improve the estimation of both the population and the replicate log spectra (see also Chau and von Sachs, 2016, for wavelet estimation for spectra of replicated time series). Krafty et al. (2011) construct a covariate-dependent model for multiple stationary time series in which the log spectrum has a mixed effects representation, where the effects are functions over the frequency domain.

Iannaccone and Coles (2001) describe a Bayesian extension of the work of Diggle and al Wasel (1997), in which the population spectral density is modeled using a nonparametric structure as in Carter and Kohn (1997), and MCMC is used for estimation. Macaro and Prado (2014) propose

a Bayesian model for multiple stationary time series with a covariate-dependent spectral density composed as a sum of spectral densities corresponding to different levels of two or more factors and, following [Choudhuri et al. \(2004\)](#), model the spectral density associated with these factors via Bernstein-Dirichlet priors. [Krafty et al. \(2017\)](#) present a Bayesian model for stationary multivariate time series based on the work of [Rosen and Stoffer \(2007\)](#), where multiple (multivariate) time series from different subjects are available, and subjects have an associated single covariate. [Cadonna et al. \(2019\)](#) model multiple stationary time series using a Bayesian hierarchical model. The log-periodogram of a single stationary time series is modeled as a mixture of Gaussian distributions where the mixture weights and mean functions are frequency-dependent. The hierarchical model for multiple time series is constructed by setting the mean functions to be common to all time series while letting the weights vary between time series.

For multiple nonstationary time series, [Qin et al. \(2009\)](#), hereafter referred to as QGL09) present a frequentist method for estimating the covariate-dependent time-varying spectral density of a collection of time series with covariates. The covariates are allowed to vary with time, and the local log spectral density at each time within each series is a linear combination of time-frequency surfaces, where the coefficients on the combination are the covariates. This has similarities to the approach for stationary time series used by [Diggle and al Wasel \(1997\)](#), but random variation in the replicates is not accounted for. Estimation is performed in two stages. First, the time period of the study is split into blocks, and the local log periodogram is calculated within each block and for each replicate. The covariate-dependent time-varying spectral density is then estimated using an approximate functional model over the local log periodograms. Also frequentist, [Fiecas and Ombao \(2016\)](#) perform spectral estimation for replicated nonstationary multivariate time series, where the spectrum changes slowly between replications. The application is to multiple time series collected from the same subject over time. The estimation technique, which has similarities to that of QGL09, first splits the time series into blocks and computes local periodogram matrices, and then smooths over the blocks to estimate the spectrum. [Bruce et al. \(2018\)](#) present a method for multiple nonstationary time series with a single covariate that is referred to as conditional adaptive Bayesian spectrum analysis (CABS), which adaptively partitions both time- and covariate-space, modeling the spectrum within each partition by smoothing splines (as in [RWS12](#)).

Augmenting this literature, we present methodology, referred to as AdaptSPEC-X, which com-

bines four features: multiple time series, nonstationarity in both mean and spectrum, multiple covariates, and missing data. We demonstrate the method on simulated data, and show how it can be used to estimate the mean and spectra in two application areas: Australian rainfall data, and measles incidence in the United States. Software implementing AdaptSPEC-X is available in the R package BayesSpec.¹

The AdaptSPEC-X methodology constitutes an advance of the existing state-of-the-art in several respects. A near-universal feature in the literature on spectral estimation is to assume the mean is known or fixed over time (see, e.g., [Diggle and al Wasel, 1997](#)), whereas AdaptSPEC-X allows for the mean to be unknown and to change over time. A fixed-mean assumption is appropriate in many settings, but not universally so. In particular, the assumption would be inappropriate in the applications presented in the paper. In many cases, particularly those with highly variable short time series, there may also be considerable uncertainty regarding the mean, so assuming it is known would lead to overconfident and potentially different inference about other aspects of the time series. Another advance is AdaptSPEC-X's treatment of missing values, which are accommodated via the MCMC scheme; while most parametric models have little trouble with missing data, they are typically not handled by nonparametric spectral methods. These first two advances are in fact extensions of AdaptSPEC ([RWS12](#)), and are not unique to the setting of multiple nonstationary time series.

In the setting of multiple nonstationary time series, AdaptSPEC-X shares characteristics with the CABS method of [Bruce et al. \(2018\)](#), in that both are Bayesian, both take as their starting point a piecewise-stationary model with a nonparametric model for the log spectrum within each segment, and both can accommodate either slowly- or abruptly-varying time series. AdaptSPEC-X differs from CABS in that it can also handle multiple covariates, a time-varying mean, and missing values. The other comparable method to AdaptSPEC-X is that of [QGL09](#), which also allows for multiple covariates. Instead of adaptively partitioning time, as does AdaptSPEC-X to cater for both abrupt changes and (through model averaging) smooth changes to the spectrum, [QGL09](#) preprocess the time series into blocks that are assumed stationary, thus ignoring the uncertainty associated with the partition of the time series. A similar approach to this is used by [Qin and](#)

¹Available from GitHub at <https://github.com/mbertolacci/BayesSpec/>. As of publication, the version of the package on CRAN does not contain AdaptSPEC-X.

Wang (2008) for univariate time series. RWS12 contrast the adaptive partitioning approach with the block-based approach of Qin and Wang (2008), and conclude that AdaptSPEC yielded better estimates of the spectrum. One feature of the method of QGL09 that AdaptSPEC-X lacks is the ability to accommodate time-varying covariates; the converse is true for the ability to handle time-varying means and missing values. A further difference is that AdaptSPEC-X is fully Bayesian, and, through its use of thin-plate splines, incorporates covariates in a more general fashion than that provided by the additive model used by QGL09.

The paper proceeds as follows. Section 2 describes AdaptSPEC, the model for single nonstationary time series forming the basis for the analysis of multiple nonstationary time series. AdaptSPEC-X, a covariate-dependent infinite mixture model, is presented in Section 3. Section 4 outlines the MCMC scheme used to estimate the model parameters. Section 5 presents simulation studies that examine the performance of the method. Section 6 describes the application areas. Australian rainfall is analyzed in Section 6.1, and measles incidence in the US is discussed in Section 6.2. The article is accompanied by a set of Appendices in the supplementary material. Appendix A provides details of the conditional distributions necessary for the sampling scheme, and Appendix B expands on the covariance structure used to derive the conditional distribution of the missing values.

2 Model for single time series

RWS12 present the AdaptSPEC method for modeling single nonstationary time series which we summarize in this section. Let $\mathbf{x} = (x_1, \dots, x_n)'$ be a time series of length n . Assume for ease of notation that n is even, and suppose initially that \mathbf{x} is a realization from a stationary process $\{X_t\}$ with constant mean μ and a bounded positive spectral density $f(\omega)$ for $\omega \in (-\frac{1}{2}, \frac{1}{2}]$. Whittle (1957) shows that, for large n , the likelihood of \mathbf{x} can be approximated as

$$p(\mathbf{x} \mid \mu, f) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\prod_{k=1}^n f(\omega_k)^{1/2}} \exp \left\{ -\frac{1}{2} \sum_{k=1}^n \frac{I_k}{f(\omega_k)} \right\}, \quad (1)$$

where $\omega_k = \frac{k-1}{n}$ for $k = 1, \dots, n$ are the Fourier frequencies, $I_k = |d_k|^2$ is the periodogram at ω_k ,

and

$$d_k = \frac{1}{\sqrt{n}} \sum_{t=1}^n (x_t - \mu) e^{-2\pi i \omega_k (t-1)} \quad (2)$$

is the discrete Fourier transform (DFT) at ω_k , with $i = \sqrt{-1}$. RWS12 follow Wahba (1990) by expressing $\log f$ as

$$\log f(\omega) = \alpha_0 + h(\omega), \quad (3)$$

and placing a smoothing spline prior on $h(\omega)$. Due to the evenness of $f(\omega)$ and the periodogram, $h(\omega)$ is modeled on the domain $\omega \in [0, 0.5]$, corresponding to the first $\frac{n}{2} + 1$ Fourier frequencies. This prior is expressed via a linear combination of J basis functions, where $J < \frac{n}{2} + 1$ is chosen to balance prior flexibility and computational resources. See Appendix A.1 in the supplementary material and RWS12 for details.

Next we allow the underlying process $\{X_t\}$ to be nonstationary. Let a time series consist of a number of segments, m , and let $\xi_{s,m}$ be the end of the s th segment, $s = 1, \dots, m$, where $\xi_{0,m} = 0$ and $\xi_{m,m} = n$. Then we assume that $\{X_t\}$ is piecewise stationary, with

$$X_t = \sum_{s=1}^m X_t^s \delta_{s,m}(t), \quad (4)$$

where the processes $\{X_t^s\}$ are independent and stationary with means $\mu_{s,m}$, spectral densities $f_{s,m}(\omega)$, and $\delta_{s,m}(t) = 1$ iff $t \in (\xi_{s-1,m}, \xi_{s,m}]$. Consider a realization \mathbf{x} from (4). RWS12 approximate the likelihood of \mathbf{x} by

$$g(\mathbf{x} | \Theta) = \prod_{s=1}^m p(\mathbf{x}_{s,m} | \mu_{s,m}, f_{s,m}), \quad (5)$$

where $\Theta = \{m, \boldsymbol{\xi}_m, \boldsymbol{\mu}_m, f_{1,m}, \dots, f_{m,m}\}$, $\boldsymbol{\xi}_m = (\xi_{1,m}, \dots, \xi_{m,m})'$, $\boldsymbol{\mu}_m = (\mu_{1,m}, \dots, \mu_{m,m})'$, $\mathbf{x}_{s,m} = \{x_t : \delta_{s,m}(t) = 1\}$ are the data for the s th segment, and $p(\mathbf{x}_{s,m} | \mu_{s,m}, f_{s,m})$ is the Whittle likelihood (1). The values of m , $\boldsymbol{\xi}_m$ and $f_{s,m}$ for $s = 1, \dots, m$ are considered unknown and are assigned priors. For $f_{s,m}$, the prior in (3) is used, while m is given the discrete uniform prior between 1 and M . (See RWS12, for the details of the prior on $\boldsymbol{\xi}_m$). RWS12 consider $\boldsymbol{\mu}_m$ to be known and equal to zero, but this work considers it unknown and assigns to $\mu_{s,m}$ a uniform prior with support $\mu_- < \mu_{s,m} < \mu_+$. A minimum segment length t_{\min} is set to ensure that there are sufficient time periods within each segment so that the Whittle likelihood approximation is appropriate.

RWS12 describe a reversible jump Markov chain Monte Carlo algorithm (Green, 1995) that

samples from the posterior distribution of this model that allows AdaptSPEC to handle both abruptly and slowly varying nonstationary time series, as well as identify whether a time series is stationary. AdaptSPEC forms the basis of our spectral estimation technique for multiple time series.

3 Model for multiple time series

In this section we extend the model of Section 2 to multiple time series. Suppose now that the stochastic process $\{X_t\}$ has associated covariates $\mathbf{u} = (u_1, \dots, u_P)'$. We model $\{X_t\}$ with a covariate-dependent mixture structure

$$\{X_t\} \sim \sum_{h=1}^H \pi_h(\mathbf{u}) g_h(\{X_t\} \mid \Theta_h), \quad (6)$$

where the mixture component distributions g_h are instances of AdaptSPEC (Equation (5)) with parameters $\Theta_h = \{m^h, \boldsymbol{\xi}_m^h, \boldsymbol{\mu}_m^h, f_{1,m}^h, \dots, f_{m,m}^h\}$, $2 \leq H \leq \infty$, and the mixture weights $\pi_h(\cdot)$ satisfy $0 \leq \pi_h(\cdot) \leq 1$ and $\sum_{h=1}^H \pi_h(\mathbf{u}) = 1$. Equation (6) implies that $\{X_t\}$'s distribution is determined by its covariates \mathbf{u} , which, importantly, do not vary with time. The purpose of the mixture structure in Equation (6) is to induce covariate-dependence in a flexible, semi-parametric manner, and we do not use this structure to perform inference about clustering or the number of clusters among multiple time series.

Let $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ be a finite collection of N time series, of length n each, where each time series $\mathbf{x}_j = (x_{1,j}, \dots, x_{n,j})'$ has covariates $\mathbf{u}_j = (u_{1,j}, \dots, u_{P,j})'$ for $j = 1, \dots, N$. Assuming independence conditional on $\pi_h(\cdot)$ and Θ_h , it follows from Equation (6) that the joint distribution of the collection is

$$p(\mathbf{x}_1, \dots, \mathbf{x}_N) = \prod_{j=1}^N \sum_{h=1}^H \pi_h(\mathbf{u}_j) g_h(\mathbf{x}_j \mid \Theta_h). \quad (7)$$

3.1 Model for mixture weights

For the mixture weights $\pi_h(\mathbf{u})$ in Equation (6), we use the LSBP of [Rigon and Durante \(2021\)](#), according to which $\pi_h(\mathbf{u})$ is given by

$$\pi_h(\mathbf{u}) = v_h(\mathbf{u}) \prod_{h'=1}^{h-1} (1 - v_{h'}(\mathbf{u})), \quad (8)$$

where $\text{logit } v_h(\mathbf{u}) = w_h(\mathbf{u})$, so that $w_h(\mathbf{u})$ are the covariate-dependent log odds. This prior allows for $2 \leq H \leq \infty$ mixture components, with $v_H(\mathbf{u}) = 1$ when $H < \infty$.

As in [Rigon and Durante \(2021\)](#), this construction can be interpreted via sequential (continuation-ratio) logits ([Agresti, 2018](#)). Let $z_j \in \{1, 2, \dots, H\}$ be a latent indicator such that $(\mathbf{x}_j \mid z_j = h) \sim g_h(\mathbf{x}_j \mid \Theta_h)$. Then the LSBP can be represented in a generative manner as a sequence of decisions, where $p(z_j = 1) = v_1(\mathbf{u}_j) = (1 + \exp(-w_1(\mathbf{u}_j)))^{-1}$, $p(z_j = 2 \mid z_j > 1) = v_2(\mathbf{u}_j) = (1 + \exp(-w_2(\mathbf{u}_j)))^{-1}$, and so on, such that in general, $p(z_j = h \mid z_j > h - 1) = (1 + \exp(-w_h(\mathbf{u}_j)))^{-1}$.

The model in equations (7) and (8) has a similar structure to the classic mixture of experts model ([Jacobs et al., 1991](#)), in which the weights of a finite mixture depend on covariates through multinomial logits. Our motivation for choosing the LSBP (Equation (8)) over multinomial logits is to obviate the choice of the number of mixture components. The model in equations (7) and (8) is in principle an infinite mixture and so the question of the number of components becomes irrelevant. In practice, however, it is common to truncate the infinite representation at a suitably high but finite K ([Ishwaran and James, 2001](#)). The LSBP is analogous to the probit stick-breaking process ([Chung and Dunson, 2009](#)), where a probit link function is used in place of the logit.

3.2 Model for log odds

We model the log odds by a Gaussian process (GP) prior $w_h(\mathbf{u}) \sim \text{GP}(\beta_{0,h} + \mathbf{u}'\boldsymbol{\beta}_h, \tau_h^2 \Omega(\cdot, \cdot))$, where $\beta_{0,h}$ is an intercept, $\boldsymbol{\beta}_h = (\beta_{1,h}, \dots, \beta_{P,h})'$ is a vector of regression coefficients, τ_h^2 is a smoothing parameter, and $\Omega(\mathbf{u}, \mathbf{u}')$ is the covariance kernel constructed via the reproducing kernel Hilbert space defined by a P -dimensional thin-plate Gaussian process prior (see [Wood, 2013](#)). For a finite collection of N time series $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, with associated covariates $\{\mathbf{u}_1, \dots, \mathbf{u}_N\}$, the log odds

vector, $\mathbf{w}_h = (w_h(\mathbf{u}_1), \dots, w_h(\mathbf{u}_N))'$, has a multivariate normal distribution

$$\mathbf{w}_h \sim \text{N}(\beta_{0,h}\mathbf{1}_N + U\boldsymbol{\beta}_h, \tau_h^2\Sigma_w), \quad (9)$$

where $\mathbf{1}_N$ is an $n \times 1$ vectors of ones, $U = (\mathbf{u}_1, \dots, \mathbf{u}_N)'$ is an $N \times P$ matrix, and Σ_w is an $N \times N$ matrix whose j_1j_2 th entry is equal to $\Omega(\mathbf{u}_{j_1}, \mathbf{u}_{j_2})$. To facilitate the posterior sampling scheme in Section 4, we transform the problem via a basis expansion (Wood, 2013). Let $\Sigma_w = QDQ'$ be the eigenvalue decomposition of Σ_w , where Q is an $N \times N$ orthogonal matrix whose columns are the eigenvectors of Σ_w , and D is a diagonal matrix containing the eigenvalues of Σ_w . Define $U^\dagger = (\mathbf{1}_N, U, QD^{1/2})$ by columnwise concatenation, let $\boldsymbol{\beta}_h^{\text{GP}}$ be an $N \times 1$ vector, and $\boldsymbol{\beta}_h^\dagger = (\beta_{0,h}, \boldsymbol{\beta}'_h, \boldsymbol{\beta}_h^{\text{GP}'})'$. The first column of U^\dagger is a vector of ones, the next P columns of U^\dagger (equal to U) are the original covariates, while the last N columns (equal to $QD^{1/2}$) are basis functions, with associated coefficients $\boldsymbol{\beta}_h^{\text{GP}}$. For computational convenience and parsimony, we truncate the basis expansion to the first $B < N$ basis functions, so that U^\dagger is $N \times (P + B + 1)$ and $\boldsymbol{\beta}_h^\dagger$ is $(P + B + 1) \times 1$. Equation (9) now takes the form

$$\begin{aligned} \mathbf{w}_h &= U^\dagger \boldsymbol{\beta}_h^\dagger, \\ \boldsymbol{\beta}_h^{\text{GP}} &\sim \text{N}(\mathbf{0}_B, \tau_h^2 I_B), \end{aligned} \quad (10)$$

where $\mathbf{0}_B$ is an $B \times 1$ vector of zeros, and I_B is the $B \times B$ identity matrix. This basis expansion, combined with the interpretation via sequential logits given in the previous section, facilitate the development of the posterior sampling scheme presented in Section 4.

The expression for $v_h(\mathbf{u}_j)$ becomes $\text{logit } v_h(\mathbf{u}_j) = \mathbf{u}_j^{\dagger'} \boldsymbol{\beta}_h^\dagger$, where $\mathbf{u}_j^{\dagger'}$ is the j th row of U^\dagger . The prior placed on $(\beta_{0,h}, \boldsymbol{\beta}'_h)$ is $\text{N}(\boldsymbol{\mu}_\beta, \Sigma_\beta)$, where $\boldsymbol{\mu}_\beta$ is a $(P+1) \times 1$ vector and Σ_β is a $(P+1) \times (P+1)$ covariance matrix, so that

$$\boldsymbol{\beta}_h^\dagger \sim \text{N} \left(\begin{pmatrix} \boldsymbol{\mu}_\beta \\ \mathbf{0}_B \end{pmatrix}, \begin{pmatrix} \Sigma_\beta & 0 \\ 0 & \tau_h^2 I_B \end{pmatrix} \right).$$

Finally, to complete the model specification, we assign τ_h a half- t distribution (Gelman, 2006) with

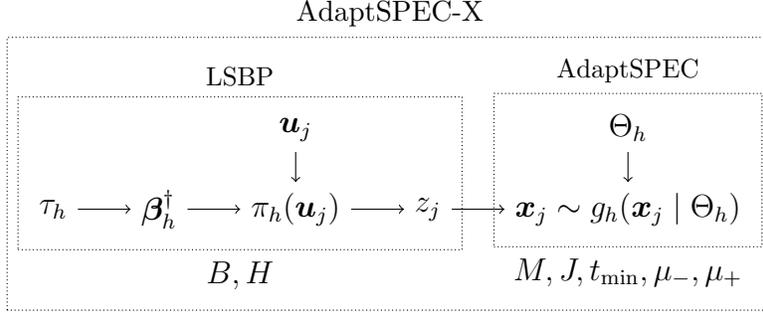


Figure 1: A graphical representation of AdaptSPEC-X. The bottom row lists the main hyperparameters.

density

$$p(\tau_h) \propto \left(1 + \frac{1}{\nu_\tau} \left(\frac{\tau_h}{A_\tau}\right)^2\right)^{-(\nu_\tau+1)/2}, \tau_h > 0,$$

where A_τ and ν_τ are scale and degrees of freedom parameters, respectively. As described in Appendix A in the supplementary material, the half- t distribution can be expressed as a scale mixture of inverse Gamma distributions, which simplifies the sampling of τ_h . For the results in this paper, we set $\boldsymbol{\mu}_\beta = \mathbf{0}$, $\Sigma_\beta = 100I_{P+1}$, $\nu_\tau = 3$ and $A_\tau = 10$.

Figure 1 displays a graphical summary of the model in Equations (5) to (10), showing the dependence between the data, covariates, parameters, and hyperparameters.

3.3 Missing values

The model can accommodate missing values by exploiting the fact that the Whittle likelihood describes a multivariate normal distribution (Whittle, 1953). Under the assumption that the missing data mechanism is ignorable (see Rubin, 1976, for minimal sufficient conditions under which this is true), we show below that this property of the Whittle likelihood implies that the conditional distribution of the missing values is also multivariate normal, and we use this fact to accommodate missing values by integrating them out as part of the MCMC scheme in Section 4. Ignorability of the missing data mechanism is sufficient for inference on the model parameters, but further assumptions, such as that the data are missing completely at random (MCAR; see Rubin, 1976), are required if inference is made on the missing values themselves.

For ease of exposition, in this section we first return to the case where \boldsymbol{x} is a single stationary time series with spectral density f , then later describe how this is extended to multiple nonsta-

tionary time series. Define the $n \times n$ matrix V with entries $V_{tk} = \frac{1}{\sqrt{n}} \exp(-2\pi i(t-1)\omega_k)$ for $t = 1, \dots, n$ and $k = 0, \dots, n-1$. It then follows that $V'(\mathbf{x} - \mu) = (d_1, \dots, d_n)'$, where d_k is given in (2). Noting that V is a unitary matrix, (1) may be rewritten as

$$p(\mathbf{x} \mid \mu, f) = \frac{1}{(2\pi)^{n/2}} |R|^{1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mu)' V R V^* (\mathbf{x} - \mu) \right\},$$

where $R = \text{diag}(\mathbf{r})$, $\mathbf{r} = (1/f(\omega_1), \dots, 1/f(\omega_n))'$, and V^* is the conjugate transpose of V . The precision matrix $\Lambda = V R V^*$ is symmetric and circulant, and is thus defined by its first column, with entries $\Lambda_{t,1} = \frac{1}{n} \sum_{k=1}^n \frac{1}{f(\omega_k)} e^{-2\pi i(t-1)\omega_k}$ (see Appendix B in the supplementary material for a derivation). Suppose some values of \mathbf{x} are missing, and write $\mathbf{x} = (\mathbf{x}'_{\text{mis}}, \mathbf{x}'_{\text{obs}})'$, where \mathbf{x}_{mis} and \mathbf{x}_{obs} are the missing and observed values, respectively. From standard multivariate normal conditioning results and under the assumption of ignorability,

$$(\mathbf{x}_{\text{mis}} \mid \mathbf{x}_{\text{obs}}, \mu, f) \sim N(\mu_{\text{mis}|\text{obs}}, \Lambda_{\text{mis}|\text{obs}}^{-1}), \quad (11)$$

where

$$\begin{aligned} \mu_{\text{mis}|\text{obs}} &= \mu - \Lambda_{\text{mis},\text{mis}}^{-1} \Lambda_{\text{mis},\text{obs}} (\mathbf{x}_{\text{obs}} - \mu), \\ \Lambda_{\text{mis}|\text{obs}} &= \Lambda_{\text{mis},\text{mis}}, \end{aligned} \quad (12)$$

and the quantities in Equation (12) can be obtained from expressing Λ as

$$\Lambda = \begin{bmatrix} \Lambda_{\text{mis},\text{mis}} & \Lambda_{\text{mis},\text{obs}} \\ \Lambda'_{\text{mis},\text{obs}} & \Lambda_{\text{obs},\text{obs}} \end{bmatrix} = \begin{bmatrix} V_{\text{mis}} R V_{\text{mis}}^* & V_{\text{mis}} R V_{\text{obs}}^* \\ V_{\text{obs}} R V_{\text{mis}}^* & V_{\text{obs}} R V_{\text{obs}}^* \end{bmatrix}, \quad (13)$$

in which V_{mis} and V_{obs} are matrices made up of the rows of V corresponding to the missing and observed times, respectively. The quantities in Equation (13) can be computed efficiently by the fast Fourier transform. When simulating from Equation (11), the most computationally intensive step is the inversion, $\Lambda_{\text{mis},\text{mis}}^{-1}$, in Equation (12). In a general framework for spectral estimation with stationary time series, Guinness (2019) describes computationally efficient methods for missing data imputation that could be used to simulate from Equation (11). For this article we compute $\Lambda_{\text{mis},\text{mis}}^{-1}$ the usual way, i.e., via its Cholesky decomposition.

Missing values may be accommodated in AdaptSPEC by partitioning the data in each segment, $\mathbf{x}_{s,m}$, into missing and observed times as above, and sampling from (11) for each segment as part of the MCMC scheme. As shown in the next section, this is extended to the full AdaptSPEC-X model by conditioning on the latent indicators \mathbf{z} .

The Whittle likelihood is used in AdaptSPEC-X in two ways: first, as a nonparametric technique based on its asymptotic properties (see Section 2 and RWS12), and second, in the assumption that the missing values follow a multivariate normal distribution. We justify the latter assumption in several ways. First, Guinness (2019), who makes a similar assumption regarding missing values, found through both theory and numerical experiments that the assumption did not have a deleterious effect on spectral estimation. Second, it is hard to avoid making some assumption about the distribution of the missing values, and multivariate normality is arguably the most parsimonious choice, as it requires only the first two central moments to be specified (the minimum necessary to have mean and spectrum consistent with the observed data). Finally, as shown above and in Guinness (2019), the assumption is computationally convenient.

4 Sampling scheme

Define $\mathbf{z} = (z_1, \dots, z_N)'$, $\mathbf{x}^{\text{all}} = (\mathbf{x}'_1, \dots, \mathbf{x}'_N)'$, $\boldsymbol{\beta}^\dagger = \{\boldsymbol{\beta}^\dagger_1, \dots, \boldsymbol{\beta}^\dagger_{H-1}\}$, $\Theta = \{\Theta_1, \dots, \Theta_H\}$, and $\boldsymbol{\tau} = (\tau_1, \dots, \tau_{H-1})'$. Let $\mathbf{x}^{\text{all}} = (\mathbf{x}^{\text{all}}_{\text{mis}} \mathbf{x}^{\text{all}}_{\text{obs}})'$ be the decomposition of \mathbf{x}^{all} into missing and observed times, respectively. Values produced in steps 2, 3, 4, and 5 of the MCMC sampling scheme below are indicated by the superscript $[l + 0.5]$, and are then used in a label swapping move in Step 6 to produce the $(l + 1)$ th iteration. As described below, this move improves convergence of the following sampling scheme.

Step 1. $(\mathbf{x}^{\text{all}[l+1]}_{\text{mis}} \mid \boldsymbol{\beta}^\dagger[l], \Theta[l], \mathbf{z}^{[l]}, \mathbf{x}^{\text{all}}_{\text{obs}})$ as per Section 3.3.

Step 2. $(\Theta_h^{[l+0.5]} \mid \mathbf{x}^{\text{all}[l+1]}_{\text{mis}}, \mathbf{z}^{[l]}, \mathbf{x}^{\text{all}}_{\text{obs}})$ for each $h = 1, \dots, H$. This step is potentially transdimensional, and uses the reversible-jump MCMC scheme of RWS12, with two modifications. The first one samples the segment means, $\boldsymbol{\mu}_m^h$, as described in Appendix A.1.1 in the supplementary material. The second modification incorporates a Riemann manifold Hamiltonian Monte Carlo (RMHMC for short, see Girolami and Calderhead, 2011) step to accelerate convergence, as described in Appendix A.1.2.

Step 3. $(\mathbf{z}^{[l+0.5]} \mid \boldsymbol{\beta}^{\dagger[l]}, \Theta^{[l+0.5]}, \mathbf{x}_{\text{mis}}^{\text{all}[l+1]}, \mathbf{x}_{\text{obs}}^{\text{all}})$.

Step 4. $(\boldsymbol{\beta}_h^{\dagger[l+0.5]} \mid \boldsymbol{\tau}^{[l]}, \mathbf{z}^{[l+0.5]})$ for each $h = 1, \dots, H$. This uses the Polya-Gamma data augmentation scheme developed by Polson et al. (2013), as applied to the LSBP by Rigon and Durante (2021).

Step 5. $(\tau_h^{[l+0.5]} \mid \boldsymbol{\beta}^{\dagger[l+0.5]})$ for each $h = 1, \dots, H$.

Step 6. $(\Theta^{[l+1]}, \mathbf{z}^{[l+1]}, \boldsymbol{\beta}^{\dagger[l+1]}, \boldsymbol{\tau}^{[l+1]} \mid \mathbf{x}_{\text{mis}}^{\text{all}[l+1]}, \mathbf{x}_{\text{obs}}^{\text{all}})$ using a label swapping step, described below.

The details of steps 2, 3, 4, and 5 are presented in Appendix A in the supplementary material, while Step 1, for \mathbf{x}_{mis} , is described in Section 3.3. For Step 6, we adapt a label swapping move from Hastie et al. (2015), who find that it improves convergence in the context of MCMC samplers for Dirichlet process mixture models. The label swapping step is composed of the following substeps:

Step 6a. Pick uniformly at random components $h_1, h_2 \in \{1, \dots, H\}$, $h_1 < h_2$, to swap.

Step 6b. Construct proposal component indicators \mathbf{z}^{swap} and Θ^{swap} such that

$$\mathbf{z}_j^{\text{swap}} = \begin{cases} h_2 & \text{if } z_j^{[l+0.5]} = h_1, \\ h_1 & \text{if } z_j^{[l+0.5]} = h_2, \\ z_j^{[l+0.5]} & \text{otherwise,} \end{cases} \quad \Theta_h^{\text{swap}} = \begin{cases} \Theta_{h_2}^{[l+0.5]} & \text{if } h = h_1, \\ \Theta_{h_1}^{[l+0.5]} & \text{if } h = h_2, \\ \Theta_h^{[l+0.5]} & \text{otherwise.} \end{cases}$$

Step 6c. Construct proposal $\boldsymbol{\tau}^{\text{swap}}$ by setting

$$\tau_h^{\text{swap}} = \begin{cases} \tau_{h_2}^{[l+0.5]} & \text{if } h = h_1 \\ \tau_{h_1}^{[l+0.5]} & \text{if } h = h_2 \\ \tau_h^{[l+0.5]} & \text{otherwise,} \end{cases}$$

and sample proposal $\boldsymbol{\beta}_{h_1}^{\dagger\text{swap}}, \boldsymbol{\beta}_{h_2}^{\dagger\text{swap}}$ from

$$q(\boldsymbol{\beta}_h^{\dagger\text{swap}} \mid \mathbf{z}^{[l+0.5]}, \boldsymbol{\tau}^{[l+0.5]}) \sim \text{N}(\mu_h^{\text{mode}}, \Sigma_h^{\text{mode}}),$$

where μ_h^{mode} and Σ_h^{mode} are the mode and the negative inverse of the Hessian, respectively, of $\log p(\beta_h^{\dagger\text{swap}} \mid \mathbf{z}^{\text{swap}}, \boldsymbol{\tau}^{\text{swap}})$.

Step 6d. Accept the swap with probability equal to the Metropolis-Hastings ratio

$$\min \left\{ 1, \frac{p(\beta^{\dagger\text{swap}}, \mathbf{z}^{\text{swap}}, \Theta^{\text{swap}}, \boldsymbol{\tau}^{\text{swap}} \mid \mathbf{x}_{\text{mis}}^{\text{all}[l+1]}, \mathbf{x}_{\text{obs}}^{\text{all}})}{p(\beta^{\dagger[l+0.5]}, \mathbf{z}^{[l+0.5]}, \Theta^{[l+0.5]}, \boldsymbol{\tau}^{[l+0.5]} \mid \mathbf{x}_{\text{mis}}^{\text{all}[l+1]}, \mathbf{x}_{\text{obs}}^{\text{all}})} \frac{q(\beta_{h_1}^{\dagger[l+0.5]} \mid \mathbf{z}^{\text{swap}}, \boldsymbol{\tau}^{\text{swap}})}{q(\beta_{h_1}^{\dagger\text{swap}} \mid \mathbf{z}^{[l+0.5]}, \boldsymbol{\tau}^{[l+0.5]})} \frac{q(\beta_{h_2}^{\dagger[l+0.5]} \mid \mathbf{z}^{\text{swap}}, \boldsymbol{\tau}^{\text{swap}})}{q(\beta_{h_2}^{\dagger\text{swap}} \mid \mathbf{z}^{[l+0.5]}, \boldsymbol{\tau}^{[l+0.5]})} \right\}.$$

If accepted, set $\beta^{\dagger[l+1]}, \mathbf{z}^{[l+1]}, \Theta^{[l+1]}$ and $\boldsymbol{\tau}^{[l+1]}$ equal to $\beta^{\dagger\text{swap}}, \mathbf{z}^{\text{swap}}, \Theta^{\text{swap}}$ and $\boldsymbol{\tau}^{\text{swap}}$, respectively. Otherwise, $\beta^{\dagger[l+1]}, \mathbf{z}^{[l+1]}, \Theta^{[l+1]}$ and $\boldsymbol{\tau}^{[l+1]}$ are set to $\beta^{\dagger[l+0.5]}, \mathbf{z}^{[l+0.5]}, \Theta^{[l+0.5]}$ and $\boldsymbol{\tau}^{[l+0.5]}$, respectively.

In Step 6b, the labels of the component indicators for the chosen pair h_1, h_2 are swapped, as are the corresponding mixture component parameters Θ_h , leaving the likelihood unchanged. Step 6c swaps the smoothing spline parameters τ_{h_1}, τ_{h_2} and samples new values of β_h^{\dagger} from a normal approximation centered on its conditional mode. The latter new values are necessary because due to the sequential nature of the LSBP, merely swapping the values of $\beta_{h_1}^{\dagger}$ and $\beta_{h_2}^{\dagger}$ is unlikely to result in an acceptable proposal.

AdaptSPEC-X is implemented in the latest version of the R package BayesSpec.² The implementation is in R and C++, and can take advantage of multiple processor cores to reduce the running time of the analysis.

5 Simulation studies

This section reports on simulation results to study the proposed method, as well as to compare it with another method in the literature. Section 5.1 focuses on the advantages of AdaptSPEC-X for joint modeling of multiple nonstationary time series, comparing it to separate applications of AdaptSPEC to multiple time series. To this end, multiple time series with time varying means and spectra at both observed and unobserved covariate values are generated repeatedly. This

²Available from GitHub at <https://github.com/mbertolacci/BayesSpec/>. As of publication, the version of the package on CRAN does not contain AdaptSPEC-X.

simulation demonstrates AdaptSPEC-X’s ability, quantified by mean squared error, to estimate time varying means and spectra at both observed and unobserved covariate values. This study also illustrates AdaptSPEC-X’s ability to provide out-of-sample predictions, which are not available when AdaptSPEC is fitted separately to individual time series. Section 5.2 compares AdaptSPEC-X to the conditional adaptive Bayesian spectrum analysis (CABS) method of Bruce et al. (2018) which is capable of estimating the time-varying spectra of a panel of time series, given a single covariate. Our simulation results demonstrate superior performance of AdaptSPEC-X in terms of mean squared error.

5.1 Joint versus separate modeling

Let $U = (\mathbf{u}_1, \dots, \mathbf{u}_{100})'$ be a 100×2 design matrix corresponding to $N = 100$ subjects, each with two covariates, where the \mathbf{u}_j , $j = 1, \dots, N$, are sampled uniformly from $[0, 1] \times [0, 1]$. Each \mathbf{u}_j is mapped deterministically to a region $r_j \in \{1, 2, 3, 4\}$, according to the plot shown in Figure 2(a), which also includes the locations of the 100 sampled points, denoted by crosses. Four locations are chosen as example time series, marked in green circles, and labeled D1 through to D4 (corresponding to $r_j = 1$ through $r_j = 4$, respectively). Four more locations are marked in red diamonds labeled T1 through to T4 (again for $r_j = 1$ to 4). These values of \mathbf{u} have no corresponding time series and are used as test points to evaluate the predictive inferences. The four different regions correspond to four different data generating processes. Each time series \mathbf{x}_j , within region r_j , $j = 1, \dots, 100$, is a realization of length $n = 256$ from

$$(x_{j,t} - \mu_{r_j,t}) = \phi_{r_j,1,t}(x_{j,t-1} - \mu_{r_j,t-1}) + \phi_{r_j,2,t}(x_{j,t-2} - \mu_{r_j,t-2}) + \epsilon_{j,t}, \quad (14)$$

where $\epsilon_{j,t} \sim N(0, 1)$, and the values of $\mu_{r_j,t}$ and $\phi_{r_j,p,t}$ are given in the following table:

	$t \leq 128$			$t > 128$		
	$\mu_{r_j,t}$	$\phi_{r_j,1,t}$	$\phi_{r_j,2,t}$	$\mu_{r_j,t}$	$\phi_{r_j,1,t}$	$\phi_{r_j,2,t}$
$r_j = 1$	-1.5	1.5	-0.75	-2	-0.8	0
$r_j = 2$	1	-0.8	0	-1	-0.8	0
$r_j = 3$	0	1.5	-0.75	0	1.5	-0.75
$r_j = 4$	1	0.2	0	1	1.5	-0.75

Thus, time series with $r_j = 1$ have two segments with different means and different spectra, those with $r_j = 2$ have two segments with different means but with the same spectra, time series

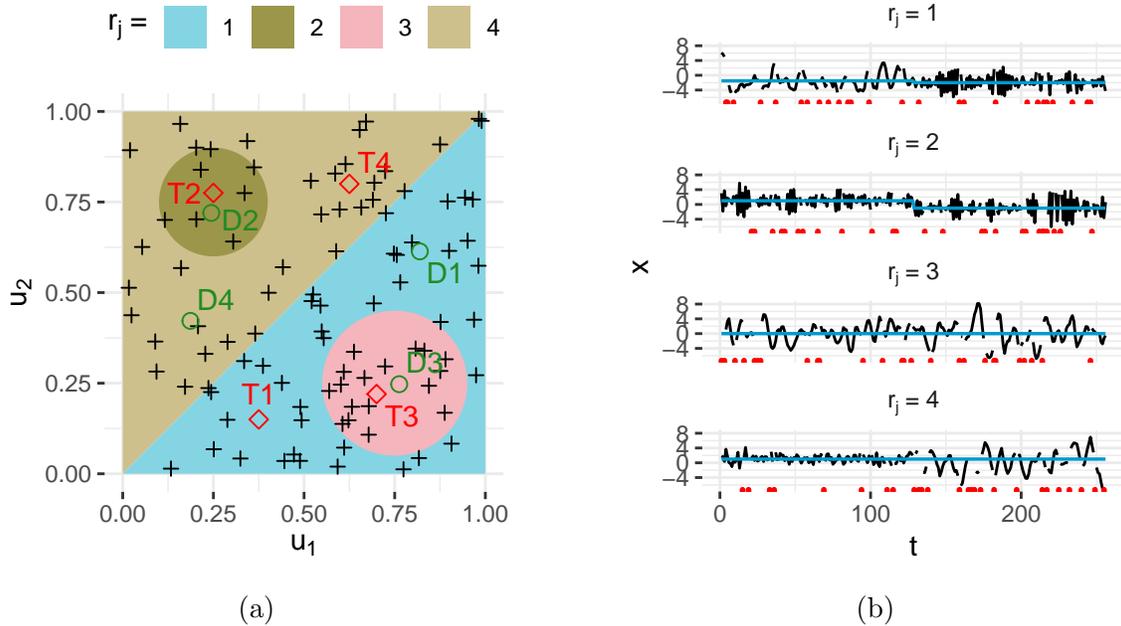


Figure 2: (a) Underlying surface mapping \mathbf{u}_j to r_j for Process (14), where the 100 sampled locations are shown as crosses. Regions, r_j , for $j = 1, \dots, 4$, are identified by colors. Four points are used as examples in the paper and are denoted by green circles and labeled D1 to D4. Four test points labeled T1 to T4 are shown with red diamonds. (b) Example realizations from Process (14) corresponding to $r_j = 1$ at the top through $r_j = 4$ at the bottom. Black lines give the values of the time series, blue lines the underlying time-varying mean $\mu_{r_j,t}$, and red ticks on the bottom axis mark missing values.

with $r_j = 3$ have only one stationary segment, and those with $r_j = 4$ have two segments with the same mean but with different spectra. In each time series, 10% of the times are set as missing. Figure 2(b) displays example realizations from Process (14) for $r_j = 1, 2, 3$ and 4, showing the time series values, underlying time-varying mean, and the times at which values are missing.

We sample 100 replicates from Process (14), and to each fit AdaptSPEC-X using the MCMC sampling scheme of Section 4. We run 50,000 iterations of the MCMC scheme, where the first 10,000 are discarded as burn-in. Each mixture component has $M = 4$, as the maximum number of segments, $t_{\min} = 40$, as the minimum segment length, $J = 25$, as the number of basis functions for the smoothing spline prior on the log spectra (see Section 6.1, where we discuss how the selection of J relates to that of t_{\min}), and $(\mu_-, \mu_+) = (-10, 10)$, as the support of the prior on $\mu_{s,m}^h$. The LSBP is truncated at $H = 25$ components, and has $B = 10$ basis functions. The MCMC scheme for each replicate takes around 1.5 hours to run when using two cores on a computer with an Intel

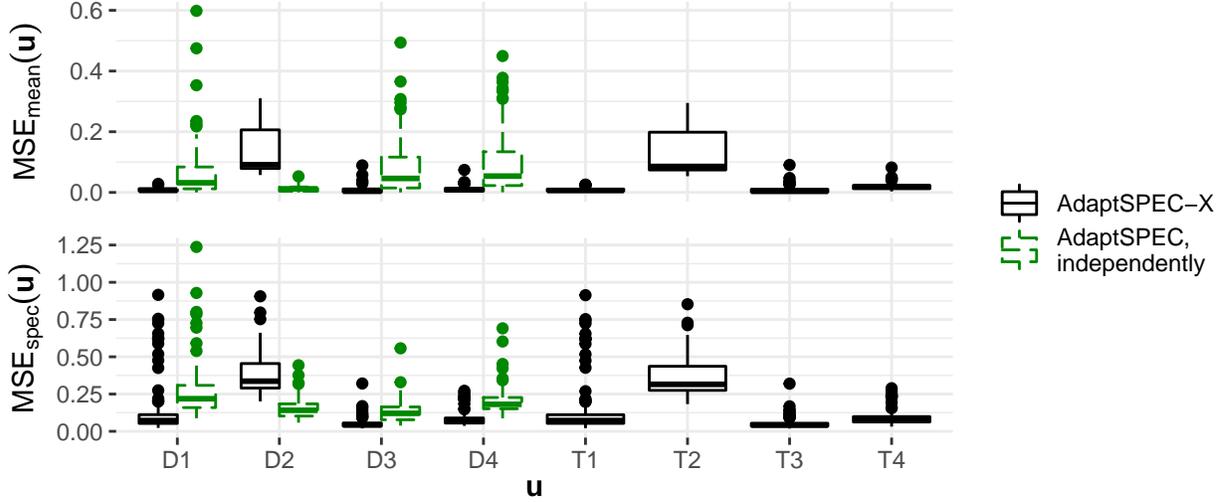


Figure 3: MSE across 100 replications for estimates of mean (top) and spectrum (bottom) for Process (14) at the observed $\mathbf{u} = \text{D1-D4}$ and the unobserved $\mathbf{u} = \text{T1-T4}$ from left to right, respectively. For the observed locations, green boxplots show the MSE from estimates made using independent fits of AdaptSPEC to the replicates.

Core i9-9900K CPU with a clock speed of 3.60 GHz.

To assess the quality of the estimated time-varying mean, we define the mean squared error (MSE) for the mean as

$$\text{MSE}_{\text{mean}}(\mathbf{u}) = \frac{1}{n} \sum_{t=1}^n [\hat{\mu}(t, \mathbf{u}) - \mu(t, \mathbf{u})]^2, \quad (15)$$

where $\hat{\mu}(t, \mathbf{u})$ is the estimate of $\mu(t, \mathbf{u})$, the true time-varying mean at covariates \mathbf{u} . Similarly, we define the MSE for the spectrum as

$$\text{MSE}_{\text{spec}}(\mathbf{u}) = \frac{1}{n} \frac{1}{k_{\max}} \sum_{t=1}^n \sum_{k=1}^{k_{\max}} \left[\log \hat{f} \left(t, \frac{k-1}{2k_{\max}-2}, \mathbf{u} \right) - \log f \left(t, \frac{k-1}{2k_{\max}-2}, \mathbf{u} \right) \right]^2, \quad (16)$$

where $k_{\max} = 128$, and $\log \hat{f}(t, \omega, \mathbf{u})$ is the estimate of $\log f(t, \omega, \mathbf{u})$, the true time-varying log spectral density at location \mathbf{u} .

Figure 3 presents boxplots of MSE_{mean} (top) and MSE_{spec} (bottom), at each observed location, D1–D4, and unobserved test location, T1–T4, from left to right, respectively. We now fit the AdaptSPEC model (RWS12) to each replicate of D1–D4 individually, and the green boxplots in Figure 3 show MSE_{mean} and MSE_{spec} for those fits.

The median MSE_{mean} of AdaptSPEC-X is substantially smaller than that of the individual

AdaptSPEC fits at D1, D3, and D4, and its interquartile range is narrower. Similarly, the median MSE_{spec} of AdaptSPEC-X is smaller than that of individual AdaptSPEC fits at D1, D3, and D4. This demonstrates the benefits of applying a joint model to multiple time series, in that the ability to group them together via the covariates can lead to improved estimation performance. The exception to this is the performance at D2, for which AdaptSPEC-X does worse, and we discuss possible reasons for this below. For the test points T1–T4, the predictive performance of AdaptSPEC-X as measured by MSE_{mean} and MSE_{spec} is comparable to the observed points D1–D4. The test points are unobserved, so individual fits of AdaptSPEC to the observed series do not allow prediction of their time-varying mean and spectra. This is another benefit of jointly modeling the data by including covariates.

AdaptSPEC-X’s estimates of the time-varying mean and spectrum corresponding to the median MSE values are shown in figures 4 and 5, respectively. These qualitatively match the MSE_{mean} and MSE_{spec} scores, in that the estimates for points other than D2 and T2 are visually very close to the truth, while for D2 and T2 some differences are visible.

There are several explanations for the worse performance for D2 and T2 relative to the other points. One reason is that the spatial model has relatively little information to identify the existence of a region: these points belong to region two, which is represented by only eight time series (in contrast to regions one, three, and four, which have 41, 18, and 33 members, respectively), and which has the smallest spatial area in the study. Another reason is that the process mean for region two (equal to 1 in the first half and -1 in the second half) is similar to that of the surrounding region four (constant mean of 1), making it harder to distinguish between the regions. For both reasons, it is not surprising to see worse performance for T2 and D2, which corresponds to good model behavior in the sense that it represents genuine model uncertainty. This is seen in Figure 4, where the $\hat{\mu}(t, \mathbf{u})$ for T2 and D2 in the second half of the time series is shrunk towards that of region four, the enclosing region. It is also worth noting that the median MSE_{mean} and MSE_{spec} for T2 and D2 are small relative to the scales of their mean and log spectrum, respectively. This can also be seen qualitatively by the similarity between the true and estimated mean and spectra in figures 4 and 5.

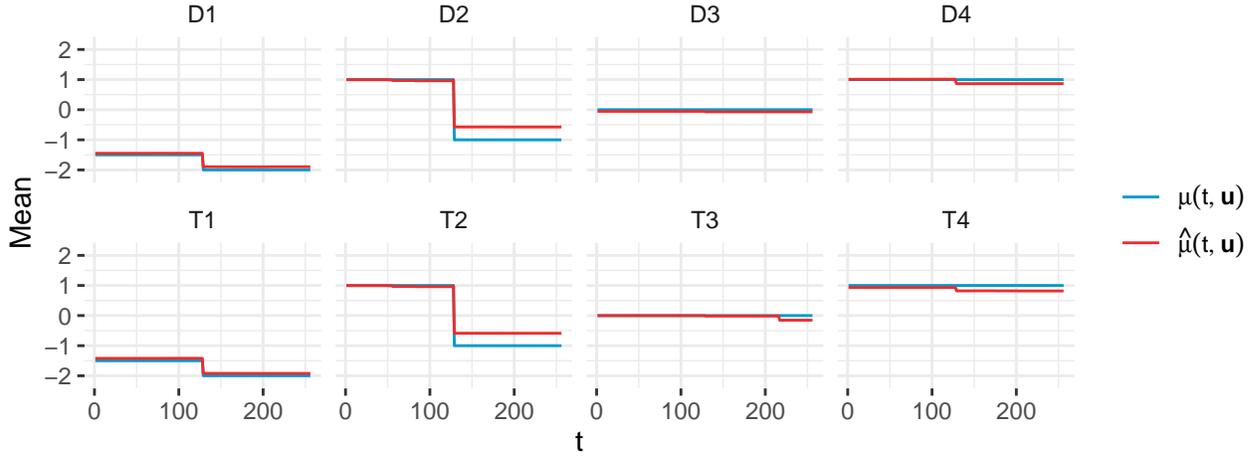


Figure 4: Estimated mean $\hat{\mu}(t, \mathbf{u})$ corresponding to the median $\text{MSE}_{\text{mean}}(\mathbf{u})$ (red) and true mean $\mu(t, \mathbf{u})$ (blue) for Process (14). The first row shows the estimates for $\mathbf{u} = \text{D1-D4}$ from left to right, respectively, while the second row shows the estimates for test points $\mathbf{u} = \text{T1-T4}$.

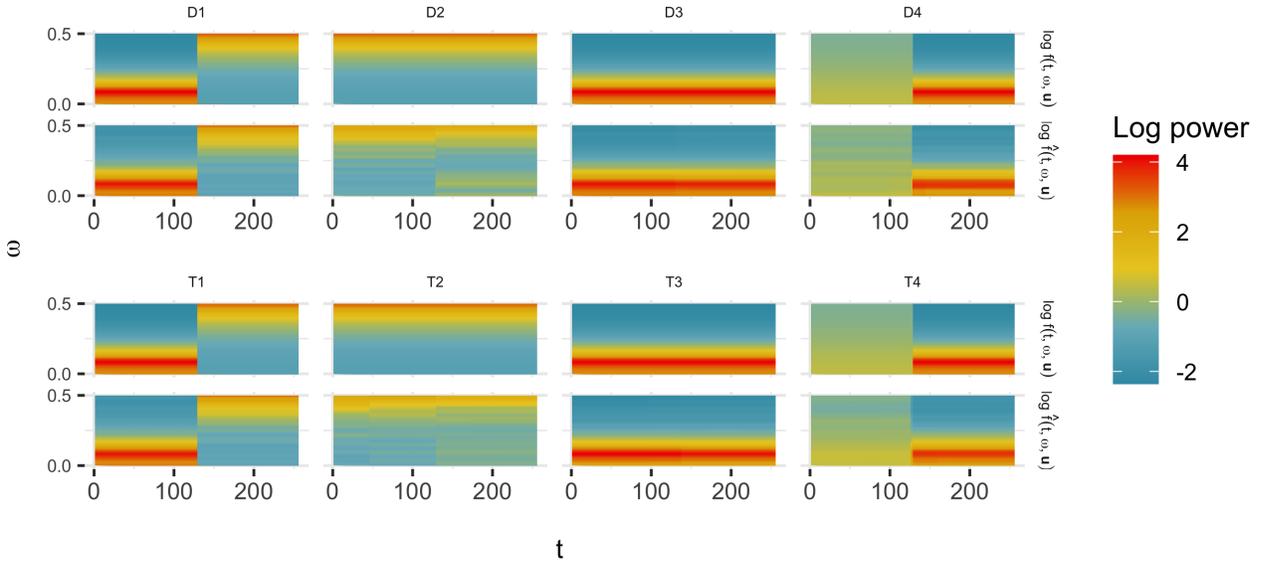


Figure 5: Estimated time-varying log spectra $\log \hat{f}(t, \omega, \mathbf{u})$ corresponding to the median $\text{MSE}_{\text{spec}}(\mathbf{u})$ and true time-varying log spectra $\log f(t, \omega, \mathbf{u})$ for Process (14). The first row shows $\log f(t, \omega, \mathbf{u})$ for $\mathbf{u} = \text{D1-D4}$ from left to right, respectively, while the second row shows the estimates $\log \hat{f}(t, \omega, \mathbf{u})$. The third and fourth rows display the analogous quantities for the test points $\mathbf{u} = \text{T1-T4}$.

5.2 Comparison of AdaptSPEC-X and CABS

CABS is a method for estimating the time-varying spectra of a panel of time series, $\{\mathbf{x}_j : j = 1, \dots, N\}$, for $\mathbf{x}_j = (x_{1,j}, \dots, x_{n,j})'$, where each time series is associated with a single covariate value u_j . It extends AdaptSPEC to multiple time series by partitioning the covariate in the same way that AdaptSPEC partitions the time series, such that the spectral density is constant as a function of the covariate within a ‘rectangle’ in time and covariate space. Unlike CABS, AdaptSPEC-X is able to handle an arbitrary number of covariates, because there is no need to adaptively divide the covariate space. AdaptSPEC-X is richer in features than CABS, but for the purposes of comparison we evaluate AdaptSPEC-X and CABS in a setting that is applicable to both methods. To this end, we reproduce the simulation study presented by [Bruce et al. \(2018\)](#), which examined the ability of CABS to estimate the time-varying spectra of a collection of time series drawn from known processes. The simulation study considers two cases: one in which time series have abrupt changes in their spectra, and another where changes occur slowly. In both cases, there are N time series of length n each, with a single covariate, taking the values $u_j = (j - 1)/(N - 1)$, $j = 1, \dots, N$, that in turn determine the spectra of the corresponding time series. The j th abruptly varying time series is simulated from a piecewise AR(1) processes, $\{x_{t,j} : t = 1, \dots, n\}$, for which

$$x_{t,j} = \begin{cases} -\phi_j x_{t-1,j} + \epsilon_{t,j}, & 1 \leq t \leq n/2 \\ \phi_j x_{t-1,j} + \epsilon_{t,j}, & n/2 < t \leq n \end{cases}, \text{ where } \phi_j = \begin{cases} 0.5, & 0 \leq u_j \leq 0.5 \\ 0.9, & 0.5 < u_j \leq 1 \end{cases} \quad (17)$$

and $\epsilon_{t,j} \sim \text{i.i.d. } N(0, 1)$. The j th slowly varying time series is simulated from a time-varying AR(1) process with

$$x_{t,j} = \phi_{t,j} x_{t-1,j} + \epsilon_{t,j}, \text{ where } \phi_{t,j} = \begin{cases} -0.5 + t/n, & 0 \leq u_j \leq 0.5 \\ -0.9 + 9t/(5n), & 0.5 < u_j \leq 1 \end{cases} \quad (18)$$

and $\epsilon_{t,j} \sim \text{i.i.d. } N(0, 1)$.

[Bruce et al. \(2018\)](#) apply CABS to 100 replicates from processes (17) and (18) over the nine (N, n) combinations, where $N \in \{20, 40, 80\}$ and $n \in \{1000, 2000, 4000\}$, and evaluated its ability to estimate the time-varying log spectra of the time series. Here we repeat this evaluation using AdaptSPEC-X. To facilitate comparison, for quantities that are common to CABS and AdaptSPEC-

X we use the same values: $t_{\min} = 50$ (minimum segment length); $M = 10$ (the maximum number of segments); and $J = 7$ (the number of basis functions for the smoothing spline on the log spectrum; B in the notation of [Bruce et al. \(2018\)](#)). The number of segments in covariate space in CABS is set equal to eight. In AdaptSPEC-X, we set the number of mixture components to $H = 10$, and the number of basis functions in covariate space to $B = 10$. [Bruce et al. \(2018\)](#) run MCMC chains for 5,000 iterations with a warm-up of 1,000 iterations; we run AdaptSPEC-X for 10,000 iterations with a warm-up of 1,000 iterations. For AdaptSPEC-X, running the MCMC scheme took between five minutes (for $N = 20, n = 1000$) and 30 minutes (for $N = 80, n = 4000$) using one core of a desktop with an Intel Core i9-9900K CPU with a clock speed of 3.60 GHz.

To evaluate the performance of CABS at estimating the true time-varying log spectrum, [Bruce et al. \(2018\)](#) use the mean-squared error (MSE), calculated according to equation (16). For a particular combination (n, N) , and underlying process (abrupt or slow), a mean MSE was calculated across the N time series within each of the 100 replicates. The mean and standard deviation of the mean MSE across the replicates was then calculated. Table 1 presents the mean MSE (and standard deviations in parentheses) for both CABS, as reported in [Bruce et al. \(2018\)](#), and AdaptSPEC-X. The ratio of the mean MSE for CABS to the mean MSE for AdaptSPEC-X is also presented, showing that the mean MSE for CABS is between 2 to 5 times as large as that of AdaptSPEC-X when estimating the log spectra of the abruptly varying time series, and between 1.3 to 1.5 times as large for the slowly varying process.

The fact that AdaptSPEC-X outperforms CABS in both simulation settings suggests that the mixture model of AdaptSPEC-X is able to share information between time series more efficiently than the covariate-space partitioning scheme in CABS, despite the fact that the covariate dependence in equations (17) and (18) is abrupt and should lend itself to the way CABS partitions the covariate space. It appears that the covariate-dependent infinite mixture combined with the thin-plate splines used by AdaptSPEC-X are sufficiently flexible to handle this feature.

6 Applications

In this section, we describe two applications. Section 6.1 considers Australian rainfall, while incidence counts of measles in the United States are analyzed in Section 6.2.

n	N	Abruptly varying			Slowly varying		
		CABS	AdaptSPEC-X	Ratio	CABS	AdaptSPEC-X	Ratio
1000	20	0.1190 (0.0580)	0.0339 (0.0113)	3.5	0.0353 (0.0052)	0.0245 (0.0027)	1.4
1000	40	0.1130 (0.0598)	0.0266 (0.0071)	4.3	0.0260 (0.0031)	0.0184 (0.0021)	1.4
1000	80	0.1079 (0.0557)	0.0201 (0.0027)	5.4	0.0206 (0.0026)	0.0135 (0.0015)	1.5
2000	20	0.0748 (0.0252)	0.0316 (0.0084)	2.4	0.0249 (0.0035)	0.0180 (0.0017)	1.4
2000	40	0.0755 (0.0274)	0.0242 (0.0039)	3.1	0.0185 (0.0017)	0.0130 (0.0010)	1.4
2000	80	0.0702 (0.0230)	0.0198 (0.0039)	3.5	0.0148 (0.0013)	0.0101 (0.0009)	1.5
4000	20	0.0617 (0.0121)	0.0300 (0.0076)	2.1	0.0177 (0.0019)	0.0135 (0.0010)	1.3
4000	40	0.0610 (0.0133)	0.0249 (0.0064)	2.5	0.0137 (0.0012)	0.0101 (0.0007)	1.4
4000	80	0.0607 (0.0135)	0.0192 (0.0032)	3.2	0.0110 (0.0009)	0.0082 (0.0007)	1.3

Table 1: The mean (standard deviation) of the MSEs across the 100 replicates for estimating the log spectra of the time series in the simulation study using CABS and AdaptSPEC-X. The ratio of the mean MSE for CABS to the mean MSE for AdaptSPEC-X is also presented. The values for CABS are taken from [Bruce et al. \(2018\)](#).

6.1 Australian rainfall data

Rainfall is governed in large part by cyclical processes, in particular the seasonal cycle driven by the Earth’s orbit around the sun. It has therefore historically been a natural application area for spectral methods. These have been used to study interannual variation (see, among many others, [Alter, 1924](#); [Rajagopalan and Lall, 1998](#); [Ansell et al., 2000](#)), intraannual or intraseasonal variation ([Joshi and Pandey, 2011](#)), and the connections between rainfall and other climatic processes ([Rajagopalan and Lall, 1998](#); [Ansell et al., 2000](#)). Here we focus on identifying changes in both the mean and spectrum of Australian rainfall. As part of a report on climate change tendered by several Australian government agencies, [Cai et al. \(2007\)](#) find that, since 1950, the Australian north has seen increased annual rainfall, while the southeast and southwest have experienced the opposite. The causes of these trends have been the subject of study and debate (see, among others, [Hope et al., 2006](#); [Ummenhofer et al., 2009](#); [Pook et al., 2012](#); [Risbey et al., 2013](#)). Apart from trends in overall rainfall, several authors have reported relative increases in heavy rainfall events, indicating changes in the variability of rainfall ([Cai et al., 2007](#); [Gallant et al., 2013](#)). We contribute to this literature by using AdaptSPEC-X to analyze the time-varying mean and spectrum of Australian rainfall from sites dispersed over a wide spatial field, addressing simultaneously the question of whether changes have occurred in rainfall levels and rainfall variability.

We use data from [Bertolacci et al. \(2019\)](#), who studied the climatology of Australian daily

rainfall using measurements from 17,606 sites across the continent. In particular, we use 151 of these sites characterized by having long and nearly contiguous rainfall records, the locations of which are displayed in Figure 6(a). These sites are among those identified by Lavery et al. (1992) as having high quality records suitable for monitoring and assessing climate change. The raw time series are daily, and observations are typically made at 9 am local time, recording the total rainfall in millimeters (mm) for the previous 24 hours. Aggregation to monthly data is performed by calculating the average daily rainfall for the month. To avoid artifacts, we consider as missing any month with fewer than fifteen days of measurements available (that is, not missing). We choose to aggregate the data to monthly intervals in order to focus the analysis on climatic changes to the spectrum, which occur on time scales much longer than a month. A secondary benefit is that doing so requires much fewer basis functions for the smoothing splines on the log spectrum. This is due to the fact that the dominant feature of the data is the seasonal cycle at a period of 365.25 days for the daily data, and such a low frequency spike can only be accommodated by a high value of J .

The resulting time series span the 1,078 months from September, 1914 to June, 2004 (inclusive), for a total of 162,778 observations, of which 4,095 are missing. The smallest possible measurement is 0 mm, corresponding to no rainfall for the month; this is true for 9,933 months. Time series for four example sites are displayed in Figure 6(b), and their locations are marked on inset maps (also marked in green in Figure 6(a)). The four time series span a wide range of average rainfall levels from around 1mm at site 47053 to 4mm at site 14042. They also exhibit varying levels of seasonality, where sites 10525 and 14042 have highly seasonal rainfall, while rainfall at sites 47053 and 69018 is less seasonal.

We fit AdaptSPEC-X to these data, setting $\mathbf{u}_j = (\text{lon}_j, \text{lat}_j)$, the longitude and latitude of the sites. Each mixture component has $t_{\min} = 60$ (5 years), chosen to ensure several observations of the dominant annual cycle are available. This constrains the maximum number of segments to be $M_{\max} = 17$. We choose $J = 60$ for the number of basis functions for the smoothing spline prior on the log spectra. Preliminary fits of AdaptSPEC to the individual time series indicated that using $J > 60$ makes little difference with respect to the estimated log spectra compared to $J = 60$. Also, since $t_{\min} = 60$, for short segments increasing the value of J is not likely to be beneficial. The support of the prior on $\mu_{s,m}^h$ is set to $(\mu_-, \mu_+) = (0, 30)$. The lower bound reflects the positivity of rainfall, while the upper bound is three times larger than the largest empirical mean attained in

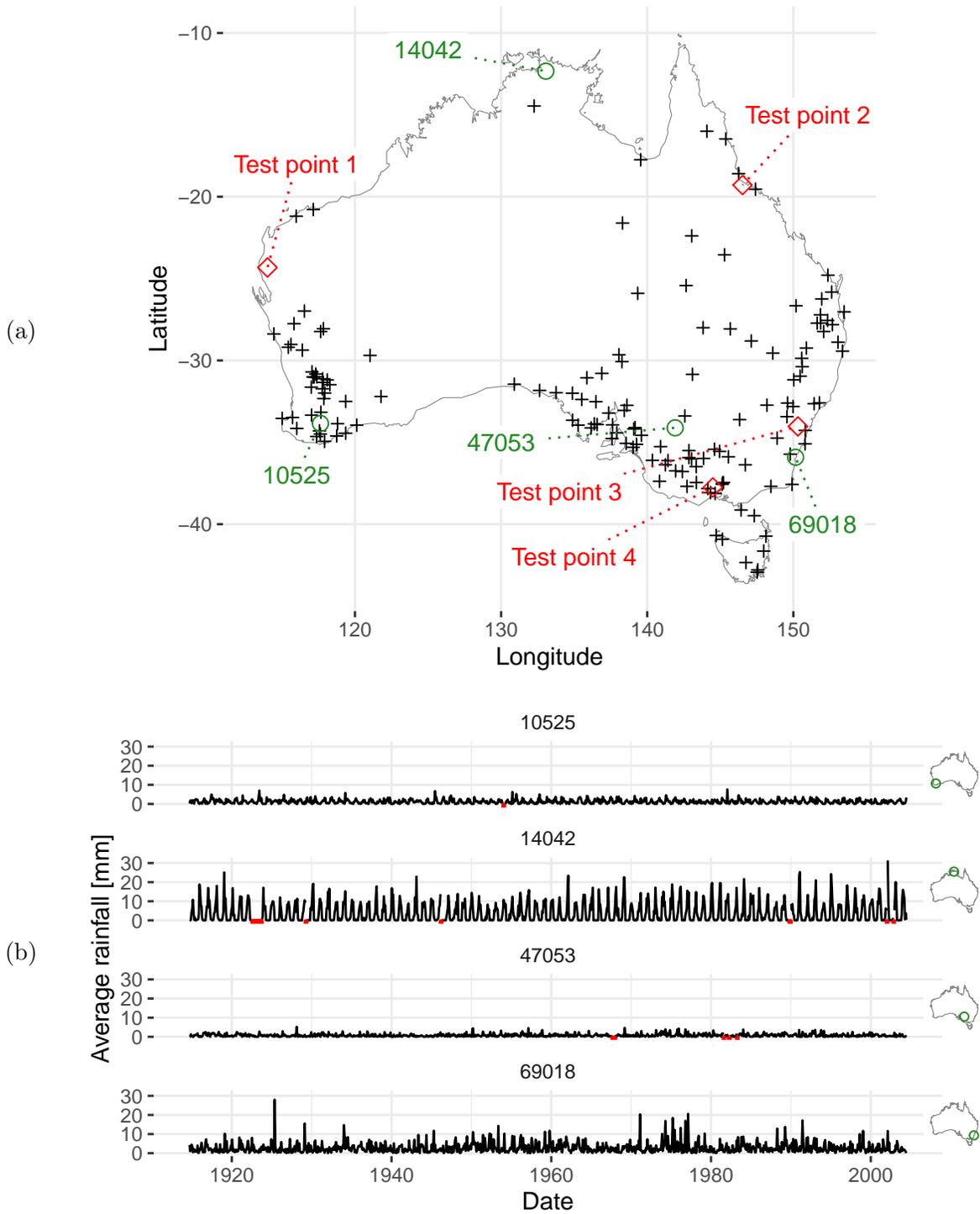


Figure 6: (a) Locations of the 151 rainfall sites. Four example sites are marked with green circles, and four test locations are marked with red diamonds. (b) Monthly rainfall records at the four example sites, whose locations are indicated by inset maps. Missing values are marked with red ticks on the bottom axis.

any 5-year period in the data. The LSBP is truncated at $H = 25$ components, as higher values of H made no difference to the estimates. Finally, the thin-plate GP prior for the LSBP has $B = 20$ basis functions, which captures more than 95% of the variation implied by the prior. In addition to the locations with measurements, we estimate the predictive time-varying mean and spectrum at four locations on the Australian landmass without measurements. These locations are indicated by red diamonds in Figure 6(a). The MCMC scheme ran for six hours using four cores of an Intel Core i9-9900K CPU with a clock speed of 3.60 GHz.

The estimated time varying means $\hat{\mu}(t, \mathbf{u})$ are shown in Figure 7, where the first two rows display the estimates for the four example sites, and the last two rows present predictive estimates for the four test locations which do not have observations. The corresponding estimated time varying log spectra $\log \hat{f}(t, \omega, \mathbf{u})$ are shown in Figure 8. Figures 7 and 8 show that the model has accommodated considerable variation between sites, with mean monthly rainfall ranging from 1mm at site 47053 to 3.5mm at 14042, and with spectra similarly varied. The dominant period at all sites is 12 months, corresponding to the Earth’s orbit around the sun. There is also considerable power at lower frequencies, indicating long-term dependence. The predictive means and spectra at unobserved sites match qualitative expectations. For example, among the unobserved locations, Test Point 2 has the heaviest mean rainfall and the most power in its spectrum, reflecting its location in the tropics.

Two major droughts occurred during the study period: the World War II drought of 1937–1945, and the Millennium drought that started in 1996 and was still ongoing by the end of the study period in 2004 (Ummenhofer et al., 2009). Table 2 presents estimated posterior probabilities of changes in the mean $\mu(t, \mathbf{u})$ or variance $\sigma^2(t, \mathbf{u}) = 2 \int_0^{1/2} f(t, \omega, \mathbf{u}) d\omega$ around these times. Specifically, it shows that $\hat{P}(\mu_{1940} < \mu_{1950}) > 0.9$ and $\hat{P}(\sigma_{1940}^2 < \sigma_{1950}^2) > 0.9$ at all four sites. However, $\hat{P}(\mu_{1940} < \mu_{1930}) < 0.7$ and $\hat{P}(\sigma_{1940}^2 < \sigma_{1930}^2) < 0.57$ except for site 69018 for which these probabilities are greater than 0.8. The Millennium drought is associated with drops in $\mu(t, \mathbf{u})$, $\sigma^2(t, \mathbf{u})$ or both at sites 14042, 47053 and 69018 as can be seen from the estimated probabilities at these sites: $\hat{P}(\mu_{2004} < \mu_{1990}) > 0.9$ and $\hat{P}(\sigma_{2004}^2 < \sigma_{1990}^2) > 0.93$. Site 10525 in the southwest of the continent does not exhibit a drop with probability greater than 0.9, consistent with the fact that the drought principally affected southeastern Australia (Ummenhofer et al., 2009).

Cai et al. (2007) report large trends in rainfall since 1950. For southeast Australia, they

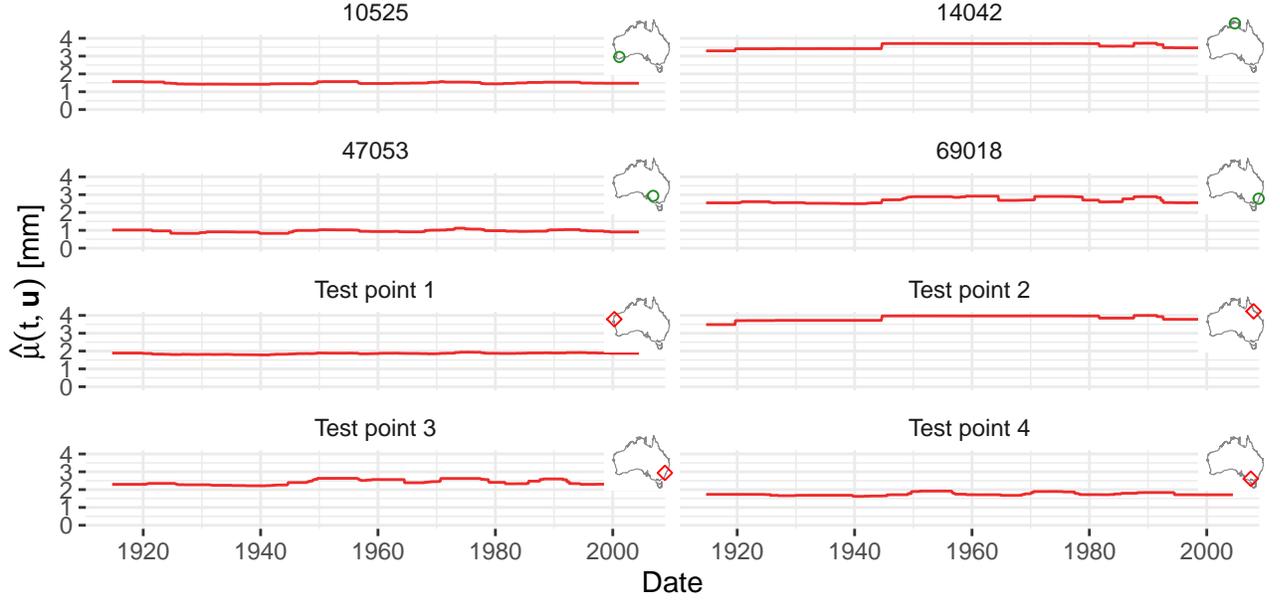


Figure 7: Estimated time-varying means $\hat{\mu}(t, \mathbf{u})$ for four monthly rainfall sites (first two rows), and four locations without observations (last two rows).

Event	$\hat{p}(\cdot \mathbf{x})$	Site (\mathbf{u})			
		10525	14042	47053	69018
WW2 drought	$\mu(1940-01, \mathbf{u}) < \mu(1930-01, \mathbf{u})$	0.540	0.442	0.697	0.801
	$\sigma^2(1940-01, \mathbf{u}) < \sigma^2(1930-01, \mathbf{u})$	0.574	0.391	0.508	0.833
	$\mu(1940-01, \mathbf{u}) < \mu(1950-01, \mathbf{u})$	0.917	0.980	0.997	0.993
	$\sigma^2(1940-01, \mathbf{u}) < \sigma^2(1950-01, \mathbf{u})$	0.997	0.996	1.000	0.997
Millenium drought	$\mu(2004-01, \mathbf{u}) < \mu(1990-01, \mathbf{u})$	0.878	0.901	0.974	0.942
	$\sigma^2(2004-01, \mathbf{u}) < \sigma^2(1990-01, \mathbf{u})$	0.747	0.930	0.999	0.978
Long term	$\mu(2004-01, \mathbf{u}) < \mu(1950-01, \mathbf{u})$	0.820	0.906	0.939	0.973
	$\sigma^2(2004-01, \mathbf{u}) < \sigma^2(1950-01, \mathbf{u})$	0.961	0.955	1.000	0.989

Table 2: Estimated posterior probabilities $\hat{p}(\cdot | \mathbf{x})$ of various events for monthly rainfall at the four example sites. The events correspond to the WW2 drought (first four rows), the Millenium drought (second two rows), and long term change (last two rows). For each event and site (\mathbf{u}), the table presents probabilities that the mean $\mu(t, \mathbf{u})$ and the variance $\sigma^2(t, \mathbf{u})$ changed before or after the event.

report reductions in annual rainfall corresponding to 10–15mm/decade for the site 47053 and 50mm/decade for 69018. Consistent with this, Table 2 shows that, for these sites, AdaptSPEC-X estimates that both $\mu(t, \mathbf{u})$ and $\sigma^2(t, \mathbf{u})$ declined between January, 1950 and January, 2004 with probability greater than 0.94. The estimated drop in $\mu(t, \mathbf{u})$ for site 47053 corresponds to a reduction of 1–16mm/decade³ (10th–90th percentile), consistent with Cai et al.’s estimate. However, for site 69018, the estimated reduction is 6–42mm/decade, substantially less than 50mm/decade. For the site 14042 in the tropical north, Cai et al. estimate increases of around 40mm/decade since 1950, while the estimates of Table 2 indicate a decline over the same period. Finally, Cai et al. report a decline of 20–30mm/decade in the region of southwest Australia containing the site 10525, but Table 2 does not indicate a significant change in $\mu(t, \mathbf{u})$ at this location (though the variance does appear to have declined). In contrast to Cai et al. and Gallant et al. (2013), the reduction in $\sigma^2(t, \mathbf{u})$ since 1950 at all locations suggests that rainfall variability has declined. This could have resulted from the use of different definitions of variability, e.g., counts of extreme events as in Cai et al. (2007), versus our use of change in variance.

6.2 Measles incidence in the United States

Measles is a highly contagious disease that causes fever, cough, runny nose, and a rash (Moreno, 2018). Complications of measles can include pneumonia, deafness, or death (Moreno, 2018). Prior to the licensing of a vaccine in 1963, the incidence rate of measles averaged 318 cases per 100,000 population per year, with outbreaks occurring annually or every other year (van Panhuis et al., 2013). After vaccine licensure, incidence declined dramatically, and endemic measles transmission was declared eliminated from the United States in 2000 (Katz and Hinman, 2004). As part of a study on the impact of vaccination for a variety of contagious diseases, van Panhuis et al. (2013) present a unique data set by digitizing weekly surveillance reports from the United States of several nationally notifiable diseases, including measles, and have made these data available online at the Project Tycho website⁴. In this section we analyze Project Tycho’s measles data using AdaptSPEC-X.

The data comprise weekly time series of measles incidence for each state, reporting the weekly

³Calculated as $10 \times 365.25 \times \text{difference in daily average} / (2004 - 1950)$

⁴<https://www.tycho.pitt.edu/>

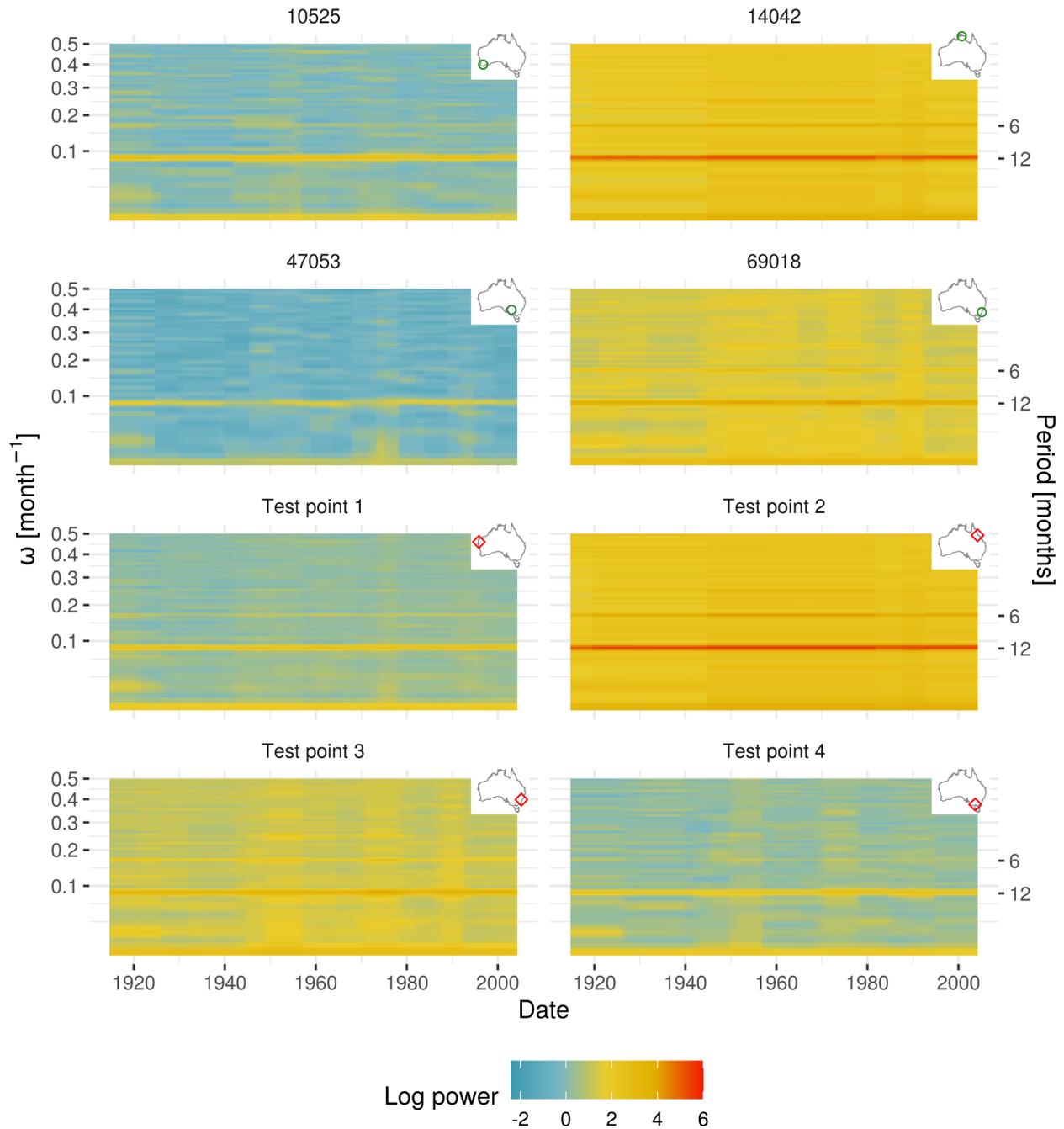


Figure 8: Estimated time-varying spectra $\log \hat{f}(t, \omega, \mathbf{u})$ for four monthly rainfall sites (first two rows) and four locations without observations (last two rows). The color indicates the log power at the corresponding time and frequency. The ω -axis is on a square-root scale. The axis on the right-hand side displays the period ($1/\omega$).

(where the week starts on a Sunday) incidence rate per 100,000 population. In this work we use time series from the continental United States (that is, excluding Hawaii and Alaska), plus the District of Columbia. These span the 3,914 weeks from week one of 1928 (which we write as 1928-01) to week one of 2003 (2003-01). Across all 49 time series there are 191,786 observations. Of these, 50,067 (26%) are missing, and 30,439 (16%) have incidence equal to zero. The time series are shown in Figure 9, where each panel presents the series for a state, and the incidence axis is on a square-root scale (note however that the spectral analysis performed later is applied to the untransformed data). The layout of the panels roughly matches the geographic distribution of the states. The most striking aspect of these plots is the dramatic decline in both the level and volatility of incidence starting in 1963, the year of vaccine licensure.

We set $\mathbf{u}_j = (\text{lon}_j, \text{lat}_j)$, the longitude and latitude of the centroid of each state, and fit AdaptSPEC-X to the measles time series. Each mixture component has $t_{\min} = 208$ (4 years) as the minimum segment length. This was chosen to ensure four observations of the annual cycle in each segment. The maximum number of segments was set as $M_{\max} = 18$, the maximum allowed by the combination of t_{\min} and the number of weeks in the data. The support of the prior on $\mu_{s,m}^h$ was set to $(\mu_-, \mu_+) = (0, 20)$; the lower bound represents the known positivity of the incidence rates (and is particularly helpful to constrain $\mu_{s,m}^h$ in periods with very small counts), while the upper bound is twice as large as the posterior mean of this parameter. As in the rainfall application, preliminary fits of AdaptSPEC to individual measles time series indicated that the estimated log spectra did not change much for $J > 60$, and so we set $J = 60$ in this application as well. The LSBP is truncated at $H = 10$ components, as we found higher values did not change the results. Finally, the thin-plate GP prior for the LSBP has $B = 20$ basis functions, which captures more than 95% of the variation implied by the prior. The MCMC scheme took three hours to run on a computer using four cores of an Intel Core i9-9900K CPU with a clock speed of 3.60 GHz. The resulting estimated time-varying means and spectra are shown in figures 10 and 11, respectively.

As in Figure 9, the post-vaccine drop in the mean and power of incidence is the most obvious feature. This drop occurs in steps, starting with a dramatic drop following the licensing of the Edmonston vaccine in 1963, stalling around 1970, then dropping again around 1980. This corresponds to the waxing and waning of government funding and effort targeted at measles elimination, which culminated in an intensified elimination drive (Hinman et al., 1979; Atkinson et al., 1992). An out-

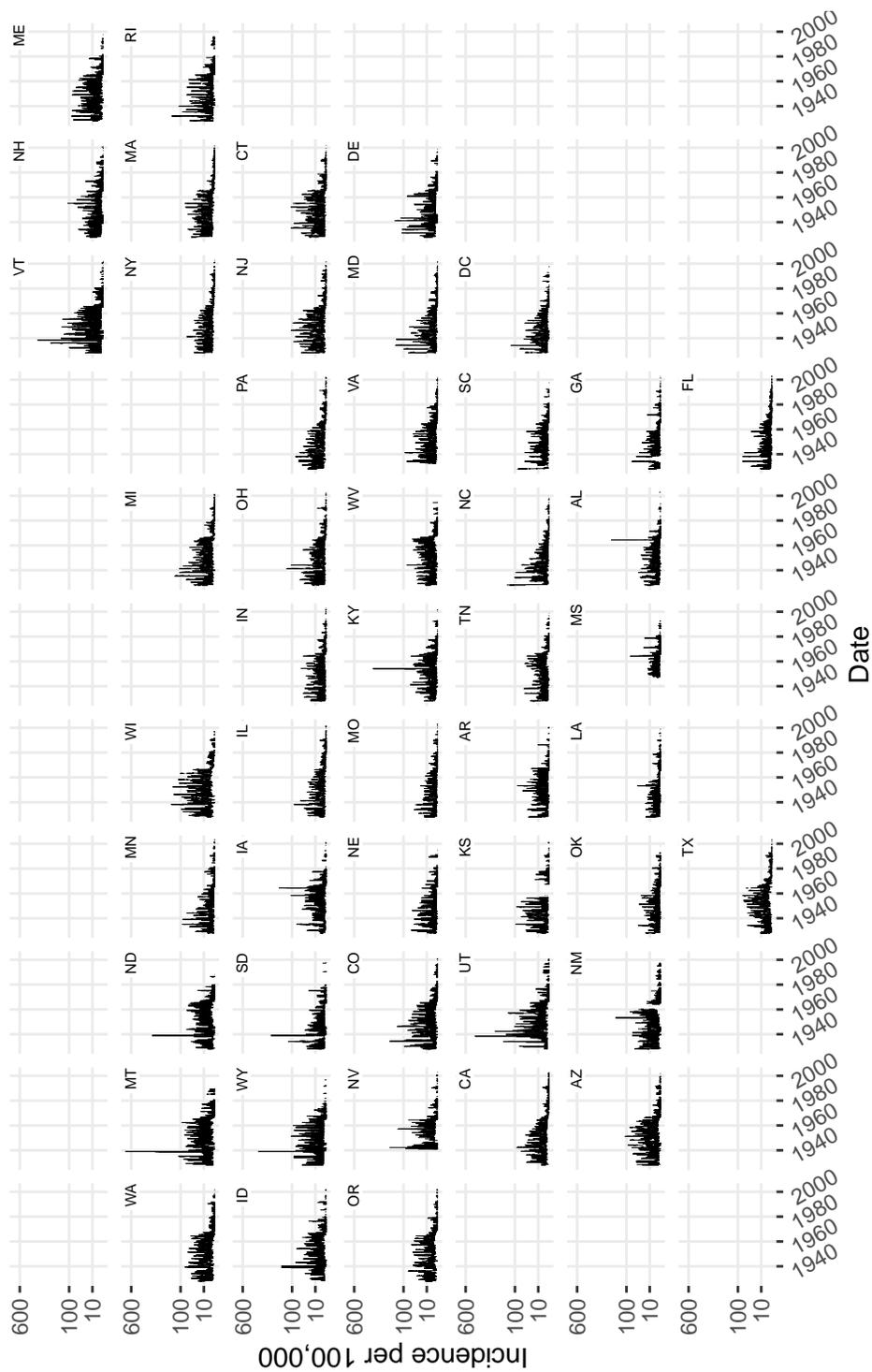


Figure 9: Measles incidence rate per 100,000 population for the continental US (that is, excluding Hawaii and Alaska), plus the District of Columbia. Each panel shows one state, where the layout of the panels roughly matches the geographic distribution of the states. The incidence axis is square-root transformed.

break during the early 1990s is visible as an increase in power in Figure 11; this outbreak received much attention and resulted in changes to the immunization schedule for children (Atkinson et al., 1992).

In the pre-vaccine period, the spectra in Figure 11 have peaks around frequencies $1/52$ and 0 , indicating annual seasonality and long-term dependence, respectively. After the introduction of the vaccine, the annual peak disappears. Grenfell et al. (2001) identify a biennial cycle in similar measles data for the UK, but this does not appear to be a feature of the US data. Figures 10 and 11 also indicate changes in the mean and spectrum during the pre-vaccine years, where all states exhibit periods of increased mean incidence and power centered around 1940 and 1955. This concords with van Panhuis et al. (2013), who note that incidence rates had variable patterns in the pre-vaccine time period, speculating that these may have been due to sanitation, hygiene, or demographic factors.

Because of the extreme nonstationarity introduced by the vaccine, the time-varying spectra in Figure 11 span such a wide range of powers that the spectra for all states look almost identical. This is not the case: a deeper inspection, for example by magnifying Figure 11, reveals that the spectra highlights the existence of geographic heterogeneity between states, where higher power is more typical of the west and north, compared to the south and east.

Since the elimination of endemic measles in the US in 2000, there have been a number of outbreaks associated with individuals ‘importing’ measles by acquiring the disease while outside the US and spreading it upon their return (Parker et al., 2006; CDC, 2019). Phadke et al. (2016) associated several of these outbreaks with individuals unvaccinated for nonmedical reasons, which they term as vaccine refusal. Some authors have even declared that a resurgence of measles has occurred (Lynfield and Daum, 2014). Using the AdaptSPEC-X fits, we tested whether the mean $\mu(t, \mathbf{u})$ or variance $\sigma^2(t, \mathbf{u})$ of measles increased from 1995-01, a few years after the big outbreak in the early 1990s, to 2003-01, the last time period in the data. We find no evidence of increase in $\mu(t, \mathbf{u})$ in any state, which can be seen from the fact that the highest posterior probability of an increase equals 0.68 in Oklahoma. As for $\sigma^2(t, \mathbf{u})$, the posterior probabilities of an increase in all states range between 0.81–0.87, which we consider to be weak evidence of change. Unfortunately, because the data end in 2003, it is not possible to assess changes to the mean or spectrum of measles incidence in more recent years.

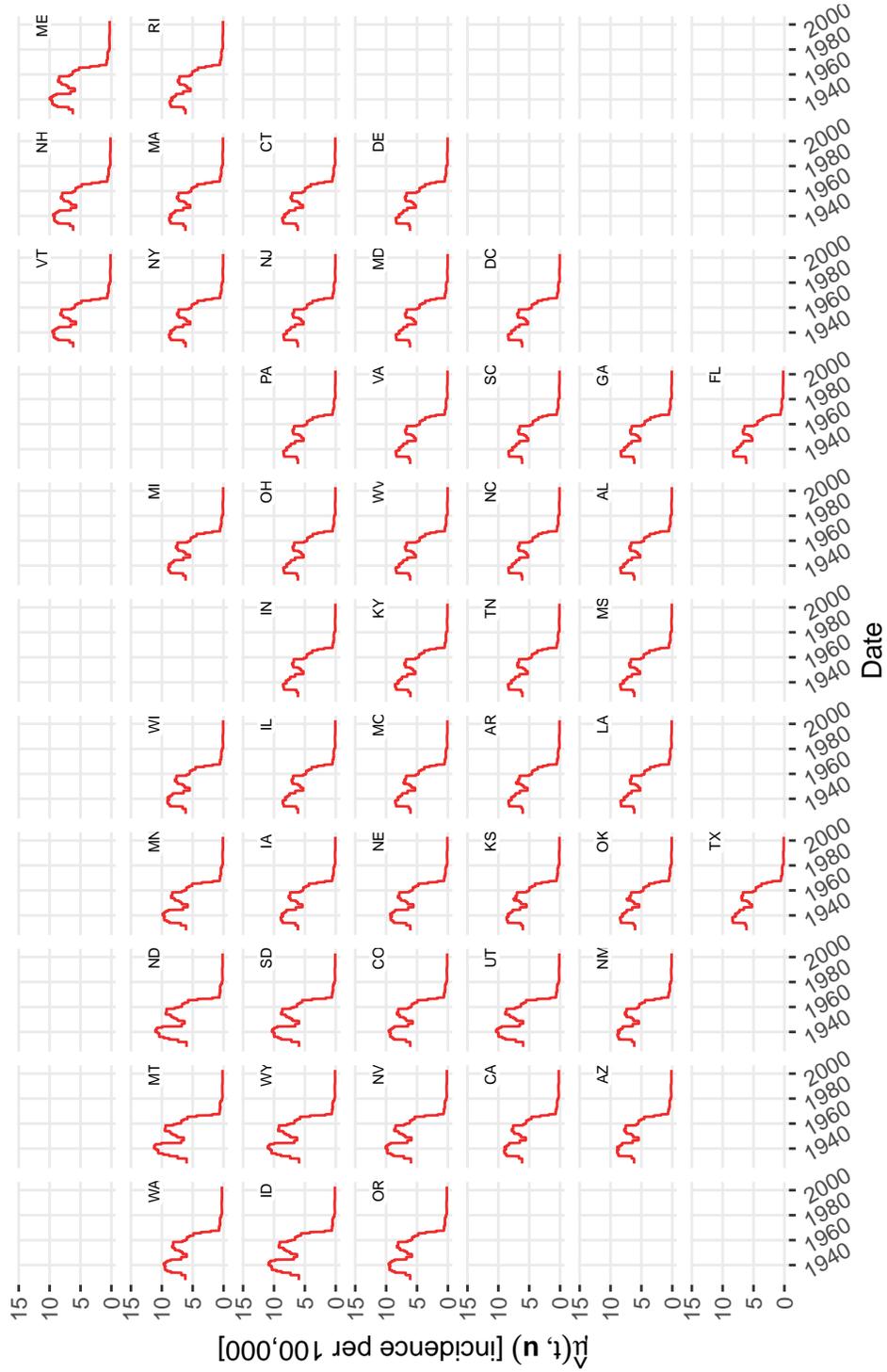


Figure 10: Estimated time-varying mean $\hat{\mu}(t, \mathbf{u})$ for measles incidence, where each panel shows the estimate for one state.

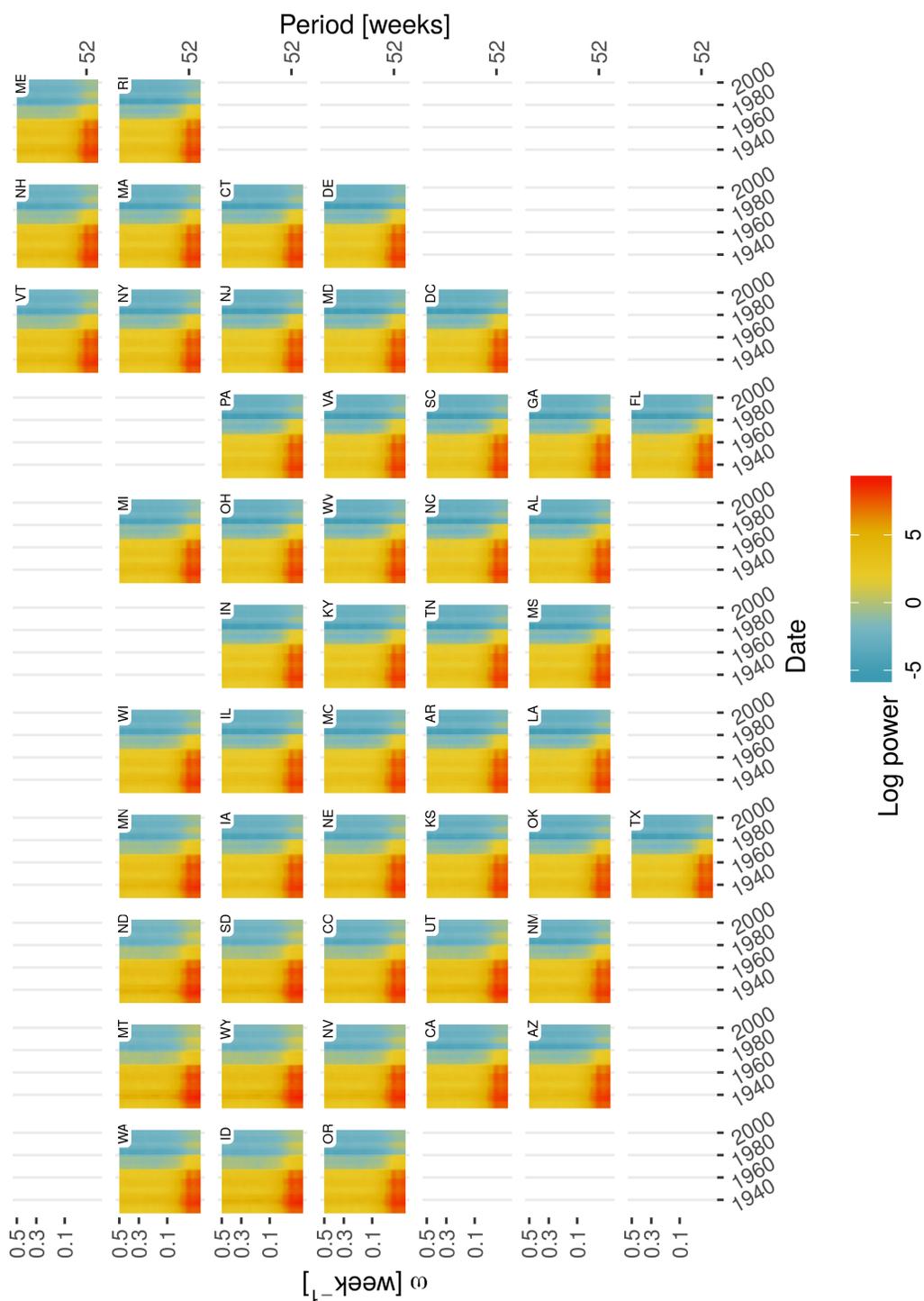


Figure 11: Estimated time-varying log spectra, $\log \hat{f}(t, \omega, \mathbf{u})$ for measles incidence, where each panel shows the estimate for one state. Colors indicate the log power at the corresponding date and ω . The ω -axis is on a square-root scale. The top axis displays the period ($1/\omega$).

7 Discussion

This article has presented AdaptSPEC-X, a Bayesian method for analyzing a panel of possibly non-stationary time series using a covariate-dependent infinite mixture model, with mixture components parameterized by their time-varying mean and spectrum. AdaptSPEC-X extends AdaptSPEC to accommodate multiple time series, each with its own covariate values. Specifically, the covariates, which are assumed to be time-independent, are incorporated via the mixture weights using the logistic stick breaking process. The mixture components are based on AdaptSPEC, which handles a single nonstationary time series. In particular, it partitions a time series into an unknown but finite number of segments, and estimates the spectral density within each segment by smoothing splines. New features which have been added to the AdaptSPEC components include estimation of time-varying means and handling of missing observations. The model and sampling scheme can accommodate large panels, such as that of the measles application. In addition to estimating time-varying spectra for each time series in the panel, AdaptSPEC-X allows inference about the underlying process at unobserved covariate values, enabling predictive inference. Efficient software implementing AdaptSPEC-X is available in the R package BayesSpec.

In Section 3.2, the log odds of the LSBP, which determine the mixture weights, are modeled using a thin-plate GP prior. While this prior is flexible, it is also smooth and stationary. This property may be inappropriate in settings where changes in the mean or spectrum of the individual time series occur abruptly over the covariate space. An extension to a nonstationary prior for the log odds, or a piecewise prior as in Bruce et al. (2018), may be of interest in these cases.

AdaptSPEC (and therefore AdaptSPEC-X) relies on Whittle’s approximation to the likelihood (Equation (1)). The Whittle likelihood is asymptotically correct for both Gaussian and non-Gaussian time series (Hannan, 1973), but is known to be inefficient for small sample sizes (Contreras-Cristán et al., 2006). Several methods exist in the literature to ameliorate this problem (see, for example, Sykulski et al., 2019), and these methods might produce useful extensions to AdaptSPEC for settings with short time series or small segment lengths (i.e., small t_{\min}). Another feature of AdaptSPEC that is inherited by AdaptSPEC-X is the use of a smoothing splines prior to model the log spectrum within a segment. While this is a very flexible model, it may not be optimal in all settings. For instance, if the rainfall application were repeated at the daily scale, the

smoothing splines would require a very large number of basis functions to accommodate the low frequency spike at $1 / 365.25$ days. Future work into alternative prior structures would be valuable for this and similar settings.

Neither AdaptSPEC nor AdaptSPEC-X account explicitly for measurement error. For i.i.d. Gaussian measurement error, this should not cause a problem, as the added variance would appear as an added constant (that is, white noise) in the spectrum. On the other hand, if the measurement error is changing over time, this may be estimated as false nonstationary, in the sense that the underlying process is not actually changing. Adding an explicit layer for measurement error to AdaptSPEC-X would be a useful extension for these cases. More specialized forms of measurement error may be useful in other settings. For example, in the measles application of Section 6.2, the data are incidence rates per 100,000 population. While rates are continuous quantities, they are constructed from counts, so the measurement process is in fact discrete. A hierarchical extension accounting for this would be an interesting addition to AdaptSPEC-X.

At present, it is not obvious how to assess the fit to data of AdaptSPEC-X, but we provide some preliminary thoughts about how this might be done. [Diggle and al Wasel \(1997\)](#) formulate spectral estimation of a stationary time series as a generalized linear model (GLM) with a log link function and a gamma family, where the dependent variable is the periodogram. [Wahba \(1980\)](#) uses this formulation as well. This leads us to think that GLM diagnostics may be possible for assessing the model fit in our case. However, since our model is piecewise stationary (within mixture components) with variable partition points, implementing such an approach would require more research.

AdaptSPEC-X allows for covariates that do not vary with time, and an extension to a more general framework accommodating time-varying covariates would facilitate new types of inference. For example, external time-varying climate indicators such as the Southern Oscillation Index are known to influence rainfall patterns ([Bertolacci et al., 2019](#)). A time-varying covariate influencing the mean or the spectrum might improve predictive performance. One challenge with such an approach would be to determine how a time-varying covariate could interact with the segmentation approach used by AdaptSPEC to handle nonstationarity. It would also be interesting to allow the mean to be time-varying and covariate-dependent within segments, not only between segments. Future research will focus on these extensions.

Reproducibility

Data and code reproducing the figures and tables in this manuscript are available online at <https://github.com/mbertolacci/adaptspecx>.

Funding

E. Cripps and M. Bertolacci were supported by the ARC Industrial Transformation Research Hub for Offshore Floating Facilities which is funded by the Australian Research Council, Woodside Energy, Shell, Bureau Veritas and Lloyds Register (Grant No. 140100012). S. Cripps and E. Cripps were supported by the ARC Industrial Transformation Training Hub, Data Analytics for Resources and Environments (IC190100031), funded by the Australian Government. O. Rosen was supported in part by grants NSF DMS-1512188 and NIH 2R01GM113243-05.

References

- Adak, S. (1998), “Time-dependent spectral analysis of nonstationary time series,” *Journal of the American Statistical Association*, 93, 1488–1501.
- Agresti, A. (2018), *An Introduction to Categorical Data Analysis, 3rd Edition*, NJ: Wiley.
- Alter, D. (1924), “Application of Schuster’s periodogram to long rainfall records, beginning 1748,” *Monthly Weather Review*, 52, 479–487.
- Ansell, T. J., Reason, C. J. C., Smith, I. N., and Keay, K. (2000), “Evidence for decadal variability in southern Australian rainfall and relationships with regional pressure and sea surface temperature,” *International Journal of Climatology*, 20, 1113–1129.
- Atkinson, W. L., Orenstein, W. A., and Krugman, S. (1992), “The resurgence of measles in the United States, 1989–1990,” *Annual Review of Medicine*, 43, 451–463.
- Bertolacci, M., Cripps, E., Rosen, O., Lau, J. W., and Cripps, S. (2019), “Climate inference on daily rainfall across the Australian continent, 1876–2015,” *Annals of Applied Statistics*, 13, 683–712.

- 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.
- Cai, W. J., Jones, D. A., Harle, K., Cowan, T. D., Power, S. B., Smith, I. N., Arblaster, J. M., and Abbs, D. J. (2007), “Past climate change,” in *Climate Change in Australia*, eds. Pearce, K., Holper, P., Hopkins, M., Bouma, W., Whetton, P., Hennessy, K., and Power, S., Victoria, Australia: CSIRO Marine and Atmospheric Research, pp. 17–28.
- 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.
- CDC (2019), “Centers for Disease Control and Prevention: Measles Cases and Outbreaks,” <https://www.cdc.gov/measles/cases-outbreaks.html>, accessed: 2019-03-21.
- 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.
- Choudhuri, N., Ghosal, S., and Roy, A. (2004), “Bayesian estimation of the spectral density of a time series,” *Journal of the American Statistical Association*, 99, 1050–1059.
- Chung, Y. and Dunson, D. B. (2009), “Nonparametric Bayes conditional distribution modeling with variable selection,” *Journal of the American Statistical Association*, 104, 1646–1660.
- Contreras-Cristán, A., Gutiérrez-Peña, E., and Walker, S. G. (2006), “A note on Whittle’s likelihood,” *Communications in Statistics - Simulation and Computation*, 35, 857–875.
- Dahlhaus, R. (1997), “Fitting time series models to nonstationary processes,” *The Annals of Statistics*, 25, 1–37.
- (2012), “Locally Stationary Processes,” in *Handbook of Statistics*, eds. Rao, T. S., Rao, S. S., and Rao, C. R., Amsterdam: Elsevier, vol. 30, chap. 13, pp. 351–413.
- Davis, R. A., Lee, T. C. M., and Rodriguez-Yam, G. A. (2006), “Structural break estimation for nonstationary time series models,” *Journal of the American Statistical Association*, 101, 223–239.

- 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.
- Fiecas, M. and Ombao, H. (2016), “Modeling the evolution of dynamic brain processes during an associative learning experiment,” *Journal of the American Statistical Association*, 111, 1440–1453.
- Freyermuth, J.-M., Ombao, H., and von Sachs, R. (2010), “Tree-structured wavelet estimation in a mixed effects model for spectra of replicated time series,” *Journal of the American Statistical Association*, 105, 634–646.
- Gallant, A. J. E., Reeder, M. J., Risbey, J. S., and Hennessy, K. J. (2013), “The characteristics of seasonal-scale droughts in Australia, 1911–2009,” *International Journal of Climatology*, 33, 1658–1672.
- Gelman, A. (2006), “Prior distributions for variance parameters (Comment on article by Browne and Draper),” *Bayesian Analysis*, 1, 515–534.
- Girolami, M. and Calderhead, B. (2011), “Riemann manifold Langevin and Hamiltonian Monte Carlo methods,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73, 123–214.
- Green, P. J. (1995), “Reversible jump Markov chain Monte Carlo computation and Bayesian model determination,” *Biometrika*, 82, 711–732.
- Grenfell, B. T., Bjørnstad, O. N., and Kappey, J. (2001), “Travelling waves and spatial hierarchies in measles epidemics,” *Nature*, 414, 716–723.
- Guinness, J. (2019), “Spectral density estimation for random fields via periodic embeddings,” *Biometrika*, 106, 267–286.
- Guo, W., Dai, M., Ombao, H. C., and von Sachs, R. (2003), “Smoothing spline ANOVA for time-dependent spectral analysis,” *Journal of the American Statistical Association*, 98, 643–652.
- Hannan, E. J. (1973), “The asymptotic theory of linear time-series models,” *Journal of Applied Probability*, 10, 130–145.

- Hastie, D. I., Liverani, S., and Richardson, S. (2015), “Sampling from Dirichlet process mixture models with unknown concentration parameter: mixing issues in large data implementations,” *Statistics and Computing*, 25, 1023–1037.
- Hinman, A. R., Brandling-Bennett, A. D., and Nieburg, P. I. (1979), “The opportunity and obligation to eliminate measles from the United States,” *J. Am. Med. Assoc.*, 242, 1157–1162.
- Hope, P. K., Drosowsky, W., and Nicholls, N. (2006), “Shifts in the synoptic systems influencing southwest Western Australia,” *Climate Dynamics*, 26, 751–764.
- Iannaccone, R. and Coles, S. (2001), “Semiparametric models and inference for biomedical time series with extra-variation,” *Biostatistics*, 2, 261–276.
- Ishwaran, H. and James, L. F. (2001), “Gibbs sampling methods for stick-breaking priors,” *Journal of the American Statistical Association*, 96, 161–173.
- Jacobs, R. A., Jordan, M. I., Nowlan, S. J., and Hinton, G. E. (1991), “Adaptive mixtures of local experts,” *Neural Comput.*, 3, 79–87.
- Joshi, M. K. and Pandey, A. C. (2011), “Trend and spectral analysis of rainfall over India during 1901–2000,” *Journal of Geophysical Research: Atmospheres*, 116.
- Katz, S. L. and Hinman, A. R. (2004), “Summary and conclusions: Measles elimination meeting, 16–17 March 2000,” *The Journal of Infectious Diseases*, 189, S43–S47.
- Kitagawa, G. and Gersch, W. (1996), *Smoothness Priors Analysis of Time Series*, NY: Springer.
- 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, 1405–1416.
- Lavery, B., Kariko, A., and Nicholls, N. (1992), “A historical rainfall data set for Australia,” *Australian Meteorological Magazine*, 40, 33–39.

- 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.
- Lynfield, R. and Daum, R. S. (2014), “The complexity of the resurgence of childhood vaccine-preventable diseases in the United States,” *Current Pediatrics Reports*, 2, 195–203.
- Macaro, C. and Prado, R. (2014), “Spectral decompositions of multiple time series: A Bayesian non-parametric approach,” *Psychometrika*, 79, 105–129.
- Moreno, M. A. (2018), “Measles,” *JAMA Pediatrics*, 172, 896–896.
- Nason, G. P., Sachs, R. V., and Kroisandt, G. (2000), “Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62, 271–292.
- Ombao, H. C., Raz, J. A., von Sachs, R., and Malow, B. A. (2001), “Automatic statistical analysis of bivariate nonstationary time series,” *Journal of the American Statistical Association*, 96, 543–560.
- Parker, A. A., Staggs, W., Dayan, G. H., Ortega-Sánchez, I. R., Rota, P. A., Lowe, L., Boardman, P., Teclaw, R., Graves, C., and LeBaron, C. W. (2006), “Implications of a 2005 measles outbreak in Indiana for sustained elimination of measles in the United States,” *New England Journal of Medicine*, 355, 447–455.
- Phadke, V. K., Bednarczyk, R. A., Salmon, D. A., and Omer, S. B. (2016), “Association between vaccine refusal and vaccine-preventable diseases in the United States: A review of measles and pertussis,” *Journal of the American Medical Association*, 315, 1149–1158.
- Polson, N. G., Scott, J. G., and Windle, J. (2013), “Bayesian inference for logistic models using Pólya-Gamma latent variables,” *Journal of the American Statistical Association*, 108, 1339–1349.
- Pook, M. J., Risbey, J. S., and McIntosh, P. C. (2012), “The synoptic climatology of cool-season rainfall in the Central Wheatbelt of Western Australia,” *Monthly Weather Review*, 140, 28–43.
- Priestley, M. B. (1965), “Evolutionary spectra and non-stationary processes,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 27, 204–237.

- 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, referred to in the text as QGL09.
- Qin, L. and Wang, Y. (2008), “Nonparametric spectral analysis with applications to seizure characterization using EEG time series,” *The Annals of Applied Statistics*, 2, 1432–1451.
- Rajagopalan, B. and Lall, U. (1998), “Interannual variability in western US precipitation,” *Journal of Hydrology*, 210, 51 – 67.
- Rigon, T. and Durante, D. (2021), “Tractable Bayesian density regression via logit stick-breaking priors,” *Journal of Statistical Planning and Inference*, 211, 131–142.
- Risbey, J. S., McIntosh, P. C., and Pook, M. J. (2013), “Synoptic components of rainfall variability and trends in southeast Australia,” *International Journal of Climatology*, 33, 2459–2472.
- Rosen, O. and Stoffer, D. S. (2007), “Automatic estimation of multivariate spectra via smoothing splines,” *Biometrika*, 94, 335–345.
- Rosen, O., Stoffer, D. S., and Wood, S. (2009), “Local spectral analysis via a Bayesian mixture of smoothing splines,” *Journal of the American Statistical Association*, 104, 249–262.
- Rosen, O., Wood, S., and Stoffer, D. S. (2012), “AdaptSPEC: Adaptive spectral estimation for nonstationary time series,” *Journal of the American Statistical Association*, 107, 1575–1589, referred to in the text as RWS12.
- Rubin, D. B. (1976), “Inference and missing data,” *Biometrika*, 63, 581–592.
- Sykulski, A. M., Olhede, S. C., Guillaumin, A. P., Lilly, J. M., and Early, J. J. (2019), “The debiased Whittle likelihood,” *Biometrika*, 106, 251–266.
- Ummenhofer, C. C., England, M. H., McIntosh, P. C., Meyers, G. A., Pook, M. J., Risbey, J. S., Gupta, A., and Taschetto, A. S. (2009), “What causes southeast Australia’s worst droughts?” *Geophysical Research Letters*, 36, L04707.

- van Panhuis, W. G., Grefenstette, J., Jung, S. Y., Chok, N. S., Cross, A., Eng, H., Lee, B. Y., Zadorozhny, V., Brown, S., Cummings, D., and Burke, D. S. (2013), “Contagious diseases in the United States from 1888 to the present,” *New England Journal of Medicine*, 369, 2152–2158.
- Wahba, G. (1980), “Automatic smoothing of the log periodogram,” *Journal of the American Statistical Association*, 75, 122–132.
- (1990), *Spline Models for Observational Data*, PA: Society for Industrial and Applied Mathematics.
- West, M., Prado, R., and Krystal, A. D. (1999), “Evaluation and comparison of EEG traces: Latent structure in nonstationary time series,” *Journal of the American Statistical Association*, 94, 1083–1095.
- Whittle, P. (1953), “Estimation and information in stationary time series,” *Arkiv för Matematik*, 2, 423–434.
- (1957), “Curve and periodogram smoothing,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 19, 38–63.
- Wood, S. (2013), “Applications of Bayesian Smoothing Splines,” in *Bayesian Theory and Applications*, eds. Damien, P., Dellaportas, P., Polson, N. G., and Stephens, D. A., Oxford University Press, pp. 309–336.
- Yang, J. and Zhou, Z. (2020), “Spectral inference under complex temporal dynamics,” *Journal of the American Statistical Association*, 0, 1–23.
- Yang, W.-H., Holan, S. H., and Wikle, C. K. (2016), “Bayesian lattice filters for time-varying autoregression and time–frequency analysis,” *Bayesian Analysis*, 11, 977–1003.