Hierarchical Nonlinear, Multivariate, and Spatially-Dependent Time-Frequency Functional Models

---

A Dissertation presented to

the Faculty of the Graduate School

at the University of Missouri

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Wen-Hsi Yang

Dr. Scott H. Holan and Dr. Christopher K. Wikle, Dissertation Supervisors

July 2013

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled:

Hierarchical Nonlinear, Multivariate, and
Spatially-Dependent Time-Frequency Functional Models

presented by Wen-Hsi Yang,
a candidate for the degree of Doctor of Philosophy and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Scott H. Holan

Dr. Christopher K. Wikle

Dr. Athanasios Micheas

Dr. Subharup Guha

Dr. Mark L. Wildhaber

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Notions of time and frequency are important constituents of most scientific inquiries, providing complimentary information. In an era of "big data," methodology for analyzing functional and/or image data is increasingly important. This dissertation develops methodology at the cross-section of time-frequency analysis and functional data and consists of three distinct, but related, contributions. First, we propose nonparametric methodology for nonlinear multivariate time-frequency functional data. In particular, we consider polynomial nonlinear functional data models that accommodate higher dimensional functional covariates, including time-frequency images, along with their interactions. The necessary dimension reduction for model estimation proceeds through carefully chosen basis expansions (empirical orthogonal functions) and feature-extraction stochastic search variable selection (SSVS). Properties of the methodology are examined through an extensive simulation study. Finally, we illustrate the approach through an application that attempts to characterize spawning behavior of shovelnose sturgeon in terms of high-density depth and temperature profiles. The second contribution proposes model-based time-frequency estimation through Bayesian lattice filter time-varying autoregressive models. In this context, we take a fully Bayesian approach and allow both the autoregressive coefficients and innovation variance to vary over time. Importantly, our model is estimated within the partial autocorrelation domain (i.e., through the partial autocorrelation coefficients). Additionally, all of the full conditional distributions required for our algorithm are of standard form and thus can be easily implemented using a Gibbs sampler. Further, as a by-product of the lattice filter recursions, our approach avoids

undesirable matrix inversions. As such, estimation is computationally efficient and stable. We conduct a comprehensive simulation study that compares our method with other competing methods and find that, in most cases, our approach performs superior in terms of average squared error between the estimated and true time-varying spectral density. Lastly, we demonstrate our methodology through several real case studies. The final project of the dissertation develops models that accommodate spatially dependent functional responses with spatially dependent image predictors. The methodology is motivated by a soil science study that seeks to model spatially correlated water content functionals as a function of electro-conductivity images. The water content curves are measured at different locations within the study field and at various depths, whereas the electro-conductivity images are spatially referenced images of wavelength by depth. Estimation is facilitated by taking a Bayesian approach, where the necessary dimension reduction for model implementation proceeds using basis function expansions along with SSVS. Finally, the methodology is illustrated through an application to our motivating data.

Chapter 1

Introduction

1.1 Introduction

Time-frequency analysis has become a fundamental tool in the analysis of nonstationary signals. Due to the lack of applicability surrounding the global Fourier transform for such processes, there has been significant research on time-frequency methodology. In the time-frequency domain, crucial aspects of the underlying process of interest often become apparent that would otherwise go undetected (Kestin et al., 1998; Cranstoun et al., 2002; Ansari-Asl et al., 2005; Holan et al., 2010). For example, Holan et al. (2010) consider time-frequency analysis in the context of animal communication signals, whereas Ansari-Asl et al. (2005) investigate electroencephalograms (EEGs) in the time-frequency domain. In contrast, Yang et al. (2013) (cf., Chapter 2) considers time-frequency analysis to characterize spawning behavior of shovelnose sturgeon on the Lower Missouri River. Importantly, analyses of these types may hold the keys to

unraveling some of the mysteries of animal behavior and phenotypic selection, as well as unlocking the secrets of brain function. Nevertheless, often, time-frequency analyses are implemented in a "stand-alone" fashion with no quantification of uncertainty, thus hindering inferential capacity.

Additionally, in many contexts, multiple signals and their interactions are necessary for addressing core scientific questions. Furthermore, these time-frequency predictors may relate to the response nonlinearly. As a result, development of general methodology for time frequency functional data is timely and relevant. Critically, methodology for modeling complex processes using time-frequency functional predictors is needed that accurately quantifies uncertainty, accommodates multivariate time-frequency predictors and allows for nonlinear behavior. Devoid of such methodology, substantive signal aspects may go undiscovered during many scientific investigations. For example, in the time-frequency domain, Holan et al. (2010) discover several features of *Enchenopa* treehopper communication signals that are important indicators of mating success (such as increased energy in the broadband clicks at the start of the signal); these features were overlooked in previous time-domain analyses. Similarly, Holan et al. (2012) utilize time-frequency functional models in the context of macroeconomics/finance to reveal important features of daily NASDAQ returns corresponding to economic recessions.

Recently, attention has been directed toward modeling data having image predictors (see Reiss and Ogden, 2010; Morris et al., 2011; Holan et al., 2010, and the references therein). In contrast, relatively few research efforts have been devoted to functional models with time-frequency functional predictors. Two notable exceptions are given by Qin et al. (2009) and Holan et al. (2010). In the case of the former, the

authors develop a smoothing spline ANOVA model for the time-frequency coefficients. Estimation is carried out using a two-stage approach and uncertainty is determined through an asymptotic equivalence that yields a posterior Bayesian confidence interval (Wahba, 1983). To implement this approach it is also necessary to choose block sizes for calculating local periodograms. Holan et al. (2010) presents the first hierarchical functional generalized linear model with time-frequency predictors. Specifically, Holan et al. (2010) represents the first instance of treating the time-frequency representation (spectrogram) as a spatial image with the goal of incorporating the time-frequency image as a regressor in a predictive model. Importantly, their methodology allows the practitioner to draw statistical inference in a rigorous manner based on a dimension-reduced posterior mean *difference-spectrogram* (i.e., posterior mean of dimension reduced time-frequency representation of successful mating minus that of the unsuccessful mating), along with measures of uncertainty.

Many other efforts have been made to analyze time-frequency behavior (see, for example, Ombao et al., 2002, 2005; Wolfe et al., 2004); however, these methods typically analyze the nonstationary time series directly rather than using the time-frequency representation as a predictor to describe a more complicated process and often consider asymptotic arguments to assess uncertainty. Consequently, methodology is lacking for nonlinear time-frequency functional data as well as for spatially-dependent time-frequency functional data. Thus, given the current state of time-frequency analysis and functional data modeling, there is a definite need for the development of flexible methodology for time-frequency functional data that can extract salient data features while fully accounting for uncertainty.

1.2 Time-Frequency Transformation

Estimation of the time-frequency surface of a signal can be achieved using various approaches. The most common approach is to use the short-time Fourier transform (STFT) to form a spectrogram or evolutionary spectrum (see Gröchenig, 2001; Oppenheim and Schafer, 2009, and the references therein). Specifically, STFT segments the signal and then computes the local Fourier transform to produce a local time-frequency representation that characterizes local signal properties. This segmentation results in artificially large Fourier coefficients at high frequencies and therefore some form of smoothing (windowing) is often introduced (see Feichtinger and Strohmer, 1998, for further discussion). Several popular windows have been proposed for this purpose (e.g., Hamming, Hanning, Bartlett and Kaiser windows). Consequently, in practice, care must be taken to assess the sensitivity to the choice of window. Another approach uses smooth localized complex exponentials (SLEX) functions (Ombao et al., 2002). The SLEX functions are simultaneously orthogonal and localized in time and frequency by applying a projection operator (Ombao et al., 2002). Similar to the STFT, this method requires segmentation and subsequent smoothing. Automatic methods of SLEX have been proposed (Cranstoun et al., 2002), but may be difficult to integrate into a fully Bayesian framework. Alternatively, one can use continuous wavelet transforms (Mallat, 2008; Percival and Walden, 2000; Vidakovic, 1999) and Gabor frames (Feichtinger and Strohmer, 1998; Fitzgerald et al., 2000; Wolfe et al., 2004) to directly produce a time-frequency representation. These approaches use the theory of frames and overcomplete dictionaries, introducing redundancy into the basis functions. As a result, the time-frequency representation may simultaneously provide better time and frequency resolution as a consequence of the *Balian-Low Theorem*

(Gröchenig, 2001).

Instead of using nonparametric methods, one can alternatively employ a model-based approach. In particular, the time-frequency representation of a nonstationary signal can be obtained from the coefficients of a time-varying autoregressive model (TVAR) along with the innovation variance (Kitagawa, 2009; Prado and West, 2010; Priestley, 1981; Shumway and Stoffer, 2006). In contrast to nonparametric methods, the model-based approach can be easily integrated into a fully Bayesian framework.

1.3   Dimension Reduction

One can consider the time-frequency representation as a spatial map, where the time and frequency axis correspond to the "latitude" and "longitude", respectively, and the value of each pixel is the observed value on a regular lattice. Consequently, the time-frequency representation includes pixels that are spatially dependent and the dimension is extremely high. For example, Figure 1.1 contains the spectrogram of a treehopper mating signal, with 87 time points and 129 frequencies, corresponding to 11,223 possible time-frequency pixels. Clearly, if one were considering individual pixels as covariates, there would be too many covariates to consider even with model selection approaches.

To reduce the dimension of time-frequency representation, one could reduce the number of time points of the signal in the time domain (e.g., through decimation) and/or consider fewer frequencies in the time-frequency domain. However, because the pixels in the time-frequency representation are spatially dependent, such an approach could possibly cause a loss of important information. Alternatively, to suf-

ficiently reduce the dimensionality, one could use basis functions to decompose the spectrogram (e.g., splines, wavelets, empirical orthogonal functions (EOFs), among others). For example, one can vectorize the spectrogram and use a Karhunen-Loéve decomposition (Wikle and Cressie, 1999; Jolliffe, 2010) to find EOFs and then use the EOFs to produce expansion coefficients in low-dimensional feature space.

As mentioned above, the time-frequency decomposition of a signal represented as a high-dimensional spectrogram can be projected into a lower dimensional feature space. The lower dimensional space is important for identifying significant components of the underlying complex process and for reprojecting these components back into the time-frequency and/or time domain. Within the lower dimensional space variable selection methods such as stochastic search variable selection (SSVS) are used to facilitate such an identification. For example, Holan et al. (2010) use a stochastic search variable selection (SSVS) procedure (George and McCulloch, 1993, 1997) to identify important EOFs in terms of predicting mating success in *Enchenopa* tree hoppers.

1.4   Overview

This dissertation is organized as follows. Chapter 2 introduces a class of nonlinear multivariate time-frequency functional models using a fully Bayesian approach. This methodology is motivated by an application to fisheries biology that seeks to characterize spawning behavior of shovelnose sturgeon in terms of high-density depth and temperature profiles. We conduct an extensive simulation study to examine the properties of model and then we apply the approach to the motivating ecological data.

6

Chapter 3 introduces Bayesian lattice filter time-varying autoregressive models to provide model-based time-frequency estimation. The proposed models allow both the autoregressive coefficients and innovation variance to vary over time in order to capture the complex dependence structure of nonstationary signals. Importantly, estimation is performed in the partial autocorrelation domain. In this domain, our approach avoids calculation of high-dimensional inverse matrices. A comprehensive simulation study demonstrates that, in most cases, our approach has superior performance in terms of estimating the time-frequency representation of various nonstationary signals relative to existing state-of-the-art methodologies. We demonstrate our methodology through several real-world examples. Chapter 4 develops models that accommodate spatially dependent functional responses with spatially dependent image predictors. The methodology is motivated by a soil science study that seeks to model spatially correlated water content functionals as a function of electro-conductivity images. The water conductivity curves are measured at different locations within the study field and at various depths, whereas the electro-conductivity images are spatially referenced images of wavelength by depth. Estimation is facilitated taking a Bayesian approach, where the necessary dimension reduction for model implementation proceeds using basis function expansions along with SSVS. Finally, the methodology is illustrated through an application to our motivating data. The dissertation is concluded with a brief discussion in Chapter 5.
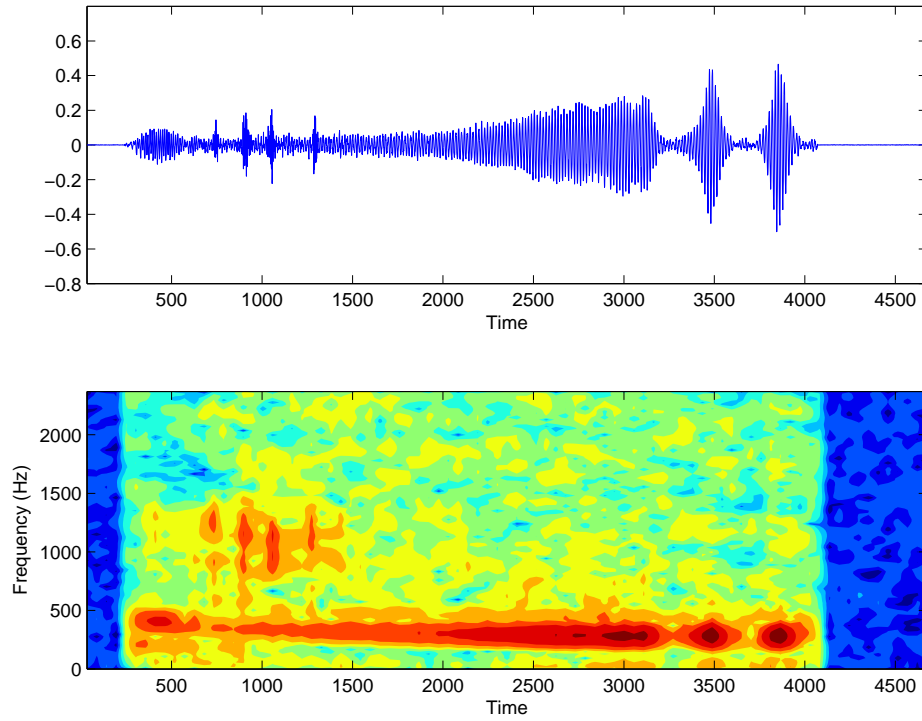
Figure 1.1: Representative treehopper mating signal in waveform (top) and an associated time-frequency representation using the short-time Fourier transform (bottom); note that the red colored portions of the image are associated with higher energy portions of the time-frequency domain

Chapter 2

Ecological Prediction with Nonlinear Multivariate Time-Frequency Functional Data Models

2.1   Introduction

The study of ecological processes has recently benefited from new data sources such as remotely sensed satellite observations, long-term networks, telemetry, and data storage tags. Such data sources typically bring vast amounts of data to bear on the problem of interest, which in turn, requires the development of new methodology to accommodate such high-dimensional data. We are interested in high-frequency time series data and its use as functional covariates in predictive models for general scientific processes, especially ecological. In particular, we consider the case where such data are potentially nonstationary in time and where it is unknown *a priori* what

features of the series are important to the ecological process. We are specifically interested in the case where we may have multiple such time series and it is their functional interaction that is fundamentally important.

In addition to the extensive development of time series methodology within the time domain, time-frequency analysis has become a fundamental component of many scientific inquiries. Crucial aspects of the underlying process of interest often become apparent in the time-frequency domain that would otherwise go undetected (Ansari-Asl et al., 2005; Cranstoun et al., 2002; Kestin et al., 1998; Holan et al., 2010). For example, salient features found in animal communication signals and electroencephalograms (EEGs) arise naturally in the time-frequency domain (Ansari-Asl et al., 2005). These features hold the keys to unraveling some of the mysteries of animal behavior and phenotypic selection, as well as unlocking the secrets of brain function.

Time-frequency analyses are often implemented in a "stand-alone" fashion with no quantification of uncertainty, thus hindering inferential capacity. In many contexts, multiple signals and their interactions are necessary for addressing core scientific questions. These time-frequency predictors may relate to the response nonlinearly. Therefore, methodology for modeling complex processes using time-frequency functional predictors is needed that accurately quantifies uncertainty, accommodates multivariate time-frequency predictors and allows for nonlinear behavior. Substantive signal aspects may go undiscovered during many scientific investigations, if such methodology is not available. For example, in the time-frequency domain Holan et al. (2010) discover several features of *Enchenopa* treehopper communication signal found to be important indicators of mating success (such as increased energy in the broad-

band clicks at the start of the signal); these features were overlooked in previous time-domain analyses. Similarly, Holan et al. (2012) utilize time-frequency functional models in the context of macroeconomics/finance to reveal important features of daily NASDAQ returns corresponding to economic recessions.

Attention has recently been directed toward modeling data having image predictors (e.g., two-dimensional functional data) (see Reiss and Ogden, 2010; Morris et al., 2011; Holan et al., 2010, and the references therein), though relatively few research efforts have been devoted to functional models with time-frequency functional predictors. Two notable exceptions are given by Qin et al. (2009) and Holan et al. (2010). In the case of the former, the authors develop a smoothing spline ANOVA model for the time-frequency coefficients. Estimation is carried out using a two-stage approach and uncertainty is determined through an asymptotic equivalence that yields a posterior Bayesian confidence interval (Wahba, 1983) (i.e., confidence intervals based on the posterior covariance function of the estimates). To implement this approach it is also necessary to choose block sizes for calculating local periodograms Holan et al. (2010) presents the first hierarchical functional generalized linear model with time-frequency predictors. Specifically, Holan et al. (2010) represents the first instance of treating the time-frequency representation (spectrogram) as a spatial image with the goal of incorporating the time-frequency image as a regressor in a predictive model. Their methodology allows the practitioner to draw statistical inference in a rigorous manner based on a dimension-reduced posterior mean *difference-spectrogram* (i.e., posterior mean of dimension reduced time-frequency representation of successful maters minus that of the unsuccessful maters), along with measures of uncertainty.

Recently, Martinez et al. (2013) studied chirp syllables of free-tailed bats using

Bayesian functional mixed models. In contrast to Holan et al. (2010), the analysis of Martinez et al. (2013) treats the spectrogram image (i.e., spectrograms of the chirps) as the response. Their method provides an excellent modeling approach that is complementary to Holan et al. (2010) and allows for modeling between-spectrogram correlation.

Many other efforts have been made to analyze time-frequency behavior (see, for example, Ombao et al., 2002, 2005; Wolfe et al., 2004); however, these methods typically analyze the nonstationary time series directly rather than using the time-frequency representation as a predictor to describe a more complicated process and often consider asymptotic arguments to assess uncertainty. Consequently, methodology is lacking for multivariate and/or nonlinear time-frequency functional data and there is a definite need for the development of flexible methodology for time-frequency functional data that can extract salient data features while fully accounting for uncertainty.

The approach proposed here extends the methodology of Holan et al. (2010), in a fully Bayesian context, to the case of multivariate and/or nonlinear time-frequency predictors. Central to this proposed framework is the probabilistically consistent hierarchical approach to incorporating multivariate time-frequency image predictors. It is reasonable to consider a time-frequency representation of a nonstationary signal as an image if one acknowledges that "nearby" time-frequency atoms are likely to be dependent and can thus be thought of analogously to pixels in an image. Heuristically, there is an extra data stage in our hierarchical model, the so-called "signal-data" stage (see Sections 2.3.2 for details). In the signal-data stage of the hierarchy the signal is transformed from the time-domain to the time-frequency domain by projection

onto various pre-specified basis functions (e.g., Fourier or wavelet transforms, etc.) or frames (e.g., using overcomplete dictionaries such as Gabor or continuous wavelet transforms, etc.). The time-frequency image is then incorporated as a predictor in lower stages of the model hierarchy, after suitable dimension reduction (see Section 2.3.3 for details). At these stages, the time-frequency "image" predictor experiences subsequent dimension reduction using feature-extraction stochastic search variable selection (SSVS; George and McCulloch, 1993, 1997). Fundamental to our approach is the coherent hierarchical formulation that accounts for the uncertainty induced by the lower-dimensional transformation in the signal-data stage of the model and avoids potential bias typically encountered from incorporating two-stage estimation procedures. Furthermore, our methodology accounts for potential nonlinear interactions between time-frequency and/or functional covariates.

This chapter is organized as follows. Section 2.2 describes the motivating ecological problem of interest. This is followed by a description of the methodology in Section 2.3 and the Bayesian inference and prediction in Section 2.4. Section 2.5 presents simulated examples that illustrate the effectiveness of our approach. Our ecological data analysis, predicting spawning success of shovelnose sturgeon, is provided in Section 2.6 and Section 2.7 concludes with discussion.

2.2 Shovelnose Sturgeon Spawning Behavior

To help elucidate the model specifications that follow, we consider the motivating example of tracking sturgeon spawning behavior in the Lower Missouri River. Human activities such as flood control and power generation may have direct effects on the

spawning success of sturgeon in the Missouri River (Galat and Lipkin, 2000). Such variations in habitat may cause changes in the Missouri River native fish assemblage, most notably in the decline of shovelnose sturgeon (*Scaphirhynchus platorynchus*) and endangered pallid sturgeon (*Scaphirhynchus albus*) (Funk and Robinson, 1974; Hesse and Sheets, 1993). Hence, to benefit sturgeon in the Lower Missouri River, several management activities such as channel reconstruction and reservoir releases have been introduced (U.S. Fish and Wildlife Service, 2000). Moreover, to understand the spawning behavior of these species, the U.S. Geological Survey (USGS) has implemented detailed studies in which data storage tags (DSTs) have been surgically implanted in gravid, female shovelonse sturgeon in the Lower Missouri River (DeLonay et al., 2007). Specifically, the DSTs provide high frequency depth (i.e., pressure) and temperature records, with measurements recorded every 15 minutes throughout the spawning season (approximately 2-4 months).

This technology provides an opportunity to develop new models that take advantage of the resulting high frequency data in order to extract crucial information on sturgeon spawning behavior and success, in the form of egg release. DeLonay et al. (2007) presented an exploratory analysis in which a change point modeling approach was introduced to examine possible timing of spawning events, whereas Holan et al. (2009) developed a hierarchical Bayesian Markov switching model with GARCH dynamics for predicting spawning success using DST depth time series data collected during 2004-2006. Subsequently, Wildhaber et al. (2011b) applied the Bayesian Markov switching model to data collected during 2007. Although these last two efforts successfully argue for the use of DSTs in ichthyology, neither extracts the full utility of the high frequency depth time series in the presence of temperature time

series data. Incorporating the temperature profile into the statistical model provides a unique challenge since temperature is one of multiple environmental factors considered key to spawning success (Wildhaber et al. (2011a) and the references therein), and therefore, is not likely to alone be a significant predictor of spawning success. This is further complicated by the fact that sturgeon, like most fish, are ectothermic with their body temperature directly correlated to surrounding water temperature and that there is strong homogeneity in water temperature in a large river system such as the Lower Missouri River. As such, there is little separation in the temperature profiles of the successful and unsuccessful spawners. However, the critical nature of temperature to development, maturation, and final release of eggs supports the hypothesis that, with all other environmental cues at sufficient levels, on reaching a certain temperature range the sturgeon settle down to lay their eggs (i.e., a decrease in the variability in the depth profile). This suggests that temperature should be an environmental cue to be considered in this model in combination with other factors (Holan et al., 2009; Wildhaber et al., 2011b). Therefore, one can better make use of this information by incorporating an interaction effect between a time-frequency representation of depth and the temperature time series.

As an example of the data of interest here, Figure 2.1 displays the time series of the (pointwise) mean values of the DST depth and temperature time series as well as the mean spectrogram time-frequency plot of the depth for 40 gravid shovelnose sturgeon. It can be seen that the time-frequency representation of depth illustrates interesting features in the data that are difficult to glean from the original time series data. For example, the time-frequency representation associated with the unsuccessful spawners presents a more pronounced time-frequency representation around May 15th when

compared with other regions and time periods. This attribute of the time-frequency plot manifests itself at low frequencies, indicating that the difference in unsuccessful spawners relative to spawners can be represented by a linear trend in depth as a function of time, which is precisely what is seen in the corresponding time-plot.

The biological hypothesis of interest focuses on changes in the variability in the depth profile as a function of time, which naturally present themselves through the time series spectral representation. In this case, the appropriate quantity is time-frequency, rather than just considering the spectral density, as the depth time series are nonstationary. This nonstationary behavior can be seen from the nonconstant mean and variance, slow decay of the autocovariance function (not shown), and other attributes. Therefore, although we can use the time series data of depth and temperature directly to study how their interaction effects corresponds to spawning success of shovelnose sturgeon, it may be more illuminating to model the depth spectrogram and the temperature time series and their potential interactions.

2.3   Methodology

We consider a general class of nonlinear multivariate time-frequency functional models. The method we propose is motivated by Holan et al. (2010), Wikle and Holan (2011), Wikle and Hooten (2010) and Yao and Müller (2010). Wikle and Holan (2011) develop polynomial nonlinear integro-difference equation models for discrete time spatio-temporal processes whereas Wikle and Hooten (2010) present a general class of models known as "general quadratic nonlinearity" for spatio-temporal processes. In contrast, Yao and Müller (2010) introduce functional polynomial regression

16

models, though the exposition is mostly developed in terms of the quadratic case. Additionally, Yao and Müller (2010) proceed from a classical perspective, quantifying uncertainty through asymptotic arguments. Our probabilistically coherent approach seamlessly embeds the nonlinear time-frequency representation into a flexible Bayesian hierarchical model and allows inclusion of both multivariate time-frequency and nonsignal predictors. As a result, the framework proposed here is extremely flexible and manages uncertainty in complex functional modeling problems typical of those encountered when adopting time-frequency predictors.

2.3.1   Time-Frequency Representations

Time-frequency analysis has become an important component in the analysis of many nonstationary signals, though it is infrequently used in the context of ecological modeling. Nonstationary signals such as depth profiles coming from DSTs necessitate local (in time) frequency analysis. In particular, the global Fourier transform will be of limited use when analyzing the spectrum of DST depth profiles, as corroborated by the type of analysis conducted in Holan et al. (2009).

Following Holan et al. (2010, 2012), the time-frequency analysis we consider utilizes the spectrogram and therefore relies on the short-time Fourier transform (STFT). Specifically, by restricting the signal (f) to a particular interval and then computing the Fourier transform, the STFT produces a local time-frequency representation that characterizes local signal properties. The segmentation of the signal, f, results in a function that is not periodic. Consequently, taking the Fourier transform will result in boundary effects that lead to large Fourier coefficients at high frequencies. To alleviate these artifacts, one can appeal to smoothing (or windowing; see Feichtinger and

Strohmer, 1998, for further discussion). The window used in our analysis (Section 2.6) is the Hamming window. A comprehensive discussion surrounding the Hamming and other windows and their usage can be found in Oppenheim and Schafer (2009).

For completeness, let $g \neq 0$ denote a fixed window function. Then, at time t and frequencies ω $(t, \omega \in \mathbb{R}^d)$, the STFT of a function f with respect to g is defined as

$$V_g f(t, \omega) = \int_{\mathbb{R}^d} f(x) \overline{g(x - t)} e^{-2\pi i \cdot \omega} dx,$$

where $i = \sqrt{-1}$, and $\overline{g(-x)} = g^*(x)$ denotes complex conjugation (cf., Definition 3.1.1, Page 37, Gröchenig, 2001). Now, assuming $g \in L^2(\mathbb{R}^d)$ is a window function and $||g||_2 = 1$, then the spectrogram of f with respect to g is defined by $\text{SPEC}_g f(t, \omega) \equiv |V_g f(t, \omega)|^2$ (cf., Definition 4.1.1, Page 60 Gröchenig, 2001). See Holan et al. (2010) and Holan et al. (2012) for further discussion.

It is important to note that the model we propose in the following sections uses a time-frequency image. However, the methodology we present allows for general object regression; e.g., images and/or functions. Moreover, our model can accommodate functional data of arbitrary dimensions (i.e., scalars, curves, images, etc.) as well as extract information from complex functional interactions.

2.3.2 Data Models

Using similar notation to Crainiceanu et al. (2009) and Yao and Müller (2010), we assume that the outcome $Y_i, i = 1, \ldots, N$ is a member of the exponential family (EF) with linear or nonlinear predictor ϑ and dispersion parameter α (i.e., ϑ is related to

the expected value of the response through a link function),

$$Y_i | \vartheta_i, \alpha \sim \mathrm{EF}(\vartheta_i, \alpha).$$

In addition, we assume that the signal predictors may be observed with error. Let $s_{ij}(\mathbf{u}^{(j)})$ denote the j-th observed signal, $j = 1, \ldots, J$ for the i-th response, $i = 1, \ldots, N$, at time-frequency location $\mathbf{u}^{(j)} \equiv (\omega^{(j)}, t^{(j)})'$, where $\omega^{(j)}$ and $t^{(j)}$ are the frequency and time indices, respectively, for the j-th signal. Note, in cases where the signal predictor is stationary in time, we would have simply $\mathbf{u}^{(j)} \equiv \omega^{(j)}$. Define $x_{ij}(\mathbf{u}^{(j)})$ to be the analogous "true" signal component associated with the observed signal component $s_{ij}(\mathbf{u}^{(j)})$. We assume the measurement (data) model

$$s_{ij}(\mathbf{u}^{(j)}) | x_{ij}(\mathbf{u}^{(j)}), \sigma_{\epsilon,j}^2 \sim N(x_{ij}(\mathbf{u}^{(j)}), \sigma_{\epsilon,j}^2),$$

for $i = 1, \ldots, N$, $j = 1, \ldots, J$, and $\{\mathbf{u}^{(j)} = (\omega^{(j)}, t^{(j)})' : \omega^{(j)} \in \Omega; t^{(j)} \in \mathcal{T}\}$. Although the true time-frequency signal elements are dependent, it is typically reasonable to assume that the signal observations are conditionally independent given the true signal. We note that dependence could be readily considered, if warranted, depending on dimension constraints.

In practice, one typically is interested in a finite set of time-frequency observations. In this case, if we let $\mathbf{s}_{ij} \equiv (s_{ij}(\mathbf{u}_1^{(j)}), \ldots, s_{ij}(\mathbf{u}_{M_j}^{(j)}))'$, where $M_j = m_j \times T_j$, corresponding to m_j frequencies and T_j times, then

$$\mathbf{s}_{ij} = \mathbf{x}_{ij} + \boldsymbol{\varepsilon}_j, \quad \boldsymbol{\varepsilon}_j \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\epsilon,j}),$$

where $\mathbf{x}_{ij}$ is an M_j-dimensional vector of true signal components analogous to $\mathbf{s}_{ij}$, and we have written the more general covariance matrix $\boldsymbol{\Sigma}_{\epsilon,j}$ for the j-th signal. Typically, $\mathbf{s}_{ij}$ come from the implementation of the spectrogram smoothing procedure described in Section 2.3.1; e.g., the k-th element of the vector $\mathbf{s}_{ij}$ is equal to $\text{SPEC}_g f(t_k^{(j)}, \omega_k^{(j)})$. As mentioned above, the errors are typically assumed to be independent and thus, in this case, $\boldsymbol{\Sigma}_{\epsilon,j} \equiv \sigma^2_{\epsilon,j}\mathbf{I}$. We also assume that, for each response Y_i, there are potentially n_z scalar covariates, given by $\mathbf{z}_i$, an n_z-dimensional vector.

2.3.3 Nonlinear Processes

Most real-world processes are nonlinear and thus, it is essential that methodology be developed that is flexible enough to model this complexity. The time-frequency functional polynomial nonlinear (FPNL) models proposed here offer such flexibility. For specificity, we first consider the quadratic case and subsequently present the general case. Let $\beta_j^{(1)}(\mathbf{u}^{(j)})$ and $\beta_{j,j'}^{(2)}(\mathbf{u}^{(j)}, \mathbf{u}^{(j')})$ be square integrable functions associated with the linear and quadratic terms in the model, respectively. Specifically,

$$
\begin{aligned}
\vartheta_i \;\; = \;\; & \sum_{j=1}^{J} \int_{D_j} \beta_j^{(1)}(\mathbf{u}^{(j)}) x_{ij}(\mathbf{u}^{(j)}) d\mathbf{u}^{(j)} \\
& + \;\; \sum_{j=1}^{J} \sum_{j'=j}^{J} \int_{D_j} \int_{D_{j'}} \beta_{j,j'}^{(2)}(\mathbf{u}^{(j)}, \mathbf{u}^{(j')}) x_{ij}(\mathbf{u}^{(j)}) x_{ij'}(\mathbf{u}^{(j')}) d\mathbf{u}^{(j)} d\mathbf{u}^{(j')} + \mathbf{z}_i'\boldsymbol{\delta}, \;\; (2.1)
\end{aligned}
$$

where $\boldsymbol{\delta}$ is an n_z-dimensional vector of regression coefficients associated with the covariates, $\mathbf{z}_i$. It is important to clarify that the $\mathbf{x}$'s in this formulation are random latent processes. Thus, the quadratic interaction between two values of $\mathbf{x}$ is considered "random," unlike a classical regression analysis where the $\mathbf{x}$'s are fixed.

Assume that $\{\boldsymbol{\phi}_k^{(j)}(\mathbf{u}^{(j)}) : k = 1, \ldots, \infty\}$ form a complete orthonormal basis corresponding to the j-th signal, $j = 1, \ldots, J$. Then, we have the unique representations of the signals

$$x_{ij}(\mathbf{u}^{(j)}) = \sum_{k=1}^{\infty} \xi_{ij}(k)\phi_k^{(j)}(\mathbf{u}^{(j)}),$$

where $\xi_{ij}(k)$ are expansion coefficients associated with the i-th response and j-th signal. In addition, we also have the unique representations

$$\beta_j^{(1)}(\mathbf{u}^{(j)}) = \sum_{k_1=1}^{\infty} b_j^{(1)}(k_1)\phi_{k_1}^{(j)}(\mathbf{u}^{(j)})$$

$$\beta_{j,j'}^{(2)}(\mathbf{u}^{(j)}, \mathbf{u}^{(j')}) = \sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty} b_{j,j'}^{(2)}(k_1, k_2)\phi_{k_1}^{(j)}(\mathbf{u}^{(j)})\phi_{k_2}^{(j')}(\mathbf{u}^{(j')}),$$

where $b_j^{(1)}(k_1)$ and $b_{j,j'}^{(2)}(k_1, k_2)$ are expansion coefficients associated with square integrable functions of the linear and quadratic terms in the model, respectively, and the ϕ's denote the orthonormal basis functions, as described above. Applying the orthonormality property of the basis functions, and assuming $b_j^{(1)}(k_1)$ and $b_{j,j'}^{(2)}(k_1, k_2)$ are square summable, (2.1) can be alternatively expressed as a function of expansion coefficients $\xi_{ij}(k)$ and the additional covariates

$$\vartheta_i = \sum_{j=1}^{J}\sum_{k_1=1}^{\infty} b_j^{(1)}(k_1)\xi_{ij}(k_1) + \sum_{j=1}^{J}\sum_{j'=j}^{J}\sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty} b_{j,j'}^{(2)}(k_1, k_2)\xi_{ij}(k_1)\xi_{ij}(k_2) + \mathbf{z}_i'\boldsymbol{\delta}. \quad (2.2)$$

Similar to applications in functional data analysis, we consider a finite approximation to the infinite summations in (2.2) and implement further dimension reduction through SSVS (see Section 2.3.4). Also, note that the quadratic interaction of the $\boldsymbol{\xi}$'s is considered nonlinear because $\boldsymbol{\xi}$ are random due to the randomness of the $\mathbf{x}$'s.

Although we focus on the quadratic nonlinear case, we note that higher order interactions follow similarly. For example, consider the more general case of p-th order $(p \geq 3)$ polynomial regression. In this case,

$$
\begin{aligned}
\vartheta_i =\ & \sum_{j_1=1}^{J} \int_{D_{j_1}} \beta_{j_1}^{(1)}(\mathbf{u}_{j_1}) x_{ij_1}(\mathbf{u}_{j_1}) d\mathbf{u}_{j_1} \\
& + \sum_{j_1 \leq j_2}^{J} \int_{D_{j_1} \times D_{j_2}} \beta_{j_1,j_1}^{(2)}(\mathbf{u}_{j_1}, \mathbf{u}_{j_2}) x_{ij_1}(\mathbf{u}_{j_1}) x_{ij_2}(\mathbf{u}_{j_2}) d\mathbf{u}_{j_1} d\mathbf{u}_{j_2} \\
& + \sum_{j_1 \leq j_2 \leq j_3}^{J} \int_{D_{j_1} \times D_{j_2} \times D_{j_3}} \beta_{j_1,j_2,j_3}^{(3)}(\mathbf{u}_{j_1}, \mathbf{u}_{j_2}, \mathbf{u}_{j_3}) x_{ij_1}(\mathbf{u}_{j_1}) x_{ij_2}(\mathbf{u}_{j_2}) x_{ij_3}(\mathbf{u}_{j_3}) d\mathbf{u}_{j_1} d\mathbf{u}_{j_2} d\mathbf{u}_{j_3} \\
& + \ldots + \sum_{j_1 \leq \ldots \leq j_p}^{J} \int_{D_{j_1} \times \ldots \times D_{j_p}} \beta_{j_1,\ldots,j_p}^{(p)}(\mathbf{u}_{j_1}, \ldots, \mathbf{u}_{j_p}) x_{ij_1}(\mathbf{u}_{j_1}) \ldots x_{ij_p}(\mathbf{u}_{j_p}) d\mathbf{u}_{j_1} \ldots d\mathbf{u}_{j_p} \\
& + \mathbf{z}_i' \boldsymbol{\delta},
\end{aligned}
$$

$$(2.3)$$

where $\beta^{(\ell)}(\ell = 1, \ldots, p)$ are square integrable functions associated with the linear, quadratic and ℓ-th order terms in the model, respectively. Similar to the quadratic case, (2.3) can be written in terms of the expansion coefficients

$$
\begin{aligned}
\vartheta_i =\ & \sum_{j_1=1}^{J} \sum_{k_1=1}^{\infty} b_{j_1}^{(1)}(k_1) \xi_{ij_1}(k_1) + \sum_{j_1 \leq j_2}^{J} \sum_{k_1=1,k_2=1}^{\infty} b_{j_1,j_2}^{(2)}(k_1, k_2) \xi_{ij_1}(k_1) \xi_{ij_2}(k_2) \\
& + \sum_{j_1 \leq j_2 \leq j_3}^{J} \sum_{k_1=1,k_2=1,k_3=1}^{\infty} b_{j_1,j_2,j_3}^{(3)}(k_1, k_2, k_3) \xi_{ij_1}(k_1) \xi_{ij_2}(k_2) \xi_{ij_3}(k_3) \\
& + \cdots + \sum_{j_1 \leq \cdots \leq j_p}^{J} \sum_{k_1=1,\ldots,k_p=1}^{\infty} b_{j_1,\ldots,j_p}^{(p)}(k_1, \ldots, k_p) \xi_{ij_1}(k_1) \ldots \xi_{ij_p}(k_p) \\
& + \mathbf{z}_i' \boldsymbol{\delta},
\end{aligned}
$$

where the terms are defined analogously to the quadratic case. As with the quadratic case, the infinite summations are approximated by the finite summations and further dimension reduction is achieved through SSVS as described in Section 2.3.4. The model proposed here is extremely general and can easily accommodate multiple time and/or time-frequency functionals. Process specific interaction coefficients can then be accommodated through dependence priors in the SSVS procedure. Note that the nonsignal covariates could also be allowed to interact with signal predictors.

It is important to note that no form for the complete orthonormal basis functions were specified. In fact, there are many choices for the basis functions among the class of orthogonal (or bi-orthogonal) bases (e.g., Fourier, splines, wavelets, empirical orthogonal functions (EOFs), among others). The choice of basis function is somewhat subjective, but there can be advantages and disadvantages depending on the application. In the remainder, we use the Karhunen-Loéve decomposition (Wikle and Cressie, 1999; Jolliffe, 2010) to provide the following infinite decompositions for the j-th signal

$$x_{ij}(\mathbf{u}^{(j)}) = \sum_{k=1}^{\infty} \xi_{ij}(k)\phi_k^{(j)}(\mathbf{u}^{(j)})$$

where

$$\xi_{ij}(k) = \int x_{ij}(\mathbf{u}^{(j)})\phi_k^{(j)}(\mathbf{u}^{(j)})d\mathbf{u}_j$$

with $\mathrm{E}(\xi_{ij}) = 0$ and $\mathrm{Var}(\xi_{ij}) = \lambda_k^{(j)}$. Thus, $\lambda_k^{(j)}$ and $\{\phi_k^{(j)}(\cdot)\}$ are eigenvalues and eigenfunctions of the covariance function of the j-th signal, respectively. In addition, we assume a normal distribution for $\xi_{ij} \sim N(0, \lambda_k^{(j)})$. In practice, we can calculate eigenvalues and eigenfunctions from the empirical covariance matrix of the signals. The obtained eigenfunctions are called empirical orthogonal functions (EOFs), as is

common in spatial and spatio-temporal statistics (Cressie and Wikle, 2011, Chapter 5).

2.3.4 Stochastic Search Variable Selection

Although the basis expansion formulation is used to simplify the FPNL integrals and facilitate parameter dimension reduction, the dimensionality still increases not only by the number of signals but also by an order of magnitude for each additional polynomial interaction. In addition, there is no reason that all time-frequency features should be important relative to the responses (Holan et al., 2010; Wikle and Holan, 2011; Holan et al., 2012). Therefore, the FPNL model we propose makes use of a SSVS algorithm. Such algorithms provide an effective means of model selection when interest lies in considering a large number of potential submodels (see George, 2000, for a detailed overview).

In general, there are two primary types of SSVS priors (George and McCulloch, 1997). One comes from a mixture of normals and the other is a mixture of a Dirac function at zero and a normal distribution. Many variations of these two basic algorithms have been proposed and applied across a broad range of applications (see O'Hara and Sillanpää, 2009; Stingo et al., 2012, and the references therein). Both of these prior formulations can be incorporated into our model. For the sake of brevity, we only describe the latter approach here. Comprehensive details on the first approach can be found in George and McCulloch (1993, 1997); Holan et al. (2010), and Wikle and Holan (2011).

Let $\boldsymbol{\zeta}^{(\ell)}$ be the vector of linear coefficients when $\ell = 1$, the vector of quadratic coefficients when $\ell = 2$, and the non-signal coefficients when $\ell = 3$ (e.g., the vectors

$\mathbf{b}^{(1)}$, $\mathbf{b}^{(2)}$ and $\boldsymbol{\delta}$, respectively, in the quadratic model). Note, although we use $\ell = 3$ here, this should not be confused with the third-order interaction effects. Then, letting $\gamma_k^{(\ell)}$ denote a latent binary variable associated with the k-th element $\zeta_k^{(\ell)}$ of $\boldsymbol{\zeta}^{(\ell)}$, we have the following prior specification

$$\zeta_k^{(\ell)}|\gamma_k^{(\ell)}, \sigma_\ell^2 \sim (1 - \gamma_k^{(\ell)})\delta_0(\zeta_k^{(\ell)}) + \gamma_k^{(\ell)}N(0, \kappa_k^{(\ell)}\sigma_\ell^2), \tag{2.4}$$

where $\delta_0(\cdot)$ is the Dirac function at zero and the $\kappa_k^{(\ell)}$s are hyperparameters to be chosen. Using this prior, if $\gamma_k^{(\ell)} = 0$ then $\zeta_k^{(\ell)} = 0$ and if $\gamma_k^{(\ell)} = 1$ then $\zeta_k^{(\ell)} \sim N(0, \kappa_k^{(\ell)}\sigma_\ell^2)$. Usually, $\kappa_k^{(\ell)}$s are fixed at a common value (i.e., $\kappa_k^{(\ell)} = \kappa_{k'}^{(\ell)}$) and the conjugate prior, $IG(a_\ell, b_\ell)$, is often chosen for σ_ℓ^2 with a_ℓ and b_ℓ specified.

Prior specification requires the choice of prior distribution for $\boldsymbol{\gamma} = (\boldsymbol{\gamma}^{(1)'}, \boldsymbol{\gamma}^{(2)'}, \boldsymbol{\gamma}^{(3)'})'$, where $\boldsymbol{\gamma}^{(\ell)} = (\gamma_1^{(\ell)}, \ldots, \gamma_K^{(\ell)})'$, $\ell = 1, 2, 3$. This is often taken to be a product of independent Bernoulli distributions. However, the independence assumption for $\boldsymbol{\gamma}$ might not be appropriate for our model because the quadratic terms are likely to be related to the linear terms in the signals (e.g., similar to ANOVA models, where we assume no interaction effects are significant unless the main effects are significant). Thus, we use hierarchical priors that can describe relationships between linear and quadratic terms (Chipman, 1996). That is, we first set a common probability for the linear terms in the signals and covariate terms, $P(\gamma_k^{(1)} = 1) = P(\gamma_{k'}^{(3)} = 1) = P_1$. Note that it is not necessary that each of the elements of the linear and covariate terms have the same probability. After giving probabilities to the linear terms, we construct

conditional probabilities for the quadratic terms,

$$p(\gamma_{q,r}^{(2)} = 1 | \gamma_q^{(1)}, \gamma_r^{(1)}) = \begin{cases} P_{00}, & \text{if } (\gamma_q^{(1)}, \gamma_r^{(1)}) = (0,0); \\ P_{01}, & \text{if } (\gamma_q^{(1)}, \gamma_r^{(1)}) = (0,1) \text{ or } (1,0); \\ P_{11}, & \text{if } (\gamma_q^{(1)}, \gamma_r^{(1)}) = (1,1), \end{cases}$$

where $\gamma_{q,r}^{(2)}$ corresponds to the parameters associated with the interaction of the q- and r-th linear coefficients, i.e., $\gamma_q^{(1)}$ and $\gamma_r^{(1)}$. The choice of (P_{00}, P_{01}, P_{11}) are typically based on different principles of variable selection. For example, one might choose $(P_{00}, P_{01}, P_{11}) = (0, 0, p)$. This setting implies that for a quadratic term to be included, all corresponding linear terms must also be included in the model. Comprehensive details can be found in Chipman (1996). Many alternative hyperparameter specifications have been proposed for specific applications (see Vannucci and Stingo, 2010, and the references therein for an overview).

In the context of "big data" problems, especially those involving time-frequency and functional covariates, SSVS has been shown to be extremely effective (Holan et al., 2010; Wikle and Holan, 2011; Holan et al., 2012). Nevertheless, many other model selection methods could be employed, including deviance information criterion (DIC) (Spiegelhalter et al., 2002), posterior predictive loss (Gelfand and Ghosh, 1998), and cross-validation, among others. The advantage of using SSVS over other model selection approaches is that SSVS produces a model averaged classification while simultaneously achieving feature extraction (i.e., identifies important regions of the time-frequency representation and/or function). In contrast, other model selection approaches typically yield a single "preferred" model lacking robust feature identification. More importantly, due to the high-dimensionality, typical of the prob-

lems presented here, classical model selection methods are infeasible in searching the candidate model space.

2.4   Bayesian Inference and Prediction

As described in Section 2.3, the model we propose is a joint nonlinear generalized mixed effects model consisting of a signal measurement model and a generalized linear model with nonlinear functional predictors in the link function. However, for most nonlinear models, it is difficult to derive an explicit form for the marginal likelihood corresponding to the response variables. A two-stage approach avoids this problem with the following steps: (1) one obtains the predictors of the expansion coefficients ξ_{ij}, say $\widehat{\xi}_{ij}$, and then (2) estimates the parameters of the response model by replacing ξ_{ij} with $\widehat{\xi}_{ij}$ (see Müller and Stadtmüller, 2005; Crainiceanu et al., 2009; Holan et al., 2010; Yao and Müller, 2010; Holan et al., 2012, and the references therein). For example, in the generalized linear model setting, Holan et al. (2010) produced the expansion coefficients through EOF analysis and then implemented a Bayesian SSVS procedure on the coefficients associated with the link function with the goal of dimension reduction and feature extraction, as well as predicting the response variables. Although it is easier to implement, the two-stage procedure has potential problems. In particular, Crainiceanu et al. (2009) illustrate how such a substitution induces bias in the estimators due to the misspecification of the variability associated with the fixed and mixed effect parameters.

Due to the potential problems associated with the two-stage procedure, James (2002) uses a non-Bayesian approach which treats the latent variables (i.e., true sig-

nal) in the measurement model as missing values and makes inference jointly on both measurement model and response model through the expectation-maximization (EM) algorithm. Conversely, Crainiceanu et al. (2009) use the Bayesian framework to facilitate fitting the model jointly when the measurement model is a multilevel functional model and the response model is a generalized linear model. The joint Bayesian analysis we propose is more complicated because our model has multiple measurement models and the true process is modeled nonlinearly.

The main difference between the two-stage and joint Bayesian approach is that we need to estimate the unknown expansion coefficient ξ_{ij} through a MCMC procedure, as well as the variance $\sigma_{\epsilon,j}^2$ of the measurement error for the j-th signal. In Section 2.3.3, the expansion coefficients ξ_{ij} were assumed to have the prior, $N(0, \lambda_k^{(j)})$. Further, we use independent inverse gamma priors for $\lambda_k^{(j)}$; $\lambda_k^{(j)} \sim IG(a_{\lambda_k^{(j)}}, b_{\lambda_k^{(j)}})$, $j = 1, \ldots, J$, $k = 1, \ldots, K$ and $\sigma_{\epsilon,j}^2 \sim IG(a_{\epsilon,j}, b_{\epsilon,j})$ and specify $a_{\lambda_k^{(j)}}$, $b_{\lambda_k^{(j)}}$, $a_{\epsilon,j}$ and $b_{\epsilon,j}$ such that each prior density has a mean equal to 2 and variance equal to 10,000. Note that the prior distributions chosen are vague but proper, and thus, we maintain propriety of the posterior distribution while imparting little impact on the analysis.

To be specific, in order to analyze shovelnose sturgeon spawning success, we focus on a probit model with binary response variables. Other types of response variables can be treated similarly. Full details of the MCMC estimation algorithms are provided in the Appendix.

2.5 Simulated Examples

In this section, we consider the performance of the FPNL model through simulated examples similar to our motivating data that is analyzed in Section 2.6. Specifically, we examine a probit model and consider estimation under various levels of magnitude for the signal and noise variances (i.e., signal-to-noise ratio). In addition, we consider the performance of a two-stage implementation relative to our fully Bayesian implementation.

The response variables were simulated from a probit model with quadratic predictor

$$\Phi^{-1}(p_i) = \sum_{j=1}^{2} \sum_{k_1=1}^{K_1} b_j^{(1)}(k_1) \xi_{ij}(k_1) + \sum_{j=1}^{2} \sum_{j'=j}^{2} \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} b_{j,j'}^{(2)}(k_1, k_2) \xi_{ij}(k_1) \xi_{ij}(k_2),$$

for $i = 1, \ldots, n$, where $n = 100$. We note that K_1 and K_2 depend on j. We used the functional predictor $x_{ij}(\mathbf{u}_j) = \sum_{k=1}^{5} \xi_{ij}(k) \phi_k^{(j)}(\mathbf{u}_j)$, where $\xi_{ij} \sim D(\lambda_k^{(j)})$, with $D(\cdot)$ a distribution to be specified below. In this example, $\{\phi_k^{(1)}(\mathbf{u}_1)\}$ and $\{\phi_k^{(1)}(\mathbf{u}_2)\}$ are the sets of the top five eigenfunctions in terms of accounting for the most variation of the depth and temperature profiles from the fish tracking data example presented in Section 2.6, respectively. Choosing 5 EOFs for the simulation helps facilitate rapid computation and constitutes the entire basis set for the simulated data. We set $b_1(1) = 1.0$, $b_1(3) = -1.5$, $b_2(2) = 1.0$, and $b_{1,2}(1,2) = -1.4$, with all remaining b-parameters set to 0. Thus, the quadratic predictor can be rewritten as

$$\Phi^{-1}(p_i) = \xi_{i1}(1) - 1.5\xi_{i1}(3) + \xi_{i2}(2) - 1.4\xi_{i1}(1)\xi_{i2}(2). \tag{2.5}$$

To model the simulated data, we consider probit models via the two approaches described in Section 2.4. Specifically, denote a two-stage and joint Bayesian procedure with linear predictors as M1 and M2, respectively. In addition, let M3 and M4 denote the model with a quadratic predictor using the two-stage and joint Bayesian approach, respectively. In general, it is difficult to know the optimal choices for the SSVS parameters (see Section 2.3.4). Therefore, we conducted a sensitivity analysis that considered various possible values for these parameters and evaluated the performance under several scenarios. Specifically, for M1 and M2, we set $\kappa_j^{(1)} = \{0.001, 0.01, 0.1, 1, 10, 100, 1000\}$ with $P_1 = 0.5$. For M3 and M4, we considered the following combinations of $(\kappa_j^{(1)}, \kappa_j^{(2)})$: $(0.001, 100)$, $(0.001, 10)$, $(0.001, 1)$, $(0.01, 10)$, $(0.01, 1)$, $(0.1, 1)$, and $(1, 1)$. We considered the hierarchical prior $(P_{00}, P_{01}, P_{11}) = (0.01, 0.01, 0.5)$, referred to as "weak heredity" by Chipman (1996). When the covariate space is large, this setting is more appropriate than the "strong heredity" version, suggested by Chipman (1996), in which $P_{00} = P_{01} = 0$.

To evaluate model performance, we considered the out-of-sample classification rate using a leave-one-out-cross-validation experiment (across all of the $n = 100$ observations). The classification rate was obtained by considering posterior samples of the mean probability of "successful spawning" over all iterations of the SSVS MCMC iterations. The classification was assigned a 1 ("success") if this probability was greater than or equal to 0.5, and a 0 ("non-success") if this probability was less than 0.5. Similar to Holan et al. (2010), this classification takes advantage of the Bayesian model averaging, thereby reducing uncertainty in the model selection stage (Draper, 1995; Hoeting et al., 1999), while selecting important covariates/features through the SSVS algorithm. The SSVS sampler results are based on 35,000 iterations with a

5,000 iteration burn-in.

Consider the situation where one observes the noisy predictors $s_{ij}(\mathbf{u}_j) = x_{ij}(\mathbf{u}_j) + \epsilon_{ij}(\mathbf{u}_j)$, where $\epsilon_{ij}(\mathbf{u}_j) \sim N(0, \sigma^2_{\epsilon,j})$ is the measurement error. We set $\xi_{ij} \sim N(0, \lambda_k^{(j)})$, where $\lambda_k^{(j)}$ are eigenvalues of the depth and temperature profiles associated with the fish tracking data discussed in Section 2.6: when $j = 1$, $\lambda_1^{(1)} = (1140.3)^2$, $\lambda_2^{(1)} = (720.5)^2$, $\lambda_3^{(1)} = (530.8)^2$, $\lambda_4^{(1)} = (480.9)^2$, and $\lambda_5^{(1)} = (429.2)^2$; when $j = 2$, $\lambda_1^{(2)} = (3.0597)^2$, $\lambda_2^{(2)} = (1.7659)^2$, $\lambda_3^{(2)} = (1.4325)^2$, $\lambda_4^{(2)} = (0.9477)^2$, and $\lambda_5^{(1)} = (0.8666)^2$. To examine the effect of different noise variances, we considered three different combinations of magnitudes for the noise variances $(\sigma^2_{\epsilon,1}, \sigma^2_{\epsilon,2})$: $(1^2, 0.01^2)$, $(400^2, 1^2)$, and $(800^2, 2^2)$. We used the eigenfunctions associated with the depth and temperature profile associated with the fish tracking data discussed in Section 2.6.

Table 2.1 shows the out-of-sample classification results for each model, as well as the area under the *receiver operating characteristic* (ROC) curve (AUROC; Hosmer and Lemeshow, 2000). The ROC curve measures the true positive rate along side the false positive rate and AUROC values close to one indicate superior performance. When $(\sigma^2_{\epsilon,1}, \sigma^2_{\epsilon,2}) = (1^2, 0.01^2)$, both quadratic models M3 and M4 show superior performance compared to the linear models M1 and M2. In this case, M3 and M4 perform similarly (i.e., there is little difference between two-staged and joint Bayesian analysis). When $\sigma^2_{\epsilon,1}$ and $\sigma^2_{\epsilon,2}$ increases to 1^2 and 400^2, respectively, both M3 and M4 still show superior performance relative to M1 and M2. However, in this scenario, M4 performs better than M3. In contrast, in the last scenario, M1 is superior to all models. This is because the variance of $\sigma^2_{\epsilon,2} = 2^2$ is larger than the variance of one of the signals in the interaction term $(\lambda_2^{(2)} = (1.7659)^2)$. As expected, the noise variance significantly affects classification results, with the quadratic model performing the

best, and with the one-stage model performing slightly better in many cases.

Finally, it is of interest to evaluate models M1 and M2 relative to M3 and M4 when data arise from a linear model. Using the simulation setup previously described, we simulated data from the linear predictors of (2.5). Model performance was assessed using a sensitivity analysis with the same hyperparameter settings as the quadratic case and we considered the out-of-sample classification rate using a leave-one-out-cross-validation experiment. In summary, when simulating from a linear model, the linear and quadratic models produce similar classification results (results not shown). The reason for the similar classification is that the quadratic terms are chosen in the model with low probability (i.e., significantly less relative to the linear terms across both quadratic models and all signal to noise ratios). This aspect of the algorithm is particularly advantageous.

2.6   Sturgeon Spawning Success Analysis

To illustrate the effectiveness of our approach we provide an analysis of the motivating data on shovelnose sturgeon spawning behavior previously discussed in Section 2.2. Specifically, using DST depth and temperature data collected in 2007, we predict spawning success of shovelnose sturgeon and characterize the salient features of the DST depth and temperature profiles associated with this outcome. In this context, the measure that is used to assess spawning success is the same as in Holan et al. (2009) and is given by the recapture oocyte ratio or spawning index (SI). This index measures the ratio of mature oocytes (eggs) to early stage oocytes as an indicator of the completeness of spawning. Lower ratios indicate more complete spawning.

As a result of individual fish being tagged with DSTs at different time points (i.e., the onset times of DST recordings were different), our analysis considers the study period from April 19 to June 3. This constitutes the maximum time range that includes the majority of the data. During this period, 41 of the 99 tagged fish were recaptured. We note that one DST instrument failed and, thus, that fish was not considered in this analysis. Of the recaptured fish, 29 were successful spawners and 11 were unsuccessful spawners.

The spectrogram for each DST depth time series was calculated using the signal processing toolbox function *spectrogram* (short-time Fourier transform) in Matlab®. Specifically, we chose the default Hamming moving window (Oppenheim and Schafer, 2009) of length 64 time periods with an overlap of 10 time periods in this analysis. We also evaluated the sensitivity of this choice by examining several combinations of window length and overlap (16×8, 32×10, 32×16, 64×10, 64×32, 128×10 and 128×64). Such specifications showed no difference with regards to variable selection or classification in in-sample analysis when we considered a quadratic model with depth and temperature.

In this example, we find the first 10 and 15 EOFs of the depth spectrogram account for approximately 82.3% and 89.6% of the variation, respectively. Additionally, the first 10 and 15 EOFs of the temperature time series account for approximately 86.3% and 91.5% of the variation, respectively. We evaluate several combinations of the number of EOF components of the depth spectrogram and temperature time series (10×10, 10×15, 15×10, and 15×15). The in-sample analysis showed no difference with regards to classification for such specifications. However, the result of variable selection in the in-sample analysis indicate that EOFs 11-15 are less important to the

spawning success. Consequently, we consider the parsimonious model based on the first 10 EOFs of the depth spectrogram and temperature time series.

To predict spawning success, we considered models M1, M2, M3, and M4 described in Section 2.5. M1 and M2 include the time-frequency depth predictor and functional time series temperature predictor. M3 and M4 include both the time-frequency depth predictor, functional time series temperature predictor, and their interaction. In addition, let M5 and M6 denote the model with linear predictor including only the time-frequency depth predictor using the two-stage and joint Bayesian approach, respectively. We also conducted a sensitivity analysis using the same sets of SSVS parameters described in Section 2.5. Finally, we used leave-one-out-cross-validation to evaluate model performance. Each cross-validation procedure consists of 50,000 MCMC iterations with the first 25,000 discarded for burn-in.

Table 2.2 shows the out-of-sample classification results associated with M1, M2, M3, M4, M5 and M6. In general, all of the models have the same performance in terms of classifying unsuccessful spawners (72.73% correct), with M5 and M6 performing superior to M1 and M2. This result demonstrates that including the time series for temperature as a functional linear predictor provides diminished predictive power for successful spawners. Such a result is not surprising because the sturgeon's body temperature is highly regulated by the temperature of the river, with little separation between successful and unsuccessful spawners.

Both quadratic models M3 and M4 are superior compared to the linear models M1, M2 and the depth only models, M5 and M6. Specifically, M4 shows superior performance in classifying successful spawners relative to M3 (72.41% vs. 65.52%). Furthermore, Table 2.3 shows the 15 most commonly selected variables for M1, M2,

M3 and M4, as well as the 10 most commonly selected variables for M5 and M6, based on the posterior probability of inclusion in the model. Depth EOF 6, 7, 5, and 1 are commonly selected in all the models. This implies that M1, M2, M3, and M4 select the same important information about sturgeon spawning behavior from the DST depth profile as M5 and M6. In contrast, M1, M2, M3, and M4 selected several common temperature EOFs. However, according to the classification results, these EOFs may have less impact on spawning success. More specifically, in terms of classification, the interaction between depth EOF 1 and temperature EOF 2 that appears in models M3 and M4 results in a better predictor of successful spawning behavior. This latter finding is consistent with the notion that successful spawning may rely on multiple cues happening simultaneously; e.g., the sturgeon exhibit less variability in their depth profile as they reach a suitable body temperature. These results suggest that important information on spawning success is obtained from a model that includes depth, temperature, and importantly, their interaction.

Holan et al. (2010) demonstrate the effectiveness of signal reconstruction to distinguish the important signal features driving mating success in insect communication signals. An added facet of our approach is that it can reveal important aspects of signal interactions corresponding to the response variables (spawning success). Now, because M3 and M4 suggest that the interaction of depth EOF 1 and temperature EOF 2 is an important predictor corresponding to spawning success, we reconstructed the depth spectrogram based on the depth EOF 1, as well as temperature curve on the temperature EOF 2; see Figure 2.2. This figure shows that the spectrogram of successful spawners has less variability overall relative to the unsuccessful spawners, and that a low-frequency region in the period between May 10 and May 20 may be

an important indicator of spawning success. Moreover, the mean difference of the re-constructed temperature expresses a deep trough/minimum in the period before the lowest values of the mean difference of the reconstructed depth spectrogram, implying a potential interaction between depth variability and temperature.

The model results can be used to suggest a hypothesized time period in which the spawning may have occurred for these shovelnose in 2007 (i.e., May 10 through May 20, 2007). Spring (April-June) is when most shovelnose sturgeon spawning seems to occur in the Lower Missouri River, with temperature a key factor influencing timing (see Wildhaber et al., 2007, 2011a, and references therein). However, this is no more than a hypothesis because recapture and validation of spawning of these fish occurred weeks to months after the hypothesized time period. To provide support for development of this type of hypothesis from this model will require validation through verification of actual spawning dates for additional fish for which DST records are available.

2.7 Discussion

Ecological data are often high-dimensional in terms of the response predictors. We consider the situation where one has extremely high-dimensional nonstationary time series predictors (functions) of responses that are not necessarily time-referenced. In our case the number of subjects (fish) in the response is fairly low-dimensional but the covariates themselves can be considered "big data." High-frequency time-series predictors, such as those coming from DSTs, typically exhibit nonstationarity. Given the difficulty of establishing important features from such signals subjectively,

36

it is necessary to develop a general methodology that can select important signal components. Furthermore, most predictors do not influence responses unilaterally, and thus we consider possible nonlinear interactions between multiple time signals.

An efficient way to represent nonstationary high-frequency time series signals is in the time-frequency domain (e.g., via a spectrogram). Such representations show regions in time-frequency space that exhibit high variability. However, a spectrogram representation comes at the price of having a substantially large number of time-frequency "pixels" (typically many thousands). In practice, it is not feasible to consider each of these pixels as a potential covariate. However, the pixels are often dependent, as important features in time-frequency space occur over adjacent frequency bands and across multiple time points. Thus, it is convenient to consider the spectrogram as an "image" and utilize a low-rank spatial representation, similar to what is commonly done in spatial and spatio-temporal statistics (see Cressie and Wikle, 2011, for an overview). We demonstrate how such a methodology can be used within a time-frequency functional polynomial nonlinear (FPNL) model. In particular, the main contribution we provide is to illustrate how one can account for signal uncertainty as well as accommodate nonlinear interactions in an ecological (or other scientific) model in which the signals are extremely high-dimensional.

The effectiveness of the methodology we propose is demonstrated through the motivating example of predicting spawning success of shovelnose sturgeon. In addition, our approach produces feature extracted curves and images that highlight the important aspects of depth variability and temperature, as they relate to successful spawning. Our analysis suggests that depth, temperature, and their interaction are all important predictors for successful spawning of shovelnose sturgeon in the Lower

Missouri River and details the specific components of each.

We note that similar to implementation of low-rank spatial and spatio-temporal random effects models, the choice of basis functions is somewhat arbitrary (e.g., Wikle, 2010). An interesting extension is to consider the model performance under different basis functions. In particular, it may be that jointly specified basis functions (e.g., for depth and temperature simultaneously) may provide a more efficient dimension reduction and may facilitate superior interpretation. More investigation into the importance of the uncertainty associated with the signal, both in terms of signal-to-noise-ratio and registration, is warranted. These issues will be explored elsewhere.

2.8   Appendix: MCMC Algorithm

The MCMC algorithms differ for different types of response variables (i.e., continuous, count, and binary variables) and different model structures. For specificity, we consider a probit model (for a Bernoulli response) with quadratic predictor. In particular, suppose we have binary response variables $Y_i = \{0, 1\}$; following Albert and Chib (1993), we define a continuous latent variable Z_i such that

$$
Y_i = \begin{cases} 1; & \text{if} \quad Z_i > 0, \\ 0; & \text{if} \quad Z_i \leq 0, \end{cases}
$$

with

$$
Z_i | \cdot \sim N(\vartheta_i, 1).
$$

where

$$\vartheta_i = \sum_{j=1}^{J}\sum_{k_1=1}^{K_1} b_j^{(1)}(k_1)\xi_{ij}(k_1) + \sum_{j=1}^{J}\sum_{j'=j}^{J}\sum_{k_1=1}^{K_1}\sum_{k_2=1}^{K_2} b_{j,j'}^{(2)}(k_1,k_2)\xi_{ij}(k_1)\xi_{ij}(k_2) + \mathbf{z}_i'\boldsymbol{\delta},$$

as described in Section 2.3.3.

We specify a SSVS prior described in Section 2.3.4 for the coefficients of the model. Recall that $\boldsymbol{\zeta}^{(\ell)}$ denotes as the vector of linear coefficients when $\ell = 1$, the vector of quadratic coefficients when $\ell = 2$, and the non-signal coefficients when $\ell = 3$. In addition, $\gamma_k^{(\ell)}$ denotes a latent binary variable associated with the k-th element $\zeta_k^{(\ell)}$ of $\boldsymbol{\zeta}^{(\ell)}$. Then, we have the following prior specification

$$\zeta_k^{(\ell)}|\gamma_k^{(\ell)}, \sigma_\ell^2 \sim (1-\gamma_j^{(\ell)})\delta_0(\zeta_k^{(\ell)}) + \gamma_k^{(\ell)}N(0,\kappa_k^{(\ell)}\sigma_\ell^2),$$

where $\delta_0(\cdot)$ is the Dirac function at zero, the $\kappa_k^{(\ell)}$s have a common value, say $\kappa^{(\ell)}$, and $\boldsymbol{\gamma} = (\boldsymbol{\gamma}^{(1)'}, \boldsymbol{\gamma}^{(2)'}, \boldsymbol{\gamma}^{(3)'})'$ have hierarchical priors.

2.8.1 Full Conditionals

The full conditional of Z_i is a truncated normal distribution given by

$$Z_i|\cdot = \begin{cases} N_{0,\infty}(\vartheta_i, 1); & \text{if } Y_i = 1, \\ N_{-\infty,0}(\vartheta_i, 1); & \text{if } Y_i = 0. \end{cases}$$

However, because the SSVS prior is a mixture of a Dirac function at zero and a normal distribution, the Gibbs sampler gets trapped easily at coefficients equal to zero (George and McCulloch, 1997). To avoid this problem, we sample coefficients

and associated latent binary variable simultaneously (Geweke, 1996). Letting w.p. denote "with probability," the full joint conditional of $(\zeta_k^{(\ell)}, \gamma_k^{(\ell)})$ is

$$
\zeta_k^{(\ell)} = \begin{cases} 0, & \text{w.p.} \quad 1 - \tilde{P}_k^{(\ell)}, \\ N(\mu, (\tilde{\xi}_k^{(\ell)'} \tilde{\xi}_k^{(\ell)})^{-1}), & \text{w.p.} \quad \tilde{P}_k^{(\ell)}, \end{cases}
$$

with $\tilde{\xi}_k^{(\ell)} = [\xi_{1k}^{(\ell)}, \ldots, \xi_{Nk}^{(\ell)}, (\kappa^{(\ell)} \sigma_\ell^2)^2]'$. Here, $\{\xi_{ik}^{(\ell)}\}$ are the associated expansion coefficients, $\mu = (\tilde{\xi}_k^{(\ell)'} \tilde{\xi}_k^{(\ell)})^{-1} \tilde{\xi}_k^{(\ell)'} \mathbf{Z}^*$, where $\mathbf{Z}^* = [Z_1^*, \ldots, Z_N^*, 0]'$ and $Z_i^* = Z_i - \vartheta_i$ with $\zeta_k^{(\ell)} = 0$. Finally, $\tilde{P}_k^{(\ell)} = P_k^{(\ell)} \times R / (1 - P_k^\ell + P_k^{(\ell)} \times R)$, where $P_k^{(\ell)}$ is the prior of $\gamma_k^{(\ell)}$ and

$$
R = \left(\kappa^{(\ell)} \sigma_\ell^2 \sum_{i=1}^N (\xi_{ik}^{(\ell)})^2 + 1 \right)^{-\frac{1}{2}} \exp \left(\frac{\kappa^{(\ell)} \sigma_\ell^2 (\sum_{i=1}^N Z_i^* \xi_{ik}^{(\ell)})^2}{2(\kappa^{(\ell)} \sigma_\ell^2 \sum_{i=1}^N (\xi_{ik}^{(\ell)})^2 + 1)} \right).
$$

The full conditional of σ_ℓ^2 is

$$
\sigma_\ell^2 | \cdot \sim IG \left(\frac{n_\ell + a_\ell}{2}, \frac{\sum_{k=1}^K (\zeta_k^{(\ell)})^2}{2\kappa^{(\ell)}} + \frac{b_\ell}{2} \right),
$$

where n_ℓ is the number of coefficients not equal to zero. The full conditional of $\lambda_k^{(j)}$ is

$$
\lambda_k^{(j)} | \cdot \sim IG \left(\frac{N + a_{\lambda_k^{(j)}}}{2}, \frac{b_{\lambda_k^{(j)}} + \sum_{i=1}^N (\xi_{ij}(k))^2}{2} \right).
$$

The full conditional of $\sigma_{\epsilon,j}^2$ is

$$
\sigma_{\epsilon,j}^2 | \cdot \sim IG \left(\frac{n_j N + a_{\epsilon,j}}{2}, \frac{b_{\epsilon,j} + \sum_{i=1}^N ||\mathbf{s}_{ij} - \boldsymbol{\phi}^{(j)} \boldsymbol{\xi}_{ij}||^2}{2} \right),
$$

where n_j is the dimension of the jth signal. The full conditional of ξ_{ij} is

$$\xi_{ij}|\cdot \propto \exp\left(\frac{(Z_i - \vartheta_i)^2 - \boldsymbol{\xi}'_{ij}(\boldsymbol{\lambda}^{(j)})^{-1}\boldsymbol{\xi}_{ij}}{2} - \frac{||\mathbf{s}_{ij} - \boldsymbol{\phi}^{(j)}\boldsymbol{\xi}_{ij}||^2}{2\sigma^2_{\epsilon,j}}\right).$$

The full conditional of ξ_{ij} has no analytical form. Therefore, we used a random walk Metropolis-Hastings algorithm with a normal proposal distribution having mean equal to current value and variance chosen to yield an acceptance rate of approximately 30%.

Table 2.1: Out-of-sample prediction results of the simulated examples presented in Section 2.5 using the "model averaged" probit model via the SSVS procedure. These results are based on the leave-one-out cross-validation analysis (across all of the $n = 100$ observations). M1 and M2 denote the model with a linear predictor using the two-stage and joint Bayesian approach, respectively. M3 and M4 denote the model with a quadratic predictor using the two-stage and joint Bayesian approach, respectively. Finally, AUROC denotes area under the receiver operating characteristic curve.

		Classify To:					
$\{\sigma_{\epsilon,1}^2,\sigma_{\epsilon,2}^2\}=$		$\{1^2,0.01^2\}$		$\{400^2,1^2\}$		$\{800^2,2^2\}$	
Classify From:		0	1	0	1	0	1
M1:	0	88.46%	11.54%	76.92%	23.08%	71.15%	28.85%
	1	10.42%	89.58%	22.92%	77.08%	22.92%	77.08%
	AUROC	.9127		.8654		.7989	
M2:	0	88.46%	11.54%	75.00%	25.00%	67.31%	32.69%
	1	10.42%	89.58%	14.29%	79.17%	25.00%	75.00%
	AUROC	.9127		.8743		.7849	
M3:	0	96.15%	3.85%	78.85%	21.15%	67.31%	32.69%
	1	2.08%	97.92%	16.67%	83.33%	25.00%	75.00%
	AUROC	.9952		.8774		.7865	
M4:	0	94.23%	5.77%	80.77%	19.23%	67.31%	32.69%
	1	2.08%	97.92%	14.58%	85.42%	22.92%	77.08%
	AUROC	.9956		.8790		.7821	

Table 2.2: Out-of-sample prediction results of the sturgeon spawning success analysis from Section 2.6 using the "model averaged" probit model via the SSVS procedure. These results are based on a leave-one-out cross-validation analysis (across all of the $n = 40$ observations). Recall, M1 and M2 denote the model with a linear predictor of the time-frequency depth and functional time series temperature using the two-stage and joint Bayesian approach, respectively. The model with a quadratic predictor including time-frequency depth predictor, functional time series temperature predictor and their interaction by using the two-stage and joint Bayesian approach are denoted by M3 and M4, respectively. Finally, M5 and M6 denote the model with a linear predictor of the time-frequency depth using the two-stage and joint Bayesian approach, respectively. Note that unsuccessful and successful spawning are coded as 0 and 1, respectively, and that AUROC denotes area under the receiver operating characteristic curve.

| | Classify To: | | | | | |
| | M1 | | M2 | | M3 | |
Classify From:	0	1	0	1	0	1
0	72.73%	27.27%	72.73%	27.27%	72.72%	27.27%
1	48.28%	51.72%	48.28%	51.72%	34.48%	65.52%
AUROC	.5893		.5925		.6771	

| | Classify To: | | | | | |
| | M4 | | M5 | | M6 | |
Classify From:	0	1	0	1	0	1
0	72.73%	27.27%	72.73%	27.27%	72.72%	27.27%
1	27.59%	72.41%	42.38%	58.62%	42.38%	58.62%
AUROC	.6740		.6834		.6708	

Table 2.3: Top selected variables from the SSVS cross-validation from each model used in the analysis in Section 2.6. Here, di and tj denote the ith EOF of depth and the jth EOF of temperature, respectively. The percentages shown are the mean of the posterior probability of the variable being included in the model for each of the 40 cross-validation runs. Definitions of models M1-M6 are provided in Sections 2.5 and 2.6. Note that unsuccessful and successful spawning are coded as 0 and 1, respectively.

M1		M2		M3	
Variable	% of models	Variable	% of models	Variable	% of models
(d6,0)	77	(d6,0)	78	(d1,t2)	58
(d7,0)	67	(d7,0)	71	(d6,0)	55
(d5,0)	57	(d5,0)	61	(d7,0)	54
(t6,0)	52	(t6,0)	52	(t6,0)	51
(d1,0)	48	(t8,0)	51	(d5,0)	51
(t8,0)	45	(t10,0)	50	(d1,0)	50
(t1,0)	42	(t9,0)	50	(t8,0)	49
(t3,0)	41	(d1,0)	48	(t2,0)	48
(d3,0)	41	(t7,0)	48	(t3,0)	48
(t5,0)	38	(t5,0)	46	(t1,0)	48
(t7,0)	38	(t4,0)	43	(d3,0)	47
(t2,0)	38	(t3,0)	42	(t4,0)	47
(d4,0)	37	(t1,0)	41	(t7,0)	47
(d10,0)	37	(d3,0)	38	(t5,0)	46
(d2,0)	37	(t2,0)	38	(d9,0)	46

M4		M5		M6	
Variable	% of models	Variable	% of models	Variable	% of models
(d6,0)	56	(d6,0)	78	(d6,0)	77
(d7,0)	55	(d7,0)	71	(d7,0)	71
(d5,0)	52	(d5,0)	61	(d5,0)	61
(d1,0)	51	(d1,0)	47	(d1,0)	47
(t6,0)	50	(d3,0)	36	(d3,0)	36
(t8,0)	50	(d4,0)	34	(d4,0)	34
(t10,0)	50	(d10,0)	33	(d10,0)	33
(t9,0)	49	(d2,0)	32	(d2,0)	32
(t7,0)	49	(d8,0)	31	(d9,0)	31
(t5,0)	49	(d9,0)	31	(d8,0)	31
(d1,t2)	49				
(t2,0)	49				
(t4,0)	48				
(t3,0)	48				
(t1,0)	48				
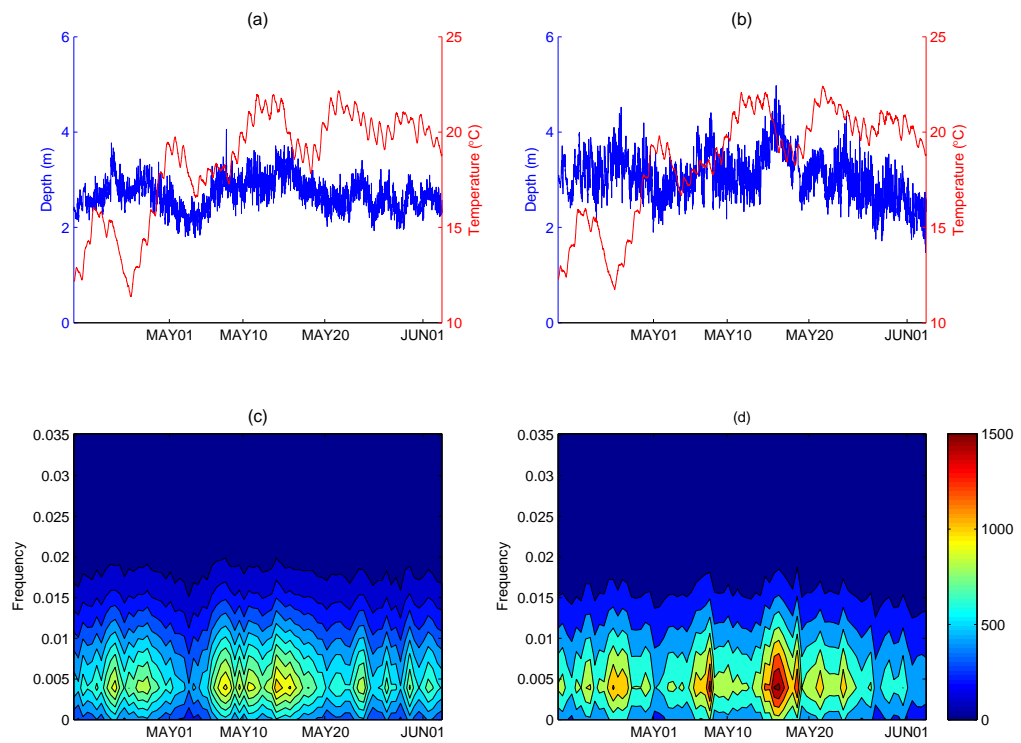
Figure 2.1: (a) and (b) are time series plots of the mean values of depth and temperature for successful and unsuccessful spawners, respectively. (c) and (d) are the mean spectrograms of depth for successful and unsuccessful spawner, respectively.
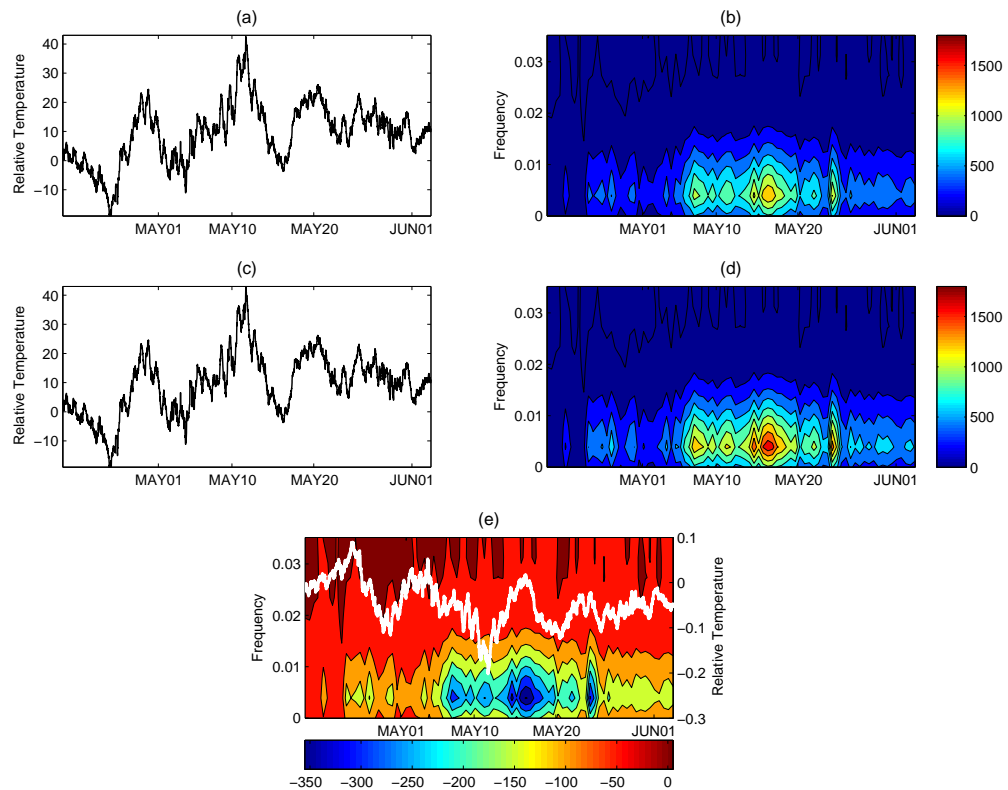
Figure 2.2: (a) and (c) are the mean reconstructed DST temperature using temperature EOF 2 for successful and unsuccessful spawners, respectively. (b) and (d) are the mean reconstructed depth spectrogram using depth EOF 1 for successful and unsuccessful spawners, respectively. (e) depicts the difference of reconstructed temperature curve ((a)-(c)) (white curve) and depth spectrogram ((b)-(d)).

Chapter 3

Bayesian Lattice Filters for Time-Varying Autoregression and Time-Frequency Analysis

3.1 Introduction

Recent advances in technology have lead to the extensive collection of complex high-frequency nonstationary signals across a wide-array of scientific disciplines. In contrast to the time-domain, the time-varying spectrum may provide better insight into important characteristics of the underlying signal (e.g., Holan et al., 2010, 2012; Rosen et al., 2012; Yang et al., 2013, Chapter 2, and the references therein). For example, Holan et al. (2010) demonstrated that features in the time-frequency domain of non-stationary *Enchenopa* treehopper mating signals may describe crucial phenotypes of sexual selection.

In general, time-frequency analyses can either proceed using a nonparametric or

model-based (i.e., parametric) approach. The most common nonparametric approach is the short-time Fourier transform (i.e., windowed Fourier transform) and produces a time-frequency representation characterizing local signal properties (Gröchenig, 2001; Oppenheim and Schafer, 2009). Common window (smoothing) functions that have been proposed for this purpose include the Hamming, Hanning, Barlett and Kaiser windows. Another path to time-frequency proceeds using smoothing splines to estimate the local spectrum via the Whittle likelihood (Rosen et al., 2009, 2012). Similarly, time-frequency can be achieved by applying smooth localized complex exponential (SLEX) functions to the observed signal (Ombao et al., 2002). The SLEX functions are simultaneously orthogonal and localized in time and frequency because they are obtained by applying a projection operator rather than a window. Alternatively, one can use the theory of frames and over-complete bases to produce a time-frequency representation. For example, continuous wavelet transforms (Vidakovic, 1999; Percival and Walden, 2000; Mallat, 2008) or Gabor frames (Wolfe et al., 2004; Feichtinger and Strohmer, 1998; Fitzgerald et al., 2000) could be used. By introducing redundancy into the basis functions, these representations may provide better simultaneous resolution over both time and frequency.

Model-based approaches typically proceed through the time-domain in order to produce a time-frequency representation for a given nonstationary signal. In this setting common approaches include fitting piecewise autoregressive (AR) models as well as time-varying autoregressive models (TVAR). The former approach assumes that the nonstationary signal is piecewise stationary. Consequently, the estimation procedure for such models attempts to identify the order of the AR models along with the location of each piecewise stationary series. For example, Davis et al. (2006)

propose the AutoPARM method using minimum description length (MDL) and a genetic algorithm (GA) to automatically locate the break points and AR model order within each segmentation. In addition to providing a time-frequency representation, this approach also locates change points. Wood et al. (2011) propose fitting mixtures of AR models in each segment via Markov chain Monte Carlo (MCMC) methods. Their approach selects a common segment length and then divides the signal into these segments prior to implementation of the fitting procedure. Although such approaches may accommodate signals with several piecewise stationary structures, they lack the capability of capturing momentary shocks to the system (i.e., changes to the evolutionary structure that only occur over relatively few time points).

For many processes, TVAR models may provide superior resolution in the time-frequency domain for both large and small scale features through modeling time-varying parameters. A TVAR model of order P for a nonstationary univariate time series x_t, $t = 1, \ldots, T$, can be expressed as

$$x_t = \sum_{m=1}^{P} a_{t,m}^{(P)} x_{t-m} + \epsilon_t, \tag{3.1}$$

where $a_{t,m}^{(P)}$ and ϵ_t are the TVAR coefficients associated with time lag m at time t and the innovation at time t, respectively. Typically, the innovations are assumed to be uncorrelated zero-mean Gaussian random variables (i.e., $\epsilon_t \sim N(0, \sigma_t^2)$, with time-varying variance σ_t^2). The model corresponds to a stationary AR form at each time t with the AR coefficients and variances evolving through time. In such settings, the model is locally stationary but nonstationary globally. As will be illustrated, the assumption of local stationarity (i.e., the forward and backward partial autocorrela-

tions need not be equal) is not required for our approach. More importantly, because this model generally allows both slow and rapid changes in the parameters, it can flexibly model stochastic pattern changes often exhibited by complex nonstationary signals.

To estimate the TVAR model coefficients, Kitagawa and Gersch (1996) treat the coefficients as a stochastic process and model them using difference equations under the assumption of a maximum fixed order of the TVAR model. Their estimation procedure for the coefficients is then based on state-space models with smoothness priors. In this context, the innovation variances are treated as constant and estimated using a maximum likelihood approach. West et al. (1999) expand this type of approach into a fully Bayesian framework by simultaneous modeling the coefficients and the innovation variances using random walk models. In contrast, by assuming constant innovation variance, Prado and Huerta (2002) model the coefficients and order of the TVAR model using random walk models. Further, to make the TVAR models stable, the constraint that the roots of the characteristic polynomial lie within the unit circle could be imposed. However, such an added condition makes estimation more complicated and computationally expensive.

To avoid these issues, we can instead work with the partial autocorrelation coefficients (i.e., in the partial correlation (PARCOR) coefficient domain) and then use the Levinson recursion to connect the PARCOR coefficients and TVAR model coefficients (Kitagawa and Gersch, 1996; Godsill et al., 2004). Godsill et al. (2004) use a sequential Monte Carlo algorithms to estimate the TVAR coefficients by modeling the PARCOR as a random walk using a truncated normal distribution. Alternatively, assuming a constant innovation variance, Kitagawa and Gersch (1996) implement the

smoothness prior within a lattice filter to estimate the PARCOR coefficients. From a statistical point of view, these two approaches introduce an indirect way of fitting the TVAR model. However, the former approach is computationally time consuming and may suffer from the degeneracy problem (i.e., the collapse of approximations of the marginal distributions). Although the latter takes advantage of the lattice form to efficiently estimate the PARCOR coefficients, without considering the true evolution of the volatility, it may be impractical for capturing the complex dependence structures inherent in real-world data.

The approach we propose for estimating the TVAR models considers the PARCOR coefficient domain and use the lattice structure to estimate the PARCOR coefficients and innovation variances. The difference between our work and the work of Kitagawa and Gersch (1996) is that our approach models both the evolution of the PARCOR coefficients and innovation variances simultaneously within a fully Bayesian context. Our approach takes advantages of dynamic linear model theory (West and Harrison, 1997; West et al., 1999) and provides full conditional distributions of standard form for both the PARCOR coefficients and innovation variances and thus Bayesian estimation is computationally advantageous. Moreover, our estimation procedure takes advantage of recursive one-dimensional equations so that, when the order $P \geq 2$, our approach avoids calculation of higher dimensional inverse matrices. Importantly, the simulation study we provide demonstrates that, in many cases, our approach has superior performance in terms of estimating the time-frequency representation of various nonstationary signals. Thus, our approach provides a stable and computationally efficient way to fit TVAR models for time-frequency analysis.

The remainder of this chapter is organized as follows. Section 3.2 briefly intro-

51

duces the lattice structure and describes our methodology along with prior specification. Section 3.3 presents a comprehensive simulation study that illustrates the effectiveness of our approach across an expansive array of nonstationary processes. Subsequently, in Section 3.4, our methodology is demonstrated through applications to modeling insect communication signals and environmental data; i.e, wind components. Lastly, Section 3.5 concludes with discussion. For convenience of exposition, details surrounding the estimation algorithms are left to the Appendix.

3.2 Methodology

3.2.1 Lattice Structures

Instead of fitting AR models directly, by estimating the AR coefficients, the lattice structures provide an indirect way of fitting AR models through the PARCOR coefficients (Kitagawa and Gersch, 1996). Using the Levinson-Durbin algorithm, there is a unique correspondence between the PARCOR coefficients and the AR coefficients (Shumway and Stoffer, 2006; Kitagawa, 2009). The lattice structure provides a means of directly estimating the PARCOR coefficients, associated with a particular AR model, using the observed time series (Hayes, 1996; Haykin, 2002). Importantly, stage P corresponds to the AR model of order P.

Let $f_t^{(P)}$ and $b_t^{(P)}$ denote the prediction error at time t for the forward and backward

AR(P) models, respectively, where

$$f_t^{(P)} = x_t - \sum_{m=1}^{P} a_m^{(P)} x_{t-m},$$

$$b_t^{(P)} = x_t - \sum_{m=1}^{P} d_m^{(P)} x_{t+m}.$$

Then, the m-th stage of the lattice filter can be described by the pair of input-output relations between the forward and backward predictions,

$$f_t^{(m-1)} = \alpha_m^{(m)} b_{t-m}^{(m-1)} + f_t^{(m)}, \tag{3.2}$$

$$b_t^{(m-1)} = \beta_m^{(m)} f_{t+m}^{(m-1)} + b_t^{(m)}, \quad m = 1, 2, \ldots, P \tag{3.3}$$

with the initial condition, $f_t^{(0)} = b_t^{(0)} = x_t$, and $\alpha_m^{(m)}$ and $\beta_m^{(m)}$ are the lag m forward and backward PARCOR coefficients, respectively. Equation (3.2) shows that the forward PARCOR coefficient of lag m is a regression coefficient of the forward prediction error $f_t^{(m-1)}$ regressed on the backward prediction error $b_{t-m}^{(m-1)}$ and the residual term f_t^m is the forward prediction error of the forward (subset) AR(m) model. On the other hand, (3.3) shows that backward PARCOR coefficient of lag m is a regression coefficient of the backward prediction error $b_t^{(m-1)}$ regressed on the forward prediction error $f_{t+m}^{(m-1)}$ and the residual term b_t^m is the backward prediction error of the backward (subset) AR(m) model. Using (3.2) and (3.3) recursively, we can derive the PARCOR coefficients for a given lag. In the stationary case, the forward and backward PARCOR coefficients are equivalent, $\alpha_m^{(m)} = \beta_m^{(m)}$.

Example: To illustrate $f_t^{(m)}$ and $b_t^{(m)}$ on the right hand side of (3.2) and (3.3) are the prediction errors of the forward and backward AR models, respectively, we consider

an example when $P = 2$. In this case, we first derive that the difference between $f_t^{(1)}$ and $\alpha_2^{(2)} b_{t-1}^{(1)}$ of (3.2) is equal to the forward prediction errors of AR(2) as follows

$$
\begin{aligned}
f_t^{(1)} - \alpha_2^{(2)} b_{t-1}^{(1)} &= f_t^{(0)} - \alpha_1^{(1)} b_{t-1}^{(0)} - \alpha_2^{(2)} (b_{t-2}^{(0)} - \beta_1^{(1)} f_{t-1}^{(0)}) \\
&= x_t - \alpha_1^{(1)} x_{t-1} - \alpha_2^{(2)} (x_{t-2} - \beta_1^{(1)} x_{t-1}) \\
&= x_t - (\alpha_1^{(1)} - \alpha_2^{(2)} \beta_1^{(1)}) x_{t-1} - \alpha_2^{(2)} x_{t-1} \\
&= f_t^{(2)}.
\end{aligned}
$$

The above derivation shows that the difference between $f_t^{(1)}$ and $\alpha_2^{(2)} b_{t-1}^{(2)}$ is equal to the prediction errors of the forward AR(2). It also can be shown that (3.3) is true. Moreover, when the signal is stationary, the forward and backward PARCOR coefficients are equal (i.e., $\alpha_m^{(m)} = \beta_m^{(m)}$). In such cases, we can change the second term of the third row to $(\alpha_1^{(1)} - \alpha_2^{(2)} \alpha_1^{(1)}) x_{t-1}$.

The PARCOR coefficients $\alpha_P^{(P)}$ are equal to the last component of the coefficients of the forward AR(P) model; i.e., $\alpha_P^{(P)} = a_P^{(P)}$. Using the Levinson-Durbin algorithm (Shumway and Stoffer, 2006; Kitagawa, 2009), the rest of the AR coefficients and innovation variance can be obtained as follows

$$
\hat{a}_m^{(P)} = \hat{a}_m^{(P-1)} - \hat{a}_P^{(P)} \hat{a}_{P-m}^{(P-1)}, \quad m = 1, \ldots, P - 1. \tag{3.4}
$$

This equation implies that once the PARCOR coefficient $a_P^{(P)}$ is estimated, the other coefficients are automatically determined as well. More importantly, the estimates of the AR coefficients are the minimum mean squared error estimate.

3.2.2 The Lattice Structure of the TVAR model

Given the assumption of second-order stationarity, the forward and backward PAR-COR coefficients are constant over time (i.e., shift-invariant). However, since most real-world signals are nonstationary, the shift-invariant PARCOR coefficients are typically inappropriate. In such cases we can modify (3.2) and (3.3) as follows

$$f_t^{(m-1)} = \alpha_{t,m}^{(m)} b_{t-m}^{(m-1)} + f_t^{(m)}, \tag{3.5}$$

$$b_t^{(m-1)} = \beta_{t,m}^{(m)} f_{t+m}^{(m-1)} + b_t^{(m)}, \quad m = 1, 2, \ldots, P \tag{3.6}$$

with both the forward and backward PARCOR coefficients $\alpha_{t,m}^{(m)}$ and $\beta_{t,m}^{(m)}$ now time dependent. Note that for notational simplicity, $f_t^{(m)}$ and $b_t^{(m)}$ here denote the prediction error at time t of the forward and backward TVAR models of order m, respectively. For locally stationary signals, we may impose the constraint that $\alpha_{t,m}^{(m)} = \beta_{t,m}^{(m)}$ at each time t. However, for general nonstationary cases, $\alpha_{t,m}^{(m)}$ and $\beta_{t,m}^{(m)}$ may not be identical at each time t. Therefore, our approach will proceed without this constraint. Also, the innovation terms, $f_t^{(m)}$ and $b_t^{(m)}$, are assumed to follow zero-mean Gaussian distributions, $N(0, \sigma_{f,m,t}^2)$ and $N(0, \sigma_{b,m,t}^2)$, respectively. Finally, for the TVAR model of order P, $\alpha_{t,m}^{(m)} = 0$ when $m > P$, and $\sigma_{f,P,t}^2$ is identical to σ_t^2 of (3.1).

For each stage m of the lattice structure, we construct the following equations to obtain the coefficients, $a_{t,k}^{(m)}$ and $d_{t,k}^{(m)}$, of the forward and backward TVAR models, respectively, (Hayes, 1996; Kitagawa and Gersch, 1996; Haykin, 2002)

$$a_{t,k}^{(m)} = a_{t,k}^{(m-1)} - a_{t,m}^{(m)} d_{t,m-k}^{(m-1)} \tag{3.7}$$

$$d_{t,k}^{(m)} = d_{t,k}^{(m-1)} - d_{t,k}^{(m)} a_{t,m-k}^{(m-1)}, \quad k = 1, 2, \ldots, m-1 \tag{3.8}$$

with $a_{t,m}^{(m)} = \alpha_{t,m}^{(m)}$ and $d_{t,m}^{(m)} = \beta_{t,m}^{(m)}$. Equations (3.7) and (3.8) describe the relationship between the coefficients of the forward and backward TVAR models. In particular, these relations illustrate that the forward coefficients at the current stage are a linear combination of the forward and backward coefficients of the previous stage, with the weights equal to the PARCOR coefficients. Importantly, such a combination also includes the stationary and locally stationary cases. For the stationary case, since $\alpha_{t,m}^{(m)} = \beta_{t,m}^{(m)}$ are constant over time, the general equations (3.7) and (3.8) can be reduced to (3.4). For locally stationary cases, since $\alpha_{t,m}^{(m)} = \beta_{t,m}^{(m)}$ at time t, (3.7) and (3.8) are identical.

3.2.3 Model Specification and Bayesian Inference

Since both the forward and backward PARCOR coefficients of (3.5) and (3.6) as well as the corresponding innovation variances require evolutionary structures, we consider random walk models for their evolutions. In such cases, the following two hierarchical components are added to (3.5) and (3.6). The evolution of the forward and backward PARCOR coefficients are modeled, respectively, as follows:

$$\alpha_{t,m}^{(m)} = \alpha_{t-1,m}^{(m)} + \epsilon_{\alpha,m,t}, \quad \epsilon_{\alpha,m,t} \sim N(0, w_{\alpha,m,t}) \tag{3.9}$$

$$\beta_{t,m}^{(m)} = \beta_{t-1,m}^{(m)} + \epsilon_{\beta,m,t}, \quad \epsilon_{\beta,m,t} \sim N(0, w_{\beta,m,t}) \tag{3.10}$$

where $w_{\alpha,m,t}$ and $w_{\beta,m,t}$ are time dependent system variances. These are defined in terms of hyperparameters $\gamma_{f,m}$ and $\gamma_{b,m}$, so called *discount factors* with range $(0, 1)$, respectively (West and Harrison, 1997). See the Appendix for further details. Usually, we treat $\gamma_{f,m} = \gamma_{b,m} = \gamma_m$ at each stage m. These two equations also imply

56

a sequential update form that the PARCOR coefficient at time $t+1$ is equal to the sum of the PARCOR coefficient at time t plus a correction. Such form is used in the gradient adaptive lattice filter (Hayes, 1996).

Similarly, both the evolution innovation variances, $\sigma^2_{f,m,t}$ and $\sigma^2_{b,m,t}$, are modeled through multiplicative random walks as follows

$$\sigma^2_{f,m,t} = \sigma^2_{f,m,t-1}(\delta_{f,m}/\eta_{f,m,t}), \quad \eta_{f,m,t} \sim Beta(g_{f,m,t}, h_{f,m,t}) \tag{3.11}$$

$$\sigma^2_{b,m,t} = \sigma^2_{b,m,t-1}(\delta_{b,m}/\eta_{b,m,t}), \quad \eta_{b,m,t} \sim Beta(g_{b,m,t}, h_{b,m,t}) \tag{3.12}$$

where $\delta_{f,m}$ and $\delta_{b,m}$ are hyperparameters (i.e., discount factors on the range $(0,1)$), and the multiplicative innovations, $\eta_{f,t,m}$ and $\eta_{b,t,m}$ follow beta distributions with parameters, $(g_{f,m,t}, h_{f,m,t})$ and $(g_{b,m,t}, h_{b,m,t})$, respectively (West et al., 1999). These parameters are defined at each time t by the discount factors, $\delta_{f,m}$ and $\delta_{b,m}$, as detailed in the Appendix. In many cases, we also assume that $\delta_{f,m} = \delta_{b,m} = \delta_m$ at each stage m. The series of stochastic error terms $\epsilon_{\alpha,m,t}$, $\epsilon_{\beta,m,t}$, $\eta_{f,m,t}$, and $\eta_{b,m,t}$ are mutually independent, and also independent of the forward and backward innovation term series, $f_t^{(m)}$ and $b_t^{(m)}$ of (3.5) and (3.6).

We specify conjugate initial priors for $\alpha_{0,m}^{(m)}$ and $\sigma^2_{f,m,0}$ at each stage m as follows

$$p(\alpha_{0,m}^{(m)}|D_{f,m,0}, \sigma^2_{f,m,0}) \sim N(\mu_{f,m,0}, c_{f,m,0}), \tag{3.13}$$

$$p(\sigma^{-2}_{f,m,0}|D_{f,m,0}) \sim G(v_{f,m,0}/2, \kappa_{f,m,0}/2) \tag{3.14}$$

where $D_{f,m,0}$ denotes the information set at time $t = 0$ (the initial time), $G(\cdot, \cdot)$ is the gamma distribution, $\mu_{f,m,0}$ and $c_{f,m,0}$ are user specified values of the mean and variance for a normal distribution, and $v_{f,m,0}/2$ and $\kappa_{f,m,0}/2$ are the shape and scale

parameters for a gamma distribution. Usually, we treat the starting values $\mu_{f,m,0}$, $c_{f,m,0}$, $v_{f,m,0}$, and $\kappa_{f,m,0}$ as a common constant over stage m, respectively. Typically, we choose $\mu_{f,m,0}$ and $c_{f,m,0}$ close to zero and one, respectively. In addition, to set $v_{f,m,0}$ and $v_{f,m,0}$, we first fix $v_{f,m,0} = 1$ and calculate the sample variance of the initial components of the signal. Given these two values, we can obtain $\kappa_{f,m,0}$ through the formula of the expectation of the gamma distribution. In such prior settings, sequential filtering and smoothing algorithms for dynamic linear models deliver the components of marginal posterior distributions (West and Harrison, 1997). Specifically, for $t = 1, \ldots, T$, with the information set $D_{f,m,T}$ up to time T, the marginal posterior distributions $p(\alpha_{t,m}^{(m)}|D_{f,m,T})$ and $p(\sigma_{f,m,t}^{-2}|D_{f,m,T})$ are the t-distribution and gamma distribution, respectively. Analogous to $\alpha_{0,m}^{(m)}$ and $\sigma_{f,m,0}^2$, the same conjugate initial priors for $\beta_{0,m}^{(m)}$ and $\sigma_{b,m,0}^2$ are specified at each stage m. Details of the sequential filtering and smoothing for the PARCOR coefficients and innovation variances for each stage m are discussed in the Appendix.

3.2.4   Model Selection

Selection of the model order and set of discount factors $\{P, \gamma_m, \delta_m; m = 1, \ldots, P\}$ is essential for our approach. First, one can assume $\gamma_m = \gamma$ and $\delta_m = \delta$, for $m = 1, \ldots, P$. Then, the analysis will run through a specified set of combinations of (P, γ, δ). Since γ is related to the variability of the PARCOR coefficients, it also affects the variability of the TVAR coefficients. Hence, one can model the variance of the time-varying coefficients and the innovation variances of the TVAR models using discount factors γ and δ, respectively (West et al., 1999). Note that, in our context, the discount factors (γ_m, δ_m) are a function of m (the lattice filter stage), whereas in

the West et al. (1999) setting there are only one set of discount factors (γ, δ) that need to be estimated. However, estimation of (P, γ, δ) using the approach of West et al. (1999) entails repeatedly having to calculate inverse matrices in the sequential filtering process. Alternatively, we consider the case that γ_m and δ_m may vary from stage to stage. In such cases, we first specify a potential maximum value of P and a set of combinations of $\{\gamma_m, \delta_m\}$ for each stage m. Given a value of P, we search for the combination of $\{\gamma_1, \delta_1\}$ maximizing the log likelihood of (3.5) at stage one. Using the selected γ_1 and δ_1, we can obtain the corresponding series $\{f_t^{(2)}\}$ and $\{b_t^{(2)}\}$, for $t = 1, \ldots, T$, as well as the value of log maximum likelihood of (3.5), ℓ_1. We then, repeat the above search procedure for stage two using the output $\{f_t^{(2)}\}$ and $\{b_t^{(2)}\}$ obtained from implementing the selected hyperparameters γ_1 and δ_1. In turn, this produces a new series of $\{f_t^{(3)}\}$ and $\{b_t^{(3)}\}$, for $t = 1, \ldots, T$, as well as a value ℓ_2. We repeat the procedure until the set of $\{\gamma_m, \delta_m, \ell_m\}$, $m = 1, \ldots, P$, has been selected.

Here, we provide both a visual method and a numerical method to select the order. Similar to the scree plot widely used in multivariate analysis (Rencher, 2002), we can plot ℓ_m against order m. When the observed series follows an AR or TVAR model, the values of ℓ_m will stop increasing after a specific lag, this lag can be chosen as the order for the estimated model. Henceforth, this plot is referred to as "BLF-scree." This type of visual order determination can be directly quantified through the relative change in the log-likelihood. Specifically, we provide a numerical method of order selection based on calculating the percent change in going from ℓ_{m-1} to ℓ_m,

$$\{(\ell_m - \ell_{m-1})/\ell_{m-1}\} * 100. \tag{3.15}$$

Based on simulation of various TVAR models we find that the value of m for which this

percentage becomes less than 0.5% reflects the "best" value of the order. That is, we have found that 0.5% provides an effective cut-off for choosing the order. Although this value provides a good guide to order selection, formal model selection in this setting remains an area of future research.

We now summarize the procedure of our approach to fitting the TVAR models. Given a set of hyperparameters $\{P, \gamma_m, \delta_m; m = 1, \ldots, P\}$, the procedure starts by setting $f_t^{(0)} = b_t^{(0)} = x_t$, for $t = 1, \ldots, T$. Next, plugging $\{f_t^{(0)}\}$ and $\{b_t^{(0)}\}$ into (3.5) and (3.6) and using sequential filtering and smoothing algorithms, we obtain a series of estimated parameters $\{\widehat{\alpha}_{t,1}^{(1)}\}$, $\{\widehat{\beta}_{t,1}^{(1)}\}$, $\{\widehat{\sigma}_{f,1,t}^2\}$, and $\{\widehat{\sigma}_{b,1,t}^2\}$, as well as the new series of forward and backward prediction errors, $\{f_t^{(1)}\}$ and $\{b_t^{(1)}\}$, for $t = 1 \ldots, T$, respectively. We then repeat the above procedure until $\{\widehat{\alpha}_{t,P}^{(P)}\}$, $\{\widehat{\beta}_{t,P}^{(P)}\}$, $\{\widehat{\sigma}_{f,P,t}^2\}$, and $\{\widehat{\sigma}_{b,P,t}^2\}$ have been obtained. Then, recursively plugging the estimates of $\{\alpha_{t,m}^{(m)}\}$ and $\{\beta_{t,m}^{(m)}\}$, from $m = 1, \ldots, P$ into (3.7) and (3.8), we obtain the estimated time-varying coefficients of (3.1). As part of this algorithm, the series of estimated innovation variances are equal to $\{\widehat{\sigma}_{f,P,t}^2\}$. Finally, for $t = 1, \ldots, T$, the time-frequency representation associated with the time-varying $AR(P)$ model can be obtained by the following equation

$$S(t, \omega) = \frac{\sigma_t^2}{\left|1 - \sum_{m=1}^P a_{t,m}^{(P)} \exp(-2\pi i m \omega)\right|^2}, \qquad -1/2 \leq \omega \leq 1/2 \qquad (3.16)$$

where $i = \sqrt{-1}$ (Kitagawa and Gersch, 1996). Plugging the estimated values $\widehat{a}_{t,m}^{(P)}$, $m = 1, \ldots, P$ and $\widehat{\sigma}_{f,P,t}^2$ into (3.16) yields the estimated time-varying $AR(P)$ time-frequency representation $\widehat{S}(t, \omega)$.

3.3 Simulation Studies

In this section, we simulate various nonstationary time series in order to compare the performance of our approach with three other approaches used to estimate the time-frequency representation. The first approach is AdaptSPEC proposed by Rosen et al. (2012). This approach adaptively segments the signal into finite pieces and then estimates the time-frequency representation using smoothing splines via a local Whittle likelihood approximation to the true likelihood. The size of a segment and the number of the spline basis function are two essential parameters for this approach. To reduce any subjectivity in our comparisons, we use the settings for these two parameters considered in Rosen et al. (2012) (with their $t_{min} = 40$), as well as the same settings for MCMC iterations and burn-in. The second approach is the AutoPARM method (Davis et al., 2006). Although this approach combines the GA and MDL to automatically search for potential break points along with the AR orders for each segment, four parameters are crucial for the GA: numbers of islands, chromosomes in each island, generations for migration, and chromosomes replaced in a migration; see Davis et al. (2006) for a comprehensive discussion. All of these parameters were chosen identical to those used in Davis et al. (2006). The last method we consider is the approach of West et al. (1999), referred to as WPK1999. This approach requires specification of three parameters: the TVAR order and two discount factors - one associated with the variance of the time-varying coefficients and the other with the innovation variances. In general, the discount factor values are in the range $0.9-0.999$ (West et al., 1999). Therefore, for our simulations, we give each discount factor a set of values from 0.8 to 1 (with equal spacing of 0.02) and, further, a set of values for the TVAR order from 1 to 15. Given these values, we choose the combination that

achieves the maximum likelihood (West et al., 1999).

Our approach uses the two selection methods discussed in Section 3.2.4 to search for the TVAR order with appropriate discount factor values. The selected combination of $\{P, \gamma, \delta\}$ with γ and δ held fixed over all stages of the Bayesian lattice filter is referred to as BLFFix. The selected combination of (P, γ_m, δ_m), for $m = 1, \ldots, P$, is referred to as BLFDyn. The candidate space of parameters for both discount factors are from 0.8 to 1 (with equal spacing of 0.02), along with orders from 1 to 15.

We consider four types of nonstationary signals: 1) TVAR of order 2 with constant innovation variance; 2) TVAR of order 6 with constant innovation variance; 3) a piecewise AR process with constant innovation variance; and 4) simulated signals based on a *Enchenopa* treehopper communication signal (Holan et al., 2010); see Section 3.4.1. Each simulation consists of 200 realizations. To evaluate the performance in estimating the various time-frequency representations, we calculate the average squared error (ASE) for each realization as follows (Ombao et al., 2001)

$$\text{ASE}_n = (TL)^{-1} \sum_{t=1}^{T} \sum_{l=1}^{L} \left(\log \widehat{S}(t, \omega_l) - \log S(t, \omega_l) \right)^2,$$

where $n = 1, \ldots, 200$, $\omega_l = 0, 0.005, \ldots, 0.5$, and T denotes the length of the simulated series. Lastly, we denote $\overline{\text{ASE}} = (1/200) \sum_{n=1}^{200} \text{ASE}_n$.

3.3.1 Time-Varying AR(2) Process

We simulate signals from the same time-varying AR(2) process (TVAR2), used in Davis et al. (2006) and Rosen et al. (2009, 2012), which is defined as follows

$$x_t = a_t x_{t-1} - 0.81 x_{t-2} + \epsilon_t,$$

$$a_t = 0.8(1 - 0.5\cos(\pi t/1024)),$$

where $\epsilon_t \overset{\text{iid}}{\sim} N(0,1)$ and $t = 1, \ldots, 1024$. Figure 3.1 shows the BLF-scree plot of the likelihood against order of the BLFDyn. This plot suggests that order two is the appropriate choice for all 200 realizations. Since the time-varying coefficient a_t, varies slowly with time, this process naturally exhibits a slowly evolving time-varying spectrum (Figure 3.2). The box-plots of the ASE values in Figure 3.2 show that the group of TVAR-based models (i.e., WPK1999, BLFFix, and BLFDyn) perform superior to the group of non-TVAR-based models (i.e., AdaptSPEC and AutoPARM), with BLFDyn performing the best. In fact, the reduction of $\overline{\text{ASE}}$ for BLFDyn is 36.57% over WPK1999, 36.80% over BLFFix, 84.33% over AutoPARM, and 87.71% over AdaptSPEC.

3.3.2 Time-Varying AR(6) Process

We consider signals from the same time-varying AR(6) process of order six (TVAR6) used in Rosen et al. (2009). This time-varying AR(6) process can be compactly expressed as $\phi_t(B)x_t = \epsilon_t$, $t = 1, \ldots, T$, in terms of a characteristic polynomial function $\phi_t(B)$, with $\epsilon_t \overset{\text{iid}}{\sim} N(0,1)$ and B the backshift operator (i.e., $B^p x_t = x_{t-p}$).

The characteristic polynomial function for this process can be factorized as

$$\phi_t(B) = (1 - a_{t,1}B)(1 - a^*_{t,1}B)(1 - a_{t,2}B)(1 - a^*_{t,2}B)(1 - a_{t,3}B)(1 - a^*_{t,3}B),$$

where the superscript $*$ denotes the complex conjugate. Also, for $p = 1, 2, 3$, let $a^{-1}_{t,p} = A_p\exp(2\pi i\theta_{t,p})$, where the $\theta_{t,p}$s are defined by

$$\theta_{t,1} = 0.05 + (0.1/(T - 1))t,$$

$$\theta_{t,2} = 0.25,$$

$$\theta_{t,3} = 0.45 - (0.1/(T - 1))t,$$

with $T = 1024$. The values of A_1, A_2, and A_3 are equal to 1.1, 1.12, and 1.1, respectively.

Figure 3.3 suggests that order six is the appropriate choice for all 200 realizations. The TVAR(6) contains three pairs of time-varying conjugate complex roots. Figure 3.4 illustrates that the TVAR(6) has a time-varying spectrum with three peaks. Similar to the analysis of TVAR(2) (Section 3.3.1), the group of TVAR-based models performs better than the group of non-TVAR-based models, with BLFDyn performing superior to the others. Again, the percent reduction in $\overline{\text{ASE}}$ is notable. The reduction of $\overline{\text{ASE}}$ for BLFDyn is 29.57% over WPK1999, 35.43% over BLFFix, 75.26% over AdaptSPEC, and 75.68% over AutoPARM.

3.3.3 Piecewise Stationary AR Process

The signals simulated here are based on the same piecewise stationary AR process, used by Davis et al. (2006) and Rosen et al. (2009, 2012) and is defined as follows

$$
x_t = \begin{cases}
0.9x_{t-1} + \epsilon_t; & \text{if} \quad 1 \leq t \leq 512, \\
1.69x_{t-1} - 0.8x_{t-2} + \epsilon_t; & \text{if} \quad 513 \leq t \leq 768, \\
1.32x_{t-1} - 0.8x_{t-2} + \epsilon_t; & \text{if} \quad 769 \leq t \leq 1024,
\end{cases}
$$

where $\epsilon_t \overset{\text{iid}}{\sim} N(0,1)$. These generated signals are referred to as PieceAR. Since it is difficult to choose the order for some realizations using Figure 3.5, we use (3.15) to choose the order. The numerical method suggests order two for some realizations and order three for the others. The true process includes three segments, with each of the segment mutually independent. The piecewise nature of this process is clearly depicted by its time-varying spectrum (Figure 3.6b). The box-plot shows that AutoPARM exhibits superior performance in terms of the smallest median ASE. However, WPK1999 and BLFFix may perform more robustly (i.e., less outlying ASE values). Although we see a 23.78% reduction in $\overline{\text{ASE}}$ for AutoPARM versus BLFFix, we find that BLFFix performs superior to the remainder of the approaches and has the smallest standard deviation across the 200 simulations. The findings here are not surprising as AutoPARM is ideally suited toward identifying and estimating piecewise AR processes and, for our approach, taking (γ, δ) fixed is advantageous for processes that are not slowly-varying.

3.3.4 Simulated Insect Communication Signals

The signals considered in this simulation are formulated such that they exhibit the same properties as an *Enchenopa* treehopper mating signal; see Section 3.4.1 for a complete discussion. Specifically, we fit a time-varying AR(6) model to obtain time-varying AR coefficients and innovation variances. Typically, with these type of nonstationary signals, the innovation variances are time dependent, which is markedly different from the previous examples where the innovation variance was constant. The signals generated by these parameters are referred as SimBugs. As expected, Figure 3.7 suggests that order six may be an appropriate choice for all 200 realizations. Panel (a) of Figure 3.8 demonstrates one realization of the SimBugs, whereas panel (b) provides box-plots that characterize the distribution of ASE_n over the 200 simulated signals. Specifically, from Figure 3.8, we see that BLFFix and BLFDyn perform better than the other approaches, in terms of median ASE. Further, we find that BLFFix and BLFDyn are similar, in terms of $\overline{\text{ASE}}$, although the median of BLFDyn (0.2675) is smaller than that of BLFFix (0.3176).

Table 3.1 summarizes the mean values and standard deviations for this simulation. From this table we see that the reduction of $\overline{\text{ASE}}$ for BLFDyn is 6.01% over BLF-Fix, 30.04% over WPK1999, 51.77% over AutoPARM, and 78.07% over AdaptSPEC. Lastly, BLFFix has the smallest sd_{ASE} of all of the approaches in this setting.

3.4 Case Studies

3.4.1 Animal Communication Signals

Understanding the dynamics of populations is an important component of evolutionary biology. Many organisms exhibit complex characteristics that intricately relate to fitness. For example, the mating signal of the *Enchenopa* treehopper represents a phenotype of the insect that is used in mate selection (see Holan et al., 2010, and the references therein). During the mating season, males in competition deliver their vibrational signals through stems of plants to females (see Cocroft and McNett, 2006, and the references therein). The data considered here comes from an experiment that was previously analyzed in Holan et al. (2010). The experiment was designed with the goal of reducing potential confounding effects between environmental and phenotypical variation. In this experiment, males signals were recorded one week prior to the start of mating. Figure 3.9 (a) displays a typical signal of from a successful mater, with length 4,739 downsampled from a registered signals of length 37,912. In this context, justification for the appropriateness of downsampling the original signal can be found in Holan et al. (2010). Also, as discussed in Holan et al. (2010), this signal shows a series of broadband clicks preceding a frequency-modulated sinusoidal component, followed by a series of pulses.

For this analysis, we used the BLFDyn approach to search for a model having both discount factors in the range of 0.8 to 1 (with equal spacing of 0.02) and an order between 1 and 25. Figure 3.10 summarizes the likelihood along with the time-varying PARCOR coefficients, AR coefficients, and variances. The likelihood in panel (a), different from the simulated time-varying AR models, increases along with the

order. Hence, we use (3.15) with the model order chosen by our 0.5% rule of thumb. This rule yields a time-varying AR model with order equal to six. In panel (b), the PARCOR coefficients of lag larger than two are close to zero following time around 0.3 (where the time access has been normalized such that $t \in (0, 1)$). Thus, the last four TVAR coefficients after time 0.3 are close to zero in panel (c). Such phenomena suggests that the period before time 0.3 has a more complex dependence structure. Panel (d) shows that the innovation variance exhibits higher volatility at the beginning signal. These bursts in the innovation variance are related to the series of broadband clicks at the beginning of the signal. Finally, Figure 3.9 (b) presents the time-frequency representation of the treehopper signal using the time-varying AR(6) model and corroborates the significance of the broadband clicks at the beginning of the signal.

3.4.2 Wind Components

We study the time-frequency representations of the east/west and north/south wind components, recorded daily at Chuuk Island in the tropical Pacific during the period of 1964 to 1994 (see Cressie and Wikle, 2011, Sections 3.5.3 and 3.5.4). The data studied are at the level of 70 hPa, which is important scientifically due to the likely presence of westward and eastward propagating tropical waves, and the presence of the quasi-biennial oscillation (QBO) (Wikle et al., 1997). Figure 3.11 shows the two wind component time series from which we can discern visually that the east/west series clearly exhibits the QBO signal, but no discernible smaller-scale oscillations are present in either series. Our interest is then whether the time-varying spectra for these series suggest the presence of time-varying oscillations, which are theorized to

be present.

We consider the same search space for the model order and discount factors as that used for the treehopper communication signal (Section 3.4.1). Figure 3.12 shows the analysis and corresponding inference for the east/west component of the wind. Panel (a) shows an increase of the likelihood so that we use (3.15) with the critical point 0.5% to choose the order equal to three. The PARCOR coefficient of lag one depends on time but the PARCOR coefficient of lag two and three appear to be constant. Panels (c) and (d) present the estimated time-varying coefficients and innovation variances of the time-varying AR(3) model. Figure 3.13 show the analysis and corresponding inference for the east/west component of the wind. Panel (a) shows a turning point at order three so that we choose the order equal to four. The PARCOR coefficients are time independent. Panel (c) and (d) present the estimated time-varying coefficients and innovation variances of the TVAR model of order four. Figure 3.14 shows the estimated time-frequency representations of the wind east/west and north/south components. The east/west component time-varying spectrum does suggest that the QBO intensity varies considerably as evidenced by the power in the low-frequencies. Perhaps more interesting is the suggestion of time-varying equatorial waves in the north/south wind component time-varying spectrum. In particular, the lower-frequency (Kelvin and Rossby) waves with frequencies between 0.1 and 0.2 show considerable variation in duration of wave activity, as well as intensity. One also sees time-variation in the likely mixed-Rossby gravity waves in the frequency band between 0.2 and 0.35. Interestingly, in some cases these are in phase with the lower-frequency wave activity but more often act in opposition. We also note the almost complete collapse of the equatorial wave activity centered on 1984.

3.5   Discussion

This chapter develops a computationally efficient method for model-based time-frequency analysis. Specifically, we consider a fully Bayesian lattice filter approach to estimating time-varying autoregressions. By taking advantage of the partial auto-correlation domain, our approach is extremely stable. Notably, the full conditional distributions arising from our approach are all of standard form and, thus, facilitate easy estimation.

The framework we propose extends the current model-based approaches to time-frequency analysis and, in most cases, provides superior performance, as measured by the average squared error between the true and estimated time-varying spectral density. In fact, for slowly-varying processes we have demonstrated significant estimation improvements from using our approach. In contrast, when the true process comes from a piecewise AR model the approach of Davis et al. (2006) performed best, with our approach a close competitor and performing second best. This is not unexpected as the AutoPARM method is a model-based segmented approach and more closely mimics the behavior of a piecewise AR.

In addition to a comprehensive simulation study we have provided two real-data examples, one from environmental science and the other from animal (insect) communication. In both cases, the exceptional time-frequency resolution obtained using our approach helps us identify salient features in the time-frequency surface. Finally, as a by-product of taking a fully Bayesian approach, we are naturally able to quantify uncertainty and, thus, use our approach to draw inference.

3.6 Appendix: Sequential Updating and Smoothing

To complete the Bayesian estimation of the forward and backward PARCOR coefficients, as well as the time-varying innovation variances, we use dynamic linear model (DLM) methodology (see, West and Harrison, 1997; West et al., 1999). Specifically, we provide the details and formulas for analysis for $\{\alpha_{t,m}^{(m)}\}$ and $\{\sigma_{f,m,t}^2\}$. The analysis of $\{\beta_{t,m}^{(m)}\}$ and $\{\sigma_{b,m,t}^2\}$ follows similarly.

For $t = 1, \ldots, T$, given the values of $f_t^{(m-1)}$ and $b_{t-m}^{(m-1)}$ at stage m, recall (3.5) of Section 3.2.2 gives

$$f_t^{(m-1)} = \alpha_{t,m}^{(m)} b_{t-m}^{(m-1)} + f_t^{(m)},$$

with $f_t^{(m)} \sim N(0, \sigma_{f,m,t}^2)$. Modeling of $\{\alpha_{t,m}^{(m)}\}$ and $\{\sigma_{f,m,t}^2\}$ proceeds by (3.9) and (3.11):

$$\alpha_{t,m}^{(m)} = \alpha_{t-1,m}^{(m)} + \epsilon_{\alpha,m,t}, \quad \epsilon_{\alpha,m,t} \sim N(0, w_{\alpha,m,t})$$

$$\sigma_{f,m,t}^2 = \sigma_{f,m,t-1}^2 (\delta_{f,m}/\eta_{f,m,t}), \quad \eta_{f,m,t} \sim Beta(g_{f,t}^{(m)}, h_{f,t}^{(m)}),$$

with $w_{\alpha,m,t} = c_{f,m,t-1}(1 - \gamma_{f,m})/\gamma_{f,m}$ (see West and Harrison, 1997, Section 6.3) and $c_{f,m,t-1}$ the scale parameter of the marginal t-distribution of $\alpha_{t-1,m}^{(m)}$ given the information set up to time $t - 1$. Moreover, $g_{f,m,t} = \delta_{f,m} \cdot v_{f,m,t-1}/2$, and $h_{f,m,t} = (1 - \delta_{f,m}) \cdot v_{f,m,t-1}/2$ (see West and Harrison, 1997, Section 10.8), where $v_{f,m,t-1}/2$ the shape parameter of the marginal gamma distribution of $\sigma_{f,m,t}^{-2}$ given the information set up to time $t - 1$ (see Section 3.6.1). Then, given specified values for the two discount factors $\gamma_{f,m}$ and $\delta_{f,m}$, as well as conditional on assuming conjugate initial priors (3.13) and (3.14), we can derive the corresponding sequential updating and

smoothing of DLM theory.

3.6.1 Sequential Updating

Using similar notation to West and Harrison (1997) and West et al. (1999), we first sequentially update the joint posterior distributions of $p(\alpha_{t,m}^{(m)}, \sigma_{f,m,t}^{-2} | D_{f,m,t})$ over $t = 1, \ldots, T$. Since the initial priors have the conjugate normal/gamma forms, $p(\alpha_{t,m}^{(m)}, \sigma_{f,m,t}^{-2} | D_{f,m,t})$ also has the normal/gamma form. Therefore, the marginal posterior distribution of $\alpha_{t,m}^{(m)}$ is a t-distribution; i.e., $p(\alpha_{t,m}^{(m)} | D_{f,m,t}) \sim T_{v_{f,m,t}}(\mu_{f,m,t}, c_{f,m,t})$, with degrees of freedom $v_{f,m,t}$, location parameter $\mu_{f,m,t}$, and scale parameter $c_{f,m,t}$. The marginal posterior distribution of $\sigma_{f,m,t}^{-2}$ is a gamma distribution $p(\sigma_{f,m,t}^{-2} | D_{f,m,t}) \sim G(v_{f,m,t}/2, \kappa_{f,m,t}/2)$, with shape parameter $v_{f,m,t}/2$ and scale parameter $\kappa_{f,m,t}/2$. We summarize the sequential updating equations of parameters for $t = 1, \ldots, T$, as follows:

$$\mu_{f,m,t} = \mu_{f,m,t-1} + z_{f,m,t} e_{f,m,t},$$

$$c_{f,m,t} = (r_{f,m,t} - z_{f,m,t}^2 q_{f,m,t})(s_{f,m,t}/s_{f,m,t-1}),$$

and

$$v_{f,m,t} = \delta_{f,m} v_{f,m,t-1} + 1,$$

$$\kappa_{f,m,t} = \delta_{f,m} \kappa_{f,m,t-1} + s_{f,m,t-1} e_{f,m,t}^2 / q_{f,m,t},$$

where

$$e_{f,m,t} = f_t^{(m)} - \mu_{f,m,t-1} b_{t-m}^{(m-1)},$$

$$z_{f,m,t} = r_{f,m,t} b_{t-m}^{(m-1)} / q_{f,m,t},$$

$$s_{f,m,t} = \kappa_{f,m,t} / v_{f,m,t},$$

$$q_{f,m,t} = r_{f,m,t} + s_{f,m,t-1},$$

$$r_{f,m,t} = c_{f,m,t} + w_{\alpha,m,t},$$

$$w_{\alpha,m,t} = c_{f,m,t-1}(1 - \gamma_{f,m})/\gamma_{f,m}.$$

Importantly, since $w_{\alpha,m,t} = c_{f,m,t-1}(1 - \gamma_{f,m})/\gamma_{f,m}$, we can reduce $r_{f,m,t}$ to $r_{f,m,t} = c_{f,m,t-1}/\gamma_{f,t}$.

3.6.2   Smoothing

After the sequential updating process, we can use a retrospective approach to derive the smoothing joint distribution of $p(\alpha_{t,m}^{(m)}, \sigma_{f,m,t}^{-2}|D_{f,m,T})$, for $t = 1, \ldots, T$, given all the information up to time T (West and Harrison, 1997; West et al., 1999). We summarize the equations for the parameters of both marginal distributions, $p(\alpha_{t,m}^{(m)}|D_{f,m,T}) \sim T_{v_{f,m,t|T}}(\mu_{f,m,t|T}, c_{f,m,t|T})$, with degrees of freedom $v_{f,m,t|T}$, location parameter $\mu_{f,m,t|T}$, and scale parameter $c_{f,m,t|T}$. Moreover, $p(\sigma_{f,m,t}^{-2}|D_{f,m,T}) \sim G(v_{f,m,t|T}/2, \kappa_{f,m,t|T}/2)$, with shape parameter $v_{f,m,t|T}/2$ and scale parameter $\kappa_{f,m,t|T}/2$. It is important to note that when $t = T$, we have $\mu_{f,m,T|T} = \mu_{f,m,T}$, $c_{f,m,T|T} = c_{f,m,T}$, $v_{f,m,T|T} = v_{f,m,T}$, and $z_{f,m,T|T} = v_{z,m,T}$ from the results of sequential updating. Additionally, the point estimate of $\sigma_{f,m,t}^2$ at time T is $s_{f,m,T|T} = s_{f,m,T}$. Then, for $t = T - 1, \ldots, 1$, we can

summarize the equations as follows:

$$\mu_{f,m,t|T} = (1 - \gamma_{f,m})\mu_{f,m,t} + (1 - \gamma_{f,m})\mu_{f,m,t+1|T},$$

$$c_{f,m,t|T} = [(1 - \gamma_{f,m})c_{f,m,t} + \gamma_{f,m}^2 c_{f,m,t+1|T}](s_{f,m,t|T}/s_{f,m,t}),$$

$$v_{f,m,t|T} = (1 - \delta_{f,m})v_{f,m,t} + \delta_{f,m}v_{f,m,t+1|T},$$

$$1/s_{f,m,t|T} = (1 - \delta_{f,m})/s_{f,m,t} + \delta_{f,m}/s_{f,m,t|T},$$

$$\kappa_{f,m,t|T} = v_{f,m,t|T}/s_{f,m,t|T}.$$

3.6.3 Algorithm to Fit TVAR Models

We summarize the algorithm of our approach to fitting a time-varying $\mathrm{AR}(P)$ model as follows:

Step 1. Give a value for order P and a set of values for $\{\gamma_m, \delta_m\}$, for $m = 1, \ldots, P$, as well as initial values of parameters at $t = 0$.

Step 2. For $t = 1, \ldots, T$, set $f_t^0 = b_t^0 = x_t$.

Step 3. Put $\{f_t^0\}$ and $\{b_t^0\}$ into (3.5) and then use sequential updating and smoothing to obtain $\{\widehat{\alpha}_{t,1}^{(1)}\}$, $\{\widehat{\sigma}_{f,1,t}\}$, and $\{f_t^1\}$.

Step 4. Put $\{f_t^0\}$ and $\{b_t^0\}$ into (3.6) and then run sequential updating and smoothing to obtain $\{\widehat{\beta}_{t,1}^{(1)}\}$, $\{\widehat{\sigma}_{b,1,t}\}$, and $\{b_t^1\}$.

Step 5. Put $\{f_t^1\}$ and $\{b_t^1\}$ into (3.5) and (3.6) and then run sequential updating and smoothing to obtain $\{\widehat{\alpha}_{t,2}^{(2)}\}$, $\{\widehat{\sigma}_{f,2,t}\}$, and $\{f_t^2\}$ as well as $\{\widehat{\beta}_{t,2}^{(2)}\}$, $\{\widehat{\sigma}_{b,2,t}\}$, and $\{b_t^2\}$.

Step 6. Repeat **Step 5** until $\{\widehat{\alpha}_{t,P}^{(P)}\}$, $\{\widehat{\sigma}_{f,P,t}\}$, $\{\widehat{\beta}_{t,P}^{(P)}\}$, and $\{\widehat{\sigma}_{b,P,t}\}$ are obtained.

Step 7. Given the set of estimated values $\{\widehat{\alpha}_{t,m}^{(m)}\}$, $\{\widehat{\beta}_{t,m}^{(m)}\}$, for $m = 1, \ldots, P$, use

(3.7) and (3.8) iteratively to get the set of estimated $\{\widehat{a}_{t,m}^{(P)}\}$, $m = 1, \ldots, P$, as well as set $\{\widehat{\sigma}_t^2 = \widehat{\sigma}_{f,P,t}\}$.

Table 3.1: The mean, $\overline{\text{ASE}}$ and standard deviation, sd_{ASE}, of the ASE values for the simulations presented in Section 3.3. Note that the bold values represent the approach having minimum $\overline{\text{ASE}}$.

| | $\overline{\text{ASE}}(\text{sd}_{ASE})$ | | | | |
	AdaptSPEC	AutoPARM	WPK1999	BLFFix	BLFDyn
TVAR2	0.1383 (0.0275)	0.1085 (0.0302)	0.0268 (0.0093)	0.0269 (0.0093)	**0.0170 (0.0084)**
TVAR6	0.2195 (0.0351)	0.2233 (0.0439)	0.0771 (0.0171)	0.0841 (0.0191)	**0.0543 (0.0276)**
PieceAR	0.1070 (0.0227)	**0.0702 (0.0392)**	0.0931 (0.0218)	0.0921 (0.0205)	0.1607 (0.1107)
SimBugs	1.4760 (0.2237)	0.6711 (0.0725)	0.4627 (0.0796)	0.3444 (0.0659)	**0.3237 (0.1075)**

Figure 3.1: BLF-scree plot of likelihood along with the TVAR order for the 200 realizations of the time-varying AR(2) process (TVAR2), given in Section 3.3.1.

Figure 3.2: (a) and (b) depict one realization and the true time-frequency representation of the time-varying AR(2) process (TVAR2), respectively (Section 3.3.1). (c) illustrates the box-plots of the average squared error (ASE) values corresponding to the time-frequency representation of TVAR2 for all of the approaches considered.
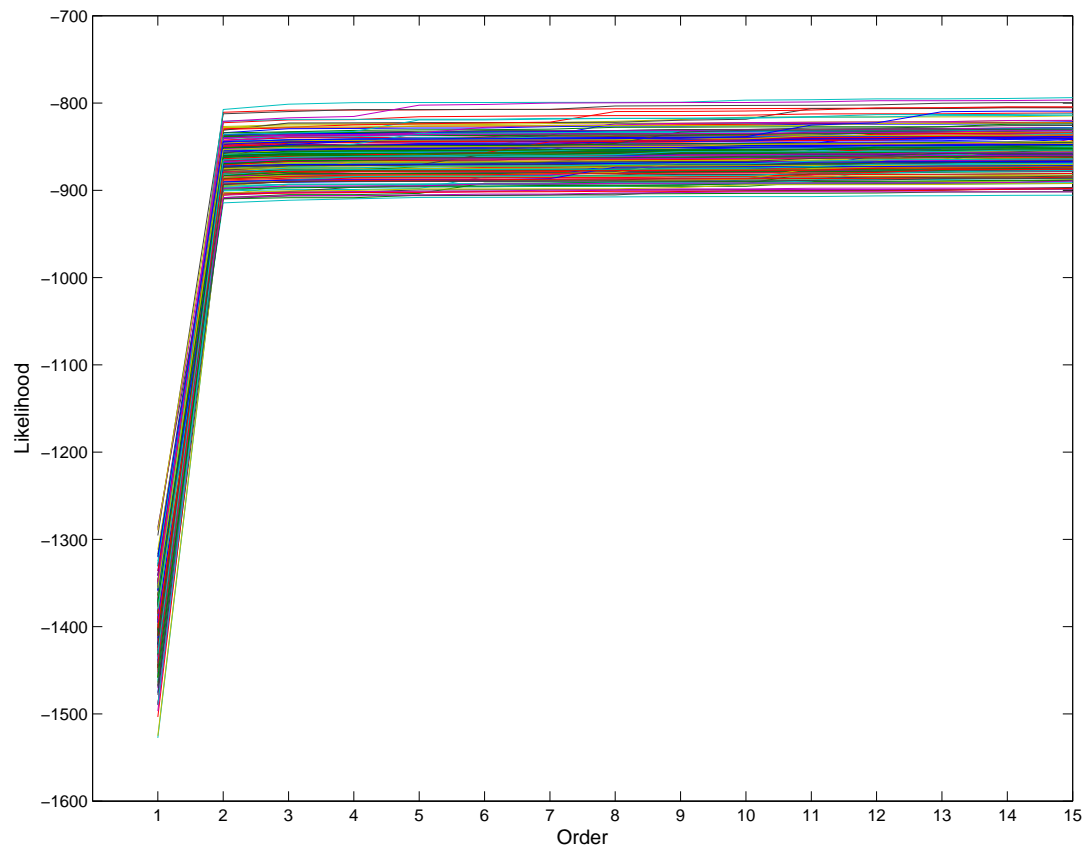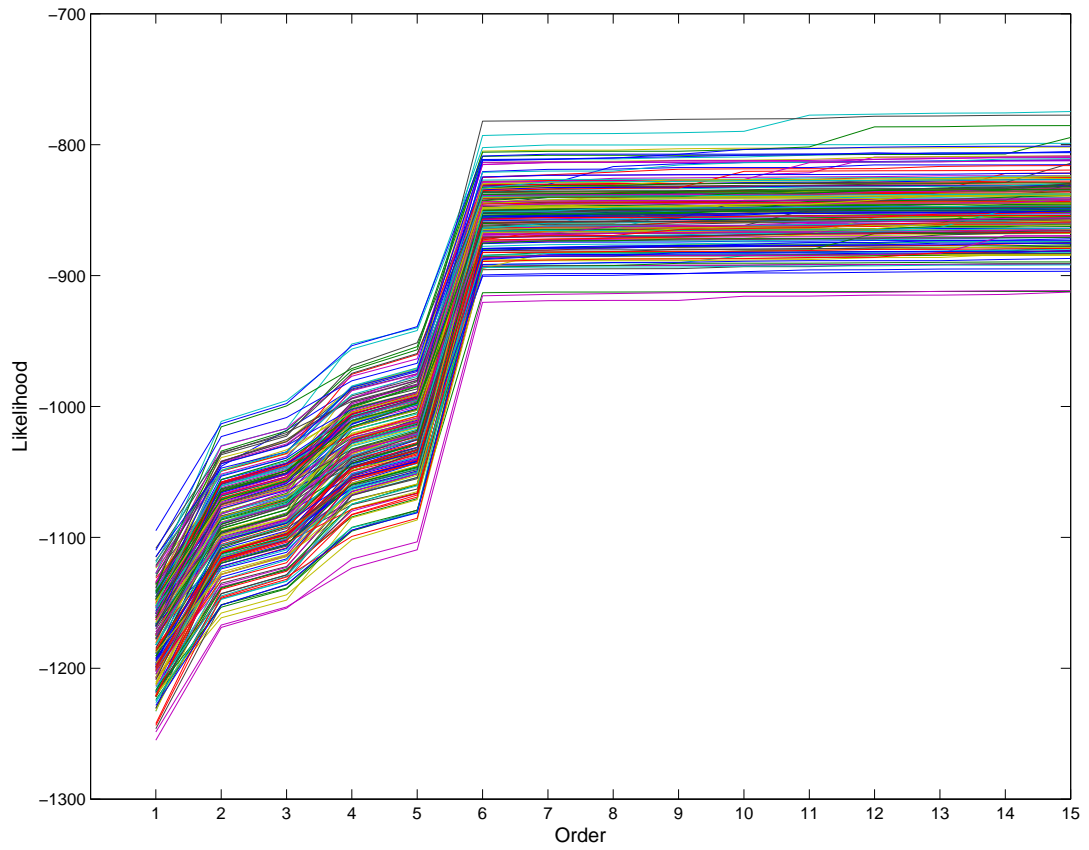
Figure 3.3: BLF-scree plot of the likelihood along with the TVAR order for the 200 realizations of the time-varying AR(6) process (TVAR6), given in Section 3.3.2.
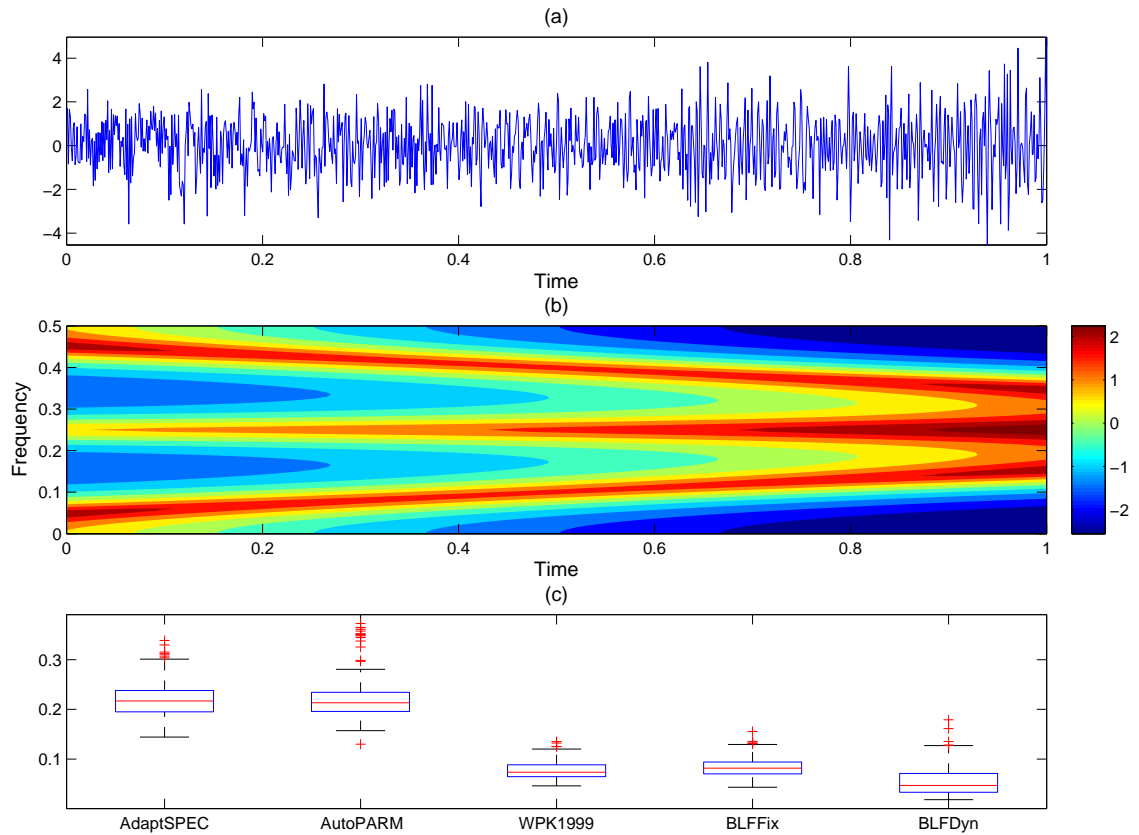
Figure 3.4: (a) and (b) depicts one realization and the true time-frequency representation of the time-varying AR(6) process (TVAR6), respectively (Section 3.3.2). (c) illustrates the box-plots of the average squared error (ASE) values corresponding to the time-frequency representation of TVAR6 for all of the approaches considered.
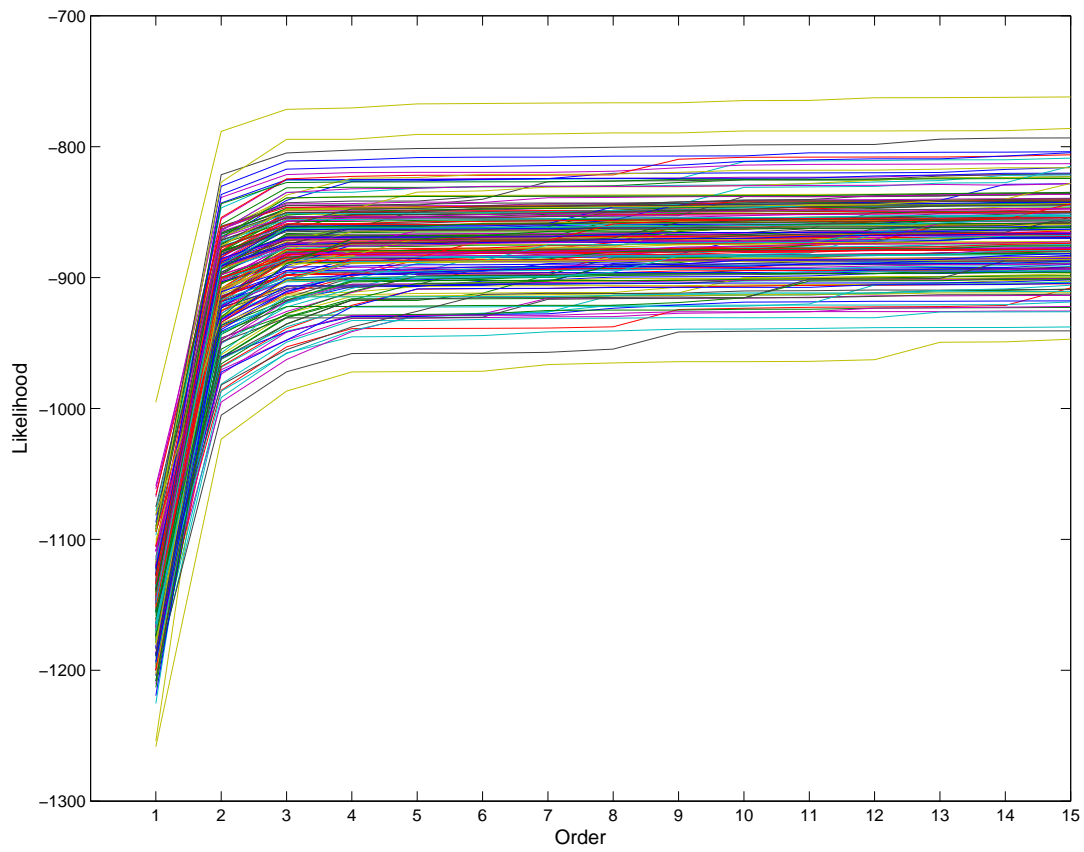
Figure 3.5: BLF-scree plot of the likelihood along with the TVAR order for the 200 realizations of the piecewise stationary AR process (PieceAR), given in Section 3.3.3.
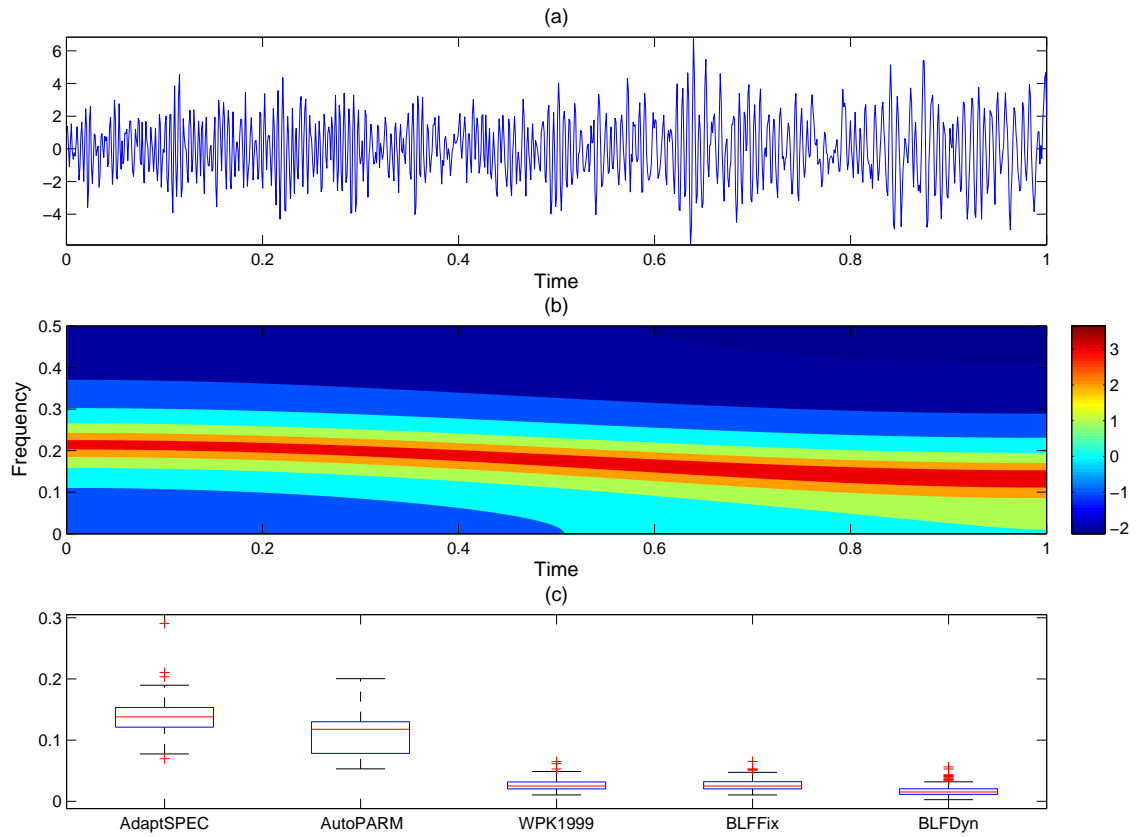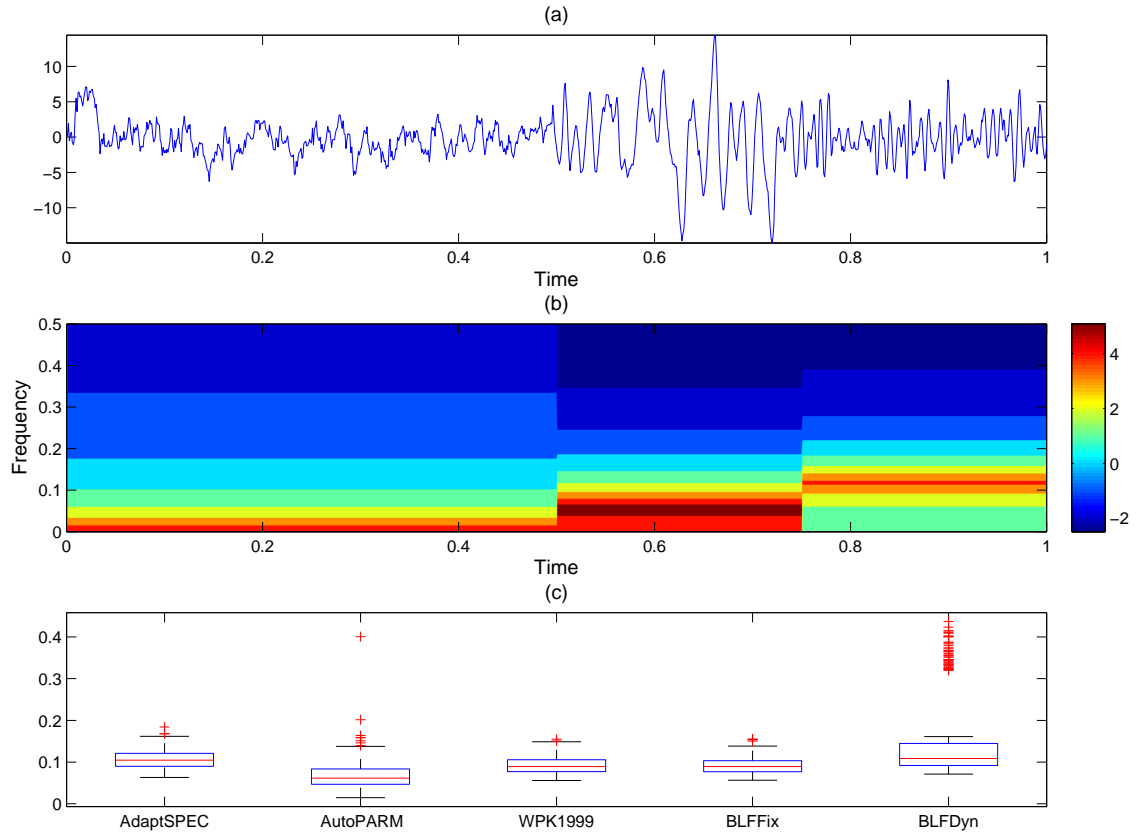
Figure 3.6: (a) and (b) depict one realization and the true time-frequency representation of the piecewise AR process (PieceAR), respectively (Section 3.3.3). (c) illustrates the box-plots of the average squared error (ASE) values corresponding to the time-frequency representation of the PieceAR for all of the approaches considered.
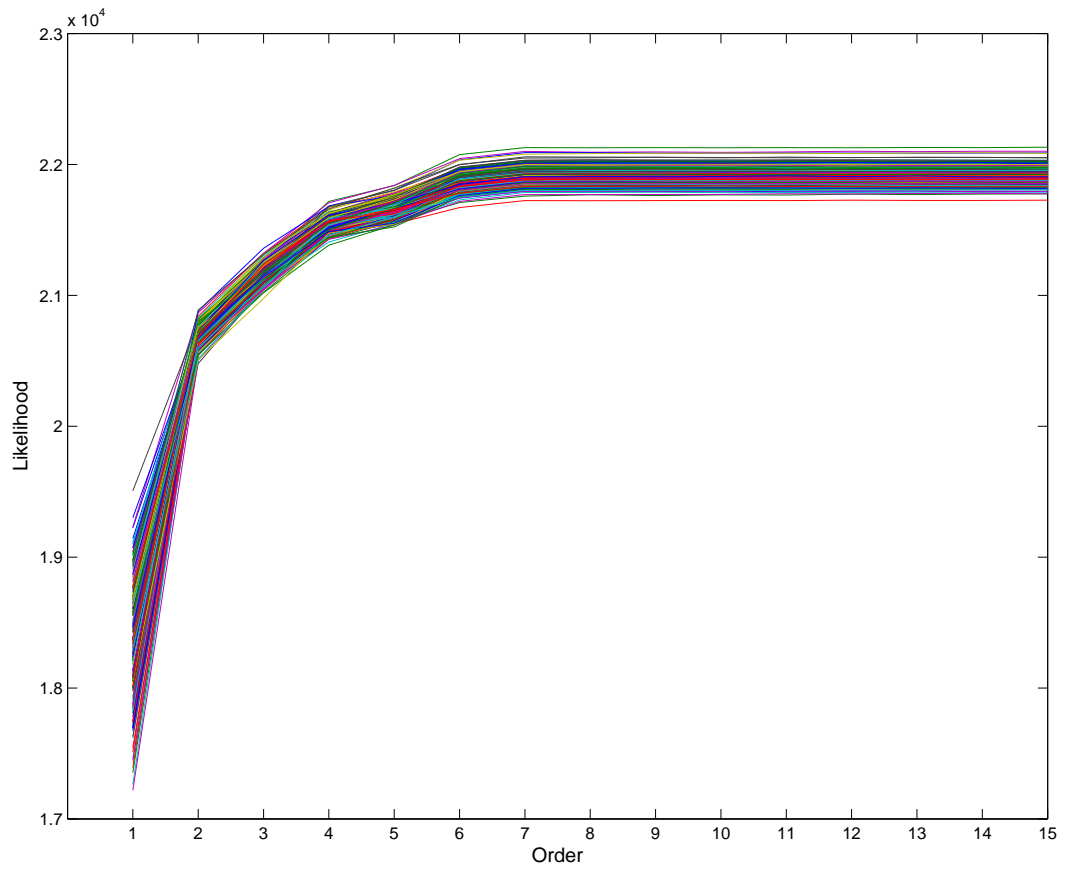
Figure 3.7: BLF-scree plot of the likelihood along with the TVAR order for the 200 realizations of the simulated insect communication signal (SimBugs), given in Section 3.3.4.
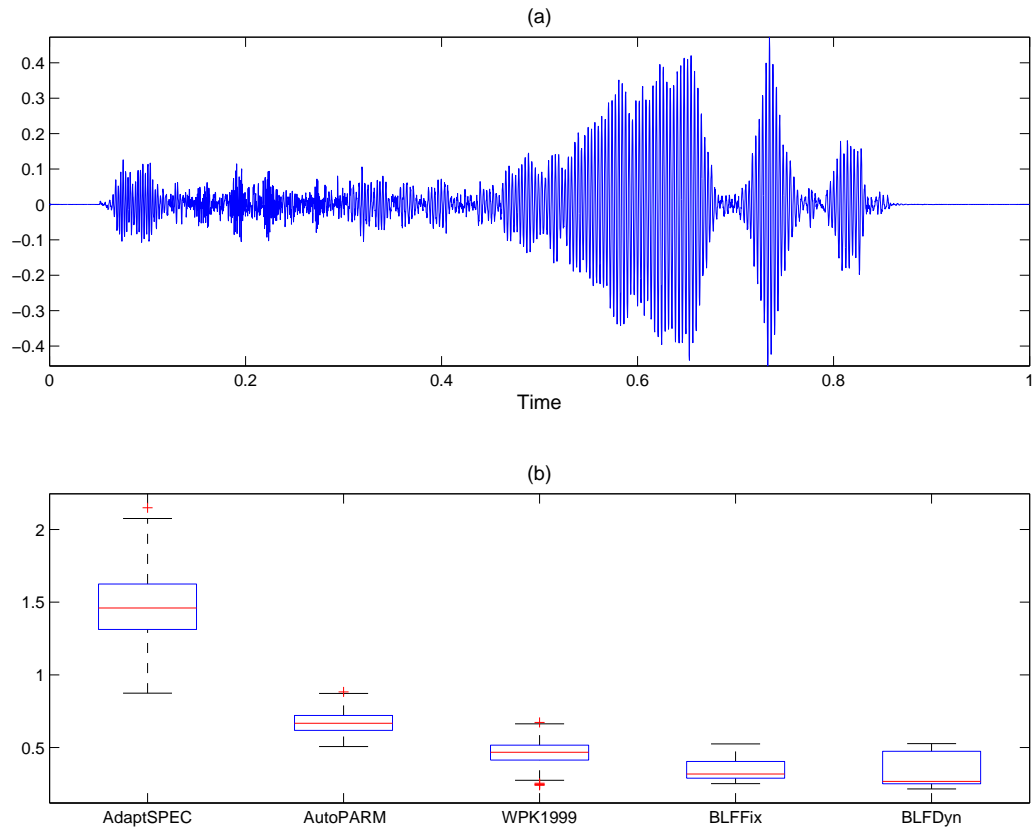
Figure 3.8: (a) depicts one realization of the simulated insect communication signals (SimBugs), Section 3.3.4. (b) illustrates the box-plots of the average squared error (ASE) values corresponding to the time-frequency representation of the SimBugs for all of the approaches considered.
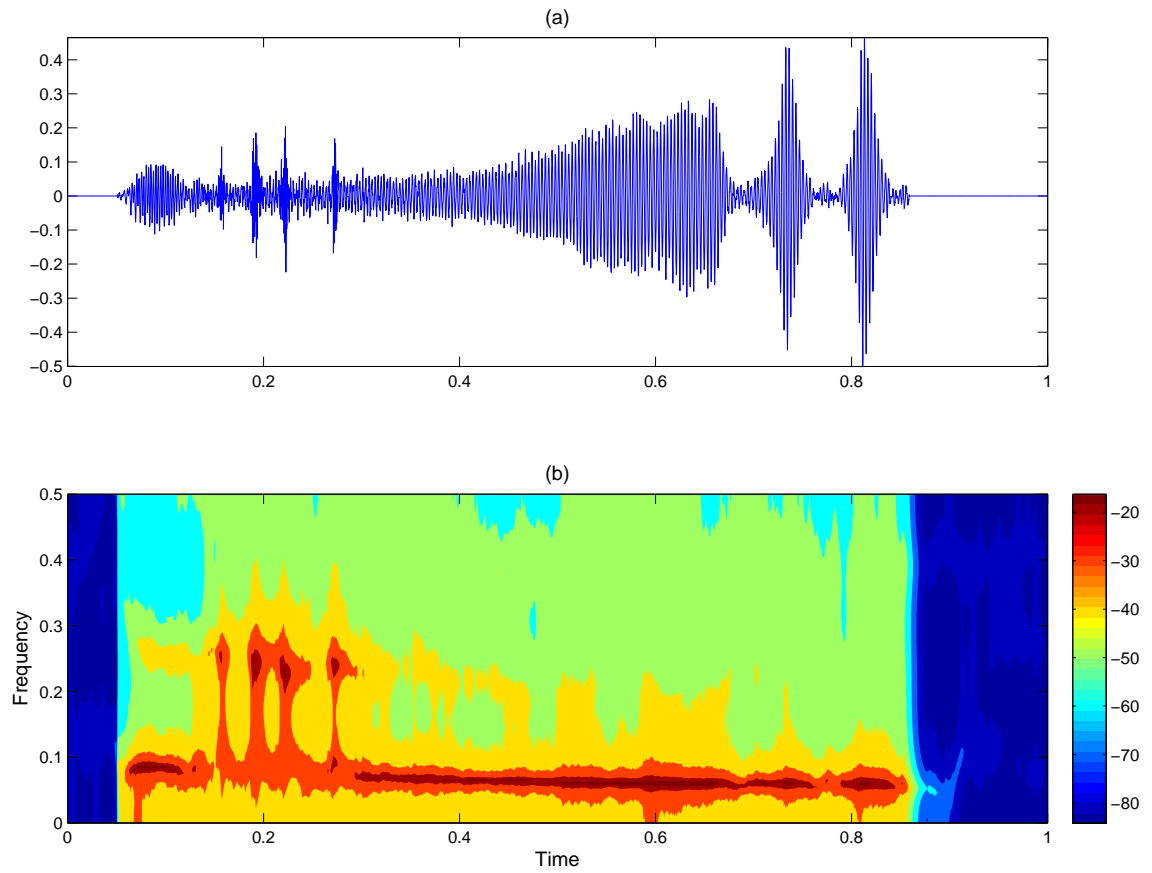
Figure 3.9: (a) An example of typical signal corresponding to a successful mater (Section 3.4.1). (b) Time-varying AR(6) spectral representation of the signal in plot (a).
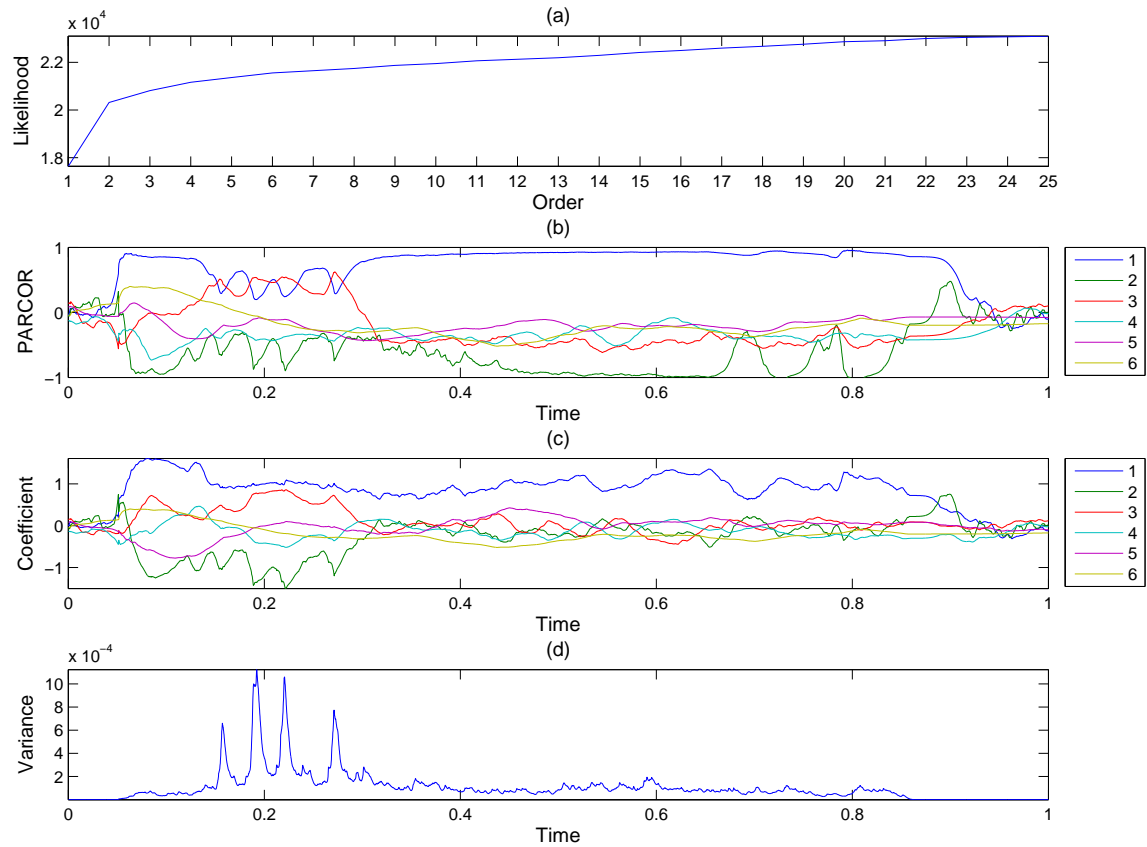
Figure 3.10: (a) shows the likelihood along with increase of the TVAR order for fitting the treehopper communication signal. (b) depicts the first six time-varying estimated PARCOR coefficients. (c) and (d) show the estimated time-varying coefficients and innovation variances of the time-varying AR(6) model.
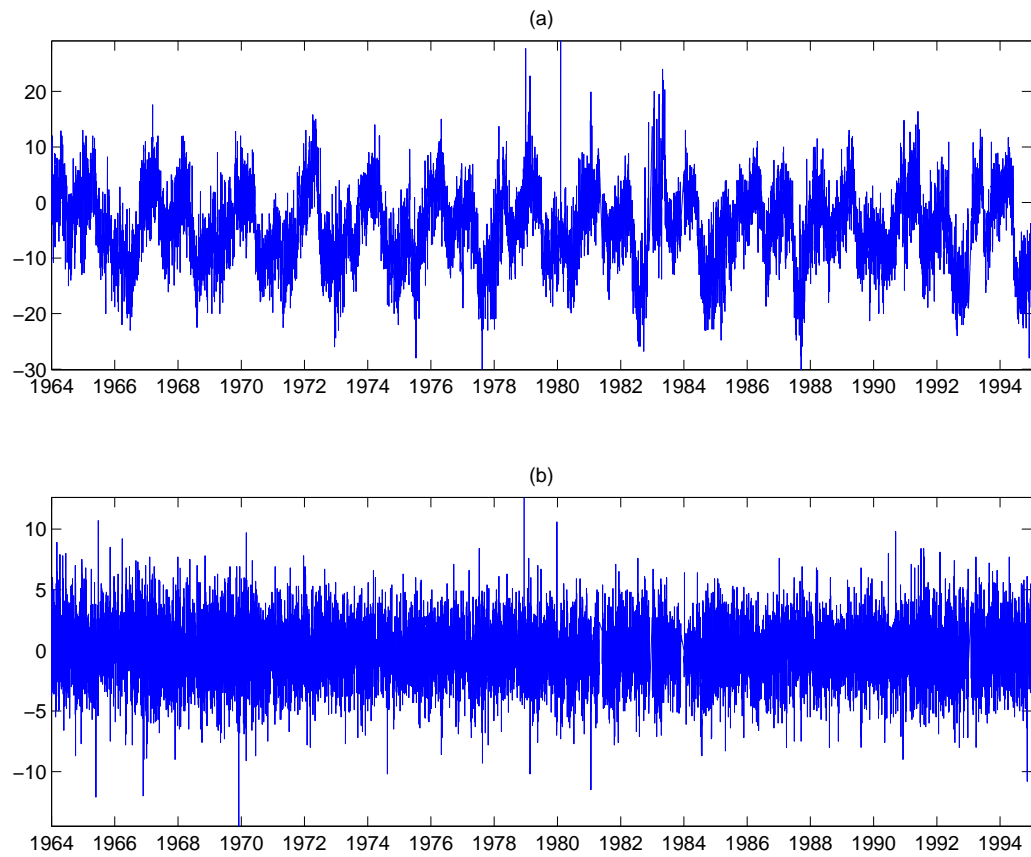
Figure 3.11: (a) and (b) show daily time series (1964-1994) of east/west and north/south components of wind, respectively.
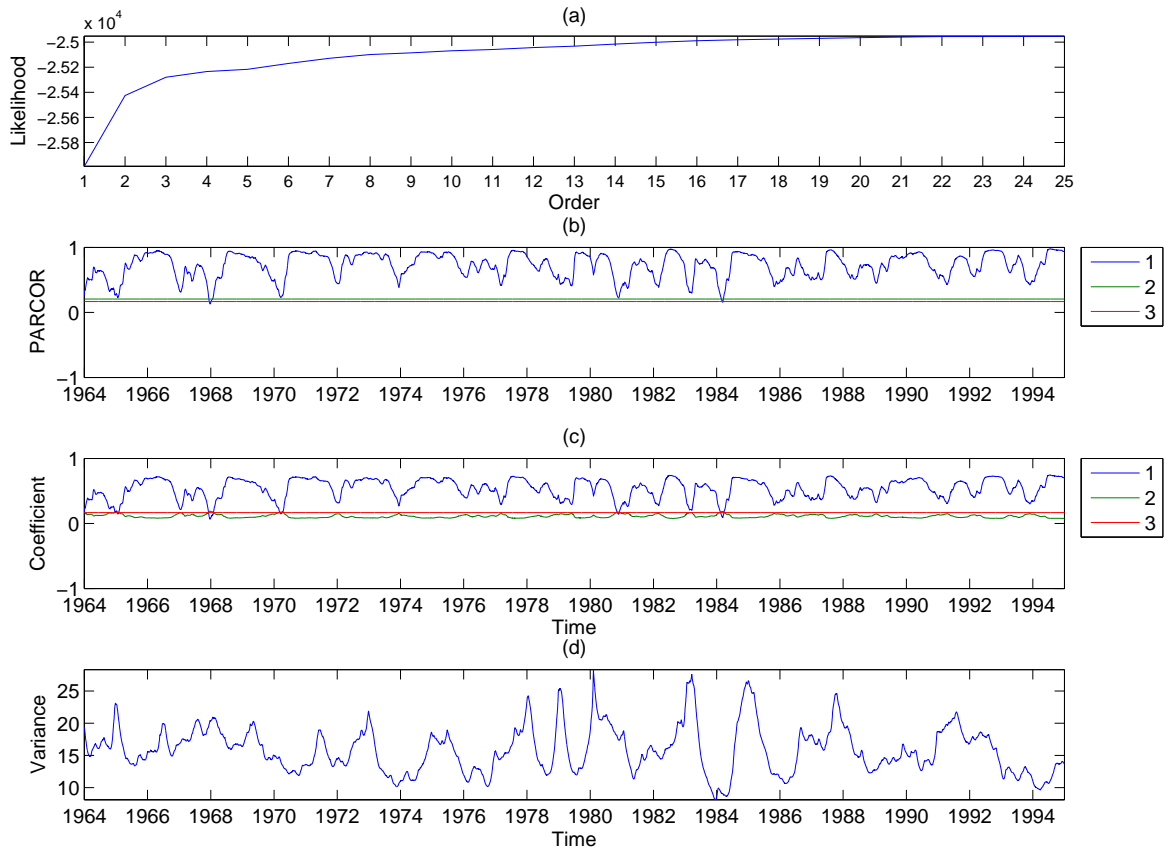
Figure 3.12: (a) shows the likelihood along with increase of the TVAR order for fitting the east/west component of the wind signal. (b) depicts the first three time-varying estimated PARCOR coefficients. (c) and (d) show the estimated time-varying coefficients and innovation variances of the time-varying AR(3) model.
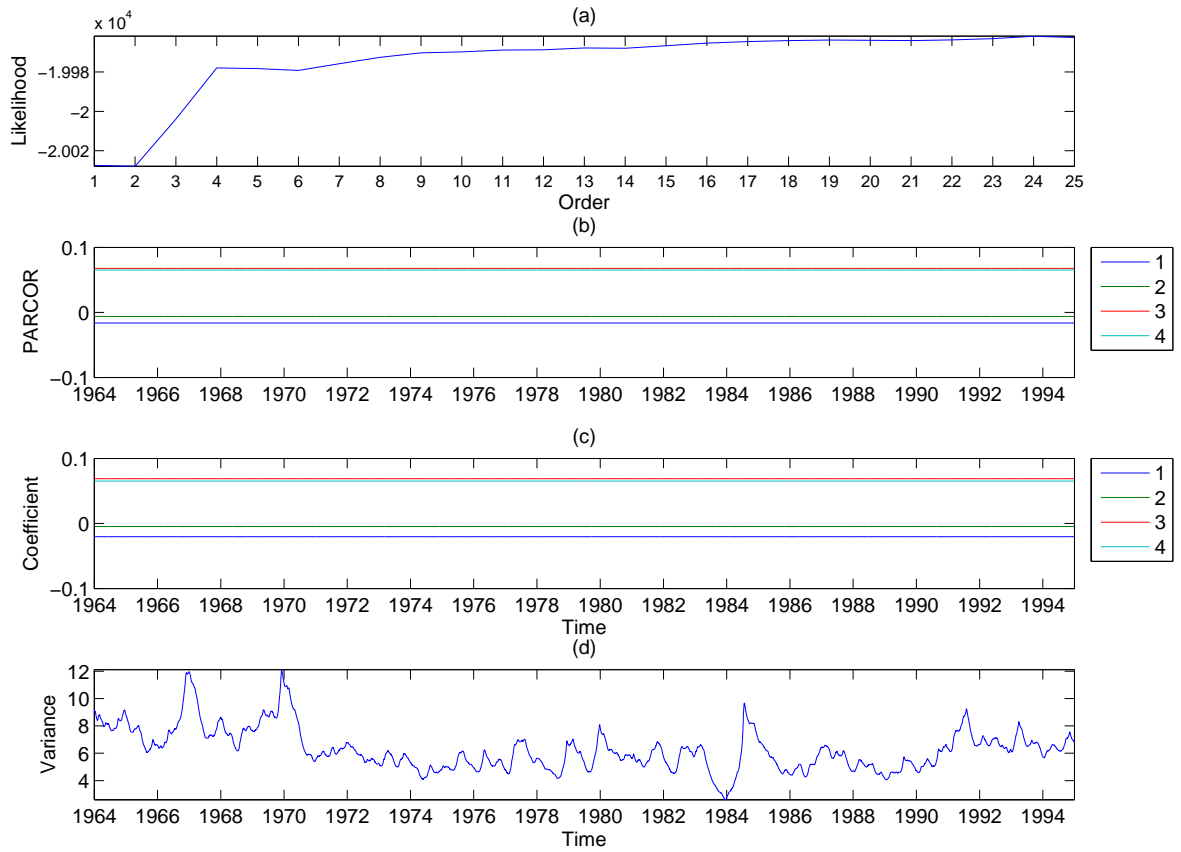
Figure 3.13: (a) shows the likelihood along with increase of the TVAR order for fitting the north/south component of the wind signal. (b) depicts the first four time-varying estimated PARCOR coefficients. (c) and (d) show the estimated time-varying coefficients and innovation variances of the time-varying AR(4).
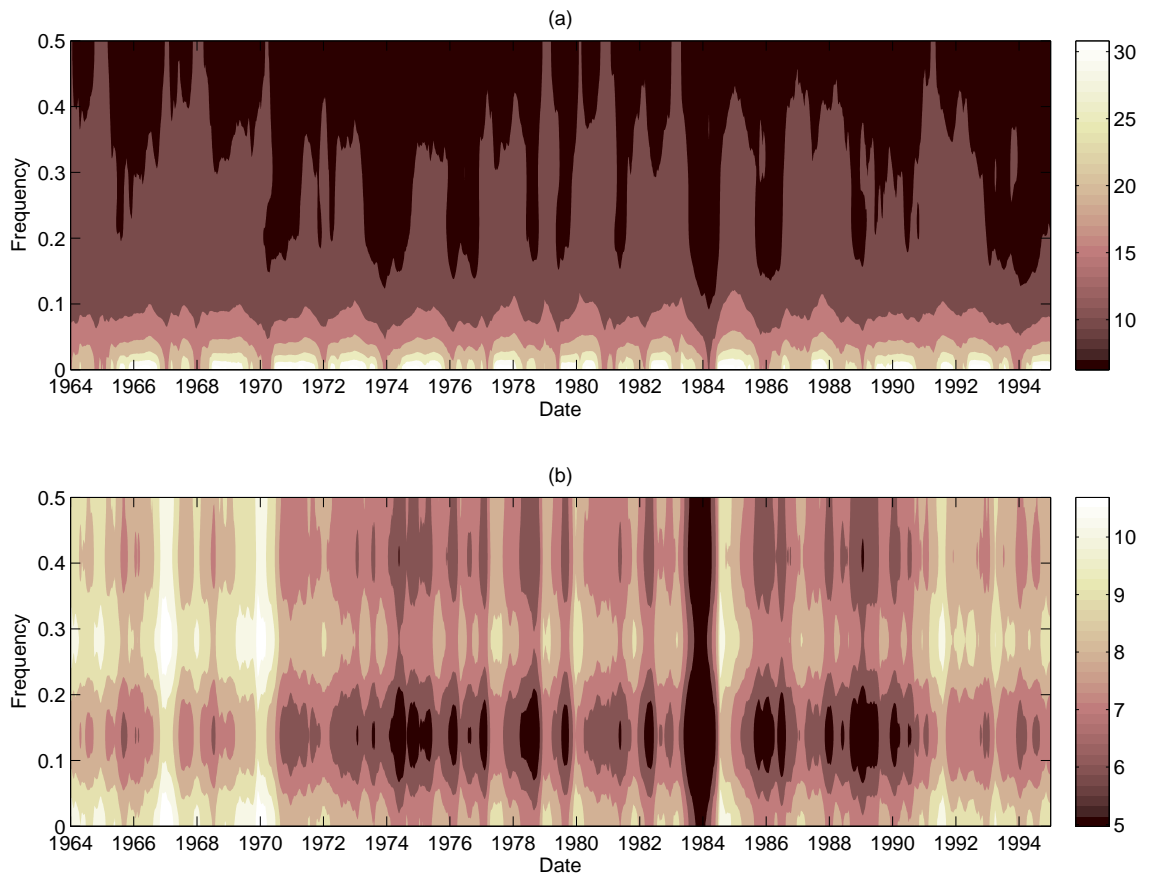
Figure 3.14: (a) and (b) displays the time-frequency representations of the wind east/west and north/south components.

Chapter 4

Bayesian Analysis of Spatially-Dependent Functional Responses with Spatially-Dependent Image Predictors

4.1 Introduction

With advances in instrumentation such as satellites, sensor networks, data storage tags, and spectroscopes, scientists are often faced with the problem of incorporating extremely high-dimensional covariates into statistical models. The efficient use of such "big data" is the subject of much active research in the statistics and computer science communities. However, such problems are only compounded when these data are collected over space and/or time, thereby introducing dependence. In cases where

the response is also spatial and/or temporal, this dependence may be accounted for by the covariates, but may also require temporal and/or spatially-explicit error structures. Another complication arises when the responses themselves are inherently functional (e.g., "curves" in time and/or space). This paper presents methodology that can accommodate high-dimensional "image" covariates that vary in space, as well as purely spatial scalar covariates, in the context of modeling spatially-dependent functional responses. The methodology is presented using a hierarchical Bayesian approach to account for uncertainties that arise in the observations, process and parameters. Inference and prediction are facilitated by the use of stochastic search variable selection (SSVS).

The use of so-called "image predictors" has seen increased utility in statistical models over the last few years (e.g., see Reiss and Ogden, 2010; Morris et al., 2011; Holan et al., 2010, 2012; Martinez et al., 2013; Yang et al., 2013, Chapter 2). Holan et al. (2010) showed in the context of insect communication that one could treat a time-frequency representation of a high-frequency nonstationary time signal as an "image" and, with suitable functional dimension reduction and SSVS, easily incorporate such big data covariates into classical generalized linear mixed models. This was subsequently considered in the context of business cycle modeling (Holan et al., 2012) and with the incorporation of nonlinear interactions in modeling spawning success of shovelnose sturgeon in Yang et al. (2013) (Chapter 2). Recently, Martinez et al. (2013) considered a functional mixed model approach to modeling acoustic signals associated with bats. To our knowledge, modeling spatially-correlated functional data with spatially-dependent image predictors has not been considered to date.

As previously alluded to, the use of spatially-dependent image predictors is com-

pounded in the case where one has responses that are spatially-dependent functions as well. Although functional responses have been considered in the context of image prediction (e.g., Morris et al., 2011), the spatially-dependent functional case has not been considered. Functional data analysis is fairly mature in statistics (e.g., see Bosq, 2000; Ramsay and Silverman, 2005, among others), yet spatial functional data analysis has just recently become an active sub-field of spatial statistics and functional data analysis. Excellent reviews of recent work in the area can be found in Delicado et al. (2010), Ruiz-Medina (2012a), and Kokoszka (2012). In general, geostatistical predictive and clustering approaches have focused on co-kriging ideas (e.g., Goulard and Voltz, 1993; Monestiez and Nerini, 2008; Giraldo et al., 2010, 2012) and the general theory of spatial autoregressive and moving average Hilbertian processes (Ruiz-Medina, 2011; Ruiz-Medina and Montes, 2011; Ruiz-Medina, 2012b; Ruiz-Medina and Espejo, 2013). In addition, a more traditional functional principal components approach to spatially-dependent functions, where interest is on estimation of mean functions, is given in Gromenko et al. (2012) and Gromenko and Kokoszka (2013). These approaches have been from the classical perspective, with relatively few Bayesian implementations. Notable exceptions include Baladandayuthapani et al. (2008) who consider a Bayesian hierarchical model with relatively simple spatial dependence on the functions at low levels of the hierarchy. This was extended by Zhou et al. (2010) to have a more flexible covariance structure, and was implemented using an expectation-maximization (E-M) algorithm. These Bayesian approaches to spatial functional data analysis have not considered SSVS, although Li and Zhang (2010) provide the groundwork for such high-dimensional, highly-structured SSVS.

The contribution of this chapter is then the development of methodology for

modeling spatially-dependent functional responses in terms of spatially-dependent functional-image predictors, along with spatially-dependent covariates, within a Bayesian paradigm that utilizes SSVS. Section 4.2 describes the methodology. A motivating example from soil science is given in Section 4.3, followed by a discussion and concluding remarks in Section 4.4.

4.2 Methodology

In this section, we introduce a class of spatially-explicit functional models. Different from traditional spatial models, the proposed models include functional responses and functional covariates. In general, these functional covariates can be curves of one dimension, images of two dimensions, or objects of higher dimensions. In this chapter, we focus on two-dimensional image covariates that are spatially dependent, although it is straightforward to generalize the model to account for other functionals and their interactions (e.g., see Yang et al., 2013, Chapter 2). Inference is performed in the Bayesian framework, utilizing stochastic search variable selection (SSVS) algorithms based on a mixture of normal priors.

4.2.1 Spatially-Dependent Functional-Image Model

We denote a response functional to be a continuous spatial process $\{Y(\mathbf{s}, d) : \mathbf{s} \in D \subset \mathfrak{R}^2, d \in \mathfrak{D} \subset \mathfrak{R}\}$, where D is a continuous spatial domain. Furthermore, $\mathfrak{D}$ represents a continuous one-dimensional domain, such as time or depth. For purposes of this exposition, we refer to this dimension as "depth" to coincide with the application in Section 4.3, but note that there are many applications in which this index

would correspond to time. Also, denote $X_j(\mathbf{s}, \mathbf{u}_j)$ the j-th observed two-dimensional functional covariate, $j = 1, \ldots, J$, at spatial location $\mathbf{s}$ with $\{\mathbf{u}_j = (d, \omega_j)' : \omega_j \in \Omega\}$ corresponding to the index of the two image dimensions of interest (e.g., depth and wavelength, respectively). Note that the j subscript on $\mathbf{u}_j$ serves to indicate that the image "pixels" may be different for the different covariates (although this is not the case in the example presented in Section 4.3). To give this notation some perspective relative to our application in Section 4.3, we let Ω correspond to the continuous frequency (or wavelength) domain.

The primary model can then be written, at location $\mathbf{s}$, as the relation between the response functional and functional covariates as follows

$$Y(\mathbf{s}, d) = \sum_{j=1}^{J} \int X_j(\mathbf{s}, \mathbf{u}_j) \beta_j(\mathbf{u}_j, d) d\mathbf{u}_j + \mathbf{z}'(\mathbf{s}) \boldsymbol{\delta}(d) + \eta(\mathbf{s}, d) + \epsilon(\mathbf{s}, d), \qquad (4.1)$$

where $\beta_j(\mathbf{u}_j, d)$ is a square integrable functional coefficient corresponding to $X_j(\mathbf{s}, \mathbf{u}_j)$, $\boldsymbol{\delta}(d)$ is a p-vector of regression coefficients associated with the $p \times 1$ spatially indexed covariate vector $\mathbf{z}(\mathbf{s})$, $\eta(\mathbf{s}, d)$ is a random effect capturing spatial and depth dependence, and $\epsilon(\mathbf{s}, d)$ is an independent Gaussian process with mean zero and variance σ_ϵ^2 (i.e., the nugget or measurement error). Typically, we specify $\eta(\mathbf{s}, d)$ to be a Gaussian process with mean zero and covariance function $C(\mathbf{s}, d)$. In a traditional spatial analysis, one might consider a three-dimensional spatial covariance function if d corresponds to the vertical dimension, or a spatio-temporal covariance function if d corresponds to time (e.g. Cressie and Wikle, 2011). Here, we deliberately keep these indices separate given that the response and covariate vary functionally in the dimension d, and this is assumed to operate on a different scale of variability relative to the two-dimensional spatial component of the process. That is, different from

traditional geostatistics that is concerned with scalar or vector variables at location $\mathbf{s}$, the proposed model considers the case that the response variable of interest is a functional (curve) in depth (or time) and some of covariates are also functionals. In this sense, the function $\beta_j(\mathbf{u}_j, d)$ can be thought of as a kernel that "distributes" the covariate $X_j(\mathbf{s}, \mathbf{u})$ to the depth of the response variable at location $\mathbf{s}$.

Assume that $\{\phi_{jk}(\mathbf{u}_j) : k = 1, \ldots, \infty\}$ form a complete orthonormal basis corresponding to the j-th functional covariate. Then, we have the unique representation of the functional covariate

$$X_j(\mathbf{s}, \mathbf{u}_j) = \sum_{k=1}^{\infty} \xi_{jk}(\mathbf{s})\phi_{jk}(\mathbf{u}_j),$$

where $\xi_{jk}(\mathbf{s})$ are expansion coefficient functions (for a given location $\mathbf{s}$) associated with the jth functional covariate. In addition, by considering the same basis we also have the unique representation

$$\beta_j(\mathbf{u}_j, d) = \sum_{k=1}^{\infty} b_{jk}(d)\phi_{jk}(\mathbf{u}_j),$$

where $b_{jk}(d)$ are expansion coefficient functions (for a given depth d) associated with the j-th square integrable function. Substituting these expressions into (4.1) and making use of the orthogonality, we obtain

$$Y(\mathbf{s}, d) = \sum_{j=1}^{J}\sum_{k=1}^{\infty} \xi_{jk}(\mathbf{s})b_{jk}(d) + \mathbf{z}'(\mathbf{s})\boldsymbol{\delta}(d) + \eta(\mathbf{s}, d) + \epsilon(\mathbf{s}, d). \qquad (4.2)$$

In addition, assume that $\{\psi_i(d) : i = 1, \ldots, \infty\}$ form a complete orthonormal

basis corresponding to depth. Then, we have the unique representations

$$Y(\mathbf{s}, d) = \sum_{i=1}^{\infty} \alpha_i(\mathbf{s})\psi_i(d),$$

$$b_{jk}(d) = \sum_{i=1}^{\infty} b_{jki}\psi_i(d),$$

$$\eta(\mathbf{s}, d) = \sum_{i=1}^{\infty} \theta_i(\mathbf{s})\psi_i(d),$$

where $\alpha_i(\mathbf{s})$ and $\theta_i(\mathbf{s})$ are expansion coefficient functions (of $\mathbf{s}$) corresponding to $Y(\mathbf{s}, d)$ and $\eta(\mathbf{s}, d)$, respectively, and b_{jki} are expansion coefficients associated with $b_{jk}(d)$. Replacing $Y(\mathbf{s}, d)$, $b_{jk}(d)$ and $\eta(\mathbf{s}, d)$ in (4.2) with these expansions, we can rewrite (4.2) as

$$\sum_{i=1}^{\infty} \alpha_i(\mathbf{s})\psi_i(d) = \sum_{j=1}^{J}\sum_{k=1}^{\infty}\sum_{i=1}^{\infty} \xi_{jk}(\mathbf{s})b_{jki}\psi_i(d) + \mathbf{z}'(\mathbf{s})\boldsymbol{\delta}(d) + \sum_{i=1}^{\infty} \theta_i(\mathbf{s})\psi_i(d) + \epsilon(\mathbf{s}, d). \quad (4.3)$$

Finally, assume that $\{w_\ell(\mathbf{s}) : \ell = 1, \ldots, \infty\}$ form a complete orthonormal basis corresponding to $\mathbf{s}$, which then gives the unique representations

$$\alpha_i(\mathbf{s}) = \sum_{\ell=1}^{\infty} w_\ell(\mathbf{s})a_{i\ell},$$

$$\xi_{jk}(\mathbf{s}) = \sum_{\ell=1}^{\infty} w_\ell(\mathbf{s})f_{jk\ell},$$

$$\theta_i(\mathbf{s}) = \sum_{\ell=1}^{\infty} w_\ell(\mathbf{s})g_{i\ell},$$

where $a_{i\ell}$, $f_{jk\ell}$, and $g_{i\ell}$ are expansion coefficients associated with $\alpha_i(\mathbf{s})$, $\xi_{jk}(\mathbf{s})$, and $\theta_i(\mathbf{s})$, respectively. Substituting the above expansions into (4.3), we obtain the up-

dated representation of (4.3),

$$\sum_{i=1}^{\infty}\sum_{\ell=1}^{\infty} w_\ell(\mathbf{s})\psi_i(d)a_{i\ell} = \sum_{j=1}^{J}\sum_{k=1}^{\infty}\sum_{i=1}^{\infty}\sum_{\ell=1}^{\infty} w_\ell(\mathbf{s})\psi_i(d)f_{jk\ell}b_{jki} + \mathbf{z}'(\mathbf{s})\boldsymbol{\delta}(d)$$
$$+ \sum_{i=1}^{\infty}\sum_{\ell=1}^{\infty} w_\ell(\mathbf{s})\psi_i(d)g_{i\ell} + \epsilon(\mathbf{s},d). \qquad (4.4)$$

Similar to applications in traditional functional data analysis, we consider a finite approximation to the infinite summation in (4.4),

$$\sum_{i=1}^{n_i}\sum_{\ell=1}^{n_\ell} w_\ell(\mathbf{s})\psi_i(d)a_{i\ell} = \sum_{j=1}^{J}\sum_{k=1}^{n_{k_j}}\sum_{i=1}^{n_i}\sum_{\ell=1}^{n_\ell} w_\ell(\mathbf{s})\psi_i(d)f_{jk\ell}b_{jki} + \mathbf{z}'(\mathbf{s})\boldsymbol{\delta}(d)$$
$$+ \sum_{i=1}^{n_i}\sum_{\ell=1}^{n_\ell} w_\ell(\mathbf{s})\psi_i(d)g_{i\ell} + \epsilon(\mathbf{s},d). \qquad (4.5)$$

In practice the truncations n_{k_j}, n_i, and n_ℓ are typically problem specific and can be chosen based on percent of variance explained, cross-validation , and/or sensitivity analysis. Further, denote $\boldsymbol{\psi}(d) \equiv [\psi_1(d), \ldots, \psi_{n_i}(d)]'$ and $\mathbf{w}(\mathbf{s}) \equiv [w_1(\mathbf{s}), \ldots, w_{n_i}(\mathbf{s})]'$. Then, we can rewrite (4.5) as

$$\mathbf{w}(\mathbf{s})'\mathbf{A}\,\boldsymbol{\psi}(d) = \sum_{j=1}^{J} \mathbf{w}(\mathbf{s})'\mathbf{F}_j\mathbf{b}_j\boldsymbol{\psi}(d) + \mathbf{z}'(\mathbf{s})\boldsymbol{\delta}(d) + \mathbf{w}'(\mathbf{s})\mathbf{G}\,\boldsymbol{\psi}(d) + \epsilon(\mathbf{s},d), \qquad (4.6)$$

where $\mathbf{A}$ and $\mathbf{G}$ are $n_\ell \times n_i$ matrices with elements $\{a_{\ell i}\}$ and $\{g_{\ell i}\}$, respectively, $\mathbf{F}_j$ is an $n_\ell \times n_k$ matrix with elements $\{f_{j\ell k}\}$, and $\mathbf{b}_j$ is an $n_{k_j} \times n_i$ matrix with elements $\{b_{jki}\}$, for $i = 1, \ldots, n_i$, $k = 1, \ldots, n_{k_j}$, and $\ell = 1, \ldots, n_\ell$.

In practice, one typically has observations at a finite collection of spatial locations and finite depths. In such a case, consider n_s spatial locations and n_d depths. We denote $\boldsymbol{\Psi} \equiv [\boldsymbol{\psi}(d_1), \ldots, \boldsymbol{\psi}(d_{n_d})]$ with the assumption of $n_d > n_i$ and

$\mathbf{W} \equiv [\mathbf{w}(\mathbf{s}_1), \ldots, \mathbf{w}(\mathbf{s}_{n_s})]'$ with the assumption of $n_s > n_\ell$. Consequently, we can rewrite (4.6) as

$$\mathbf{WA\Psi} = \sum_{j=1}^{J} \mathbf{WF}_j \mathbf{b}_j \mathbf{\Psi} + \mathbf{Z\Delta} + \mathbf{WG\Psi} + \mathbf{E}, \tag{4.7}$$

with $\mathbf{Z} \equiv [\mathbf{z}(\mathbf{s}_1), \ldots, \mathbf{z}(\mathbf{s}_{n_{ns}})]'$, $\mathbf{\Delta} \equiv [\boldsymbol{\delta}(d_1), \ldots, \boldsymbol{\delta}(d_{n_d})]$, and $\mathbf{E} \equiv [\boldsymbol{\varepsilon}(\mathbf{s}_1), \ldots, \boldsymbol{\varepsilon}(\mathbf{s}_{n_s})]'$, where $\boldsymbol{\varepsilon}(\mathbf{s}) \equiv [\epsilon(\mathbf{s}, d_1), \ldots, \epsilon(\mathbf{s}, d_{n_d})]'$. Because $\mathbf{W}$ and $\mathbf{\Psi}$ are orthogonal, it follows that $\mathbf{W}'\mathbf{W} = \mathbf{I}_{n_\ell}$ and $\mathbf{\Psi}\mathbf{\Psi}' = \mathbf{I}_{n_i}$. Then, we can simplify (4.7) such that

$$\mathbf{A} = \sum_{j=1}^{J} \mathbf{F}_j \mathbf{b}_j + \mathbf{W}'\mathbf{Z\Delta\Psi}' + \mathbf{G} + \mathbf{W}'\mathbf{E\Psi}', \tag{4.8}$$

where $\mathbf{W}$ and $\mathbf{\Psi}$ are known. Consequently, our model reduces to a multivariate multiple regression model with random effects.

To preserve the matrix structure, we utilize matrix variation normal distributions as introduced by Dawid (1981). That is, if an $m \times t$ matrix $\mathbf{U}$ follows a matrix variation normal distribution, it can be expressed as

$$\mathbf{U} \sim N_{m,t}(\mathbf{H}, \mathbf{\Sigma}_m, \mathbf{\Sigma}_t),$$

where $\mathbf{H}$ is an $m \times t$ matrix of mean values, $\mathbf{\Sigma}_m$ is an $m \times m$ covariance matrix between rows, and $\mathbf{\Sigma}_t$ is an $t \times t$ covariance matrix between columns. This can also be written as a multivariate normal distribution with $\mathrm{vec}(\mathbf{U}) \sim MVN(\mathrm{vec}(\mathbf{H}), \mathbf{\Sigma}_t \otimes \mathbf{\Sigma}_m)$, where $\mathrm{vec}(\mathbf{U})$ is the vectorization of a matrix $\mathbf{U}$, and one can see by the Kronecker product that the standard matrix normal representation implies a separable covariance structure between the matrix row and column variables. In addition, the matrix variation normal distributions have a property that the linear transforma-

tions of $\mathbf{U}$ still follow the matrix variation normal distribution such as $\mathbf{MUT'} \sim N_{m,t}(\mathbf{MHT'}, \mathbf{M\Sigma_m M'}, \mathbf{T\Sigma_t T'})$.

In our model, we then note that $\mathbf{W'E\Psi'} \sim N_{n_\ell, n_i}(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_{n_\ell}, \mathbf{I}_{n_i})$ since $\mathbf{W'W} = \mathbf{I}$ and $\mathbf{\Psi\Psi'} = \mathbf{I}$. Furthermore, we specify $\mathbf{G} \sim N_{n_\ell, n_i}(\mathbf{0}, \mathbf{\Sigma}_{n_\ell}, \mathbf{\Sigma}_{n_i})$, where $\mathbf{\Sigma}_{n_\ell} = \mathbf{W'C_s W}$ and $\mathbf{\Sigma}_{n_i} = \mathbf{\Psi C_d \Psi'}$. In this case, $\mathbf{C}_d$ and $\mathbf{C_s}$ are $n_d \times n_d$ and $n_s \times n_s$ depth and spatial covariance matrices, respectively. These may be specified, as is typical in geostatistics, according to some valid spatial covariance function (e.g., a Matérn model) or empirically, as in functional principal components analysis. Often $\mathbf{\Sigma}_{n_i}$ and $\mathbf{\Sigma}_{n_\ell}$ are then diagonal matrices (or nearly so) (e.g., see Cressie and Wikle, 2011). Note that a non-separable covariance structure could be specified between depth and space, as mentioned previously.

Upon estimation of the parameters, one can then perform prediction of the response at any location and depth by utilizing the appropriate basis expansions to obtain $\mathrm{E}(Y(\mathbf{s}, d))$. Critically, one must have basis functions that are defined at any desired spatial and depth location.

4.2.2 Stochastic Search Variable Selection

Although the basis functions simplify the integrals and reduce the dimension of the functional data and parameters, the dimensionality still can increase by the number of functional covariates. This is particularly problematic in the context of image predictors, as there are often very high numbers of "pixels" of potential interest, even in the context of a functional representation (e.g., Holan et al. 2010). Therefore, the model we propose makes use of a SSVS algorithm. Such algorithms provide an effective means of model selection when interest lies in considering a large number of

potential submodels (see George, 2000, for a detailed overview).

In general, SSVS priors include two types of mixture distributions (George and McCulloch, 1997). One comes from a mixture of normals and the other is a mixture of a Dirac function and a normal distribution. Many variations of these two basic algorithms have been proposed and applied across a broad range of applications (see O'Hara and Sillanpää, 2009; Stingo et al., 2012, and the references therein). Among them, Brown and Vannucci (1998) extend the SSVS algorithm to multivariate regression models using a mixture of a Dirac function and a normal distribution. Both of these prior formulations can be incorporated into our model. For the sake of brevity, we only describe the SSVS prior of a mixture of normals here.

We denote $\mathbf{B} = [\mathbf{b}_1', \ldots, \mathbf{b}_j', (\mathbf{\Delta \Psi}')']'$ as an $Q \times n_i$ matrix and $\boldsymbol{\gamma}$ as a latent binary vector of length Q associated with each row of $\mathbf{B}$, where $Q = (\sum_{j=1}^{J} n_{k_j} + p)$. Using similar notation to Brown and Vannucci (1998) and Brown et al. (2001), we have the prior for each row of $\mathbf{B}$

$$[\mathbf{B}]_{[q,:]}|\gamma_q \sim (1 - \gamma_q)MVN(\mathbf{0}, \tau \mathbf{\Sigma}_{n_i}) + \gamma_q MVN(\mathbf{0}, c\tau \mathbf{\Sigma}_{n_i}), \ q = 1, \ldots, Q,$$

where $\mathbf{\Sigma}_{n_i}$ was defined above, $[\mathbf{B}]_{[q,:]}$ denotes the q-th row of $\mathbf{B}$, and c and τ are hyperpriors. In this formulation, if $\gamma_q = 0$, the associated parameters are close to zero, and if $\gamma_q = 1$, the associated parameters are not necessarily close to zero. Consequently, τ is typically chosen to be close to zero and c is specified to be large. Further, $\mathbf{B}$ can be presented as a matrix variation normal distribution,

$$\mathbf{B}|\boldsymbol{\gamma} \sim N_{Q,n_i}(\mathbf{0}, \mathbf{\Sigma}_Q, \mathbf{\Sigma}_{n_i}),$$

where $\mathbf{\Sigma}_Q$ is a diagonal matrix with elements equal to τ and $c\tau$. We also specify a Bernoulli distribution for γ_q,

$$\gamma_q \sim Bern(\pi_q),$$

where π_qs are hyperparameters between 0 and 1. In general, we can give different values for π_q, $q = 1, \ldots, Q$ or specify π_q to follow a beta distribution. In this context, we specify a common value for π_q, as discussed in Section 4.3.4.

4.3 Soil Science Application

This section describes a real-world motivating application in soils science. We present the motivating problem and data, followed by the specific model choices used for implementation, Bayesian estimation, and the results.

4.3.1 Motivating Problem

Scientific understanding of soil is important for many purposes, including the improvement of sustainable agricultural production and possible remediation of atmospheric carbon dioxide through carbon sequestration. Soil science traditionally requires significant and laborious field work to collect samples as well as significant laboratory time to analyze the chemical and biological properties of those samples. Increasingly, soil sensing technology such as from precision agriculture, digital soil mapping, and optical diffuse reflectance spectroscopy (DRS) in the visible and near-infrared (VNIR) wavelength ranges ($\sim$400-2500mm) is being used to obtain high-resolution soil information (Sudduth and Hummel, 1996; Sudduth et al., 2010). In particular, this later

technology offers a great deal of information through wavelength by depth maps (or images). It is of increasing interest to be able to use such information as covariates in spatial and spatio-temporal models because it is relatively inexpensive to obtain and mitigates the need for expensive and time-consuming laboratory analyses. The purpose of the analysis presented here is to ascertain whether functional/image covariates obtained from the VNIR soil spectra depth profiles can be useful in predicting functional (in depth) response curves through the methodology described in Section 4.2.

4.3.2 Greenly Research Center Data

We consider soil data collected at 28 sites as shown in Figure 4.1. In general, we are interested in predicting soil water content as a function of depth in terms of images corresponding to the VNIR soil spectra depth profiles. At each location, water content is measured at up to 48 depth segments as shown in Figure 4.2 and the VNIR spectra are measured at wavelengths from 500 to 2,500 nanometers (nm) for each depth segment as shown in Figure 4.3. We note that some locations have missing information at the deeper depth segments. To retain as many locations as possible for model building, we thus only consider the first 37 depth segments. Consequently, we have 25 locations with complete covariate and response information and these are used to fit the model in our analysis. Of the three discarded locations, two are missing most of the VNIR spectra information and the other is missing the water content response.

Specifically, the soil profile measurements we consider were collected at the University of Missouri Greenly Research Center near Novelty Missouri, USA (Lat. 40.0302°,

Lon. -92.188°) ; see Myers et al. (2011) for a comprehensive description. The exact locations sampled within the study site are shown in Figure 4.1. Soil cores (4.5 cm x 1.2 m) were obtained from each location within the study area for ex-situ measurements. These included diffuse reflectance spectra (500 x 2500 nm, FieldSpec Pro FR, ASD Inc., Boulder, CO), miniaturized Wenner array soil electrical conductivity (ECm), and gravimetric water content, all at 2.54 cm intervals along the core length. Soil profile electrical conductivity (ECp) and penetration resistance was measured in-situ using a Veris Profiler 3000 with an insulated shaft (Veris Technologies, Salina, KS, USA) and interpolated to 2.54 cm intervals. Both in-situ and ex-situ measurements were made on in the late spring of 2007, see Myers et al. (2010).

The analysis we consider can be viewed as an illustrative example, as it currently does not include any additional spatial covariates (e.g., elevation). However, such variables could easily be included as described in Section 4.2. Rather, our purpose here is to investigate the utility of using the spatially-dependent functional VNIR image covariates to predict water content as a function of depth.

4.3.3 Model Choices

The major model choices in this framework are concerned with the basis functions – both the form and the number of functions. We choose the Karhunen-Loéve decomposition for the image and water depth functions because of their efficiency in representing a large amount of variation using a relatively small number of functions. In fact, this is the basis for functional principal components analysis, which is common in functional data analysis (Ramsay and Silverman, 2005). For both the image and depth functions, we are not interested in predicting at new depths or wavelengths, so

that such an emprically-based basis is sufficient. Thus, in practice we can calculate eigenvalues and eigenfunctions from the symmetric decomposition of the empirical covariance matrix of the water content profiles as well as the depth/wavelength VNIR images. The eigenfunctions obtained from this decomposition are called empirical orthogonal functions (EOFs), as is common in spatial and spatio-temporal statistics (see Cressie and Wikle, 2011). In this example, the first 10, 15 and 20 EOFs associated with the image covariates account for approximately 94.0%, 97.3% and 99.2% of the total variation, respectively. Additionally, the first 10, 15 and 20 EOFs of the water content depth profiles account for approximately 98.7%, 99.6% and 99.9% of the total variation, respectively. To facilitate the SSVS implementation, we thus select the first 20 EOFs for the image covariates because this is a reasonable number of functions over which to conduct SSVS and it accounts for over 99% of the variation. In addition, we consider the maximum 24 EOFs associated with the water content depth functions (we lose one degree of freedom due to subtracting the mean) so that we retain all of the variation in a relatively small number of coefficients. A sensitivity analysis with 10 EOFs showed very similar results. Note, in this case, we recognize that we do not have to project the functional responses onto a basis set, but the methodology is designed for more general (high-dimensional) applications, in which such a decomposition would be necessary.

For the spatial basis functions, we consider two-dimensional thin plate splines. This gives the flexibility to predict at locations for which we do not have observations, which is a central purpose of the methodology. We select 20 knot locations based on the recommendation in Ruppert et al. (2003). In particular we use a space filling design (Nychka and Saltzman, 1998) to select the knots as shown in Figure 4.4.

4.3.4 Bayesian Implementation

In all cases, the basis functions are normalized by the square root of their respective eigenvalues. Thus, it is reasonable to assume $\Sigma_{n_\ell} = \sigma_{n_\ell}^2 I_{n_\ell}$ and $\Sigma_{n_i} = \sigma_{n_i}^2 I_{n_i}$. We then give inverse gamma priors for $\sigma_{n_\ell}^2 \sim IG(\kappa_{n_\ell}, \nu_{n_\ell})$ and $\sigma_{n_i}^2 \sim IG(\kappa_{n_i}, \nu_{n_i})$ and specify κ_{n_ℓ}, ν_{n_ℓ}, κ_{n_i}, and ν_{n_i} equal to one. In such settings, the prior distributions are vague (i.e. non-informative relative to the scale of the data) so that they impart little impact on the analysis.

Choices of τ and c need to be made to implement the SSVS algorithm. To find suitable values, we consider various possible combinations for these hyperparameters by way of a cross-validation sensitivity study (e.g., see Holan et al., 2010). Specifically, with $\pi_q \equiv 0.5$, we considered the following combinations of (τ, c): $(0.1, 10)$, $(0.01, 100)$, $(0.01, 10)$, $(0.01, 1)$, $(0.001, 100)$, $(0.001, 10)$, $(0.001, 1)$, and $(0.0001, 100)$. The SSVS algorithm results are based on 30,000 iterations with a 5,000 iteration burn-in.

To evaluate model performance, we considered mean square prediction errors (MPSE) using a leave-one-out cross-validation experiment (across all of the 25 locations). This was calculated based on the observed depths associated with the water content functional profiles.

4.3.5 Results

Table 4.1 shows that SSVS parameters $\tau = 0.1$ and $c = 10$ give superior performance in terms of cross-validation MPSE. We use this parameterization for the results presented herein. Figure 4.5 shows the in-sample fits of the water content depth profiles given the VNIR image predictors. Specifically, panel (a) shows the posterior mean water content depth profile for each of the 25 observation locations, panel (b) shows

106

the associated posterior standard deviation, panel (c) the observed water content depth profiles, and panel (d) the associated residuals (posterior mean minus the observed). Note that many of the features at the upper and mid-depths are captured by the model, but there are distinct regions of bias; e.g., the depths between increments 1 and 7 and depths from increment 22–30 show posterior means that are too large relative to the observations, while increments from 9–12 and 32–37 show means that are too small relative to the observations.

We also consider a true out-of-sample prediction at the location that had missing information for the water content depth profile response (i.e., the blue point in Figure 4.1). Figure 4.6 shows the posterior mean prediction of the depth functional for this location, along with the pointwise 95% credible intervals. We are also interested in predicting spatial fields of the response variable at various depths. Figure 4.7 shows a prediction grid over the spatial domain of interest, along with the observation locations. The posterior predicted mean and standard deviation on this grid for water content at depth level 15 are shown in Figure 4.8. Note that the spatial field is relatively smooth as expected based on the spline basis representation. Also, note that the most uncertainty occurs generally in the southwest portion of the domain. Indeed, there is a relatively large portion of that area for which the posterior mean is predicted to be zero (or less than zero) – clearly indicative of the uncertainty associated with this relatively data poor area of the domain.

4.4 Discussion and Conclusion

Scientists are increasingly faced with very large data signals from new technologies and they are interested in relating these "big data" signals to various types of responses. We consider here a particular case where we have functional responses and so-called "image" predictors. In this case, we use the term image loosely, and consider it to be any two-dimensional continuous process, such as a time-frequency representation of a time signal or a depth-wavelength representation of a spectroscopic profile. We consider here the additional complication that both the functional response and the image covariates can be spatially-dependent. This brings together several areas of research in functional data analysis, spatial statistics, and nonstationary time series analysis.

We develop a flexible, yet fairly easy to implement, methodology for the aforementioned problem by considering several layers of basis expansions. In practice, these expansions are truncated, leading effectively to a complex Bayesian mixed-effects multivariate multiple regression model. The major complication with implementation is the potentially large number of covariates (even in the basis expansion context), which is mitigated by the use of stochastic search variable selection priors on the parameters.

We demonstrate via an illustrative example that this methodology is useful in the context of complex soil profile data. In particular, we consider a water content depth profile response variable as predicted by depth/wavelength images of VNIR measurements. Although the model fit does show some apparent biases with depth, we have yet to include additional covariates that may be more related to the water content depth profiles. We were encouraged by the ability of the model to reproduce realistic depth profiles at spatial locations for which we did not have data.

Future extensions of this work include the consideration of additional covariates that are spatially dependent, including those that lack depth information. The methodology developed here includes such a term, but we have yet to implement it in practice. In addition, a real strength of this methodology will come from helping to suggest to practitioners where they should take additional (image) observations in order to most improve the predictions of the spatial distribution of the functional response. We will consider such optimal adaptive spatial sampling design problems as an extension to the work presented here.

Table 4.1: Top selected variables from the SSVS cross-validation. The percentages shown are the mean of the posterior probability of the variable being included in the model for each of the 25 cross-validation runs. MPSE is the mean square prediction errors associated with the water content depth profiles.

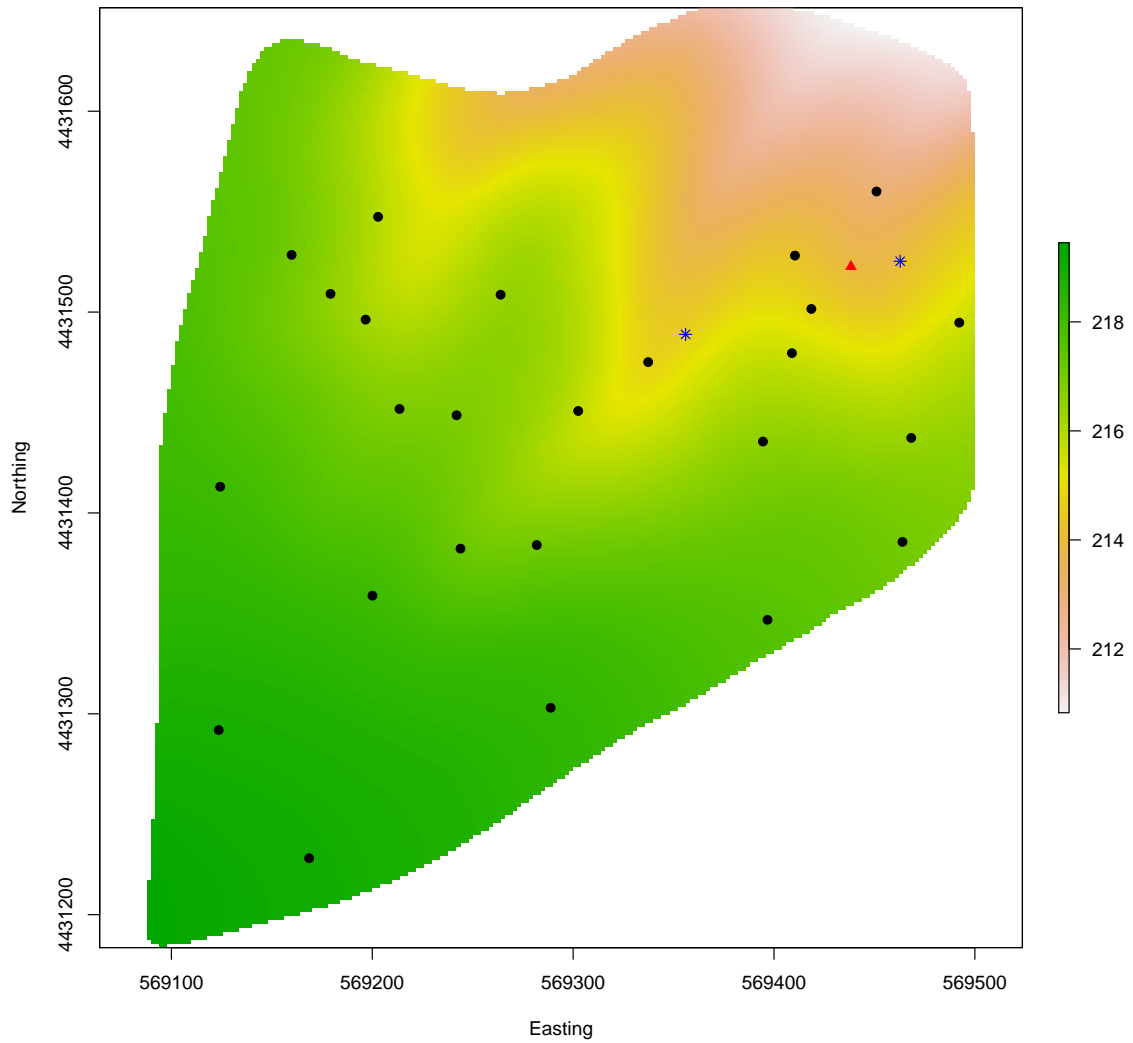| ($\tau = 0.1$, $c = 10$)		($\tau = 0.01$, $c = 100$)		($\tau = 0.01$, $c = 10$)		($\tau = 0.01$, $c = 1$)	
MPSE=0.0118		MPSE=0.0339		MPSE=0.0294		MPSE=0.0426	
Variable	% of models	Variable	% of models	Variable	% of models	Variable	% of models
19	97	20	95	11	78	13	50
14	95	15	88	20	71	18	50
15	93	14	88	15	61	10	50
16	93	11	83	13	52	11	50
20	83	19	80	19	51	15	50
17	73	13	78	14	50	1	50
13	69	18	76	9	50	19	50
9	58	7	61	16	48	3	50
12	56	17	58	17	48	5	50
18	54	16	56	12	46	14	50
11	50	12	42	7	43	9	49
7	47	9	32	18	42	2	49
10	13	8	19	8	41	4	49
8	5	6	17	6	38	16	49
5	0	5	13	10	38	7	49
($\tau = 0.001$, $c = 100$)		($\tau = 0.001$, $c = 10$)		($\tau = 0.001$, $c = 1$)		($\tau = 0.0001$, $c = 100$)	
MPSE=0.0351		MPSE=0.0432		MPSE=0.0733		MPSE=0.0442	
Variable	% of models	Variable	% of models	Variable	% of models	Variable	% of models
20	100	9	99	4	50	20	100
19	100	13	99	7	50	19	100
18	100	11	99	5	50	18	100
17	100	15	98	10	50	17	100
16	100	17	98	19	50	16	100
15	100	8	97	3	50	15	100
14	100	6	97	1	50	14	100
13	100	4	97	14	50	13	100
12	100	19	97	8	50	12	100
11	100	20	96	13	50	11	100
9	100	14	96	12	50	10	100
4	95	12	96	6	49	9	100
6	91	16	95	11	49	8	100
10	88	5	90	16	49	7	100
7	80	7	90	9	49	6	100

Figure 4.1: The elevation surface (measured in meters) of the study area. Black points are the 25 observation locations considered in the analysis; blue stars are locations with missing image covariate information, and the red triangle is the location with no water content response data.
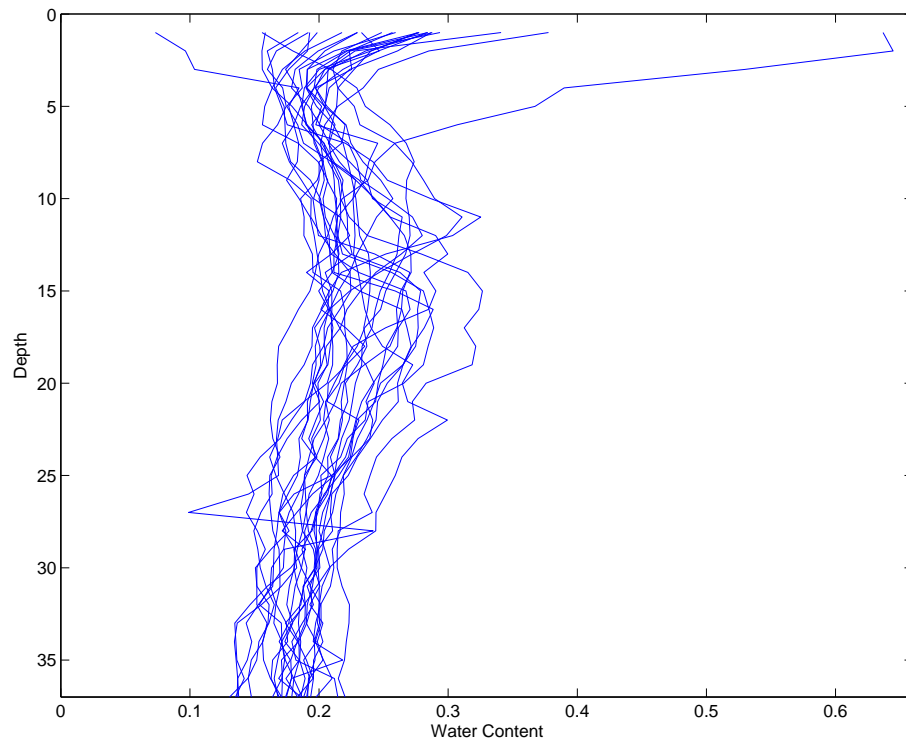
Figure 4.2: Water content (measured in percent by mass) as a function of depth as measured at the 25 locations shown in Figure 4.1.
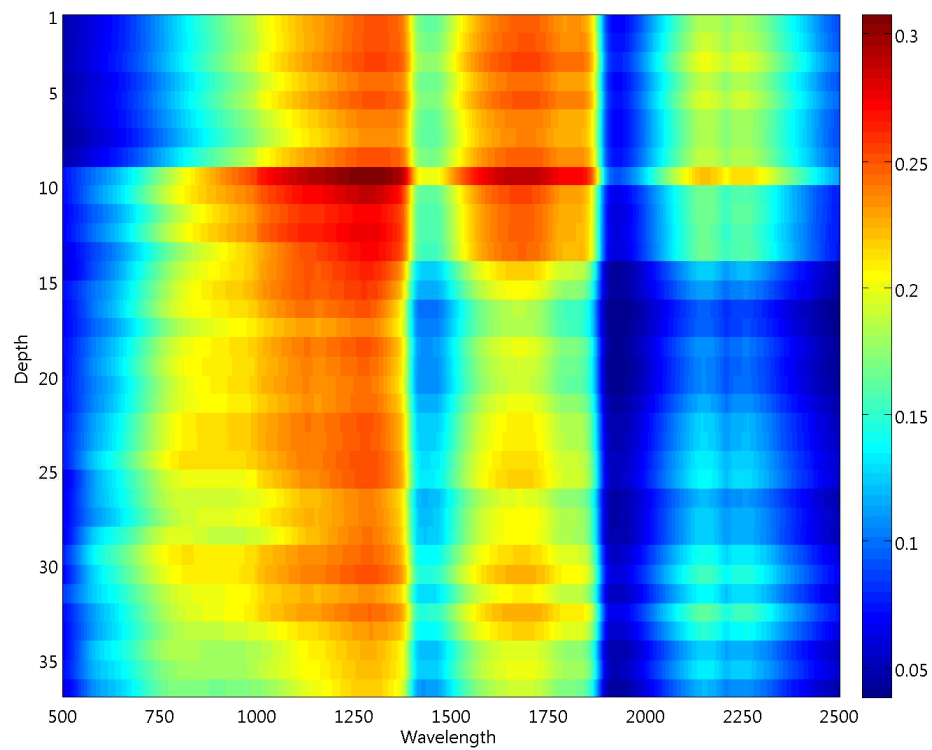
Figure 4.3: The VNIR covariate (measured in diffuse-reflectance) wavelength by depth image from a randomly selected data location. Wavelength is measured in nanometers whereas depth is measured in centimeter.
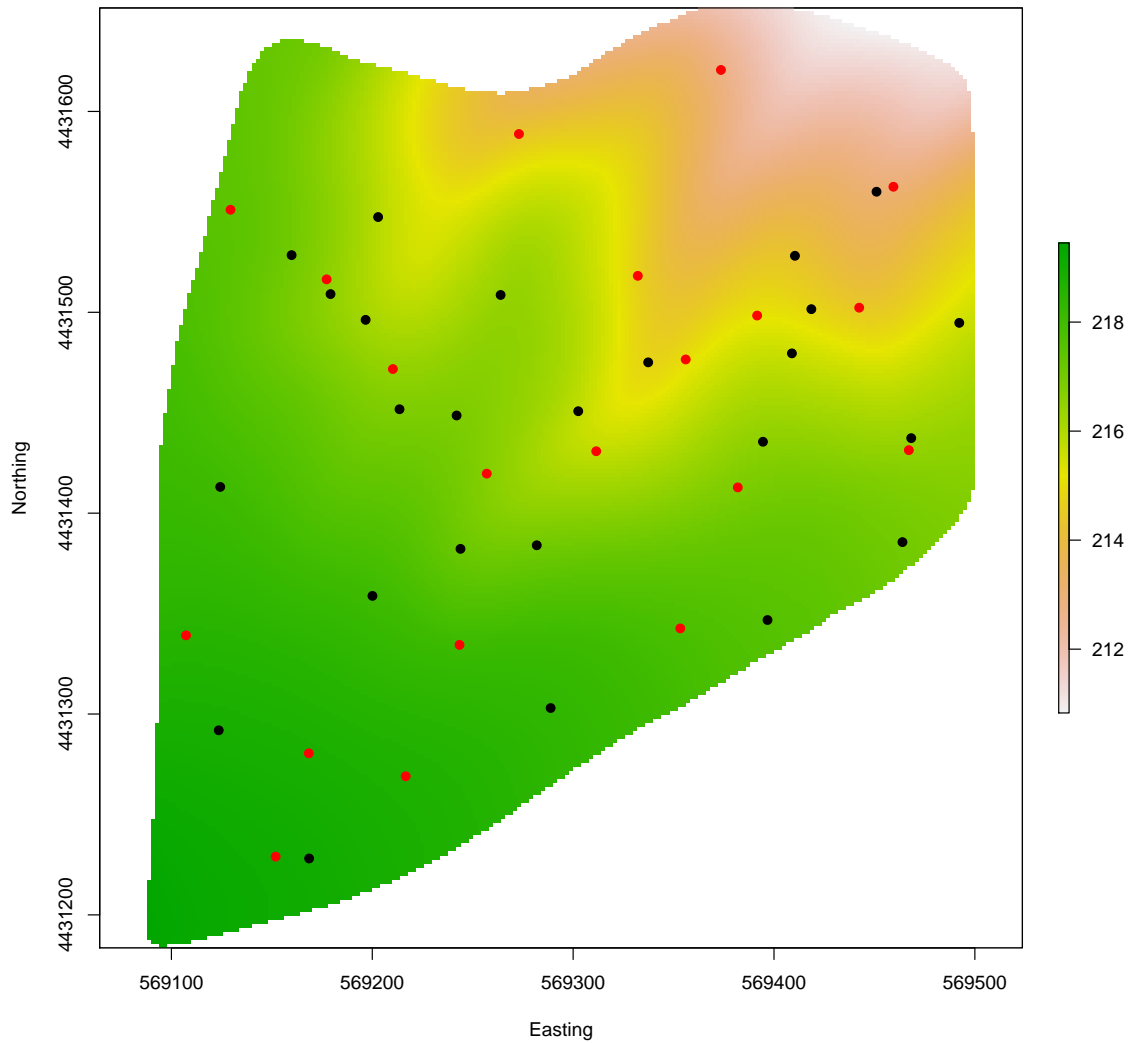
Figure 4.4: Elevation surface of the study area (measured in meters). Observation (black) and knots locations (red), as selected via a space-filling design.

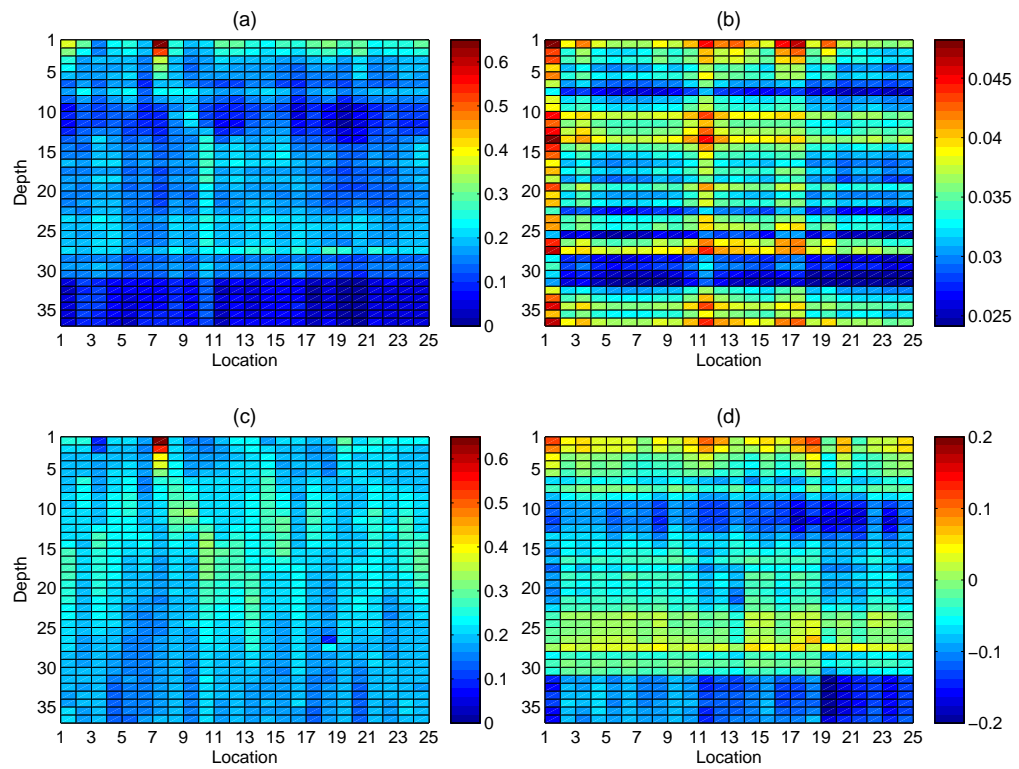Figure 4.5: In sample prediction of water content depth profiles for each of the 25 observation locations based on the $\tau = 0.1$, $c = 10$ stochastic search variable selection (SSVS) implementation. Panel (a) shows the posterior mean profile, (b) shows the posterior standard deviation, (c) shows the truth depth profile observations, and (d) shows the residuals between the posterior means and the observed depth profiles.
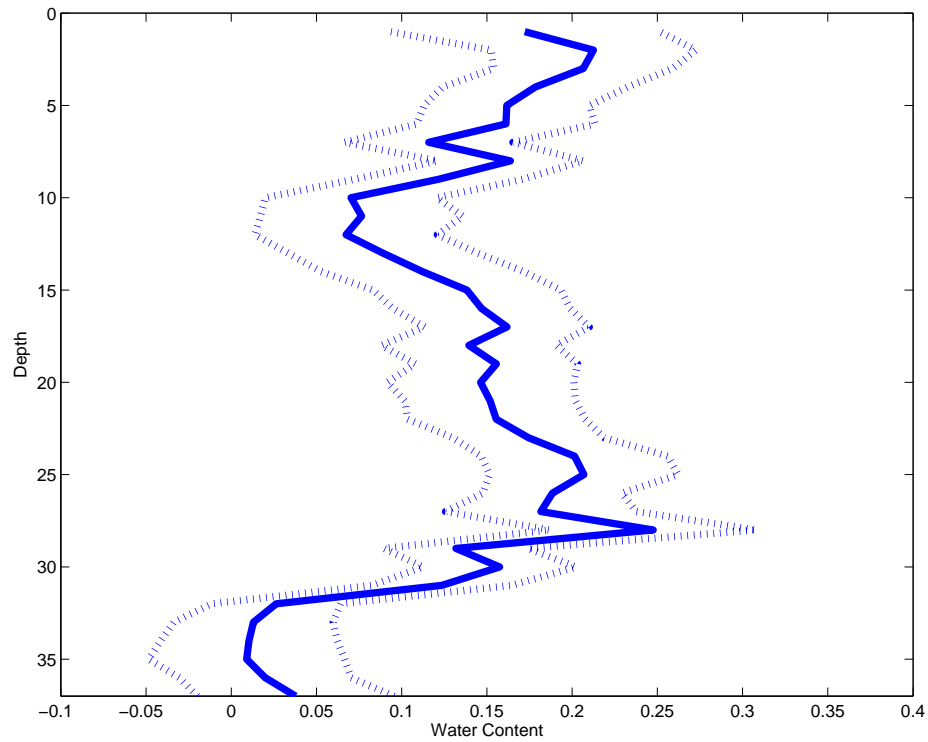
Figure 4.6: Prediction of water content depth profile (measured in percent by mass) of the location indicated in blue in Figure 4.1. The solid line refers to the posterior mean and the dotted lines show the pointwise posterior 95% credible intervals. Note that the water content is measured in percent by mass.

Figure 4.7: Elevation surface of the study area (measured in meters). Spatial prediction grid (red dots) and observation locations (black dots).

Figure 4.8: Predicted water content at depth segment 15 on the prediction grid shown in Figure 4.7. (a) Posterior mean (negative values are truncated at 0), (b) Posterior standard deviation.

Chapter 5

Conclusion and Discussion

5.1 Summary

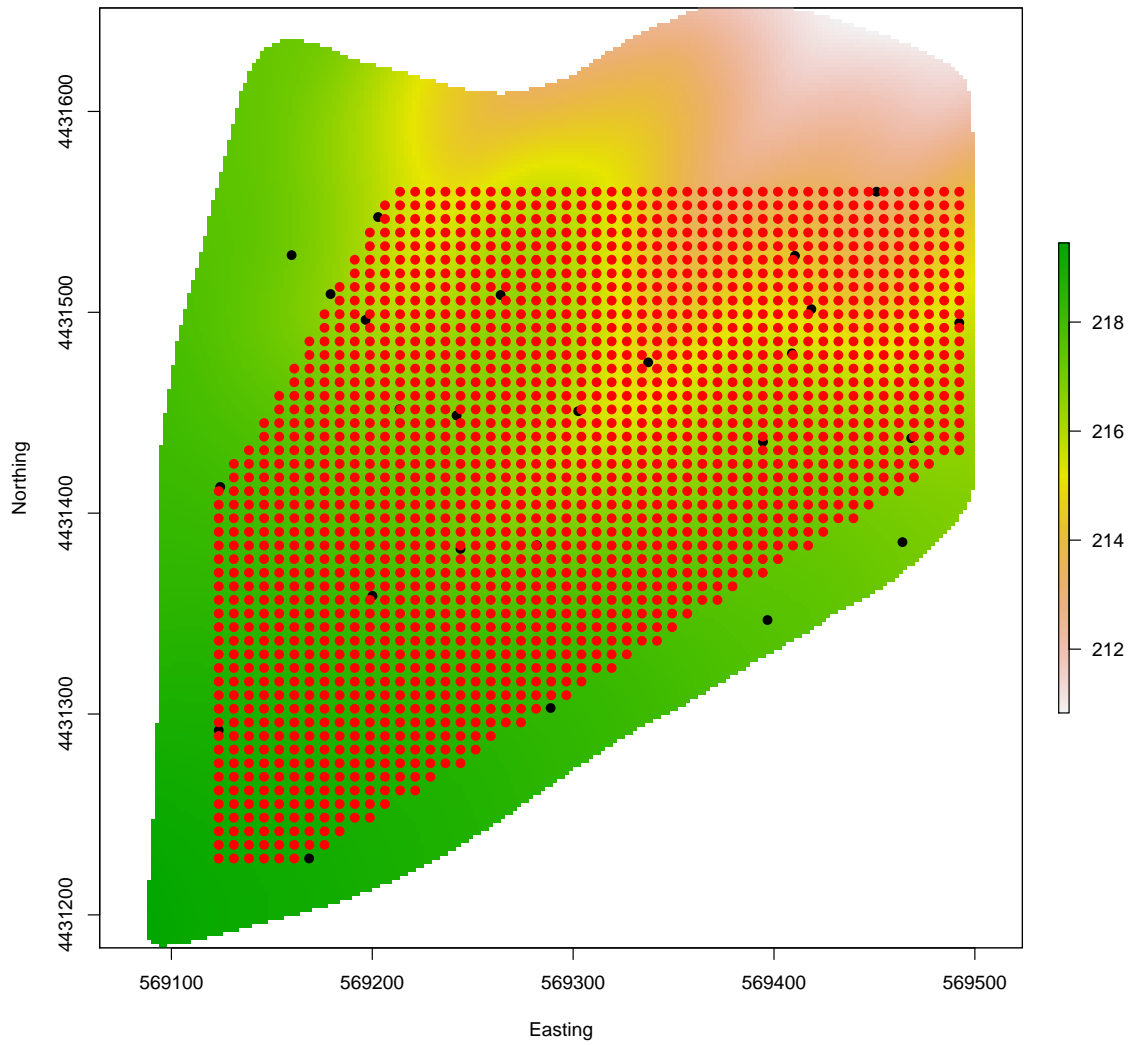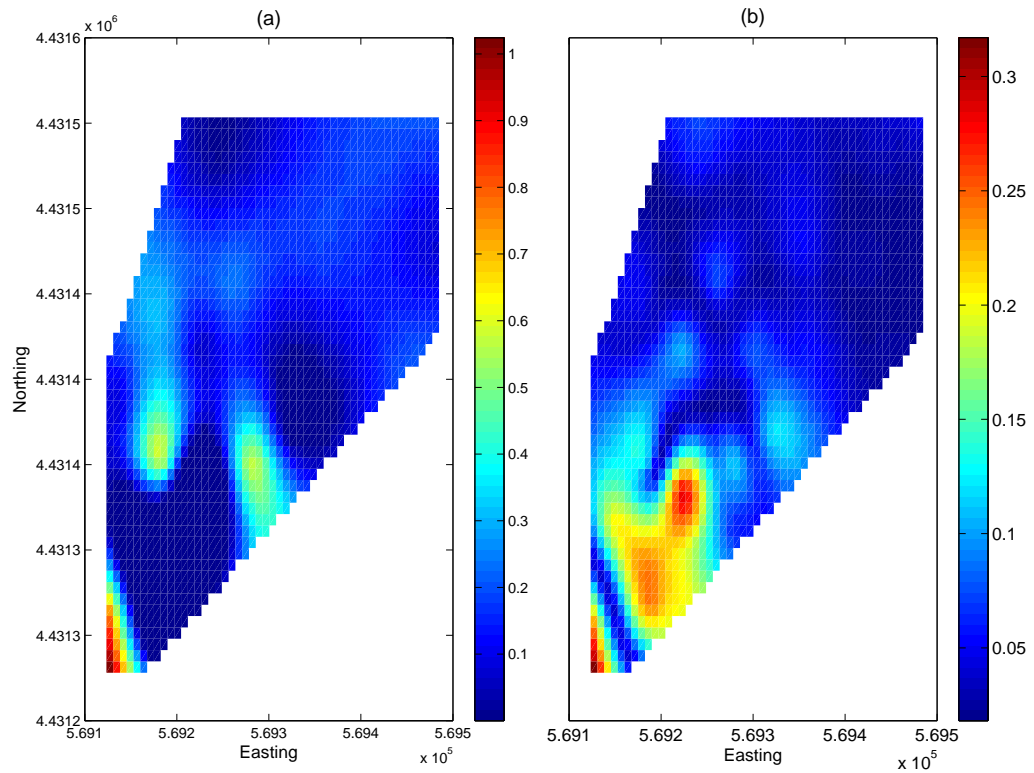When observed signals are nonstationary and sampled at high-frequency, it is often difficult to characterize important features. An efficient way to characterize such processes is in the time-frequency domain (e.g., through a spectrogram representation). Such representations can produce regions in time-frequency space that exhibit high variability.

However, utilizing a spectrogram representation usually comes with a substantially large number of time-frequency "pixels." In practice, it is not feasible to consider each of these pixels as a potential covariate. In addition to the large number of pixels, the time-frequency image often exhibits a high-degree spatial dependence. To reduce the inherent high-dimensionality, several orthogonal basis functions can be used (e.g., Fourier, splines, wavelets, and EOFs, among others). Some of them gather global

variation of a signal into low dimension feature space, whereas some gather local variation of a signal into feature space.

In many context, multiple signals and their interactions are necessary for addressing core scientific questions. These time-frequency predictors may relate to the response nonlinearly. Therefore, methodology for modeling complex processes using time-frequency functional predictors is needed that accurately quantifies uncertainty, accommodates multivariate time-frequency predictors and allows for nonlinear behavior. In Chapter 2, we illustrated time-frequency functional polynomial nonlinear (FPNL) models can account for signal uncertainty as well as accommodate nonlinear interactions through the motivating example of predicting spawning success of shovelnose sturgeon. Importantly, the approach produces feature extracted curves and images highlighting the important aspects of depth and temperature variability associated with successful spawning.

Different from nonparametric approach of time-frequency transform, model-based approaches provide potential ways to model signals while maintaining the time resolution. In Chapter 3, we provide model-based Bayesian lattice filter time-varying autoregressive models to estimate the time-frequency representation of nonstationary signals. In the simulation study, we compare our approach with other competing methods and the study shows that our approach performs superior in most simulation cases.

Chapter 4 provide a class of spatially-dependent functional models when both the response and covariates are functionals. Such models consider several layers of basis expansions. In practice, these expansions are truncated, effectively leading to a complex Bayesian mixed-effects multivariate multiple regression model. The

truncations provide sufficient dimensional reduction for both the functional response and covariates. Moreover, the stochastic search variable selection priors are used to select the important covariates in low-rank space associated with the response variable.

In many cases, time-frequency functional models have the ability to resolve complex scientific inquiries. Nevertheless, more research still needs to be conducted. For example, although the FPNL models allow multivariate signals, the dependence of signals may affect the inferences. To decrease such an affect, instead of decomposing various signals one by one, we may consider techniques to decompose them together. In addition, we can extend the Bayesian lattice filter time-varying autoregressive models to the multivariate context.

Bibliography

Albert, J. and Chib, S. "Bayesian analysis of binary and polychotomous response data." *Journal of the American Statistical Association*, 88(422):669–679 (1993).

Ansari-Asl, K., Bellanger, J., Bartolomei, F., Wendling, F., and Senhadji, L. "Time-frequency characterization of interdependencies in nonstationary signals: application to epileptic EEG." *Biomedical Engineering, IEEE Transactions on*, 52(7):1218–1226 (2005).

Baladandayuthapani, V., Mallick, B., Young Hong, M., Lupton, J., Turner, N., and Carroll, R. "Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis." *Biometrics*, 64(1):64–73 (2008).

Bosq, D. *Linear Processes in Function Spaces: Theory and Applications*, volume 149. Springer (2000).

Brown, P., Fearn, T., and Vannucci, M. "Bayesian wavelet regression on curves with application to a spectroscopic calibration problem." *Journal of the American Statistical Association*, 96(454):398–408 (2001).

Brown, P. and Vannucci, M. "Multivariate Bayesian variable selection and pre-

diction." *Journal of the Royal Statistical Society: Series B (Methodological)*, 60(3):627–641 (1998).

Chipman, H. "Bayesian variable selection with related predictors." *The Canadian Journal of Statistics*, 24(1):17–36 (1996).

Cocroft, R. and McNett, G. "Vibrational communication in treehoppers (Hemiptera: Membracidae)." In *Insect Sounds and Communication: Physiology, Ecology and Evolution*, S. Drosopoulos (Eds.), 305–317. Taylor & Francis (2006).

Crainiceanu, C., Staicu, A., and Di, C. "Generalized multilevel functional regression." *Journal of the American Statistical Association*, 104(488):1550–1561 (2009).

Cranstoun, S., Ombao, H., von Sachs, R., Guo, W., and Litt, B. "Time-frequency spectral estimation of multichannel EEG using the Auto-SLEX method." *Biomedical Engineering, IEEE Transactions on*, 49(9):988–996 (2002).

Cressie, N. and Wikle, C. *Statistics for Spatio-Temporal Data*. John Wiley and Sons (2011).

Davis, R., Lee, T., and Rodriguez-Yang, G. "Structural break estimation for non-stationary time series models." *Journal of the American Statistical Association*, 101(473):223–239 (2006).

Dawid, A. "Some matrix-variate distribution theory: notational considerations and a Bayesian application." *Biometrika*, 68:265–274 (1981).

Delicado, P., Giraldo, R., Comas, C., and Mateu, J. "Statistics for spatial functional data: some recent contributions." *Environmetrics*, 21(3-4):224–239 (2010).

DeLonay, A., Papoulias, D., Wildhaber, M., Annis, M., Bryan, J., Griffith, S., Holan, S., and Tillit, D. "Use of behavioral and physiological indicators to evaluate Scaphirhynchus sturgeon spawning success." *Journal of Applied Ichthyology*, 23:428–435 (2007).

Draper, D. "Assessment and propagation of model uncertainty (with discussion)." *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):45–97 (1995).

Feichtinger, H. and Strohmer, T. *Gabor Analysis and Algorithms: Theory and Applications*. Birkhauser (1998).

Fitzgerald, W., Smith, R., Walden, A., and Young, P. *Nonlinear and Nonstationary Signal Processing*. Cambridge University Press (2000).

Funk, J. and Robinson, J. *Changes in the Channel of the Lower Missouri River and Effects on Fish and Wildlife*. Missouri Department of Conservation, Jefferson City (1974).

Galat, D. and Lipkin, R. "Restoring ecological integrity of great rivers: historical hydrographs aid in defining reference conditions for the Missouri River." *Hydrobiologia*, 422:29–48 (2000).

Gelfand, A. E. and Ghosh, S. K. "Model choice: A minimum posterior predictive loss approach." *Biometrika*, 85(1):1–11 (1998).

George, E. "The variable selection problem." *Journal of the American Statistical Association*, 95(452):1304–1308 (2000).

George, E. and McCulloch, R. "Variable selection via Gibbs sampling." *Journal of the American Statistical Association*, 88(423):881–889 (1993).

—. "Approaches for Bayesian variable selection." *Statistica Sinica*, 7:339–374 (1997).

Geweke, J. "Variable selection and model comparison in regression." In *Bayesian Statistics 5*, J. M. Bernardo, J. O. Berger, A. P. Dawid, A. F. M. Smith (Eds.), 609–620. Oxford University Press (1996).

Giraldo, R., Delicado, P., and Mateu, J. "Continuous time-varying kriging for spatial prediction of functional data: An environmental application." *Journal of Agricultural, Biological, and Environmental Statistics*, 15(1):66–82 (2010).

—. "Hierarchical clustering of spatially correlated functional data." *Statistica Neerlandica*, 66(4):403–421 (2012).

Godsill, S., Doucet, A., and West, M. "Monte Carlo smoothing for nonlinear time series." *Journal of the American Statistical Association*, 99(465):156–168 (2004).

Goulard, M. and Voltz, M. "Geostatistical interpolation of curves: a case study in soil science." In *Geostatistics Tróia92*, 805–816. Springer (1993).

Gröchenig, K. *Foundations of Time-Frequency Analysis*. Birkhauser (2001).

Gromenko, O. and Kokoszka, P. "Nonparametric inference in small data sets of spatially indexed curves with application to ionospheric trend determination." *Computational Statistics & Data Analysis*, 59:82–94 (2013).

Gromenko, O., Kokoszka, P., Zhu, L., and Sojka, J. "Estimation and testing for

spatially indexed curves with application to ionospheric and magnetic field trends." *The Annals of Applied Statistics*, 6(2):669–696 (2012).

Hayes, M. *Statistical Digital Signal Processing and Modeling* . John Wiley & Sons (1996).

Haykin, S. *Adaptive Filter Theory (4th ed).* Pearson (2002).

Hesse, L. and Sheets, W. "The Missouri River hydrosystem." *Fisheries*, 18(5):5–14 (1993).

Hoeting, J., Madigan, D., Rafferty, A., and Volinsky, C. "Bayesian model averaging: a tutorial (with discussion)." *Statistical Science*, 14(4):382–417 (1999).

Holan, S., Davis, G., Wildhaber, M., DeLonay, A., and Papoulias, D. "Hierarchical Bayesian Markov switching models with application to predicting spawning success of shovelnose sturgeon." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58(1):47–64 (2009).

Holan, S., Wikle, C., Sullivan-Beckers, L., and Cocroft, R. "Modeling complex phenotypes: generalized linear models using spectrogram predictors of animal communication signals." *Biometrics*, 66(3):914–924 (2010).

Holan, S., Yang, W., Matteson, D., and Wikle, C. "An approach for identifying and predicting economic recessions in real-time using time-frequency functional models." *Applied Stochastic Models in Business and Industry*, 28:485–499 (2012).

Hosmer, D. and Lemeshow, S. *Applied Logistic Regression*. John Wiley & Sons (2000).

James, G. "Generalized linear models with functional predictors." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):411–432 (2002).

Jolliffe, I. *Principal Component Analysis*. Springer (2010).

Kestin, T., Karoly, D., Yano, J., and Rayner, N. "Time-frequency variability of ENSO and stochastic simulations." *Journal of Climate*, 11(9):2258–2272 (1998).

Kitagawa, G. *Introduction to Time Series Modeling*. Chapman & Hall (2009).

Kitagawa, G. and Gersch, W. *Smoothness Priors Analysis of Time Series*. Springer (1996).

Kokoszka, P. "Dependent functional data." *ISRN Probability and Statistics*, 2012 (2012).

Li, F. and Zhang, N. R. "Bayesian variable selection in structured high-dimensional covariate spaces with applications in genomics." *Journal of the American Statistical Association*, 105(491) (2010).

Mallat, S. *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press (2008).

Martinez, J. G., Bohn, K. M., Carroll, R. J., and Morris, J. S. "A Study of Mexican Free-Tailed Bat Chirp Syllables: Bayesian Functional Mixed Models for Nonstationary Acoustic Time Series." *UT MD Anderson Cancer Center Department of Biostatistics Working Paper Series. Working Paper 79* (2013).

Monestiez, P. and Nerini, D. "A cokriging method for spatial functional data with

applications in oceanology." In *Functional and Operatorial Statistics*, 237–242. Springer (2008).

Morris, J., Baladandayuthapani, V., Herrick, R., Sanna, P., and Gutstein, H. "Automated analysis of quantitative image data using isomorphic functional mixed models, with application to proteomics data." *Annals of Applied Statistics*, 5(2A) (2011).

Müller, H. and Stadtmüller, U. "Generalized functional linear models." *Annals of Statistics*, 33(2):774–805 (2005).

Nychka, D. and Saltzman, N. "Design of air-quality monitoring networks." *Lecture Notes in Statistics*, 51–76 (1998).

O'Hara, R. and Sillanpää, M. "A review of Bayesian variable selection methods: what, how and which." *Bayesian Analysis*, 4(1):85–118 (2009).

Ombao, H., Raz, J., Von Sachs, R., and Guo, W. "The SLEX model of a nonstationary random process." *Annals of the Institute of Statistical Mathematics*, 54(1):171–200 (2002).

Ombao, H., Raz, J., Von Sachs, R., and Malow, B. "Automatic statistical analysis of bivariate nonstationary time series." *Annals of the Institute of Statistical Mathematics*, 96(454):543–560 (2001).

Ombao, H., Von Sachs, R., and Guo, W. "SLEX analysis of multivariate nonstationary time series." *Journal of the American Statistical Association*, 100(470):519–531 (2005).

Oppenheim, A. and Schafer, R. *Discrete-Time Signal Processing*. Prentice Hall Signal Processing (2009).

Percival, D. and Walden, A. *Wavelet Methods for Time Series Analysis*. Cambridge University Press (2000).

Prado, R. and Huerta, G. "Time-varying autogressions with model order uncertainty." *Journal of Time Series*, 23(5):599–618 (2002).

Prado, R. and West, M. *Time Series Modeling, Inference and Forecasting*. Chapman & Hall/CRC (2010).

Priestley, M. *Spectral Analysis and Time Series*. Academic Press London (1981).

Qin, L., Guo, W., and Litt, B. "A time-frequency functional model for locally stationary time series data." *Journal of Computational and Graphical Statistics*, 18(3):675–693 (2009).

Ramsay, J. O. and Silverman, B. W. *Functional Data Analysis: 2nd ed*. Springer-Verlag (2005).

Reiss, P. and Ogden, R. "Functional generalized linear models with images as predictors." *Biometrics*, 66(1):61–69 (2010).

Rencher, A. *Methods of Multivariate Analysis: 2nd ed* . John Wiley and Son (2002).

Rosen, O., Stoffer, D., and Wood, S. "Local spectral analysis via a Bayesian mixture of smoothing splines." *Journal of the American Statistical Association*, 104(485):249–262 (2009).

Rosen, O., Wood, S., and Stoffer, D. "AdaptSPEC: Adaptive spectral estimation for nonstationary time series." *Journal of the American Statistical Association*, 107(500):1575–1589 (2012).

Ruiz-Medina, M. "New challenges in spatial and spatiotemporal functional statistics for high-dimensional data." *Spatial Statistics*, 1:82–91 (2012a).

—. "Spatial functional prediction from spatial autoregressive Hilbertian processes." *Environmetrics*, 23(1):119–128 (2012b).

Ruiz-Medina, M. and Montes, R. "Incorporating Spatial Interaction between Large Dimensional Temperature Series in Atmosphere–Ocean Modelling of Global Climate Change." *Procedia Environmental Sciences*, 7:2–7 (2011).

Ruiz-Medina, M. D. "Spatial autoregressive and moving average Hilbertian processes." *Journal of Multivariate Analysis*, 102(2):292–305 (2011).

Ruiz-Medina, M. D. and Espejo, R. "Integration of spatial functional interaction in the extrapolation of ocean surface temperature anomalies due to global warming." *International Journal of Applied Earth Observation and Geoinformation*, 22:27–39 (2013).

Ruppert, D., Wand, M., and Carroll, R. *Semiparametric regression*. Cambridge Univ Pr (2003).

Shumway, R. and Stoffer, D. *Time Series Analysis and Its Applications: 2nd ed*. Springer Verlag (2006).

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. "Bayesian

measures of model complexity and fit." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639 (2002).

Stingo, F., Vannucci, M., and Downey, G. "Bayesian wavelet-based curve classification via distribution analysis with markov random tree priors." *Statistica Sinica*, 22:465–488 (2012).

Sudduth, K. and Hummel, J. "Geographic operating range evaluation of a NIR soil sensor." *Transactions of the ASAE*, 39(5):1599–1604 (1996).

Sudduth, K., Kitchen, N., Sadler, E., Drummond, S., and Myers, D. "VNIR spectroscopy estimates of within-field variability in soil properties." In *Proximal Soil Sensing*, 153–163. Springer (2010).

U.S. Fish and Wildlife Service. *Biological Opinion on the Operation of the Missouri River Main Stem Reservoir System, Operation and Maintenance of the Missouri River Bank Stabilization and Navigation Project, and Operation of the Kansas River Reservoir System*. US Fish and Wildlife Service, Bismarck (2000).

Vannucci, M. and Stingo, F. "Bayesian models for variable selection that incorporate biological information." In *Bayesian Statistics 9*, J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West (Eds.). Oxford University Press (2010).

Vidakovic, B. *Statistical Modeling by Wavelets*. Wiley Online Library (1999).

Wahba, G. "Bayesian confidence intervals for the cross-validated smoothing spline." *Journal of the Royal Statistical Society: Series B (Methodological)*, 45(1):133–150 (1983).

West, M. and Harrison, J. *Bayesian Forecasting and Dynamic Models (2nd Ed.)*. Springer (1997).

West, M., Prado, R., and Krystal, A. "Evaluation and comparison of EEG traces: latent structure in nonstationary time series." *Journal of the American Statistical Association*, 94(446):375–387 (1999).

Wikle, C. "Low rank representations as models for spatial processes." In *Handbook of Spatial Statistics*, A. Gelfand, P. Diggle, M. Fuentes, P. Guttorp (Eds.), 107–118. Chapman and Hall/CRC (2010).

Wikle, C. and Cressie, N. "A dimension-reduced approach to space-time Kalman filtering." *Biometrika*, 86(4):815 (1999).

Wikle, C. and Holan, S. "Polynomial nonlinear spatio-temporal integro-difference equation models." *Journal of Time Series Analysis*, 32(4):339–350 (2011).

Wikle, C. and Hooten, M. "A general science-based framework for spatio-temporal dynamical models." *Test*, 19(3):417–451 (2010).

Wikle, C., Madden, R., and Chen, T.-C. "Seasonal variation of upper tropospheric lower stratospheric equatorial waves over the tropical Pacific." *Journal of the Atmospheric Sciences*, 54:1895–1909 (1997).

Wildhaber, M., DeLonay, A., Papoulias, D., Galat, D., Jacobson, R., Simpkins, D., Braaten, P., Korschegen, C., and Mac, M. "A conceptual life-history model for pallid and shovelnose sturgeon." Technical report, USGS Circular 1315 (2007).

—. "Identifying structural elements needed for development of a predictive life-history

model for pallid and shovelnose sturgeons." *Journal of Applied Ichthyology*, 27:462–469 (2011a).

Wildhaber, M., Holan, C., Davis, G., Gladish, D., DeLonay, A., Papoulias, D., and Sommerhauser, D. "Evaluating spawning migration patterns and predicting spawning success of shovelnose sturgeon in the Lower Missouri River." *Journal of Applied Ichthyology*, 27:301–308 (2011b).

Wolfe, P., Godsill, S., and Ng, W. "Bayesian variable selection and regularization for time–frequency surface estimation." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):575–589 (2004).

Wood, S., Rosen, O., and Kohn, R. "Bayesian mixtures of autoregressive models." *Journal of Computational and Graphical Statistics*, 20(1):174–195 (2011).

Yang, W., Wikle, C., Holan, S., and Wildhaber, M. "Ecological prediction with nonlinear multivariate time-frequency functional data models." *Journal of Agricultural, Biological, and Environmental Statistics*, In Press (2013).

Yao, F. and Müller, H. "Functional quadratic regression." *Biometrika*, 97(1):49–64 (2010).

Zhou, L., Huang, J. Z., Martinez, J. G., Maity, A., Baladandayuthapani, V., and Carroll, R. J. "Reduced rank mixed effects models for spatially correlated hierarchical functional data." *Journal of the American Statistical Association*, 105(489):390–400 (2010).

VITA

Wen-Hsi Yang was born in Miaoli, Taiwan, on November 14, 1978. He was raised in Miaoli, and graduated from Jian Tai Senior High School. He earned his B.S. from National Taipei University of Taiwan in Statistics in 2001 and, then M.S. in statistics at National Central University of Taiwan, in 2003. Upon completion of his Master degree, he worked as a research assistant at the Institute of Space Science in National Central University, Taiwan. Since 2008, Wen-Hsi entered the Ph.D. program in Statistics at the University of Missouri under the direction of Dr. Scott Holan and Dr. Christopher Wikle. He has accepted a postdoctoral position of Commonwealth Scientific and Industrial Research Organisation, Australia and will begin in August of 2013.