

ABSTRACT

Modeling Nonlinear, Nonstationary, Vector Time Series: Methods and Applications

Justin R. Sims, Ph.D.

Chairperson: Jane L. Harvill, Ph.D.

Methods for modeling nonlinear time series provide ways to extract and describe information from complex and dynamic processes. The class of nonlinear time series models is large. Rather than be exhaustive, we provide a review of two popular classes of nonlinear time series models: *Momentum threshold autoregressive* and *functional coefficient autoregressive models*. These models are then extended to vector time series. We illustrate utility by applying the models to real data examples in geology and photovoltaics, respectively.

The layers of speleothems (stalactites and stalagmites) hold information on ancient climates. Geologists hypothesize that the layers of a speleothem correspond to annual deposits, similar to tree rings. In these same layers, the ratios of carbon-13 isotopes and of oxygen-18 isotopes provide information on the types of vegetation, which in turn, gives information into the climate at the time that vegetation lived. We apply a *vector momentum threshold autoregressive model (VMTAR)* to the 3-dimensional series. We show a vast improvement over the linear vector autoregressive (VAR) model, both statistically and from a geological perspective, thus providing a useful tool for describing the climates during the late and middle Holocene periods.

Assessment of a utility scale photovoltaic (PV) power plant's potential performance is a critical aspect in the initial plant design and construction, and accurate

monitoring of plant efficiency is crucial to profitable plant operation. Both assessment and monitoring rely on temporally dense, but spatially sparse measurements of irradiance from sensors at the plant's location. We propose a sensor design algorithm to answer the question, "What is the optimal number and layout of sensors for predicting solar irradiance?" The algorithm makes use of *vector functional coefficient autoregressive (VFCAR) models* to determine if an optimal sensor design exists. To illustrate utility, we apply the algorithm to irradiance data collected from a 1.2 MW PV plant located in Lanai, Hawaii.

Modeling Nonlinear, Nonstationary, Vector Time Series: Methods and Applications

by

Justin R. Sims, B.S., M.S.

A Dissertation

Approved by the Department of Statistical Science

Jack D. Tubbs, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of
Doctor of Philosophy

Approved by the Dissertation Committee

Jane L. Harvill, Ph.D., Chairperson

Steven G. Driese, Ph.D.

Joon Jin Song, Ph.D.

Jack D. Tubbs, Ph.D.

Dean M. Young, Ph.D.

Accepted by the Graduate School
May 2017

J. Larry Lyon, Ph.D., Dean

Copyright © 2017 by Justin R. Sims
All rights reserved

TABLE OF CONTENTS

| | |
|-------------------------------------------------------|-----|
| LIST OF FIGURES | vii |
| LIST OF TABLES | ix |
| ACKNOWLEDGMENTS | x |
| DEDICATION | xi |
| 1 Overview | 1 |
| 1.1 Introduction | 1 |
| 1.2 Momentum Threshold Autoregressive Model | 1 |
| 1.2.1 Fitting of MTAR Models | 4 |
| 1.3 Functional Coefficient Autoregressive Model | 5 |
| 1.3.1 Fitting of FCAR Models | 6 |
| 1.3.2 Forecasting Methods of FCAR Models | 9 |
| 1.4 Overview of Dissertation | 11 |
| 2 Time Series Modeling of Ancient Climates | 12 |
| 2.1 Introduction | 12 |
| 2.2 Models | 13 |
| 2.2.1 Vector Autoregressive Models | 14 |
| 2.2.2 Vector Momentum Threshold Autoregressive Models | 14 |
| 2.3 Speleothem Data | 16 |
| 2.4 Results | 23 |
| 2.4.1 VAR Results | 23 |

| | | |
|-------|--------------------------------------------------------------------------------------|----|
| 2.4.2 | VMTAR Results | 25 |
| 2.4.3 | Geological Significance | 32 |
| 2.5 | Concluding Remarks | 35 |
| 3 | Semiparametric Approach to Optimal Sensor Design in a Photovoltaic Power Plant | 36 |
| 3.1 | Introduction | 36 |
| 3.2 | Methods | 37 |
| 3.2.1 | Functional Coefficient Autoregressive Model | 38 |
| 3.2.2 | Estimation of Functional Coefficient Matrices | 39 |
| 3.2.3 | Bootstrap Forecasts | 40 |
| 3.3 | Sensor Design Application | 41 |
| 3.4 | Application: La Ola PV Plant | 42 |
| 3.4.1 | Data | 43 |
| 3.4.2 | Results and Discussion | 46 |
| 3.5 | Concluding Remarks | 52 |
| 3.6 | Acknowledgment | 52 |
| | BIBLIOGRAPHY | 54 |

LIST OF FIGURES

| | | |
|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.1 | Map of the Raccoon Mountain Cave system. The blue dot represents the site at which speleothem RM0710-2-1 was collected. To the right of the map is a fluoresced and thorium-230 dated thin section of the stalagmite. | 18 |
| 2.2 | Images of the enlarged UVf photographs showing well preserved laminae, or rings. | 21 |
| 2.3 | Plots of number of rings (top), $\delta^{13}\text{C} \text{‰ PDB}$, (middle), and $\delta^{18}\text{O} \text{‰ PDB}$ (bottom). The asterisks (*) on the plot of number of rings represent data imputed using linear interpolation. | 22 |
| 2.4 | Plots of original data (—) and fitted values (- - -) from an order-one vector autoregressive model. The top plot is the number of rings per half-millimeter; the center plot is of $\delta^{13}\text{C} \text{‰ PDB}$; the bottom plot is of $\delta^{13}\text{C} \text{‰ PDB}$. The asterisks (*) on the plot of the number of rings represent data imputed using the fitted VAR(1) model. | 26 |
| 2.5 | Pointwise estimates of VFCAR(2) coefficient functions for $\delta^{13}\text{C} \text{‰ PDB}$. | 29 |
| 2.6 | Pointwise estimates of VFCAR(2) coefficient functions for $\delta^{18}\text{O} \text{‰ PDB}$. | 29 |
| 2.7 | Graphical output for grid search for two thresholds. The top plot is a time plot of the momentum in the number of UVf layers (the threshold variable) with threshold values ($r_1 = -4$ and $r_2 = 10$) represented via the horizontal lines. The middle graph is of the ordered momentum in the number of UVf layers with the threshold values labeled as *. The bottom graph contains the results of a grid search for the threshold values, with the pair chosen that resulted in the smallest sum of squared residuals (SSR) marked as solid triangles. | 31 |
| 2.8 | Plots of original series (—) and fitted series (- - -) from an order-two momentum threshold vector autoregressive model. The top plot is of $\delta^{13}\text{C} \text{‰ PDB}$; the bottom plot is of $\delta^{18}\text{O} \text{‰ PDB}$ | 34 |
| 3.1 | Coordinates and labels of the sixteen LiCor-200 pyranometers located at the midpoint of each tracking array and four additional locations for the La Ola PV plant in Lanai, Hawaii. | 44 |
| 3.2 | Time plots of 10-min averages (solid black) of estimated GHI irradiance measurements at a single sensor for Sensor 1 on March 8, 2010 (top plot) and March 10, 2010 (bottom plot). | 45 |

| | | |
|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 3.3 | Top graph is the time plot of 10-min averages (solid black) of irradiance measurements at a single sensor for March 8, 2010 with the local polynomial kernel estimate (dashed red) superimposed. The bottom plot is transformed irradiance (residuals after using local polynomial kernel regression to remove the diurnal trend). | 47 |
| 3.4 | Top graph is the time plot of 10-min averages (solid black) of irradiance measurements at a single sensor for March 10, 2010, with the local polynomial kernel estimate (dashed red) superimposed. The bottom plot is transformed irradiance (residuals after using local polynomial kernel regression to remove the diurnal trend). | 48 |
| 3.5 | Boxplots for the logarithm of the ratio of prediction errors across all days. The ratio is calculated as MPE of the VFCAR model divided by the MPE for the naive forecasting method. The left plot uses mean absolute prediction errors while the right plot uses mean squared prediction errors. | 51 |

LIST OF TABLES

| | | |
|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.1 | Estimated coefficients of VAR(1) model and approximate P -values (in parentheses) for tests of significance from zero. | 24 |
| 2.2 | Estimated mean square errors and estimated contemporaneous error correlations with approximate P -values for significance from zero in parentheses from order-one vector autoregressive model. | 25 |
| 2.3 | Significant VMTAR(2,3,2) coefficients with standard errors (in parentheses). Coefficients marked with an * are significantly zero at a 0.01 level of significance. Additionally, coefficients marked with an † are also significantly zero at a 0.05 level of significance. All other coefficients are not significantly different from zero. | 33 |
| 2.4 | Estimated mean square errors and contemporaneous estimated error correlations with approximate P -values for significance from zero in parentheses from order-two vector momentum threshold autoregressive model. | 33 |
| 3.1 | Sensor combinations and median of prediction error ratios across all days. The columns are the sensor combinations for sensor subsets of size $k = 2, 3$, and 4. The values in parenthesis are the median ratio of prediction errors across all days. The ratio is calculated as MPE of the VFCAR model divided by the MPE for the naive forecasting method. . . | 49 |
| 3.2 | Summary percentiles of absolute prediction error ratios for the five combinations of $k = 2$ sensors that produced the smallest median absolute prediction error ratios across all days. The ratio is calculated as MAPE of the VFCAR model divided by the MAPE for the naive forecasting method. | 50 |
| 3.3 | Summary percentiles of square prediction error ratios for the five combinations of $k = 2$ sensors that produced the smallest median square prediction error ratios across all days. The ratio is calculated as MSPE of the VFCAR model divided by the MSPE for the naive forecasting method. . . | 50 |

ACKNOWLEDGMENTS

Thank you to my dissertation committee members for your insight and for taking your valuable time to review this research. I am thankful for you, Dr. Harvill, for taking a chance and choosing to work with me. Your patience and guidance throughout this process have been a blessing. Thank you for your support both academically and personally.

I would also like to thank my professors and peers in the Department of Statistical Science at Baylor University. My years with you all have been an absolute pleasure. Thank you to Courtney, Kristen, Joyce, Forrest, Jack, Michelle, and Amy, for your friendship and collaboration.

Thank you to my colleagues at Francis Marion University, Nicole, Sharon, and Sophia. Thank you for your encouragement and always being there to listen.

Thank you, Barbara and Curtis, for taking me under your wings and helping me to grow up. I am forever grateful for you bringing the smile back to my face.

I am eternally grateful for you, Dr. McCullough, for your encouragement and supporting me on my worst of days. Thank you for instilling in me a passion for statistics early on. Most importantly, thank you for your love. You were never a surrogate, and I will always be proud to be your own.

Lastly, thank you to my parents. You knew the importance of an education and always made sure I had whatever I needed.

DEDICATION

To my mother

CHAPTER ONE

Overview

1.1 Introduction

In the early 1980s, the realization that linear time series models lacked the ability to accurately describe many complex dynamics spurred the development of nonlinear time series models. Examples of observable phenomena that can be explained by nonlinear dynamics include jump phenomenon, limit cycles, amplitude dependent frequencies, and time irreversibility. The class of nonlinear time series models is large and contains many popular parametric models including the bilinear model, the exponential autoregressive (EXPAR) model, the threshold autoregressive (TAR) model, the self-exciting threshold autoregressive (SETAR) model, Markov-switching models, and generalized autoregressive conditional heteroscedasticity (GARCH) models.

In this chapter, we present both a parametric and a semiparametric model for nonlinear time series and to adapt them to two different applications. More specifically, we shall present the momentum threshold autoregressive model as a variant of the TAR model. Also, we present the functional coefficient autoregressive model as a structure for reducing the size of the class of nonlinear models. The chapter is organized as follows. We discuss the literature on the momentum threshold autoregressive models in Section 1.2. In Section 1.3, we discuss the literature on functional coefficient autoregressive models. We then give an overview of Chapters Two and Three in Section 1.4.

1.2 Momentum Threshold Autoregressive Model

The class of threshold autoregressive (TAR) models, first introduced by Tong (1978, 1983) and Tong and Lim (1980), provides a framework for representing a process that can have multiple states. For each state, or regime, a different model is

used. The basic idea of a TAR model is to express the series through piecewise autoregressive linearization. Each of these different regimes possesses their own unique linear autoregressive dynamics. Let $\{X_t\}$ represent a zero-mean time series where the index set t is the set of integers. Then an ℓ -regime, order- p threshold autoregressive, TAR(p, ℓ), model is

$$X_t = \sum_{j=1}^p \phi_j^{(i)} X_{t-j} + \varepsilon_t^{(i)}, \quad \text{if } r_{i-1} < Z_t \leq r_i, \quad i = 1, \dots, \ell, \quad t = 1, 2, \dots, T, \quad (1.1)$$

where $\phi_j^{(i)}$, $i = 1, \dots, \ell$, is the lag j autoregressive (AR) coefficient for regime i , $\varepsilon_t^{(i)}$ are zero-mean white noise with standard deviation $\sigma^{(i)} < \infty$, and the series $\{Z_t\}$ is the thresholding time series. The thresholding series may be either the original or an exogenous time series. If $Z_t = X_{t-q}$, then q is the delay parameter where $q < p$ to avoid model unidentifiability. If Z_t is exogeneous to X_t , no delay parameter is necessary. Furthermore, $-\infty = r_0 < r_1 < \dots < r_{\ell-1} < r_\ell = \infty$ are the $\ell + 1$ non-trivial thresholds dividing the domain of $\{Z_t\}$ into ℓ different regimes. Equation (1.1) is the general form of a TAR model. It may be more beneficial to view the TAR model in its piecewise representation. For $t = 1, 2, \dots, T$,

$$X_t = \begin{cases} \sum_{j=1}^p \phi_j^{(1)} X_{t-j} + \varepsilon_t^{(1)} & Z_t \leq r_1 \\ \sum_{j=1}^p \phi_j^{(2)} X_{t-j} + \varepsilon_t^{(2)} & r_1 < Z_t \leq r_2 \\ \vdots & \vdots \\ \sum_{j=1}^p \phi_j^{(\ell)} X_{t-j} + \varepsilon_t^{(\ell)} & Z_t > r_{\ell-1}, \end{cases} \quad (1.2)$$

where the model description above still holds.

It is possible to extend the the univariate TAR model in several ways. When $Z_t = X_{t-q}$ the model is the well-studied self-exciting threshold autoregressive (SETAR) model of Tong and Lim (1980). De Gooijer (1998), Ling (1999), and Ling et al. (2007) all propose variations on threshold moving-average models. These proposals have have led to much applied and theoretical work in this model class. Tsay (1998)

extended the methodology of TAR model to a vector setting, including procedures of testing the linear hypothesis and identification of the delay and threshold parameters

In the basic TAR model, the regimes are is defined only by some thresholding series. An alternative specification allows for the threshold to depend on the magnitude of change, or momentum, in a previous period of a thresholding series. Proposed by Enders and Granger (1998) and Caner and Hansen (1998) then further developed by Enders and Siklos (2001), momentum threshold autoregressive (MTAR) models are adept at modeling asymmetric features. Let $\{Z_t\}$ be the (possibly exogenous) thresholding time series, and consider $\nabla Z_{t-q} = Z_{t-1} - Z_{t-q}$; that is, $\{\nabla Z_{t-q}\}$ is a time series that captures the q -th difference, or momentum, of the series $\{Z_t\}$. The ℓ -regime, order- p momentum threshold autoregressive, $\text{MTAR}(p, \ell, q)$, model for a zero mean time series $\{X_t\}$ is given by

$$X_t = \sum_{j=1}^p \phi_j^{(i)} X_{t-j} + \varepsilon_t^{(i)}, \text{ if } r_{i-1} < \nabla Z_{t-q} \leq r_i, \quad i = 1, \dots, \ell, \quad (1.3)$$

where $\phi_j^{(i)}$, $i = 1, 2, \dots, \ell$, is the lag j AR coefficient for regime i and $\varepsilon_t^{(i)}$ is zero mean white noise with standard deviation $\sigma^{(i)}$. Again, $-\infty = r_0 < r_1 < \dots < r_{\ell-1} < r_\ell = \infty$ are the $\ell + 1$ non-trivial thresholds dividing the domain of $\{\nabla Z_{t-q}\}$ into ℓ different regimes.

Real data applications using a MTAR model are primarily in the field of econometrics. Coakley et al. (2002) use a MTAR model to capture fast-up, slow-down dynamics in the unemployment rates of the United States, the United Kingdom and Germany during the 1970s and 1980s. Vincent and Morley (2012) investigate whether house prices in the United Kingdom react differently to negative and positive “shocks” (or changes) in housing market fundamentals. The authors’ results reveal that negative momentum in mortgage rates and personal income result in a convergence to a housing price equilibrium, but positive shocks result in a divergence from equilibrium. More recently, Pilatowska et al. (2014) use MTAR models to explore the long-run equilibrium relationship between per capita greenhouse gas emissions and

per capita real gross domestic product (GDP). The authors find strong evidence of environmental degradation in Poland from 2000-2012 for abrupt changes in income level.

1.2.1 Fitting of MTAR Models

The modeling procedure for a MTAR model is accomplished systematically. For determination of the autoregressive order p , Enders (2009) suggests the use of Akaike information criterion (AIC) or Bayesian information criterion (BIC). Alternatively, tests for white noise, such as those discussed in Tong (1983) or Granger and Teräsvirta (1993), can be used to determine the optimal AR order p . These methods use a sequence of hypotheses to test linearity against a threshold structure for various values of the AR order. If linearity is rejected for more than one value of p , the AR order producing the smallest P -value indicates the optimal p .

For selection of the number of regimes ℓ and the delay parameter q , Hansen (1997) and Tsay (1998) suggest the use of past experience and substantive information to aid in the choice of each. When this prior information is unknown, Tsay (1998) suggests that the computational complexity, especially for vector time series, restricts the number of regimes ℓ to a small value, typically two or three. The author also provided an estimation procedure of the delay parameter q using methods based on AIC as well as conventional testing methodologies such as χ^2 test statistics.

Chan (1993) and Enders and Siklos (2003) suggest a grid search algorithm for determining a super-consistent estimate of the threshold value(s). First the candidate threshold values are sorted in ascending order. In order to avoid disorganized dispersion of observations in each regime, the lowest and highest 15% of candidate threshold values are eliminated. Afterwards, one-regime MTAR models are estimated for each potential threshold values. Each sum of squared residuals of these models is considered as a function of the associated threshold value because each model is constructed according to the threshold value. Thus, any pattern of sum of squared

residuals indicates the presence of nonlinearity. In the absence of nonlinearity, there must be no relationship between threshold values and sum of squared residuals. However, if there is a threshold effect, sum of squared residuals becomes smaller as the threshold value in use is getting closer to the true threshold value. The model having the smallest sum of squared residuals will give a consistent estimate of the threshold value. If more than one threshold value is desired, the remaining candidate threshold values are further trimmed by 15% above and below the first threshold value. The algorithm is repeated using the estimated threshold value and the remaining potential values. If there is more than one threshold value, the sum of squared residuals will have the same number of local minima.

As previously discussed, a MTAR model constitutes of piecewise linear autoregressive model. Therefore, after detecting the threshold variable and the threshold value(s), the estimation procedure for $\phi_j^{(i)}$, the lag- j autoregressive coefficient in the i^{th} regime, is that of a linear autoregressive model estimation procedure. Estimations are performed regime-by-regime. Tong (1983) noted that because each group exhibits linear dynamics, simple ordinary least squares (OLS) techniques can be used for estimation of each piecewise linear AR model. Chan (1993) showed the consistency and the limiting distribution of the least squares estimators of a TAR under some regularity conditions.

1.3 Functional Coefficient Autoregressive Model

The functional coefficient autoregressive model of order p (FCAR(p)) is a flexible semiparametric model first introduced by Chen and Tsay (1993). The additive autoregressive structure of the FCAR(p) model allows the autoregressive coefficients to change as a function of lagged values of some possibly exogenous time series. Let X_t be a zero mean univariate time series observed at T time points. Define the

FCAR(p) model as

$$X_t = \sum_{j=1}^p f^{(j)}(Z_t)X_{t-j} + \varepsilon_t, \quad t = p+1, p+2, \dots, T, \quad (1.4)$$

where Z_t is the functional variable which may consist of exogenous predictors or lagged values of the series X_t , $f^{(j)}(\cdot)$, $j = 1, \dots, p$, are real-valued measurable functions that change as a function of Z_t and which have continuous second derivatives, and ε_t is zero mean white noise with variance $\sigma^2 < \infty$. If $Z_t = X_{t-q}$, then q is the delay parameter where $q < p$ to avoid model unidentifiability.

An advantage of the FCAR model is its versatility. As an example, if all functions are constant with respect to Z_t , the model in Eq. (1.4) reduces to the commonly used linear AR model of order p . In another example, the FCAR(p) model reduces to the EXPAR model of Haggan and Ozaki (1981) if for each $j = 1, 2, \dots, p$ the coefficients in Eq. (1.4) are of the form $f^{(j)}(Z_t) = \alpha_j + \beta_j \exp(-\delta Z_t^2)$. Additionally, the formulation of the FCAR model allow for a mixture of models, such as an FCAR(2) model with coefficients $f^{(1)}(Z_t) = \alpha_1$ and $f^{(2)}(Z_t) = \alpha_2 + \beta_2 \exp(-\delta Z_t^2)$. Estimation of the functional form may also provide an objective guideline for choosing a parametric model.

1.3.1 Fitting of FCAR Models

Chen and Tsay (1993) proposed a model fitting procedure that makes use of arranged local regression (ALR) when $Z_t = X_{t-q}$. To begin, the ALR procedure selects an interval length c to form a window and a minimum sample size K to control the number of observations in the window. Let $X_{(1)}, \dots, X_{(k)}$ be in the window $[X_{(1)}, X_{(1)} + c]$. Initialize the estimation procedure by fitting the linear regression

$$X_{t+s} = a_1 X_{t+s-1} + \dots + a_p X_{t+s-p} + \varepsilon_{t+s} \quad (1.5)$$

with $t = t_1, \dots, t_k$ where ε_{t+s} is the residual at time $t + q$. The OLS estimate of a_j is an estimate of $f^{(j)}(X_{(1)} + c)$ provided $k > K$. Note that $(X_{(i)} + c)$ denotes the right

end point of the window used. Move the window along the X_{t-d} axis until there is at least one new data point or a point drops out of the window. If the sample size in the new window is at least K , then refit a linear regression similar to the one in (1.5) to obtain an estimate $\hat{f}^{(j)}(X_{(i)}+c)$ of $f^{(j)}(X_{(i)}+c)$. Scatterplots of the estimates $\hat{f}^{(j)}(X)$ versus X can be used to infer the functional forms of $f^{(j)}(\cdot)$. The authors compared the FCAR model and the above model building procedure to linear and threshold autoregressive models using simulated data and the chickenpox data of Sugilara and May (1990). The FCAR model outperformed the other two models in terms of bias, but only performed better in short-term forecasts in terms of mean squared error (MSE).

Although Chen and Tsay (1993) did not use local regression techniques directly, their iterative recursive formula is reminiscent of local constant fitting. Cai, Fan, and Yao (2000) adapted local linear regression methods to estimate $f^{(j)}(\cdot)$ in (1.4). The authors further extended the model to allow the autoregressive coefficients to vary as function of more than one variable or possibly exogenous functional variables. In the local linear method, $f^{(j)}(\cdot)$ is approximated locally at z_0 by a linear function $f^{(j)}(Z_t) \approx a_j + b_j(Z_t - z_0)$. The local linear estimate is defined as $\hat{a}_j(z_0) = \hat{a}_j$ where $\left\{ \left(\hat{a}_j, \hat{b}_j \right) \right\}$ minimize

$$\sum_{t=1}^n \left[X_t - \sum_{j=1}^p \{a_j + b_j(Z_t - z_0)\} X_{t-j} \right]^2 K_h(Z_t - z_0), \quad (1.6)$$

where $K_h(\cdot) = h^{-1}K(\cdot/h)$, $K(\cdot)$ is a kernel function, and $h > 0$ is a bandwidth. It follows that for $\mathbf{X}_t = (X_{t-1}, \dots, X_{t-p})'$ the least squares solution is

$$\hat{f}^{(j)}(z_0) = \sum_{t=1}^n K_{n,j}(Z_t - z_0, \mathbf{X}_t) X_t, \quad (1.7)$$

where

$$K_{n,j}(z, \mathbf{x}) = \mathbf{e}'_{j,2p} \left(\tilde{\mathbf{X}}' \mathbf{W} \tilde{\mathbf{X}} \right)^{-1} \tilde{\mathbf{X}}' K_h(z). \quad (1.8)$$

In equation (1.8), $\mathbf{e}_{j,2p}$ is a $2p \times 1$ unit vector with 1 at the j th position, $\tilde{\mathbf{X}}$ denotes an $n \times 2p$ matrix with $(\mathbf{X}'_t, \mathbf{X}'_t(Z_t - z_0))$ as its t th row, and $\mathbf{W} = \text{diag}\{K_h(Z_1 - z_0), \dots, K_h(Z_n - z_0)\}$.

Asymptotic properties of the local linear estimator were established by Chen and Liu (2001). Furthermore, they construct pointwise confidence bands for the function estimates for a simulated EXPAR model. Their work shows that the local linear procedure is reasonably robust for estimating the functional coefficients in model (1.4).

More recently, estimation of nonlinear time series models has been accomplished using spline-backfitted kernel (SBK) methods. The SBK method was first proposed by Wang and Yang (2007) and is an adaptation of the backfitting algorithm of Hastie and Tibshirani. Lin and Yang (2010) adapt the SBK method for additive coefficient models. Patrick (2013) adapts the work of Lin and Yang (2010) to the FCAR model which is a generalized additive coefficient model.

For the FCAR model with functional variable Z_t , the SBK method finds pre-estimates $\hat{f}^{(j)}(Z_t)$ for $f^{(j)}(Z_t)$, $j = 1, 2, \dots, p$, using an under-smoothed spline procedure. These pre-estimates are then used to find pseudo-responses $\hat{X}_{t,j'}$ through a backfitting procedure given by

$$\hat{X}_{t,j'} = X_t - \sum_{1 \leq j \leq p, j \neq j'} \hat{f}^{(j)}(Z_t) X_j \quad (1.9)$$

These pseudo-responses are a proxy of the original realization, but with the effects of all $f^{(j)}(Z_t)$, $j \neq j'$, removed. Then the pseudo-responses are used to find the SBK estimate $\hat{f}^{(j')}(Z_t)$ of $f^{(j')}(Z_t)$ through a kernel estimator, e.g. the Nadaraya-Watson estimator. The procedure is then repeated using the newly obtained kernel estimate and the remaining pre-estimates. The idea behind SBK estimation is to under-smooth in the pre-estimates in order to reduce the bias. This under-smoothing leads to a larger variance which is reduced in the kernel estimation step.

1.3.2 Forecasting Methods of FCAR Models

Even in the parametric case, forecasting for nonlinear time series models is not necessarily straightforward. For a time series $\{X_t\}$, the assumed nonlinear structure of the process complicates the derivation of the expected value of X_{t+M} given X_t, \dots, X_1 when $M > 1$. When nonparametric or semiparametric methods are used to model $\{X_t\}$, this issue is compounded further. When $f^{(j)}(\cdot)$, $j = 1, 2, \dots, p$ is estimated using $Z_t = X_{t-q}$, the expectation of X_{t+M} given X_t, \dots, X_1 is no longer a simple linear operation. We now proceed to highlight some work in forecasting of nonlinear time series.

Early work of Clements and Smith (1997) investigate multi-step ahead forecasts for the SETAR model. Comparisons to forecasts from a linear AR model are made using Monte Carlo simulation and bootstrap methods assuming all parameters are unknown except for the lag order and the delay value. Through simulation results, the authors find that the Monte Carlo methods are at least as good as the bootstrap methods, but the latter is preferred due to Monte Carlo methods being computationally intensive.

Chen and Lui (1993) note that post-sample multi-step forecast accuracy can be substantially improved with the use of the FCAR model compared to other linear models and parametric nonlinear models. Similarly, Chen and Tsay (1993) obtain multi-step forecasts from an FCAR model fit using the ALR procedure described in Section 1.3.1. The authors of both paper note improvements without specifying how the multi-step ahead forecasts are obtained.

Cai, Fan, and Yao (2000) assess the post sample forecasting performance of the FCAR model fit via the local linear method using the Canadian lynx data set and Wolf's sunspot numbers data set (see e.g., Tong, 1990). They compare the local linear method with the linear AR model, the TAR model, and the ALR procedure using one-step ahead and an iterative two-step ahead forecast. The local linear method

performed better, in terms of average absolute predictive errors (AAPE), for the lynx example. The method performed just as well as the other three models in the sunspot numbers example.

Fan and Yao (2003) showed that a direct and iterative “naive plug-in” method performs well in terms of AAPE when compared to forecasting using a linear AR model. They adapt the local smoothing method of Cai, Fan, and Yao (2000) to estimate the functional coefficients in both methods. Using the direct method, forecast for X_{t+M} are computed as a function of X_t while ignoring the relationship between X_{t+M} and X_{t+M-j} , $j = 1, \dots, M - 1$. For $M \geq 2$, the iterative “naive plug-in” method simply plugs \hat{X}_{t+M-j} into the forecast equation. The form of the functional coefficient is determined using only the within-sample series values. The authors show that the iterative method outperforms the direct method for a TAR model of order two.

Huang and Shen (2004) use bootstrapping to sample the residuals of a univariate FCAR model for forecasting. The predicted values are obtained as

$$\hat{X}_{t+M} = \sum_{j=1}^p \hat{f}^{(j)}(\hat{X}_{t+M-q}) \hat{X}_{t+M-j} + \epsilon^b, \quad (1.10)$$

where ϵ^b is a bootstrapped value of the within-sample residuals from the FCAR model with functional coefficients estimated using polynomial splines. They note that care must be taken when \hat{X}_{t+M-q} falls outside or near the boundary of the range of the original X_{t-q} .

Harvill and Ray (2006) compare three methods for multi-step prediction using the univariate and vector FCAR models. The authors adapt the bootstrapping method of Huang and Sheng (2004) for forecasting FCAR models fit using local linear smoothing. They also adapt the multi-stage method of Chen (1996) and the “naive plug-in” method of Fan and Yao (2003) to vector FCAR models. The bootstrap method is found to be preferred to the other two methods. The multi-stage method tends to have larger bias especially for forecasting beyond two or three steps

ahead. Additionally, the authors also show that the bootstrap method performs well for forecasting a linear process.

1.4 Overview of Dissertation

The purpose of the dissertation is to present novel approaches to real data applications. The dissertation is organized as follows. In Chapter Two, we explore the MTAR model for multivariate time series. We present a data set consisting of isotope ratio measurements taken across the length of a stalagmite from Raccoon Mountain Cave in Tennessee. We compare linear time series models to the MTAR model. The analysis of the stalagmite data demonstrates that the ease of interpretability allows MTAR models to be powerful tools from a geological perspective.

In Chapter Three, we discuss the use of the FCAR model in the vector framework. This vector model is then used in an algorithm to determine the optimal number and arrangement of monitoring sensors at photovoltaic power plants. To illustrate utility, we apply the algorithm to a vector time series data set consisting of solar irradiance measurements from multiple sensors obtained at the La Ola photovoltaic plant in Lanai, Hawaii.

CHAPTER TWO

Time Series Modeling of Ancient Climates

2.1 Introduction

In the study of climate systems, knowledge of the climate's past evolution is becoming an increasingly necessary commodity. Typically, ancient climate reconstructions are accomplished via analysis of isotopic signatures present in paleoclimate proxies, or climate archives. These proxies include soil, ice core, and marine sediments. An isotopic signature is a ratio of non-radiogenic stable isotopes, stable radiogenic isotopes, or unstable radioactive isotopes of particular elements. Paleoclimate reconstructions, based on modeling of the isotopic signatures, aid in the understanding of both the climate's current state and future, unobserved patterns.

Speleothems, or cave formations, are particularly valuable logs of paleoclimate (Fairchild et al., 2006) and have been widely used to reconstruct paleo-vegetation and paleo-temperature (Dorale et al., 1998; Springer et al., 2008; Hardt et al., 2010; Springer et al., 2010; Li et al., 2013). In particular, stable oxygen and carbon isotopic signatures within speleothems prove to be excellent tools in describing surface climate conditions.

For an extensive review of the statistical methods and challenges in paleoclimate reconstruction see Tingley et al. (2012). We will highlight some of the methods discussed in Tingley et al. (2012), as well as discuss other modeling approaches. Li et al. (2010) and Tingley and Huybers (2010a,b) both propose Bayesian hierarchical models for reconstructions. Often modeling of isotopic signatures in speleothems rely on uranium series dating (Edwards et al., 1986) and methods for irregularly spaced time series. Fairchild and Baker (2012) discuss the popularity of linear autoregressive (AR) models for modeling isotope series in speleothems. Harvill and Ray (2006)

illustrate the use of multivariate functional coefficient autoregressive (FCAR) models in describing annual temperature and tree ring widths.

In modeling oxygen and hydrogen isotopic signatures within a stream, Birkel et al. (2012) employ Markov-switching autoregressive (MSAR) models. They find that each isotope series is modeled best using two hidden states. These hidden states may be associated with previously unknown series dynamics that linear models fail to describe. The ability of the MSAR models to better describe these states, while maintaining interpretable coefficients, is vital to the overall paleoclimate reconstruction.

We shall expand the univariate autoregressive idea of Fairchild and Baker (2012) and the regime-switching idea of Birkel et al. (2012) to a vector time series scenario. We introduce the use of vector momentum threshold autoregressive (VMTAR) models as a more sophisticated alternative by allowing for greater flexibility in model parameters by introducing a regime, or state, switching behavior. We demonstrate that the ease of interpretability allows momentum threshold models to be powerful tools from a geological perspective.

The remainder of this chapter is organized as follows. In Section 2.2, we describe a class of multivariate threshold autoregressive models for handling nonlinear time series. Section 2.3 introduces the speleothem data used in our application. We apply the model to the speleothem data set in Section 2.4 and show improvements over comparable linear vector autoregressive techniques. Concluding remarks and future research are given in Section 2.5.

2.2 Models

The time series models utilized are primarily motivated by our paleoclimate application to follow. As will be seen in Section 2.3, our data consists of paleoclimate information in addition to the isotopic signature series. Our goal is to jointly model the isotopic series, while taking advantage of this additional information. In

this section we will discuss both linear vector autoregressive and vector momentum threshold autoregressive models.

2.2.1 Vector Autoregressive Models

Define $\mathbf{X}_t = (X_{1,t}, X_{2,t}, \dots, X_{k,t})'$ as a k -variate zero mean vector time series observed at T time points. Let Φ_j be a $k \times k$ matrix of lag j autoregressive (AR) coefficients, $j = 1, 2, \dots, p$. A vector autoregressive model of order p , VAR(p), is given by

$$\mathbf{X}_t = \sum_{j=1}^p \Phi_j \mathbf{X}_{t-j} + \boldsymbol{\varepsilon}_t, \quad t = 1, 2, \dots, T, \quad (2.1)$$

where $\boldsymbol{\varepsilon}_t$ is a $k \times 1$ vector of zero mean white noise with covariance $\Sigma_{\boldsymbol{\varepsilon}} = \sigma^2 \mathbf{I}$.

The VAR(p) model has the advantage of being one of the simplest vector time series models. Note that when $k = 1$ the VAR(p) model reduces to the well-studied univariate order- p autoregressive, AR(p), model. Model parameters are easily estimated using common regression methods, such as ordinary least squares. Their simplicity leads to a popularity and better understanding across many disciplines. The model parameters are constant and independent of time resulting in an ease in interpretability of coefficients. However, the types of series dynamics and structures that may be explained by these models are limited.

2.2.2 Vector Momentum Threshold Autoregressive Models

Nonlinear time series methods may be employed to address many structural limitations of linear models. Before determining a parametric model it may be beneficial to explore the nonlinear dynamics of the series using a general class of nonlinear time series models. We consider the vector functional coefficient autoregressive (VFCAR) model in aiding with the exploratory analysis. The VFCAR of Harvill and Ray (2006), model is an extension of the univariate functional coefficient autoregressive model (see Section 1.3) introduced by Chen and Tsay (1993) and developed more extensively by Cai et al. (2000) and Chen and Liu (2001). For $\mathbf{X}_t = (X_{1,t}, X_{2,t}, \dots, X_{k,t})'$, the vector

functional coefficient autoregressive model of order p , VFCAR(p), is given by

$$\mathbf{X}_t = \sum_{j=1}^p \mathbf{f}^{(j)}(\mathbf{Z}_t) \mathbf{X}_{t-j} + \boldsymbol{\varepsilon}_t, \quad t = 1, 2, \dots, T, \quad (2.2)$$

where \mathbf{Z}_t is the functional variable of dimension $m \geq 1$ and $\boldsymbol{\varepsilon}_t$ is a $k \times 1$ vector of zero mean white noise with covariance $\boldsymbol{\Sigma}_\varepsilon = \sigma^2 \mathbf{I}$. The \mathbf{Z}_t may consist of exogenous predictors or lagged values of the series \mathbf{X}_t , and $\mathbf{f}^{(j)}$, $j = 1, \dots, p$, are $k \times k$ matrices with elements $[f_{i,t}^{(j)}]$ that are real-valued measurable functions that change as a function of \mathbf{Z}_t and which have continuous second derivatives. When all functions are constant with respect to the functional variable, the model in (2.2) reduces to (2.1).

In the basic threshold autoregressive (TAR) model seen in Section 1.2, the regime of interest is defined only by some thresholding series. An alternative specification allows for the threshold to depend on the magnitude of change, or momentum, in a previous period of a thresholding series. For convenience, we restate the momentum threshold autoregressive model here. Let $\{Z_t\}$ be the (possibly exogenous) thresholding time series, such that $\nabla Z_{t-q} = Z_{t-1} - Z_{t-q}$; that is, $\{\nabla Z_{t-q}\}$ is a time series that captures the q -th difference, or momentum, of the series $\{Z_t\}$. The ℓ -regime, order- p momentum threshold autoregressive, MTAR(p, ℓ, q), model for a zero mean time series $\{X_t\}$ is given by

$$X_t = \sum_{j=1}^p \phi_j^{(i)} X_{t-j} + \varepsilon_t^{(i)}, \quad \text{if } r_{i-1} < \nabla Z_{t-q} \leq r_i, \quad i = 1, \dots, \ell, \quad (2.3)$$

where $\phi_j^{(i)}$, $i = 1, 2, \dots, \ell$, is the lag j AR coefficient for regime i and $\varepsilon_t^{(i)} < \infty$ is zero mean white noise with standard deviation $\sigma^{(i)}$. Again, $-\infty = r_0 < r_1 < \dots < r_{\ell-1} < r_\ell = \infty$ are the $\ell + 1$ non-trivial thresholds dividing the domain of $\{\nabla Z_{t-q}\}$ into ℓ different regimes.

The speleothem application within Sections 2.3 and 2.4 will focus on the multivariate extension of the MTAR model. The vector momentum threshold autoregressive (VMTAR) model provides a mechanism for modeling differently the values of a vector time series, where different models are selected based upon the magnitude of

changes in previous values of a thresholding series. Let $\mathbf{X}_t = (X_{1,t}, X_{2,t}, \dots, X_{k,t})'$ be a zero mean vector time series, then the ℓ -regime, order- p vector momentum threshold autoregressive, VMTAR(p, ℓ, q), model for \mathbf{X}_t is given by

$$\mathbf{X}_t = \sum_{j=1}^p \Phi_j^{(i)} \mathbf{X}_{t-j} + \boldsymbol{\varepsilon}_t^{(i)}, \text{ if } r_{i-1} < \nabla Z_{t-q} \leq r_i, \quad i = 1, \dots, \ell, \quad (2.4)$$

where $\Phi_j^{(i)}$, $i = 1, \dots, \ell$ is a $k \times k$ matrix of lag j AR coefficients. The error terms $\boldsymbol{\varepsilon}_t^{(i)}$, $i = 1, \dots, \ell$ are $k \times 1$ vectors of zero mean white noise with covariance $\boldsymbol{\Sigma}_\varepsilon = \sigma^2 \mathbf{I}$. Across i , $\{\boldsymbol{\varepsilon}_t^{(i)}\}$ may be contemporaneously correlated. Further, $-\infty = r_0 < r_1 < \dots < r_{\ell-1} < r_\ell = \infty$ are the $\ell + 1$ non-trivial thresholds dividing the domain of $\{\nabla Z_{t-q}\}$ into ℓ different regimes.

2.3 Speleothem Data

Speleothems have been established as valuable tool for paleoclimate reconstructions. Through the process by which they form, speleothems possess the ability to capture climate details in the external environment. Isotopic signatures are ratios of stable or unstable chemicals in the material being investigated. Climate patterns in the external environment may be represented by chemical composition changes in the dripwater which eventually form the speleothem. Stable carbon and oxygen isotope ratios are the most commonly used signatures and help to describe both external temperature and precipitation in the reconstruction process. Furthermore, speleothems' continuous growth provide long, detailed climate archives.

The data for the application is obtained from speleothem RM0710-2-1 which was collected from Raccoon Mountain Cave near Chattanooga, Tennessee, USA. The stalagmite was collected circa 260 meters from the cave entrance (Li et al., 2014) and was actively growing at the time that it was collected. A map for the site at which the speleothem was collected is shown in Figure 2.1. The *Federal Cave Protection Act of 1988* makes it a federal crime, punishable by up to one year in prison and/or a \$10,000 fine, to disturb, deface, or destroy cave structures in a "significant cave,"

(without prior consent from the Secretary of the Interior). Raccoon Mountain Cave in Tennessee is a significant cave.

The stalagmite was dated using thorium-230 (^{230}Th) techniques similar to those described in Edwards et al. (1986). The clear and consistent preservation of the stalagmite in areas corresponding to the Middle and Late Holocene makes the specimen of particular interest. The Holocene is the most recent geological epoch spanning from approximately 11.7 thousand years (ka) before present (BP) until present day. According to Walker et al. (2012), the Middle Holocene spanned approximately 8.2 - 4.2 ka BP and lead into the Late Holocene which persists into present day. The stable isotope reconstruction will be for approximately the upper 11 centimeters of the stalagmite or roughly 5,000 years of climate records. This constitutes one of the most complete climate records for the time period. A cross-sectional image of the speleothem with accompanying ^{230}Th dates is seen in Figure 2.1.

After quarter-cutting, double-polished thin sections of the specimen were prepared. Digital ultraviolet fluorescence (UVf) photographs were taken along the edge of the thin sections and enlarged. The printed images were then manually stitched to show complete time-deposition coverage for the thin sections (Driese et al., 2016).

Isotopic signature time series were measured on powders collected each half-millimeter along the length of the stalagmite beginning 4.75 mm from the top of the speleothem. These powders are obtained when a half-millimeter in diameter drill bit is used to bore into the speleothem. Consequently, stable isotope ratio samples at each half-millimeter serve as “time averages” of multiple years’ climate records. The values of two isotopic signatures, specifically $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ both reported in parts per thousand (‰) Pee Dee Belemnite (PDB), are obtained. We shall briefly describe how $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ are interpreted in paleoclimatology; for a more detailed description see McDermott (2004).

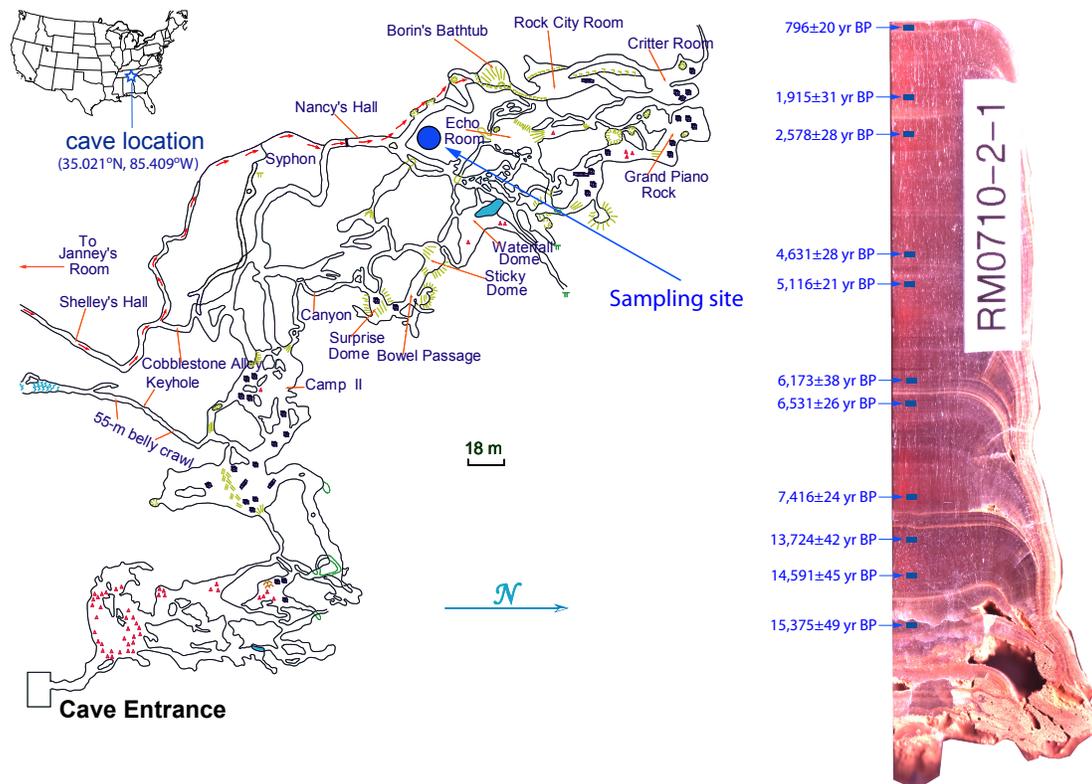


Figure 2.1: Map of the Raccoon Mountain Cave system. The blue dot represents the site at which speleothem RM0710-2-1 was collected. To the right of the map is a fluoresced and thorium-230 dated thin section of the stalagmite.

The value $\delta^{13}\text{C}$ is a measure of the ratio of stable isotopes carbon-13 to carbon-12. These values give information on changes in surface soil and vegetation dynamics, as well as indications of the cave's own microclimate structure. For example, plants living in high temperature, drought prone climates typically have higher (less negative) values of $\delta^{13}\text{C}$, while vegetation found in temperate climates produce lower (more negative) values of $\delta^{13}\text{C}$. Similarly, $\delta^{18}\text{O}$ is a measure of the ratio of stable isotopes oxygen-18 to oxygen-16. Information regarding both rainfall and temperature, among other things, are contained in the values of $\delta^{18}\text{O}$. As with $\delta^{13}\text{C}$, less negative values of $\delta^{18}\text{O}$ correspond to warmer, dryer climates; more negative values correspond to cooler, wetter climates.

The majority of speleothems are composed of calcium carbonate or calcium sulfate. They form through a series of chemical reactions. Rainwater reacts with carbon dioxide in the surface soil before traveling through the underlying bedrock. Once the solution reaches the cave, it degasses and drives precipitation of calcium carbonate. This process creates a layering effect which, over time, forms the speleothem. Typically visible after fluorescence, these layers, or rings, hold important information about the paleoclimate. Driese et al. (2016) find evidence that cooler/wetter paleoclimate conditions correspond with thinner rings and more-negative $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ values. Conversely, thicker rings and less-negative $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ values would tend to indicate a warmer/drier conditions. The authors present the saturation state of carbonate in the cave water as a likely explanation. For example, during drier climate phases, the carbonate in the cave water was increased due to greater evaporation on the surface and less water infiltration, which favored calcite precipitation, resulting in growth of thicker rings. Geologists believe that these rings correspond to annual deposit records, similar to tree rings.

Because of the well-preserved nature of speleothem RM0710-2-1, the rings are visible in the UVf images. These visible rings may be seen in Figure 2.2. Using the

manually stitched printed images, we are able to count the number of speleothem rings for each half millimeter where isotopic values were measured. Twenty-nine half millimeter intervals experience at least partial dissolution or “time loss” in the record. For these interval, linear interpolation was used to obtain estimated counts. This new series of ring counts gives the ability to incorporate information contained in the rings into any modeling of $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$. Each of the isotope measurements from the speleothem is time averaged per half-millimeter. Across the length of the specimen we observe a sequence of data points consisting of successive measurements. Therefore, distance in half-millimeters from the top of the speleothem will be used as a proxy for time. Let $d = 1, 2, \dots, 195$ represent the half-millimeter increments from the top of the speleothem beginning 4.75 mm down. Then for each value of d , the variables measured are the $X_{1,d} = \delta^{13}\text{C} \text{ ‰ PDB}$, $X_{2,d} = \delta^{18}\text{O} \text{ ‰ PDB}$, and $X_{3,d} = \text{number of rings}$; concurrently, these variables can be considered as a vector of length three written $\mathbf{X}_d = (X_{1,d}, X_{2,d}, X_{3,d})'$. Plots of the resulting series are shown in Figure 2.3. Note that older records occur at larger distances from the top of the speleothem. Therefore, lagged values of each series are given by increases in distance. For example, the lag j $\delta^{13}\text{C}$ value at some distance d is given by $X_{1,d+j}$.

An examination of the plots, especially the plots of $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$, suggests a changing system. In particular, from zero mm to 20 mm from the top of the speleothem, the system seems to be varying less than from 20 mm to around 50 mm. Additionally, there seems to be an upward trend in the two isotope ratios from 20 mm to 50 mm. From 50 mm to around 80 mm the $\delta^{13}\text{C}$ ratio seems to be decreasing, with some sharp rises and falls, while the $\delta^{18}\text{O}$ ratio seems to be remaining somewhat constant, and less variable than $\delta^{13}\text{C}$. Beyond 80 mm from the top, we see another marked change in $\delta^{13}\text{C}$, while $\delta^{18}\text{O}$ appears to vary in the same manner as from 50 mm to 80 mm.

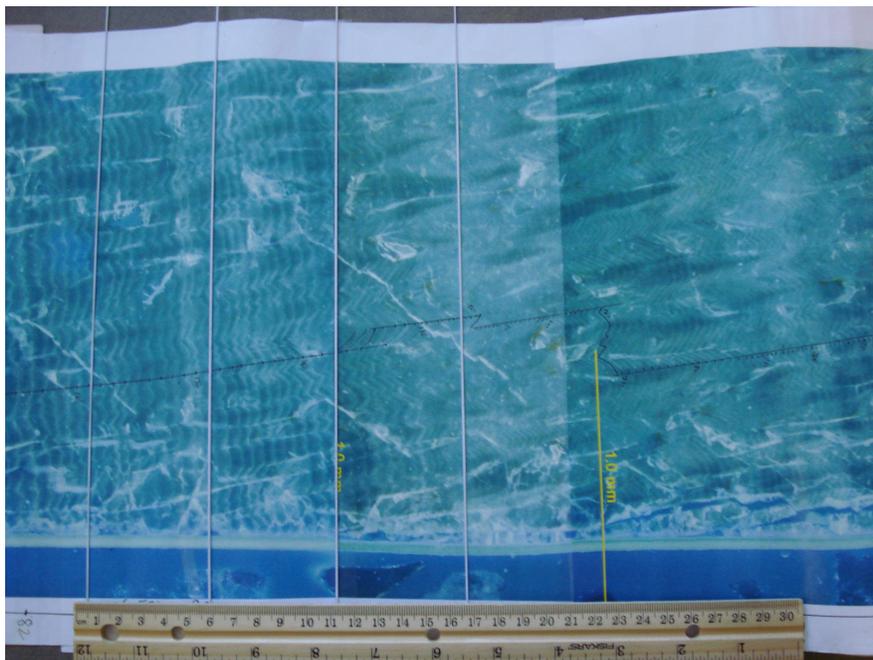
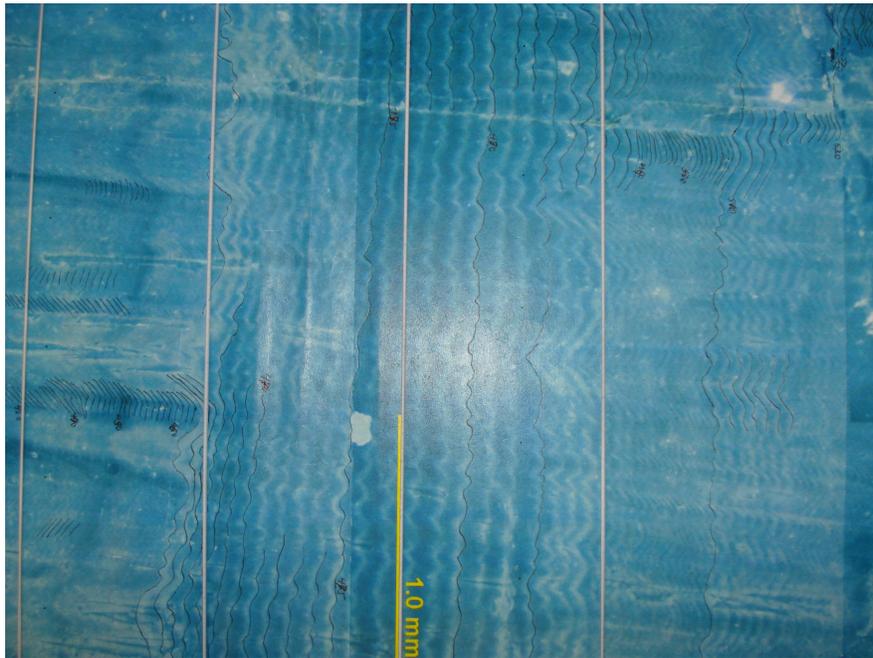


Figure 2.2: Images of the enlarged UVf photographs showing well preserved laminae, or rings.

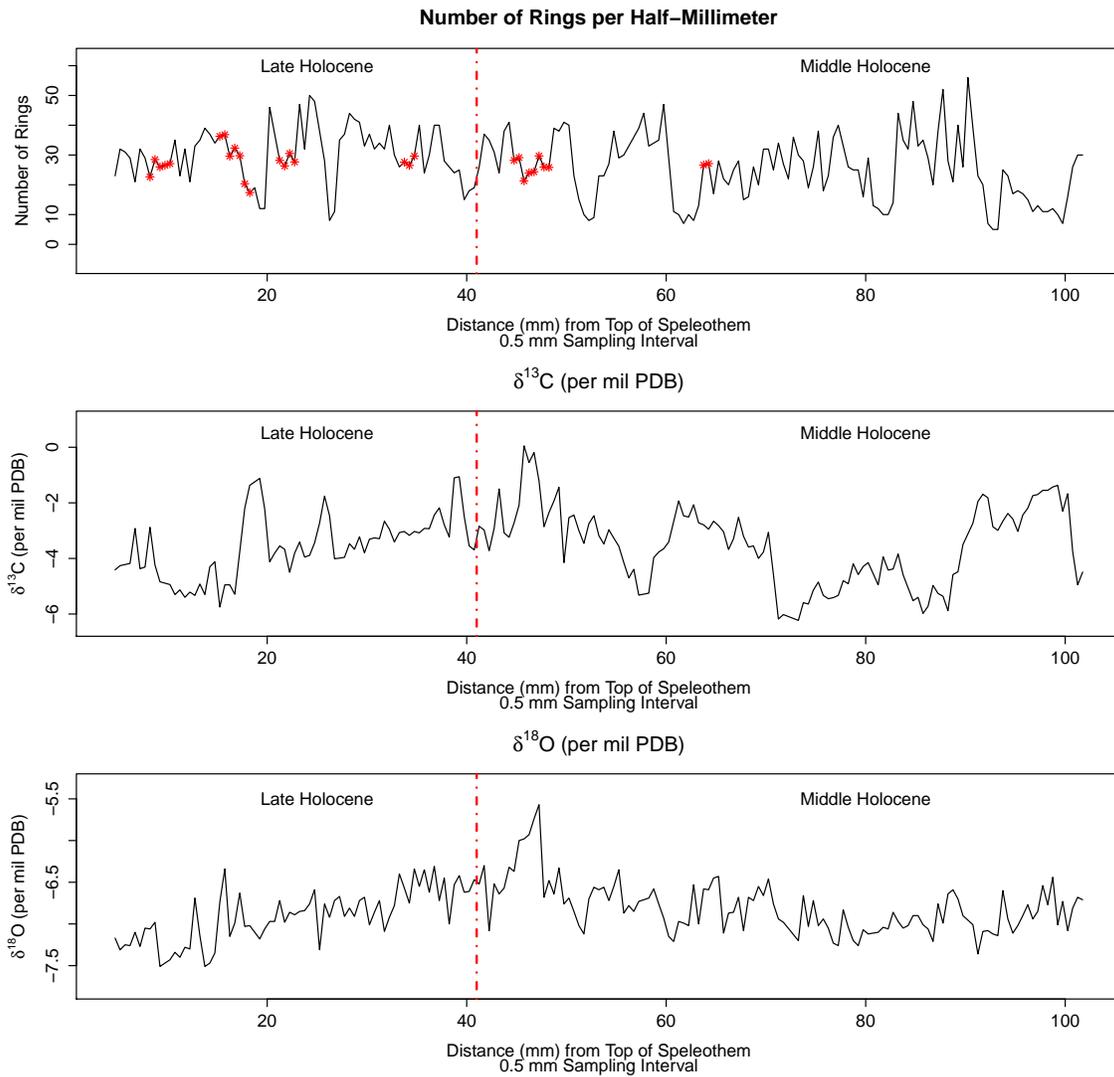


Figure 2.3: Plots of number of rings (top), $\delta^{13}\text{C}$ ‰ PDB, (middle), and $\delta^{18}\text{O}$ ‰ PDB (bottom). The asterisks (*) on the plot of number of rings represent data imputed using linear interpolation.

We also observe what seems to be a changing system in the number of rings. The most obvious change is seen in comparing the number of rings from 0 mm to 60 mm, from 60 mm to 90 mm, and then beyond 90 mm. From 0 mm to 60 mm, the number of rings has many sharp rises and falls. After any sharp rise, there appears to be a small amount of variability, until the next fall. However, beyond 60 mm, the number of rings behaves in a more regular fashion, with the exception of the sharp rise just after 80 mm. Around 92 mm from the top of the speleothem, there is another rise, with a slowly varying decline before another sharp rise.

2.4 Results

We apply methods for fitting both VAR and VMTAR models to the Raccoon Mountain speleothem data. In Section 2.4.1 we expand upon previous methods and fit a VAR model to the data. However, the complex nature of the system noted in the description of the plots in suggests the VAR model may be insufficient for describing the isotope system. In Section 2.4.2, we present the VMTAR model. We fit a three-regime VMTAR model to the data, and describe the results. The geological significance of the VMTAR model is discussed in Section 2.4.3.

2.4.1 VAR Results

Fairchild and Baker (2012) suggest the use of univariate autoregressive (AR) models when describing individual isotopic signature series in speleothems. Because at each half-millimeter we observe a vector time series, we expand the use of AR models to the vector scenario. This provides an initial direction for jointly modeling the number of rings, $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ using a vector autoregressive model. Examination of both the Akaike information criterion (AIC) and Bayesian information criterion (BIC) reveal an order-one vector autoregressive model, VAR(1), to be optimal. Recall for the d^{th} half-millimeter below the top of the speleothem ($d = 1, 2, \dots, 195$) that $X_{1,d}$ represents $\delta^{13}\text{C} \text{‰ PDB}$, $X_{2,d}$ represents $\delta^{18}\text{O} \text{‰ PDB}$, and $X_{3,d}$ represents the

number of rings. The VAR(1) model is then given by

$$\begin{aligned}
X_{1,d} &= \phi_1^{(1)} X_{1,d+1} + \phi_2^{(1)} X_{2,d+1} + \phi_3^{(1)} X_{3,d+1} + \varepsilon_d^{(1)} \\
X_{2,d} &= \phi_1^{(2)} X_{1,d+1} + \phi_2^{(2)} X_{2,d+1} + \phi_3^{(2)} X_{3,d+1} + \varepsilon_d^{(2)} \\
X_{3,d} &= \phi_1^{(3)} X_{1,d+1} + \phi_2^{(3)} X_{2,d+1} + \phi_3^{(3)} X_{3,d+1} + \varepsilon_d^{(3)},
\end{aligned} \tag{2.5}$$

where for a fixed $j = 1, 2, 3$, the $\varepsilon_d^{(j)}$ is a sequence of zero-mean, independent errors, and at any fixed half-millimeter d^* , the correlation between $\varepsilon_{d^*}^{(i)}$ and $\varepsilon_{d^*}^{(j)}$ is $\rho_{i,j}$, $i = 1, 2, 3$, $i \neq j$.

Using ordinary least square to fit the VAR(1) model to the data results in

$$\begin{aligned}
\hat{X}_{1,d} &= 0.7938X_{1,d+1} + 0.2418X_{2,d+1} - 0.0101X_{3,d+1} \\
\hat{X}_{2,d} &= 0.0513X_{1,d+1} + 0.5395X_{2,d+1} + 0.0039X_{3,d+1} \\
\hat{X}_{3,d} &= -0.3695X_{1,d+1} - 1.4780X_{2,d+1} + 0.5843X_{3,d+1}.
\end{aligned} \tag{2.6}$$

Table 2.1 contains the coefficients along with approximate P -value from corresponding t -tests for testing their significance from zero. The diagonal elements of Table 2.2 are the estimated mean square errors for each term in the model. Modeling of the $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ series is of primary importance and yields estimated mean square errors of 0.2343 and 0.0286, respectively. The estimated overall mean square error for the VAR(1) model's fit of $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ is 0.2636. The elements in the off-diagonals are estimates of the contemporaneous error correlations, $\rho_{i,j}$. The numbers in parentheses

Table 2.1: Estimated coefficients of VAR(1) model and approximate P -values (in parentheses) for tests of significance from zero.

| Variable | $\delta^{13}\text{C}$ | $\delta^{18}\text{O}$ | Number of rings |
|-----------------------|--------------------------|--------------------------|--------------------------|
| $\delta^{13}\text{C}$ | 0.7938 (< 0.0001) | 0.2418 (0.1797) | -0.0101 (0.0577) |
| $\delta^{18}\text{O}$ | 0.0513 (0.0017) | 0.5395 (< 0.0001) | 0.0039 (0.0339) |
| Number of rings | -0.3695 (0.5130) | -1.4780 (0.4980) | 0.5843 (< 0.0001) |

beneath the estimated correlations are approximate P -values of t -tests for testing significance from zero. A significant error correlation implies that one error series may be one of the driving factors for the other error series. For example, the negative estimated contemporaneous error correlation between the residuals for the number of rings and the residuals for $\delta^{13}\text{C} \text{‰ PDB}$ implies that a smaller error in number of rings is associated with an increase in error for $\delta^{13}\text{C} \text{‰ PDB}$.

A plot of the original vector series (solid line) with the fitted values (dashed line) superimposed is presented in Figure 2.4. Upon initial inspection it would appear that the VAR model is effective in modeling the isotopic signatures, especially the $\delta^{13}\text{C}$ series. Examination of both the fitted plot and significant coefficients of the $\delta^{13}\text{C}$ series indicate a fairly simple autoregressive structure. However, the fit for the $\delta^{18}\text{O}$ series does not appear as successful. In particular the range is not captured in the fitted values. For example, at approximately 45 half-millimeters below the top of the speleothem the shape of the $\delta^{18}\text{O}$ series is detected but not height of the peak. This suggest that a more sophisticated model could improve the explanation of the behavior of the $\delta^{18}\text{O}$ series.

2.4.2 VMTAR Results

The thickness of a UVf layer in a speleothem is believed to be directly related to the amount of rainfall. Consequently the number of rings per half millimeter is an

Table 2.2: Estimated mean square errors and estimated contemporaneous error correlations with approximate P -values for significance from zero in parentheses from order-one vector autoregressive model.

| Variable | $\delta^{13}\text{C}$ | $\delta^{18}\text{O}$ | Number of rings |
|-----------------------|-----------------------|-----------------------|---------------------|
| $\delta^{13}\text{C}$ | 0.2343 | 0.2229 (0.0019) | -0.2072 (0.0039) |
| $\delta^{18}\text{O}$ | 0.2229 | 0.0286 | 0.1081 (0.1355) |
| Number of rings | -0.2072 | 0.1081 | 34.8311 |

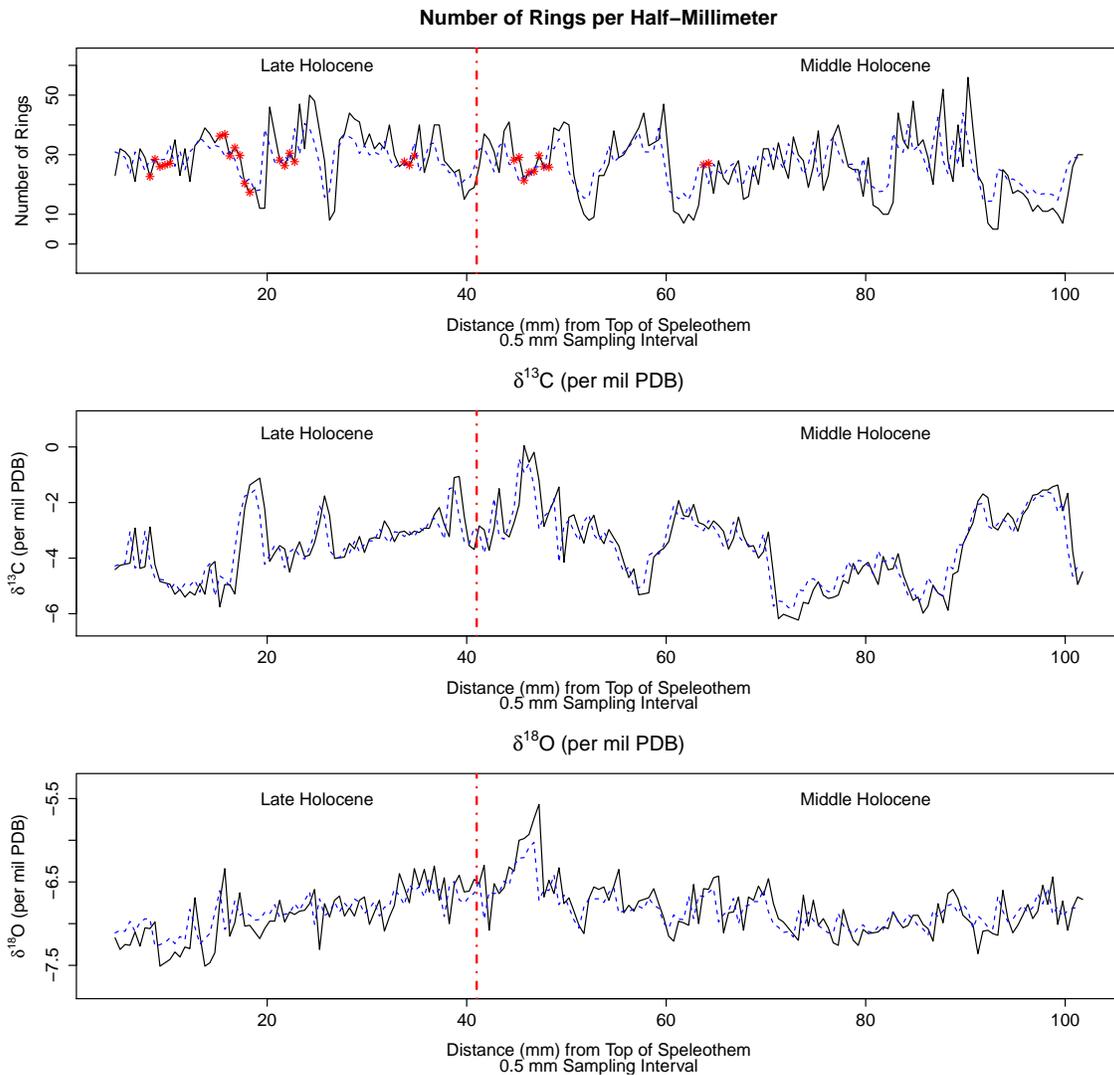


Figure 2.4: Plots of original data (—) and fitted values (---) from an order-one vector autoregressive model. The top plot is the number of rings per half-millimeter; the center plot is of $\delta^{13}\text{C}$ ‰ PDB; the bottom plot is of $\delta^{18}\text{O}$ ‰ PDB. The asterisks (*) on the plot of the number of rings represent data imputed using the fitted VAR(1) model.

excellent candidate for describing the nonlinear relationship. Define the momentum as $\nabla X_{3,d+2} = X_{3,d+1} - X_{3,d+2}$. If $\nabla X_{3,d+2} < 0$, this is an indication that the climate is entering a drier period. On the other hand, if $\nabla X_{3,d+2} > 0$, the climate is entering a wetter period.

Because the number of rings per half-millimeter will not be directly modeled, the autoregressive order is extended to order-two. Using $\nabla X_{3,d+2}$ as the functional variable, an order-two VFCAR model is fit to $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ using kernel regression techniques. The estimates of the functional coefficients are then used to explore the nonlinear dynamics of the stable isotope series. Figures 2.5 and 2.6 shows the fitted values of the functional coefficients for the VFCAR(2) model. An overall inspection of the functional coefficients reveals some non-constant functional coefficients indicating the presence of a nonlinear series structure. For modeling $\delta^{13}\text{C}$ the functional coefficient for lag-1 $\delta^{13}\text{C}$ appears significant across all values of the function variable but has a negative correlation with momentum in periods with extreme to moderate shifts toward drier climates and a positive correlation elsewhere. However, for $\delta^{18}\text{O}$ the lag-1 $\delta^{13}\text{C}$ functional coefficient appears to be constant and zero except in periods of moderate shifts toward a wetter climate where it begins to exhibit a negative correlation. The lag-2 $\delta^{13}\text{C}$ functional coefficient appears to be mostly constant and zero for both $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$. Only in the most extreme observed climate transition does the lag-2 $\delta^{13}\text{C}$ coefficient appear to have significant correlation for $\delta^{18}\text{O}$. For $\delta^{13}\text{C}$ the lag-1 $\delta^{18}\text{O}$ coefficient appears constant and zero for transitions to drier climate periods; however, a negative correlation with the functional variable is present during periods transitioning to wetter climates. Interestingly, for $\delta^{18}\text{O}$ the coefficient for lag-1 $\delta^{18}\text{O}$ appears to exhibit three distinct areas of behavior: two areas of negative correlation during moderate transitions and a constant coefficient when the climate is in a “neutral” state. Over most values of the functional variable, the functional coefficient for lag-2 $\delta^{18}\text{O}$ has a positive correlation for $\delta^{13}\text{C}$. For transitions wetter cli-

mate states the lag-2 $\delta^{18}\text{O}$ coefficient appears to be significant. For $\delta^{18}\text{O}$ the lag-2 $\delta^{18}\text{O}$ coefficient also demonstrates three states. The transitions appear to occur at approximately the same values of the functional variable. However, in the case of the lag-2 $\delta^{18}\text{O}$ coefficient there are two areas of positive correlation during moderate transitions. Again, the coefficient appears to be constant in the “neutral” climate state.

Based upon the behavior noted in the plots of the functional coefficients and the desire to preserve an ease of geological interpretability for the coefficient, we use a VMTAR model to fit the stable isotope ratios. Using $\nabla X_{3,d+2}$ as the thresholding variable, we fit a three-regime, order-two VMTAR model to $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$.

The package `tsDyn` in the *R* language was used for testing and fitting the model. The function `TVAR` was then used to fit the model to the data. To determine the optimal threshold values for defining the model regimes, the function performs the grid search proposed by Chan (1993) and Enders and Siklos (2001) and described in Section (1.2.1). Figure 2.7 contains three graphs related to this grid search. The top graph represents the momentum in number of ring ($\nabla X_{3,d+2}$) on the vertical axis versus the half-millimeter mark on the horizontal. The lower dashed horizontal line contained within the graph represent the optimum threshold, as determined by a grid search, for transitioning from a stable climate into a drier period. The value of that threshold is $r_1 = -4$. The upper dashed horizontal line contained within the graph represent the optimum threshold, as determined by a grid search, for transitioning from a stable climate into a wetter period. The value of that threshold is $r_2 = 10$. The center graph illustrates the sorted momentum with the drier and wetter thresholds superimposed using an asterisk (*). ($r_1 = -4$ and $r_2 = 10$). The bottom graph is of the sum of squared residuals (SSR) of models fit using varying thresholds versus the threshold values. The threshold of $r_1 = -4$ corresponds to a SSR equal to 91.50327.

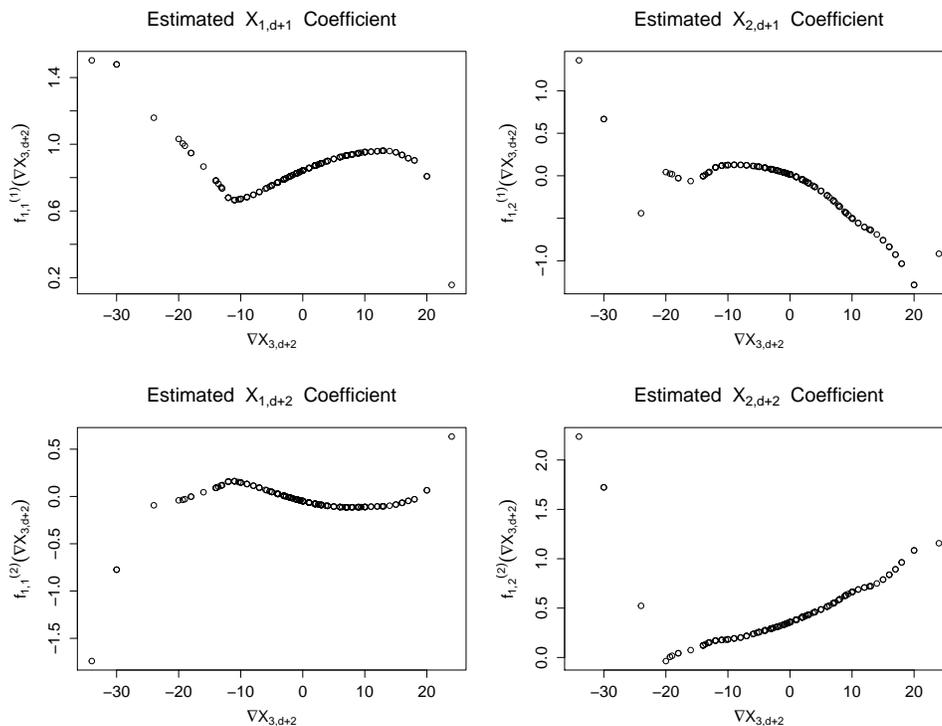


Figure 2.5: Pointwise estimates of VFCAR(2) coefficient functions for $\delta^{13}\text{C} \text{‰ PDB}$.

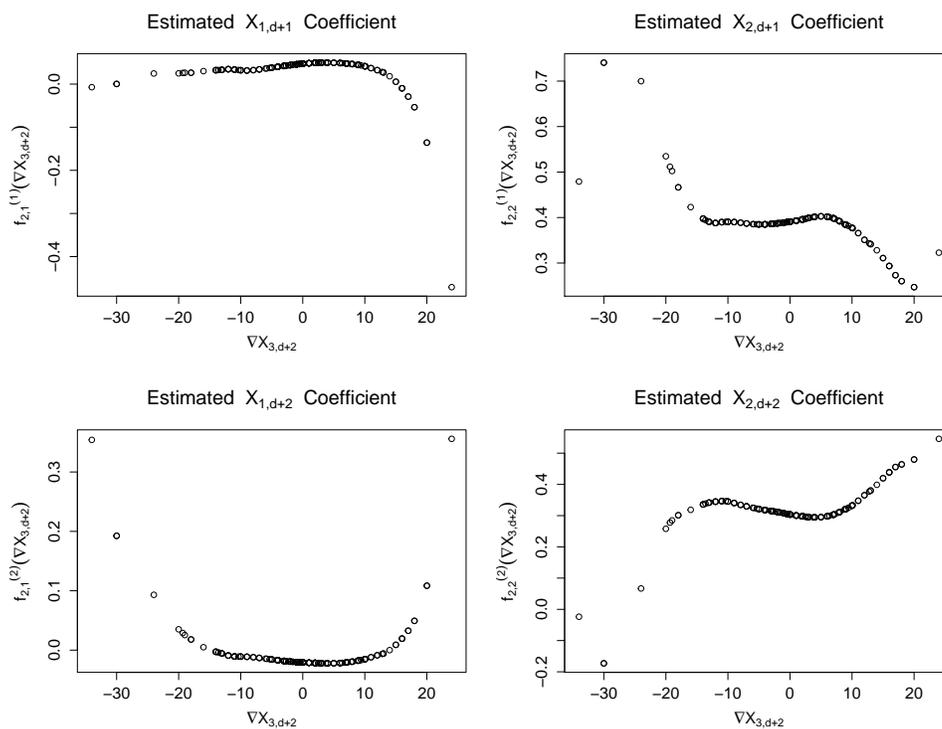


Figure 2.6: Pointwise estimates of VFCAR(2) coefficient functions for $\delta^{18}\text{O} \text{‰ PDB}$.

The threshold of $r_2 = 10$ corresponds to a SSR equal to 92.96762. These values are local minima.

The fitted VMTAR(2,3,2) model is as follows. For $\nabla X_{3,d+2} < -4$,

$$\begin{aligned}\hat{X}_{1,d} &= 0.8489X_{1,d+1} - 0.5697X_{2,d+1} + 0.1945X_{1,d+2} + 4844X_{2,d+2} \\ \hat{X}_{2,d} &= 0.0758X_{1,d+1} + 0.1651X_{2,d+1} - 0.0618X_{1,d+2} + 0.7053X_{2,d+2}.\end{aligned}$$

Of the 193 observations, 27.2% fell in a period of drier climate. For $-4 \leq \nabla X_{3,d+2} < 10$, the fitted model is

$$\begin{aligned}\hat{X}_{1,d} &= 0.8579X_{1,d+1} + 0.1021X_{2,d+1} - 0.1124X_{1,d+2} + 0.1423X_{2,d+2} \\ \hat{X}_{2,d} &= 0.0222X_{1,d+1} + 0.5320X_{2,d+1} - 0.0193X_{1,d+2} + 0.1733X_{2,d+2}.\end{aligned}$$

Of the 193 observations, 61.8% fell in a moderate climate period. For $\nabla X_{3,d+2} \geq 10$, the fitted model is

$$\begin{aligned}\hat{X}_{1,d} &= 0.4228X_{1,d+1} - 0.6199X_{2,d+1} + 0.2228X_{1,d+2} + 2.2691X_{2,d+2} \\ \hat{X}_{2,d} &= -0.1858X_{1,d+1} + 0.2401X_{2,d+1} + 0.2443X_{1,d+2} + 0.3653X_{2,d+2}.\end{aligned}$$

Of the 193 observations, 11% fell into the period of wetter climate.

Table 2.3 contains these coefficients along with corresponding standard errors (in parentheses). Each set of two rows represents one regime. Each column represents a lagged value of either $\delta^{13}\text{C}$ or $\delta^{18}\text{O}$. Coefficients found significantly different from zero at the 0.01 level are marked with an asterisk. Table 2.4 contains the estimated mean square errors and estimated contemporaneous correlations associated with the fitted VMTAR(2,3,2) model. The estimated mean square error for $\delta^{13}\text{C}$ is 0.2051 and 0.0237 for $\delta^{18}\text{O}$. Use of the VMTAR(2,3,3) model resulted in an approximate 12.5% reduction in mean square error for $\delta^{13}\text{C}$ and an approximate 18% reduction in mean square error for $\delta^{18}\text{O}$. The overall mean square error for the VMTAR(2,3,2) model is 0.2287, an approximate 13% reduction in overall mean square error. The significant coefficients from the VMTAR(2,3,2) tend to agree with the behavior noted in the

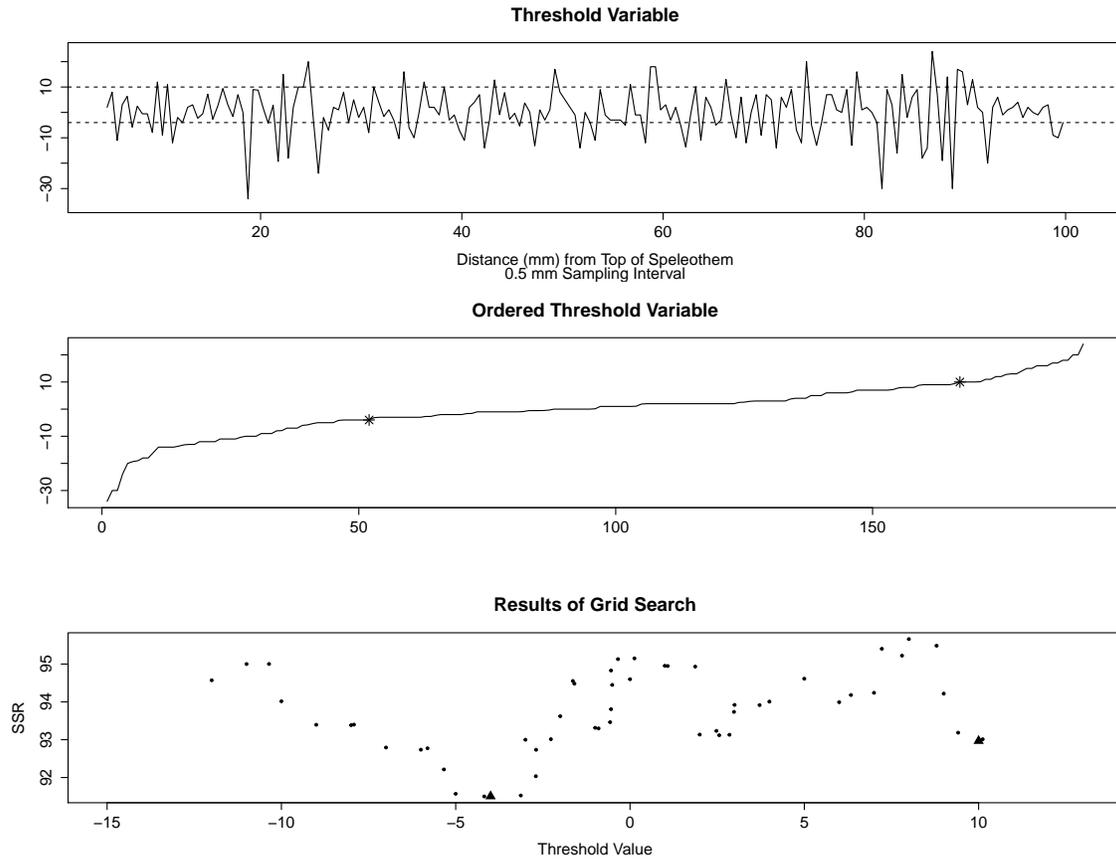


Figure 2.7: Graphical output for grid search for two thresholds. The top plot is a time plot of the momentum in the number of UVf layers (the threshold variable) with threshold values ($r_1 = -4$ and $r_2 = 10$) represented via the horizontal lines. The middle graph is of the ordered momentum in the number of UVf layers with the threshold values labeled as *. The bottom graph contains the results of a grid search for the threshold values, with the pair chosen that resulted in the smallest sum of squared residuals (SSR) marked as solid triangles.

plots of the functional coefficient. Disagreements in coefficients where significance was suspected is likely due to some regimes having a smaller percentage of observations, most notable the $\nabla X_{3,d+2} \geq 10$ regime.

Finally, in Figure 2.8, we present a plot of the original series (solid black) with the fitted values (dashed blue) from the VMTAR model superimposed. A comparison of the graphs in Figure 2.8 to the bottom two plots in Figure 2.4 reveals the superior fit of the VMTAR(2,3,2) model compared to the VAR(1) model which agree with analysis of mean square errors for both models.

2.4.3 Geological Significance

Using the VMTAR(2,3,2) we are able to quantify the relationship between $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ dependent upon the momentum in the climate as indicated by the change in the number of annual UVf layers in the speleothem. The model proves to be a powerful tool from a geological standpoint. Simple visual inspection of the UVf layers and the corresponding stable isotope values were insufficient in solely explaining these complex relationships.

The coefficients and threshold values found to optimize the performance of the VMTAR(2,3,2) model indicate the presence of previously unknown system structures. These factors influenced the behavior of the Raccoon Mountain Cave speleothem growth rate and stable isotope system. That is, the model demonstrates that abrupt shifts in climate state do cause shifts in stable isotope values.

For transitions into a drier climate state, characterized by fewer, thicker UVf layers per half-millimeter, $\delta^{13}\text{C}$ values respond directly whereas $\delta^{18}\text{O}$ values have a more delayed direct response. When transitioning to a wetter climate state, characterized by many, thinner UVf layers, $\delta^{13}\text{C}$ has a delayed behavior with respect to $\delta^{18}\text{O}$. Similarly, $\delta^{18}\text{O}$ depends on the more delayed $\delta^{13}\text{C}$ value. When the climate enters the “neutral” state, $\delta^{13}\text{C}$ changes with respect to only the last half-millimeter’s carbon isotope value, while $\delta^{18}\text{O}$ requires its previous two periods’ values. The hy-

Table 2.3: Significant VMTAR(2,3,2) coefficients with standard errors (in parentheses). Coefficients marked with an * are significantly zero at a 0.01 level of significance. Additionally, coefficients marked with an † are also significantly zero at a 0.05 level of significance. All other coefficients are not significantly different from zero.

| Threshold | Variable | $X_{1,d-1}$ | $X_{2,d-1}$ | $X_{1,d-2}$ | $X_{2,d-2}$ |
|---------------------------------|-----------------|-------------------|-------------------|-------------------|-------------------|
| $\nabla X_{3,d+2} < -4$ | $\hat{X}_{1,d}$ | 0.849* (0.133) | -0.570 (0.455) | 0.195 (0.142) | 0.488 (0.454) |
| | $\hat{X}_{2,d}$ | 0.076† (0.045) | 0.165 (0.155) | -0.062 (0.048) | 0.705* (0.154) |
| $-4 \leq \nabla X_{3,d+2} < 10$ | $\hat{X}_{1,d}$ | 0.858* (0.089) | 0.102 (0.248) | -0.112 (0.087) | 0.142 (0.236) |
| | $\hat{X}_{2,d}$ | 0.022 (0.030) | 0.532* (0.084) | -0.019 (0.030) | 0.173* (0.080) |
| $\nabla X_{3,d+2} \geq 10$ | $\hat{X}_{1,d}$ | 0.423 (0.340) | -0.620 (0.695) | 0.223 (0.292) | 2.269* (0.931) |
| | $\hat{X}_{2,d}$ | -0.186 (0.116) | 0.240 (0.237) | 0.244* (0.100) | 0.365 (0.317) |

Table 2.4: Estimated mean square errors and contemporaneous estimated error correlations with approximate P -values for significance from zero in parentheses from order-two vector momentum threshold autoregressive model.

| Variable | $\delta^{13}\text{C}$ | $\delta^{18}\text{O}$ |
|-----------------------|-----------------------|-----------------------|
| $\delta^{13}\text{C}$ | 0.2051 | 0.1688 (0.0196) |
| $\delta^{18}\text{O}$ | 0.1688 | 0.0237 |

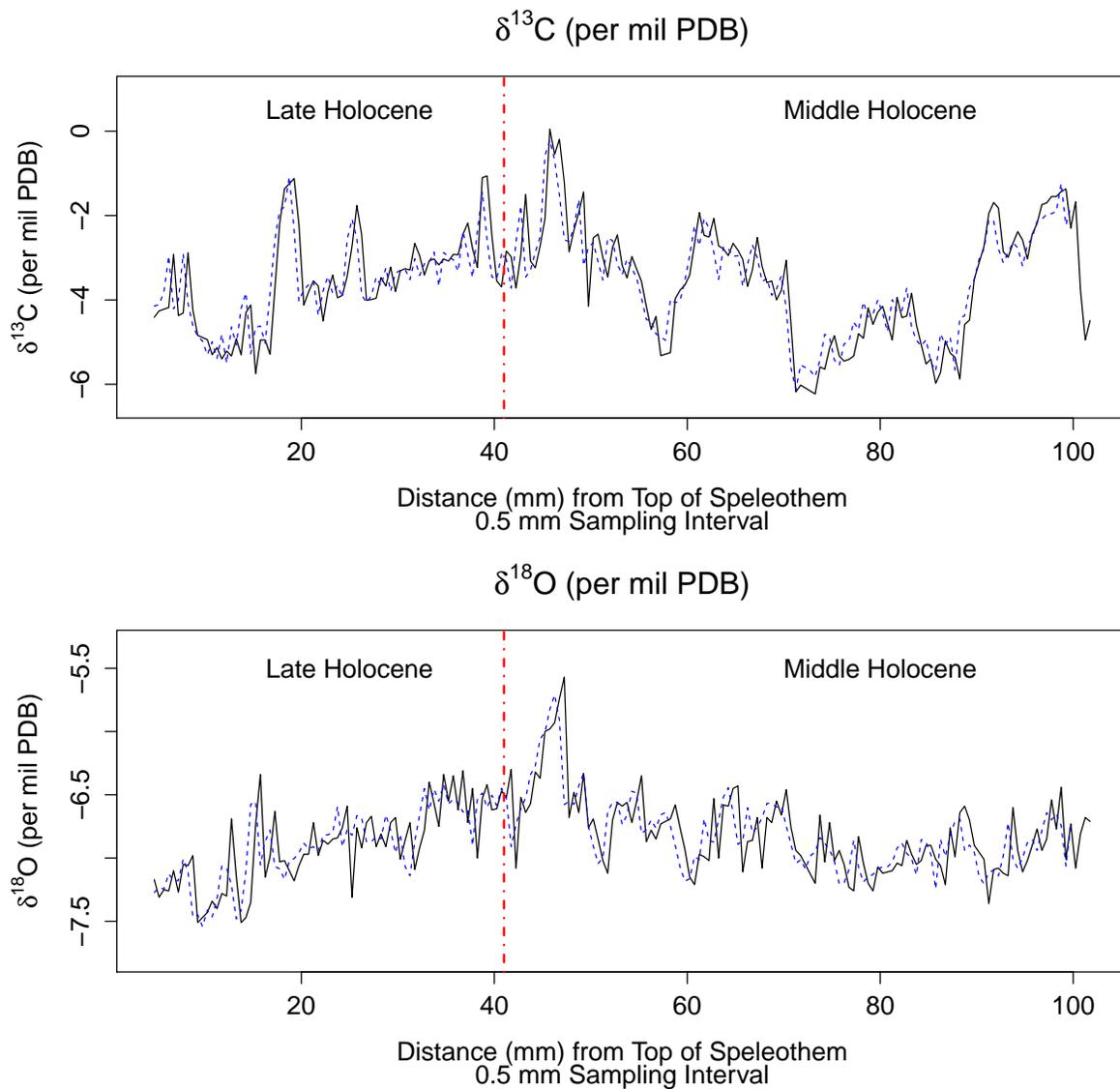


Figure 2.8: Plots of original series (—) and fitted series (- - -) from an order-two momentum threshold vector autoregressive model. The top plot is of $\delta^{13}\text{C}$ ‰ PDB; the bottom plot is of $\delta^{18}\text{O}$ ‰ PDB.

drological and geochemical basis for this behavior needs to be explored further, but could reflect complexities associated with “open and closed system” behavior at the Raccoon Mountain site, which were discussed recently by Li et al. (2014).

2.5 Concluding Remarks

We have demonstrated the capabilities of a momentum threshold autoregressive model when applied to stable isotope measures taken from speleothem. We additionally show that the relationship between the two stable isotopes (C, O) changes substantially depending upon the state of the climate. We have examined the model’s performance when compared to a linear vector autoregressive model. The VMTAR model provides superior estimation while preserving the ease on interpretability of the simpler model.

Other times trace elements, such as strontium, magnesium, or calcium, are measured along the length of the speleothem. However, these series often have a much finer resolution than the isotope series. Future work would incorporate information contained in the trace element series in the joint modeling. In this paper, we have shown that vector momentum threshold autoregressive models hold great promise for future paleoclimate reconstructions using speleothem.

CHAPTER THREE

Semiparametric Approach to Optimal Sensor Design in a Photovoltaic Power Plant

3.1 Introduction

Forecasting on a short-term horizon of solar irradiance is a crucial aspect in the design and monitoring of photovoltaic (PV) power plants. These forecasts are used in operational procedures such as switching sources, power purchases, and usage of power reserves. Past work has illustrated a correlation between the plant's power output and irradiance measures taken from sensors across a plant's footprint (Kuszamaul et al., 2010). Therefore, variations of solar irradiance in the presence of changing weather conditions will result in instability over the utility's service area. Relatively small errors in predicted irradiance may translate to significant uncertainty in projected profit because utility-scale PV plants are typically leveraged financially. Accurate monitoring of these forecasts is an integral part of the planning and construction phase of a PV plant.

There is a rich literature on efforts to apply statistical methods to the modeling and forecasting of irradiance. Early work focused on the use of conventional time-series methods such as autoregressive (AR) and autoregressive moving average (ARMA) models (e.g. Aguiar and Pereira, 1992; Mora-Lopez and Sidrach-de-Cardona, 1998; Bacher et al., 2009). However, these linear methods are not appropriate for the nonlinear structure of solar irradiance data. Glasbey (2001) explored nonlinear autoregressive analysis. Much work has been done with forecasting involving artificial neural networks (e.g. Alam et al., 2006; Paoli et al., 2010; Oudjana et al., 2014). In addition to the artificial neural networks, Paoli et al. (2010) discuss modeling based on a k -Nearest Neighbors algorithm and Bayesian inference. More recently, Ghayekhloo et al. (2015) and Ghofrani et al. (2016) propose novel clus-

tering algorithms for short-term irradiance forecasts. Patrick et al. (2016) evaluate semiparametric spatio-temporal models at high temporal and spatial resolutions.

Although extensive work has been done in the application of statistical models to predict solar irradiance on a short term resolution, we are unaware of their use to aid in the initial design of a small scale PV system. In this chapter, we extend the work of Patrick et al. (2016) to a temporally dense, but spatially sparse setting. We explore the use of vector functional coefficient autoregressive (VFCAR) models to help answer the question, “If a new PV plant is being constructed, what is the optimal number and layout of sensors for predicting solar irradiance?” We propose a sensor design algorithm that incorporates the semiparametric VFCAR model to forecast irradiance series across a small number of monitoring sensors.

The remainder of the chapter is organized as follows. In Section 3.2, we formally define the vector functional coefficient autoregressive model and briefly discuss model properties and estimation techniques. In Section 3.2.3, we will discuss a bootstrap method for obtaining M -step-ahead predictions with the VFCAR model. More specifically, the sensor algorithm is presented in Section 3.3. Section 3.4 discusses the result. Section 3.5 concludes with remarks and future research.

3.2 *Methods*

Nonlinear time series models often out-perform linear models when used to model series resulting from complex dynamics, such as solar irradiance. The class of nonlinear time series models includes many popular parametric models including the bilinear model, the exponential autoregressive (EXPAR) model, the threshold autoregressive (TAR) model, the smooth threshold autoregressive (STAR) model, and generalized autoregressive conditional heteroscedasticity (GARCH) models, among others. There is no specific class of parametric nonlinear models that is generally applicable to solar irradiance data. Therefore, it is useful to reduce the size of the class of nonlinear models. One way of accomplishing this reduction is through a versatile,

nonlinear, semiparametric model known as the functional coefficient autoregressive model.

3.2.1 Functional Coefficient Autoregressive Model

A flexible semiparametric model is the functional coefficient autoregressive model of order p (FCAR(p)), first introduced by Chen and Tsay (1993). For convenience, we restate the model here. Let X_t be a zero mean univariate time series observed at T time points. Define the FCAR(p) model as

$$X_t = \sum_{j=1}^p f^{(j)}(Z_t)X_{t-j} + \varepsilon_t, \quad t = p+1, p+2, \dots, T, \quad (3.1)$$

where Z_t is the functional variable which may consist of exogenous predictors or lagged values of the series X_t , $f^{(j)}(\cdot)$, $j = 1, \dots, p$, are real-valued measurable functions that change as a function of Z_t and which have continuous second derivatives, and ε_t is zero mean white noise with variance $\sigma^2 < \infty$. If $Z_t = X_{t-q}$, then q is the delay parameter where $q < p$ to avoid model unidentifiability.

Harvill and Ray (2006) extended the FCAR idea to the vector autoregressive framework. Define $\mathbf{X}_t = (X_{1,t}, X_{2,t}, \dots, X_{k,t})'$ as a k -variate zero mean vector time series observed at T time points. A vector functional coefficient autoregressive model of order p , VFCAR(p), is given by

$$\mathbf{X}_t = \sum_{j=1}^p \mathbf{f}^{(j)}(\mathbf{Z}_t)\mathbf{X}_{t-j} + \boldsymbol{\varepsilon}_t, \quad t = p+1, p+2, \dots, T, \quad (3.2)$$

where \mathbf{Z}_t is the functional variable of dimension $m \geq 1$ which may consist of exogenous predictors or lagged values of the series \mathbf{X}_t and $\boldsymbol{\varepsilon}_t$ is a $k \times 1$ vector of zero mean white noise with covariance $\boldsymbol{\Sigma}_\varepsilon = \sigma^2 \mathbf{I}$. The matrices $\mathbf{f}^{(j)}(\cdot)$, $j = 1, \dots, p$, are $k \times k$ matrices with elements $[f_{i,l}^{(j)}(\cdot)]$ that are real-valued measurable functions that change as a function of \mathbf{Z}_t and which have continuous second derivatives. If $\mathbf{Z}_t = \mathbf{X}_{t-q}$, then q is the delay parameter where $q < p$ to avoid model unidentifiability. A review of the indentifiability conditions can be found in Huang and Shen (2004) and Harvill and Ray (2006).

3.2.2 Estimation of Functional Coefficient Matrices

A review of methods for fitting the FCAR(p) model, as well as inferential procedures, are found in Fan and Yao (2003). In this section, we will briefly review the local linear methods for fitting the VFCAR(p) model as presented by Harvill and Ray (2006).

Elements of the matrices $\mathbf{f}^{(j)}$, $j = 1, \dots, p$ are functions that are estimated using locally constant or local linear multivariate regression in a neighborhood of \mathbf{Z}_t determined by a specified kernel and bandwidth matrix. Chen and Liu (2001) and Cai, Fan, and Yao (2000) successfully demonstrate these methods in the univariate framework.

At time t , denote the kp -vector of predictors by \mathbf{Y}_t ; that is, let $\mathbf{Y}_t = [\mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-p}]'$, where $\mathbf{X}_t = [X_{1,t-j}, X_{2,t-j}, \dots, X_{k,t-j}]$, for $j = 1, 2, \dots, p$, and let $\mathbf{f}(\mathbf{Z}_t)$ be defined as $\mathbf{f}(\mathbf{Z}_t) = [\mathbf{f}^{(1)}(\mathbf{Z}_t), \mathbf{f}^{(2)}(\mathbf{Z}_t), \dots, \mathbf{f}^{(p)}(\mathbf{Z}_t)]'$. The model in Eq. (3.2) may then be written as

$$\mathbf{X}_t = \mathbf{f}(\mathbf{Z}_t) \mathbf{Y}_t + \boldsymbol{\varepsilon}_t, \quad t = p+1, p+2, \dots, T. \quad (3.3)$$

In our application to PV power plants, we will restrict the dimension of the functional variable to be $m = 1$. Because the elements of $\mathbf{f}^{(j)}(\cdot)$ have continuous second-order derivatives, each $f_{i,l}^{(j)}(\cdot)$ may be approximated locally at z_0 by the linear function $f_{i,l}^{(j)}(z) = \alpha_{il}^{(j)} + \beta_{il}^{(j)}(z - z_0)$. Writing the coefficient matrices in the form $[\boldsymbol{\alpha}|\boldsymbol{\beta}]$, the local linear least squares kernel estimator of $\mathbf{f}(\mathbf{Z}_t)$ is defined as $\hat{\mathbf{f}}(z_0) = \hat{\boldsymbol{\alpha}}$, where $[\hat{\boldsymbol{\alpha}}|\hat{\boldsymbol{\beta}}]$ is the solution to $[\boldsymbol{\alpha}|\boldsymbol{\beta}]$ that minimizes the weighted sum of squares

$$\sum_{t=p+1}^T \left[\mathbf{X}_t - [\boldsymbol{\alpha}|\boldsymbol{\beta}] \begin{pmatrix} \mathbf{Y}_t \\ \mathbf{U}_t \end{pmatrix} \right] \left[\mathbf{X}_t - [\boldsymbol{\alpha}|\boldsymbol{\beta}] \begin{pmatrix} \mathbf{Y}_t \\ \mathbf{U}_t \end{pmatrix} \right]' K_h(\mathbf{Z}_t - z_0). \quad (3.4)$$

Here, \mathbf{U}_t is the result of multiplying the elements of \mathbf{Y}_t by $(\mathbf{Z}_t - z_0)$, K is a specified kernel function, $h > 0$ is the bandwidth, and $K_h(u) = h^{-1}K(u/h)$. The solution to Eq. (3.4) is the least squares problem having

$$\begin{bmatrix} \hat{\boldsymbol{\alpha}} \\ \hat{\boldsymbol{\beta}} \end{bmatrix} = (\mathbf{U}'\mathbf{W}\mathbf{U})' \mathbf{U}'\mathbf{W}\mathbf{Y}, \quad (3.5)$$

assuming $\mathbf{U}'\mathbf{W}\mathbf{U}$ is non-singular, where

$$\mathbf{U} = \begin{bmatrix} \mathbf{Y}_{p+1} & \mathbf{Y}_{p+1}(Z_{p+1} - z_0) \\ \vdots & \vdots \\ \mathbf{Y}_T & \mathbf{Y}_T(Z_T - z_0) \end{bmatrix},$$

$$\mathbf{W} = \text{diag} \{K_h(Z_{p+1} - z_0), \dots, K_h(Z_n - z_0)\}.$$

Thus $\hat{\mathbf{f}}(z_0) = \hat{\boldsymbol{\alpha}}$.

For fitting the model, modified multifold cross-validation is used to determine an optimal bandwidth by finding the value of h that minimize the accumulated prediction error (APE). It is recommended that the autoregressive order p and functional variable \mathbf{Z}_t be chosen based on knowledge of the underlying physical process. Alternatively, each may be found using a data-driven criteria, such as minimizing the APE as a function of p . We refer the reader to Harvill and Ray (2006) for an illustration of the use of the VFCAR model in fitting a simulated vector EXPAR series and a data-driven application to annual temperatures and tree ring widths.

3.2.3 Bootstrap Forecasts

Methods for multi-step forecasting using univariate and vector FCAR models are presented in Harvill and Ray (2005). The authors' bootstrap predictor is shown to be preferred for multi-step prediction using an FCAR model. The bootstrap predictor expands on the naive plug-in predictor of Fan and Yao (2003) by using only within-sample values to estimate the functional coefficients. The predictor evaluates the coefficients at the predicted values.

As in the previous section, we restrict the dimension of our functional variable to $m = 1$ for our application. As will be discussed in Section 3.3, we take Z_t to be a linear combination of lagged values of the k component time series of the original series \mathbf{X}_t . The M -step-ahead predicted values are obtained as

$$\hat{\mathbf{X}}_{t+M} = \sum_{j=1}^p \hat{\mathbf{f}}(\hat{Z}_{t+M}) \hat{\mathbf{X}}_{t+M-j} + \boldsymbol{\epsilon}^b, \quad (3.6)$$

where $\hat{\mathbf{X}}_t = \mathbf{X}_t$ and $\hat{Z}_t = Z_t$ if $t \leq T$. The vector ϵ^b is a bootstrapped vector of the within-sample residuals from the fitted VFCAR model with functional coefficients estimated using local linear multivariate regression. The bootstrap forecast is obtained for $b = 1, \dots, B$ and the average across all bootstrap predictions is used as the M -step-ahead point forecast. Note that should the value of \hat{Z}_{t+M} fall near the boundary or outside of the original range of Z_t the estimated functional coefficient matrix may be unreliable.

3.3 Sensor Design Application

We apply the VFCAR methods of Section 3.2 to the development of a sensor design algorithm for PV power plants. The algorithm makes use of a vector series of solar irradiance measurements in order to determine an optimal number and layout of sensors at a PV plant.

Making use of the methodology discussed in previous section, the sensor design algorithm is as follows:

- (1) For a fixed day, detrend and trim the time-averaged irradiance vector series. Because there is a negligible amount of irradiance prior to sunrise and after sunset, we trim the vector series 30 minutes prior to sunrise and 30 minutes after sunset. The series $\mathbf{X}_t = (X_{1,t}, \dots, X_{K,t})'$ is defined to be the K -dimensional vector series of all detrended and trimmed irradiance measures at K available sensors.
- (2) For a fixed day and number of sensors, $k = 2, \dots, K$, obtain the $\binom{K}{k}$ k -dimensional vector time series $\mathbf{X}_{\ell,t} = (X_{\ell,1,t}, \dots, X_{\ell,k,t})'$, $\ell = 1, \dots, \binom{K}{k}$. In the context of a PV plant, $\mathbf{X}_{\ell,t}$ is the vector series consisting of detrended irradiance measures from the ℓ -th subset of k sensors at time t .

- (3) Define $\nabla X_{t-q} = X_{t-1} - X_{t-q}$ as the q -th difference, or momentum, of a series $\{X_t\}$. Thus, the functional variable $\nabla \bar{X}_{\ell,t-2} = k^{-1} \sum_{i=1}^k \nabla X_{\ell,i,t-2}$, represents the average momentum in the detrended irradiance series at time t .
- (4) Using the first 75% of each sensor combination's vector series estimate the VFCAR(2) model with functional variable $Z_t = \nabla \bar{X}_{\ell,t-2}$.
- (5) For each model obtain three-step-ahead (30-minute-ahead) bootstrap forecasts. Unless otherwise noted, the number of bootstrap replications used is $B = 500$.
- (6) Using the VFCAR forecasts compute the mean prediction error (MPE) for each sensor combination. In the absence of statistical models, it is common to naively forecast irradiance based on the last observed measurement. We also compute the MPE using these naive forecasts. To determine and improvement in prediction using the VFCAR model, we take the ratio

$$\frac{\text{MPE for VFCAR}}{\text{MPE for naive method}} \quad (3.7)$$

A ratio less than one indicates that the VFCAR model performs better at predicting the detrended irradiance than the naive method.

- (7) Repeat Steps 2-6 for each day.
- (8) For each sensor combination, aggregate the ratios of Step 6 across all days.
- (9) Repeat Steps 2-8 for each reasonable k . Typically k is kept small due to the expense of sensors. Federal regulations require a minimum of two monitoring sensors.

3.4 Application: La Ola PV Plant

To illustrate the utility of the algorithm, we apply it to data collected from from a 1.2 MV La Ola PV plant on the island of Lanai, Hawaii. The La Ola plant

is located near the southern end of the island, atop the plateau in a subtropical dry forest climate (U.S. Forest Service, 2011). Designed by Sandia National Laboratories and SunPower Corporation, the La Ola system’s purpose was in part to study the effects of the movement of cloud shadows across the PV array on power output. Although the island is relatively dry due to its position in the rainfall shadow of Maui, measured irradiance shows that cloud movement is a dominating factor in the variability of irradiance (Johnson et al., 2012).

3.4.1 Data

The La Ola PV plant contains three columns and four rows of 12 single-axis tracked arrays. Figure 3.1 displays the latitude and longitude coordinates and label of each sensor at the La Ola PV plant. For one year (January 1, 2010 to December 31, 2010), plane-of-array (POA) irradiance (in W/m^2) is measured at the midpoint of each tracking array in addition to four additional locations at the corners of one central tracking array using LiCor-200 pyranometers. POA irradiance is the measurement when the sensor moves and tracks the sun across the sky. This contrasts global horizontal irradiance (GHI) in which the measurements come from a sensor that is stationary. In order to demonstrate a more general application of our method, we choose to model GHI rather than POA irradiance. A discussion of the methods used to translate measured POA irradiance to estimated GHI irradiance can be found in Patrick et al. (2016).

Little to no variability is observed from one irradiance measurement to the next at one second intervals. One second intervals are too fine of a time-scale for the irradiance process under consideration. As a result of the almost identical measurements from one second to the next, the resulting matrices were unstable and would not invert. Consequently, we investigated time-averages of lengths of 30 seconds, 1 minute, 5 minutes, and 10 minutes. We chose to use a 10-minute average for our methods to reduce computational burdens. The time plots in Figure 3.2 contains the 10-minute

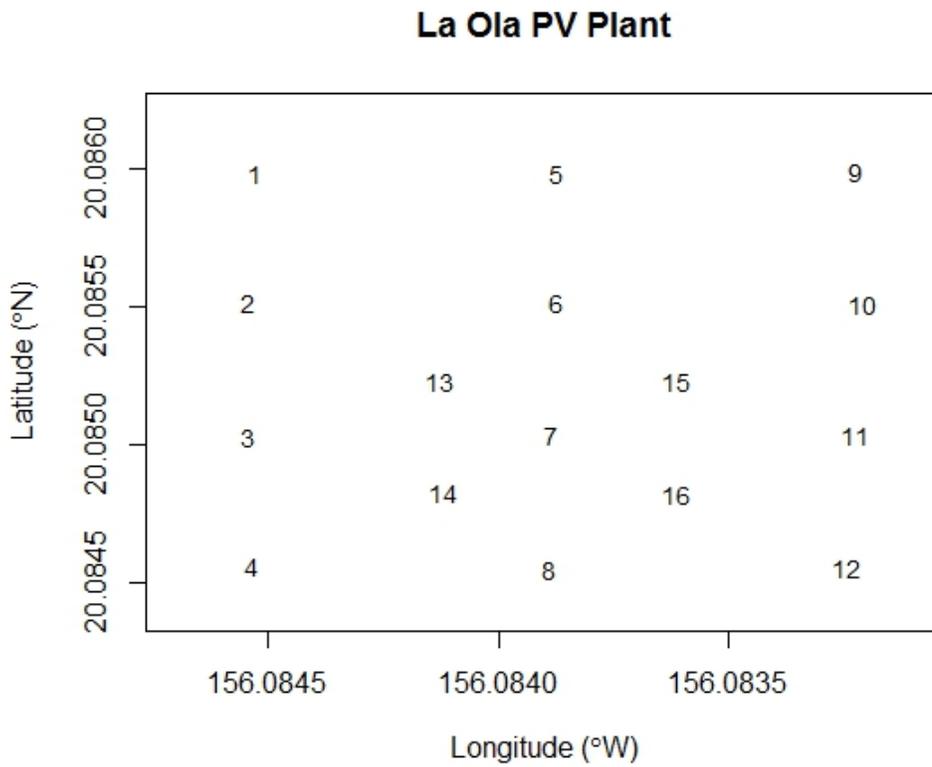


Figure 3.1: Coordinates and labels of the sixteen LiCor-200 pyranometers located at the midpoint of each tracking array and four additional locations for the La Ola PV plant in Lanai, Hawaii.

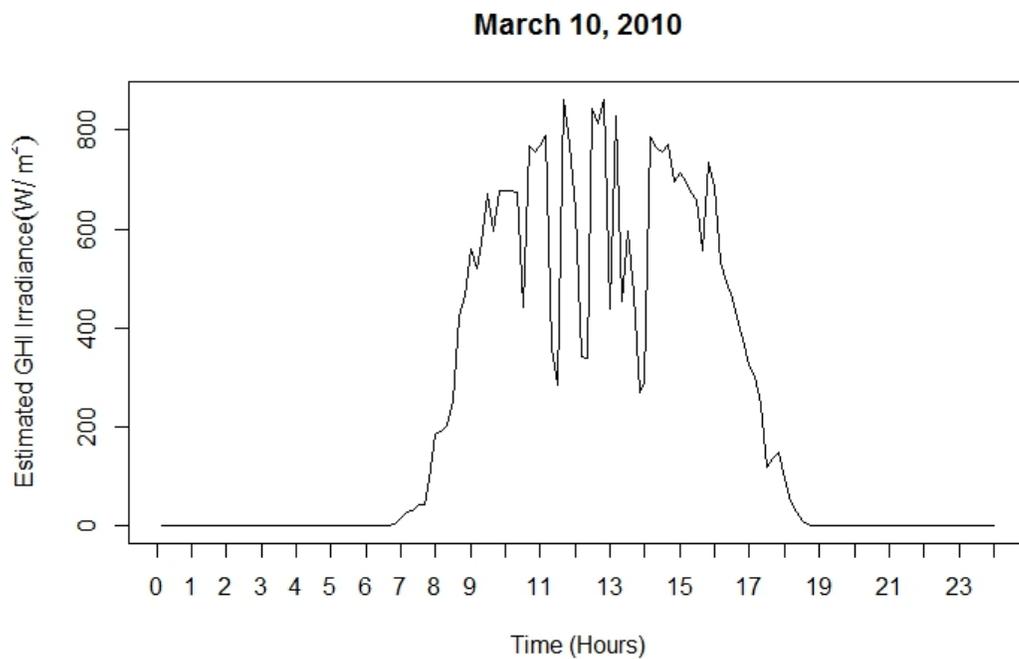
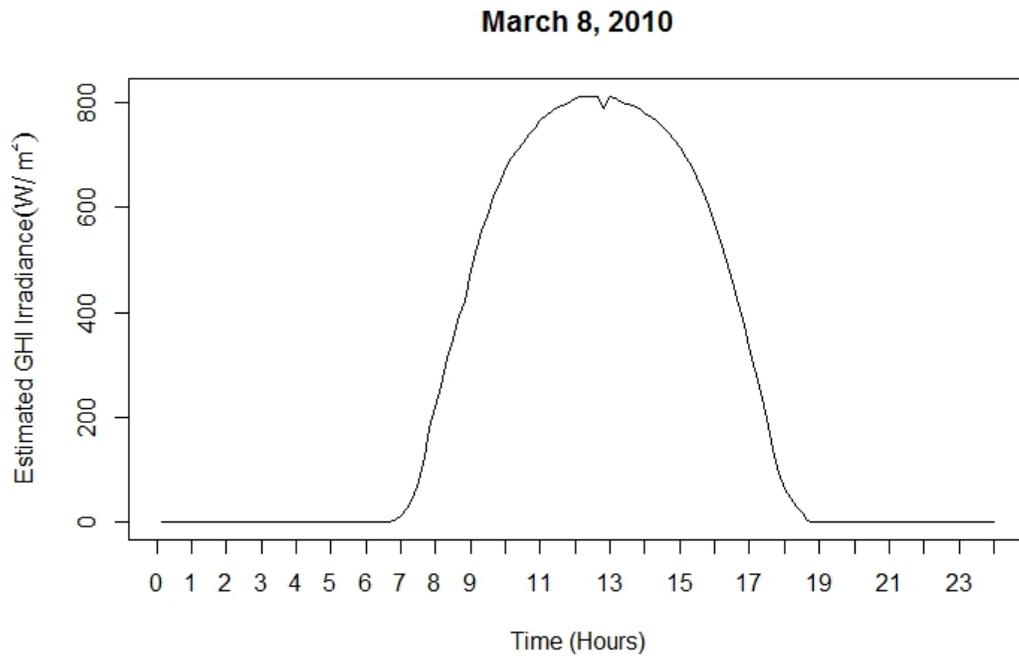


Figure 3.2: Time plots of 10-min averages (solid black) of estimated GHI irradiance measurements at a single sensor for Sensor 1 on March 8, 2010 (top plot) and March 10, 2010 (bottom plot).

time averages of estimated GHI in solid black for a Sensor 1 on March 8, 2010 (top plot) and March 10, 2010 (bottom plot). Visual examination of time plots of estimated GHI would indicate that March 8, 2010, could be considered a clear day due to its smoothly changing irradiance. However, March 10, 2010, could be considered a partly cloud day due to the variation in estimated GHI. The flexibility of the VF-CAR model allows us to model and predict for both types of days without the need to specify different parametric forms for varying weather conditions.

Before the data can be modeled in the design algorithm, it is necessary to remove the diurnal trend. A review of clear-sky models for removing trends in measured GHI can be found in Reno et al. (2012). As noted previously, the Lanai data set consists of measured POA irradiance measurements that is translated to approximate GHI measurements. This transformation is not exact and resulted in the clear-sky models performing poorly in removing the diurnal trend. Consequently, the diurnal trend in the estimated GHI is removed by using local polynomial kernel regression implemented in the `KernSmooth` package (Wand, 2012) in the R programming software.

The top time plots in Figure 3.3 and Figure 3.4 contain the 10-minute time averages of estimated GHI in solid black superimposed with the local polynomial kernel regression estimate in dashed red for March 8, 2010, and March 10, 2010, respectively. The bottom time plots contain the residuals, hereafter referred to as “detrended irradiance,” obtained after removing the diurnal trend.

3.4.2 Results and Discussion

We now illustrate the algorithm proposed in Section 3.3 using the data gathered at the La Ola PV plant. Due to the relatively small size of the plant, we only consider sensor subsets of size $k = 2, 3$, or 4. Table 3.1 provides the sensor combinations which resulted in the lowest median ratio of prediction errors across all days. The mean absolute prediction error (MAPE) is a popular criteria of evaluating forecasts in the

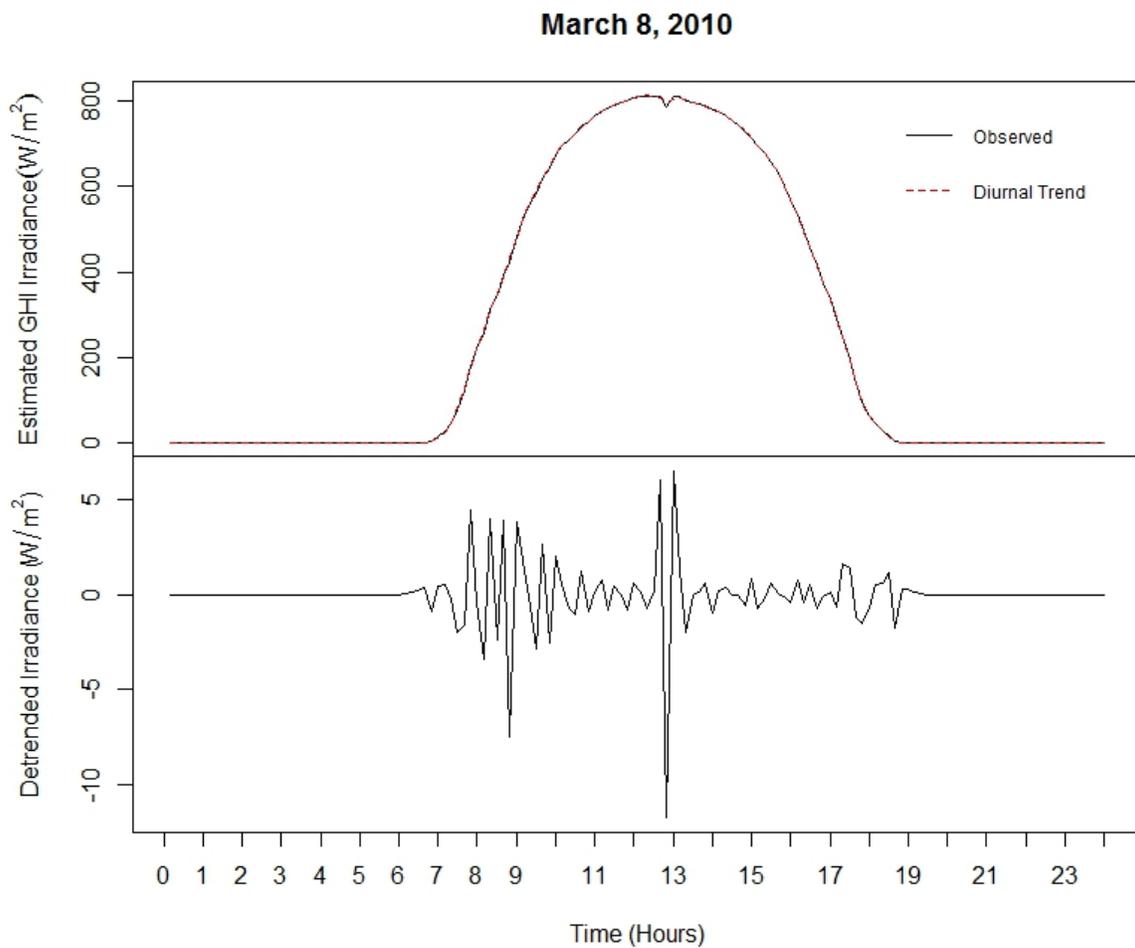


Figure 3.3: Top graph is the time plot of 10-min averages (solid black) of irradiance measurements at a single sensor for March 8, 2010 with the local polynomial kernel estimate (dashed red) superimposed. The bottom plot is transformed irradiance (residuals after using local polynomial kernel regression to remove the diurnal trend).

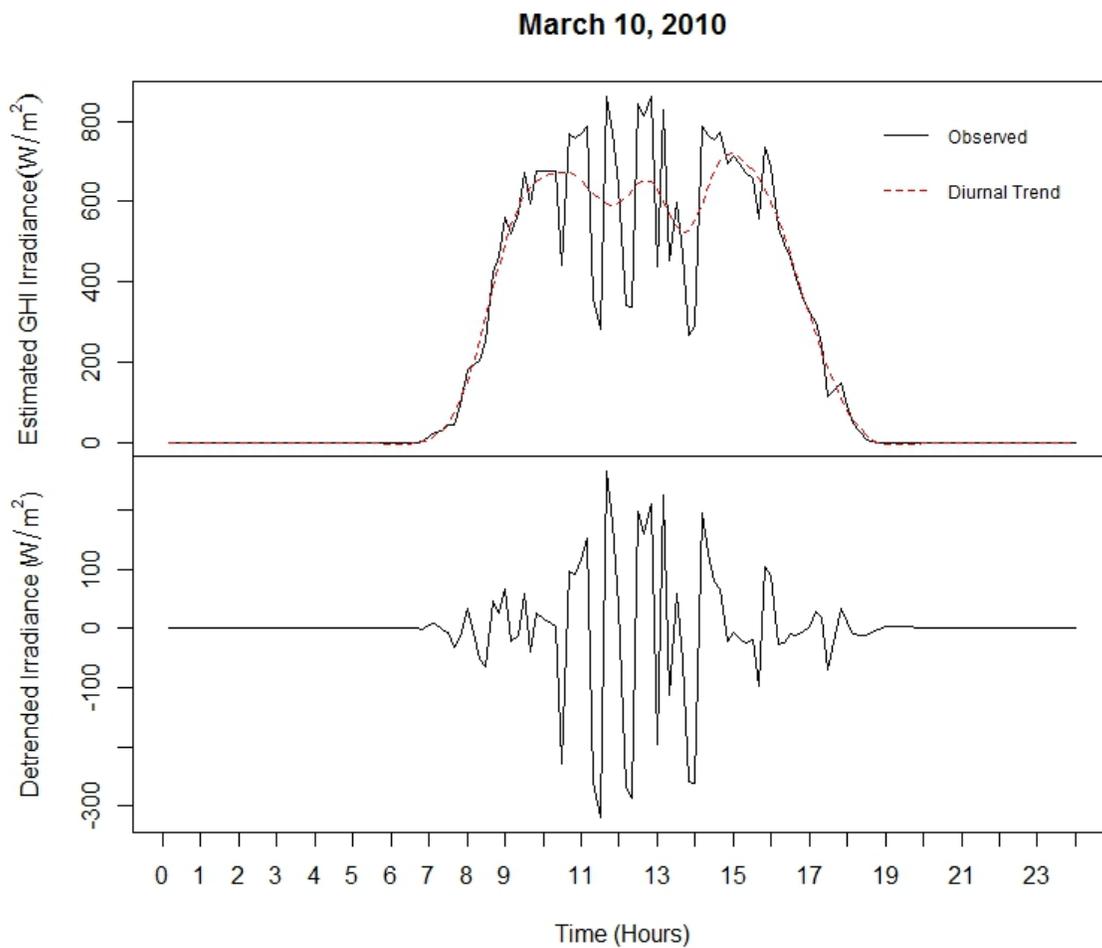


Figure 3.4: Top graph is the time plot of 10-min averages (solid black) of irradiance measurements at a single sensor for March 10, 2010, with the local polynomial kernel estimate (dashed red) superimposed. The bottom plot is transformed irradiance (residuals after using local polynomial kernel regression to remove the diurnal trend).

solar energy literature. We consider both MAPE and the more universal criteria mean squared prediction error (MSPE) in our analysis.

Both absolute and square prediction error ratios are highly skewed for the La Ola data. Tables 3.2 and 3.3 provide summary values for each set of ratios across all days. The skewness in these ratios drives our use of the median as an aggregate measure across all days. A 5% and 10% trimmed mean were also considered; however, the severity of the skewness continued to overwhelm the trimmed mean. The plots in Figure 3.5 display the distributions of the logarithm of prediction error ratios for the five combinations of sensors producing the lowest median ratio. Log-ratios smaller than zero indicate that the VFCAR model outperformed the naive method for 30-minute ahead forecasts of solar irradiance.

Table 3.1 shows that as the number of sensors increases, the ratio of prediction errors also increases, indicating that the increase in model complexity lowers the predicting power versus the naive method. However, multiple sensor combinations of two or three sensors outperform the naive forecasts. Consider the placement of sensors

Table 3.1: Sensor combinations and median of prediction error ratios across all days. The columns are the sensor combinations for sensor subsets of size $k = 2, 3$, and 4. The values in parenthesis are the median ratio of prediction errors across all days. The ratio is calculated as MPE of the VFCAR model divided by the MPE for the naive forecasting method.

| k | 2 | 3 | 4 |
|------|-----------------|---------------------|------------------------|
| MAPE | 4, 9 (0.9086) | 7, 10, 15 (0.9613) | 5, 10, 13, 15 (1.0834) |
| | 6, 13 (0.9094) | 10, 13, 15 (0.9731) | 1, 8, 10, 14 (1.0908) |
| | 3, 6 (0.9097) | 7, 10, 13 (0.9771) | 1, 3, 12, 15 (1.1005) |
| | 7, 8 (0.9108) | 12, 14, 16 (0.9799) | 5, 10, 12, 15 (1.1091) |
| | 11, 12 (0.9121) | 1, 3, 14 (0.9852) | 3, 7, 10, 15 (1.1095) |
| MSPE | 11, 12 (0.7460) | 7, 10, 15 (0.8849) | 2, 5, 15, 16 (1.1968) |
| | 4, 7 (0.7762) | 2, 3, 13 (0.9470) | 7, 10, 14, 15 (1.2033) |
| | 12, 15 (0.7831) | 3, 14, 15 (0.9511) | 2, 10, 11, 15 (1.2399) |
| | 1, 2 (0.7889) | 4, 8, 15 (0.9521) | 1, 2, 8, 13 (1.2461) |
| | 11, 15 (0.7917) | 7, 14, 16 (0.9530) | 3, 11, 13, 16 (1.2499) |

Table 3.2: Summary percentiles of absolute prediction error ratios for the five combinations of $k = 2$ sensors that produced the smallest median absolute prediction error ratios across all days. The ratio is calculated as MAPE of the VFCAR model divided by the MAPE for the naive forecasting method.

| Sensors | 4, 9 | 6, 13 | 3, 6 | 7, 8 | 11, 12 |
|---------|---------|----------|---------|---------|-----------|
| Minimum | 0.1732 | 0.1221 | 0.1493 | 0.1481 | 0.0710 |
| 1% | 0.2070 | 0.1842 | 0.1955 | 0.2016 | 0.1489 |
| 10% | 0.4331 | 0.4092 | 0.4283 | 0.4211 | 0.3770 |
| 25% | 0.6274 | 0.6288 | 0.6242 | 0.6229 | 0.6225 |
| 50% | 0.9086 | 0.9094 | 0.9097 | 0.9108 | 0.9121 |
| 75% | 1.2889 | 1.3334 | 1.2607 | 1.3451 | 1.2825 |
| 90% | 2.0986 | 2.2575 | 2.2748 | 2.6416 | 2.2239 |
| 99% | 9.2666 | 9.9180 | 15.1819 | 8.0972 | 7.3501 |
| Maximum | 21.8608 | 408.8727 | 37.9938 | 17.7292 | 1202.7520 |

Table 3.3: Summary percentiles of square prediction error ratios for the five combinations of $k = 2$ sensors that produced the smallest median square prediction error ratios across all days. The ratio is calculated as MSPE of the VFCAR model divided by the MSPE for the naive forecasting method.

| Sensors | 11, 12 | 4, 7 | 12, 15 | 1, 2 | 11, 15 |
|---------|---------|----------|------------|------------|----------|
| Minimum | 0.0046 | 0.0051 | 0.0312 | 0.0211 | 0.0066 |
| 1% | 0.0356 | 0.0654 | 0.0503 | 0.0456 | 0.1489 |
| 10% | 0.1854 | 0.2224 | 0.2002 | 0.1826 | 0.3770 |
| 25% | 0.3935 | 0.4367 | 0.4314 | 0.4473 | 0.4251 |
| 50% | 0.7460 | 0.7762 | 0.7831 | 0.7889 | 0.7917 |
| 75% | 1.6942 | 1.6601 | 1.5958 | 1.6906 | 1.6343 |
| 90% | 5.6944 | 4.6447 | 4.6166 | 5.2944 | 5.1678 |
| 99% | 84.6108 | 41.5952 | 139.2586 | 8.0972 | 99.1327 |
| Maximum | 5346050 | 523.7162 | 17700.3830 | 12116.1373 | 218.6609 |

7, 10 and 15. There is an approximate 4% reduction in MAPE and an approximate 11.5% reduction in MSPE using the VFCAR model.

With respect to MAPE, 119 of the 120 possible combinations of two sensors outperform the naive method, whereas only 21 of the 560 subsets of three sensors outperform the naive method. None of the subsets of four sensors outperform the naive method. Similarly, with respect to MSPE, all 120 combinations of two sensors outperform the naive method, whereas only 38 of the 560 subsets of three sensors outperform the naive method. Again, none of the subsets of four sensors outperform the naive method.

For predicting solar irradiance at the La Ola PV plant using a VFCAR model, the suggested optimal number of pyranometers is two. In particular, Sensors 11 and 12 performed well both in terms of absolute prediction error and square prediction error. As seen in Figure 3.1 these sensors are located in the southeastern corner of the PV array. This grouping in a single corner is contrary to intuition for predicting

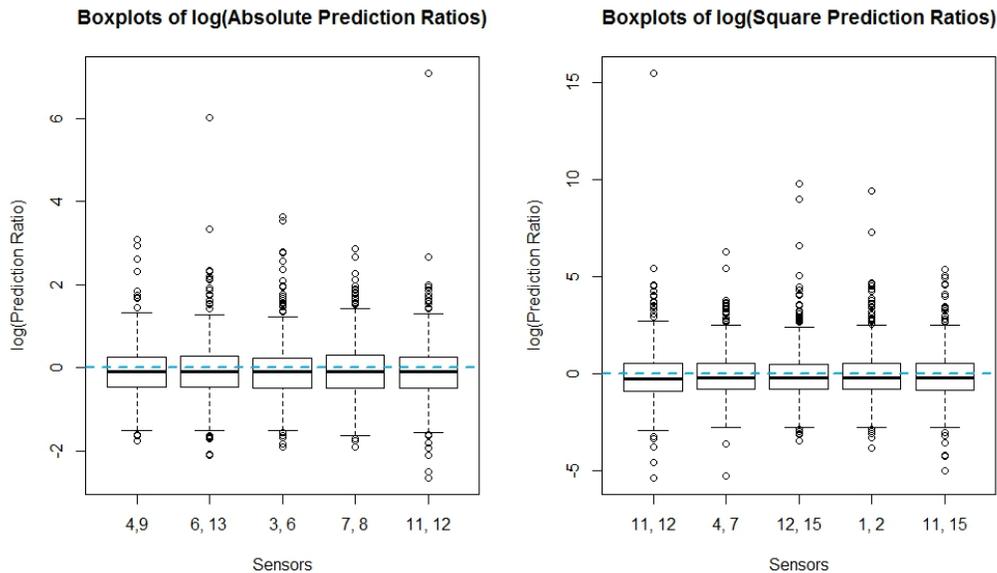


Figure 3.5: Boxplots for the logarithm of the ratio of prediction errors across all days. The ratio is calculated as MPE of the VFCAR model divided by the MPE for the naive forecasting method. The left plot uses mean absolute prediction errors while the right plot uses mean squared prediction errors.

across the entire plant’s footprint. Although the combination of Sensors 4 and 9 performed the best in terms of absolute prediction error, it was in the lower half of square prediction error ratios.

If desired, the combination of Sensors 7, 10, and 15 would be the optimal three-sensor combination. This set of sensors outperforms all other groups of three for both absolute and square prediction error. This subset, as well as other well performing combinations of three sensors, makes use of the more centralized locations when compared to the combination of Sensors 11 and 12.

3.5 Concluding Remarks

We have presented a novel approach to the use of semiparametric time series models in aiding in the determination of an optimal number and layout of sensors at a PV plant. The VFCAR(p) model provides a flexible framework for fitting irradiance data and making predictions. The sensor design algorithm is adaptable to PV plants of much larger size and varying climate regions. For the La Ola PV plant, the algorithm provides many sensor arrangements that are shown to provide more optimal forecasts than the naive method of irradiance forecasting.

The addition of a covariate accounting for a weather condition, such as cloud cover, may be explored in future work. Patrick et al. (submitted for publication) examine the addition of a cloud cover covariate in the univariate framework using an FCAR model fit via spline-backfitted local linear smoothing. This method performs well in both estimation and forecasting. Additionally, we hope to apply the algorithm to to a larger PV system, a system without a regular layout of sensors, and a system located in a different climate zone.

3.6 Acknowledgment

This research was part of research that was performed under contract (PO 1303122) with Sandia National Laboratories, a multi-program laboratory managed

and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

BIBLIOGRAPHY

- Aguiar, R. and Collares-Pereira, M. (1992), “TAG: A time-dependent, autoregressive, Gaussian model for generating synthetic hourly radiation,” *Solar Energy*, 49, 167–174.
- Alam, S., Kaushik, S., and Garg, S. (2006), “Computation of beam solar radiation at normal incidence using artificial neural network,” *Renewable Energy*, 31, 1483–1491.
- Bacher, P., Madsen, H., and Nielsen, H. A. (2009), “Online short-term solar power forecasting,” *Solar Energy*, 83, 1772–1783.
- Birkel, C., Paroli, R., Spezia, L., Dunn, S. M., Tetzlaff, D., and Soulsby, C. (2012), “A new approach to simulating stream isotope dynamics using Markov switching autoregressive models,” *Advances in Water Resources*, 46, 20–30.
- Cai, Z., Fan, J., and Yao, Q. (2000), “Functional-coefficient regression models for nonlinear time series,” *Journal of the American Statistical Association*, 95, 941–956.
- Caner, M. and Hansen, B. E. (2001), “Threshold autoregression with a unit root,” *Econometrica*, 69, 1555–1596.
- Chan, K. S. (1993), “Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model,” *The Annals of Statistics*, 520–533.
- Chan, K. S. and Tsay, R. S. (1998), “Limiting properties of the least squares estimator of a continuous threshold autoregressive model,” *Biometrika*, 85, 413–426.
- Chen, R. (1996), “A nonparametric multi-step prediction estimator in Markovian structures,” *Statistica Sinica*, 603–615.
- Chen, R. and Liu, L.-M. (2001), “Functional coefficient autoregressive models: estimation and tests of hypotheses,” *Journal of Time Series Analysis*, 22, 151–173.
- Chen, R. and Tsay, R. S. (1993), “Functional-coefficient autoregressive models,” *Journal of the American Statistical Association*, 88, 298–308.
- Clements, M. P. and Smith, J. (1997), “The performance of alternative forecasting methods for SETAR models,” *International Journal of Forecasting*, 13, 463–475.
- Coakley, J. and Fuertes, A.-M. (2002), “Asymmetric dynamics in UK real interest rates,” *Applied Financial Economics*, 12, 379–387.

- Di Narzo, A., Aznarte, J., and Stigler, M. (2008), “`tsDyn`: Time series analysis based on dynamical systems theory,” *R package version 0.6-0*, URL <http://CRAN.R-project.org/package=tsDyn>.
- Dorale, J. A., Edwards, R. L., Ito, E., and Gonzalez, L. A. (1998), “Climate and vegetation history of the midcontinent from 75 to 25 ka: a speleothem record from Crevice Cave, Missouri, USA,” *Science*, 282, 1871–1874.
- Driese, S. G., Li, Z.-H., Cheng, H., Harvill, J. L., and Sims, J. (2016), “16. High-resolution rainfall records for middle and late Holocene based on speleothem annual UV fluorescent layers integrated with stable isotopes and U/Th dating, Raccoon Mountain Cave, Tennessee, USA,” *Geological Society of America Special Papers*, 516, 231–246.
- Edwards, R. L., Chen, J., and Wasserburg, G. (1987), “ ^{238}U - ^{234}U - ^{230}Th - ^{232}Th systematics and the precise measurement of time over the past 500,000 years,” *Earth and Planetary Science Letters*, 81, 175–192.
- Enders, W. (2008), *Applied Econometric Time Series*, John Wiley & Sons.
- Enders, W. and Granger, C. W. J. (1998), “Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates,” *Journal of Business & Economic Statistics*, 16, 304–311.
- Enders, W. and Siklos, P. L. (2001), “Cointegration and threshold adjustment,” *Journal of Business & Economic Statistics*, 19, 166–176.
- Fairchild, I. J. and Baker, A. (2012), *Speleothem Science: From Process to Past Environments*, vol. 3, John Wiley & Sons.
- Fairchild, I. J., Smith, C. L., Baker, A., Fuller, L., Spötl, C., Matthey, D., McDermott, F., et al. (2006), “Modification and preservation of environmental signals in speleothems,” *Earth-Science Reviews*, 75, 105–153.
- Fan, J. and Yao, Q. (2003), *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer Verlag.
- Ghayekhloo, M., Ghofrani, M., Menhaj, M., and Azimi, R. (2015), “A novel clustering approach for short-term solar radiation forecasting,” *Solar Energy*, 122, 1371–1383.
- Ghofrani, M., Ghayekhloo, M., and Azimi, R. (2016), “A novel soft computing framework for solar radiation forecasting,” *Applied Soft Computing*, 48, 207–216.
- Glasbey, C. (2001), “Non-linear autoregressive time series with multivariate Gaussian mixtures as marginal distributions,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50, 143–154.
- Gooijer, J. D. (1998), “On threshold moving-average models,” *Journal of Time Series Analysis*, 19, 1–18.

- Granger, C. W., Terasvirta, T., et al. (1993), “Modelling non-linear economic relationships,” *OUP Catalogue*.
- Haggan, V. and Ozaki, T. (1981), “Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model,” *Biometrika*, 68, 189–196.
- Hansen, B. E. et al. (1997), “Inference in TAR models,” *Studies in Nonlinear Dynamics and Econometrics*, 2, 1–14.
- Harvill, J. L. and Ray, B. K. (2005), “A note on multi-step forecasting with functional coefficient autoregressive models,” *International Journal of Forecasting*, 21, 717–727.
- (2006), “Functional coefficient autoregressive models for vector time series,” *Computational Statistics & Data Analysis*, 50, 3547–3566.
- Huang, J. Z. and Shen, H. (2004), “Functional Coefficient Regression Models for Non-linear Time Series: A Polynomial Spline Approach,” *Scandinavian Journal of Statistics*, 31, 515–534.
- Kuszamaul, S., Ellis, A., Stein, J., and Johnson, L. (2010), “Lanai high-density irradiance sensor network for characterizing solar resource variability of MW-scale PV system,” in *Photovoltaic Specialists Conference (PVSC), 2010 35th IEEE*, IEEE, pp. 000283–000288.
- Li, B., Nychka, D. W., and Ammann, C. M. (2010), “The value of multiproxy reconstruction of past climate,” *Journal of the American Statistical Association*, 105, 883–895.
- Li, Z. H., Driese, S. G., and Cheng, H. (2014), “A multiple cave deposit assessment of suitability of speleothem isotopes for reconstructing palaeo-vegetation and palaeo-temperature,” *Sedimentology*, 61, 749–766.
- Ling, S., Tong, H., Li, D., et al. (2007), “Ergodicity and invertibility of threshold moving-average models,” *Bernoulli*, 13, 161–168.
- Ling, S. et al. (1999), “On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model,” *Journal of Applied Probability*, 36, 688–705.
- McDermott, F. (2004), “Palaeo-climate reconstruction from stable isotope variations in speleothems: a review,” *Quaternary Science Reviews*, 23, 901–918.
- Mora-Lopez, L. and Sidrach-de Cardona, M. (1998), “Multiplicative ARMA models to generate hourly series of global irradiation,” *Solar Energy*, 63, 283–291.
- Oudjana, S. H., Hellal, A., and Mahammed, I. H. (2014), “Neural network based photovoltaic electrical forecasting in south Algeria,” *Applied Solar Energy*, 50, 273–277.

- Patrick, J., Harvill, J., and Sims, J. (2015), “Spline-backfitted kernel forecasting for functional-coefficient autoregressive models,” *arXiv preprint arXiv:1502.03486*.
- Patrick, J. D., Harvill, J. L., and Hansen, C. W. (2016), “A semiparametric spatio-temporal model for solar irradiance data,” *Renewable Energy*, 87, 15–30.
- Pilatowska, M., Włodarczyk, A., and Zawada, M. (2015), “The Environmental Kuznets Curve in Poland—Evidence from Threshold Cointegration Analysis,” *Dynamic Econometric Models*, 14, 51–70.
- Reno, M. J., Hansen, C. W., and Stein, J. S. (2012), “Global horizontal irradiance clear sky models: implementation and analysis,” *SANDIA Report SAND2012-2389*.
- Springer, G. S., Rowe, H. D., Hardt, B., Edwards, R. L., and Cheng, H. (2008), “Solar forcing of Holocene droughts in a stalagmite record from West Virginia in east-central North America,” *Geophysical Research Letters*, 35.
- Springer, G. S., White, D. M., Rowe, H. D., Hardt, B., Mihimdukulasooriya, L. N., Cheng, H., and Edwards, R. L. (2009), “Multiproxy evidence from caves of Native Americans altering the overlying landscape during the late Holocene of east-central North America,” *The Holocene*.
- Tingley, M. P., Craigmile, P. F., Haran, M., Li, B., Mannshardt, E., and Rajaratnam, B. (2012), “Piecing together the past: statistical insights into paleoclimatic reconstructions,” *Quaternary Science Reviews*, 35, 1–22.
- Tingley, M. P. and Huybers, P. (2010), “A Bayesian algorithm for reconstructing climate anomalies in space and time. Part I: Development and applications to paleoclimate reconstruction problems,” *Journal of Climate*, 23, 2759–2781.
- Tong, H. (1978), *On a Threshold Model*, Sijthoff & Noordhoff, Amsterdam.
- (1983), “A note on a delayed autoregressive process in continuous time,” *Biometrika*, 710–712.
- (1990), *Non-linear Time Series: A Dynamical System Approach*, Oxford University Press.
- (2012), *Threshold Models in Non-linear Time Series Analysis*, vol. 21, Springer Science & Business Media.
- Tong, H. and Lim, K. S. (1980), “Threshold autoregression, limit cycles and cyclical data,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 245–292.
- Tsay, R. S. (1998), “Testing and modeling multivariate threshold models,” *Journal of the American Statistical Association*, 93, 1188–1202.

- U.S. Forest Service (2011), *Life Zone Map of Lanai'i*, http://www.fs.fed.us/psw/topics/ecosystem_processes/tropical/restoration/lifezone/hawaii/Lanai.jpg.
- Vincent, L. and Morley, B. (2012), “Asymmetric Adjustment and Intervention in the UK Housing Market,” *Bath Economics Research Papers*.
- Walker, M., Gibbard, P., Berkelhammer, M., Bjorck, S., Cwynar, L., Fisher, D., Long, A., Lowe, J., Newnham, R., Rasmussen, S., et al. (2014), “Formal Subdivision of the Holocene Series/Epoch,” in *STRATI 2013*, Springer, pp. 983–987.
- Wand, M. and Ripley, B. (2006), “**KernSmooth**: Functions for kernel smoothing for Wand & Jones (1995),” *R package version*, 2, 22–19.
- Wang, J. and Yang, L. (2009), “Efficient and fast spline-backfitted kernel smoothing of additive models,” *Annals of the Institute of Statistical Mathematics*, 61, 663–690.
- Wang, L., Yang, L., et al. (2007), “Spline-backfitted kernel smoothing of nonlinear additive autoregression model,” *The Annals of Statistics*, 35, 2474–2503.