

The Pennsylvania State University
The Graduate School
College of Earth and Mineral Sciences

**MARKOV-SWITCHING MODELS FOR PROBABILISTIC SOLAR
RESOURCE ASSESSMENT AND FORECASTING**

A Dissertation in
Energy and Mineral Engineering
by
Vivek Srikrishnan

© 2018 Vivek Srikrishnan

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2018

The dissertation of Vivek Srikrishnan was reviewed and approved* by the following:

Jeffrey R. S. Brownson
Associate Professor of Energy & Mineral Engineering
Dissertation Advisor, Chair of Committee

George S. Young
Professor of Meteorology

Mort D. Webster
Associate Professor of Energy Engineering

Uday V. Shanbhag
Gary and Sheila Chaired Professor of Industrial and Manufacturing Engineering

Luis F. Ayala H.
Professor of Petroleum and Natural Gas Engineering
Head of the Graduate Program

*Signatures are on file in the Graduate School.

Abstract

This work proposes and analyzes a Markov-switching autoregression model structure for joint probabilistic modeling of the beam and global components of solar irradiance, which are important to simulate the performance of a variety of solar energy conversion devices, including solar photovoltaics. The ability of this model to assess the hourly solar resource is tested, using both a version of the model that is calibrated using all-year data and a version of the model that combines individual seasonally-calibrated models. While this simple model does not fully capture the behavior of the solar resource, an analysis of the posterior predictive distribution reveals strategies for improvement. A version of this model is also used to forecast both irradiance components for a 15-minute lead time while assimilating geostationary satellite data into an inhomogeneous transition probability specification. The inhomogeneous specifications produce sharper predictive distributions than an analogous homogeneous model, but all have similar skill relative to a smart persistence forecast.

Table of Contents

List of Figures	viii
List of Tables	xi
List of Symbols	xii
Acknowledgments	xv
Chapter 1	
Introduction	1
Chapter 2	
Statistical Modeling of Solar Irradiance	6
2.1 Solar Radiation Components	6
2.1.1 Beam Irradiance	6
2.1.2 Diffuse Irradiance	7
2.1.3 Global Irradiance	8
2.2 Dimensionless Quantification of the Solar Resource	9
2.3 Distributions of Irradiance Components	9
2.4 Empirical Correlations Between Irradiance Components	13
2.5 Statistical Time Series Models for Solar Irradiance Components . .	14
2.5.1 Time Series Modeling of the Solar Resource	14
2.5.2 Autoregressive Time Series Models	19
2.5.3 Synthetic Generation of Solar Irradiance Series	20
2.5.4 Forecasting of Solar Irradiance Series	23
2.5.5 Contribution of This Dissertation	24
Chapter 3	
A Markov-Switching Vector Autoregressive Model for Solar Irradiance	26

3.1	Modeled Variables	26
3.2	Regime-Switching in Solar Irradiance Series	26
3.3	Markov Chains	28
3.4	Markov-Switching Models	30
3.4.1	MSAR Models	30
3.4.2	MSVAR Models	32
3.5	Model Definition	32
3.5.1	Variable Transformation	32
3.5.2	Model Specification	33
Chapter 4		
	Bayesian Inference for MSVAR Models	36
4.1	Bayesian Inference	36
4.2	Sampling from the Posterior Distribution	37
4.3	Identifiability	38
4.4	Likelihood and Prior Specification	39
4.5	MCMC Sampler for Markov-Switching Models	40
4.6	Sampling State Transition Probabilities	42
4.6.1	Time-Homogeneous Transition Probabilities	42
4.6.2	Time-Inhomogeneous Transition Probabilities	42
4.6.2.1	Data augmentation for the RUM expression	43
4.6.2.2	Data augmentation for the dRUM expression	44
4.6.2.3	Sampling scheme for the dRUM	45
4.7	Sampling the Model Parameters	46
4.8	Sampling the State Indicators	48
4.9	Model Selection	50
4.9.1	Predictive Information Criteria	51
4.9.2	“True” Data Generation Criteria	52
4.9.2.1	Loss Functions and Marginal Likelihoods	52
4.9.2.2	Approximations to the Marginal Likelihood	53
Chapter 5		
	Hourly Solar Resource Assessment Modeling	55
5.1	Data	55
5.2	Model Specification	58
5.3	Results	59
5.3.1	Model Selection	59
5.3.2	Identifiability Constraints	60
5.3.3	Posterior Predictive Checks	61
5.4	Discussion	66

Chapter 6	
Forecasting Components of Solar Irradiance	69
6.1 Data	70
6.1.1 Satellite Data	70
6.1.2 Irradiance Data	71
6.2 Statistical Model	72
6.2.1 Model Specification	72
6.2.2 Transition Probability Covariates	73
6.2.3 Missing Data	74
6.2.4 Missing Data Sampling Scheme	77
6.2.5 Prior Distributions	78
6.3 Forecast Validation Metrics	78
6.4 Results	81
6.4.1 Model Selection	81
6.4.2 Identifiability Constraints	82
6.4.3 Forecast Performance	82
6.5 Discussion	86
Chapter 7	
Conclusions and Future Work	87
Appendix A	
Bridge Sampling	89
A.1 Unsupervised Construction of an Importance Density	91
A.2 Relative Standard Errors	92
Appendix B	
Resource Assessment Posterior Distribution Plots	93
B.1 All-Year Model	93
B.2 Winter Model	96
B.3 Spring Model	99
B.4 Summer Model	102
B.5 Fall Model	105
Appendix C	
Forecast Posterior Distribution Plots	108
C.1 Homogeneous Model	108
C.2 Inhomogeneous (Current) Model	111
C.3 Inhomogeneous (Taylor) Model	115

List of Figures

1.1	Components of solar radiation.	2
1.2	Diffuse fraction vs. clearness index and Erbs model fit.	4
2.1	Marginal and joint distributions of hourly-averaged k_t and k_b values	10
2.2	Marginal and joint distributions of 15-minute averaged k_t and k_b values	11
2.3	Temporal and spatial scales of weather phenomena	18
3.1	Selected daily time series for k_t and k_b	27
3.2	Graphical representation of a time-homogeneous Markov chain	29
3.3	Digraph representation of a realization of a Markov-switching autoregressive model	31
3.4	Selected daily time series for k_t and k_b demonstrating the trend under clear sky conditions for an entire day. Plot i) is a 15-minute averaged series, while plot ii) is an hourly-averaged series	35
5.1	Distributions of hourly k_t by month	56
5.2	Distributions of hourly k_b by month	57
5.3	Distributions of hourly k_t and k_b by season	57
5.4	Point process representation of the posterior samples	61
5.5	Posterior predictive check of hourly k_t and k_b densities	62
5.6	Posterior predictive check of hourly GHI and DNI densities	63
5.7	Posterior predictive check of daily GHI and DNI densities	64
5.8	Posterior predictive check of k_t and k_b auto- and cross-correlations .	65
5.9	Posterior predictive check of GHI and DNI auto- and cross-correlations	66
6.1	Results of the cloud masking procedure for a sample GOES image .	71
6.2	Frozen-field forecast flowchart	73
6.3	Example of advection vector estimation	74
6.4	Number and percentage of missing satellite images by hour of the day	75
6.5	Pairs plots for covariates and their lagged values	76

6.6	Marginal calibration plots for the three probabilistic models. The red crosses are observations, and the black dots are forecast ensemble means.	83
6.7	Boxplot of forecast residuals	85
6.8	Marginal CRPS plotted against observed values.	85
B.1	Posterior distributions of the innovation covariance Σ components for the all-year resource assessment model.	93
B.2	Posterior distributions of the regression coefficients β for the all-year resource assessment model.	94
B.3	Posterior distributions of the transition probabilities ξ for the all-year resource assessment model.	95
B.4	Posterior distributions of the innovation covariance Σ components for the winter component of the seasonal resource assessment model.	96
B.5	Posterior distributions of the regression coefficients β for the winter component of the seasonal resource assessment model.	97
B.6	Posterior distributions of the transition probabilities ξ for the winter component of the seasonal resource assessment model.	98
B.7	Posterior distributions of the innovation covariance Σ components for the spring component of the seasonal resource assessment model.	99
B.8	Posterior distributions of the regression coefficients β for the spring component of the seasonal resource assessment model.	100
B.9	Posterior distributions of the transition probabilities ξ for the spring component of the seasonal resource assessment model.	101
B.10	Posterior distributions of the innovation covariance Σ components for the summer component of the seasonal resource assessment model.	102
B.11	Posterior distributions of the regression coefficients β for the summer component of the seasonal resource assessment model.	103
B.12	Posterior distributions of the transition probabilities ξ for the summer component of the seasonal resource assessment model.	104
B.13	Posterior distributions of the innovation covariance Σ components for the fall component of the seasonal resource assessment model.	105
B.14	Posterior distributions of the regression coefficients β for the fall component of the seasonal resource assessment model.	106
B.15	Posterior distributions of the transition probabilities ξ for the fall component of the seasonal resource assessment model.	107
C.1	Posterior distributions of the innovation covariance Σ components for the homogeneous forecast model.	108

C.2	Posterior distributions of the regression coefficients β for the homogeneous forecast model.	109
C.3	Posterior distributions of the transition probabilities ξ for the homogeneous forecast model.	110
C.4	Posterior distributions of the innovation covariance Σ components for the inhomogeneous forecast model using current transition covariates.	111
C.5	Posterior distributions of the regression coefficients β for the inhomogeneous forecast model using current transition covariates.	112
C.6	Posterior distributions of the transition probability covariates κ for the inhomogeneous forecast model using current transition covariates.	113
C.7	Posterior distribution of the mixture covariates α^I for the inhomogeneous forecast model using current transition covariates.	113
C.8	Posterior distribution of the missing-data imputation coefficients α^z for the inhomogeneous forecast model using current transition covariates.	114
C.9	Posterior distribution of the missing-data imputation innovation covariance Σ^z for the inhomogeneous forecast model using current transition covariates.	114
C.10	Posterior distributions of the innovation covariance Σ components for the inhomogeneous forecast model using Taylor transition covariates.	115
C.11	Posterior distributions of the regression coefficients β for the inhomogeneous forecast model using Taylor transition covariates.	116
C.12	Posterior distributions of the transition probability covariates κ for the inhomogeneous forecast model using Taylor transition covariates.	117
C.13	Posterior distribution of the mixture covariates α^I for the inhomogeneous forecast model using Taylor transition covariates.	117
C.14	Posterior distribution of the missing-data imputation coefficients α^z for the inhomogeneous forecast model using Taylor transition covariates.	118
C.15	Posterior distribution of the missing-data imputation innovation covariance Σ^z for the inhomogeneous forecast model using Taylor transition covariates.	118

List of Tables

4.1	Parameters for an approximation of the logistic density by a 6-component mixture of normal densities	46
5.1	Prior distributions for the resource model	58
5.2	Estimates of the model marginal log-likelihood	59
5.3	Estimates of the standard error of the marginal log-likelihood estimates	60
6.1	Prior distributions for the forecast model	78
6.2	WAIC values for the forecast models	82
6.3	Univariate validation metrics for the forecast models	84
6.4	Multivariate validation metrics for the forecast models	84

List of Symbols

Abbreviations & Acronyms

ANN Artificial Neural Network, [page 16](#)

AR Autoregressive Model, [page 19](#)

ARCH Autoregressive Conditional Heteroskedasticity, [page 19](#)

ARIMA Autoregressive-Integrated-Moving Average Model, [page 19](#)

ARMA Autoregressive-Moving Average Model, [page 19](#)

BHI Beam Horizontal Irradiance, [page 7](#)

CRPS Continuous Ranked Probability Score, [page 79](#)

DHI Diffuse Horizontal Irradiance, [page 7](#)

DNI Direct Normal Irradiance, [page 6](#)

dRUM difference random utility model, [page 43](#)

DS determinant sharpness, [page 79](#)

EE Euclidean error, [page 80](#)

ES Energy Score, [page 80](#)

GARCH Generalized Autoregressive Conditional Heteroscedasticity, [page 23](#)

GHI Global Horizontal Irradiance, [page 8](#)

GVAR GOES Variable Format, [page 70](#)
 HMM Hidden Markov Models, [page 20](#)
 i.i.d. Independent and Identically Distributed, [page 19](#)
 MA Moving Average Model, [page 19](#)
 MAE Mean Absolute Error, [page 80](#)
 MSAR Markov-switching autoregressive model, [page 30](#)
 MSVAR Markov-Switching Vector Autoregressive Model, [page 32](#)
 POA Plane-of-Array, [page 8](#)
 PV Photovoltaic, [page 3](#)
 SECS Solar Energy Conversion System, [page 6](#)
 TMY Typical Meteorological Year, [page 14](#)
 TOA Top-of-Atmosphere, [page 7](#)

Notation

η Markov transition probabilities, [page 20](#)
 $\mathcal{D}(\cdot; \cdot)$ Dirichlet distribution, [page 42](#)
 $\mathcal{N}(\cdot, \cdot)$ multivariate normal distribution, [page 32](#)
 $\mathcal{W}^{-1}(\cdot, \cdot)$ Inverse-Wishart distribution, [page 47](#)
 θ_z solar zenith angle, [page 3](#)
 ξ Markov chain transition probability, [page 28](#)
 k_b beam index, [page 7](#)
 k_D diffuse ratio, [page 9](#)

k_d	diffuse horizontal index, page 7
k_t	clearness index, page 8
$p(\cdot)$	probability distribution or density, page 28
S	Markov state, page 28

Acknowledgments

This dissertation would not have been possible without the support and input of many people. My advisor, Dr. Jeffrey Brownson, allowed me wide latitude in finding and shaping this work, and was there to support and guide me when necessary. I learned a lot from all of my committee members, Dr. George Young, Dr. Mort Webster, and Dr. Uday Shanbhag, all of whom have not only shaped the work I have done over the last few years, but have also given me invaluable research and career advice. Many aspects of this work were also heavily influenced or inspired by conversations with Dr. Mesude Bayracki-Boz while she was working on her Ph.D. I also must thank Dr. Klaus Keller, from whom I learned a lot about applied statistics and modeling, in addition to receiving more general research and career advice.

Dr. Ben Shaby and Benjamin Seiyon Lee, from the Penn State Department of Statistics, also played important roles in the completion of this dissertation. Both provided me with invaluable statistical advice. The members of the Center for Sustainable Climate Risk Management at Penn State both provided moral and research support. These people include Dr. Robert Nicholas, Dr. Tony Wong, Dr. Caitlin Spence, Dr. Jared Oyler, Dr. Giacomo Marangoni, Ben Lee, Robert Fuller, Kelsey Ruckert, and Joel Roop-Eckhart.

The departmental administrative staff in Energy & Mineral Engineering made my graduate experience seamless, and I cannot extend to them enough gratitude and appreciation for the hard work they put in on my behalf.

Finally, no list of acknowledgements would be complete without thanking my family. My parents, Kris and Jaya, and my sister, Maya, have always supported me and pushed me to continue to learn. It should (but won't) go without saying that I owe an incredible debt of gratitude to my wife, Julie, who has been incredibly patient and supportive. And lastly, Whitey, James, and Harry always allowed me to relax with them, distracted me when I needed to be distracted, and at times "helped" with programming.

This dissertation was partially supported by the National Science Foundation

through the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507 and the Penn State Center for Climate Risk Management. Any conclusions and recommendations expressed in this material are those of the author and do not necessarily reflect the views of the funding agency.

Chapter 1 |

Introduction

The proportion of electricity generated in the United States from solar energy is increasing rapidly. As of 2016, the United States had more than 56,000 GWh of net electricity generation from solar energy (including both utility-scale and small-scale generation), almost double the amount of solar energy generation in 2014 (EIA, [2017](#), Table 1.1A). This growth is not expected to slow down in the near future: the state of California is aiming for 33% renewable generation by 2020, and 38 states have renewable standards or targets; Pennsylvania is targeting 0.5% solar photovoltaic generation by 2020-21) (Durkay, [2017](#)).

The increasing number of solar projects require methods to accurately quantify solar resource uncertainty and variability. This need may be driven by financial risk (McMahan et al., [2013](#)) or power plant or system simulations for planning or operations (Lave et al., [2013](#); Dall’Anese et al., [2015](#)). Forecast uncertainty is the result of intermittency in the solar resource, which is predominately driven by cloud transients. Resource variability occurs on time scales ranging from inter-annually (Müller et al., [2014](#)) to seasonally and monthly (Liu and Jordan, [1960](#); Bendt et al., [1981](#)) to hourly and sub-hourly (Curtright and Apt, [2008](#)) to minutes and seconds (Ela et al., [2013](#)). Variability on longer scales (monthly through inter-annually) is the result of climatological variability as well as variations in the solar cycle, while variability on shorter scales is the result of the stochastic nature of cloud field evolution and other sources of atmospheric losses (Calinoiu et al., [2014](#)).

Both solar resource variability and uncertainty present system planners with challenges. Resource variability leads to fluctuations in generated power, increasing financial risk (McMahan et al., [2013](#)), as well as introducing constraints on power

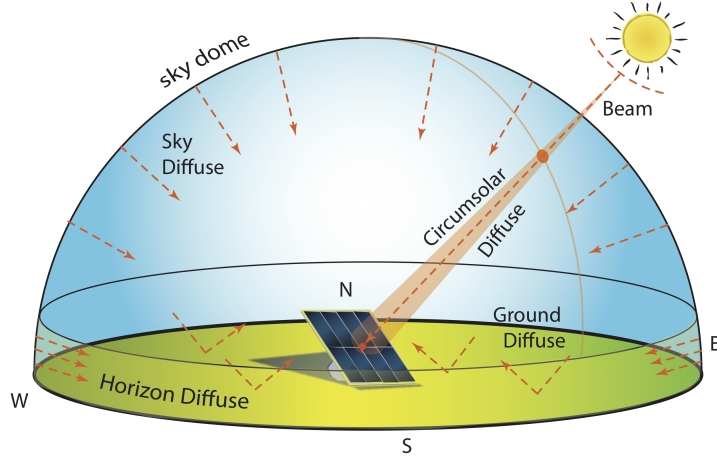


Figure 1.1: The components of solar radiation incident on a PV panel, including beam (direct) radiation and radiation diffused by multiple sources. Reproduced from Brownson (2014).

dispatchability that are unique from those related to stock-type resources such as coal (Von Meier, 2011). Forecast uncertainty increases power systems operating and production costs due to the need to have a sufficient amount of backup generating capacity available to compensate for shortfalls (Ela et al., 2013; Eber and Corbus, 2013).

The availability of the solar resource is important for disciplines other than electric power generation, due to the fundamental energetic role of solar radiation in the earth’s natural processes. Many traditional aspects of human societies were shaped by a need to take advantage of solar energy for applications such as heating and agriculture (Perlin, 2013), and buildings serve as large solar collectors, with solar radiation serving as a primary driver for cooling and heating loads (Witmer, 2014). Plant sciences rely upon solar resource modeling to understand photosynthesis and optocalorics (*e.g.* Gates, 1968; Ashenfelter and Storchmann, 2006; Bois et al., 2008). As a result, solar resource assessments and forecasts which capture the variability and uncertainty of the resource across multiple time scales are essential for simulations across many disciplines. Indeed, as will be seen in Chapter 2, probabilistic solar resource models are becoming more common.

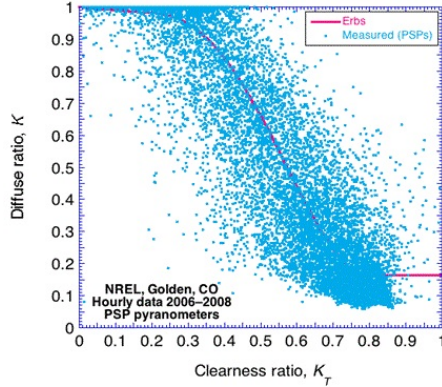
The available solar resource at the surface can be broadly divided into *beam* (or *direct*) and *diffuse* radiation, though as shown in Figure 1.1 “diffuse” radiation

can come from many different sources. The total of the available beam and diffuse radiation is referred to as *global* radiation. The fraction of global irradiance attributable to diffuse sources is the *diffuse fraction* or *diffuse ratio*. The difference between the beam and diffuse components of solar radiation incident upon a surface is relevant for many applications. While solar photovoltaics (PV) can convert beam (or *direct*) and diffuse radiation (the combination of the two incident upon a surface is referred to as *global* irradiance), they are typically tilted at an angle, making the relationship between the tilt angle and the solar zenith angle θ_z relevant for calculating the beam irradiance incident on the PV surface as well as the radiation reflected by the PV’s glass coating.

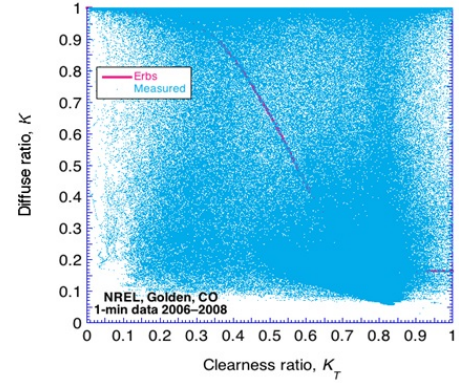
For many applications, including solar concentrating technologies and plant optocalorics, beam irradiance is the primary relevant component for their solar resource conversion. However, as will be documented in Chapter 2, beam irradiance is often modeled in a way which either collapses the joint probability distribution of global and direct irradiance onto the marginal distribution of global irradiance or models the availability of direct irradiance as a separate process. The separation of global and direct irradiance modeling is largely motivated by differences in application areas. For example, while solar PV convert global irradiance, concentrating solar technologies can only make use of the direct component. However, this approach results in models which are less useful for application-agnostic solar resource modeling.

While the limitations of separating direct and global irradiance modeling are hopefully clear, modeling beam irradiance as largely deterministic conditional on global irradiance is extremely limited. Gueymard (2010) plotted the relationship between global irradiance and the diffuse fraction on hourly and one-minute time scales (Figure 1.2). While the Erbs et al. (1982) diffuse-sky model (based on an empirical correlation) captures the general trend on an hourly scale (though a large amount of diffuse fraction variability is left unexplained), there is no visible relationship between the two variables on a one-minute scale. Therefore, the joint variability of the two components can not be properly incorporated into an analysis depending on simulations of global and direct irradiance based on similar methods.

This dissertation uses bivariate probability models to jointly model global and direct irradiance (or, more specifically, non-dimensional versions of these variables). These models, and similar models, can facilitate the use of Monte



(a) Diffuse fraction vs. clearness index for hourly data.



(b) Diffuse fraction vs. clearness index for one-minute data.

Figure 1.2: Plots of the diffuse fraction vs. the clearness index (the ratio of surface-level global irradiance to top-of-the-atmosphere global irradiance; see Section 2.1) for hourly and one-minute averaged data, along with the empirical relationship between the diffuse fraction and clearness index quantified by the Erbs et al. (1982) diffuse-sky model. Reproduced from Gueymard (2010).

Carlo simulation methods for application-specific analyses that require global and direct irradiance as resource inputs. As far as we know, this is the first work to probabilistically model both global and direct irradiance as marginal components of a higher-dimensional stochastic process. We do this using Markov-switching models (defined in Chapter 3), which can probabilistically capture the uncertainty associated with cloud-regime intermittency. Further, this dissertation uses Bayesian methods to tune and calibrate these models, allowing us to quantify uncertainty on the level of model parameters. The Bayesian framework also allows for the explicit incorporation of meteorological knowledge via prior distributions into the model selection and calibration processes. The remainder of this dissertation is structured as follows.

- **Chapter 2 – Statistical Modeling of Solar Irradiance:**

Chapter 2 provides an overview of the state of statistical (as opposed to physical) modeling of the solar resource available at the surface as well as more formal definitions of solar irradiance components. The current state-of-the-art methods, as well as their evolution, are described along with their limitations.

- **Chapter 3 – A Markov-Switching Vector Autoregression Model for**

Solar Irradiance:

Chapter 3 defines the model used in this dissertation and provides background on Markov chains and Markov-switching autoregression models.

- **Chapter 4 – Bayesian Inference for MSVAR Models:**

Chapter 4 discusses Bayesian inference and provides algorithms for a Gibbs sampler for Markov-switching vector autoregression models.

- **Chapter 5 – Hourly Solar Resource Assessment Modeling:**

Chapter 5 uses a Markov-switching autoregressive model to model the solar resource at Rock Springs, PA on an hourly time scale. We compare the fit of a single model fit to all-year data to that of a model which models each season separately.

- **Chapter 6 – Forecasting Components of Solar Irradiance:**

Chapter 6 uses the Markov-switching autoregressive model framework to forecast 15-minute-ahead irradiance components. We compare a forecasting model driven by a time-homogeneous Markov chain to one with an underlying time-inhomogeneous Markov chain, where the time-varying transition probabilities are based on covariates derived from geostationary satellite images. We also compare the performance of these models against a persistence benchmark.

- **Chapter 7 – Conclusions and Future Work:**

Chapter 7 concludes this dissertation by discussing themes which unify the problems of probabilistic solar resource assessment and forecasting, as well as proposing future research directions which can capitalize on the methods used in this work.

Chapter 2 | Statistical Modeling of Solar Irradiance

2.1 Solar Radiation Components

Solar radiation is either absorbed, reflected, or refracted/transmitted as it travels through the atmosphere. This results in the radiation arriving at the surface either through a direct beam or after being scattered by a number of sources (see Figure 1.1). Decomposition of total irradiance into various components is critical for anisotropic sky models as well as for solar energy conversion systems (SECS) which involve only a single component (such as concentrating solar technologies or plant optocalorics). In this section, we will discuss the components of the solar resource in more detail.

2.1.1 Beam Irradiance

Beam irradiance is typically represented as *direct normal irradiance* (DNI). DNI is the beam component of solar radiation incident on a surface facing the sun, or the beam component normal to the earth's surface. DNI is of particular interest to concentrating solar technologies, but it is also of interest more generally. For example, as solar PV devices typically have glass surfaces, it is necessary to understand the angle of incidence of beam radiation relative to the PV surface to compute the amount of radiation reflected by the PV surface.

Directly measuring DNI is relatively expensive, as only the radiation within a small angle of the center of the solar disc must be measured, requiring pyrheliometers

to be mounted on sun-tracking devices (see Srikrishnan (2015, Section 1.1) for a brief overview of solar irradiance measurement procedures and costs, and Sengupta et al. (2015) for more details and best practices).

To remove the diurnal trend from measurements of DNI (often for the purposes of time-series analysis), normalization with respect to top-of-atmosphere (TOA) radiation results in the direct normal clearness index (or direct normal transmittance index) k_n . As the other solar components are often quantified with respect to a horizontal surface, it can be useful to measure beam irradiance in terms of *beam horizontal irradiance* (BHI), via the relationship

$$\text{BHI} = \text{DNI} \times \cos \theta_z, \quad (2.1)$$

where θ_z is the solar zenith angle, Normalizing BHI with respect to TOA radiation results in the beam index k_b ,

$$k_b = \text{BHI}/G_0. \quad (2.2)$$

Tovar-Pescador (2008) refers to k_b as the “direct index” (we use the terminology “beam index” to avoid confusion with the DNI-based k_n , as used in Sengupta et al. (2015)).

2.1.2 Diffuse Irradiance

The diffusion of radiation can be represented isotropically or anisotropically (Brownson, 2014). The isotropic diffuse model combines an isotropic sky model with an estimate for the radiation diffused by the ground (Liu and Jordan, 1960), while anisotropic models account for the different contributions of various sky sources of diffusion, such as circumsolar and horizon diffuse radiation (Perez et al., 1986).

Diffuse irradiance is typically represented as *diffuse horizontal irradiance* (DHI). DHI is the total diffuse irradiation incident on a horizontal surface, regardless of the mechanism of diffusion. While DHI doesn’t include ground-diffused radiation (unless this radiation is further reflected back towards the horizontal collector) , it integrates contributions from multiple diffusing mechanisms in the sky dome. DHI normalized by TOA irradiance is the diffuse horizontal index k_d ,

$$k_d = \text{DHI}/G_0, \quad (2.3)$$

as compared to the diffuse fraction k_D ,

$$k_D = DHI/GHI.$$

The amount of diffusion by the sky dome (the magnitude of k_d) depends on a number of factors, including atmospheric turbidity and the types and arrangement of clouds.

2.1.3 Global Irradiance

Global irradiance is most often measured as *global horizontal irradiance*, or GHI, which is the total irradiance incident on a horizontal surface, regardless of whether the radiation is direct or diffuse. GHI is the sum of BHI and DHI, or as more commonly represented,

$$GHI = DHI + DNI \times \cos \theta_z. \quad (2.4)$$

As solar PV devices can convert both direct and diffuse radiation, GHI is a standard input for modeling PV performance. GHI is also relatively easy to measure at low cost, and GHI measurements are more commonly available than DNI or DHI measurements. As a result, when DNI and/or DHI measurements are desired (such as for plane-of-array (POA) measurements of global irradiance on a tilted collector), they are often derived from GHI measurements or forecasts using empirical correlations (see Section 2.4).

The clearness index k_t is GHI normalized by TOA radiation. Some authors use k_t to refer to the hourly-averaged clearness index and K_t to refer to the daily-averaged clearness index, but in this dissertation we use k_t to denote the clearness index regardless of data-averaging period, largely for the sake of consistency. Another common normalization is the clear-sky index k_{cs} , where GHI is normalized by the estimated clear-sky GHI (using a clear-sky model such as those surveyed by Gueymard (2012a)). One advantage of using the clear-sky GHI as a normalizing factor is that it eliminates non-stationarity due to variations in clear-sky optical depth.

2.2 Dimensionless Quantification of the Solar Resource

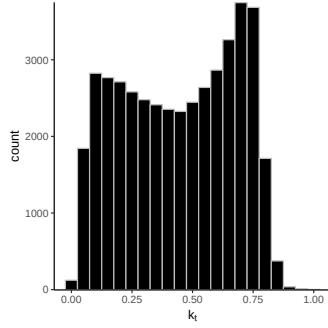
Quantification of the available surface-level solar resource requires knowledge of GHI, DNI, and DHI. Due to the relation expressed by Equation (2.4), values for two of these three components are sufficient if combined with the solar zenith angle θ_z . Ultimately, we are interested in expressing the amount of attenuation and scattering that occurs as the TOA radiation travels through the atmosphere. The instantaneous TOA irradiance G_0 is known as a function of locale (Brownson, 2014). The standard set of state variables (following Liu and Jordan (1960)) is $\{G_0, \theta_z, k_t, k_D\}$, where $k_D = \text{DHI}/\text{GHI}$ is the diffuse ratio.

However, when treating GHI, DNI, and DHI as random variables, the use of k_D is not ideal as both elements in the ratio are unknown quantities (in particular, there is a dependence of k_D on k_t). k_D can be written as the ratio k_t/k_d , or equivalently $k_t/(k_t - k_b)$. As k_t and k_b are independent quantities (we use k_b instead of k_d as we are directly interested in DNI), the state variable set $\{G_0, \theta_z, k_t, k_b\}$ is preferable for modeling. Moreover, k_t and k_b are non-dimensional, with their dimensional variants GHI and BHI depending only on the deterministic quantities G_0 and θ_z . This allows us to model only the stochastic components of the downwelling radiative transfer process (k_t and k_b), rather than those emerging from sun-observer geometry (G_0 and θ_z).

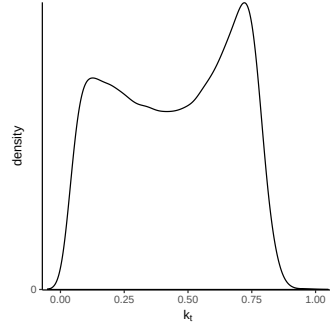
A more formal argument of the sufficiency of k_t and k_b is possible using the Buckingham π theorem (Buckingham, 1914). By the argument in the preceding paragraph, the set of dimensional state variables $\{G_0, \theta_z, \text{GHI}, \text{DNI}\}$ characterizes the overall downwelling radiative transfer process (from the top of the atmosphere to surface-level). There are two dimensions, W and m (M and L in MLT units), and so $4 - 2 = 2$ non-dimensional variables are sufficient to describe the system. While the Buckingham π theorem is not constructive, it is clear that k_t and k_b are two non-dimensional quantities that characterize the overall downwelling attenuation and scattering processes.

2.3 Distributions of Irradiance Components

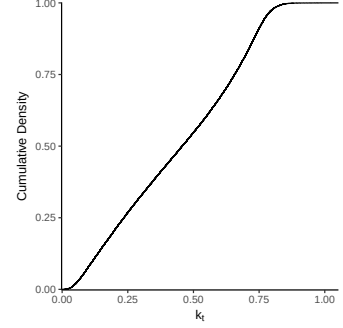
As might be expected, the length of time over which irradiance values are averaged will affect their distribution (Muneer et al., 2004). Figures 2.1 and 2.2 show several



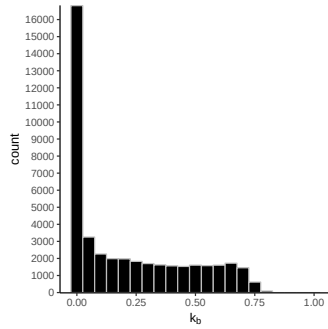
(a) Histogram of hourly-averaged k_t values.



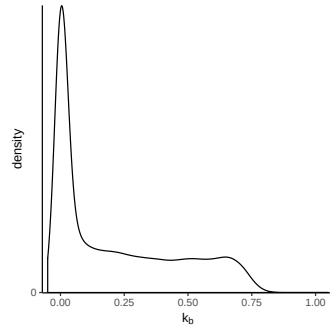
(b) Kernel density estimate of hourly-averaged k_t values.



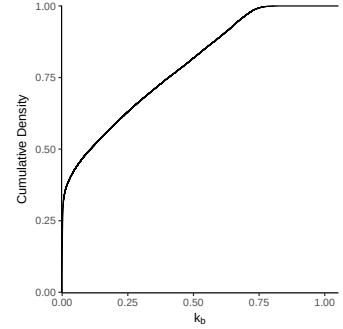
(c) Empirical CDF of hourly-averaged k_t values.



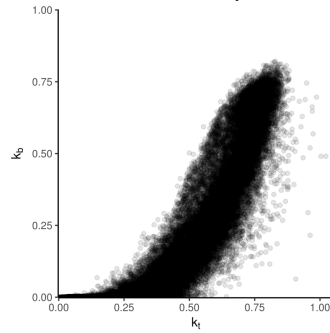
(d) Histogram of hourly-averaged k_b values.



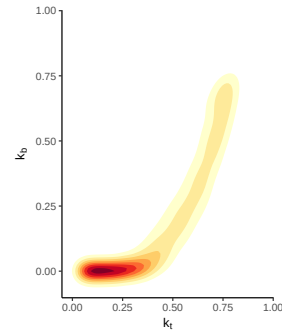
(e) Kernel density estimate of hourly-averaged k_b values.



(f) Empirical CDF of hourly-averaged k_b values.

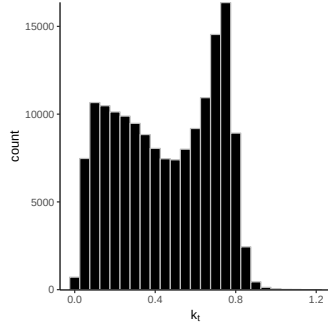


(g) Scatterplot of hourly-averaged k_t vs. k_b values.

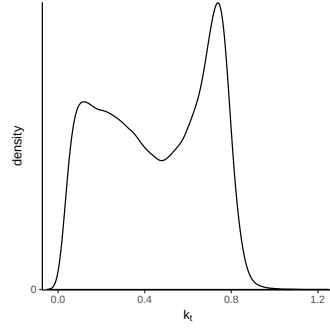


(h) Kernel density estimate of the joint distribution of hourly-averaged k_t and k_b values.

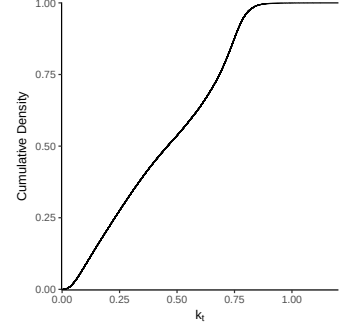
Figure 2.1: Marginal and joint distributions of hourly-averaged k_t and k_b values. Data collected at the SURFRAD site at Rock Springs, PA, USA from January 1, 2006 through December 31, 2017.



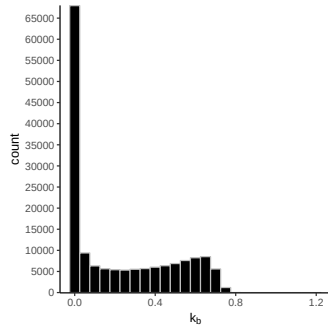
(a) Histogram of 15 minute-averaged k_t values.



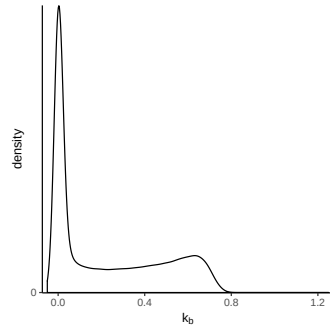
(b) Kernel density estimate of 15-minute averaged k_t values.



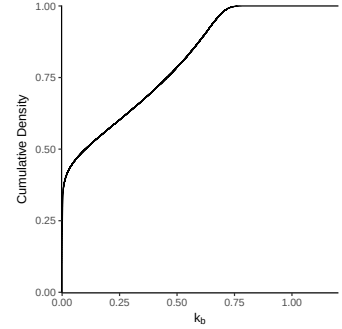
(c) Empirical CDF of 15-minute averaged k_t values.



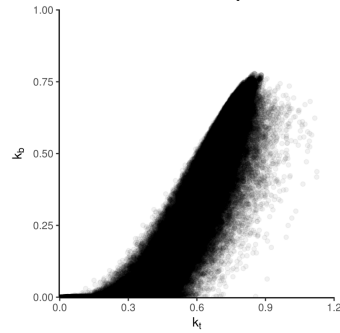
(d) Histogram of 15-minute averaged k_b values.



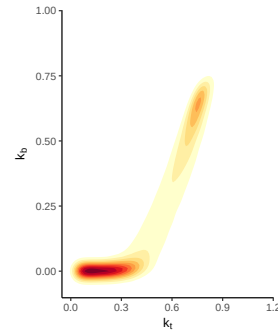
(e) Kernel density estimate of 15-minute averaged k_b values.



(f) Empirical CDF of 15-minute averaged k_b values.



(g) Scatterplot of 15-minute averaged k_t vs. k_b values.



(h) Kernel density estimate of the joint distribution of 15-minute averaged k_t and k_b values.

Figure 2.2: Marginal and joint distributions of 15-minute averaged k_t and k_b values. Data collected at the SURFRAD site at Rock Springs, PA, USA from January 1, 2006 through December 31, 2017.

visualizations of the marginal and joint distributions for hourly- and 15-minute-averaged k_t and k_b , respectively, from the SURFRAD site at Rock Springs, PA. Both sets of marginal distributions are clearly multi-modal, with modes that could roughly be described as “cloudy” and “clear.” It is important to note, however, that distributions with two modes can be generated by mixtures with more than two components, and so it might be a mistake to assume that these distributions ought to be generated by two unimodal distributions.

The scatterplots in Figures 2.1g and 2.2g show that the 15-minute k_t distribution has a longer right tail than the hourly distribution. There is also more k_t variability associated with each k_b slice over the shorter averaging period, and shorter averages also result in a higher density for larger-valued combinations of k_t and k_b (roughly corresponding to $k_t = k_b = 0.75$; compare Figures 2.1h and 2.2h). Thus, characterizations of the distribution of k_t and k_b values depend on the time period over which the data is averaged.

The pioneering work of Liu and Jordan (1960) discovered that the distribution of daily k_t values over a month, conditioned on the monthly average, are effectively similar across locations. A similar relationship between hourly k_t values and the daily average was also observed. Hollands and Huget (1983) developed unimodal probability distributions for hourly k_t conditioned on the daily average k_t . Aguiar and Collares-Pereira (1992a) clustered distributions for hourly k_t into various ranges of daily k_t averages, and found that these distributions were non-stationary throughout each day with respect to the solar zenith angle.

Stuart and Hollands (1988) examined the probability distribution of the hourly beam index k_b . This distribution was found to be a mixture of a point distribution at $k_b = 0$ and another component for $0 < k_b \leq 0.8$, with the shape of the latter component depending on the daily mean. This combination of a spike at $k_b = 0$ and another distribution with support on $0 < k_b \leq 0.8$ is also seen in the marginal distribution of k_b in Figures 2.1 and 2.2. Skartveit and Olseth (1992) modeled probability distributions for k_b and k_t for 5 minute-averaged data.

2.4 Empirical Correlations Between Irradiance Components

Due to the ease of measuring GHI and the relative expense of measuring DNI and DHI, statistical models (commonly referred to as empirical correlations) are often used to translate between GHI values (measured or modeled) and DNI and/or DHI values. Liu and Jordan (1960) introduced correlations between k_t and the diffuse fraction $\text{DHI/GHI} = k_d/k_t$ for clear and cloudy conditions. Over time, many polynomial and logistic fits have been proposed between the diffuse fraction and k_t . The Erbs et al. (1982) fit is shown in Figure 1.2. Others (including Skartveit et al. (1998), Boland et al. (2013), and Engerer (2015)) use additional variables to capture some of the variability seen in Figure 1.2. Torres et al. (2010), in a review of several models for the diffuse fraction, noted that models which include an autoregressive component perform better than polynomial or logistic models. Not surprisingly, models which incorporate more of the system dynamics (and therefore capture more of the diffuse fraction variability) perform better than those which deterministically map k_t values to diffuse fraction values (Torres et al., 2010; Engerer, 2015).

Several models also exist for estimating k_b directly from k_t (Skartveit and Olseth, 1987; Maxwell, 1987; Louche et al., 1991; Perez et al., 1992). Batlles et al. (2000) found that the correlations tended to perform better under clear conditions with a high solar elevation.

One problem with generalizing many empirical correlations is that they tend to be developed for a relatively spatially limited set of data, when the prevailing meteorological conditions (and therefore the radiative transfer patterns) in the locations used to fit the model may differ from those in the location of interest. Models which incorporate additional covariates may be able to cope with this difficulty so long as the training data set is representative, but the operational use of these models then requires additional streams of data, which may or may not be available.

2.5 Statistical Time Series Models for Solar Irradiance Components

2.5.1 Time Series Modeling of the Solar Resource

Time series models of the solar resource are generally produced for one of two purposes, which might be more broadly described as *planning* and *operations*. Planning (*e.g.* sizing, assessment of risk, and integrations planning) requires modeled data which has the statistical features of the “true” solar resource. Operations (*e.g.* maximum power-point tracking, economic dispatch, and unit commitment), on the other hand, is based on forecasted values of anticipated solar irradiance.

A common class of modeled data sets is the Typical Meteorological Year (TMY), which is intended to capture the “average” properties of the meteorological conditions, including solar irradiance, at the location of interest based on several years of data. TMY datasets were originally intended for use in building heating and cooling load models (Sengupta et al., 2015). Klein et al. (1976) demonstrated that the appropriateness of “typical” or “average” datasets as inputs depends on the sensitivity of the SECS to variations in the available solar resource. Indeed, Sengupta et al. (2015) state that TMY datasets are not appropriate for optimizing SECS performance, and that while many computational models use TMY datasets to predict the typical performance of a SECS, this was not the original purpose for which TMY datasets were designed. Indeed, for many simulation objectives, such as financial risk assessment or electric system planning, “typical” behavior may be no more important than characterizing extremes in convertible solar irradiance (particularly on the low end of the distribution), and it is critical to capture variability on the time scales of interest (Vignola et al., 2012).

Knight et al. (1991) classifies synthetic weather generation schemes into two classes. The generation of a TMY-type “representative” year is Type II in this classification. The aim of this type of data set is to contain within one simulated year sufficient variability to simulate SECS performance in realistic conditions with relatively limited computational expense. Type I schemes involve the stochastic generation of many data sets which, taken in aggregate, reproduce the long-run statistics of the observed record. The latter approach is less problematic today due to increases in computational power, as simulating and analyzing SECS perfor-

mance over many years of observed or synthetic data is not generally prohibitively expensive.

One possible objection to the goal of replicating the long-run statistics of the observed record is that there may be many locations in which there is not a sufficiently long observed record for the frequentist statistics to be known with appropriate levels of confidence. Vignola et al. (2012) observe that for DNI, 30 years of high-quality measured data was required in the Pacific Northwest to gain confidence in a statistically significant trend, though Liu and Jordan (1960) point out that this length depends on location. In any case, it is difficult to know *a priori* how much data will be required. It is important to consider that the observed record consists of a series of realizations of what may be considered to be a stochastic process, particularly when data is relatively limited. This dissertation adopts the Bayesian perspective that the goal of synthetic generation ought to be to quantify the uncertainty surrounding the observed record, rather than trying to strictly replicate the statistics of the observed record (this point is discussed further in Chapter 4). This approach allows SECS performance to be probabilistically evaluated against an ensemble of replications, all of which are consistent with the observed data.

Operations and management of SECS requires forecasts of the expected observed future values of solar irradiance, rather than the synthetic (hypothetical) series used by simulations intended for planning purposes. This is because while planning requires an understanding of how the SECS can be expected to behave over longer time scales, operations work is aimed at understanding how the SECS will behave over shorter time scales, over which changing meteorological conditions can be forecasted. In one sense, the distinction between the data sets required by planning simulations and those required by operations simulations can be expressed as the difference between climatology and meteorology; planning/climatology is concerned with the longer-term variability and behavior of the system at a particular locale, while operations/meteorology is focused with the current and upcoming system state and behavior.

While the general structure of irradiance models may be similar regardless of purpose, the selection of a particular model, calibration, and the metrics used for verification may depend on the intended purpose; we will explicitly consider these factors in Chapters 3, 5, and 6. In the following review of the solar modeling litera-

ture, we will distinguish between those models which are aimed at the generation of synthetic series and those which aimed at the generation of forecasted series (with the recognition that some work has been useful for both purposes).

Starting from datasets of meteorological inputs, physical models using explicit or emulated radiative transfer code will resolve beam attenuation and scattering directly, but are often computationally expensive, and running them multiple times may be computationally prohibitive. Statistical characterizations of the solar resource are therefore often used to model solar resource availability or forecast irradiance. Most of these models can be broadly divided into two classes: machine learning models (a class which includes artificial neural networks (ANN) and support vector machines) and parametric time-series models, though there are some methods which cross over between the two methods, and some which include methods that are not clearly classifiable in either category.

In this section, we review the literature as it pertains to solar irradiance modeling and forecasting, with a view towards identifying the current state-of-the-art and motivating the models that are used in this dissertation. While ANNs (see *e.g.* Bishop (1995) for an overview) have been a popular tool due to their ability to mine complex patterns from large datasets (*e.g.* Elizondo et al., 1994; Mellit et al., 2006; Mellit and Pavan, 2010; Marquez and Coimbra, 2011; Wang et al., 2009; Linares-Rodriguez et al., 2013; Marquez et al., 2013; Marquez and Coimbra, 2013; McCandless et al., 2013; Yang et al., 2014; Chu et al., 2014; Pedro and Coimbra, 2015; McCandless et al., 2016b; Aguiar et al., 2016; Gutierrez-Corea et al., 2016; Chu et al., 2016; Chen and Kartini, 2017; Monjoly et al., 2017), we focus our attention on models involving parametric time series models, for the following reasons:

1. Parametric models have been observed to outperform ANNs for GHI forecasts over time horizons up to one hour, though this likely depends on the choice of parametrization and which effects are represented by the model structure (Reikard, 2009; Reikard et al., 2017; David et al., 2016).
2. Parametric models allow for more direct assimilation of prior information about the problem into the model. In general, this is through the specification of the likelihood (through model selection), though Bayesian modeling procedures allow for prior information to be included through the prior

distributions.

3. One reason to model irradiance (rather than modeling PV power, or any other SECS output) is that irradiance may be of more scientific interest and provides the inputs for a conversion simulation (such as the System Advisor Model (SAM) (Blair et al., 2014)). Due to the complex structure of ANN models and their black-box nature, it is also not necessarily clear how to interpret the coefficients. Parametric models are more “transparent,” in the sense that the parameters (and their uncertainties) are more readily interpretable. It is also clearer how the decisions made in the modeling and calibration process affect the fit of parametric models. ANNs, on the other hand, may be sensitive to choices in the training procedure, and those choices do not necessarily provide guidance in terms of how to improve the model.
4. Probabilistic forecasts quantify the uncertainty in a prediction, which is an essential for “optimal” (up to the choice of loss function) decision-making procedures (Robert, 2007). While not all stochastic decision methods utilize uncertainties to the same level of detail (an overview of some methods is provided in Zheng et al. (2015)), more informative quantifications can be “downscaled” to the level appropriate for a particular decision algorithm (see *e.g.* Mehrotra and Papp (2013)), while less informative quantifications cannot have more information added to them. It is less clear how to approach probabilistic modeling and forecasting with ANNs, as each ANN represents a deterministic function. Marquez and Coimbra (2011) obtained prediction intervals for their ANN forecasts under the assumption that the residuals were Gaussian-distributed with mean zero; another ANN was trained to estimate the standard deviation of the residuals. A unique approach is the model tree estimation of GHI variability by McCandless et al. (2015). While both of these uncertainty forecasting methods are an improvement over Quantile regression neural networks (Cannon, 2011) have not been used in the solar literature to date, but could be useful for producing probabilistic forecasts.

Regardless of whether they are intended as synthetic generation models or forecasting models, solar irradiance models should account for the intended time scale of their application. Figure 2.3 shows the relationship between aspects of electric power systems operations and the spatiotemporal extent of meteorological

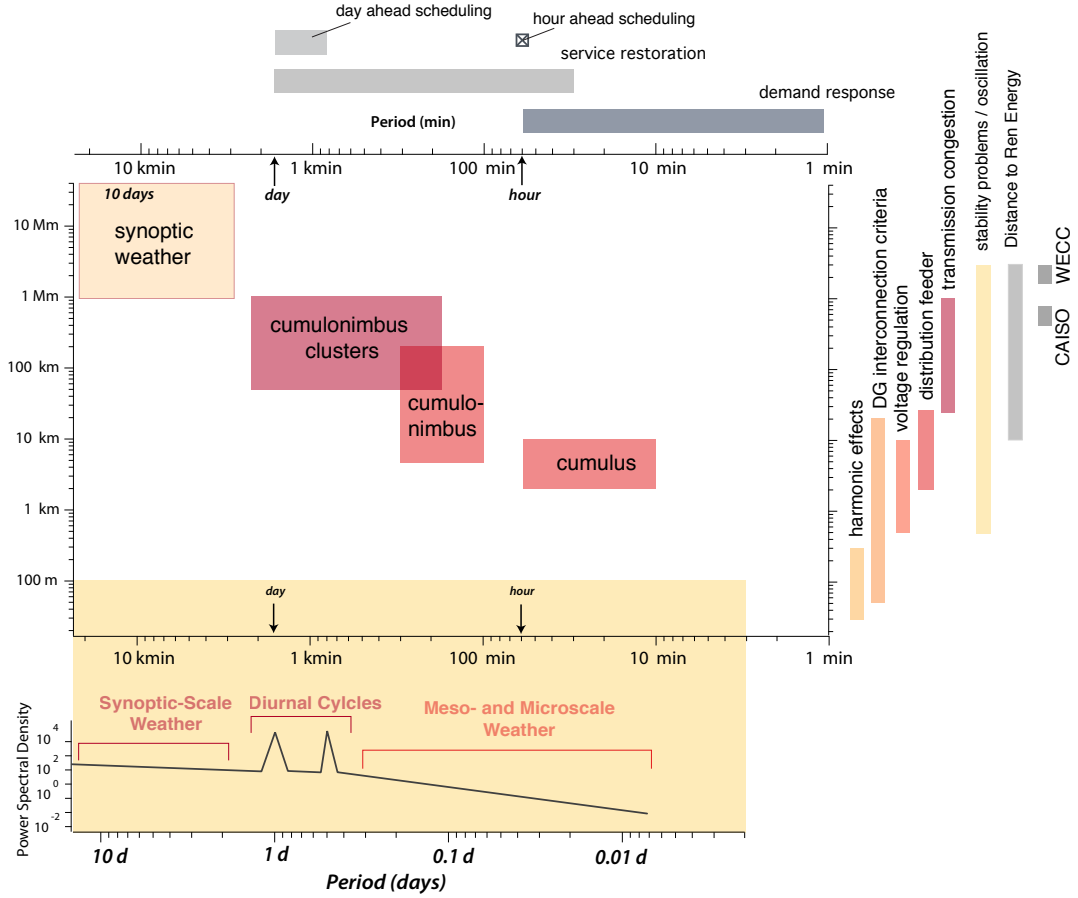


Figure 2.3: Temporal and spatial scales of weather phenomena along with those of electric power systems considerations and operations. Meteorological data based on data from Fujita (1981) and power systems information based on Von Meier (2011). Reproduced from Srikrishnan et al. (2017).

phenomena. Models which are intended for use on sub-hourly time scales must contend with the highly stochastic nature of cumulus cloud fields, while models for more aggregated data (such as daily averages) should reflect synoptic-scale variability and persistence. Aerosol optical depth also varies differently on different time scales, which will affect the variability of the associated time series (Gueymard, 2012b).

2.5.2 Autoregressive Time Series Models

The most common types of parametric time series models used in the solar modeling literature are autoregressive models (AR), autoregressive-moving average models (ARMA), and autoregressive-integrated-moving average models (ARIMA). These models explicitly take into account the autocorrelation of the observations. The AR model of order p (AR(p)) is

$$y_t = \mu + \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t,$$

where μ is a constant and $\varepsilon_t \sim N(0, \Sigma)$ is a white noise process. The stochastic disturbances introduced by ε_t are referred to either as errors or innovations. In addition to lagged observations, exogenous variables may be used as additional regressors (these models are sometimes denoted as ARX models, though we do not distinguish between AR and ARX models in this dissertation).

ARMA models of orders p, q (ARMA(p, q)) combine an AR(p) process with a moving average (MA) process of order q . MA models model the observation as the sum of lagged innovations, with the MA(q) model being

$$y_t = \mu + \varepsilon_t + \sum_{j=1}^q \psi_j \varepsilon_{t-j}.$$

MA models describe stationary time series, while series described by AR models need not be stationary. The ARMA(p, q) model is

$$y_t = \mu + \varepsilon_t + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \psi_j \varepsilon_{t-j}.$$

In many papers discussed below, the more general ARMA($p, 0$) notation is used to describe AR(p) models. ARIMA(p, r, q) models are more general still, modeling the data after r first-differences with an ARMA(p, q) model.

These autoregressive models assume independent and identically distributed (i.i.d.) Gaussian innovations. Correlated innovations may be modeled by modifying the model structure. For example, the autoregressive conditional heteroskedasticity

(ARCH) model (Engle, 1982)

$$y_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=0}^q \alpha_i y_{t-i}^2,$$

where $\varepsilon_t \sim N(0, 1)$, captures volatility clusters in innovation variance.

Other time series models capture autocorrelation in other ways. For example, hidden Markov models (HMM) model the data as being emitted from a mixture of processes, where the component indicator of the mixture is generated by an underlying Markov chain. A Markov chain is a set of states $S_i, i \in \mathcal{I}$, where $\mathcal{I}$ is some index set, equipped with a probabilistic structure governing transitions between these states as time progresses. The case where $\mathcal{I}$ is finite (and thus the states are discrete and finite) and where time is also treated as discrete is of most relevance to both this dissertation and the solar modeling literature at large. In this setting, the probabilistic structure on $\{S_i\}$ is specified by a matrix of transition probabilities $\eta(t)$, where $\eta_{i,j}(t)$ is the probability of transitioning from S_i to S_j at time t . More details on Markov chains are provided in Chapter 4. Autocorrelation can then be modeled through the persistence of the mixture component indicator or through temporal dependence of future values on previous observations.

2.5.3 Synthetic Generation of Solar Irradiance Series

Autoregressive models have been used to model solar irradiance since the 1970s (Goh and Tan, 1977; Brinkworth, 1977; Mustacchi et al., 1979; Biga and Rosa, 1981; Bartoli et al., 1981; Boileau, 1983; Graham et al., 1988; Graham and Hollands, 1990; Knight et al., 1991; Aguiar and Collares-Pereira, 1992b). Mustacchi et al. (1979) recognized the unsuitability of a direct application of autoregressive models of order p (AR(p)) for clearness indices due to the Gaussian likelihood for each innovation. While k_t may exceed a value of 1 (that is, $\text{GHI} > \text{G0}$ due to cloud lensing, this effect is less likely the longer the period over which the data is averaged, and so we can observe that the domain of k_t is some bounded set $[0, 1 + \tau]$. From a purely statistical perspective, assuming that the expected value of ε_t is zero, its distribution is dependent on the previous values of k_t (in particular, the skewness of the distribution must be related to the distance between the linear combination

of k_t values in the model equation and the boundary values). Mustacchi et al. acknowledged that transforming the observations to be Gaussian-distributed using the empirical CDF prior to using an AR(2) model produced satisfactory results. This approach was adopted by Graham et al. (1988) to generate synthetic daily k_t series with an AR(1) model.

Graham and Hollands (1990), in a seminal paper, decomposed hourly k_t series into a deterministic trend component k_{tm} and a stochastic component α ,

$$k_t = k_{tm} + \alpha.$$

k_{tm} was modeled as an exponential with respect to air mass m , $k_{tm} = a_0 + a_1 \exp(-km)$, based on a clear-sky analysis. To extend this analysis to all-sky conditions, the linear and exponential coefficients a_0, a_1 , and k were taken to be functions of the daily average k_t . The stochastic component $\alpha \sim \text{Beta}(a, b)$ was simulated using an AR(1) process after mapping the beta distribution to the standard Gaussian distribution via $\Phi^{-1} \circ I_x(a, b)$, where Φ is the CDF of the standard normal distribution and $I_x(a, b)$ is the CDF of $\text{Beta}(a, b)$.

An alternate approach to dealing with the non-normality of the marginal k_t distribution when using parametric autoregressive models is to model the z -scores of the series, which are distributed according to the standard Gaussian distribution. Aguiar and Collares-Pereira (1992b) used an AR(1) model on the z -scores of the k_t distribution for each hour, where the mean and standard deviation of each hourly k_t distribution are functions of daily average k_t and the hour. This allows the daily trend to be captured by the changing moments of the hourly distributions. One advantage of the z -score approach versus that of Graham and Hollands (1990) is that the former does not require an auxiliary air mass model to generate synthetic sequences. While it does require a process to generate sequences of daily k_t values, models for daily k_t were studied by Biga and Rosa (1981), Bartoli et al. (1981), and Aguiar et al. (1988), among others. Vindel and Polo (2014) found that Markov chains can be used to generate sequences of daily global and direct irradiation.

Hontoria et al. (2002) followed Graham and Hollands (1990) in modeling k_t series as a stationary stochastic series around a trend process, but used ANNs to generate both the trend and the innovations. The recurrent ANN for the trend acts as a hierarchical AR(3) model, using the previous three lagged k_t values as

inputs, along with the hour of the day, the daily average k_t , and other inputs intended to ensure a coherent monthly and daily trend. The stochastic component α is simulated by combining ideas from Graham and Hollands (1990) and Aguiar and Collares-Pereira (1992b). Rather than directly simulating α using the ANN, its z -scores are modeled using an AR(1) model, with the ANN generating the autoregressive coefficient and the variance of the innovations.

Glasbey (2001) modeled 30-second k_t values more directly using non-linear time series models, including Markov-switching state-space models (Kim, 1994; West and Harrison, 1997) and Markov-switching autoregressive models. The use of Markov-switching methods produces marginal distributions which are mixtures of more basic distributions (depending on how the model in each state is defined).

More recent synthetic generation models have gone beyond only simulating the clearness index to also generate beam and/or diffuse irradiance, both to enable computation of POA irradiance without reliance on empirical correlations and to facilitate planning for concentrating SECS. Polo et al. (2011) models 10-minute beam and global irradiance averages from hourly mean averages by assuming the 10-minute averages are perturbations of the hourly value. The variance of the perturbations are sampled from a beta distribution.

Other models intended to model multiple irradiance components more explicitly model the cloud cover fraction. Laslett et al. (2014) uses an AR(1) model for the average hourly cloudiness z -score, where the average cloudiness distribution for each hour is conditioned on the average daily cloudiness and the solar altitude angle.

Morf (1998) uses a two-state Markov chain to drive the model, with the states indicating whether or not the sun is behind a cloud. These sequences of cloudy and sunny periods assume no particular temporal duration for these events. Global irradiance is then divided into the sum of beam and diffuse irradiance. It is assumed that when the sun is obscured, there is no beam component. Morf (2013) expands on this model by stochastically generating the cloud cover fraction (using the model described in Morf (2011)) in addition to the sun-obscured indicator, allowing for smoother transitions between clear-sky and overcast conditions.

2.5.4 Forecasting of Solar Irradiance Series

A number of parametric time series methods have been used to forecast solar irradiance over multiple time scales. Initially, AR models were used across multiple time scales: Goh and Tan (1977) forecast hour-ahead GHI with an AR(1) model, Boileau (1983) forecast daily-averaged GHI with an AR(1) model, and Moreno-Muñoz et al. (2008) used an AR(4) model to forecast 5-minute average GHI). ARIMA models have also been used by Yang et al. (2012). Sun et al. (2015) found that an ARMA model equipped with a GARCH (generalized ARCH) residual structure outperformed ARMA models with identically and independently distributed (i.i.d.) residuals and ANN models for forecasting monthly mean GHI. Other parametric statistical forecasting methods have included linear regression on exogenous variables derived from satellite images (Fu and Cheng, 2013), exponentially smoothed time series models (Dong et al., 2013; Yang et al., 2015; Prema and Rao, 2015), Kalman filtering (Soubdhan et al., 2014), and Markov-switching regression (Shakya et al., 2017).

Simultaneous forecasts of multiple irradiance components that do not require the use of empirical correlations such as those described in Section 2.4 are rare in the literature despite the joint uncertainties illustrated in Figure 1.2. Marquez and Coimbra (2011) trained separate ANNs to forecast same-day and day-ahead hourly GHI and DNI from numerical weather prediction forecasts, using different meteorological covariates for each model. Yang et al. (2012) found slightly improved performance when forecasting DNI and DHI separately than directly forecasting GHI from a forecasted cloud cover index (the separate DNI and DHI components were combined into a GHI forecast). Alonso-Montesinos et al. (2015) forecasted all three irradiance components (GHI, DNI, and DHI) for forecast horizons up to 180 minutes-ahead using data from a total sky imager.

The number of methods for the probabilistic forecasting of irradiance components have increased in recent years (several papers have directly forecasted PV power production in a probabilistic manner). Marquez and Coimbra (2011) used ANNs to forecast the standard deviation of GHI and DNI, which were then used to produce prediction intervals under the assumption of normally-distributed residuals. Iverson et al. (2014) used stochastic differential equations to produce a probabilistic model for GHI. Silva et al. (2015) issued probabilistic forecasts for next-day hourly GHI by

clustering daily profiles using k -means clustering, then computing the probability that the next day would have a particular profile based on the sequence of cluster memberships for the previous days. This model performed more poorly than a naive Bayes classifier (using the naive Bayes assumption that the previous values were conditionally independent) when data was relatively limited (2 years), but performed equally well with 10 years of data. Torregrossa et al. (2016) computed GHI prediction intervals for sub-second forecast horizons using GARCH time-series innovations. McCandless et al. (2015) used a model tree approach to forecast GHI spatial and temporal variability. The Markov-switching method used by Shakya et al. (2017) is naturally probabilistic due to the Markov chain transition probabilities. Chu and Coimbra (2017) used k -nearest neighbor clustering to generate an ensemble GHI forecast.

2.5.5 Contribution of This Dissertation

The purpose of the work in this dissertation is to begin to fill in some gaps in the existing literature. Almost all of the existing models for synthetic generation and forecasting use frequentist methods such as maximum-likelihood estimation and expectation-maximization for calibration. It is not standard practice in the solar literature to provide confidence intervals for parameter values, making quantification of parametric uncertainty difficult. Even when provided, these intervals are generally produced under assumptions of asymptotic normality, which may not be warranted, as it is difficult to know how much data is required for the asymptotic approximation to hold. Indeed, the use of several decades of ground-measured or satellite-derived data may be required for appropriate levels of confidence, the former of which may not be available (Gueymard and Wilcox, 2011; Vignola et al., 2012). The methods that are not frequentist in nature often rely on clustering and explicitly searching the available data for similarities, which may not perform well when data is relatively scarce (for example, the clustering method of Silva et al. (2015) performs relatively poorly when only 2 years of data are available, and performs as well as a naive Bayes classifier when 10 years of data is available).

Further, joint forecasting and variability/uncertainty quantification of multiple irradiance components, which is important for multiple applications, is atypical, with Marquez and Coimbra (2011) standing out as the major attempt to simultaneously

forecast GHI and DNI from a common set of covariates while also forecasting prediction intervals. The use of these methods for POA power forecasting, to name one application, then requires the use of empirical correlations, which may not fully capture joint uncertainty.

We address both of these concerns using Bayesian methods to model GHI and DNI simultaneously, through the use of Markov-switching vector autoregressions (see Chapter 4 for more details). Bayesian methods allow for the representation of parametric uncertainty through the joint posterior distribution of the model parameters. The specification of prior distributions allow for parametric regularization, the introduction of external information about parameter values, or both. In the context of solar radiation, prior information about local radiative transfer patterns may be useful in complementing a somewhat scarce data set. Moreover, in addition to producing probabilistic models, the examination of posterior predictive distributions may provide further insights into how to improve the model if the data does not appear to be well-represented by the posterior predictive distribution.

Chapter 3 |

A Markov-Switching Vector Autoregressive Model for Solar Irradiance

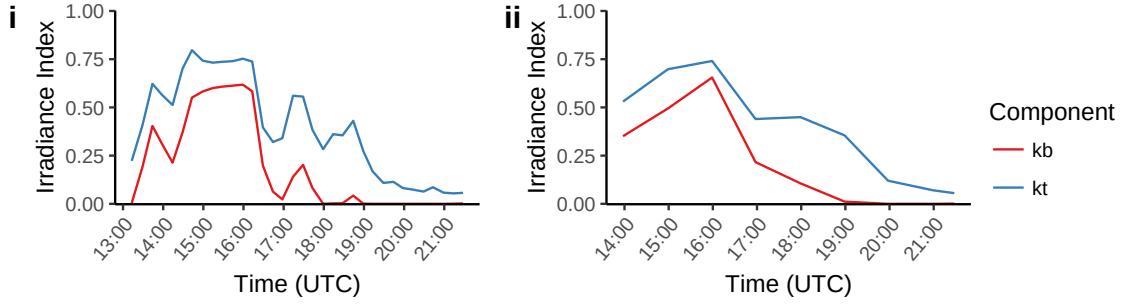
3.1 Modeled Variables

Rather than work directly with the raw irradiance components GHI and DNI, we will instead model the normalized (by TOA irradiance G_0) component indices k_t and k_b . These two variables are sufficient for a complete dimensionless characterization of the relevant properties of the atmospheric state for surface-level SECS, as described in Section 2.2.

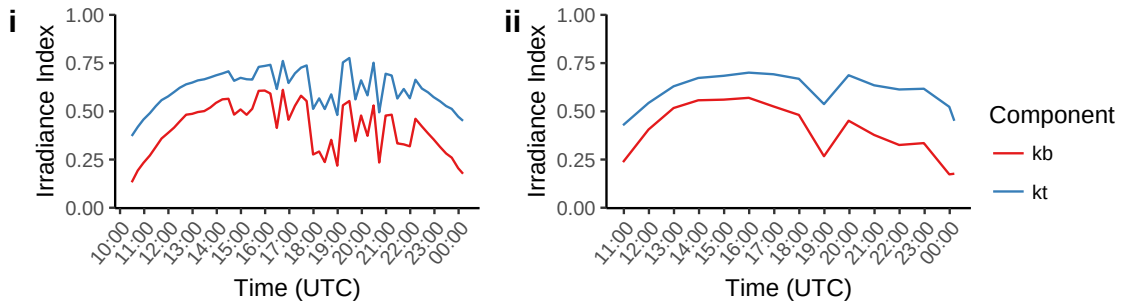
This normalization mostly removes the diurnal trend (an air mass effect, related to the path length of incoming radiation, still remains). We can then directly model the attenuation of incoming radiation through the atmosphere, rather than including the deterministic and exogenous diurnal trend stemming from first-order sun-earth geometry.

3.2 Regime-Switching in Solar Irradiance Series

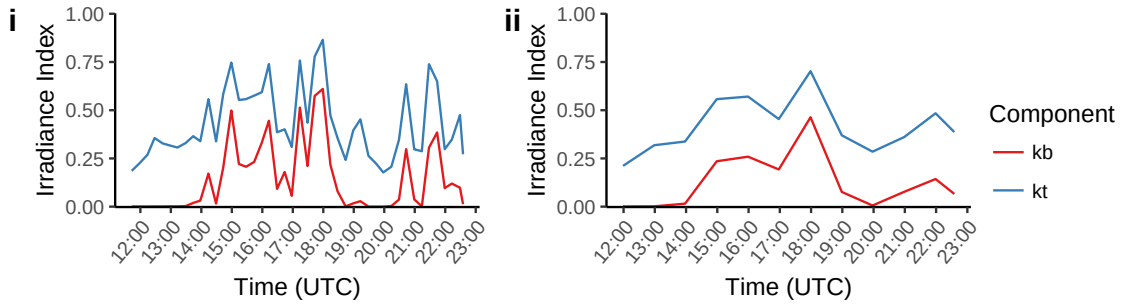
Different cloud fields have different levels of stochasticity (and therefore different levels of intermittency) and transmit and scatter downwelling solar radiation differently. Regime-switching is therefore a natural property of solar irradiance time series. Examples of the resulting non-stationarity across different time scales



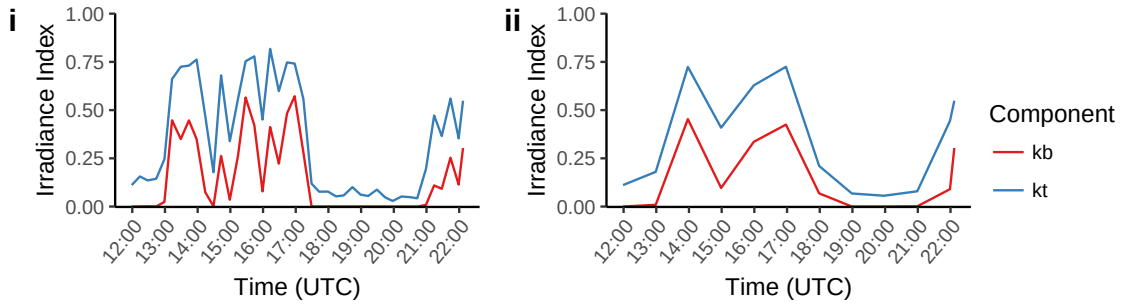
(a) January 11, 2012.



(b) July 01, 2012.



(c) September 23, 2012.



(d) October 10, 2012.

Figure 3.1: Selected daily time series for k_t and k_b demonstrating regime switching. For each date, plot i) is a series of 15-minute averaged data, while plot ii) is a series of hourly-averaged data.

can be seen in Figure 3.1. Dealing with this regime-switching behavior is the core difficulty of solar irradiance forecast and resource modeling, particularly for statistical methods; ideally, forecast models would be able to anticipate changes of regime, while resource assessment models capture the typical duration of each regime and accurately capture the transition probabilities between regimes.

The statistically relevant regimes for the purpose of this modeling work feature varying levels of volatility, covariance, and mean level for both series. For example, while a two-state model might feature states which broadly correspond to “clear” and “cloudy” regimes, the states in a three-state model might be interpreted as “clear,” “cloudy,” and “partially cloudy.” These states may or may not correspond precisely to intuitive meteorological categories, however.

3.3 Markov Chains

Regime-switching behavior in time series is typically modeled through the use of a latent Markov chain structure (Frühwirth-Schnatter, 2006). A Markov chain is a sequence of random variables $Y_1, \dots, Y_n, \dots$ for which the probability of future values is conditionally dependent only on the previous value, so that

$$p(Y_n = y_n | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}) = p(Y_n = y_n | Y_{n-1} = y_{n-1}). \quad (3.1)$$

The conditional dependence only on the previous value (or *state*) defined by Equation (3.1) is known as the first-order Markovian property (generalizations exist to higher-order dependence). The transition probabilities $P(Y_n = y_n | Y_{n-1} = y_{n-1})$ are defined by a sequence of probability structures (ξ_n). While Markov chains can be defined on continuous or discrete time index sets, the latter is of primary relevance to this work. We will write $\xi_n = \xi_t$ to reflect that the transitions between states occur between discrete time intervals. We further restrict our attention to Markov chains with a finite state space $S_t \in \{1, \dots, K\}$, though Markov chains can be defined on countable state index sets. Then the transition probability structures (ξ_t) can be represented as $K \times K$ -dimensional matrices, where $\xi_{jk,t}$ is the probability of transition from state j to state k at time t ; therefore each row of ξ satisfies $\sum_{k=1}^K \xi_{jk,t} = 1$ and all $\xi_{jk,t} \geq 0$.

Markov chains are *time-homogeneous* if the transition probabilities are inde-

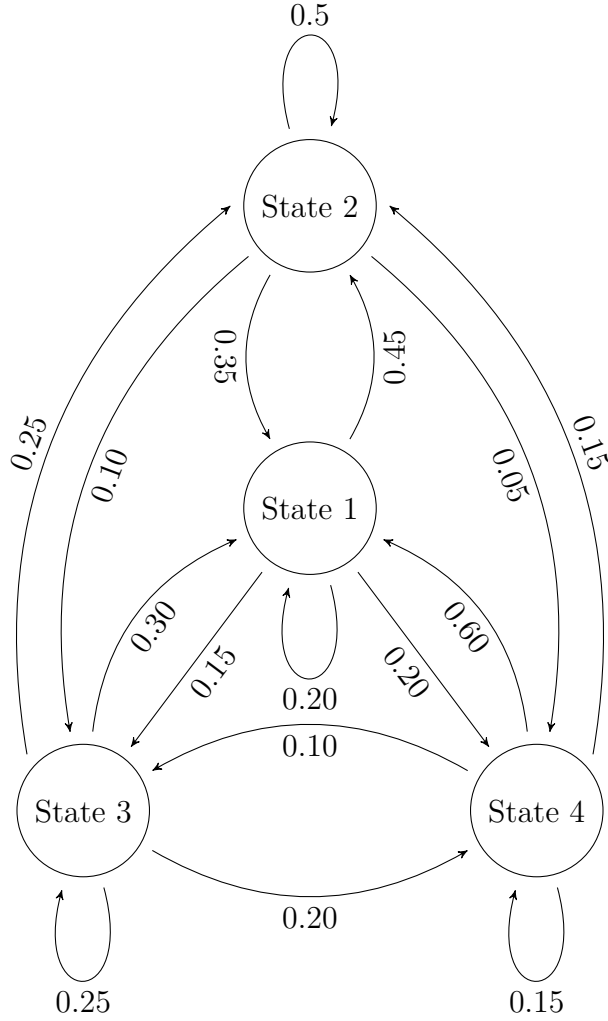


Figure 3.2: Graphical representation of a time-homogeneous Markov chain. The numbers are the probabilities of the associated state transitions.

pendent of time, *i.e.* $\xi_t = \xi$ for all t . Figure 3.2 is a graphical representation of a time-homogeneous Markov chain. In this case, if the Markov chain is *ergodic* (every state has a positive expected return time and is aperiodic), there exists a probability distribution π over the states that is invariant under the action of the transition matrix, so that

$$\xi\pi = \pi. \quad (3.2)$$

Any such distribution π is known as the *stationary*, or *invariant*, distribution of the Markov chain. Stationary distributions are of interest because if the state S_{t-1} is drawn according to an invariant distribution π , from Equation (3.2) it follows

that the state S_t is also drawn according to $\boldsymbol{\pi}$:

$$p(S_t = k) = \sum_{j=1}^K p(S_t = k | S_{t-1} = j, \boldsymbol{\xi}) p(S_{t-1} = j | \boldsymbol{\xi}) = \sum_{j=1}^K \xi_{jk} \pi_j = \pi_k.$$

Equation (3.2) also shows that $\boldsymbol{\pi}$ is the eigenvector of $\boldsymbol{\xi}$ associated with eigenvalue 1, normalized so that $\|\boldsymbol{\pi}\|_1 = 1$.

3.4 Markov-Switching Models

Due to their ability to probabilistically capture structural changes in system dynamics, Markov-switching models (sometimes referred to as hidden Markov models), have found wide application in areas such as speech recognition (Juang and Rabiner, 1991), evolutionary biology (Mitchison and Durbin, 1995), weather modeling (Dorrestijn et al., 2013; Yang et al., 2014; Salvucci et al., 2002), sensor networks (Liu et al., 2009), electricity price forecasting (Janczura and Weron, 2012), and pattern recognition (Yang et al., 1995). Models with Markov-switching dynamics have been used in solar irradiance modeling (Mustacchi et al., 1979; Aguiar et al., 1988; Morf, 1998; Muselli et al., 2001; Hocaoglu, 2011; Bhardwaj et al., 2013; Morf, 2014; Ngoko et al., 2014; Vindel and Polo, 2014; Shakya et al., 2017).

3.4.1 MSAR Models

Markov-switching autoregressive models are mixtures of autoregressive models, where the latent mixture indicator is driven by a Markov chain, rather than being sampled i.i.d. from the distribution of mixture components. Consider the following Markov-switching autoregressive structure of order p (MSAR(p)):

$$y_t = \boldsymbol{\beta}_{S_t} \mathbf{x}_t' + \varepsilon_t, \quad (3.3)$$

where $\mathbf{x}$ is a matrix of covariates, including p autoregressive terms and possibly including an intercept term, $\boldsymbol{\beta}_{S_t}$ are the regression coefficients associated with the Markov state S_t , and $\varepsilon_t \sim N(0, \sigma_{S_t}^2)$ is the white noise process for the innovations. Equation (3.3) is the most general form for an MSAR(k) model, as all of the

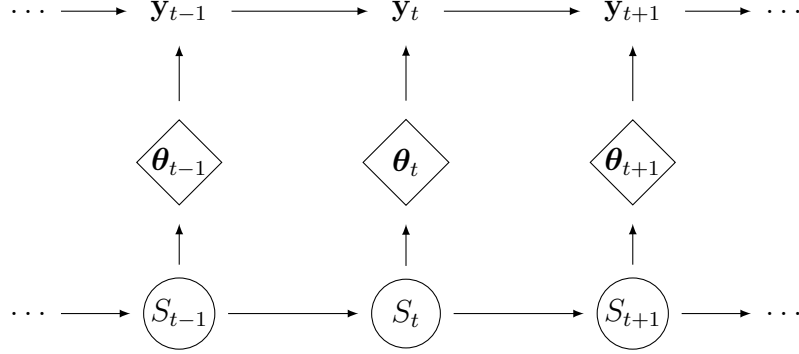


Figure 3.3: Digraph representation of a realization of a Markov-switching autoregressive model with $p = 1$. The states S_t follow a first-order Markov process. Model parameters $\theta_t = \theta_{S_t}$ are associated with each state.

model parameters are allowed to switch, and more constrained models, in which only certain parameters (the innovation variance σ^2) switch, also fit into the same framework. Figure 3.3 is a graphical representation of an MSAR(1).

For the remainder of this dissertation, we adopt the following notational convention: subscript time indices, *e.g.* y_t , refer to parameter values at time t , while superscript indices, *e.g.* y^t , refer to the history of the parameter through time t , so that $y^t = (y_1, \dots, y_{t-1})$.

If the Markov chain $\{S_t\}$ satisfies the first-order Markovian property, the MSAR model in Equation (3.3) satisfies the following conditional independence property: the conditional density of y_t given the history y^t and the model parameters θ only depends on the current state S_t and the observation history y^{t-1} :

$$p(y_t | y^{t-1}, \mathbf{S}^t, \theta) = p(y_t | y^{t-1}, S_t, \theta). \quad (3.4)$$

One consequence of the conditional independence on the state history S^{t-1} in Equation (3.4) is simplified statistical inference, as compared to models with longer state memories. (Frühwirth-Schnatter, 2006).

Markov-switching autoregressive models were first introduced in econometrics by Hamilton (1989) and have also been used in wind forecasting (Pinson and Madsen, 2012; Trombe et al., 2012; Ailliot and Monbet, 2012) and ecology (Pinto and Spezia, 2016), among other fields.

3.4.2 MSVAR Models

The multivariate extension of the MSAR model (3.3) is the Markov-Switching Vector Autoregressive Model (MSVAR):

$$\begin{aligned}\mathbf{y}_t &= X_t \boldsymbol{\beta}_{S_t} + \boldsymbol{\varepsilon}_t \\ \boldsymbol{\varepsilon}_t &\sim \mathcal{N}(\mathbf{0}, \Sigma_{S_t}),\end{aligned}\tag{3.5}$$

where $\mathcal{N}(\cdot, \cdot)$ is the multivariate normal distribution. This totally-switching specification allows a common set of covariates X (including the intercept terms) to influence multiple observations $\mathbf{y}$, with the possibility of a non-zero covariance between the terms of the multivariate white-noise process $\boldsymbol{\varepsilon}$. As with the MSAR model, the MSVAR model can be interpreted as a mixture of multivariate normal autoregressions, with the mixture indicator determined by a Markov chain instead of an i.i.d process.

3.5 Model Definition

3.5.1 Variable Transformation

To facilitate re-dimensionalization, rather than computing the average G_0 and θ_z for each average block, we used the G_0 and θ_z corresponding to the beginning of the averaging period when producing k_t and k_b . This can lead to marked visual differences in the 15-minute and hourly series during periods of rapid increase or decrease in GHI or DNI. We also limit our series to exclude observations (prior to averaging) corresponding to low-sun conditions, so that $\theta_z \leq 85^\circ$.

One complication is that while the resulting k_t and k_b values are bounded on an interval $[0, 1 + \tau]$, the MSVAR specification (3.5) has a multivariate normal likelihood,

$$\mathbf{y}_t \sim \mathcal{N}(X_t \boldsymbol{\beta}_{S_t}, \Sigma_{S_t}).$$

While beta regression (Ferrari and Cribari-Neto, 2004; Rocha and Cribari-Neto, 2009; Billio and Casarin, 2011; Figueroa-Zúñiga et al., 2013; Souza and Moura, 2016) could be used to directly model k_t and k_b , we will instead map these variables onto the real line and use a generalized linear model. This can be done using a

number of transformations, including the logit map

$$x \mapsto \log \left(\frac{x}{1-x} \right).$$

One drawback of using the logit as a link function is its highly nonlinear behavior near the boundaries $x = 0$ and $x = 1$. While the observed values of k_t and k_b rarely get close to the maximum value $x = 1$ (after normalization to the unit interval), k_b in particular has a large density of points at or near 0 (Figures 2.2h and 2.1h). This nonlinearity may cause problems when the states are not *a priori* well-constrained. Instead, we first transform each variable x_j using the probability integral transform

$$x_j \mapsto F_j(x_j),$$

where F_j is the empirical CDF of variable x_j . This maps the empirical domain of each irradiance component onto the unit interval, with each value x mapped to its quantile. We then obtain the transformed variables

$$\mathbf{y} = \begin{pmatrix} \hat{k}_t \\ \hat{k}_b \end{pmatrix} = \Phi^{-1} \begin{pmatrix} F_{k_t}(k_t) \\ F_{k_b}(k_b) \end{pmatrix},$$

where Φ is the standard multivariate normal CDF.

3.5.2 Model Specification

A key modeling choice involves the form of the autoregression terms. Consider the two $\text{AR}(\rho)$ models

$$y_t = \beta_0 + \sum_{i=1}^{\rho} \beta_i y_{t-i} + \sigma_t \quad (3.6)$$

and

$$y_t - \mu = \sum_{i=1}^{\rho} \beta_i (y_{t-i} - \mu) + \sigma_t, \quad (3.7)$$

and the associated $\text{MSAR}(\rho)$ forms

$$y_{S_t} = \beta_{S_t,0} + \sum_{i=1}^{\rho} \beta_{S_t,i} y_{t-i} + \sigma_t \quad (3.8)$$

and

$$y_t - \mu_{S_t} = \sum_{i=1}^{\rho} \beta_{S_t,i} (y_{t-i} - \mu_{S_t}) + \sigma_t. \quad (3.9)$$

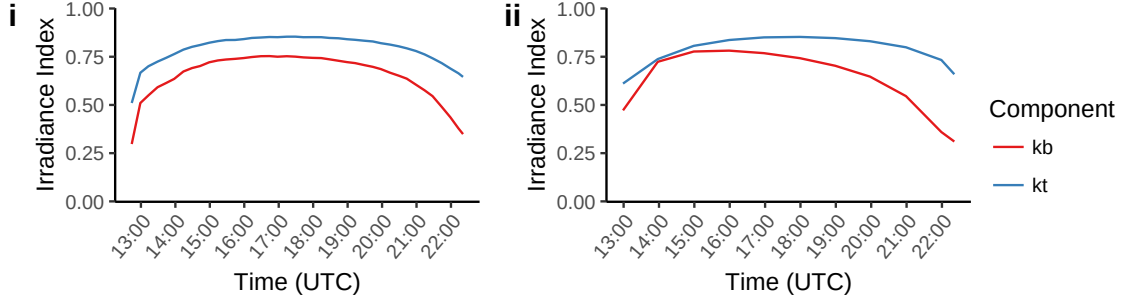
For non-switching AR(p) models, parameterizations (3.6) and (3.7) are equivalent, as $\beta_0 = \mu(1 - \sum_{i=1}^{\rho} \beta_i)$. However, the Markov-switching parameterizations (3.8) and (3.9) are not equivalent. When a state transition from S_{t-1} to S_t occurs, model (3.9) immediately shifts mean level from $\mu_{S_{t-1}}$ to μ_{S_t} , while model (3.8) more smoothly changes mean level over several time periods. While model (3.9) more accurately reflects the sudden regime changes observed in Figure 3.1 (particularly over shorter time scales, such as 15 minutes), inference (see Chapter 4) is more complicated due to the dependence on ρ previous latent states $S_{t-1}, \dots, S_{t-\rho}$ (which makes inference easier when recasting the model in switching state-space form (Kim, 1994), while inference for (3.8) depends only on the current state S_t (as in (3.4)). In the spirit of developing a benchmark probabilistic model, we opt for the intercept form (3.9) as the basic model structure. Additionally, this form is consistent with the work of Knight et al. (1991) and Aguiar and Collares-Pereira (1992b).

Given our choice of working with k_t and k_b rather than their clear-sky equivalents, we have to account for the path-length effect observed in the clear-sky series shown in Figure 3.4. With respect to the transmissivity indices k_t and k_b , this effect is additive. Thus, we include $\cos(\theta_z)$ as a covariate, which also normalizes θ_z . We also want to let all of the model parameters switch, as different cloud regimes have different mean levels of transmissivity and scattering as well as different levels of volatility (ranging from the relatively constant totally-clear and totally-overcast cloud fields to highly stochastic scattered cumulus fields). This yields the following multivariate model specification:

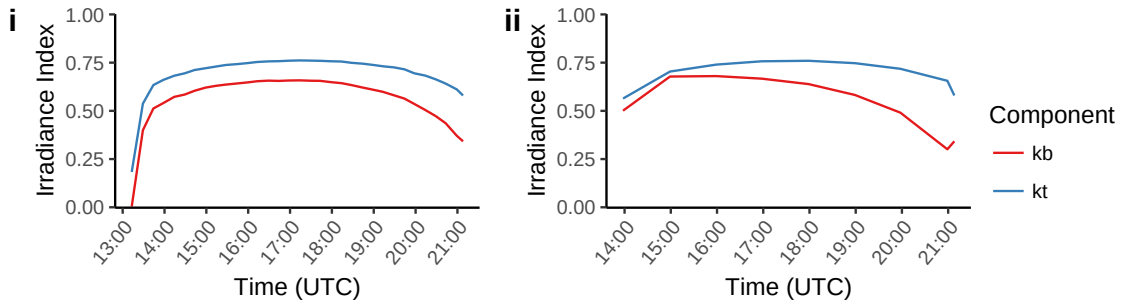
$$\mathbf{y}_t = \beta_{0,S_t} + \beta_{1,S_t} \cos(\theta_z) \cdot \mathbf{1}_2 + \sum_{i=1}^{\rho} \beta_{i+1,S_t} \mathbf{y}_{t-i} + \boldsymbol{\varepsilon}_t, \quad (3.10)$$

where $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \Sigma_{S_t})$, $\mathbf{1}_2 = (1, 1)^T$, and $\mathbf{y} = (\hat{k}_t, \hat{k}_b)^T$.

Another assumption that we make is that each daily series is an independent replicate of the stochastic process (3.10). This somewhat contradicts the solar resource assessment literature on the correlation between consecutive mean daily values (Biga and Rosa, 1981; Bartoli et al., 1981; Aguiar et al., 1988), but removes a hierarchical level from the model. For the purpose of validation, initial values



(a) February 20, 2012.



(b) December 13, 2012.

Figure 3.4: Selected daily time series for k_t and k_b demonstrating the trend under clear sky conditions for an entire day. Plot i) is a 15-minute averaged series, while plot ii) is an hourly-averaged series

$y_1, \dots, y_p$ will be drawn from the validation data set, but an easy extension of the model would involve fitting a mixture model to these values from which samples can be generated.

Finally, we will be flexible with regard to the homogeneity of the latent Markov chain (S_t). In the spirit of simplicity, for the benchmark resource assessment model in Chapter 5, we assume homogeneity, though this neglects diurnal meteorological trends. For the forecast model in Chapter 6, the Markov process will be inhomogeneous, so that the transition probabilities depend on some external covariates, with the aim of using these variables to improve the implicit cloud regime forecast.

Chapter 4 |

Bayesian Inference for MSVAR Models

4.1 Bayesian Inference

In this dissertation, we adopt a Bayesian (rather than a frequentist) approach to statistical inference and model calibration. While frequentist inference attempts to identify the most probable value(s) of the data-generating model parameters (by maximizing the likelihood function, or data distribution, $p(\mathbf{y}|\boldsymbol{\theta})$), Bayesian inference produces a probability distribution (the *posterior* distribution, $p(\boldsymbol{\theta}|\mathbf{y})$) over the model parameter space by using the likelihood to update a *prior* distribution, $p(\boldsymbol{\theta})$. This updating (also known as Bayesian inversion) is done using Bayes' rule (Bayes, 1763):

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{y}) &= \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})} \\ &\propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}), \end{aligned} \tag{4.1}$$

where $p(\mathbf{y}) = \int p(\mathbf{y}|\boldsymbol{\theta})d\boldsymbol{\theta}$ is a normalizing constant. Construction of a Bayesian model therefore requires specifying both the prior distribution for the model parameters and the sampling distribution for the data-generating process conditioned on the parameter values. Specific choices of prior distributions can be made for several reasons, including ease of computation, parameter regularization, and prior knowledge about reasonable or likely parameter values (Gelman et al., 2013).

As Bayesian inference is based on the posterior distribution, Bayesian modeling is inherently probabilistic, though point estimates can be obtained by using particular

parameter values, such as the posterior mean or MAP (maximum *a posteriori*) estimate. Bayesian predictive distributions (the *posterior predictive distribution*) for unknown data $\tilde{y}$ are generated by integrating over the posterior distribution:

$$p(\tilde{y}|\mathbf{y}) = \int p(\tilde{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}.$$

As a result, Bayesian methods lend themselves naturally to uncertainty quantification (viewing probability as the measure of uncertainty), including probabilistic forecasting.

Complex statistical models are likely to have dependencies between the parameters. Hierarchical models can describe certain types of dependencies, where the data is modeled conditionally on some parameters, which in turn have a probabilistic specification in terms of other parameters. For example, Figure 3.3 can be interpreted as a hierarchical model, where the likelihood defined by parameters $\boldsymbol{\theta}_t$ depends on the state indicator S_t . Another common reason to use hierarchical models is the probabilistic specification of hyperparameters.

4.2 Sampling from the Posterior Distribution

In general, the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$ will not be available in closed form, and therefore cannot be sampled from directly. One class of computational methods are Markov chain simulation methods, commonly referred to as Markov Chain Monte Carlo (MCMC) methods. These iterative methods rely on the property that ergodic Markov chains converge to their stationary distributions (see Robert and Casella (2004, Chapter 6) for a detailed mathematical treatment). Upon convergence, a sequence from this chain is then distributed according to the stationary distribution.

In this case, the goal is to construct a Markov chain with a suitable transition kernel such that the stationary distribution is equivalent to the posterior distribution. A sampled sequence is then representative of the posterior distribution (requiring that the chain is run sufficiently long after convergence to fully explore the stationary distribution). The so-called “burn-in” portion of the chain, corresponding to the pre-convergence portion, can be discarded if desired (or the chain can be run for a longer period of time). The length of the burn-in portion will depend on the choice of initial value $\boldsymbol{\theta}^0$. The chain can also be thinned by keeping every k th element,

which reduces the dependency between the retained samples resulting from the Markov chain transition kernel at the expense of reducing computational efficiency.

The main MCMC algorithm is the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970; Robert, 2014). A special case of the Metropolis-Hastings algorithm (albeit with a different historical motivation) is the Gibbs sampler (Geman and Geman, 1984; Gelfand and Smith, 1990). Divide the parameter set $\boldsymbol{\theta}$ into subsets, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_l)$. A Gibbs sampling step (at iteration t) for $\boldsymbol{\theta}_i$ involves sampling directly from the conditional posterior distribution, $\boldsymbol{\theta}_i^t \sim p(\boldsymbol{\theta}_i | \hat{\boldsymbol{\theta}}_{-i}, \mathbf{y})$, where $\hat{\boldsymbol{\theta}}_{-i}$ is the currently sampled vector of the components other than $\boldsymbol{\theta}_i$,

$$\hat{\boldsymbol{\theta}}_{-i} = (\boldsymbol{\theta}_1^t, \dots, \boldsymbol{\theta}_{i-1}^t, \boldsymbol{\theta}_{i+1}^{t-1}, \dots, \boldsymbol{\theta}_l^{t-1}).$$

Gibbs sampling therefore requires that the prior distribution $p(\boldsymbol{\theta}_i)$ is selected so that the conditional posterior

$$p(\boldsymbol{\theta}_i | \hat{\boldsymbol{\theta}}_{-i}, \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\theta}_i, \hat{\boldsymbol{\theta}}_{-i}) p(\boldsymbol{\theta}_i)$$

is the closed-form density of a “standard” probability distribution. Such pairings of likelihood densities and prior densities are called conjugate distributions.

4.3 Identifiability

A family of parametric distributions (indexed by $\boldsymbol{\theta} \in \Theta$) is *identifiable* if two densities $p(\mathbf{y} | \boldsymbol{\theta}_1)$ and $p(\mathbf{y} | \boldsymbol{\theta}_2)$ are almost everywhere identical only if the parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are identical (Rothenberg, 1971). Frühwirth-Schnatter (2006) distinguishes two main types of non-identifiability for mixture densities (including those resulting from MSAR models).

1. *Component Relabeling Invariance*: Permutations of component labels produce the same mixture density, though the resulting parameter vectors are permuted, rather than being identical. This type of non-identifiability complicates parameter estimation due to label-switching, and will bias function estimates.
2. *Overfitting*: Non-identifiability results from the potential embedding of the space of mixtures with K components into the space of mixtures with $K + 1$

components. This non-identifiability class results from the potential addition of empty components or the splitting of one component into two identical components.

Both of these classes of non-identifiability can be resolved by imposing positive constraints on the mixture weights (or a positive transition probability into each state for a Markov-switching model) and inequality constraints on the parameter space, so that a unique labeling is enforced and components cannot be identical. These constraints should reflect the structure of the unidentified posterior to ensure that label-switching does not occur, as this results in biasing state-dependent posterior estimates towards the global estimate (Frühwirth-Schnatter, 2001).

4.4 Likelihood and Prior Specification

The parameters in the regression model (3.5) are collected as $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\xi})$, where we abuse notation to allow ambiguity in the definition of $\boldsymbol{\theta}$ depending on whether the latent Markov chain is homogeneous (in which case $\boldsymbol{\xi}$ are the homogeneous transition probability vectors) or inhomogeneous (we restrict our attention to cases in which the transition probabilities are independent of prior values of the data and depend only on exogenous variables, in which case $\boldsymbol{\xi}$ denotes the regression parameters $\boldsymbol{\kappa}$). The extended parameter vector $\boldsymbol{\psi} = (\boldsymbol{\theta}, S^T)$ includes the model parameters and the latent state vector.

The complete-data likelihood function for a Markov-switching model is

$$p(\mathbf{y}, \mathbf{S} | \boldsymbol{\theta}) = p(\mathbf{y} | \mathbf{S}, \boldsymbol{\theta}) p(\mathbf{S} | \boldsymbol{\theta}). \quad (4.2)$$

Markov-switching models can be viewed as having a hierarchical structure, as seen in Figure 3.3. The conditional mixture likelihood, which we use for parameter estimation, is specified in a hierarchical fashion:

1. The conditional (on the realization S^T of the state process and the model parameters $\boldsymbol{\theta}$) likelihood is:

$$f(\mathbf{y}^T | S^T, X^T \boldsymbol{\theta}) = \prod_{t=1}^T f(\mathbf{y}_t | S_t, X_t, \boldsymbol{\theta}), \quad (4.3)$$

with

$$f(\mathbf{y}_t|X_t, S_t, \phi) \sim \mathcal{N}(X_t' \beta_{S_t}, \Sigma_{S_t}).$$

2. Conditional on the Markov transition probabilities ξ , the prior density for the state vector factorizes (via a repeated application of Bayes' rule (4.1)) as

$$p(S^T|\xi) \propto \prod_{t=1}^T p(S_t|S_{t-1}, \xi) p(S_0),$$

where the prior distribution of S_0 can be treated as an exogenously known distribution or can be inferred from the data.

The structure of the state-transition likelihood $p(S_t|S_{t-1}, \eta)$ (and the prior selection) depends on whether the Markov chain is time-homogeneous or -inhomogeneous (see Section 4.6).

3. An independence prior is used for the model parameters θ :

$$p(\theta) = p(\beta)p(\Sigma)p(\xi).$$

Conditioned on the state S_t , (3.5) is a multiple linear regression model, and so conjugate priors for the state-dependent regression parameters β_k can be used (assuming independence between states): $\beta_{i,k} \sim \mathcal{N}(\mathbf{b}_i, B_i)$ (for $i = 1, 2$ corresponding to the observation components) and $\Sigma_k \sim \mathcal{W}^{-1}(\mathbf{I}_2, 3)$. While uniform priors could be placed on the β (corresponding to a lack of prior information about the regression coefficients), these would result in an improper posterior (Gelman et al., 2013).

4.5 MCMC Sampler for Markov-Switching Models

Bayesian inference for Markov-switching models involves data augmentation to sample the extended parameter vector $\psi = (\mathbf{S}, \theta)$ from the posterior distribution:

$$p(\mathbf{S}, \theta|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{S}, \theta) p(\mathbf{S}|\theta) p(\theta).$$

Sampling from the posterior involves three steps (this procedure is not specific to MSAR models).

Algorithm 4.5.1 (Unconstrained Data-Augmentation MCMC Sampler for Markov-Switching Models (Frühwirth-Schnatter, 2006)). *Start from some initial state process realization $\mathbf{S}^{(0)}$ and iterate over the following steps for $m = 1, \dots, M$:*

- a) Sample the transition probabilities $\boldsymbol{\xi}^{(m)}$ from the conditional posterior $p(\boldsymbol{\xi}|\mathbf{S}^{(m-1)})$.*
- b) Sample the model parameters (possibly blocked as $\boldsymbol{\theta}_1^{(m)}, \dots, \boldsymbol{\theta}_L^{(m)}$ from the conditional posterior $p(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_L|\mathbf{y}, \mathbf{S}^{(m-1)})$.*
- c) Sample the latent state indicator chain $\mathbf{S}^{(m)}$ from the conditional posterior $p(\mathbf{S}|\mathbf{y}, \boldsymbol{\psi}^{(m)})$.*
- d) Permute all state-dependent quantities, elements of the transition matrix $\boldsymbol{\xi}^{(m)}$, and the state sequence $\mathbf{S}^{(m)}$ according to some permutation of the state labels $\{1, \dots, K\}$.*

Discard the first M_0 iterations corresponding to the burn-in period.

The procedure for sampling the transition probabilities $\boldsymbol{\xi}$ in step a) varies significantly depending on if the underlying Markov chain is time-homogeneous or -inhomogeneous; our approach to both of these cases is discussed further in Section 4.6. Step b), sampling the model parameters, is discussed further for the MSVAR models used in this dissertation in Section 4.7. While we use Gibbs sampling, Metropolis-Hastings methods could also be used for other choices of priors.

Sampling the state indicators $\mathbf{S}$ in step c) is discussed in Section 4.8 for models such as (3.3) which only depend on the current value of S ; this is more complex for models with longer dependencies, for which inference is easier when they are expressed as switching state space models (Frühwirth-Schnatter, 2006).

The permutation step d) increases mixing of the MCMC chain by jumping between labeling subspaces (Frühwirth-Schnatter, 2001). This allows for an appropriate identifiability constraint to be found which respects the geometry of the posterior distribution. While *a priori* constraints can be used, these may result in label switching or for the MCMC chain to be trapped in a particular posterior mode. Once an appropriate set of constraints is identified, the MCMC output can be postprocessed so that the constraints are satisfied at each iteration. It is also important to note that if any priors are state dependent, they must also be permuted at this step.

4.6 Sampling State Transition Probabilities

Step a) in Algorithm 4.5.1 involves sampling the transition probabilities ξ . This procedure is different depending on if the Markov chain is time-homogeneous or -inhomogeneous due to the differences in the transition probability structure and likelihood. In the former case, the transition probabilities can be directly sampled, as described in Section 4.6.1, while in the latter case, the sampling scheme depends on the specification of the transition probability model (see Section 4.6.2).

4.6.1 Time-Homogeneous Transition Probabilities

For each pair of states (l, m) , let n_{lm} denote the number of one-step transitions in the sampled state vector S^T , $n_{lm} = \sum_{t=1}^{T-1} \#(S_{l,t}S_{m,t+1})$. Conditional on the transition probabilities $\xi_l = (\xi_{l1}, \dots, \xi_{lK})$, the vector $\mathbf{n}_l$ is distributed according to a multinomial distribution with parameters $\sum_m n_{lm}$ and ξ_l (Wasserman, 2004). As a result, the likelihood for $\mathbf{n}$ conditional on ξ is a product of multinomial likelihoods

$$p(\mathbf{n}|\xi) = \prod_{l=1}^K \binom{\sum_m n_{lm}}{n_{l1}, \dots, n_{lK}} \xi_{l1}^{n_{l1}} \dots \xi_{lK}^{n_{lK}}. \quad (4.4)$$

A conjugate prior for the transition matrix ξ is a product of K independent Dirichlet distributions,

$$p(\xi) \propto \prod_{l=1}^K \left(\prod_{m=1}^K \xi_{lm}^{\gamma_{lm}-1} \right). \quad (4.5)$$

Assuming that the rows ξ_i are independent *a priori*, if the prior on row ξ_l is $\text{Dir}(K; \gamma_l)$, Equations (4.4) and (4.5) show that the posterior distribution of ξ_i is the Dirichlet distribution $\mathcal{D}(K; \gamma_l + \mathbf{n}_l)$. Prior knowledge about transition probabilities can be incorporated through the selection of the Dirichlet parameters γ_{lm} .

4.6.2 Time-Inhomogeneous Transition Probabilities

We use a logistic regression model (binomial if the number of states $K = 2$, multinomial if $K > 2$) to specify time-varying transition probabilities in terms of

some external covariates $\mathbf{z}$:

$$\begin{aligned} p(S_t = m | S_{t-1} = l, \mathbf{z}_{t-1}, \boldsymbol{\kappa}) &= \xi_{lm,t} = \text{logit}^{-1} \left(\kappa_{lm,0} + \sum_i z_{t-1,i} \kappa_{lm,i} \right) \\ &= \frac{\exp(\kappa_{lm,0} + \sum_i z_{t-1,i} \kappa_{lm,i})}{\sum_{k=1}^K \exp(\kappa_{lk,0} + \sum_i z_{t-1,i} \kappa_{lk,i})}. \end{aligned} \quad (4.6)$$

To allow the model to be identifiable, the regression parameters associated with transition to a reference state k_0 , $k_0 \in \mathcal{K} = \{1, \dots, K\}$ are fixed at 0, so that

$$p(S_t = k_0 | S_{t-1} = l, \mathbf{z}_{t-1}, \boldsymbol{\kappa}) = \frac{1}{\sum_{\mathcal{K} - \{k_0\}} \exp(\kappa_{lk,0} + \sum_i z_{t-1,i} \kappa_{lk,i})}, \quad (4.7)$$

where $\mathcal{K} - \{k_0\}$ is set-theoretic subtraction. The covariates $\mathbf{z}_t$ can have a strongly nonlinear effect on the transition probabilities (Kaufmann, 2015).

Following the ideas in Frühwirth-Schnatter and Frühwirth (2010), this section contains the derivations of the data augmentation steps for the dRUM (difference random utility model) expression of the multinomial logit model (4.6) as well as the sub-steps required to sample the transition parameters using Gibbs sampling. Frühwirth-Schnatter and Frühwirth (2010) found that dRUM-based samplers (either using a Metropolis-Hastings algorithm or Gibbs sampling) are more efficient than various other Metropolis-Hastings-based sampling methods.

4.6.2.1 Data augmentation for the RUM expression

Writing Equation (4.6) as a random utility model (RUM) requires augmentation by a latent utility for each state:

$$\begin{aligned} S_{k,t}^u &= \mathbf{z}_t' \boldsymbol{\kappa}_k + \nu_{k,t} \quad \forall k \in \mathcal{K}, \nu_{k,t} \stackrel{iid}{\sim} \text{Gumbel} \\ S_t = k &\Leftrightarrow S_{k,t}^u = \max_{j \in \mathcal{K}} S_{j,t}^u, \end{aligned} \quad (4.8)$$

so that the observed state is the state with the maximum utility. This data augmentation scheme is useful, as the latent utilities can be sampled simultaneously, but the resulting sampler is less efficient than those using the dRUM expression (Frühwirth-Schnatter and Frühwirth, 2010; Kaufmann, 2015).

4.6.2.2 Data augmentation for the dRUM expression

The dRUM representation involves expressing the utilities for each non-reference state as the difference between the state utility and the reference state utility. This model becomes

$$\begin{aligned} s_{k,t} &= \mathbf{z}'_t \boldsymbol{\kappa}_k + \tau_{k,t}, \quad \tau_{k,t} \sim \text{Logistic}, \quad \forall k \in \mathcal{K}_{-k_0}, \\ S_t &= \begin{cases} k_0 & \text{if } \max_{j \in \mathcal{K}_{-k_0}} s_{j,t} < 0 \\ k \neq k_0 & \text{if } s_{k,t} = \max_{j \in \mathcal{K}_{-k_0}} s_{j,t} > 0. \end{cases} \end{aligned} \quad (4.9)$$

where $s_{k,t} = S_{k,t}^u - S_{k_0,t}^u$ and $\tau_{k,t} = \nu_{k,t} - \nu_{k_0,t}$. However, unlike the errors $\nu_{k,t}$ in the RUM representation (4.8), the errors $\tau_{k,t}$ in the dRUM representation (4.9) are not independent across states, and so Frühwirth-Schnatter and Frühwirth (2010) use a different representation. Rather than expressing the difference from the reference state utility $S_{k_0,t}^u$, the latent utilities are expressed as the difference from the maximum utility attained across the other states. Following the notation from Kaufmann (2015),

$$\begin{aligned} S_{k,t}^* &= S_{k,t}^u - S_{-k,t}^u \\ D_{k,t} &= I\{S_t = k\}, \end{aligned}$$

where $S_{-k,t}^u = \max_{j \in \mathcal{K}_{-k}} S_{j,t}^u$.

To obtain a closed form expression for $S_{k,t}^*$, first denote $\lambda_{k,t} = \exp(\mathbf{z}'_t \boldsymbol{\kappa}_k)$. Observe that

$$\exp(-S_{-k,t}^u) \sim \text{Exp}(\lambda_{-k,t}),$$

where $\lambda_{-k,t} = \sum_{j \in \mathcal{K}_{-k}} \lambda_{j,t}$, which follows from $\exp(-S_{k,t}^u) \sim \text{Exp}(\lambda_{k,t})$ (as is implied by the Gumbel-distributed errors in Equation (4.8)). Therefore we obtain the partial representation (partial because rather than using the actual class membership, the only information that is used is whether or not y_t is in class k):

$$\begin{aligned} S_{k,t}^* &= S_{k,t}^u - S_{-k,t}^u = \mathbf{z}'_t \boldsymbol{\kappa}_k + \log(\lambda_{-k,t}) + \nu_{k,t} - \nu_{k_0,t} \\ &= \mathbf{z}'_t \boldsymbol{\kappa}_k - \log(\lambda_{-k,t}) + \tau_{k,t}, \quad \tau_{k,t} \stackrel{\text{iid}}{\sim} \text{Logistic}, \\ D_{k,t} &= I\{S_t = k\}. \end{aligned} \quad (4.10)$$

As the constant term $\log(\lambda_{-k,t})$ depends only on $\boldsymbol{\kappa}_{-k}$ (and is therefore independent of $\boldsymbol{\kappa}_k$, Equation (4.10) is a linear regression (with respect to $\boldsymbol{\kappa}_k$) with logistic errors.

4.6.2.3 Sampling scheme for the dRUM

Here we describe the sub-steps involved in sampling the transition probability coefficients $\boldsymbol{\kappa}$.

1. Sample the differenced utilities S^* :

In light of Equations (4.9) and (4.10), $S_{k,t}^* | \boldsymbol{\kappa}_k, S_t$ is distributed according to a logistic distribution truncated to $[0, \infty)$ if $S_t = k$ and $(-\infty, 0]$ if $S_t \neq k$. Therefore, to sample $\mathbf{S}_k^*$, independently draw T uniformly distributed values $W_{k,t} \sim \mathcal{U}([0, 1])$ and set

$$S_{k,t}^* = \mathbf{z}_t' \boldsymbol{\kappa}_k - \log(\lambda_{-k,t}) + F^{-1}(D_{k,t} + W_{k,t}(1 - D_{k,t} - p_{k,t})),$$

where $p_{k,t} = p(D_{k,t} = 1 | \boldsymbol{\kappa}) = 1 - F(-\mathbf{z}_t' \boldsymbol{\kappa}_k + \log(\lambda_{-k,t}))$ and $F(\cdot)$ is the cumulative distribution function of the logistic distribution, with $F^{-1}(q) = \log(q) - \log(1 - q)$.

2. Sample the mixture components R from $p(R | S^*, \boldsymbol{\kappa})$:

Conditioned on $\mathbf{S}^*$, the logistic distribution of the errors $\tau_{k,t}$ is approximated by a mixture of $M = 6$ normal distributions. The components $R_{k,t}$, $k \in \mathcal{K}_{-k_0}$ are drawn from the multinomial distribution with probabilities

$$p(R_{k,t} = i | S_{k,t}^*, \boldsymbol{\kappa}_k) \propto \frac{w_i}{s_i} \exp \left\{ -\frac{1}{2} \left(\frac{S_{k,t}^* + \log(-\lambda_{-k,t}) - \mathbf{z}_t' \boldsymbol{\kappa}_k}{s_i} \right)^2 \right\},$$

where $i = 1, \dots, 6$ and the standard deviations s_i and weights w_i are given in (Frühwirth-Schnatter and Frühwirth, 2010, Table 1), with the relevant table elements reproduced here for convenience as Table 4.1.

3. Sample $\boldsymbol{\kappa}$ from $p(\boldsymbol{\kappa} | S^*, R)$.

With the auxiliary mixture approximation to the logistic errors, conditioned on $\mathbf{S}_k^*$ and $\mathbf{R}_k$, the linear regression (4.10) is normal in $\boldsymbol{\kappa}_k$:

$$\begin{aligned} \tilde{S}_{k,t}^* &= S_{k,t}^* + \log(\lambda_{-k,t}) = \mathbf{z}_t' \boldsymbol{\kappa}_k + \tau_{k,t}, \\ \tau_{k,t} | R_{k,t} &\sim N(0, s_{R_{k,t}}^2), \quad k \in \mathcal{K}_{-k_0}. \end{aligned}$$

r	s_r^2	$100w_r$
1	0.68159	1.8446
2	1.2419	17.268
3	2.2388	37.393
4	4.0724	31.697
5	7.4371	10.89
6	13.772	0.90745

Table 4.1: Variances s_r^2 and weights w_r (multiplied by 100 for expositional convenience) for an approximation of the logistic density by a 6-component mixture of normal densities with mean 0. Reproduced from Frühwirth-Schnatter and Frühwirth (2010, Table 1).

Given a normal prior $p(\boldsymbol{\kappa}_k) = \mathcal{N}(\mathbf{g}_0, G_0)$, the posterior is therefore also normal, $p(\boldsymbol{\kappa}_k | \mathbf{S}_k^*, \mathbf{R}_k) = \mathcal{N}(\hat{\mathbf{g}}_k, \Sigma_k^\kappa)$, with

$$\begin{aligned}\hat{\mathbf{g}}_k &= \Sigma_k^\kappa \left(\sum_{t=1}^T \frac{\mathbf{z}_{t-1} \tilde{S}_{k,t}^*}{s_{R_{k,t}}^2} + G_0^{-1} g_0 \right) \\ \Sigma_k^\kappa &= \left(\sum_{t=1}^T \frac{\mathbf{z}_t' \mathbf{z}_t}{s_{R_{k,t}}^2} + G_0^{-1} \right)^{-1}.\end{aligned}\tag{4.11}$$

4.7 Sampling the Model Parameters

For each state, the remaining parameters are sampled in two blocks: the regression coefficients $\boldsymbol{\beta} = \{\boldsymbol{\beta}_k\}$ and the state-specific error variances $\boldsymbol{\Sigma} = \Sigma_k$, $k = 1, \dots, K$. It is convenient to rewrite Equation (3.5) as

$$\mathbf{y}_t = \sum_{k=1}^K \xi_{k,t} X_t B_k + \varepsilon_t,\tag{4.12}$$

where

$$\xi_{k,t} = \begin{cases} 1 & S_t = k \\ 0 & S_t \neq k \end{cases}$$

is the indicator variable corresponding to the state,

$$B_k = \begin{pmatrix} \boldsymbol{\beta}_{k,1} & \boldsymbol{\beta}_{k,2} \end{pmatrix}$$

and

$$X_t = \begin{pmatrix} \mathbf{x}'_{t,1} \\ \mathbf{x}'_{t,2} \end{pmatrix}.$$

Equation (4.12) can be rewritten by pooling across all t :

$$\hat{\mathbf{y}} = \text{vec}(\mathbf{y}') = \sum_{k=1}^K (I_2 \otimes \Xi_k) \hat{X} \text{vec}(B_k) + \boldsymbol{\epsilon}_t, \quad (4.13)$$

where $\Xi_k = \text{diag}(\boldsymbol{\xi}_k)$, $\hat{X}$ is the block-diagonal matrix

$$\hat{X} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix},$$

where each X_i ,

$$X_i = \begin{pmatrix} \mathbf{x}_{i,1} \\ \vdots \\ \mathbf{x}_{i,T} \end{pmatrix},$$

is the covariate matrix for component $\mathbf{y}_i$, and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \Sigma)$, where

$$\Sigma = \sum_{k=1}^K \Xi_k \otimes \Sigma_k.$$

The conditional distribution of Σ_k is given by

$$p(\Sigma_k | S^T, \mathbf{y}^T, \beta) \propto |\Sigma_k|^{-N_k/2} \exp\left(-\frac{1}{2} \text{tr}(\Lambda_k^{-1} \Sigma_k^{-1})\right), \quad (4.14)$$

where $N_k = \sum_{t=1}^T D_t^{(k)}$ is the total number of observations in state S_k and $\Lambda_k = (U_k' U_k)^{-1}$, in which U_k is a $T \times 2$ matrix collecting the state-conditional errors $\boldsymbol{\epsilon}'_{k,t} = \mathbf{y}'_t - B_k' X_t'$, so that

$$\Lambda_k^{-1} = \begin{pmatrix} |\boldsymbol{\epsilon}_k^{(1)}| & \boldsymbol{\epsilon}_k^{(1)} \cdot \boldsymbol{\epsilon}_k^{(2)} \\ \boldsymbol{\epsilon}_k^{(1)} \cdot \boldsymbol{\epsilon}_k^{(2)} & |\boldsymbol{\epsilon}_k^{(2)}| \end{pmatrix}.$$

(4.14) is the kernel of an inverse-Wishart distribution, so with a conjugate inverse-

Wishart prior $\mathcal{W}^{-1}(\nu_0, \Sigma_0)$, the posterior distribution for Σ_k is

$$\Sigma_k | S^T, \mathbf{y}, \boldsymbol{\beta}_k \sim \mathcal{W}^{-1}(\nu, \Sigma_k^\varepsilon),$$

where

$$\begin{aligned}\nu &= \nu_0 + \frac{N_k}{2} \\ \Sigma_k^\varepsilon &= \Sigma_0 + U_k' \Xi_k U_k.\end{aligned}$$

With a conjugate normal prior $\mathcal{N}(\boldsymbol{\mu}_k, \Omega_k)$ on B_k , the conditional posterior $p(B_k | \mathbf{y}, S^T, \Sigma_k)$ is also normal. At each iteration, the Gibbs sampler draws regression coefficients $B_k \sim \mathcal{N}(\hat{\boldsymbol{\beta}}_k, \hat{\Sigma}_k^\beta)$ and error covariance matrices $\Sigma_k \sim \mathcal{IW}(a_k, \hat{\Sigma}_k)$. In the pooled form (4.13), the moments of the posterior for $\boldsymbol{\beta}_k$ as

$$\begin{aligned}\hat{\Sigma}_k^\beta &= \left(\hat{X}' (\Xi_k \otimes \Sigma_k^{-1}) \hat{X} + \Omega_k^{-1} \right) \\ \hat{\boldsymbol{\beta}}_k &= \hat{\Sigma}_k^\beta \left(\hat{X}' (\Xi_k \otimes \Sigma_k^{-1}) \hat{\mathbf{y}} + \Omega_k^{-1} \boldsymbol{\mu}_k \right).\end{aligned}$$

4.8 Sampling the State Indicators

Viewing the MSVAR model (3.5) as a hierarchical model, the first step in the hierarchy is the latent state Markov process. In Algorithm 4.5.1, sampling realizations of the state indicators is step c), and the other two steps are performed conditionally on those realizations.

Broadly speaking, sampling algorithms for the hidden Markov chain can be considered *single-move* or *multi-move*. Single-move samplers sample the state S_t from the conditional posterior $p(S_t | \mathbf{S}_{-t}, \mathbf{y}, \boldsymbol{\theta})$. Multi-move samplers sample the entire sequence $\mathbf{S}$ from the joint posterior $p(\mathbf{S} | \mathbf{y}, \boldsymbol{\theta})$, which requires knowledge of the filtered probabilities $p(S_t | \mathbf{y}, \boldsymbol{\theta})$. Single-move samplers avoid the need to run a filtering algorithm, but mix more slowly. Many more details on single-move and multi-move samplers can be found in Frühwirth-Schnatter (2006). In this section, we reproduce the multi-move sampling algorithm from Frühwirth-Schnatter (2006, Algorithm 11.5).

Algorithm 4.8.1 (Multi-Move Sampler for State Posterior Chains).

1. Given some initial state distribution $p(S_0 = k|\boldsymbol{\xi})$, the filter is initialized at $t = 0$ with

$$p(S_1 = l|\mathbf{y}^0, \boldsymbol{\phi}) = \sum_{k=1}^K \xi_{kl}(0)p(S_0 = k|\boldsymbol{\xi}), \quad (4.15)$$

where $\xi_{kl}(0) = p(S_1 = l|S_0 = k, \boldsymbol{\phi})$.

Conditional on the model parameters $\boldsymbol{\phi}$, recursively follow the steps below for $t = 1, \dots, T$.

2. Predict S_t conditional on $\mathbf{y}^{t-1}$:

$$p(S_t = l|\mathbf{y}^{t-1}, \boldsymbol{\phi}) = \sum_{k=1}^K \xi_{kl}(t-1)p(S_{t-1} = k|\mathbf{y}^{t-1}, \boldsymbol{\phi}),$$

for $l = 1, \dots, K$, and where $\xi_{kl}(t-1) = p(S_t = l|S_{t-1} = k, \mathbf{y}^{t-1}, \boldsymbol{\phi})$.

3. Filter for S_t :

$$p(S_t = l|\mathbf{y}^t, \boldsymbol{\phi}) = \frac{p(y_t|S_t = l, \mathbf{y}^{t-1}, \boldsymbol{\phi})p(S_t = l|\mathbf{y}^{t-1}, \boldsymbol{\phi})}{p(y_t|\mathbf{y}^{t-1}, \boldsymbol{\phi})},$$

where

$$p(y_t|\mathbf{y}^{t-1}, \boldsymbol{\phi}) = \sum_{k=1}^K p(y_t|S_t = k, \mathbf{y}^{t-1}, \boldsymbol{\phi})p(S_t = k|\mathbf{y}^{t-1}, \boldsymbol{\phi})$$

is the “total likelihood” of the data y_t given the model parameters and the data history.

4. Sample $S_T^{(m)}$ from the filtered probability distribution $p(S_T = j|\mathbf{y}^T, \boldsymbol{\phi})$.
5. For $t = T-1, \dots, 0$, sample $S_t^{(m)}$ from the conditional distribution

$$p(S_t = j|S_{t+1}^{(m)}, \mathbf{y}^t, \boldsymbol{\phi}) = \frac{\xi_{j,l_m}(t)p(S_t = j|\mathbf{y}^t, \boldsymbol{\phi})}{\sum_{k=1}^K \xi_{k,l_m}(t)p(S_t = k|\mathbf{y}^t, \boldsymbol{\phi})}.$$

The initial distribution can be determined *a priori* (e.g., a uniform distribution over the states), or might be estimated from the data. The use of the stationary distribution for the Markov chain complicates the analysis due to the loss of independence between the initial parameter distribution and the transition probabilities in Equation (4.15).

4.9 Model Selection

The flexibility of mixture and Markov-switching models complicates model selection due to the number of possible models with similar structures. Various models specified by (3.3) have differing numbers of latent states and different autoregressive orders. These uncertain features of model specification should be selected simultaneously to reduce the possibility of overfitting (Frühwirth-Schnatter, 2006).

In general, the number of states K and the autoregressive order ρ are not *a priori* known, though there may be domain-specific information about the former. In the context of solar irradiance, the states may be interpreted as different classes of cloud regimes, each of which attenuate and scatter incoming radiation in a statistically similar fashion. For example, if $K = 2$, the states could be (coarsely) classified as “mostly-to-all clear” and “mostly cloudy-to-overcast,” reflecting relatively higher or lower values of k_t and k_n , respectively. If $K = 3$, the states might be interpreted as “clear,” “partially cloudy,” and “cloudy,” with the middle state exhibiting moderate values of k_t and k_n and higher levels of volatility (lower correlations between k_t and k_n and higher variances) than the other two. However, it is not clear which of these classifications better fits the data, or to decide if even more states are required to properly capture the intertemporal dynamics.

A common method for model comparison involves the use of various “information criteria,” such as the Akaike information criterion (AIC) (Akaike, 1974), Bayesian information criterion (BIC) (Schwarz, 1978), deviance information criterion (DIC) (Spiegelhalter et al., 2002), or Watanabe-Akaike information criterion (WAIC) (Watanabe, 2010). Each of these criteria have a particular niche, based on their derivation and the parameters used to estimate the model score. All penalize models that are overly complex, and therefore possibly overfit the data.

However, while AIC, DIC, and WAIC are intended to score models based on an estimate of their out-of-sample predictive accuracy, BIC is aimed at identifying the most probable data-generating model among the candidates by approximating the marginal likelihood $p(\mathbf{y})$ of the model (Gelman et al., 2014). DIC and WAIC are Bayesian methods, while AIC and BIC use maximum-likelihood estimates of the model parameters. Gelman et al. (2014) prefers WAIC to DIC as WAIC uses simulations from the joint posterior distribution, while DIC uses a point estimate of the model parameters (the posterior means).

As a result of this dichotomy, we should consider separately the classes of metrics that are intended for predictive models and those that are intended to distinguish the most “true” data-generating model out of the set of models under consideration.

4.9.1 Predictive Information Criteria

Predictive criteria are designed to estimate the quality of predictive fit of the model to an out-of-sample data point $\tilde{y}$. The most commonly used predictive criteria, such as AIC, DIC, and WAIC, are built off of the log predictive density $p(\tilde{y}|\boldsymbol{\theta})$. Model selection using these criteria is based on the observation that, in the large-sample limit, the model with the highest log predictive density will have the highest posterior probability (based on Kullback-Leibler information) (Gelman et al., 2014). The log predictive density of the model is then estimated by computing log pointwise predictive density across the dataset (this may not literally be for each data point, but may involve a coarser partitioning of the data, depending on considerations such as computational time or spatial or temporal dependence) and adding a bias-correcting factor which is intended to penalize overfit resulting from the use of the same dataset for inference and validation.

One important point (mentioned in the review by Gelman et al. (2014)) is that these criteria use the *predictive* density of the model, rather than the *posterior* density. This avoids the prior density from biasing the fit of the model output to data.

AIC is the most commonly used predictive information criterion. However, AIC uses the maximum likelihood parameter estimate, and therefore may be difficult to employ in the context in Bayesian inference. The asymptotic properties underlying AIC’s convergence to the Kullback-Leibler information also may fail to hold for singular models, a class which includes most hierarchical and mixture models (Watanabe, 2010). DIC is a Bayesian information criterion that uses the posterior means for its calculation; however, following Gelman et al. (2014), we use WAIC, which estimates the log predictive density over the entire posterior, rather than collapsing the posterior onto a point estimate.

The formula for WAIC that we use is

$$\text{WAIC} = -2 \left(\sum_{i=1}^n \log \left(\frac{1}{S} \sum_{s=1}^S p(y_i | \boldsymbol{\theta}^s) \right) - \sum_{i=1}^n V^S \left(\log p(y_i | \boldsymbol{\theta}^2) \right) \right), \quad (4.16)$$

where $\boldsymbol{\theta}^s$ are an ensemble of draws from the posterior distribution, and V^S denotes the sample variance. The first term is the expected log pointwise predictive density, and the second corrects for overfitting. Gelman et al. (2014) interprets the correction term as approximately penalizing the number of “unconstrained” parameters in the model, so that parameters that are fully penalized are those which are estimated entirely from the data. The factor -2 puts the value of WAIC on the deviance scale, so that its value is comparable to AIC or DIC, and as a result, the model with the lowest WAIC value has the best predictive fit.

4.9.2 “True” Data Generation Criteria

4.9.2.1 Loss Functions and Marginal Likelihoods

Let $\mathcal{M}_1, \dots, \mathcal{M}_{L_{\max}}$ be the set of models under consideration. In the setting of trying to identify the “true” data-generating model, decision-theoretic criterion for selecting a particular model $\mathcal{M}_{\hat{L}}$ is that the loss associated with selecting an incorrect model should be minimized. Denote by $R(\mathcal{M}_{\hat{L}}, \mathcal{M}_L)$ the loss corresponding to model $\mathcal{M}_{\hat{L}}$ if $\mathcal{M}_L$ is “true.” The loss associated with model $\mathcal{M}_L$ is a random quantity, and its expectation may be estimated in terms of the posterior evidence for each alternative model $p(\mathcal{M}_L|\mathbf{y})$ (called the posterior probability of model $\mathcal{M}_L$):

$$E(\mathcal{M}_{\hat{L}}) = \sum_{L=1}^{L_{\max}} R(\mathcal{M}_{\hat{L}}, \mathcal{M}_L) p(\mathcal{M}_L|\mathbf{y}).$$

The expected loss is of course sensitive to the selection of the loss function $R(\cdot, \cdot)$, and different models with similar marginal likelihoods may be selected under different imposed loss functions. A standard choice in the absence of additional information is the 0/1 loss function (Frühwirth-Schnatter, 2006),

$$R(\mathcal{M}_{\hat{L}}, \mathcal{M}_L) = \begin{cases} 0 & \hat{L} = L \\ 1 & \hat{L} \neq L, \end{cases}$$

for which the expected loss simplifies to

$$E(\mathcal{M}_{\hat{L}}) = \sum_{L=1, L \neq \hat{L}}^{L_{\max}} p(\mathcal{M}_L|\mathbf{y}) = 1 - p(\mathcal{M}_{\hat{L}}|\mathbf{y}),$$

so that the model with the highest posterior probability minimizes the risk of selecting an incorrect model. By Bayes' rule (4.1), this probability factorizes

$$p(\mathcal{M}_{\hat{L}}|\mathbf{y}) = p(\mathbf{y}|\mathcal{M}_{\hat{L}})p(\mathcal{M}_{\hat{L}}), \quad (4.17)$$

so that if all models have the same prior probability $p(\mathcal{M}_{\hat{L}})$, choosing the model with the highest marginal likelihood $p(\mathbf{y}|\mathcal{M}_{\hat{L}})$ will minimize the expected loss. However, computing the marginal likelihood for a Markov switching model is a non-trivial integration problem in general:

$$p(\mathbf{y}|\mathcal{M}_L) = \int_{\mathcal{S}_K \times \Phi_K} p(\mathbf{y}|\mathbf{S}, \phi_K) p(\mathbf{S}|\xi) p(\phi_K) d(\mathbf{S}, \phi_K).$$

Another approach is to use variable-dimension methods, such as reversible jump MCMC (Richardson and Green, 1997). These methods jointly estimate the posterior probabilities of the model structure and the model parameters. We do not explore these methods due to their increased computational cost, though reversible jump MCMC is relatively tractable when comparing nested models such as we do here (Robert and Casella, 2004).

Finally, one could use Bayesian model averaging methods, in which model output is weighted based some probability distribution. For example, when the loss function is quadratic (rather than the 0-1 loss above), the models are weighted by their posterior probabilities (Robert and Casella, 2004). Reversible jump MCMC also produces the posterior probability of each model structure along with the parameter posterior distributions.

4.9.2.2 Approximations to the Marginal Likelihood

BIC is a commonly used (asymptotic) approximation to the marginal likelihood, but is known to perform poorly for complex (and singular) models such as Markov-Switching GARCH models (Chuffart, 2015). Instead, to select K and ρ for a given model, we will directly estimate the marginal likelihoods $p(\mathbf{y})$ (Fruhwirth-Schnatter, 2004)). Denoting the set of candidate models by $\mathcal{M}_1, \dots, \mathcal{M}_{L_{\max}}$, consider the posterior probability of a model $p(\mathcal{M}_{\hat{L}}|\mathbf{y})$. The factorized posterior (4.17) shows that, under a uniform prior distribution, the model with the greatest marginal likelihood minimizes the 0-1 loss. Of course, other prior distributions could be

used other than the uniform distribution, for example, to help select for more parsimonious models (in accordance with Occam’s razor). Another possibility is to use the estimated model posterior probabilities as model averaging weights (Wong et al., 2017), though we do not do so here.

Computing $p(\mathbf{y})$ is in general a difficult problem. Following Frühwirth-Schnatter (2001) and Frühwirth-Schnatter (2004), we estimate the marginal likelihood using the method of bridge sampling (Meng and Wing, 1996), described in more detail in Appendix A. The bridge sampling method requires computation of the mixture likelihood, which can be expressed as a mixture of complete-data likelihoods (4.2),

$$p(\mathbf{y}|\boldsymbol{\theta}) = \sum_{\mathbf{S} \in \mathcal{S}} p(\mathbf{y}|\mathbf{S}, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K) p(\mathbf{S}|\boldsymbol{\xi}), \quad (4.18)$$

where $\mathcal{S} = \{1, \dots, K\}^{T+1}$ is the space of all realizations of $\mathbf{S}$. As a result, the expression (4.18) for the mixture likelihood becomes intractable quickly as the number of states and length of the data series grows. Instead, by considering the product of one-step ahead predictive densities, we obtain

$$p(\mathbf{y}|\boldsymbol{\theta}) = \pi_{t=1}^T \left(\sum_{k=1}^K p(y_t|S_t = k, \boldsymbol{\theta}) p(S_t = k|\mathbf{y}^{t-1}, \boldsymbol{\theta}) \right),$$

where the terms in the left-hand side product are the normalizing constants $p(y_t|\mathbf{y}^{t-1}, \boldsymbol{\theta})$ from running the filtering steps of Algorithm 4.8.1.

Marginal likelihoods, just like any other model comparison method, should not be treated as a black box. Indeed, the marginal likelihood is highly sensitive to the choice of prior (Gelman et al., 2013). Further, when comparing the posterior model probability given by 4.17, there is no necessary reason to insist on the use of a uniform distribution over all models, in which case the model with the lowest marginal likelihood would not necessarily be the one with the lowest posterior probability. Rather, we use marginal likelihood estimates to narrow the range of potentially suitable models, and then examine their point process representations (viewing model parameter simulated values as if they were points in Euclidean space) visually and using k-means clustering to look for signs of overfitting (Frühwirth-Schnatter, 2006). Those models with evidentiary support which are identifiable are then calibrated, and the statistical properties of an ensemble of simulated irradiance series are then compared to the validation data set.

Chapter 5 |

Hourly Solar Resource Assessment Modeling

As discussed in Section 2.5.1, characterization of the solar resource at a site is important for the design and simulation of a SECS. Many SECS (for example, plants or buildings) are less influenced by sub-hourly variations in incident solar irradiation than by longer-scale trends. In their seminal work, Liu and Jordan (1960) observed that much of the variability in the daily average clearness index k_t between months can be explained by the variations in the monthly average k_t , so that months with similar monthly averages have similarly distributed daily averages. Hollands and Huget (1983) drew the same conclusion about the relationship between the distribution of hourly average k_t within a day and the daily average.

In this chapter, we investigate the differences between a benchmark model trained on an all-year dataset versus a collection of models trained on individual local seasons (in this case, winter, spring, fall, and summer). The choice to use seasons, rather than downscale further to months, is based on the identification of seasons with dominant climatological regimes. Figures 5.1 and 5.2 show generally consistent distributions of k_t and k_b across months within the same season (with a large outlier in December), while Figure 5.3 indicates that the seasons correspond to distinct k_b and k_t distributions.

5.1 Data

Solar irradiance data for this study was obtained from the NOAA SURFRAD site at Rock Springs, PA (Augustine et al., 2000) from January 1, 2007 through

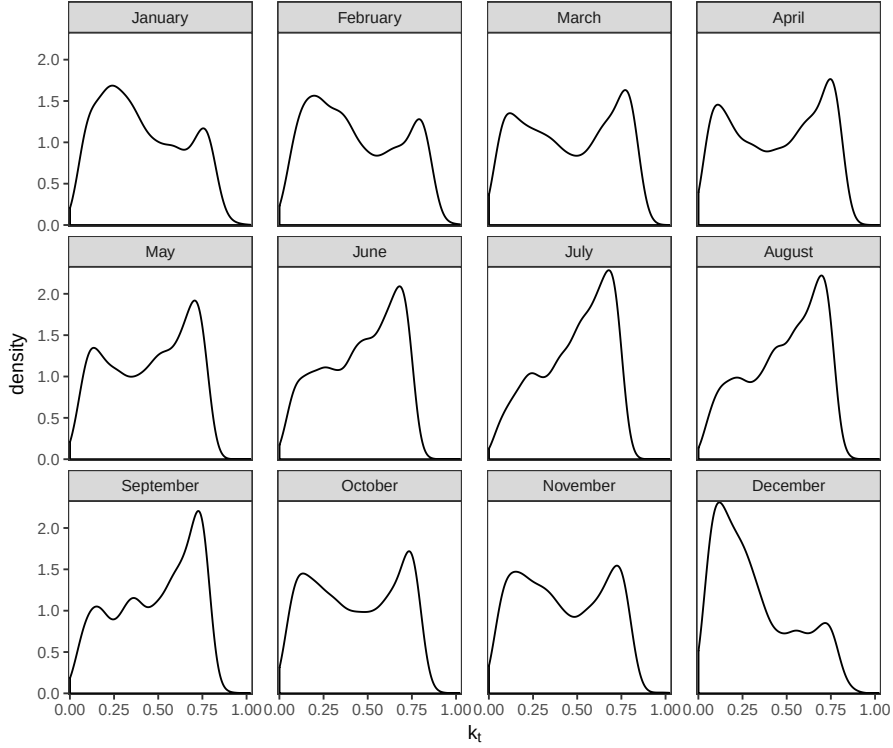


Figure 5.1: Distributions of hourly k_t values by month. Data was collected at the Rock Springs, PA SURFRAD site, from January 1, 2007 through December 31, 2016.

December 31, 2016. The site is located at 40.72° N and 77.93° W. DNI is measured using an Eppley Normal Incidence Pyrheliometer (NIP). While GHI is directly measured, due to issues relating to thermal calibration it is advised to compute GHI using DNI (after conversion to BHI) and DHI (which is measured using an Eppley 8-48 pyranometer). All measured values are averaged over 1 minute before reporting. Quality control for the data was handled using the flags included with the data.

The resulting data set was randomly divided into three sets by calendar year, in accordance with the “tune/train/test” separation paradigm (Haupt et al., 2009). The tuning/model selection dataset consisted of data from 2007, 2010, and 2015, the model calibration dataset consisted of data from 2009, 2011, 2013, and 2016, and the validation dataset consisted of data from 2008, 2012, and 2014.

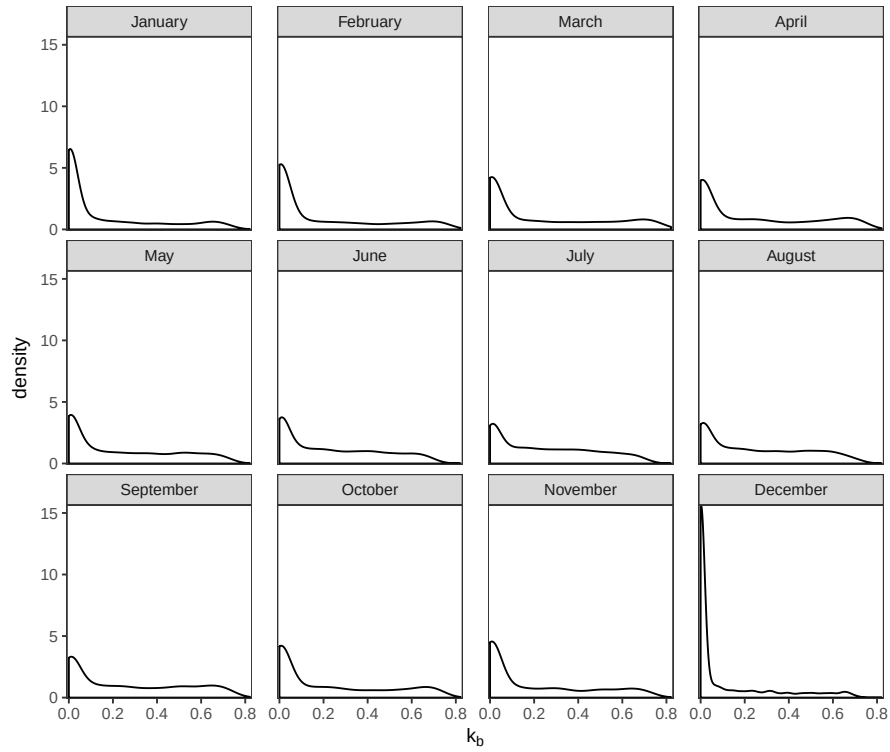


Figure 5.2: Distributions of hourly k_b values by month. Data was collected at the Rock Springs, PA SURFRAD site, from January 1, 2007 through December 31, 2016.

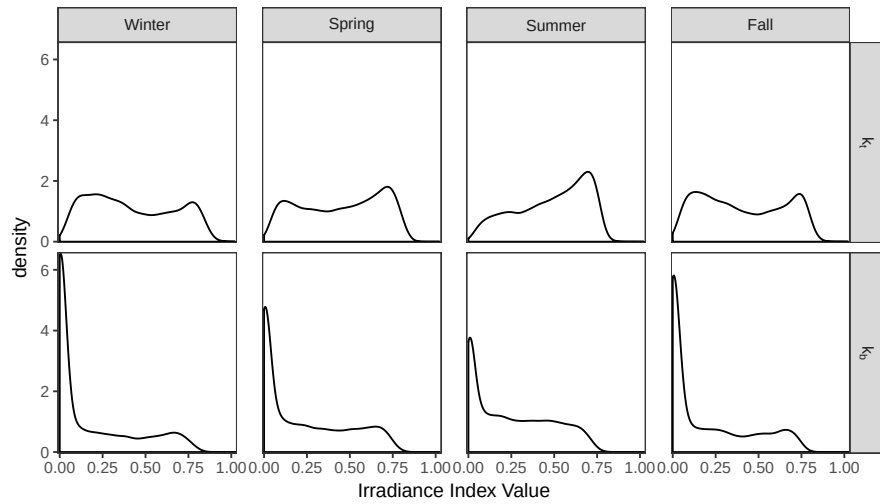


Figure 5.3: Distributions of hourly k_t and k_b values by season. Data was collected at the Rock Springs, PA SURFRAD site, from January 1, 2007 through December 31, 2016.

Table 5.1: Prior distributions for the resource model parameters. B_k denotes the vectorized regression parameters β_k , as described in Section 4.7.

Parameter	Prior Distribution	Comments
ξ_k	$\mathcal{D}(K; e_{k1}, \dots, e_{kK})$	$e_{kk'} = \begin{cases} 3 & k = k' \\ 2 & k \neq k' \end{cases}$
B_k	$\mathcal{N}(\mu_k, I)$	$\mu_k = \left(-2 + \frac{4(k-1)}{(K-1)}, \mathbf{0}_{\rho+1}, -2 + \frac{4(k-1)}{(K-1)}, \mathbf{0}_{\rho+1}\right)$
Σ_k	$\mathcal{W}^{-1}(3, I)$	

5.2 Model Specification

For the reasons discussed in Section 3.5.2, our model is specified as follows:

$$\mathbf{y}_t = \beta_{0,S_t} + \beta_{1,S_t} \cos(\theta_z) \cdot \mathbf{1}_2 + \sum_{i=1}^{\rho} \beta_{i+1,S_t} \mathbf{y}_{t-i} + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \Sigma_{S_t})$, $\mathbf{1}_2 = (1, 1)^T$, and $\mathbf{y} = (\hat{k}_t, \hat{k}_b)^T$. Further, we assume that the latent Markov chain generating the state sequence S_t is homogeneous, so that the transition probabilities ξ are constant.

Our selection of prior distributions, summarized in Table 5.1, is largely determined by our desire to use Gibbs sampling in the sampling steps of Algorithm 4.5.1, which requires conjugate distributions. The choice between Gibbs sampling and the use of Metropolis-Hastings steps with other, perhaps more informative priors, represents a tradeoff between computational effort and the strength of the prior information. In general, relatively little prior meteorological information is included in the prior distributions, both to help this model serve as a benchmark and to avoid giving the states an *a priori* physical interpretation. This latter consideration is not crucial except that this work is conducted with a view towards model structural uncertainty (in the autoregressive order ρ and number of states K), and so it is not necessarily clear how the states ought to correspond to particular meteorological regimes as the number of states varies. Stronger priors could be used if such an *a priori* correspondence was established.

The Dirichlet prior for ξ_k is defined to bound the Markov-switching process away from an i.i.d state process (by setting $e_{kk} > e_{kk'}$), in which case path-dependency would be eliminated. The non-zero prior means for the intercepts $\beta_{0,k}$ are intended to separate the states, though experimentation suggests that the prior mean has

relatively little influence on the posterior distribution. The 0 mean value for the other regression coefficients are intended primarily to regularize the sampled values. Finally, the number of degrees of freedom for the inverse-Wishart prior for Σ_k is selected to leave the prior only weakly informative, though inverse-Wishart priors are known to be relatively restrictive (Gelman et al., 2013).

5.3 Results

5.3.1 Model Selection

Models with AR coefficients $\rho = 1, 2$ and number of states $K = 2, \dots, 5$ were compared using the bridge-sampling estimate of their marginal likelihoods (Appendix A). The MCMC run for the model selection step used the tuning dataset, and was run for 75,000 iterations. The first 25,000 were discarded as burn-in (convergence of the sampler was tested using the Gelman-Rubin diagnostic (Gelman and Rubin, 1992)). An ensemble of 10,000 posterior samples were then used to compute the marginal likelihood estimate. The resulting log-likelihood estimates and relative standard errors are displayed in Tables 5.2 and 5.3.

Table 5.2: Bridge-sampling estimates of the marginal log-likelihoods $\log \hat{p}(\mathbf{y})$ for models with number of states $K = 2, \dots, 5$ and autoregressive order $\rho = 1, 2$. The bolded values correspond to the maximum marginal likelihood for each dataset.

Season	ρ	K			
		2	3	4	5
winter	1	-2021.6	-1969.9	-1904.0	-1880.5
	2	-2067.0	-1995.5	-1999.3	-1879.6
spring	1	-2336.6	-2373.7	-2251.9	-2221.9
	2	-2029.9	-1885.9	-2055.3	-1987.3
summer	1	-1642.7	-1691.3	-1658.5	-1575.5
	2	-1451.5	-1299.6	-1553.8	-1419.8
fall	1	-1582.3	-1547.3	-1577.7	-1372.7
	2	-1384.2	-1419.0	-1393.8	-1349.0
all	1	-6804.3	-6965.9	-5903.3	-6380.9
	2	-5926.7	-5814.9	-6066.0	-5622.2

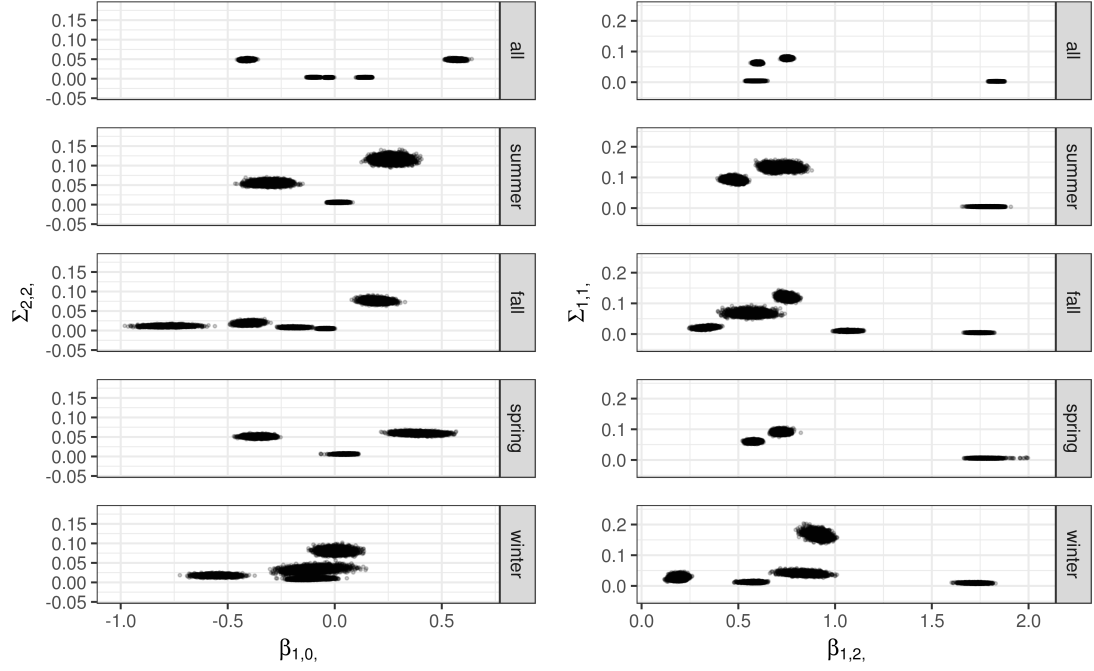
Table 5.3: Estimates of the expected relative mean-squared error for the log-marginal likelihood estimates $\log \hat{p}(\mathbf{y})$ in Table 5.2.

Season	ρ	K			
		2	3	4	5
winter	1	0.0024	0.0033	0.0054	0.0089
	2	0.0024	0.0030	0.0066	0.0067
spring	1	0.0020	0.0024	0.0038	0.0041
	2	0.0019	0.0028	0.0032	0.0041
summer	1	0.0019	0.0024	0.0037	0.0041
	2	0.0019	0.0029	0.0035	0.0045
fall	1	0.0017	0.0019	0.0032	0.0037
	2	0.0021	0.0027	0.0035	0.0042
all	1	0.0020	0.0024	0.0030	0.0061
	2	0.0020	0.0023	0.0036	0.0193

Based on Table 5.2 and a visual inspection of the point process representation of the posterior draws (to rule out overfitting), MSAR(2) models with $K = 3$ were selected for the spring and summer seasons, and MSAR(2) models with $K = 5$ were selected for the other models.

5.3.2 Identifiability Constraints

To calibrate the models, the MCMC sampler was run on the training dataset for 100,000 iterations, with a burn-in period of 30,000 iterations (convergence was again diagnosed with the Gelman-Rubin diagnostic). The resulting point process representations are shown in Figure 5.4. While most of the states are clearly separated, the winter model requires a separate set of constraints than the others. Figure 5.4 suggests that for the non-winter models, the states can be ordered by the first-component (corresponding to k_t) intercept $\beta_{0,1}$. The winter model can be identified by a constraint combining the lag-2 first-component AR coefficient $\beta_{1,2}$, with the first-component variance $\Sigma_{1,1}$.

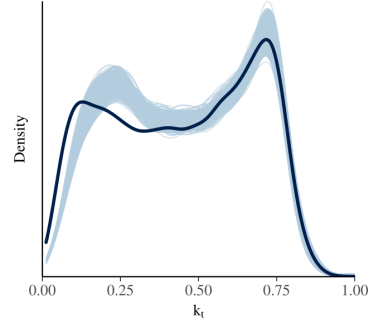


(a) Point process representation of $\beta_{1,0}$ vs. $\Sigma_{2,2}$. (b) Point process representation of $\beta_{1,2}$ vs. $\Sigma_{1,1}$.

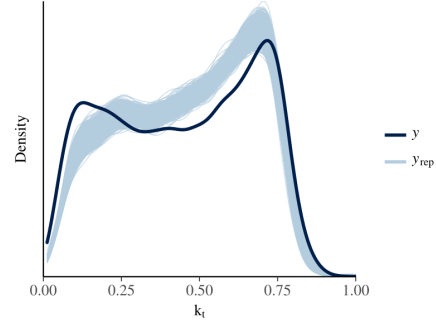
Figure 5.4: Point process representation of the posterior samples for the selected models.

5.3.3 Posterior Predictive Checks

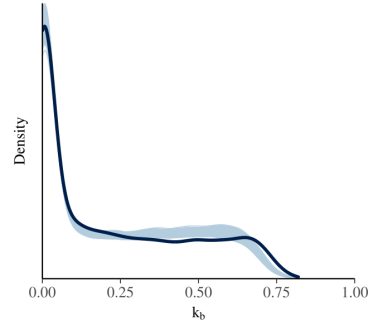
To test how well the seasonal and all-year models perform as stochastic solar resource generators, we rely on graphical representations of simulated datasets from the posterior. While the use of the Kolmogorov-Smirnov integral is recommended for resource assessment model validation (Sengupta et al., 2015), our goal is not to reproduce the validation dataset (which may be, and in this case is, significantly shorter than the data lengths required for appropriate levels of confidence in a frequentist sense (Vignola et al., 2012)), but rather to examine the uncertainty of simulations from the posterior. The validation dataset serves as a benchmark (under the assumption that it does not consist of outlying data); with a sufficient number of realizations, the validation data (and any test statistics of interest) ought to be enveloped by the posterior predictive distribution. We focus in this section on the dimensional GHI and DNI values, as those are the quantities of interest for



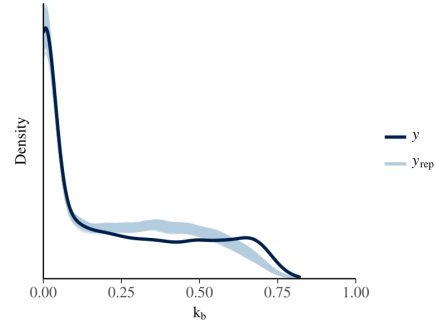
(a) Hourly-averaged k_t density for simulations from the all-year model.



(b) Hourly-averaged k_t density for simulations from the seasonal models.



(c) Hourly-averaged k_b density for simulations from the all-year model.



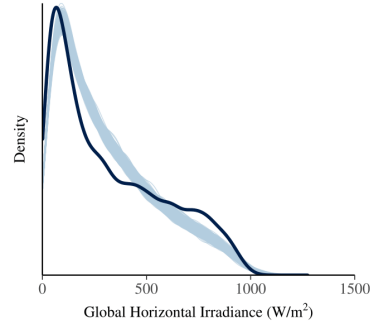
(d) Hourly-averaged k_b density for simulations from the seasonal models.

Figure 5.5: Densities of hourly k_t and k_b for the validation dataset (y , the black line) and the simulated datasets (y_{rep} , the blue lines).

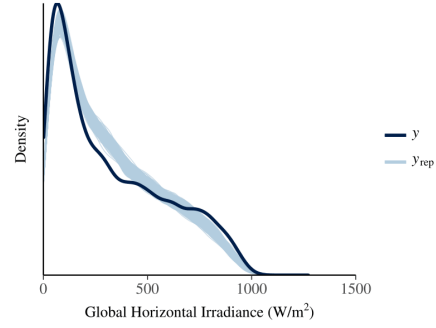
SECS modeling.

We generated 1,000 simulated datasets (each dataset corresponding to three years). The initial k_t and k_b values for each day in the simulated datasets were the same as those for the corresponding days in the validation dataset, though the initial state S_0 was sampled from a uniform distribution over the states. The empirical CDF used to transform k_t and k_b values to quantiles, and vice versa, was the empirical CDF from the training dataset, to avoid overfitting the simulations to the validation distribution.

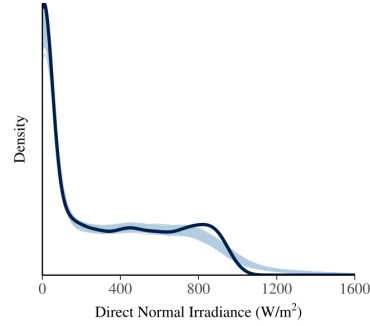
Figure 5.5 compares the marginal distributions of k_t and k_b values from the validation dataset to those from the simulated datasets. In general, the all-year model does a better job fitting the validation data, which is not surprising given it



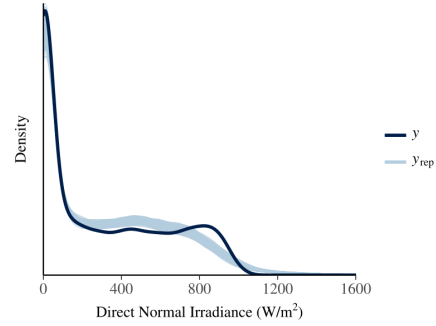
(a) Hourly-averaged GHI density for simulations from the all-year model.



(b) Hourly-averaged GHI density for simulations from the seasonal models.



(c) Hourly-averaged DNI density for simulations from the all-year model.

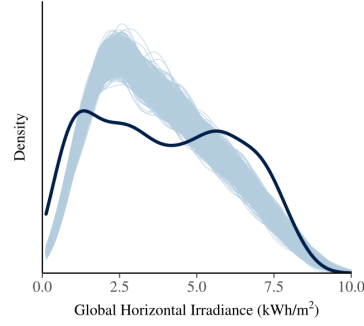


(d) Hourly-averaged DNI density for simulations from the seasonal models.

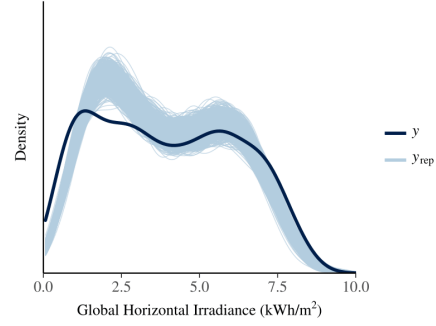
Figure 5.6: Densities of hourly-averaged GHI and DNI for the validation dataset (y , the black line) and the simulated datasets (y_{rep} , the blue lines).

was fit on four times the data of any of the individual seasonal models. However, when examining the hourly GHI and DNI marginal distributions in Figure 5.6, there does not appear to be a strong difference between the two models, suggesting that the effect of the solar diurnal cycle is more important than the differences between the dimensionless output of the two models. Both models produce extreme GHI values too infrequently, shifting the probability mass towards the center of the distribution. Both models also fail to capture the upper DNI mode and have right tails that are too long, though they do capture the point mass at 0 W/m^2 .

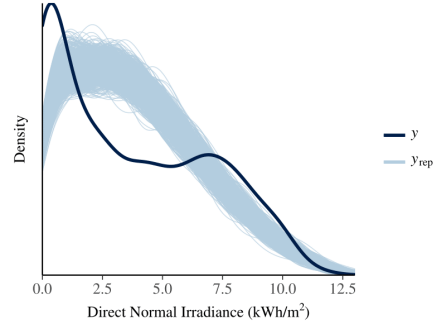
We are also interested in testing the model's time-series properties. The daily GHI and DNI sums demonstrate whether the intradaily dynamics of each series are reasonable approximation of those found in the validation data. Figure 5.7



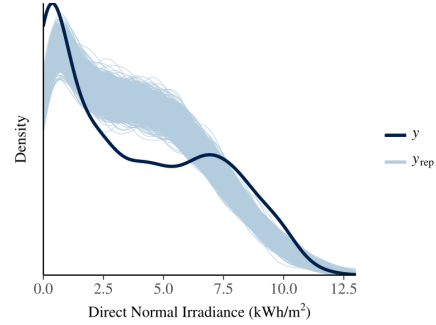
(a) Daily GHI densities for simulations from the all-year model.



(b) Daily GHI densities for simulations from the seasonal models.



(c) Daily DNI densities for simulations from the all-year model.



(d) Daily DNI densities for simulations from the seasonal models.

Figure 5.7: Densities of daily GHI and DNI for the validation dataset (y , the black line) and the simulated datasets (y_{rep} , the blue lines).

shows that there is considerably more uncertainty than for the hourly series. The all-year model produces completely overcast days with a lower frequency than the seasonal model, and neither produce them often enough to match the validation data, with more mass on days with 2-3 kWh/m². The seasonal model does capture the bimodality (and the upper mode) of the daily GHI distribution, but the right tail is thinner than the validation data. The all-year model, on the other hand, produces a unimodal GHI distribution.

Neither model appropriately captures the frequency of high-beam (DNI) days, and as with the GHI results, neither captures days with effectively no beam irradiance. The seasonal model does a better job of placing the appropriate mass on lower DNI days (likely due to the presence of states corresponding to highly overcast conditions in the winter and fall models, which are absent in the all-year model), but both models overpredict days with moderate DNI.

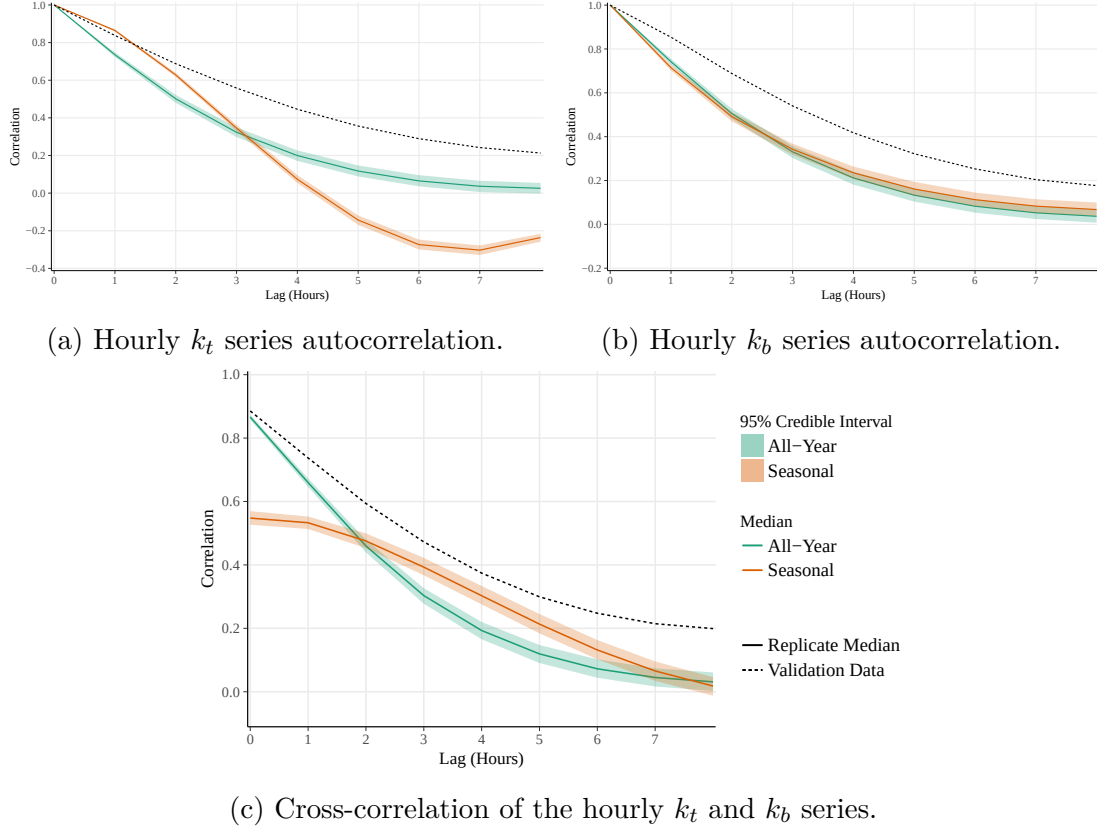


Figure 5.8: k_t and k_b auto- and cross-correlations for the validation data and the simulated datasets. The ribbons correspond to the 95% credible interval across the simulations.

Both models also feature too little autocorrelation within their dimensional and non-dimensional replicates, as seen in Figures 5.8 and 5.9. This under-correlation is much more pronounced in the DNI series than the GHI series (which may help to explain the failure of both models to properly capture daily DNI extremes), even though the seasonal model's k_t autocorrelation deviates significantly from the observed data beginning at a lag of 3 hours, suggesting that closeness of the autocorrelations in the dimensional series is the result of the diurnal cycle. On the other hand, the effect of dimensionalization is to increase the distance between the observed DNI autocorrelation and the modeled DNI autocorrelation relative to that of the dimensionless k_b series. The joint distribution of GHI and DNI is also off, as the lag-0 cross-correlation between the two series is much lower than in the validation data.

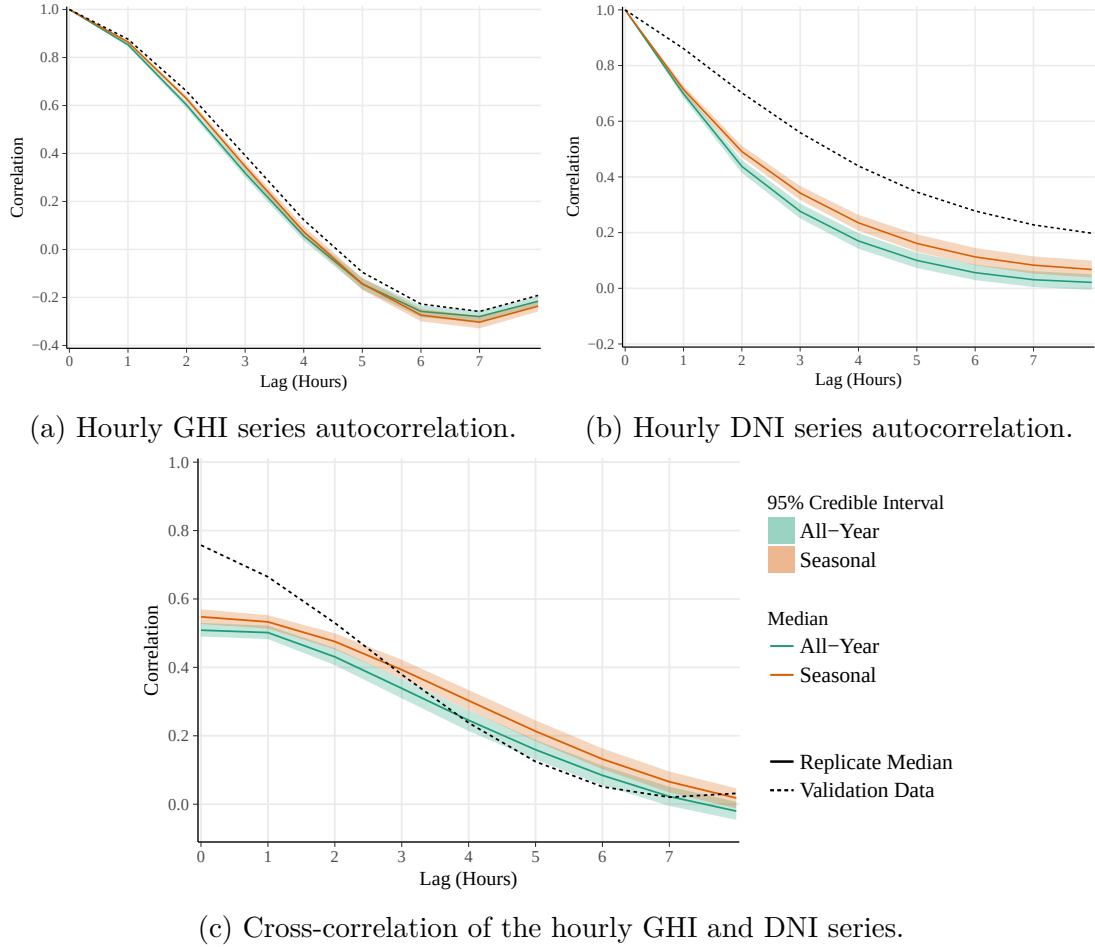


Figure 5.9: GHI and DNI auto- and cross-correlations for the validation data and the simulated datasets. The ribbons correspond to the 95% credible interval across the simulations.

5.4 Discussion

The goal of this chapter was to provide a benchmark, through the use of a relatively simple model, for a probabilistic Markov-switching approach to solar resource assessment modeling, to allow for the assessment of the marginal contributions of more complex models. The results from Section 5.3.3 demonstrate that the model used here, with homogeneous transition probabilities and little prior meteorological information, does not properly capture the probability of extreme GHI or DNI hours or days and fails to properly reproduce the temporal properties of real GHI and DNI series. There are a number of possible ways that these shortcomings can

be addressed by the next round of modeling.

First, the use of the empirical CDF of the training data likely resulted in some degree of overfit of the simulated data to the distributions found in the training dataset. This is more likely to be a problem when the training data is relatively limited. While the generalized linear modeling approach (using the logit as a link function) was observed to not work well in this, relatively unconstrained, study, a more constrained model (requiring stationary dynamics around a switching mean, for example, as in Equation (3.9) with constrained autoregression coefficients) might work well, particularly if the prior distributions are more informative. Inference for models requiring this type of dependence on previous states is easier if they are cast in the form of a switching state-space model (Kim, 1994; Frühwirth-Schnatter, 2006). This framework also facilitates the incorporation of random walk processes for the mean or correlated innovations within the continuous state process, which is typically estimated using the Kalman filter (Kalman, 1960; Kalman and Bucy, 1961).

Second, homogeneous transition probabilities result in daily series that cannot properly capture diurnal trends, which further affects the ability of the model to represent realistic daily series. A new version of this model should incorporate the fractional time of day into the Markov-switching transition process, and the transition probabilities could also depend on the duration of the current regime. An additional consideration is that the hourly clearness indices are known to be dependent on the daily-average clearness index (Liu and Jordan, 1960; Hollands and Huget, 1983), which can be modeled by an AR(1) process (Bartoli et al., 1981; Graham et al., 1988). An AR(1) process for the daily-average k_t and k_b could be used as an additional hierarchical level upon which the transition probabilities are further conditioned.

Finally, it will be important to find the initial distribution of the latent states, $p(\mathbf{S}_0)$. In this work, we have assumed that this distribution was uniform, which is obviously unrealistic. This inference could be done endogenously, using the smoothed state probabilities for each day (and possibly conditioned on the daily average k_t and k_b), or could be done exogenously, *e.g.* using an analysis of satellite images. While the latter approach requires an *a priori* identification of the states with particular sets of meteorological conditions, we feel that this is necessary in any case, as it allows for more informative prior distributions to be incorporated

into the model. This also allows for the latent state sequences to be identified exogeneously during the training-data period.

Chapter 6 |

Forecasting Components of Solar Irradiance

As the penetration of renewable sources of energy increases, the uncertainty of their flows result in a need for more frequent power systems dispatching (Ela et al., 2013). One consequence has been an increase in interest in solar irradiance forecasting for sub-hourly forecast horizons. The uncertainty associated with solar irradiance forecasts makes probabilistic forecasting attractive, as probabilistic information can be collapsed to point estimates if the decision-making procedure requires these. In this chapter, we use the MSVAR model from Chapter 3 to probabilistically forecast GHI and DNI for a 15-minute ahead time horizon.

One difference between the model used for forecasting in this chapter and that used for resource assessment in Chapter 5 is that here we allow covariates derived from geostationary satellites to drive the transition probabilities for the latent state process. Our motivation in using satellite imagery in this fashion (rather than deriving irradiance estimates directly from the image values, as in Perez et al. (1992) and Perez et al. (2002)), is to incorporate uncertainty about future transitions. These transitions between cloud regimes can be the result of a combination of advection (transport) and evolution (cloud building/dying). The temporal and spatial scales over which these two drivers dominate vary depending on the cloud regime, as quantified by data collected by Fujita (1981) and visualized in Figure 2.3.

For the most stochastic cloud fields, deterministic frozen-field forecasts from geostationary satellite images have a reasonable cloud-forecasting skill through no more than a 30-minute lead time, which is consistent with Figure 2.3 (Srikrishnan et al., 2017). Rather than taking this approach, by assimilating satellite data into

the transition process, we are able to account for uncertainties related to both advection and evolution.

6.1 Data

6.1.1 Satellite Data

The satellite data used for this study is obtained from the NOAA GOES-13 geosynchronous satellite (GOES-EAST); specifically, images from the visible channel of the GOES imager, with a spatial resolution of 1 km per pixel. Both CONUS (Continental United States) and Extended Northern Hemisphere scans are used to allow images to be available every 15 minutes. This data is formatted as 10-bit GVAR (GOES Variable Format) counts, which are linearly related to the visible radiance of the area represented by a pixel.

The area covered by the images used in this work covers latitudes 34° – 47° N and longitudes 74° – 92° W. This area is large enough to account for wind vectors coming from the west, southwest, and northwest corresponding to a maximum wind speed of 133.33 m/s (120 km per 15 min), which is greater than the maximum winter jet speed in the region (Imhoff, 2015). Images were obtained corresponding to the daytime hours of 11:00–24:00 UTC. While it might be desirable for forecasts to be generated at nighttime for use in the time period shortly after sunrise, this would require the use of the infrared imager channels.

These raw images are corrected to adjust for the first-order projection effects due a varying solar zenith angle (Perez et al., 2002), which removes the diurnal trend resulting from solar geometry from the mean image intensity. The images are then subjected to a cloud-masking procedure to increase the contrast between cloud and clear-sky pixels. Under the assumption that the distribution of pixel values is generated by a “clear” process and a “cloudy” process, a two-component Gaussian mixture model (McLachlan and Peel, 2000) is fit to the GVAR distribution. Pixels are flagged as clear if their value is no greater than the value at which the tails of the two Gaussians cross. Any pixels flagged as clear then had their value set to 0. An example of this cloud masking procedure can be seen in Figure 6.1.

The assumption that the GVAR distribution is generated by two Gaussians is plausible so long as there are no snow-covered areas in the image domain. We

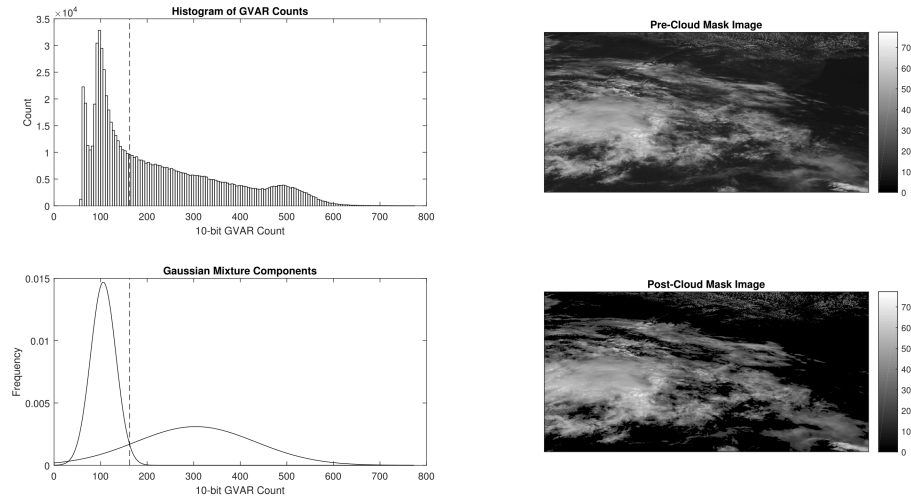


Figure 6.1: Results of the cloud masking procedure for a sample GOES image. Shown are the pre-cloud masking GVAR histogram, the fit Gaussian mixture components, and the pre- and post-cloud masking images. The dashed line in the histogram and the mixture fit images is the value separating “cloudy” from “clear” pixels. Reproduced from Srikrishnan et al. (2017).

avoid the issue of identifying snow-covered pixels due to the difficulty of separating snow cover from clouds using a single satellite image channel (Miller et al., 2005). Images are treated as missing if the MODIS MOD10_L2 snow cover product Hall and Riggs, 2016 flags pixels within certain areas as having snow (more details in Section 6.2.2. Operationally, this is far from ideal, but solutions to cope with snow cover (for example, using infrared channel images following Perez et al. (2010)) can be added into the method with no other changes to the routine.

While it is common in frozen-cloud field solar forecasting to derive a “cloud index” from these counts to capture the degree of cloud cover (see *e.g.* Cano et al., 1986; Hammer et al., 1999; Perez et al., 2002), we will use the GVAR counts directly as external covariates for the regime transition probabilities, as outlined in Section 6.2.1.

6.1.2 Irradiance Data

As in Chapter 5, irradiance data was collected at the Rock Springs, PA SURFRAD site (Augustine et al., 2000), in this case from April 1–October 31 during calendar

year 2012 (for a total of 190 days). The site is located at 40.72° N and 77.93° W. DNI is measured using an Eppley Normal Incidence Pyrheliometer (NIP). While GHI is directly measured, due to issues relating to thermal calibration it is advised to compute GHI using DNI (after conversion to BHI) and DHI (which is measured using an Eppley 8-48 pyranometer). All measured values are averaged over 1 minute before reporting. Quality control for the data was handled using the flags included with the data. The 190 days in the dataset was randomly divided into three groups, in accordance with the tuning/training/testing paradigm (Haupt et al., 2009).

6.2 Statistical Model

6.2.1 Model Specification

As in Chapter 5, our model is specified as follows:

$$\mathbf{y}_t = \beta_{0,S_t} + \beta_{1,S_t} \cos(\theta_z) \cdot \mathbf{1}_2 + \sum_{i=1}^{\rho} \beta_{i+1,S_t} \mathbf{y}_{t-i} + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \Sigma_{S_t})$, $\mathbf{1}_2 = (1, 1)^T$, and $\mathbf{y} = (\hat{k}_t, \hat{k}_b)^T$. Unlike the resource assessment model, however, here we provide a time-varying specification for the transition probabilities:

$$\begin{aligned} p(S_t = m | S_{t-1} = l, \mathbf{z}_{t-1}, \boldsymbol{\kappa}) &= \xi_{lm} = \text{logit}^{-1} \left(\kappa_{lm,0} + \sum_i z_{t-1,i} \kappa_{lm,i} \right) \\ &= \frac{\exp(\kappa_{lm,0} + \sum_i z_{t-1,i} \kappa_{lm,i})}{\sum_{k=1}^K \exp(\kappa_{lk,0} + \sum_i z_{t-1,i} \kappa_{lk,i})}, \end{aligned} \quad (6.1)$$

where $\mathbf{z}$ are exogenous satellite-derived covariates (see Section 6.2.2). As the transition probabilities are only specified up to differences between coefficients, identifiability requires the regression parameters associated with transition to a reference state k_0 , $k_0 \in \mathcal{K} = \{1, \dots, K\}$ to be fixed at 0, so that

$$p(S_t = k_0 | S_{t-1} = l, \mathbf{z}_{t-1}, \boldsymbol{\kappa}) = \frac{1}{\sum_{\mathcal{K} - \{k_0\}} \exp(\kappa_{lk,0} + \sum_i z_{t-1,i} \kappa_{lk,i})}, \quad (6.2)$$

where $\mathcal{K} - \{k_0\}$ is set-theoretic subtraction. An example of the nonlinear effect of the covariates $\mathbf{z}_t$ on state persistence is provided by Kaufmann (2015).

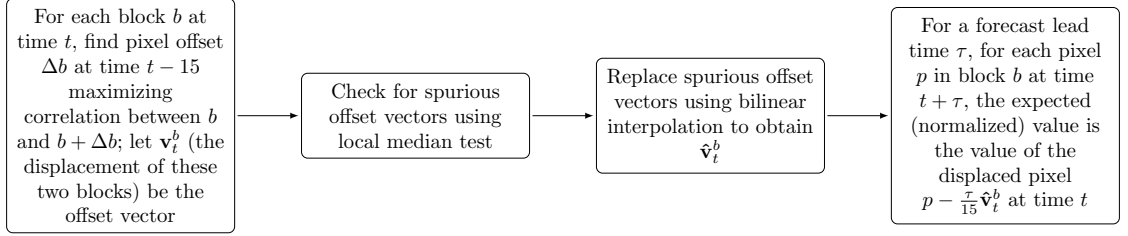


Figure 6.2: Flowchart of the Taylor/frozen-field forecasting procedure. After estimation of the advection vector, the forecast is issued by assuming that the cloud field will remain frozen for the next 15 minutes, so that the forecasted cloud field is the result solely of advective transport. Adapted from Srikrishnan et al. (2017).

6.2.2 Transition Probability Covariates

Two sets of transition covariates are considered. The first set consists of an intercept term along with the mean and standard deviation of the $50 \text{ km} \times 50 \text{ km}$ block centered on the Rock Springs site (referred to as the “current” covariates). 50 km is used as the block size to allow for a sufficient sampling of the cloud field.

The second set adds to the current covariates a set of “Taylor” covariates, generated using a frozen-field forecast (or “Taylor” forecast, due to the assumption of the Taylor hypothesis, that changes to the cloud field resulting from advection dominate changes resulting from local evolution (Taylor, 1938)). A flowchart summarizing the forecasting procedure is provided by Figure 6.2.

To issue a forecast at time t for block b , the image corresponding to time $t - 15$ is first divided into non-overlapping blocks of size $50\text{km} \times 50\text{km}$. The estimated advection vector is the displacement vector between the block at time $t - 15$ that has the highest correlation with b (this process is visualized in Figure 6.3). After advection vectors are estimated within the 2-block neighborhood of b , spurious vectors are identified using a local median test (Westerweel and Scarano, 2005). If the vector corresponding to b is flagged as spurious, blocks within the 1-block neighborhood of b are also examined. Any spurious vectors are replaced using bilinear interpolation.

Due to the cloud masking procedure, any completely clear blocks will have only 0-valued pixels within them. Since the maximum-correlation procedure will fail for these blocks, any block that is sufficiently clear is assigned an advection vector of $\mathbf{0}$ (as if the procedure were tracking the terrain features of the unmasked block), so that the forecast is also for a clear sky. The setting of this threshold is important:

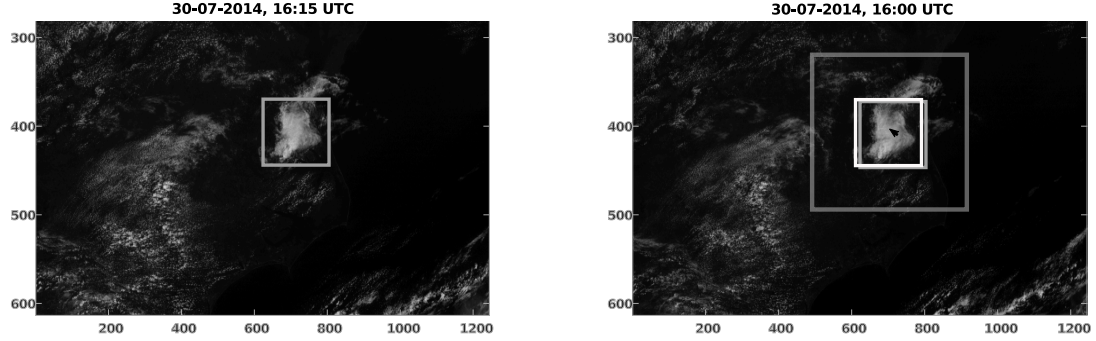


Figure 6.3: Example of advection vector estimation. The left panel is a portion of the forecast-issuance time GOES image. To obtain an estimate of the advection vector used in forecasting, pixel offsets within a search radius are used to maximize the correlation between the GVAR counts in the forecast-issuance time cell (the light gray box in both panels) and the offset cell. The search radius is represented by the dark gray box in the right panel, which is the image 15 minutes prior to the forecast-issuance time. The correlation-maximizing offset cell is bounded in the right panel by the off-white box, with the advection vector shown by the arrow. Reproduced from Srikrishnan et al. (2017).

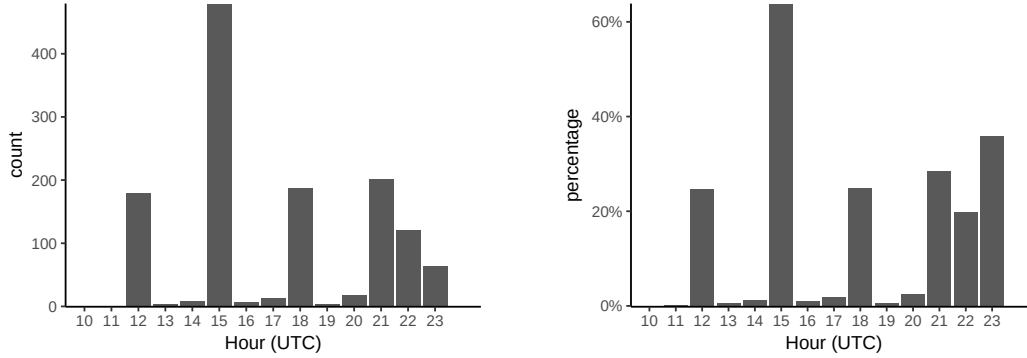
if it is too high, random correlation peaks are likely to be generated for blocks with sparse cloud features, while if it is too low, too many cells will be assigned an advection vector of $\mathbf{0}$. We follow Srikrishnan et al. (2017) in using a threshold value of 96%.

The Taylor covariates are then the mean and standard deviation of the block at time $t + 15$ with the appropriate displacement (the estimated advection vector or its quality control replacement). The satellite-derived covariates are then normalized by dividing by 1024, the maximum possible GVAR count value.

The transition covariates are matched to the 15-minute irradiance averages by using the most recent satellite image available during the averaging period. If no such image is available, the image is treated as missing.

6.2.3 Missing Data

For a variety of reasons (including missing satellite images and the presence of snow), a large amount ($\sim 25\%$) of the satellite data are missing or treated as missing (including some portion of the data from every day in the data set). These missing data are spread throughout the data set with respect to hour of the day, as seen



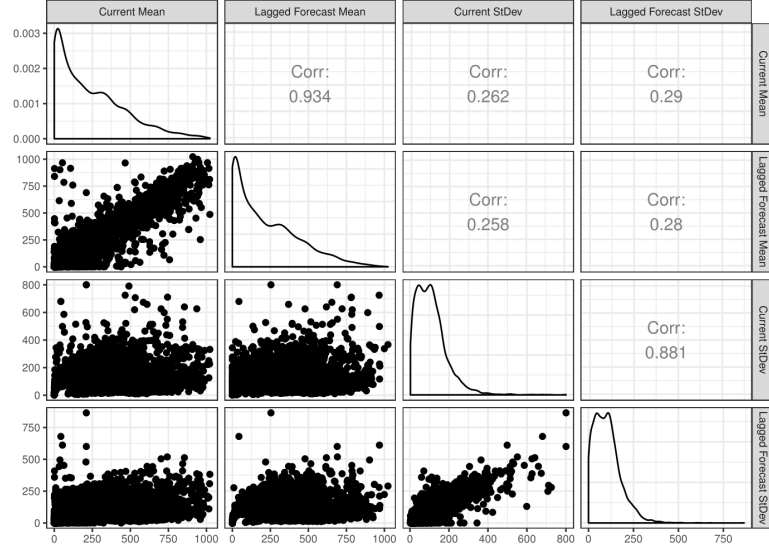
(a) Number of missing satellite images by hour of the day (UTC). (b) Percentage of missing satellite images by hour of the day (UTC).

Figure 6.4: Number and percentage of missing satellite images by hour of the day (UTC).

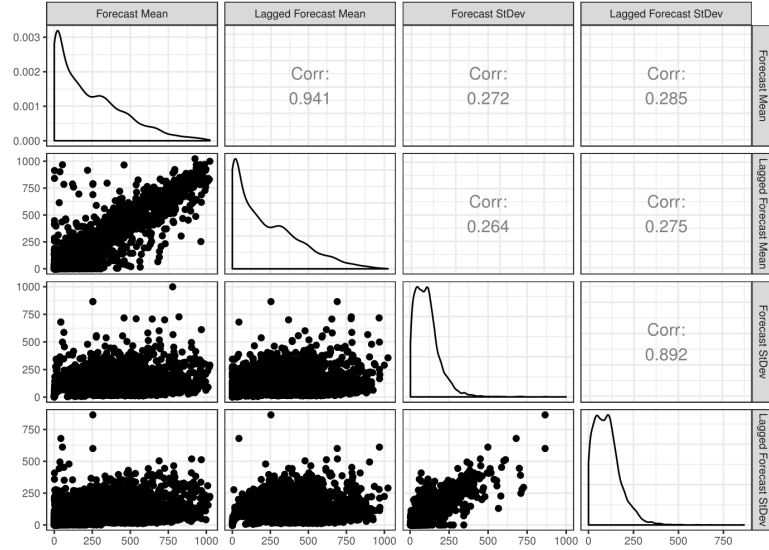
in Figure 6.4a. The proportion of the missing satellite data by hour is shown in Figure 6.4b. We cannot treat this missing data as ignorable as there are no complete cases, and even if there were, restricting statistical analyses to complete-case analysis can result in biased estimates when the data are not missing completely at random (in other words, if the probability of missingness is independent of the missing or observed data) (Little and Rubin, 2002).

Missing data imputation can be easily incorporated into the Bayesian framework via data augmentation (Tanner and Wong, 1987). Missing covariate values can be sampled at each iteration of the MCMC procedure. The remainder of the sampling steps are then conditioned on those imputed values. One approach is to place a prior distribution over the missing data. However, the spatiotemporal dynamics of the cloud field induce a time series structure on the derived covariates, suggesting that model-based imputation may be more appropriate.

When the satellite image cell is sufficiently clear, so that there are not enough cloud features for a maximum-correlation method to be effective, the covariates are set to zero. Otherwise, there is a strong correlation between the moments of the relevant time- t satellite image cells (the “current” cell, overhead of the measurement site, and the “forecast” cell obtained by the maximum-correlation advection method) and the moments of the lagged (time- $t - 1$) forecast cell, as seen



(a) Pairs plot for the time- t “current” mean and standard deviation and the lagged (time $(t - 1)$) Taylor forecast cell mean and standard deviation.



(b) Pairs plot for the time- t Taylor forecast cell mean and standard deviation and the lagged (time- $(t - 1)$) forecast cell mean and standard deviation.

Figure 6.5: Pairs plots relating the time- t covariate values with the lagged (time- $(t - 1)$) covariates derived from the advection-only satellite forecasting method. Only time- t values corresponding to images that were not too clear for the cloud-filtering procedure are included.

in Figure 6.5. Thus, we can think of the imputation model as a two-stage process

$$\mathbf{z}_t = I_t^{\mathbf{z}} \mathbf{z}_t^{\text{pos}}, \quad (6.3)$$

where $I_t^{\mathbf{z}} = 1$ if $\mathbf{z}_t \neq 0$ and 0 otherwise, and $\mathbf{z}_t^{\text{pos}} = \mathbf{z}_t = (\mathbf{z}_t^{\text{cur}}, \mathbf{z}_t^{\text{for}})$ if $\mathbf{z}_t \neq 0$.

The model (6.3) generates $\mathbf{z}_t$ as a mixture of two components: a point distribution at $\mathbf{0}$ and a truncated multivariate normal whose mean depends on the lagged values $\mathbf{z}_{t-1}^{\text{for}}$. The probability of the indicator variable $I_t^{\mathbf{z}}$ is the mixture weight associated with non-zero component. The first stage of the model is a logistic regression for $I_t^{\mathbf{z}}$,

$$\Pr(I_t^{\mathbf{z}} = 1) = \text{logit}^{-1}(\mathbf{z}_{t-1}^{\text{cur}} \boldsymbol{\alpha}^I) \quad (6.4)$$

based on the covariates corresponding to “current” conditions.

The second stage involves calibrating the non-zero process for $\mathbf{z}_t^{\text{pos}}$, with each vector $\mathbf{z}_t^{\text{cur}}$ and $\mathbf{z}_t^{\text{for}}$ given by a constrained vector AR(1) process. After normalizing $\mathbf{z}$ onto the unit interval (which we denote by $\hat{\mathbf{z}}$),

$$\hat{\mathbf{z}}_t = \hat{\mathbf{z}}_{t-1} \boldsymbol{\alpha}^{\mathbf{z}} + \epsilon^{\mathbf{z}}. \quad (6.5)$$

Note that, as written, the model (6.5) would allow for values of $\hat{\mathbf{z}}_t$ that are outside the unit interval (so that $\mathbf{z}_t$ can be less than zero or greater than 1024), but this is not problematic given that $\hat{\mathbf{z}}_t$ are used as logistic covariates for the state transition probabilities: values close to, but less than, 0 will result in similar probabilities to zero values, and similarly near 1. Therefore, we censor the AR(1) output to keep all values on the unit interval.

6.2.4 Missing Data Sampling Scheme

The missing-data imputation process adds extra steps to the MCMC sampler described in Algorithm 4.5.1.

1. Sample the missing-covariate parameters $\boldsymbol{\alpha}^I, \boldsymbol{\alpha}^{\mathbf{z}}, \Sigma^{\mathbf{z}}$.

The first stage of the missing-data imputation process described in Section 6.2.3 is the binomial logistic regression model (6.4) for the process component $I^{\mathbf{z}}$. The Gibbs sampling method used to sample the logistic regression coefficients for the inhomogeneous transition probabilities (Section 4.6.2)

Table 6.1: Prior distributions for the forecast model parameters. B_k denotes the vectorized regression parameters β_k , as described in Section 4.7.

Parameter	Prior Distribution	Comments
α^I	$\mathcal{N}(\mathbf{0}, I)$	
α^z	$\mathcal{N}(\mathbf{0}, I)$	
Σ^z	$\mathcal{W}^{-1}(3, I)$	
κ_k	$\mathcal{N}(\mathbf{0}, I)$	
B_k	$\mathcal{N}(\mu_k, I)$	$\mu_k = \left(-2 + \frac{4(k-1)}{(K-1)}, \mathbf{0}_{\rho+1}, -2 + \frac{4(k-1)}{(K-1)}, \mathbf{0}_{\rho+1}\right)$
Σ_k	$\mathcal{W}^{-1}(3, I)$	

is also used for the missing-data process mixture coefficients α^I .

Then the AR coefficients α^z and innovation variance Σ^z are sampled (conditioned on the previously imputed values) as a standard linear regression model.

2. Impute missing transition covariates $\mathbf{z}_t$.

Once the missing-data process indicator I_t^z is determined by sampling ι_t , missing covariates $\mathbf{z}_t$ are imputed as $\mathbf{z}_t = \mathbf{0}$ if $I_t^z = 0$ and according to Equation (6.5) otherwise.

6.2.5 Prior Distributions

The priors for this forecasting model are listed in Table 6.1. As in Section 5.2, our selection of priors is based on our desire to use Gibbs sampling steps. For the shared parameters between the two models (B_k and Σ_k), the justifications for the priors are identical. For the regression (logistic or linear) parameters unique to this model, the primary reason for the specification of these particular priors is for coefficient regularization. The imputation innovation variance Σ^z prior was chosen to keep it only weakly informative.

6.3 Forecast Validation Metrics

Validating a probabilistic forecast is more complex than validating a point forecast, for which the only concern is the distance between the forecast and the realization. For a probabilistic forecast, the validation question can be thought of as asking two

questions of the predictive distribution. First, is the observation consistent with the predictive distribution? Does the observation seem like a random draw from the forecast? If so, the forecast is *calibrated*. Second, how uncertain is the forecast? The question of how widely the probability mass is distributed within the forecast density refers to the *sharpness* of the forecast.

While sharpness is a property of the predictive distribution only, calibration is a joint property of the distribution and the realized observation. Calibration has a direct analogue in the validation of a point forecast, while sharpness is a property unique to density forecasting (including prediction intervals surrounding point forecasts, which can be thought of as uniform predictive distributions centered on the point forecast). Gneiting et al. (2007) posits that an ideal probabilistic forecast is maximally sharp subject to calibration. In other words, the ideal predictive density assigns the correct level of uncertainty to the observation, and is as sharp as can be given the “true” uncertainty about the outcome.

The *marginal calibration* (Gneiting et al., 2007) of a predictive distribution refers to the compatibility of the empirical distributions of the forecasts and observations. The marginal calibration of a forecast density can be assessed visually by, for example, plotting the realizations alongside the forecast means. This allows for a quick identification of forecast biases or over- or under-dispersion.

One metric for the sharpness of a multivariate predictive distribution is the determinant sharpness (DS) Gneiting et al. (2008):

$$\text{DS} = (\det \Sigma)^{1/(2d)}, \quad (6.6)$$

where $\Sigma \in \mathbb{R}^{d \times d}$ is the covariance matrix of the forecast density or ensemble. The DS (6.6) generalizes the standard deviation of a univariate forecast.

In the univariate setting, an omnibus metric (or *scoring rule*) for assessing both the calibration and sharpness of a predictive distribution is the Continuous Ranked Probability Score (CRPS) (Gneiting and Raftery, 2007). If the forecast density is denoted by p and the realized observation by $\mathbf{x}$, the CRPS is defined as

$$\begin{aligned} \text{CRPS}(p, x) &= \int_{-\infty}^{\infty} (F(y) - \mathbb{I}\{y \geq x\})^2 dy \\ &= \mathbb{E}_p |X - x| - \frac{1}{2} \mathbb{E}_p |X - X'|, \end{aligned} \quad (6.7)$$

where X and X' are independent random variables distributed according to p . When the forecast is a point forecast at x_0 , the CRPS reduces to the mean absolute error (MAE) $|x_0 - x|$.

While (6.7) is in general difficult to compute when the predictive density p is not a “nice” density (such as a Gaussian density), when the density is represented by a size- k ensemble of forecast elements x_i , the integral has the Monte Carlo approximation

$$\hat{\text{CRPS}}(p, x) = \frac{1}{k} \sum_{i=1}^k |x_i - x| - \frac{1}{2k^2} \sum_{i=1}^k \sum_{j=1}^k |x_i - x_j|, \quad (6.8)$$

which is derived using the empirical CDF of $\{x_i\}$ (Krüger et al., 2016). Lower values of the CRPS (when expressed using (6.7)) are better.

As a multivariate analogue of the CRPS, Gneiting et al. (2008) proposed the so-called Energy Score (ES):

$$\text{ES}(p, \mathbf{x}) = \mathbb{E}_p \|\mathbf{X} - \mathbf{x}\| - \frac{1}{2} E_p \|\mathbf{X} - \mathbf{X}'\|, \quad (6.9)$$

where $\|\cdot\|$ is the Euclidean (L_2) norm. For a point forecast $\mathbf{x}_0$, the ES is equivalent to the Euclidean error (EE) $\|\mathbf{x}_0 - \mathbf{x}\|$.

As with the CRPS, the ES (6.9) admits a Monte Carlo approximation

$$\hat{\text{ES}}(p, \mathbf{x}) = \frac{1}{k} \sum_{i=1}^k \|\mathbf{x}_i - \mathbf{x}\| - \frac{1}{2k^2} \sum_{i=1}^k \sum_{j=1}^k \|\mathbf{x}_i - \mathbf{x}_j\|. \quad (6.10)$$

Given a set of observations, the score (CRPS or ES) for a given model is the mean score over the observations.

Both of the multivariate metrics DS and ES are best used when there is a common scale for all of the elements of the forecasted vector, so that each dimension is weighted equally. The units of the CRPS are the same as those of the forecasted quantity x . For the ES, this may be the case if the vector components have the same units; otherwise, the ES ought to be dimensionless after component-wise normalization.

Advantages of the CRPS and the ES are that they are proper scoring rules (Gneiting and Raftery, 2007), which means that when scoring predictive distributions p using these metrics, in expectation a forecaster has no incentive to issue any

other forecast if p represents the forecaster’s best (in the sense of calibration and sharpness) forecast. Moreover, they facilitate the comparison of probabilistic and deterministic forecasts due to their equivalence to the MAE in the point-forecast setting.

The last forecast measure that we will consider is the skill score, which measures the forecast’s improvement relative to some reference forecast (typically persistence or climatological, depending on the application). In this case, we compute the skill with respect to the CRPS (in the univariate setting) and the ES (in the multivariate setting), so that the skill score is

$$\text{Skill} = 1 - \frac{\text{Error}_{\text{forecast}}}{\text{Error}_{\text{reference}}}.$$

The reference forecast for the skill score is *smart persistence* as the reference forecast in this study. Smart persistence assumes that the non-dimensional index values k_t and k_b persist from forecast-issuance time t to $t + 1$, but allow G_0 and θ_z to progress according to their diurnal cycles, unlike regular persistence forecasts. As this is a point forecast, we use the MAE instead of the CRPS and the EE instead of the ES in the skill computations.

Forecast performance will be evaluated with respect to the re-dimensionalized GHI and DNI values, rather than the modeled k_t and k_b indices. This is both to facilitate comparisons with other papers in the solar forecasting literature and because these are more directly the quantities of interest.

6.4 Results

6.4.1 Model Selection

In addition to the inhomogeneous models using the “current” and “Taylor” transition covariates, for the purposes of comparison we also fit a homogeneous model with constant transition probabilities. For each model under consideration, the number of states K and the autoregressive order ρ are selected using WAIC (Section 4.9.1). Using the tuning dataset, the MCMC sampler was run for 100,000 iterations, with the first 45,000 iterations discarded as burn-in (convergence was diagnosed using the Gelman-Rubin diagnostic (Gelman and Rubin, 1992)). The WAIC was computed

Table 6.2: Watanabe-Akaike information criterion (WAIC) values for the 15-minute ahead forecast models, as in Equation (4.16). Lower values correspond to a better estimated predictive fit. The WAIC values for the models which are selected are in bold.

Model	ρ	K				
		2	3	4	5	6
Homogeneous	1	1964	1433	544	-303	-283
	2	1394	864	410	-77	-385
Inhomogeneous: Current	1	2231	1325	989	6576	9539
	2	1664	871	4082	10678	8791
Inhomogeneous: Taylor	1	2270	158	1039	6126	9270
	2	1662	167	2566	9767	7325

using an ensemble of 10,000 posterior samples. The resulting values are shown in Table 6.2.

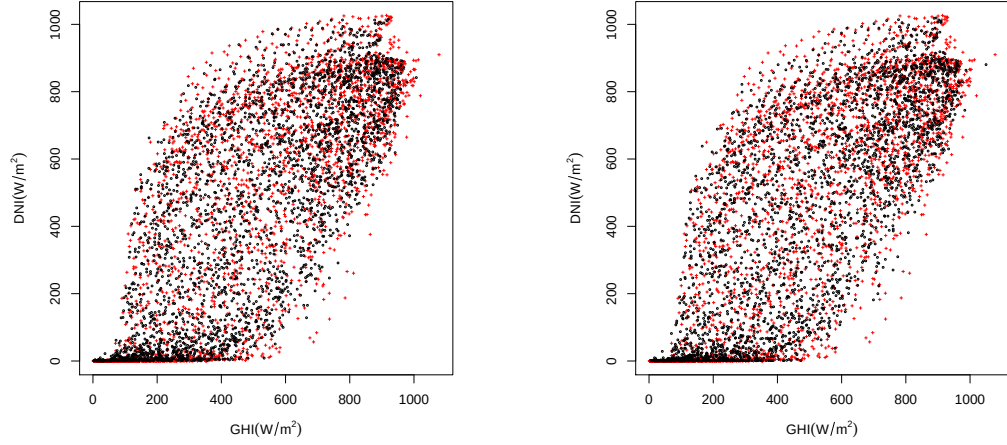
6.4.2 Identifiability Constraints

Using the training dataset, the MCMC sampler was run for 150,000 iterations, with the first 50,000 discarded as burn-in, using the Gelman-Rubin diagnostic to diagnose convergence. Each of the three models had very different constraints imposed on the MCMC output for identifiability.

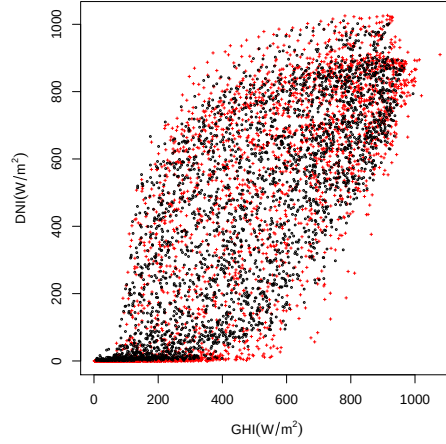
The homogeneous model was identified using a combination of constraints on the k_b covariance $\Sigma_{2,2}$ and the k_b intercept $\beta_{2,0}$, as can be seen in the posterior distributions in Appendix C. The inhomogeneous model using the current covariates was identified using a combination of constraints on the joint covariance $\Sigma_{2,1}$ and the k_b AR(1) coefficient $\beta_{2,2}$, while the inhomogeneous model using the Taylor covariates was identified by ordering the k_t intercept $\beta_{1,0}$.

6.4.3 Forecast Performance

Forecast ensembles were then generated for each point in the validation dataset, with each ensemble containing 10,000 forecast elements. Figure 6.6 shows the marginal calibration plots for all three models. Visually, there appear to be no serious issues with bias or over- or under-dispersion for any of the three models.



(a) Inhomogeneous model using current co- (b) Inhomogeneous model using Taylor co-
variates. variates.



(c) Homogeneous model.

Figure 6.6: Marginal calibration plots for the three probabilistic models. The red crosses are observations, and the black dots are forecast ensemble means.

The forecast validation metrics shown in Tables 6.3 and 6.4 include both univariate and multivariate measures. The univariate measures are the CRPS and the univariate skill score for each component (GHI and DNI). The multivariate measures are the ES, the DS, and the multivariate skill score. Both inhomogeneous models perform similarly to the homogeneous model, with skill of approximately 25% in both univariate component forecasts and the multivariate forecast (for comparison, the 15-minute ahead time-varying ARIMA GHI forecast in Reikard

Table 6.3: Univariate validation metrics for the homogeneous and inhomogeneous Markov-switching forecast models.

Model	Univariate Metrics			
	CRPS (GHI)	CRPS (DNI)	Skill (GHI)	Skill (DNI)
Homogeneous	43.5	54.0	0.26	0.27
Inhomogeneous (Current)	44.1	53.5	0.25	0.27
Inhomogeneous (Taylor)	44.5	54.2	0.25	0.26

Table 6.4: Multivariate validation metrics for the homogeneous and inhomogeneous Markov-switching forecast models.

Model	Multivariate Metrics		
	ES	DS	Skill
Homogeneous	74.0	73.9	0.26
Inhomogeneous (Current)	74.2	59.6	0.26
Inhomogeneous (Taylor)	74.9	58.2	0.25

et al. (2017) has a skill of approximately 49%). However, the inhomogeneous models are significantly sharper than the homogeneous model, likely due to the increased number of states used for the homogeneous model and the limited ability of that model to shift the probability weight associated with each mode.

All of the models have a slight negative correlation between the solar zenith angle and the ES (≈ 0.23), suggesting increased difficulty in forecasting near solar noon (this is primarily the result of the distribution of outliers with ES values greater than 400 W/m^2 , and is therefore likely due to the increased uncertainty in the state transition probabilities due to variability in how the cloud field evolves around mid-day). The inhomogeneous models also have a similar negative correlation (≈ 0.26) between the solar zenith angle and the DS, while the homogeneous model’s DS has a correlation of -0.49 between the solar zenith angle and the DS. Perhaps surprisingly, neither of the inhomogeneous models have a significant increase in ES or DS related to the missingness of satellite images.

While the median residual values of all of the models are located near 0 W/m^2 , the homogeneous model has a wider distribution of marginal DNI residuals than either of the inhomogeneous models, as seen in Figure 6.7. Figure 6.8 suggests that there is no particular relationship between the observed irradiance values and the component-wise CRPS, though extremely low GHI values ($< 100 \text{ W/m}^2$) are forecasted more accurately than higher values. The models appear to have similar

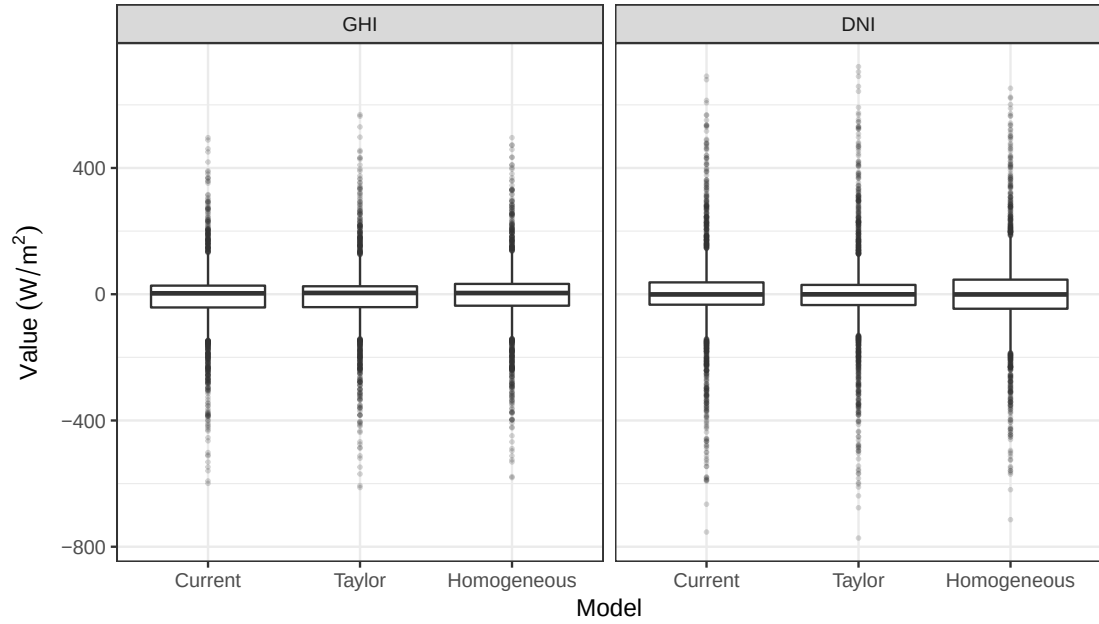
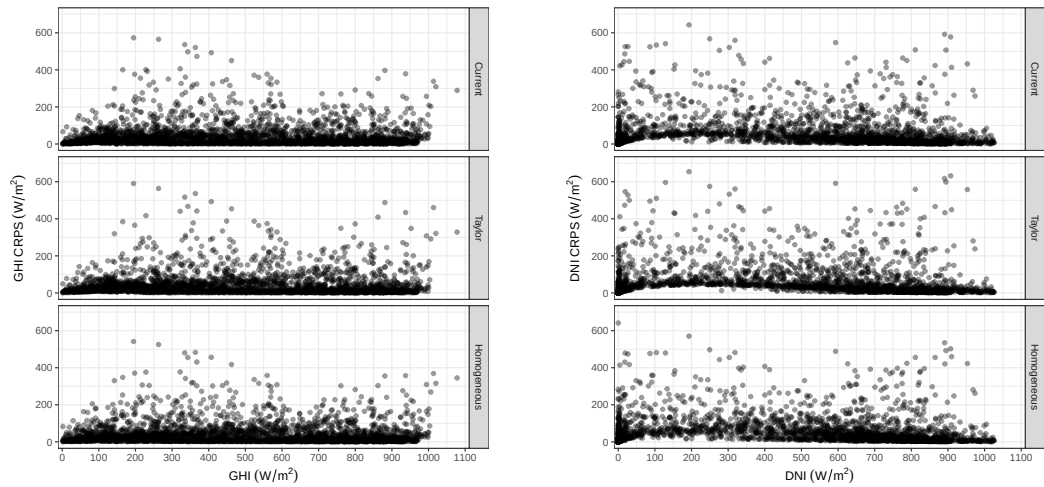


Figure 6.7: Boxplot of forecast residuals. The hinges are located at the 25% and 75% quantiles, and the whiskers extend out 1.5 times the interquartile range.



(a) CRPS for GHI marginal forecast vs. observed GHI. (b) CRPS for DNI marginal forecast vs. observed DNI.

Figure 6.8: Marginal CRPS plotted against observed values.

trends with respect to the relationship between the marginal CRPS values and the observations, suggesting that there is no bias or decrease in uncertainty that is unique to any particular specification other than the increased sharpness for the homogeneous model.

6.5 Discussion

In this chapter, we have compared the forecasting performance of inhomogeneous Markov-switching models which use satellite-derived transition covariates with that of a homogeneous Markov-switching model. No model stood out in terms of skill relative to a smart persistence forecast, and all three had only half the GHI-forecasting skill of an ARIMA model with time-varying coefficients (Reikard et al., 2017). However, the homogeneous model forecasts were significantly less sharp than those of the inhomogeneous models. As there was relatively little difference between the two inhomogeneous models, in the future, we are likely to use the “current” covariates, as these require smaller satellite images, less computational time for image processing, and a smaller parameter set, as well as not having an additional source of error from the Taylor cloud field forecasting method.

Many of the model structural changes discussed in Section 5.4 are likely to also be valuable in this forecast setting. In particular, modeling the dynamics as being stationary around a switching mean is likely to have a greater impact over shorter temporal horizons, as seen in Figure 3.1. Future versions of this method may want to include forecasts of the average daily clearness index in the transition probabilities, whether these are obtained from WRF models or included in the hierarchical model as an autoregressive process from the previous day.

Chapter 7 |

Conclusions and Future Work

In this dissertation, we have begun to develop probabilistic models for jointly forecasting the direct and global components of solar irradiance. We have opted to use time-series models, rather than artificial intelligence methods such as ANNs. Part of the reason for this choice is that time-series models allow us to quantify the uncertainty associated with a forecast, and the Bayesian methods that we have used facilitate working with missing data. Once we have a model that performs satisfactorily, we can also quantify how SECS will respond to particular deviations from climatology. We also believe there is a critical advantage to being able to assimilate meteorological prior information into solar resource models, though we have done relatively little of that in this work.

However, machine-learning techniques remain a valuable modeling tool. One possibility for reconciling the two approaches is to do initial data exploration using ANNs, allowing for patterns and initial parameter estimates to be mined from the exploratory data set, and then fit a time series model using those insights. For example, the work of McCandless et al. ([2016a](#)) can provide us with certain insights into the number of statistically important meteorological states, as well as how they might be characterized.

Probabilistic forecasting will only increase in value and importance as computational power and the ability of stochastic optimization methods to cope with large ensembles of samples increases. Nevertheless, more sophisticated point forecasting methods will and should continue to be developed. To facilitate comparisons between probabilistic and deterministic techniques, we hope that point forecasters will use validation metrics that generalize to the probabilistic setting, such as the MAE. We also hope that more emphasis is placed on modeling both the direct and

global components of irradiance, rather than just the global component, so that solar resource assessments and forecasts are ready to be used for a variety of SECS models.

While the model parameterizations used in this dissertation were selected primarily for simplicity, it is clear that more sophisticated models are necessary to achieve adequate levels of performance. Some of the changes that we have proposed draw on some of the original insights in the solar literature, such as the dependence of hourly k_t series on the daily average index value, and we plan to pursue these modeling directions in the near future. An advantage of the Bayesian framework is that it is relatively straightforward to add hierarchical layers and to quantify uncertainty as it propagates “downstream,” so to speak.

Once switching models perform satisfactorily at a single site, we also hope to extend these models to a spatial setting, allowing for the simulation of a portfolio of SECS. This could be achieved, for example, by generating the latent states as the output of a Gaussian process, where the covariance matrix might reflect the typical spatial scales associated with local meteorology.

Appendix A |

Bridge Sampling

Fruhwirth-Schnatter (2004) investigated the use of bridge sampling to estimate the marginal likelihood, and this appendix replicates that analysis for convenience. Bridge sampling is a generalization of importance sampling which combines i.i.d. draws from an importance density with draws from the posterior density (Meng and Wing, 1996). One advantage of bridge sampling compared to importance sampling is that the variance of the estimator depends on a ratio which is bounded regardless of the choice of importance density (in particular, regardless of the tail behavior of the importance density). Denoting by $p(\boldsymbol{\psi}|\mathbf{y})$ the normalized posterior density, $p^*(\boldsymbol{\psi}|\mathbf{y})$ the non-normalized posterior density, and $q(\boldsymbol{\psi})$ an importance density, let $\alpha(\boldsymbol{\psi})$ be an arbitrary function such that

$$C^\alpha = \int \alpha(\boldsymbol{\psi}) p(\boldsymbol{\psi}|\mathbf{y}) q(\boldsymbol{\psi}) \nu(d\boldsymbol{\psi}) > 0. \quad (\text{A.1})$$

Bridge sampling is based on the identity

$$1 = \frac{E_q(\alpha(\boldsymbol{\psi}) p(\boldsymbol{\psi}|\mathbf{y}))}{E_p(\alpha(\boldsymbol{\psi}) q(\boldsymbol{\psi}))}, \quad (\text{A.2})$$

which follows trivially from Equation (A.1) by switching the order of $p(\cdot)$ and $q(\cdot)$ in the integrand. In general, the normalized constant for the density $p(\cdot)$ is not known, but substituting the equation linking the normalized and non-normalized posterior densities,

$$p(\mathbf{y}|\boldsymbol{\psi}) = \frac{p^*(\mathbf{y}|\boldsymbol{\psi})}{p(\mathbf{y})},$$

into Equation (A.2) yields

$$p(\mathbf{y}) = \frac{E_q(\alpha(\boldsymbol{\psi})p^*(\boldsymbol{\psi}|\mathbf{y}))}{E_p(\alpha(\boldsymbol{\psi})q(\boldsymbol{\psi}))}. \quad (\text{A.3})$$

Both of the expectations on the right-hand side of Equation (A.3) can be estimated using sample averages, either using i.i.d draws $\tilde{\boldsymbol{\psi}}^{(m)}$, $m = 1, \dots, M$, from the importance density $q(\cdot)$ or MCMC draws $\boldsymbol{\psi}^{(n)}$, $n = 1, \dots, N$, from $p^*(\cdot)$, resulting in the estimator

$$\hat{p}(y) = \frac{\hat{E}_q(\alpha(\boldsymbol{\psi})p^*(\boldsymbol{\psi}|\mathbf{y}))}{\hat{E}_p(\alpha(\boldsymbol{\psi})q(\boldsymbol{\psi}))} = \frac{M^{-1} \sum_{m=1}^M \alpha(\tilde{\boldsymbol{\psi}}^{(m)})}{N^{-1} \sum_{n=1}^N \alpha(\boldsymbol{\psi}^{(n)})}. \quad (\text{A.4})$$

The estimator $\hat{p}(\mathbf{y})$ is the general form of the bridge sampling estimator, though more specific samplers result from specific choices of $\alpha(\cdot)$ (Meng and Wing, 1996). For example, choosing $\alpha(\boldsymbol{\psi}) = 1/q(\boldsymbol{\psi})$ yields an importance sampling estimator (Fruhwirth-Schnatter, 1995)

$$\hat{p}_{\text{IS}}(y) = \frac{1}{M} E_q \left(\frac{p^*(\boldsymbol{\psi}|\mathbf{y})}{q(\boldsymbol{\psi})} \right), \quad (\text{A.5})$$

or choosing $\alpha(\boldsymbol{\psi}) = 1/p^*(\boldsymbol{\psi}|\mathbf{y})$ yields the reciprocal importance estimator (Gelfand and Dey, 1994)

$$\hat{p}_{\text{RI}}(y) = \left\{ \frac{1}{N} E_p \left(\frac{q(\boldsymbol{\psi}|\mathbf{y})}{p^*(\boldsymbol{\psi})} \right) \right\}^{-1}. \quad (\text{A.6})$$

In particular, Meng and Wing (1996) discuss an asymptotically optimal choice of $\alpha(\cdot)$, which is optimal in the sense of minimizing the expected relative error of $\hat{p}(y)$ for i.i.d. draws from both $p(\boldsymbol{\psi}|\mathbf{y})$ and $q(\boldsymbol{\psi})$:

$$\alpha(\boldsymbol{\psi}) \propto \frac{1}{Mq(\boldsymbol{\psi}) + Np(\boldsymbol{\psi}|\mathbf{y})}.$$

The resulting estimator is denoted by $\hat{p}_{\text{BS}}(y)$. This $\alpha(\cdot)$ is based on the normalized posterior $p(\cdot|\mathbf{y})$, so Meng and Wing (1996) used an iterative procedure to obtain $\hat{p}_{\text{BS}}$. An estimate of the normalized posterior is obtained based on a previous

estimate of the normalizing constant $\hat{p}_{\text{BS},t-1}(y)$,

$$\hat{p}_{t-1}(\boldsymbol{\psi}|\mathbf{y}) = \frac{p^*(\boldsymbol{\psi}|\mathbf{y})}{\hat{p}_{\text{BS},t-1}(y)}.$$

The recursion is then given by

$$\hat{p}_{\text{BS},t}(y) = \hat{p}_{\text{BS},t-1}(y) \frac{M^{-1} \sum_{m=1}^M \frac{\hat{p}_{t-1}(\tilde{\boldsymbol{\psi}}^{(m)}|\mathbf{y})}{Mq(\tilde{\boldsymbol{\psi}}^{(m)}) + N\hat{p}_{t-1}(\tilde{\boldsymbol{\psi}}^{(m)}|\mathbf{y})}}{N^{-1} \sum_{n=1}^N \frac{q(\boldsymbol{\psi}^{(n)})}{Mq(\boldsymbol{\psi}^{(n)}) + N\hat{p}_{t-1}(\boldsymbol{\psi}^{(n)}|\mathbf{y})}}, \quad (\text{A.7})$$

Other estimators, such as that from Equation (A.5) or (A.6), may be used as the starting value $\hat{p}_{\text{BS},0}(y)$. As the densities in Equation (A.7) may be computed ahead of time, this iteration converges relatively quickly.

A.1 Unsupervised Construction of an Importance Density

As it is impractical to manually tune importance densities for each model under consideration, Fruhwirth-Schnatter (2004) recommends the following importance density, which is constructed in an unsupervised fashion from the MCMC output of an unconstrained sampler. With respect to some measure ν on the parameter space $\mathcal{A}$,

$$p(\boldsymbol{\psi}|\mathbf{y}) = \int_{\mathcal{A}} K(\boldsymbol{\psi}|\boldsymbol{\psi}')p(\boldsymbol{\psi}'|\mathbf{y})\nu(d\boldsymbol{\psi}'),$$

where $K(\boldsymbol{\psi}|\boldsymbol{\psi}')$ is the MCMC transition kernel. Splitting $\boldsymbol{\psi} = (\boldsymbol{\phi}, \boldsymbol{\xi})$, where $\boldsymbol{\phi}$ are the model parameters and $\boldsymbol{\xi}$ are the transition parameters, when the kernel $K(\boldsymbol{\psi}|\boldsymbol{\psi}')$ factorizes as

$$K(\boldsymbol{\psi}|\boldsymbol{\psi}') = K_{\boldsymbol{\phi}}(\boldsymbol{\phi}|S^T, \mathbf{y}, \boldsymbol{\phi}')p(\boldsymbol{\xi}|\mathbf{y})p(S^T|(\boldsymbol{\phi}, \boldsymbol{\xi})', \mathbf{y})$$

and the densities $K_{\boldsymbol{\phi}}(\boldsymbol{\phi}|S^T, \mathbf{y}, \boldsymbol{\phi}')$ can be sampled directly from densities

$$K_{\boldsymbol{\phi}}(\boldsymbol{\phi}|S^T, \mathbf{y}, \boldsymbol{\phi}') = \prod_{d=1}^D p(\boldsymbol{\phi}_d|\boldsymbol{\phi}_{j<d}, \boldsymbol{\phi}'_{j>d}, S^T, \mathbf{y}), \quad (\text{A.8})$$

the importance density $q(\phi)$ can be constructed from MCMC draws as a mixture density by Rao-Blackwellization (Gelfand and Smith, 1990; Robert and Casella, 2004). Specifically, take MCMC samples $\boldsymbol{\psi}^{(1)}, \dots, \boldsymbol{\psi}^{(L)}$. Then

$$q(\boldsymbol{\theta}, \boldsymbol{\xi}) = \frac{1}{S} \sum_{l=2}^L p\left(\boldsymbol{\xi} \mid (S^T)^{(l)}\right) \prod_{d=1}^D p\left(\phi_d \mid \phi_{j < d}^{(l)}, \phi_{j > d}^{(l-1)}, (S^T)^{(l)}, \mathbf{y}\right). \quad (\text{A.9})$$

A.2 Relative Standard Errors

As the bridge sampling estimate $\hat{p}_{\text{BS}}(y)$ is a random quantity, quantifying its variability for different posterior and importance samples and starting values is of interest to help characterize the degree to which one model has a lower expected loss than another. Fruhwirth-Schnatter (2004) approximates the estimator's relative expected mean-squared error

$$\text{RE}^2(\hat{p}(y)) = \frac{\mathbb{E}(\hat{p}(y) - p(y))^2}{p(y)^2}, \quad (\text{A.10})$$

which is approximation to the expected absolute mean-squared error of the log-marginal likelihood estimate $\log \hat{p}(y)$. Using the simulated importance and posterior draws $\tilde{\boldsymbol{\phi}}^{(m)}$ and $\boldsymbol{\phi}^{(n)}$, respectively, the approximation of Equation (A.10) for the bridge-sampling estimate $\hat{p}_{\text{BS}}(y)$ is given by

$$\text{RE}^2(\hat{p}_{\text{BS}}(y)) = \frac{1}{M} \frac{\text{Var}_q\left(\frac{p(\boldsymbol{\phi}|\mathbf{y})}{w_q q(\boldsymbol{\phi}) + w_p p(\boldsymbol{\phi}|\mathbf{y})}\right)}{\mathbb{E}_q^2\left(\frac{p(\boldsymbol{\phi}|\mathbf{y})}{w_q q(\boldsymbol{\phi}) + w_p p(\boldsymbol{\phi}|\mathbf{y})}\right)} + \frac{\rho_f(0)}{N} \frac{\text{Var}_p\left(\frac{q(\boldsymbol{\phi})}{w_q q(\boldsymbol{\phi}) + w_p p(\boldsymbol{\phi}|\mathbf{y})}\right)}{\mathbb{E}_p^2\left(\frac{q(\boldsymbol{\phi})}{w_q q(\boldsymbol{\phi}) + w_p p(\boldsymbol{\phi}|\mathbf{y})}\right)}, \quad (\text{A.11})$$

where $w_q = M/(M + N)$, $w_p = N/(M + N)$, and $\rho_f(0)$ is the normalized spectral density at frequency 0 of the process $f(\boldsymbol{\phi}^{(n)}) = q(\boldsymbol{\phi}^{(n)})/(w_q q(\boldsymbol{\phi}^{(n)}) + w_p p(\boldsymbol{\phi}^{(n)}|\mathbf{y}))$, which accounts for the possible autocorrelation of the posterior draws $\boldsymbol{\phi}^{(n)}$ (the importance draws $\tilde{\boldsymbol{\phi}}^{(m)}$ are independent).

Appendix B |

Resource Assessment Posterior Distribution Plots

B.1 All-Year Model

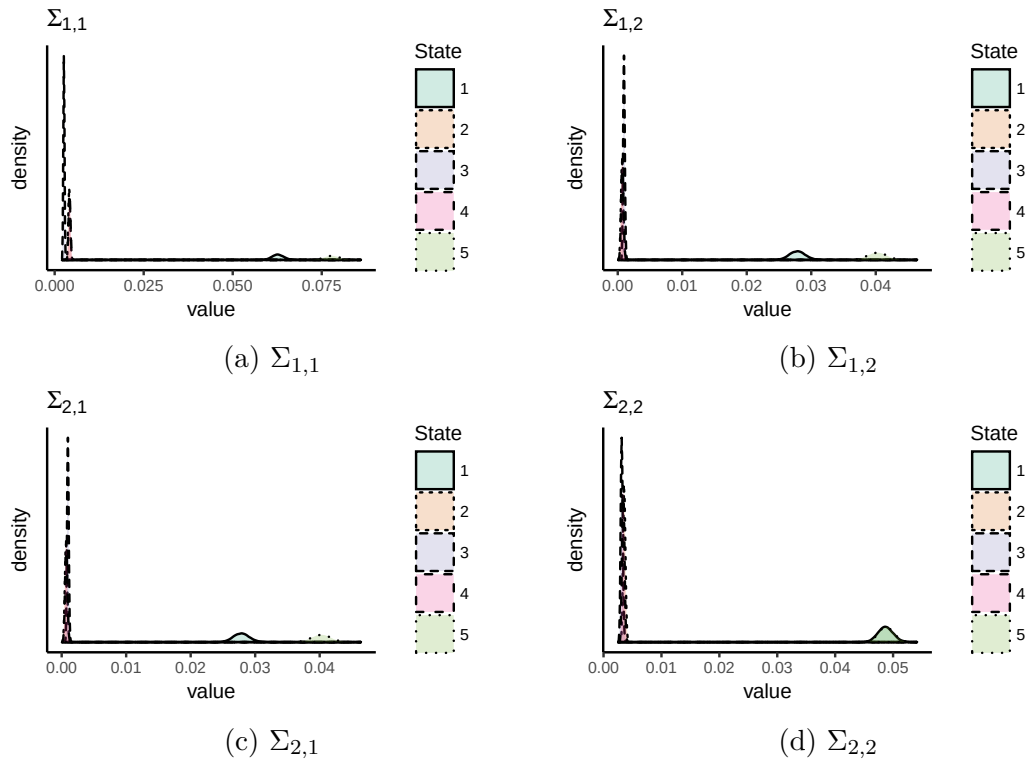


Figure B.1: Posterior distributions of the innovation covariance Σ components for the all-year resource assessment model.

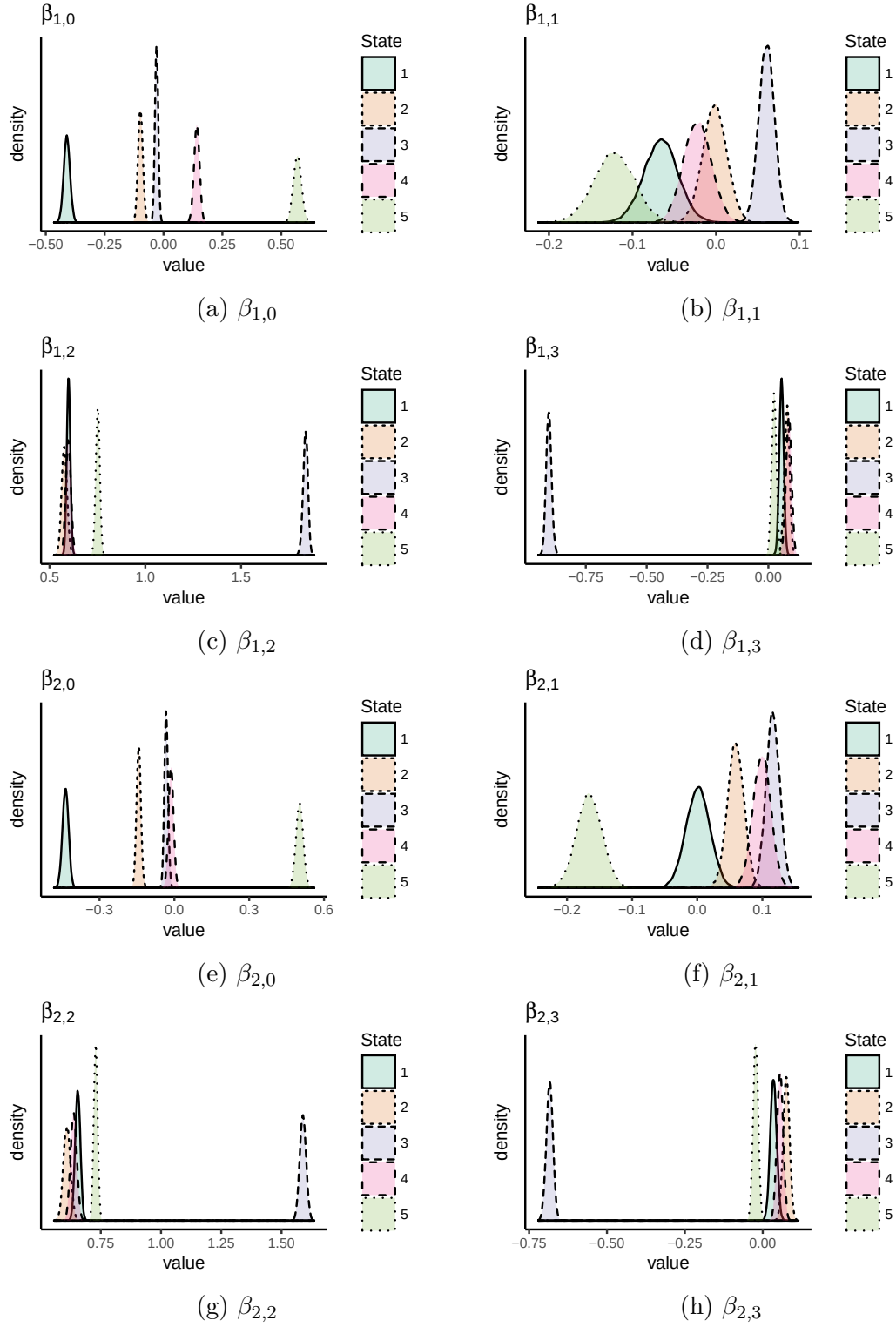


Figure B.2: Posterior distributions of the regression coefficients β for the all-year resource assessment model.

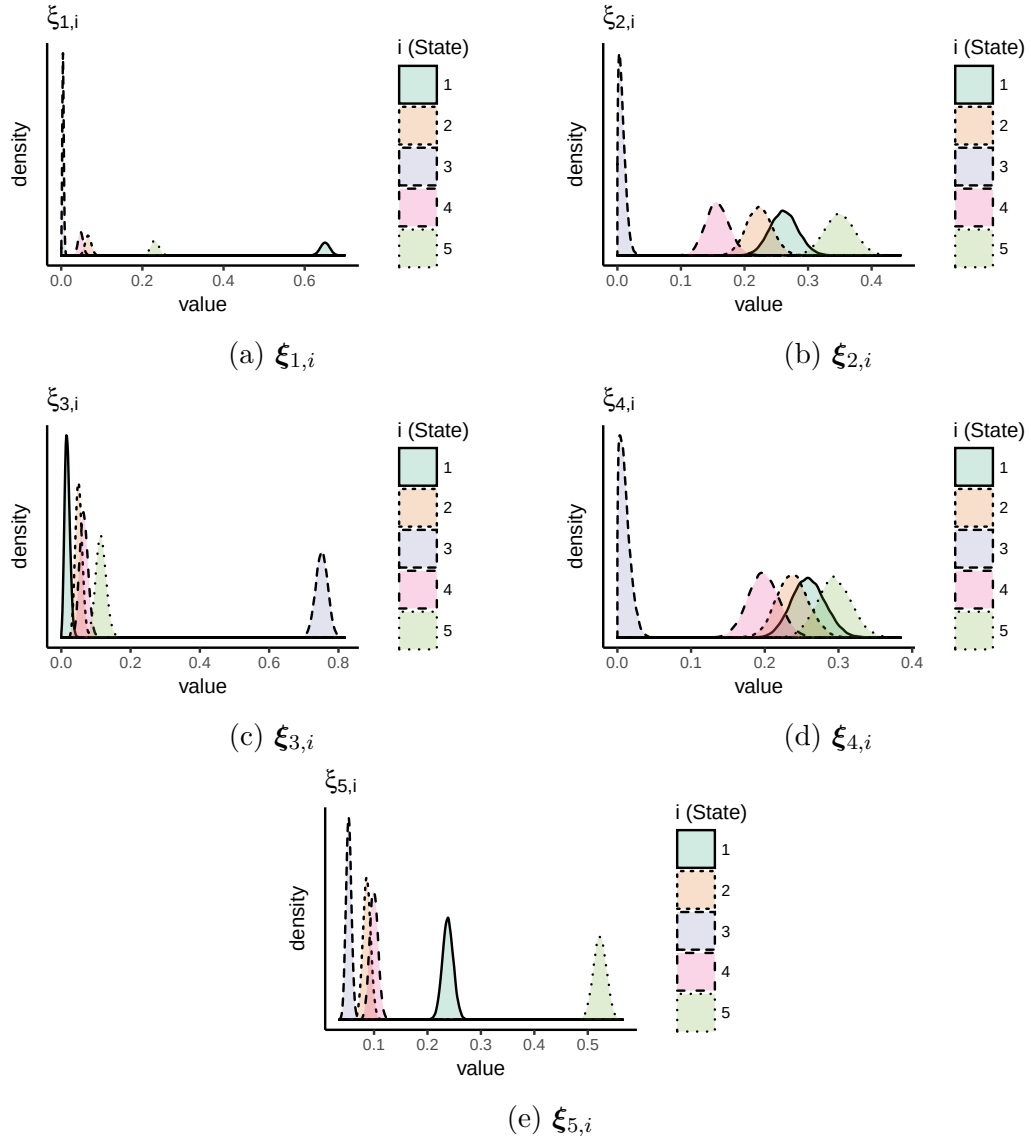


Figure B.3: Posterior distributions of the transition probabilities ξ for the all-year resource assessment model.

B.2 Winter Model

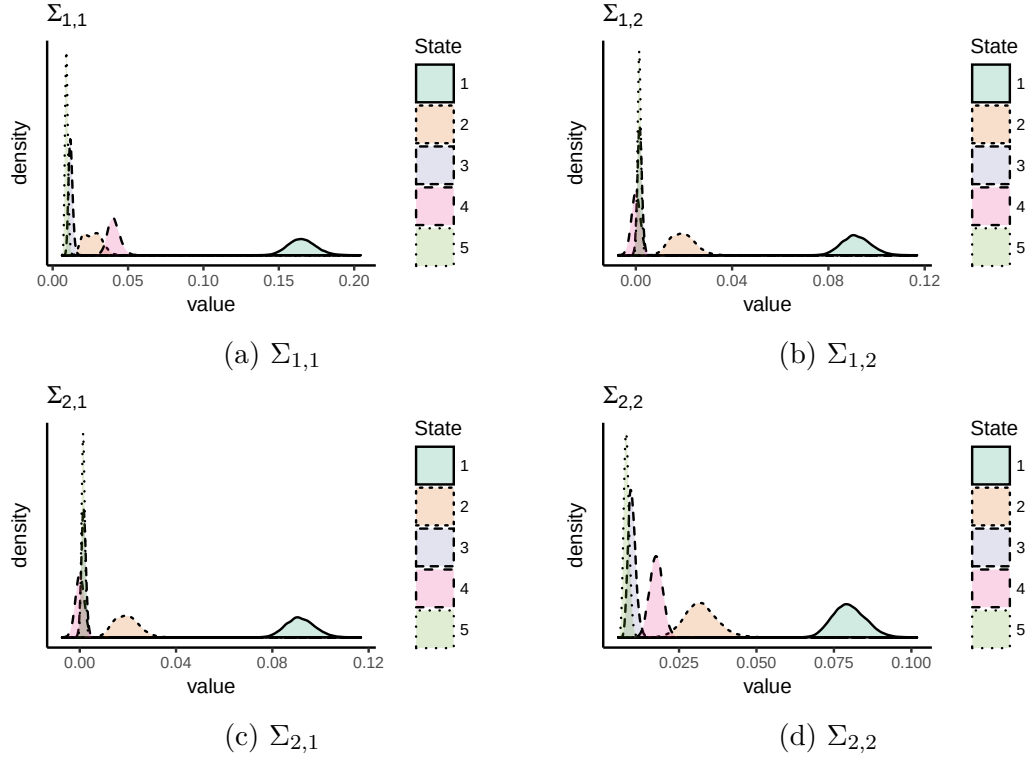


Figure B.4: Posterior distributions of the innovation covariance Σ components for the winter component of the seasonal resource assessment model.

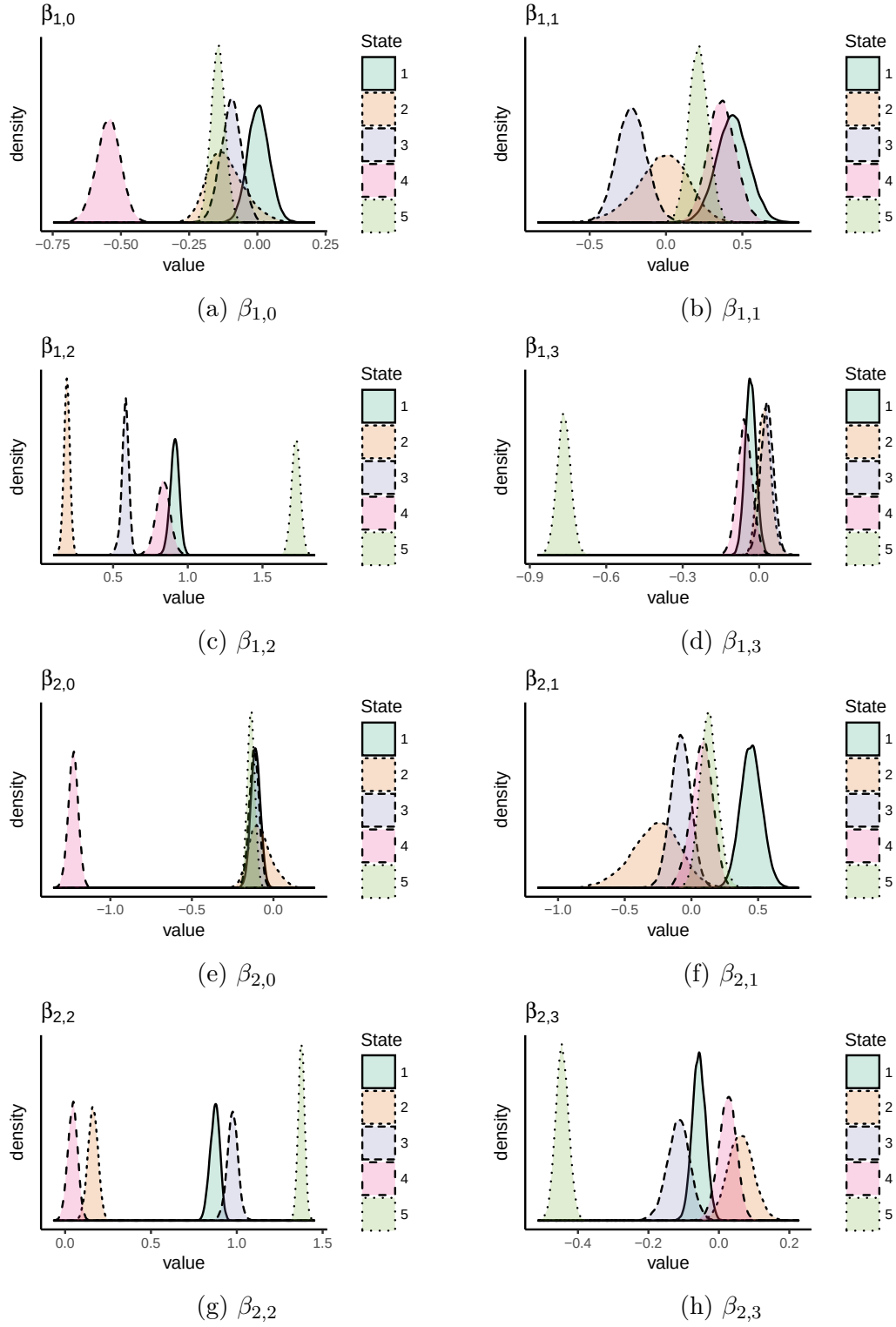


Figure B.5: Posterior distributions of the regression coefficients β for the winter component of the seasonal resource assessment model.

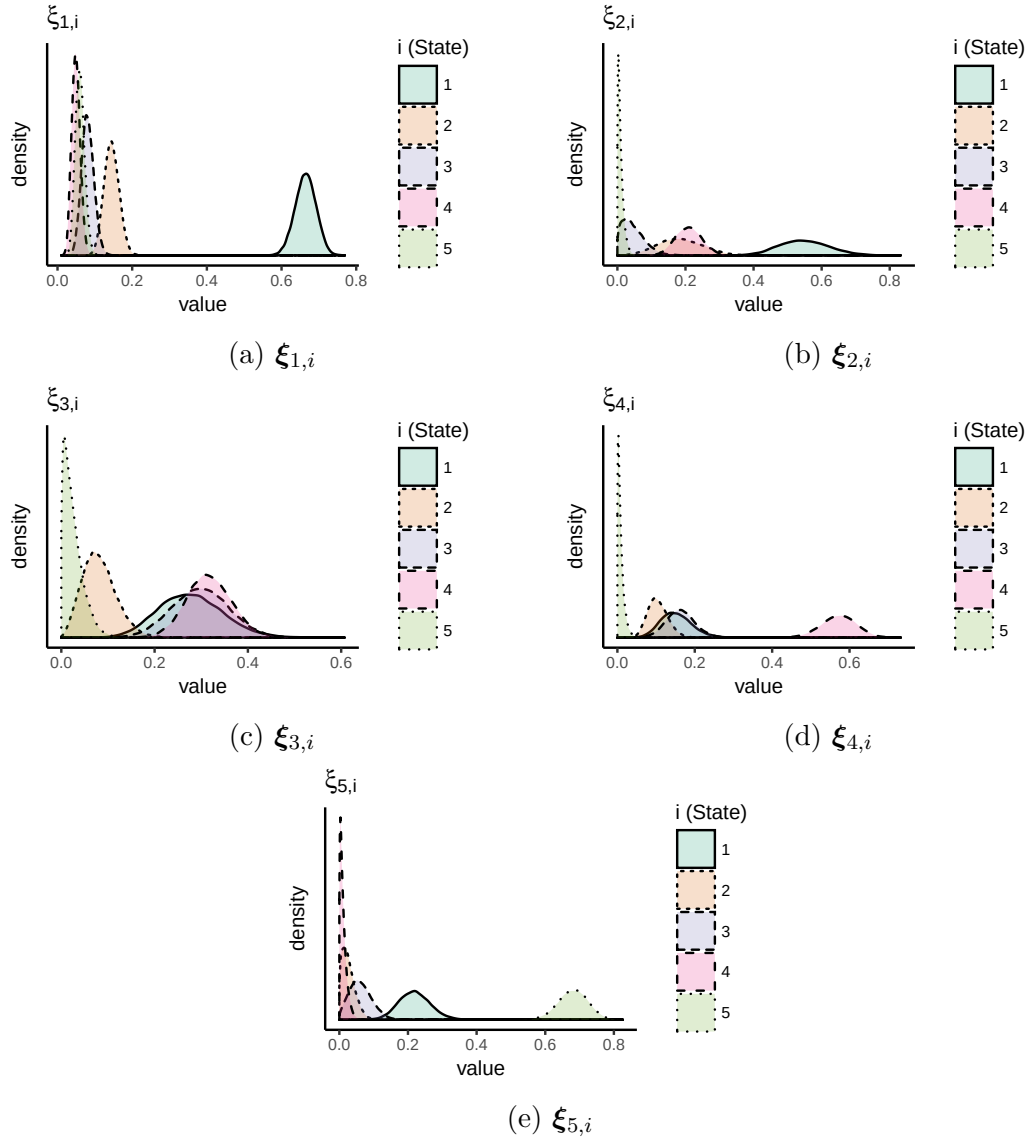


Figure B.6: Posterior distributions of the transition probabilities ξ for the winter component of the seasonal resource assessment model.

B.3 Spring Model

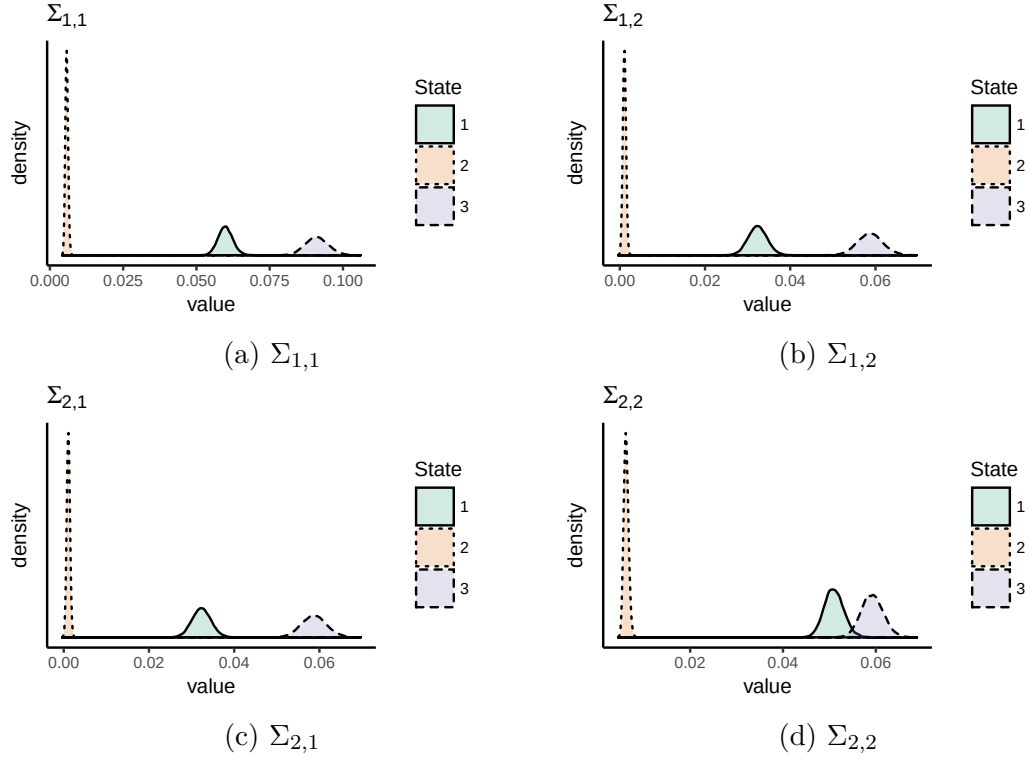


Figure B.7: Posterior distributions of the innovation covariance Σ components for the spring component of the seasonal resource assessment model.

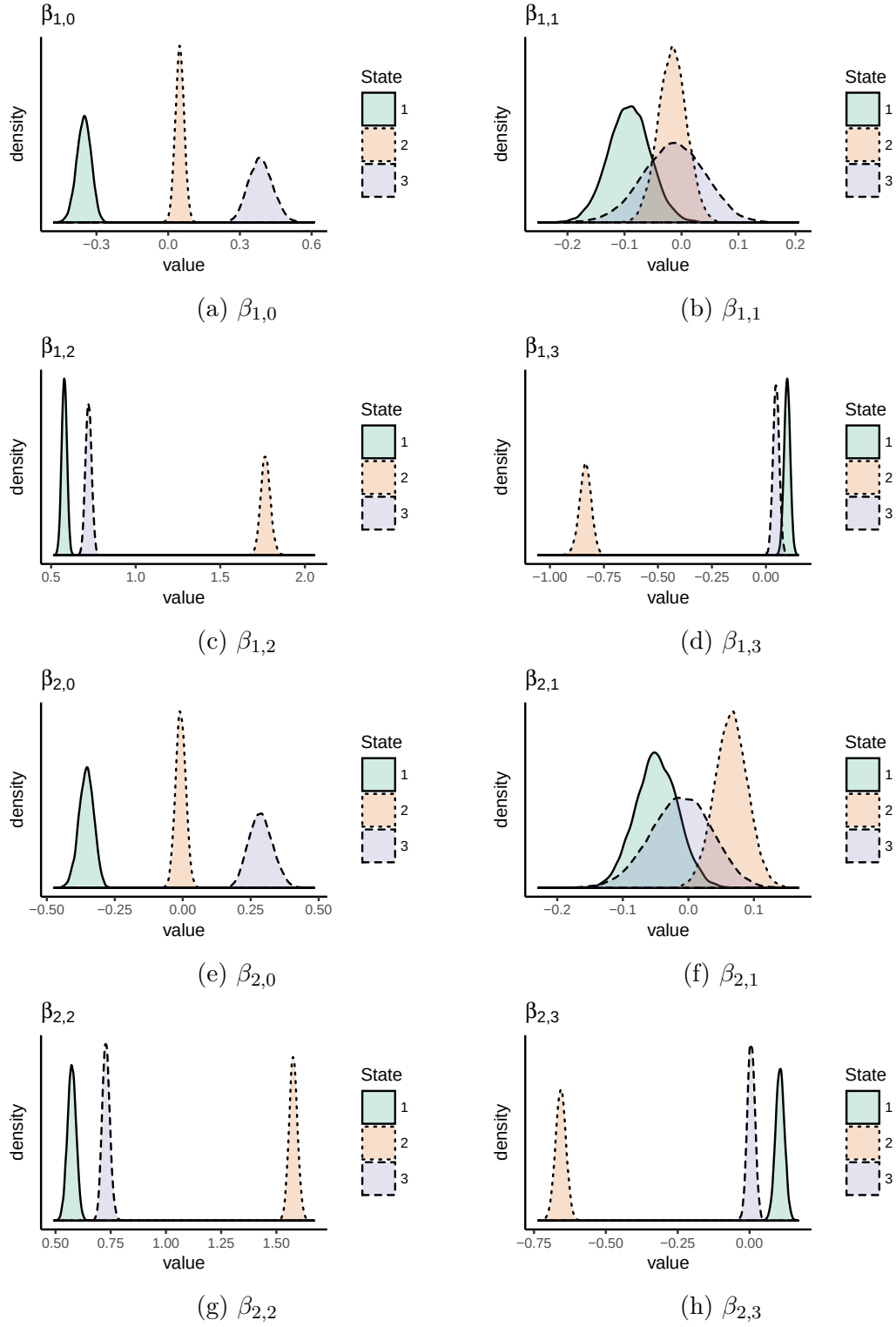


Figure B.8: Posterior distributions of the regression coefficients β for the spring component of the seasonal resource assessment model.

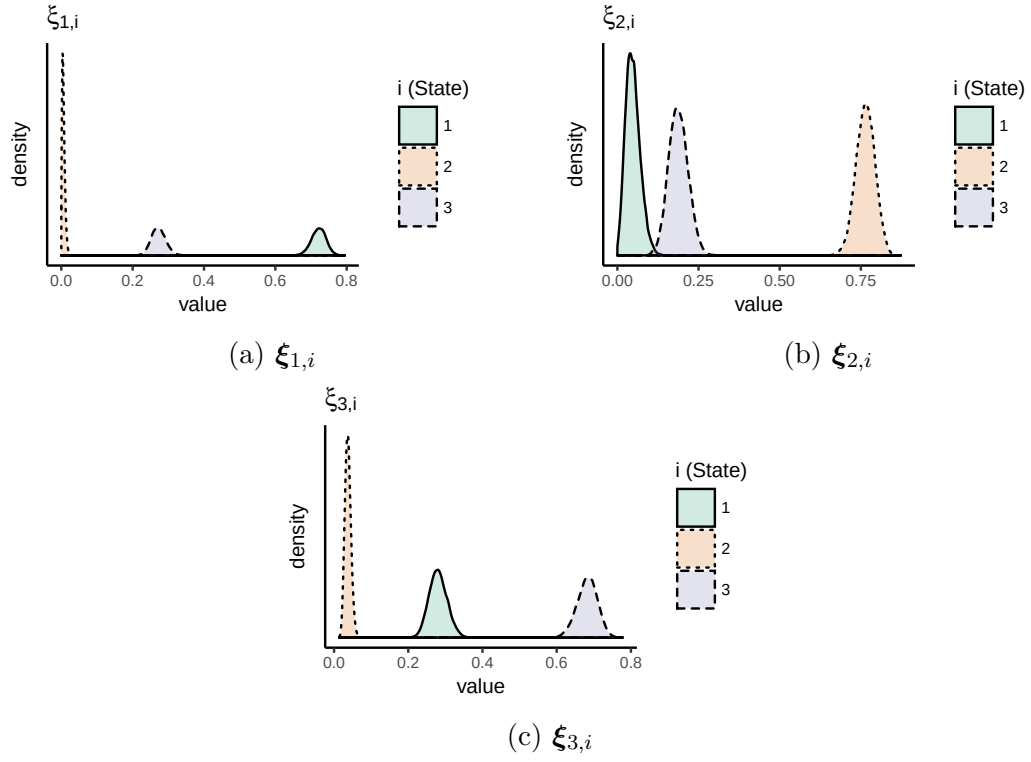


Figure B.9: Posterior distributions of the transition probabilities ξ for the spring component of the seasonal resource assessment model.

B.4 Summer Model

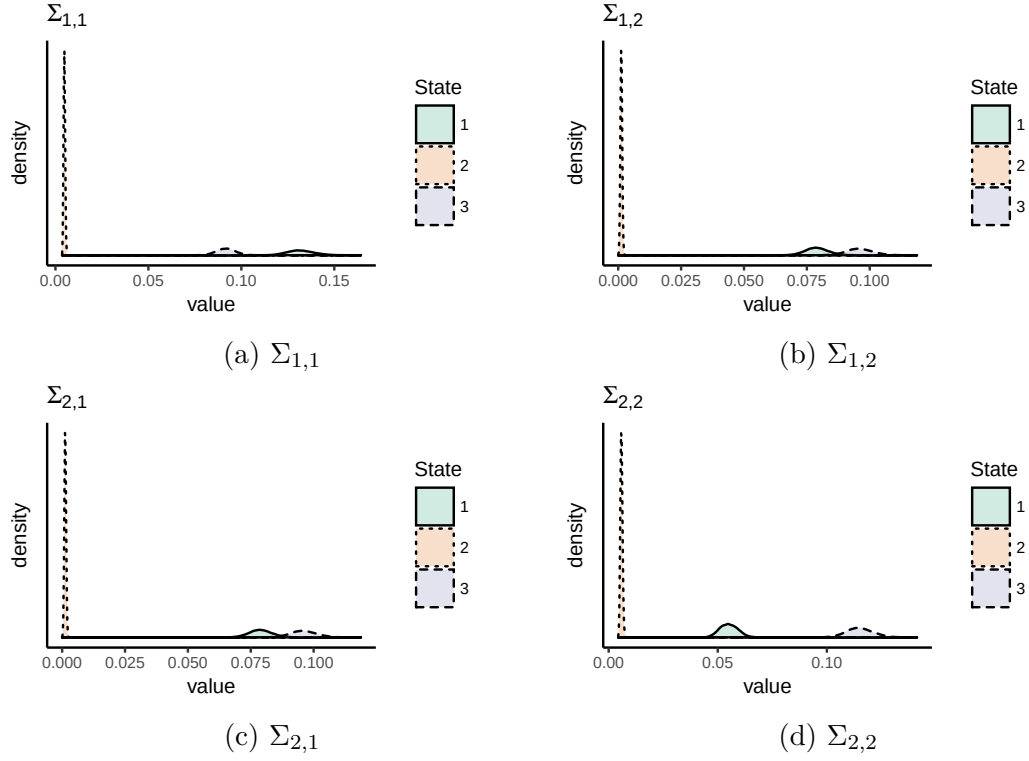


Figure B.10: Posterior distributions of the innovation covariance Σ components for the summer component of the seasonal resource assessment model.

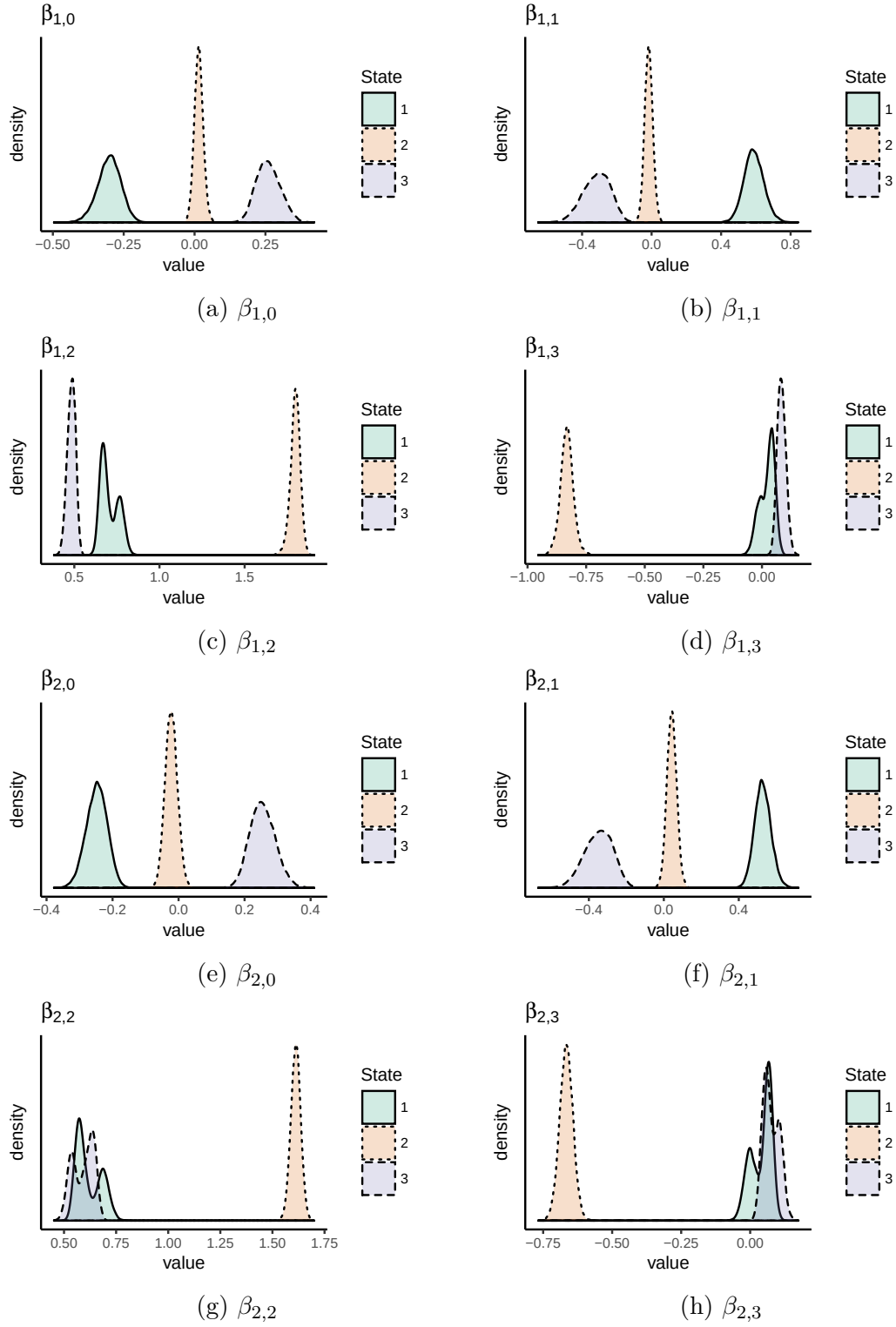


Figure B.11: Posterior distributions of the regression coefficients β for the summer component of the seasonal resource assessment model.

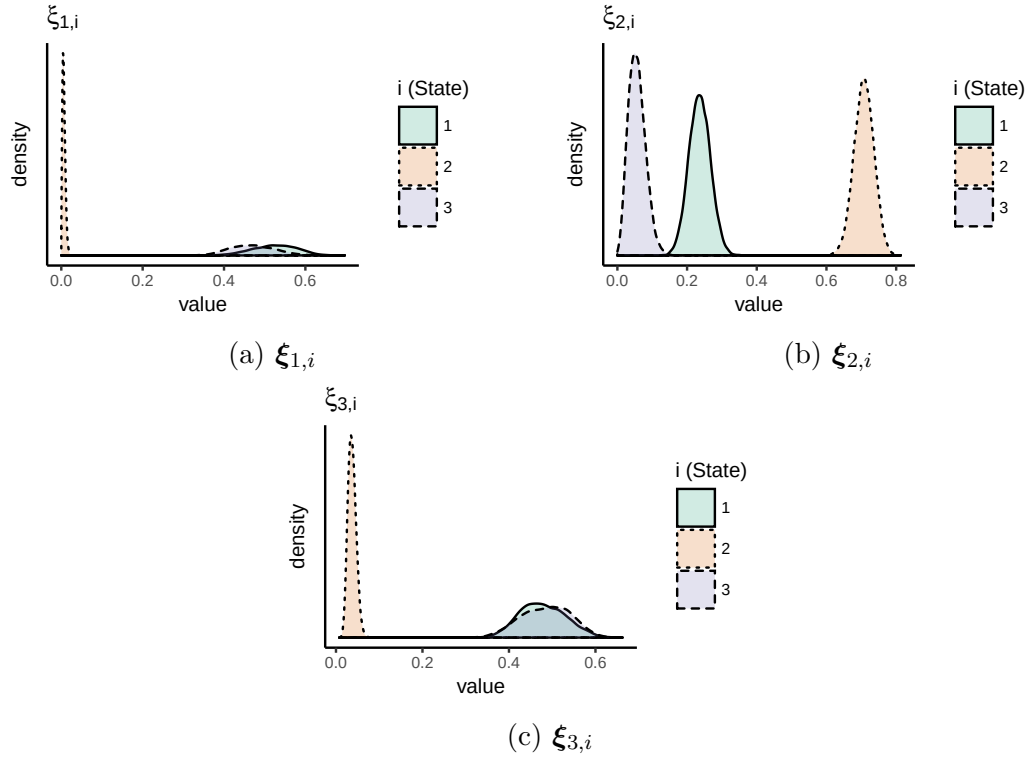


Figure B.12: Posterior distributions of the transition probabilities ξ for the summer component of the seasonal resource assessment model.

B.5 Fall Model

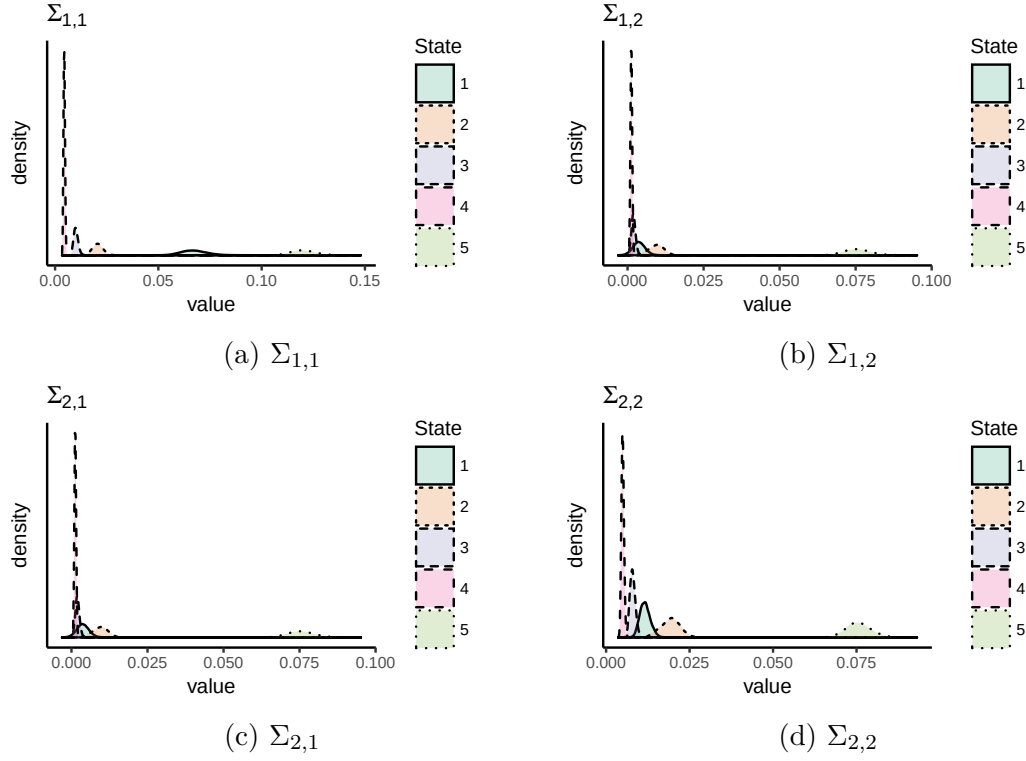


Figure B.13: Posterior distributions of the innovation covariance Σ components for the fall component of the seasonal resource assessment model.

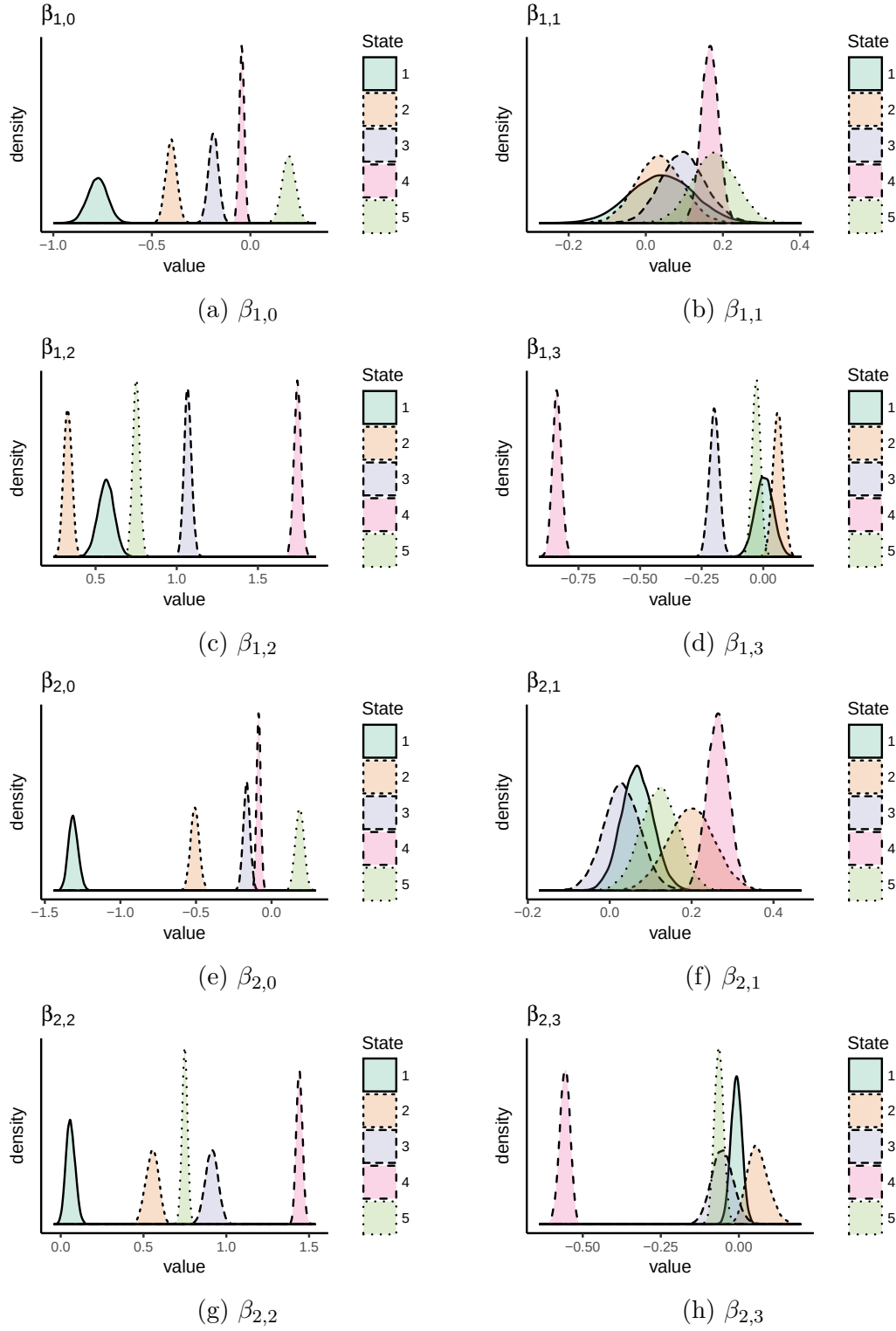


Figure B.14: Posterior distributions of the regression coefficients β for the fall component of the seasonal resource assessment model.

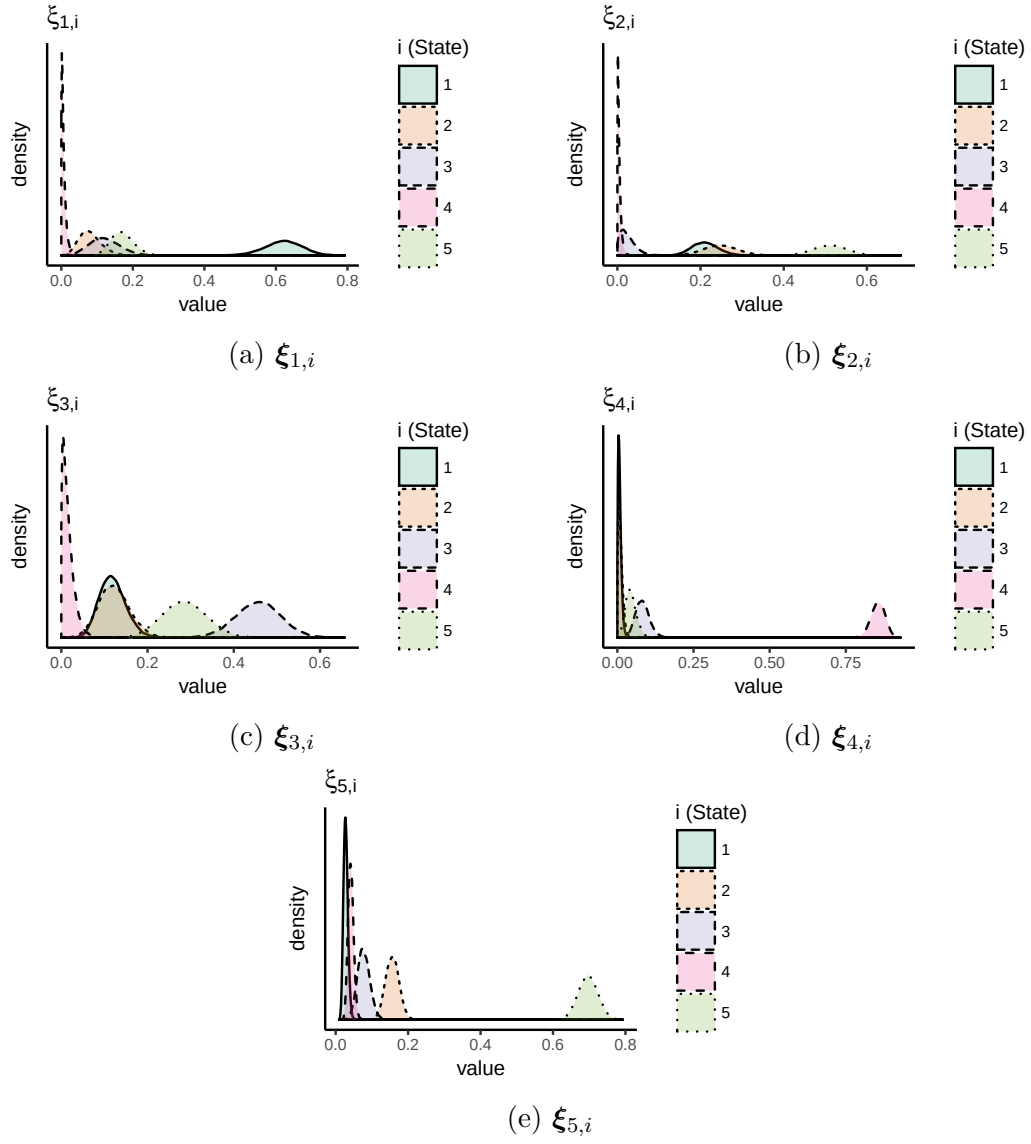


Figure B.15: Posterior distributions of the transition probabilities ξ for the fall component of the seasonal resource assessment model.

Appendix C

Forecast Posterior Distribution Plots

C.1 Homogeneous Model

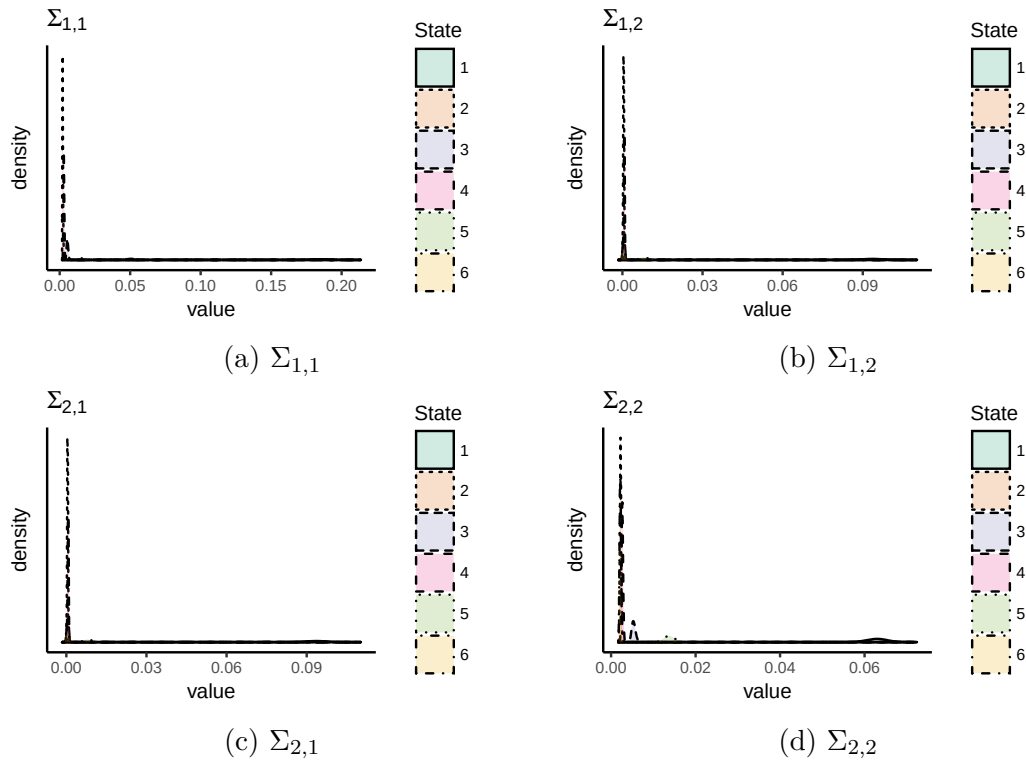


Figure C.1: Posterior distributions of the innovation covariance Σ components for the homogeneous forecast model.

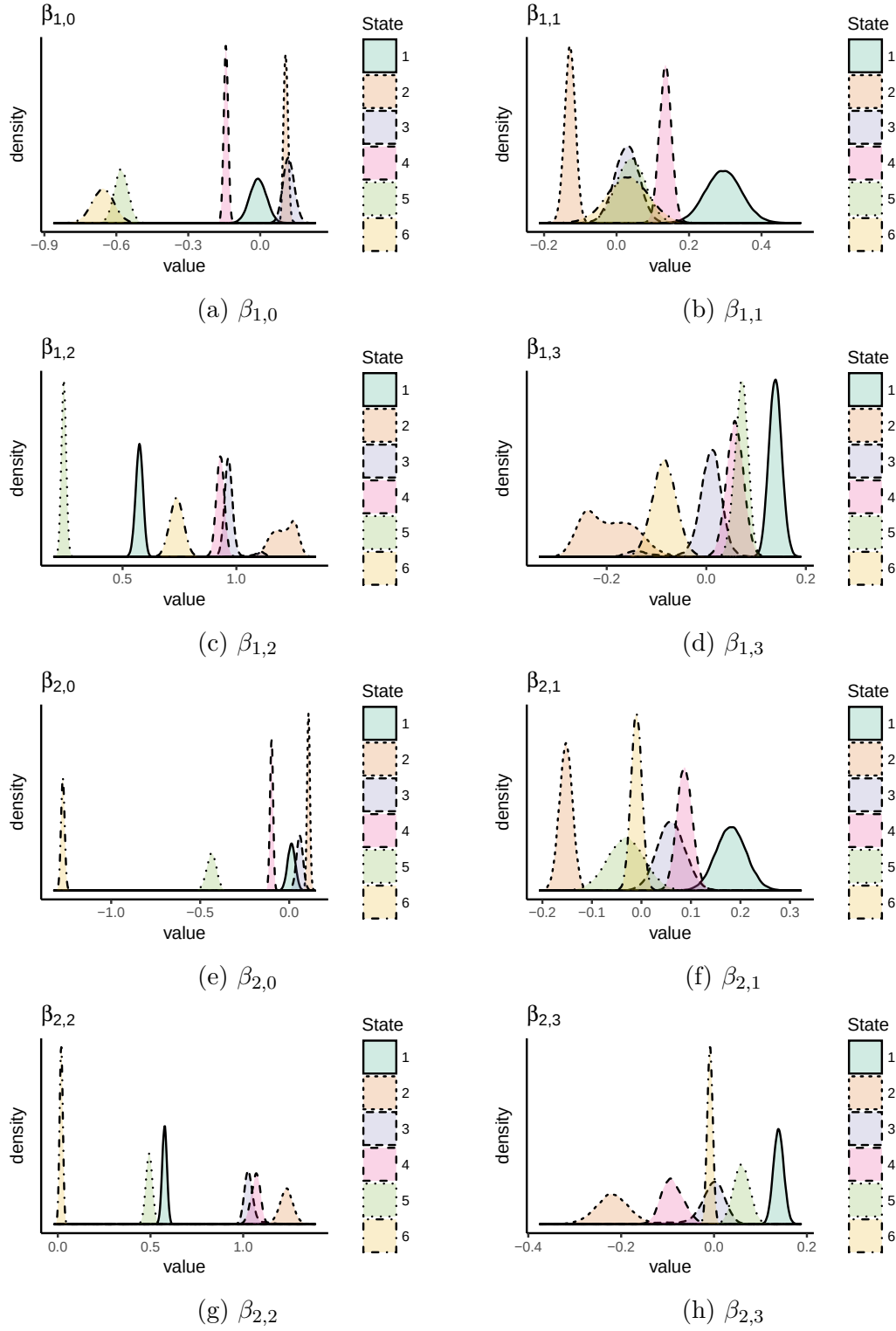


Figure C.2: Posterior distributions of the regression coefficients β for the homogeneous forecast model.

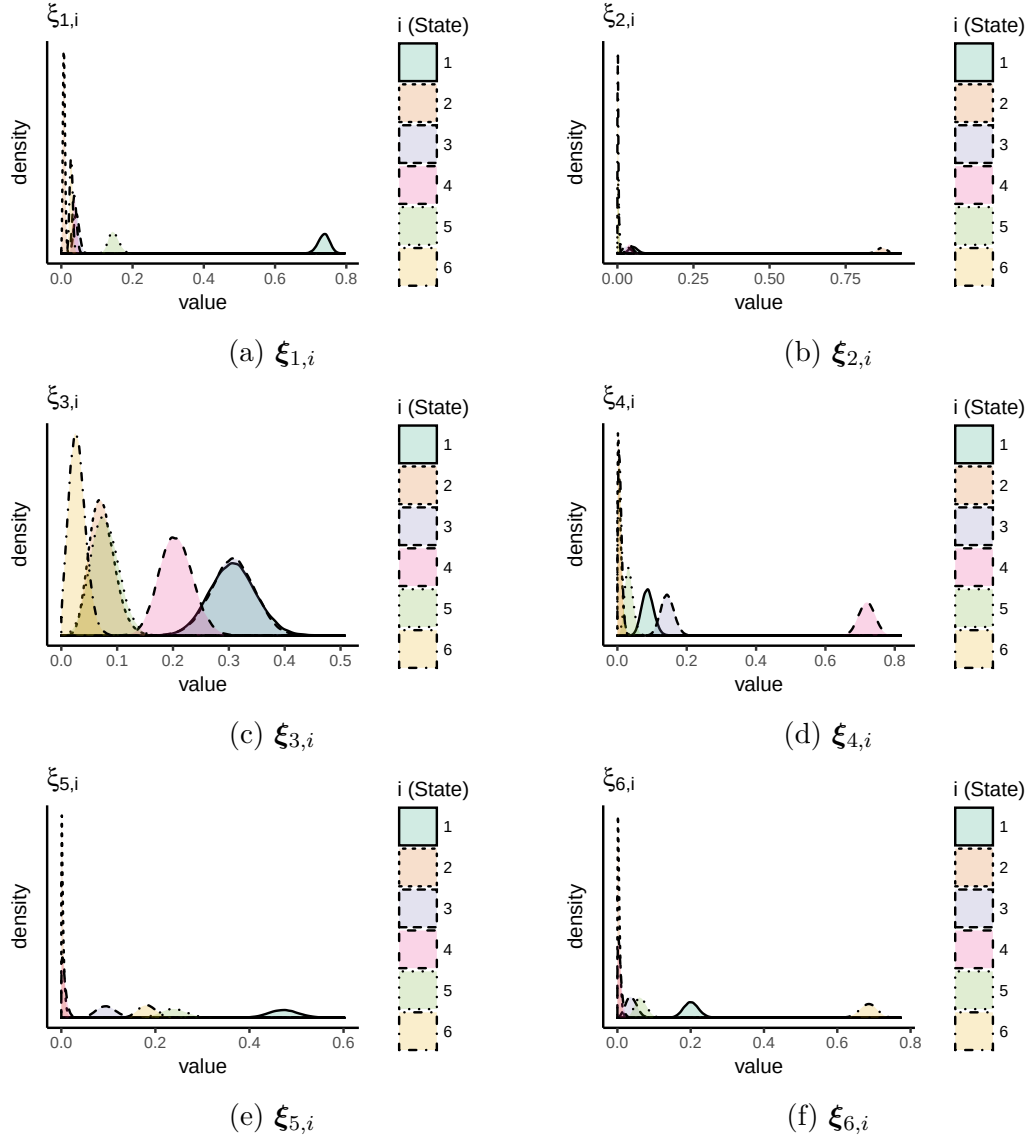


Figure C.3: Posterior distributions of the transition probabilities ξ for the homogeneous forecast model.

C.2 Inhomogeneous (Current) Model

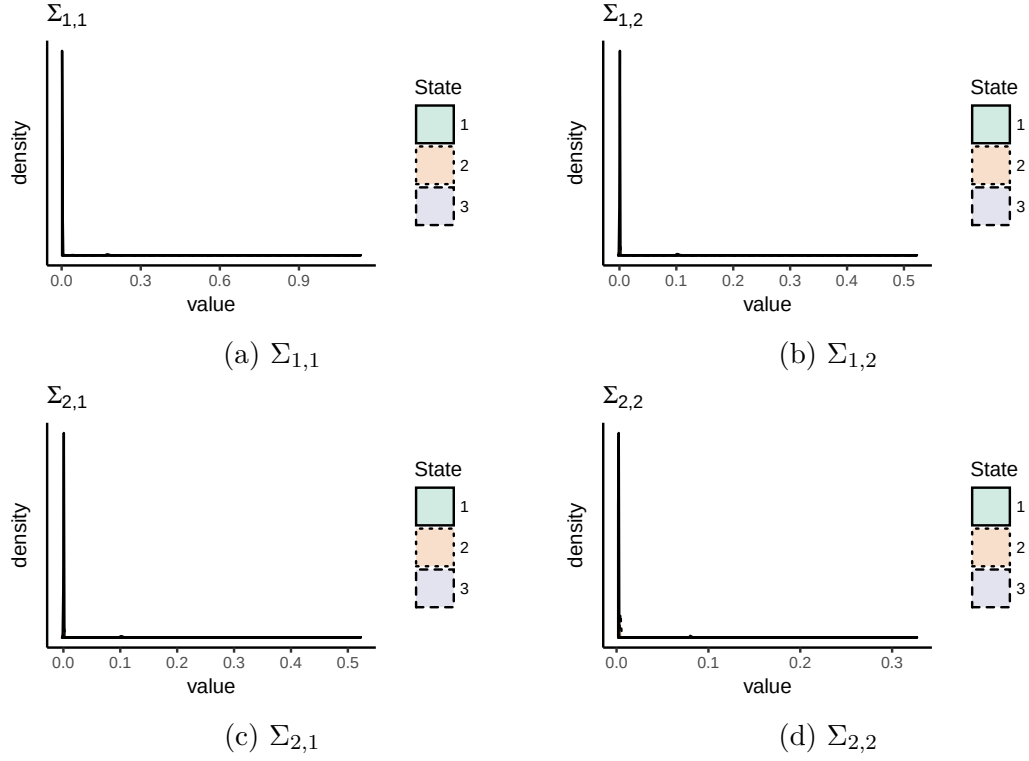


Figure C.4: Posterior distributions of the innovation covariance Σ components for the inhomogeneous forecast model using current transition covariates.

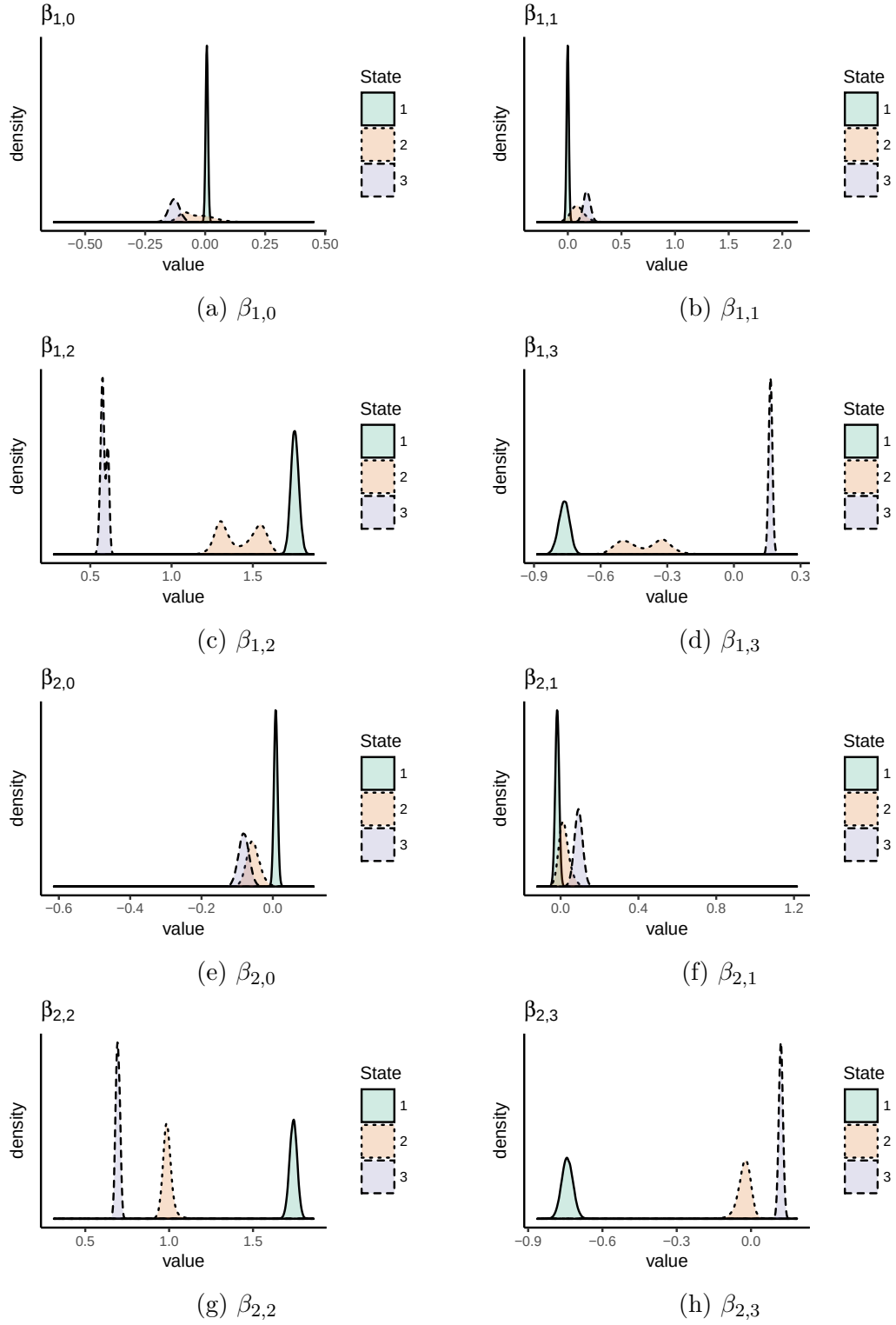


Figure C.5: Posterior distributions of the regression coefficients β for the inhomogeneous forecast model using current transition covariates.

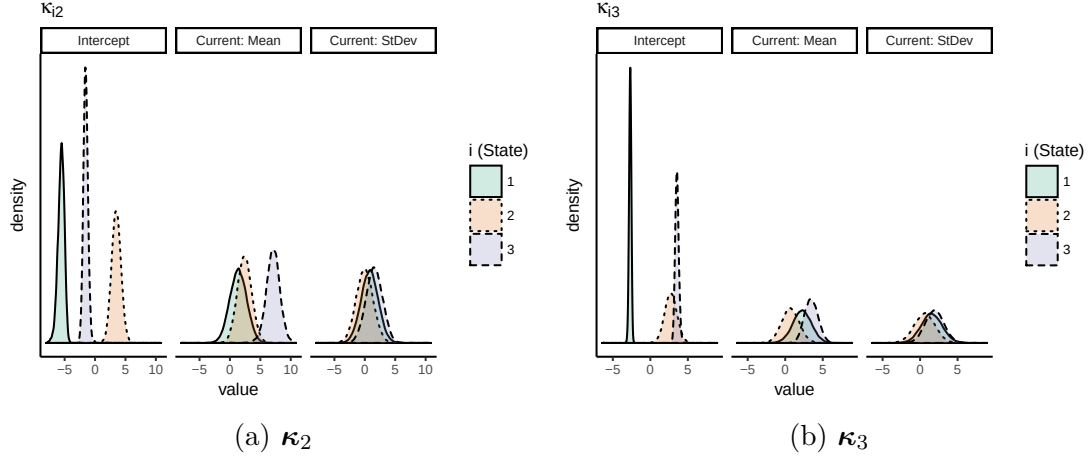


Figure C.6: Posterior distributions of the transition probability covariates κ for the inhomogeneous forecast model using current transition covariates.

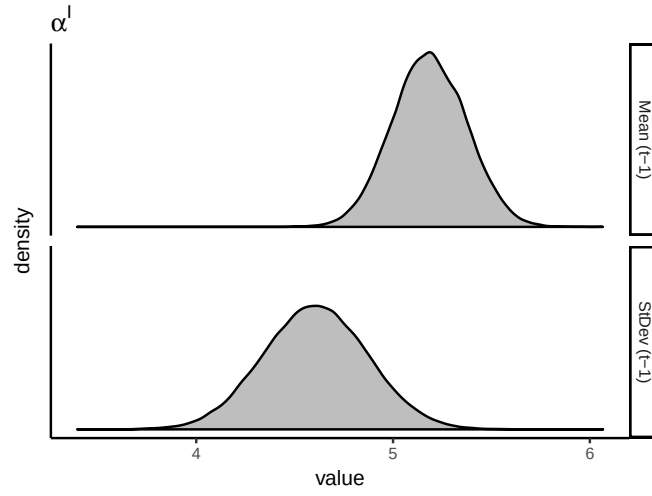


Figure C.7: Posterior distribution of the mixture covariates α^I for the inhomogeneous forecast model using current transition covariates.

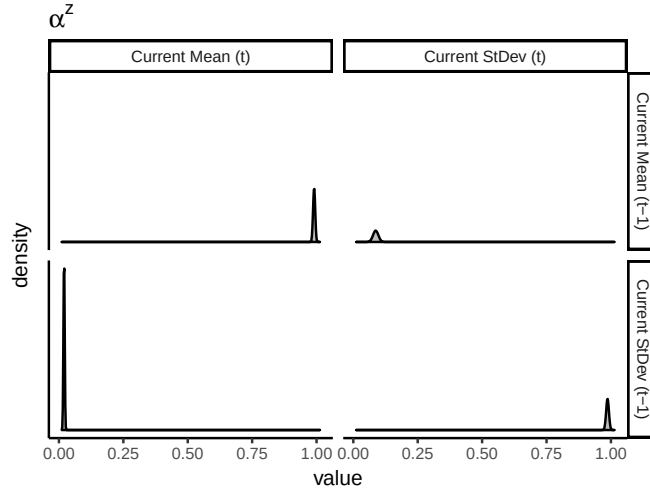


Figure C.8: Posterior distribution of the missing-data imputation coefficients α^z for the inhomogeneous forecast model using current transition covariates.

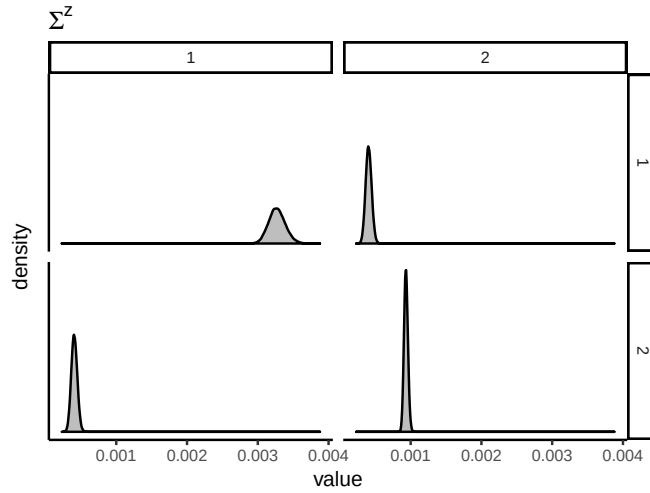


Figure C.9: Posterior distribution of the missing-data imputation innovation covariance Σ^z for the inhomogeneous forecast model using current transition covariates.

C.3 Inhomogeneous (Taylor) Model

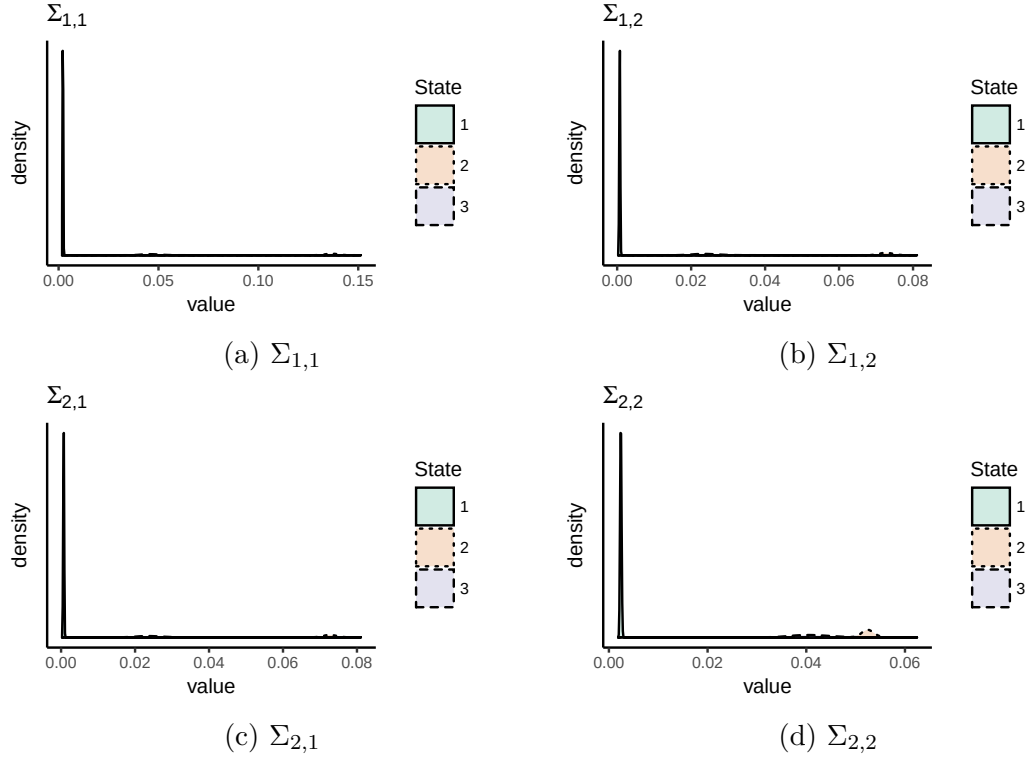


Figure C.10: Posterior distributions of the innovation covariance Σ components for the inhomogeneous forecast model using Taylor transition covariates.

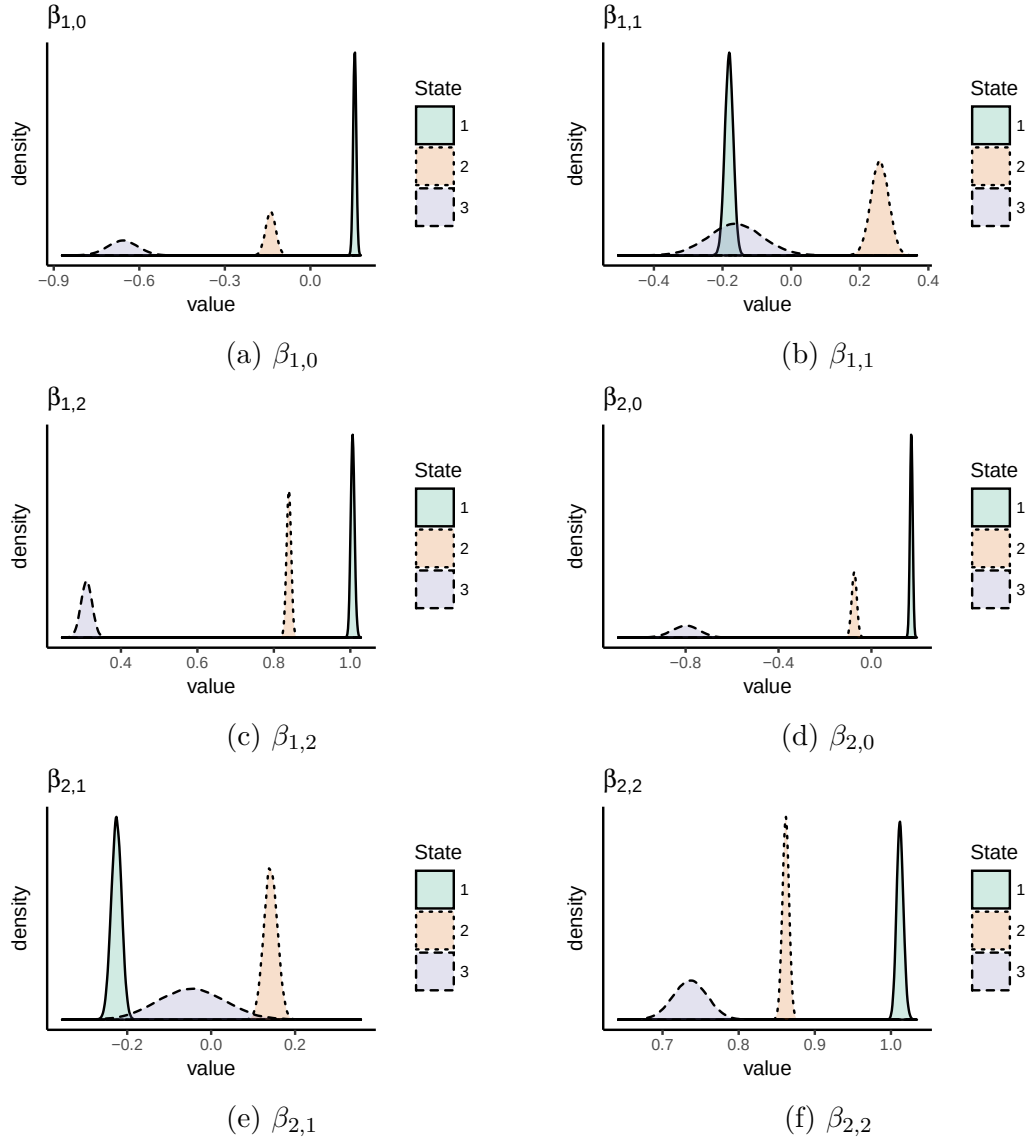


Figure C.11: Posterior distributions of the regression coefficients β for the inhomogeneous forecast model using Taylor transition covariates.

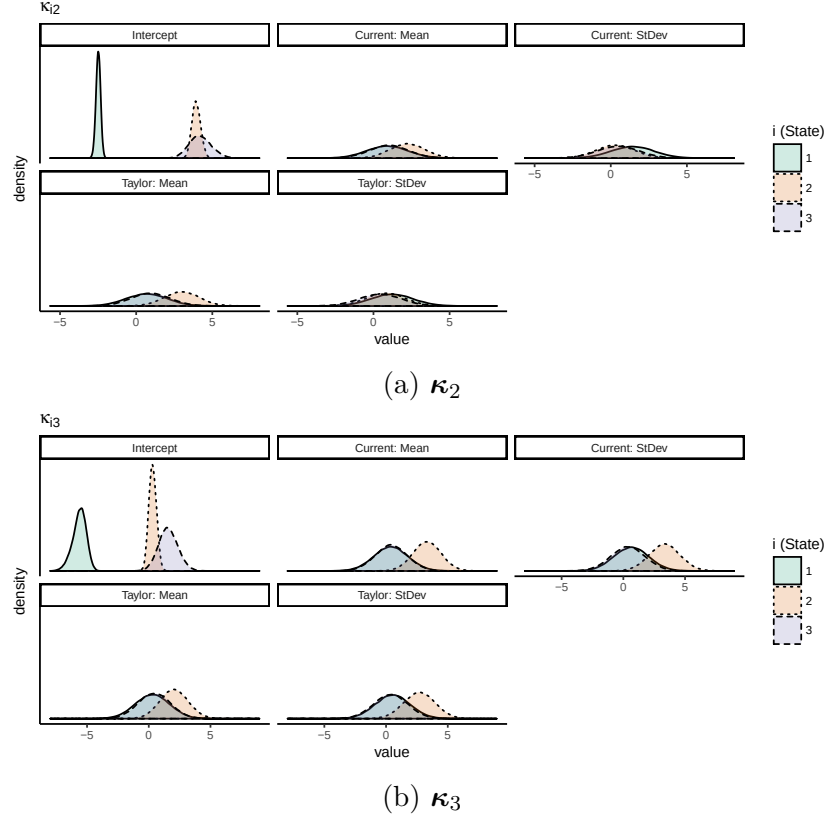


Figure C.12: Posterior distributions of the transition probability covariates κ for the inhomogeneous forecast model using Taylor transition covariates.

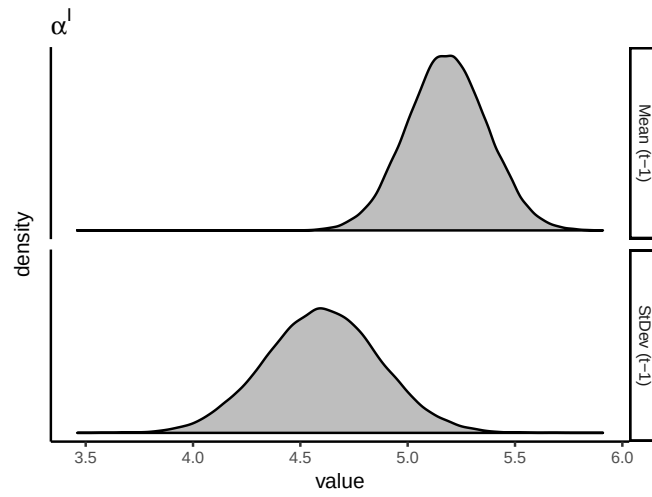


Figure C.13: Posterior distribution of the mixture covariates α^I for the inhomogeneous forecast model using Taylor transition covariates.

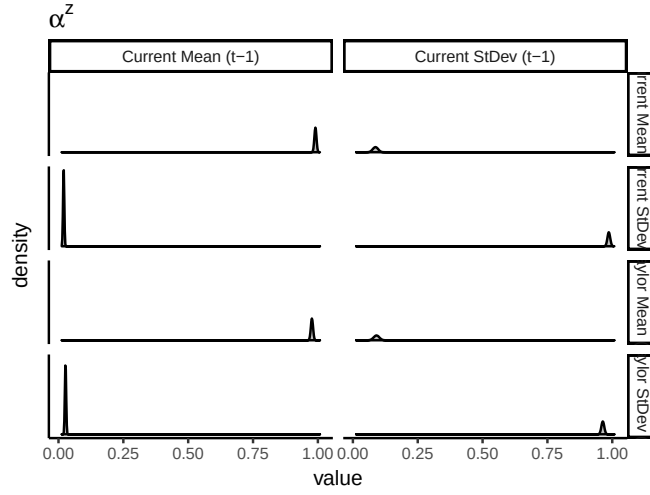


Figure C.14: Posterior distribution of the missing-data imputation coefficients α^z for the inhomogeneous forecast model using Taylor transition covariates.

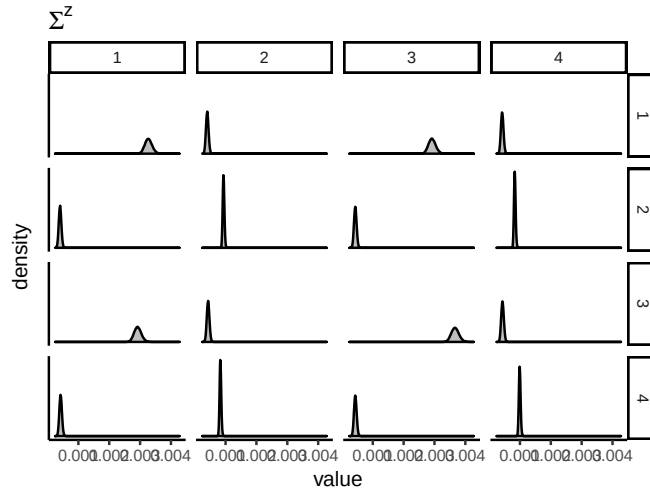


Figure C.15: Posterior distribution of the missing-data imputation innovation covariance Σ^z for the inhomogeneous forecast model using Taylor transition covariates.

References

- Aguiar, L. M. et al. (2016). “Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting”. en. In: *Renewable Energy* 97, pages 599–610. ISSN: 0960-1481. DOI: [10.1016/j.renene.2016.06.018](https://doi.org/10.1016/j.renene.2016.06.018).
- Aguiar, R. J. and M. Collares-Pereira (1992a). “Statistical properties of hourly global radiation”. In: *Solar Energy* 48.3, pages 157–167. ISSN: 0038-092X. DOI: [10.1016/0038-092X\(92\)90134-V](https://doi.org/10.1016/0038-092X(92)90134-V).
- (1992b). “TAG: A time-dependent, autoregressive, Gaussian model for generating synthetic hourly radiation”. In: *Solar Energy* 49.3, pages 167–174. ISSN: 0038-092X. DOI: [10.1016/0038-092X\(92\)90068-L](https://doi.org/10.1016/0038-092X(92)90068-L).
- Aguiar, R. J., M. Collares-Pereira, and J. P. Conde (1988). “Simple procedure for generating sequences of daily radiation values using a library of Markov transition matrices”. In: *Solar Energy* 40.3, pages 269–279. ISSN: 0038-092X. DOI: [10.1016/0038-092X\(88\)90049-7](https://doi.org/10.1016/0038-092X(88)90049-7).
- Ailliot, P. and V. Monbet (2012). “Markov-switching autoregressive models for wind time series”. In: *Environmental Modelling & Software* 30.Supplement C, pages 92–101. ISSN: 1364-8152. DOI: [10.1016/j.envsoft.2011.10.011](https://doi.org/10.1016/j.envsoft.2011.10.011).
- Akaike, H. (1974). “A New Look at the Statistical Model Identification”. In: *IEEE Transactions on Automatic Control* 19.6, pages 716–723. ISSN: 1558-2523. DOI: [10.1109/TAC.1974.1100705](https://doi.org/10.1109/TAC.1974.1100705). arXiv: [arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).
- Alonso-Montesinos, J., F. J. Batlles, and C. Portillo (2015). “Solar irradiance forecasting at one-minute intervals for different sky conditions using sky camera images”. en. In: *Energy Convers. Manage.* 105, pages 1166–1177. ISSN: 0196-8904, 1879-2227. DOI: [10.1016/j.enconman.2015.09.001](https://doi.org/10.1016/j.enconman.2015.09.001).
- Ashenfelter, O. and K. Storchmann (2006). “Using a Hedonic Model of Solar Radiation to Assess the Economic Effect of Climate Change: The Case of Mosel Valley Vineyards”.
- Augustine, J. A., J. J. DeLuisi, and C. N. Long (2000). “SURFRAD – A National Surface Radiation Budget Network for Atmospheric Research”. In: *Bulletin of the American Meteorological Society* 81.10, pages 2341–2357. ISSN: 0003-0007. DOI: [10.1175/1520-0477\(2000\)081<2341:SANSRB>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<2341:SANSRB>2.3.CO;2).

- Bartoli, B. et al. (1981). “Autocorrelation of daily global solar radiation”. en. In: *Nuovo Cimento: C: Geophys. Space Phys.* 4.2, pages 113–122. ISSN: 1124-1896, 0390-5551. DOI: [10.1007/BF02507395](https://doi.org/10.1007/BF02507395).
- Batlles, F. J. et al. (2000). “Empirical modeling of hourly direct irradiance by means of hourly global irradiance”. In: *Energy* 25.7, pages 675–688. ISSN: 0360-5442. DOI: [10.1016/S0360-5442\(00\)00007-4](https://doi.org/10.1016/S0360-5442(00)00007-4).
- Bayes, T. (1763). “An Essay towards Solving a Problem in the Doctrine of Chance”. In: *Philosophical Transactions of the Royal Society of London* 53, pages 370–418.
- Bendt, P., M. Collares-Pereira, and A. Rabl (1981). “The frequency distribution of daily insolation values”. In: *Solar Energy* 27.1, pages 1–5. ISSN: 0038-092X. DOI: [10.1016/0038-092X\(81\)90013-X](https://doi.org/10.1016/0038-092X(81)90013-X).
- Bhardwaj, S. et al. (2013). “Estimation of solar radiation using a combination of Hidden Markov Model and generalized Fuzzy model”. In: *Solar Energy* 93.Supplement C, pages 43–54. ISSN: 0038-092X. DOI: [10.1016/j.solener.2013.03.020](https://doi.org/10.1016/j.solener.2013.03.020).
- Biga, A. J. and R. Rosa (1981). “Statistical behaviour of solar irradiation over consecutive days”. In: *Solar Energy* 27.2, pages 149–157. ISSN: 0038-092X. DOI: [10.1016/0038-092X\(81\)90037-2](https://doi.org/10.1016/0038-092X(81)90037-2).
- Billio, M. and R. Casarin (2011). “Beta Autoregressive Transition Markov-Switching Models for Business Cycle Analysis”. In: *Stud. Nonlinear Dyn. Econom.* 15.4. ISSN: 1558-3708. DOI: [10.2202/1558-3708.1856](https://doi.org/10.2202/1558-3708.1856).
- Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. en. Clarendon Press. ISBN: 9780198538646.
- Blair, N. et al. (2014). *System Advisor Model, SAM 2014.1.14: General Description*. Technical report NREL/TP-6A20-61019. National Renewable Energy Laboratory.
- Boileau, E. (1983). “Use of some simple statistical models in solar meteorology”. In: *Solar Energy* 30.4, pages 333–339. ISSN: 0038-092X. DOI: [10.1016/0038-092X\(83\)90187-1](https://doi.org/10.1016/0038-092X(83)90187-1).
- Bois, B. et al. (2008). “Using remotely sensed solar radiation data for reference evapotranspiration estimation at a daily time step”. In: *Agric. For. Meteorol.* 148.4, pages 619–630. ISSN: 0168-1923. DOI: [10.1016/j.agrformet.2007.11.005](https://doi.org/10.1016/j.agrformet.2007.11.005).
- Boland, J., J. Huang, and B. Ridley (2013). “Decomposing global solar radiation into its direct and diffuse components”. In: *Renewable Sustainable Energy Rev.* 28.Supplement C, pages 749–756. ISSN: 1364-0321. DOI: [10.1016/j.rser.2013.08.023](https://doi.org/10.1016/j.rser.2013.08.023).
- Brinkworth, B. J. (1977). “Autocorrelation and stochastic modelling of insolation sequences”. In: *Solar Energy* 19.4, pages 343–347. ISSN: 0038-092X. DOI: [10.1016/0038-092X\(77\)90004-4](https://doi.org/10.1016/0038-092X(77)90004-4).
- Brownson, J. R. S. (2014). *Solar Energy Conversion Systems*. Oxford: Elsevier Academic Press. ISBN: 9780123970213.

- Buckingham, E. (1914). “On Physically Similar Systems; Illustrations of the Use of Dimensional Equations”. In: *Physical Review* 4.4, pages 345–376. DOI: [10.1103/PhysRev.4.345](#).
- Calinoiu, D.-G. et al. (2014). “Evaluation of errors made in solar irradiance estimation due to averaging the Angstrom turbidity coefficient”. In: *Atmos. Res.* 150.Supplement C, pages 69–78. ISSN: 0169-8095. DOI: [10.1016/j.atmosres.2014.07.015](#).
- Cannon, A. J. (2011). “Quantile regression neural networks: Implementation in R and application to precipitation downscaling”. In: *Comput. Geosci.* 37.9, pages 1277–1284. ISSN: 0098-3004. DOI: [10.1016/j.cageo.2010.07.005](#).
- Cano, D. et al. (1986). “A method for the determination of the global solar radiation from meteorological satellite data”. In: *Solar Energy* 37.1, pages 31–39. ISSN: 0038-092X. DOI: [10.1016/0038-092X\(86\)90104-0](#).
- Chen, C.-R. and U. T. Kartini (2017). “k-Nearest Neighbor Neural Network Models for Very Short-Term Global Solar Irradiance Forecasting Based on Meteorological Data”. en. In: *Energies* 10.2. ISSN: 1996-1073. DOI: [10.3390/en10020186](#).
- Chu, Y. and C. F. M. Coimbra (2017). “Short-term probabilistic forecasts for Direct Normal Irradiance”. en. In: *Renewable Energy* 101, pages 526–536. ISSN: 0960-1481. DOI: [10.1016/j.renene.2016.09.012](#).
- Chu, Y., M. Li, and C. F. M. Coimbra (2016). “Sun-tracking imaging system for intra-hour DNI forecasts”. en. In: *Renewable Energy* 96, pages 792–799. ISSN: 0960-1481. DOI: [10.1016/j.renene.2016.05.041](#).
- Chu, Y. et al. (2014). “A Smart Image-Based Cloud Detection System for Intrahour Solar Irradiance Forecasts”. en. In: *J. Atmos. Ocean. Technol.* 31.9, pages 1995–2007. ISSN: 0739-0572, 1520-0426. DOI: [10.1016/j.renene.2016.01.020](#).
- Chuffart, T. (2015). “Selection Criteria in Regime Switching Conditional Volatility Models”. In: *Econometrics* 3.2, pages 289–316. ISSN: 2225-1146. DOI: [10.3390/econometrics3020289](#).
- Curtright, A. E. and J. Apt (2008). “The Character of Power Output from Utility-Scale Photovoltaic Systems”. In: *Prog. Photovoltaics Res. Appl.* 16, pages 241–247. DOI: [10.1002/pip](#).
- Dall’Anese, E. et al. (2015). “Optimal Dispatch of Residential Photovoltaic Inverters Under Forecasting Uncertainties”. en. In: *IEEE Journal of Photovoltaics* 5.1, pages 350–359. ISSN: 2156-3381. DOI: [10.1109/JPHOTOV.2014.2364125](#).
- David, M. et al. (2016). “Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models”. en. In: *Solar Energy* 133, pages 55–72. ISSN: 0038-092X. DOI: [10.1016/j.solener.2016.03.064](#).
- Dong, Z. et al. (2013). “Short-term solar irradiance forecasting using exponential smoothing state space model”. en. In: *Energy* 55, pages 1104–1113. ISSN: 0360-5442, 1873-6785. DOI: [10.1016/j.energy.2013.04.027](#).

- Dorrestijn, J. et al. (2013). “Stochastic parameterization of shallow cumulus convection estimated from high-resolution model data”. In: *Theor. Comput. Fluid Dyn.* 27, pages 133–148. ISSN: 0935-4964. DOI: [10.1007/s00162-012-0281-y](https://doi.org/10.1007/s00162-012-0281-y).
- Durkay, J. (2017). *State Renewable Portfolio Standards and Goals*. URL: <http://www.ncsl.org/research/energy/renewable-portfolio-standards.aspx>.
- Eber, K. and D. Corbus (2013). *HAWAII SOLAR INTEGRATION STUDY: Executive Summary*. Technical report NREL/TP-5500-57215. National Renewable Energy Laboratory, pages 1–40.
- EIA (2017). *Electric Power Monthly*. Technical report. Washington, DC: U.S. Energy Information Administration. URL: <https://www.eia.gov/electricity/monthly/archive/february2017.pdf>.
- Ela, E. et al. (2013). *Impacts of Variability and Uncertainty in Solar Photovoltaic Generation at Multiple Timescales*. Technical report NREL/TP-5500-58274. National Renewable Energy Laboratory.
- Elizondo, D., G. Hoogenboom, and R. W. McClendon (1994). “Development of a neural network model to predict daily solar radiation”. In: *Agric. For. Meteorol.* 71.1, pages 115–132. ISSN: 0168-1923. DOI: [10.1016/0168-1923\(94\)90103-1](https://doi.org/10.1016/0168-1923(94)90103-1).
- Engerer, N. A. (2015). “Minute resolution estimates of the diffuse fraction of global irradiance for southeastern Australia”. In: *Solar Energy* 116. Supplement C, pages 215–237. ISSN: 0038-092X. DOI: [10.1016/j.solener.2015.04.012](https://doi.org/10.1016/j.solener.2015.04.012).
- Engle, R. F. (1982). “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation”. In: *Econometrica* 50.4, pages 987–1008. ISSN: 0012-9682.
- Erbs, D. G., S. A. Klein, and J. A. Duffie (1982). “Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation”. In: *Solar Energy* 28.4, pages 293–302. ISSN: 0038-092X. DOI: [10.1016/0038-092X\(82\)90302-4](https://doi.org/10.1016/0038-092X(82)90302-4).
- Ferrari, S. and F. Cribari-Neto (2004). “Beta Regression for Modelling Rates and Proportions”. In: *J. Appl. Stat.* 31.7, pages 799–815. ISSN: 0266-4763. DOI: [10.1080/0266476042000214501](https://doi.org/10.1080/0266476042000214501).
- Figuerola-Zúñiga, J. I., R. B. Arellano-Valle, and S. L. P. Ferrari (2013). “Mixed beta regression: A Bayesian perspective”. In: *Comput. Stat. Data Anal.* 61. Supplement C, pages 137–147. ISSN: 0167-9473. DOI: [10.1016/j.csda.2012.12.002](https://doi.org/10.1016/j.csda.2012.12.002).
- Fruhwirth-Schnatter, S. (1995). “Bayesian Model Discrimination and Bayes Factors for Linear Gaussian State Space Models”. In: *J. R. Stat. Soc. Series B Stat. Methodol.* 57.1, pages 237–246. ISSN: 1369-7412.
- (2004). “Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques”. In: *Econom. J.* 7.1, pages 143–167. ISSN: 1368-4221. DOI: [10.1111/j.1368-4221.2004.00125.x](https://doi.org/10.1111/j.1368-4221.2004.00125.x).
- Frühwirth-Schnatter, S. (2001). “Markov chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models”. In: *J. Am. Stat. Assoc.* 96.453, pages 194–209. ISSN: 0162-1459. DOI: [10.1198/016214501750333063](https://doi.org/10.1198/016214501750333063).

- Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. New York: Springer-Verlag New York, pages 1–492. ISBN: 9780387329093.
- Frühwirth-Schnatter, S. and R. Frühwirth (2010). “Data Augmentation and MCMC for Binary and Multinomial Logit Models”. In: *Statistical Modelling and Regression Structures: Festschrift in Honour of Ludwig Fahrmeir*. Edited by T. Kneib and G. Tutz. Heidelberg: Physica-Verlag, pages 111–132. ISBN: 9783790824124. DOI: [10.1007/978-3-7908-2413-1](https://doi.org/10.1007/978-3-7908-2413-1).
- Fu, C.-L. and H.-Y. Cheng (2013). “Predicting solar irradiance with all-sky image features via regression”. en. In: *Solar Energy* 97, pages 537–550. ISSN: 0038-092X. DOI: [10.1016/j.solener.2013.09.016](https://doi.org/10.1016/j.solener.2013.09.016).
- Fujita, T. T. (1981). “Tornadoes and downbursts in the context of generalized planetary scales”. In: *J. Atmos. Sci.* 38.8, pages 1511–1534. ISSN: 0022-4928. DOI: [10.1175/1520-0469\(1981\)038<1511:TADITC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1981)038<1511:TADITC>2.0.CO;2).
- Gates, D. M. (1968). “Transpiration and Leaf Temperature”. In: *Annu. Rev. Plant. Physiol.* 19.1, pages 211–238.
- Gelfand, A. E. and D. K. Dey (1994). “Bayesian Model Choice : Asymptotics and Exact Calculations”. In: *J. R. Stat. Soc. Series B Stat. Methodol.* 56.3, pages 501–514. ISSN: 1369-7412.
- Gelfand, A. and A. F. M. Smith (1990). “Sampling-Based Approaches to Calculating Marginal Densities”. In: *J. Am. Stat. Assoc.* 85.410, pages 398–409. ISSN: 0162-1459. DOI: [10.2307/2289776](https://doi.org/10.2307/2289776).
- Gelman, A., J. Hwang, and A. Vehtari (2014). “Understanding predictive information criteria for Bayesian models”. In: *Statistics and Computing* 24.6, pages 997–1016. ISSN: 15731375. DOI: [10.1007/s11222-013-9416-2](https://doi.org/10.1007/s11222-013-9416-2). arXiv: [1307.5928](https://arxiv.org/abs/1307.5928).
- Gelman, A. and D. B. Rubin (1992). “Inference from Iterative Simulation Using Multiple Simulations”. In: *Stat. Sci.* 7.4, pages 457–511. ISSN: 0883-4237. DOI: [10.1214/ss/1177011136](https://doi.org/10.1214/ss/1177011136).
- Gelman, A. et al. (2013). *Bayesian Data Analysis*. Third. CRC Press. ISBN: 9781439840955.
- Geman, S. and D. Geman (1984). “Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6.6, pages 721–741. DOI: [10.1109/TPAMI.1984.4767596](https://doi.org/10.1109/TPAMI.1984.4767596).
- Glasbey, C. A. (2001). “Non-linear autoregressive time series with multivariate Gaussian mixtures as marginal distributions”. en. In: *Journal of the Royal Statistical Society Series C-Applied Statistics* 50, pages 143–154. ISSN: 0035-9254. DOI: [10.1111/1467-9876.00225](https://doi.org/10.1111/1467-9876.00225).
- Gneiting, T., Fadoua Balabdaoui, and A. E. Raftery (2007). “Probabilistic Forecasts, Calibration and Sharpness”. In: *J. R. Stat. Soc. Series B Stat. Methodol.* 69.2, pages 243–268. ISSN: 1369-7412, 1467-9868.

- Gneiting, T. and A. E. Raftery (2007). “Strictly Proper Scoring Rules, Prediction, and Estimation”. In: *J. Am. Stat. Assoc.* 102.477, pages 359–378. ISSN: 0162-1459.
- Gneiting, T. et al. (2008). “Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds”. en. In: *Test* 17.2, pages 211–234. ISSN: 1133-0686, 1863-8260. DOI: [10.1007/s11749-008-0122-x](https://doi.org/10.1007/s11749-008-0122-x).
- Goh, T. N. and K. J. Tan (1977). “Stochastic modeling and forecasting of solar radiation data”. In: *Solar Energy* 19.6, pages 755–757. ISSN: 0038-092X. DOI: [10.1016/0038-092X\(77\)90041-X](https://doi.org/10.1016/0038-092X(77)90041-X).
- Graham, V. A. and K. G. T. Hollands (1990). “A method to generate synthetic hourly solar radiation globally”. In: *Solar Energy* 44.6, pages 333–341. ISSN: 0038-092X. DOI: [10.1016/0038-092X\(90\)90137-2](https://doi.org/10.1016/0038-092X(90)90137-2).
- Graham, V. A., K. G. T. Hollands, and T. E. Unny (1988). “A time series model for Kt with application to global synthetic weather generation”. In: *Solar Energy* 40.2, pages 83–92.
- Gueymard, C. A. (2010). “Progress in direct irradiance modeling and validation”. In: *The Solar 2010 Conference*. Phoenix, AZ. ISBN: 9781617822698.
- (2012a). “Clear-sky irradiance predictions for solar resource mapping and large-scale applications: Improved validation methodology and detailed performance analysis of 18 broadband radiative models”. In: *Solar Energy* 86.8, pages 2145–2169. ISSN: 0038-092X. DOI: [10.1016/j.solener.2011.11.011](https://doi.org/10.1016/j.solener.2011.11.011).
- (2012b). “Temporal variability in direct and global irradiance at various time scales as affected by aerosols”. In: *Solar Energy* 86.12, pages 3544–3553. ISSN: 0038-092X. DOI: [10.1016/j.solener.2012.01.013](https://doi.org/10.1016/j.solener.2012.01.013).
- Gueymard, C. A. and S. M. Wilcox (2011). “Assessment of spatial and temporal variability in the US solar resource from radiometric measurements and predictions from models using ground-based or satellite data”. In: *Solar Energy* 85.5, pages 1068–1084. ISSN: 0038-092X. DOI: [10.1016/j.solener.2011.02.030](https://doi.org/10.1016/j.solener.2011.02.030).
- Gutierrez-Corea, F.-V. et al. (2016). “Forecasting short-term solar irradiance based on artificial neural networks and data from neighboring meteorological stations”. en. In: *Solar Energy* 134, pages 119–131. ISSN: 0038-092X. DOI: [10.1016/j.solener.2016.04.020](https://doi.org/10.1016/j.solener.2016.04.020).
- Hall, D. K. and G. A. Riggs (2016). *MODIS/Terra Snow Cover 5-Min L2 Swath 500m, Version 6*. Accessed: 2017-3-3. Boulder, CO, USA. DOI: [10.5067/MODIS/MOD10_L2.006](https://doi.org/10.5067/MODIS/MOD10_L2.006).
- Hamilton, J. D. (1989). “A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle”. In: *Econometrica* 57.2, pages 357–384. ISSN: 0012-9682, 1468-0262.
- Hammer, A. et al. (1999). “Short-term forecasting of solar radiation: a statistical approach using satellite data”. In: *Solar Energy* 67.1-3, pages 139–150. ISSN: 0038-092X. DOI: [10.1016/S0038-092X\(00\)00038-4](https://doi.org/10.1016/S0038-092X(00)00038-4).

- Hastings, W. K. (1970). "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". In: *Biometrika* 57.1, pages 97–109. ISSN: 0006-3444. DOI: [10.2307/2334940](#).
- Haupt, S. E., A. Pasini, and C. Marzban, editors (2009). *Artificial Intelligence Methods in the Environmental Sciences*. 1st edition. Springer Netherlands. ISBN: 978-1-4020-9117-9. DOI: [10.1007/978-1-4020-9119-3](#).
- Hocaoğlu, F. O. (2011). "Stochastic approach for daily solar radiation modeling". In: *Solar Energy* 85.2, pages 278–287. ISSN: 0038-092X. DOI: [10.1016/j.solener.2010.12.003](#).
- Hollands, K. G. T. and R. G. Huget (1983). "A probability density function for the clearness index, with applications". In: *Solar Energy* 30.3, pages 195–209. ISSN: 0038-092X. DOI: [10.1016/0038-092X\(83\)90149-4](#).
- Hontoria, L., J. Aguilera, and P. Zufria (2002). "Generation of hourly irradiation synthetic series using the neural network multilayer perceptron". In: *Solar Energy* 72.5, pages 441–446. ISSN: 0038-092X. DOI: [10.1016/S0038-092X\(02\)00010-5](#).
- Imhoff, K. (2015). *Maximum Wind Speed*. private communication. Pennsylvania State Climatologist.
- Iverson, E. B. et al. (2014). "Probabilistic forecasts of solar irradiance using stochastic differential equations". en. In: *Environmetrics* 25, pages 152–164. ISSN: 1099-095X, 1099-095X. DOI: [10.1002/10.1002/env.2267](#).
- Janczura, J. and R. Weron (2012). "Efficient estimation of Markov regime-switching models: An application to electricity spot prices". In: *AStA Adv. Stat. Anal.* 96.3, pages 385–407. ISSN: 1863-8171. DOI: [10.1007/s10182-011-0181-2](#).
- Juang, B. H. and L. R. Rabiner (1991). "Hidden Markov Models for Speech Recognition". In: *Technometrics* 33.3, pages 251–272. ISSN: 0040-1706. DOI: [10.2307/1268779](#).
- Kalman, R. E. (1960). "A new approach to linear filtering and prediction problems". In: *Journal of Basic Engineering* 82.1, pages 35–45.
- Kalman, R. E. and R. S. Bucy (1961). "New results in linear filtering and prediction theory". In: *Journal of Basic Engineering* 83.1, pages 95–108.
- Kaufmann, S. (2015). "K-state switching models with time-varying transition distributions — Does loan growth signal stronger effects of variables on inflation?". In: *J. Econom.* 187.1, pages 82–94. ISSN: 0304-4076. DOI: [10.1016/j.jeconom.2015.02.001](#).
- Kim, C.-J. (1994). "Dynamic linear models with Markov-switching". In: *J. Econom.* 60.1, pages 1–22. ISSN: 0304-4076. DOI: [10.1016/0304-4076\(94\)90036-1](#).
- Klein, S. A., W. A. Beckman, and J. A. Duffie (1976). "A design procedure for solar heating systems". In: *Solar Energy* 18, pages 113–127.
- Knight, K. M., S. A. Klein, and J. A. Duffie (1991). "A methodology for the synthesis of hourly weather data". In: *Solar Energy* 46.2, pages 109–120. ISSN: 0038-092X. DOI: [10.1016/0038-092x\(91\)90023-p](#).

- Krüger, F. et al. (2016). “Probabilistic Forecasting and Comparative Model Assessment Based on Markov Chain Monte Carlo Output”. In: arXiv: [1608.06802 \[stat.ME\]](#).
- Laslett, D., C. Creagh, and P. Jennings (2014). “A method for generating synthetic hourly solar radiation data for any location in the south west of Western Australia, in a world wide web page”. In: *Renewable Energy* 68.Supplement C, pages 87–102. ISSN: 0960-1481. DOI: [10.1016/j.renene.2014.01.015](#).
- Lave, M., J. Kleissl, and J. Stein (2013). “Quantifying and Simulating Solar-Plant Variability Using Irradiance Data”. In: *Solar Energy Forecasting and Resource Assessment*. Edited by J. Kleissl. San Diego, CA: Elsevier Academic Press. Chapter 7, pages 150–169. ISBN: 9780123971777.
- Linares-Rodriguez, A. et al. (2013). “An artificial neural network ensemble model for estimating global solar radiation from Meteosat satellite images”. In: *Energy* 61.Supplement C, pages 636–645. ISSN: 0360-5442. DOI: [10.1016/j.energy.2013.09.008](#).
- Little, R. J. and D. B. Rubin (2002). *Statistical Analysis with Missing Data*. 2nd edition. John Wiley & Sons, Inc. ISBN: 9780471183860.
- Liu, B. Y. H. and R. C. Jordan (1960). “The interrelationship and characteristic distribution of direct, diffuse and total solar radiation”. In: *Solar Energy* 4.3, pages 1–19. ISSN: 0038-092X. DOI: [10.1016/0038-092X\(60\)90062-1](#).
- Liu, S. et al. (2009). “Pushback: A hidden Markov model based scheme for energy efficient data transmission in sensor networks”. In: *Ad Hoc Networks* 7.2008, pages 973–986. ISSN: 1570-8705. DOI: [10.1016/j.adhoc.2008.09.001](#).
- Louche, A. et al. (1991). “Correlations for direct normal and global horizontal irradiation on a French Mediterranean site”. In: *Solar Energy* 46.4, pages 261–266. ISSN: 0038-092X. DOI: [10.1016/0038-092X\(91\)90072-5](#).
- Marquez, R. and C. F. M. Coimbra (2011). “Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database”. In: *Solar Energy* 85, pages 746–756. ISSN: 0038-092X. DOI: [10.1016/j.solener.2011.01.007](#).
- (2013). “Intra-hour DNI forecasting based on cloud tracking image analysis”. In: *Solar Energy* 91, pages 327–336. ISSN: 0038-092X. DOI: [10.1016/j.solener.2012.09.018](#).
- Marquez, R., H. T. C. Pedro, and C. F. M. Coimbra (2013). “Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs”. en. In: *Solar Energy* 92, pages 176–188. ISSN: 0038-092X. DOI: [10.1016/j.solener.2013.02.023](#).
- Maxwell, E. L. (1987). *A quasi-physical model for converting hourly global horizontal to direct normal insolation*. Technical report SERI/TR-215-3087. Solar Energy Research Institute.

- McCandless, T. C., S. E. Haupt, and G. S. Young (2015). “A model tree approach to forecasting solar irradiance variability”. en. In: *Solar Energy* 120, pages 514–524. ISSN: 0038-092X. DOI: [10.1016/j.solener.2015.07.020](https://doi.org/10.1016/j.solener.2015.07.020).
- (2016a). “A regime-dependent artificial neural network technique for short-range solar irradiance forecasting”. en. In: *Renewable Energy* 89, pages 351–359. ISSN: 0960-1481. DOI: [10.1016/j.renene.2015.12.030](https://doi.org/10.1016/j.renene.2015.12.030).
- McCandless, T. C. et al. (2016b). “Regime-Dependent Short-Range Solar Irradiance Forecasting”. en. In: *J. Appl. Meteorol. Climatol.* 55.7, pages 1599–1613. ISSN: 1558-8424, 1558-8432. DOI: [10.1175/JAMC-D-15-0354.1](https://doi.org/10.1175/JAMC-D-15-0354.1).
- McCandless, T. C., S. E. Haupt, and G. S. Young (2013). “Short Term Solar Radiation Forecasts Using Weather Regime Dependent Artificial Intelligence Techniques”. In: *12th Conference on Artificial and Computational Intelligence and its Applications to the Environmental Sciences*, pages 1–7.
- McLachlan, G. J. and D. Peel (2000). *Finite mixture models*. English. New York: Wiley. ISBN: 047165406X;9780471654063.
- McMahan, A. C., C. N. Grover, and F. E. Vignola (2013). “Evaluation of Resource Risk in Solar-Project Financing”. In: *Solar Energy Forecasting and Resource Assessment*. Edited by J. Kleissl. San Diego, CA: Elsevier Academic Press. Chapter 4, pages 81–95. ISBN: 9780123971777.
- Mehrotra, S. and D. Papp (2013). “Generating Moment Matching Scenarios Using Optimization Techniques”. In: *SIAM J. Optim.* 23.2, pages 963–999. ISSN: 1052-6234.
- Mellit, A., M. Benghaneim, and S. A. Kalogirou (2006). “An adaptive wavelet-network model for forecasting daily total solar-radiation”. In: *Appl. Energy* 83.7, pages 705–722. ISSN: 0306-2619, 1872-9118. DOI: [10.1016/j.apenergy.2005.06.003](https://doi.org/10.1016/j.apenergy.2005.06.003).
- Mellit, A. and A. M. Pavan (2010). “A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy”. en. In: *Solar Energy* 84.5, pages 807–821. ISSN: 0038-092X. DOI: [10.1016/j.solener.2010.02.006](https://doi.org/10.1016/j.solener.2010.02.006).
- Meng, X. L. and H. W. Wing (1996). “Simulating ratios of normalizing constants via a simple identity: a theoretical exploration”. In: *Stat. Sin.* 6, pages 831–860. ISSN: 1017-0405.
- Metropolis, N. et al. (1953). “Equation of State Calculations by Fast Computing Machines”. In: *J. Chem. Phys.* 21.6, pages 1087–1092. ISSN: 0021-9606, 1089-7690. DOI: [10.1063/1.1699114](https://doi.org/10.1063/1.1699114).
- Miller, S. D., T. F. Lee, and R. L. Fennimore (2005). “Satellite-Based Imagery Techniques for Daytime Cloud / Snow Delineation from MODIS”. In: *Journal of Applied Meteorology* 44.7, pages 987–997. ISSN: 08948763. DOI: [10.1175/JAM2252.1](https://doi.org/10.1175/JAM2252.1).

- Mitchison, G. and R. Durbin (1995). “Tree-based maximal likelihood substitution matrices and hidden Markov models”. en. In: *J. Mol. Evol.* 41.6, pages 1139–1151. ISSN: 0022-2844, 1432-1432. DOI: [10.1007/BF00173195](https://doi.org/10.1007/BF00173195).
- Monjoly, S. et al. (2017). “Hourly forecasting of global solar radiation based on multiscale decomposition methods: A hybrid approach”. en. In: *Energy* 119, pages 288–298. ISSN: 0360-5442, 1873-6785. DOI: [10.1016/j.energy.2016.11.061](https://doi.org/10.1016/j.energy.2016.11.061).
- Moreno-Muñoz, A. et al. (2008). “Short term forecasting of solar radiation”. In: *IEEE International Symposium on Industrial Electronics*, pages 1537–1541. ISSN: 0160-8371. DOI: [10.1109/ISIE.2008.4676880](https://doi.org/10.1109/ISIE.2008.4676880).
- Morf, H. (1998). “The Stochastic two-state solar irradiance model (STSIM)”. In: *Solar Energy* 62.2, pages 101–112. ISSN: 0038-092X. DOI: [10.1016/S0038-092X\(98\)00004-8](https://doi.org/10.1016/S0038-092X(98)00004-8).
- (2011). “The Stochastic Two-State Cloud Cover Model STSCCM”. In: *Solar Energy* 85.5, pages 985–999. ISSN: 0038-092X. DOI: [10.1016/j.solener.2011.02.015](https://doi.org/10.1016/j.solener.2011.02.015).
- (2013). “A stochastic solar irradiance model adjusted on the Ångström–Prescott regression”. In: *Solar Energy* 87. Supplement C, pages 1–21. ISSN: 0038-092X. DOI: [10.1016/j.solener.2012.10.005](https://doi.org/10.1016/j.solener.2012.10.005).
- (2014). “Sunshine and cloud cover prediction based on Markov processes”. In: *Solar Energy* 110, pages 615–626. ISSN: 0038-092X.
- Müller, B. et al. (2014). “Rethinking solar resource assessments in the context of global dimming and brightening”. In: *Solar Energy* 99, pages 272–282. ISSN: 0038-092X. DOI: [10.1016/j.solener.2013.11.013](https://doi.org/10.1016/j.solener.2013.11.013).
- Muneer, T., C. Gueymard, and H. Kambezidis (2004). *Solar Radiation and Daylight Models*. Routledge.
- Muselli, M. et al. (2001). “First order Markov chain model for generating synthetic ‘typical days’ series of global irradiation in order to design photovoltaic stand alone systems”. In: *Energy Convers. Manage.* 42.6, pages 675–687. ISSN: 0196-8904. DOI: [10.1016/S0196-8904\(00\)00090-X](https://doi.org/10.1016/S0196-8904(00)00090-X).
- Mustacchi, C., V. Cena, and M. Rocchi (1979). “Stochastic simulation of hourly global radiation sequences”. In: *Solar Energy* 23.1, pages 47–51. ISSN: 0038-092X. DOI: [10.1016/0038-092X\(79\)90042-2](https://doi.org/10.1016/0038-092X(79)90042-2).
- Ngoko, B. O., H. Sugihara, and T. Funaki (2014). “Synthetic generation of high temporal resolution solar radiation data using Markov models”. In: *Solar Energy* 103, pages 160–170. ISSN: 0038-092X. DOI: [10.1016/j.solener.2014.02.026](https://doi.org/10.1016/j.solener.2014.02.026).
- Pedro, H. T. C. and C. F. M. Coimbra (2015). “Short-term irradiance forecastability for various solar micro-climates”. en. In: *Solar Energy* 122, pages 587–602. ISSN: 0038-092X. DOI: [10.1016/j.solener.2015.09.031](https://doi.org/10.1016/j.solener.2015.09.031).
- Perez, R. R. et al. (1992). “Dynamic global-to-direct irradiance conversion models”. In: *ASHRAE Transactions* 98.1, pages 354–369.

- Perez, R. et al. (2002). “A new operational model for satellite-derived irradiances: Description and validation”. In: *Solar Energy* 73.5, pages 307–317. ISSN: 0038-092X. DOI: [10.1016/S0038-092X\(02\)00122-6](https://doi.org/10.1016/S0038-092X(02)00122-6).
- Perez, R. et al. (2010). “Validation of short and medium term operational solar radiation forecasts in the US”. In: *Solar Energy* 84.12, pages 2161–2172. ISSN: 0038-092X. DOI: [10.1016/j.solener.2010.08.014](https://doi.org/10.1016/j.solener.2010.08.014).
- Perez, R. et al. (1986). “An anisotropic hourly diffuse radiation model for sloping surfaces: Description, performance validation, site dependency evaluation”. In: *Solar Energy* 36.6, pages 481–497. ISSN: 0038-092X. DOI: [10.1016/0038-092X\(86\)90013-7](https://doi.org/10.1016/0038-092X(86)90013-7).
- Perlin, J. (2013). *Let It Shine: The 6,000-Year Story of Solar Energy*. en. New World Library. ISBN: 9781608681334.
- Pinson, P. and H. Madsen (2012). “Adaptive modelling and forecasting of offshore wind power fluctuations with Markov-switching autoregressive models”. In: *J. Forecast.* 31.4, pages 281–313. ISSN: 0277-6693, 1099-131X. DOI: [10.1002/for.1194](https://doi.org/10.1002/for.1194).
- Pinto, C. and L. Spezia (2016). “Markov switching autoregressive models for interpreting vertical movement data with application to an endangered marine apex predator”. In: *Methods Ecol. Evol.* 7.4, pages 407–417. ISSN: 2041-210X. DOI: [10.1111/2041-210X.12494](https://doi.org/10.1111/2041-210X.12494).
- Polo, J. et al. (2011). “A simple approach to the synthetic generation of solar irradiance time series with high temporal resolution”. In: *Solar Energy* 85.5, pages 1164–1170. ISSN: 0038-092X. DOI: [10.1016/j.solener.2011.03.011](https://doi.org/10.1016/j.solener.2011.03.011).
- Prema, V. and K. U. Rao (2015). “Development of statistical time series models for solar power prediction”. en. In: *Renewable Energy* 83, pages 100–109. ISSN: 0960-1481. DOI: [10.1016/j.renene.2015.03.038](https://doi.org/10.1016/j.renene.2015.03.038).
- Reikard, G. (2009). “Predicting solar radiation at high resolutions: A comparison of time series forecasts”. In: *Solar Energy* 83.3, pages 342–349. ISSN: 0038-092X. DOI: [10.1016/j.solener.2008.08.007](https://doi.org/10.1016/j.solener.2008.08.007).
- Reikard, G., S. E. Haupt, and T. Jensen (2017). “Forecasting ground-level irradiance over short horizons: Time series, meteorological, and time-varying parameter models”. In: *Renewable Energy* 112, pages 474–485. ISSN: 0960-1481. DOI: [10.1016/j.renene.2017.05.019](https://doi.org/10.1016/j.renene.2017.05.019).
- Richardson, S. and P. J. Green (1997). “On Bayesian Analysis of Mixtures with an Unknown Number of Components”. In: *Journal of the Royal Statistical Society Series B* 59.4, pages 731–792. DOI: [10.2307/2346101](https://doi.org/10.2307/2346101).
- Robert, C. P. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*. 2nd. New York: Springer. ISBN: 9780387715988.
- (2014). “The Metropolis–Hastings Algorithm”. In: *Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons, Ltd. ISBN: 9781118445112. DOI: [10.1002/9781118445112.stat07834](https://doi.org/10.1002/9781118445112.stat07834).

- Robert, C. P. and G. Casella (2004). *Monte Carlo Statistical Methods*. Second. New York: Springer-Verlag. ISBN: 9780387212395.
- Rocha, A. V. and F. Cribari-Neto (2009). “Beta autoregressive moving average models”. In: *Test* 18.3, pages 529–545. ISSN: 1133-0686. DOI: [10.1007/s11749-008-0112-z](#).
- Rothenberg, T. J. (1971). “Identification in Parametric Models”. In: *Econometrica* 39.3, pages 577–591. ISSN: 0012-9682, 1468-0262. DOI: [10.2307/1913267](#).
- Salvucci, G. D., J. A. Saleem, and R. Kaufmann (2002). “Investigating soil moisture feedbacks on precipitation with tests of Granger causality”. In: *Adv. Water Resour.* 25, pages 1305–1312. ISSN: 0309-1708. DOI: [10.1016/S0309-1708\(02\)00057-X](#).
- Schwarz, G. (1978). “Estimating the Dimension of a Model”. In: *The Annals of Statistics* 6.2, pages 461–464. DOI: [10.1525/irqr.2013.6.2.210.This](#).
- Sengupta, M. et al. (2015). *Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications*. Technical report. National Renewable Energy Laboratory. DOI: [10.18777/ieashc-task46-2015-0001](#).
- Shakya, A. et al. (2017). “Using Markov Switching Model for Solar Irradiance Forecasting in Remote Microgrids”. en. In: *Ieee Transactions on Sustainable Energy* PP.99, pages 895–905. ISSN: 1949-3029. DOI: [10.1109/TSTE.2016.2629974](#).
- Silva, C. V. A. et al. (2015). “Probabilistic Models for One-Day Ahead Solar Irradiance Forecasting in Renewable Energy Applications”. en. In: *2015 Ieee 14th International Conference on Machine Learning and Applications (icmla)*. Sara Burgerhartstraat 25, Po Box 211, 1000 Ae Amsterdam, Netherlands: Elsevier Science Bv, pages 1163–1168. ISBN: 9781509002870. DOI: [10.1109/ICMLA.2015.137](#).
- Skartveit, A. and J. A. Olseth (1992). “The probability density and autocorrelation of short-term global and beam irradiance”. In: *Solar Energy* 49.6, pages 477–487. ISSN: 0038-092X. DOI: [10.1016/0038-092X\(92\)90155-4](#).
- Skartveit, A. and J. A. Olseth (1987). “A model for the diffuse fraction of hourly global radiation”. In: *Solar Energy* 38.4, pages 271–274. ISSN: 0038-092X. DOI: [10.1016/0038-092X\(87\)90049-1](#).
- Skartveit, A., J. A. Olseth, and M. E. Tuft (1998). “An hourly diffuse fraction model with correction for variability and surface albedo”. In: *Solar Energy* 63.3, pages 173–183. ISSN: 0038-092X. DOI: [10.1016/S0038-092X\(98\)00067-X](#).
- Soubdhan, T., J. Ndong, and H. Ould-Baba (2014). “A robust forecasting framework based on the Kalman filtering approach with a twofold parameter tuning procedure: Application to solar and photovoltaic prediction”. en. In: *Renewable Energy* 131, pages 475–484. ISSN: 0960-1481. DOI: [10.1016/j.solener.2016.02.036](#).

- Souza, D. F. and F. A. S. Moura (2016). “Multivariate Beta Regression with Application in Small Area Estimation”. In: *J. Off. Stat.* 32.3, page 1609. ISSN: 0282-423X, 2001-7367. DOI: [10.1515/jos-2016-0038](https://doi.org/10.1515/jos-2016-0038).
- Spiegelhalter, D. J. et al. (2002). “Bayesian measures of model complexity and fit”. In: *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 64.4, pages 583–616. ISSN: 13697412. DOI: [10.1111/1467-9868.00353](https://doi.org/10.1111/1467-9868.00353). arXiv: [arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).
- Srikrishnan, V. (2015). “Estimation of direct normal irradiance with multi-pyranometer arrays and artificial neural networks”. Master’s thesis. Pennsylvania State University.
- Srikrishnan, V., G. S. Young, and J. R. S. Brownson (2017). “Skill and Skill Prediction of Cloud-Track Advection-Only Forecasting Under a Cumulus-Dominated Regime”. In: *J. Appl. Meteorol. Climatol.* 56.3, JAMC-D-16-0224.1. ISSN: 1558-8424. DOI: [10.1175/JAMC-D-16-0224.1](https://doi.org/10.1175/JAMC-D-16-0224.1).
- Stuart, R. W. and K. G. T. Hollands (1988). “A probability density function for the beam transmittance”. In: *Solar Energy* 40.5, pages 463–467. ISSN: 0038-092X. DOI: [10.1016/0038-092X\(88\)90101-6](https://doi.org/10.1016/0038-092X(88)90101-6).
- Sun, H. et al. (2015). “Empirical investigation on modeling solar radiation series with ARMA-GARCH models”. en. In: *Energy Convers. Manage.* 92, pages 385–395. ISSN: 0196-8904, 1879-2227. DOI: [10.1016/j.enconman.2014.12.072](https://doi.org/10.1016/j.enconman.2014.12.072).
- Tanner, M. A. and W. H. Wong (1987). “The Calculation of Posterior Distributions by Data Augmentation”. In: *Journal of the American Statistical Association* 82.398, pages 528–540.
- Taylor, G. I. (1938). “The spectrum of turbulence”. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 164.919, pages 476–490. DOI: [10.1098/rspa.1938.0032](https://doi.org/10.1098/rspa.1938.0032).
- Torregrossa, D., J.-Y. Le Boudec, and M. Paolone (2016). “Model-free computation of ultra-short-term prediction intervals of solar irradiance”. en. In: *Solar Energy* 124, pages 57–67. ISSN: 0038-092X. DOI: [10.1016/j.solener.2015.11.017](https://doi.org/10.1016/j.solener.2015.11.017).
- Torres, J. L. et al. (2010). “Comparative study of various models in estimating hourly diffuse solar irradiance”. In: *Renewable Energy* 35.6, pages 1325–1332. ISSN: 0960-1481. DOI: [10.1016/j.renene.2009.11.025](https://doi.org/10.1016/j.renene.2009.11.025).
- Tovar-Pescador, J. (2008). “Modelling the Statistical Properties of Solar Radiation and Proposal of a Technique Based on Boltzmann Statistics”. In: *Modeling Solar Radiation at the Earth’s Surface: Recent Advances*. Edited by V. Badescu. Berlin, Heidelberg: Springer Berlin Heidelberg, pages 55–91. ISBN: 9783540774556. DOI: [10.1007/978-3-540-77455-6_3](https://doi.org/10.1007/978-3-540-77455-6_3).
- Trombe, P.-J., P. Pinson, and H. Madsen (2012). “A General Probabilistic Forecasting Framework for Offshore Wind Power Fluctuations”. en. In: *Energies* 5.3, pages 621–657. DOI: [10.3390/en5030621](https://doi.org/10.3390/en5030621).

- Vignola, F. et al. (2012). “Building a bankable solar radiation dataset”. In: *Solar Energy* 86.8, pages 2218–2229. ISSN: 0038-092X. DOI: [10.1016/j.solener.2012.05.013](https://doi.org/10.1016/j.solener.2012.05.013).
- Vindel, J. M. and J. Polo (2014). “Markov processes and Zipf’s law in daily solar irradiation at earth’s surface”. In: *J. Atmos. Sol. Terr. Phys.* 107. Supplement C, pages 42–47. ISSN: 1364-6826. DOI: [10.1016/j.jastp.2013.10.017](https://doi.org/10.1016/j.jastp.2013.10.017).
- Von Meier, A. (2011). “Integration of renewable generation in California: Coordination challenges in time and space”. In: *Proceeding of the International Conference on Electrical Power Quality and Utilisation, EPQU*. Lisbon, Portugal: IEEE, pages 768–773. ISBN: 9781467303798. DOI: [10.1109/EPQU.2011.6128888](https://doi.org/10.1109/EPQU.2011.6128888).
- Wang, J. et al. (2009). “Impact of Wind Power Forecasting on Unit Commitment and Dispatch”. In: *Univ. Polytechnica Bucharest Sci. Bull. Ser. C: Electr. Eng.* ISSN: 1454-234X.
- Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics. Springer.
- Watanabe, S. (2010). “Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory”. In: *Journal of Machine Learning Research* 11, pages 3571–3594. ISSN: 1532-4435. arXiv: [1004.2316](https://arxiv.org/abs/1004.2316). URL: <http://arxiv.org/abs/1004.2316>.
- West, M. and J. Harrison (1997). *Bayesian Forecasting and Dynamic Models*. Second. Springer Series in Statistics. Springer-Verlag New York.
- Westerweel, J. and F. Scarano (2005). “Universal outlier detection for PIV data”. In: *Exp. Fluids* 39.6, pages 1096–1100. ISSN: 0723-4864. DOI: [10.1007/s00348-005-0016-6](https://doi.org/10.1007/s00348-005-0016-6).
- Witmer, L. (2014). “The Impact of Solar Radiation on the Heating and Cooling of Buildings”. PhD thesis. The Pennsylvania State University.
- Wong, T. E. et al. (2017). “Neglecting Model Structural Uncertainty Underestimates Upper Tails of Flood Risk”. In: arXiv: [1709.08776](https://arxiv.org/abs/1709.08776) [stat.AP].
- Yang, D., P. Jirutitijaroen, and W. M. Walsh (2012). “Hourly solar irradiance time series forecasting using cloud cover index”. In: *Solar Energy* 86.12, pages 3531–3543. ISSN: 0038-092X. DOI: [10.1016/j.solener.2012.07.029](https://doi.org/10.1016/j.solener.2012.07.029).
- Yang, D. et al. (2015). “Forecasting of global horizontal irradiance by exponential smoothing, using decompositions”. In: *Energy* 81, pages 111–119. ISSN: 0360-5442. DOI: [10.1016/j.energy.2014.11.082](https://doi.org/10.1016/j.energy.2014.11.082).
- Yang, H. et al. (2014). “Solar irradiance forecasting using a ground-based sky imager developed at UC San Diego”. In: *Solar Energy* 103, pages 502–524. ISSN: 0038-092X. DOI: [10.1016/j.solener.2014.02.044](https://doi.org/10.1016/j.solener.2014.02.044).
- Yang, L., B. K. Widjaja, and R. Prasad (1995). “Application of hidden Markov models for signature verification”. In: *Pattern Recognit.* 28.2, pages 161–170. ISSN: 0031-3203. DOI: [10.1016/0031-3203\(94\)00092-Z](https://doi.org/10.1016/0031-3203(94)00092-Z).

Zheng, Q. P., J. Wang, and A. L. Liu (2015). “Stochastic Optimization for Unit Commitment – A Review”. In: *IEEE Trans. Power Syst.* 30.4, pages 1913–1924. ISSN: 0885-8950. DOI: [10.1109/TPWRS.2014.2355204](https://doi.org/10.1109/TPWRS.2014.2355204).

Vita

Vivek Srikrishnan

Education:

- Ph.D., Energy and Mineral Engineering, The Pennsylvania State University, May 2018 (Expected)
- M.S., Energy and Mineral Engineering, The Pennsylvania State University, August 2015
- B.S., Mathematics, University of Illinois at Urbana-Champaign, August 2004
- B.A., Philosophy, University of Illinois at Urbana-Champaign, August 2004

Publications:

1. J. Morris, **V. Srikrishnan**, M. Webster, and J. Reilly (2018, pre-press), “Hedging strategies: electricity investment decisions under policy uncertainty,” *The Energy Journal* 39 (1). DOI: [10.5547/01956574.39.1.jmor](https://doi.org/10.5547/01956574.39.1.jmor).
2. P. .C. Oddo, B. S. Lee, G. G. Garner, **V. Srikrishnan**, P. .M. Reed, C. E .Forest, and K. Keller (2017, pre-press), “Deep uncertainties in sea-level rise and storm surge projections: Implications for coastal flood risk management,” *Risk Analysis*. DOI: [10.1111/risa.12888](https://doi.org/10.1111/risa.12888).
3. T. E. Wong, **V. Srikrishnan**, D. Hadka, and K. Keller (2017), “A multi-objective decision-making approach to the journal submission problem,” *PLOS ONE*. DOI: [10.1371/journal.pone.0178874](https://doi.org/10.1371/journal.pone.0178874).
4. **V. Srikrishnan**, G. S. Young, and J. R. S. Brownson (2017), “Skill and skill prediction of cloud-track advection-only forecasting under a cumulus-dominated regime,” *Journal of Applied Meteorology and Climatology* 56 (3), 651–665. DOI: [10.1175/JAMC-D-16-0224.1](https://doi.org/10.1175/JAMC-D-16-0224.1).
5. **V. Srikrishnan**, G. S. Young, L. Witmer, and J. R. S. Brownson (2015), “Using multi-pyranometer arrays and neural networks to estimate direct normal irradiance,” *Solar Energy* 119, 531–542. DOI: [10.1016/j.solener.2015.06.004](https://doi.org/10.1016/j.solener.2015.06.004).