Inference on Climate Indices Using Bayesian Variable Selection in Quantile Regression

DILANI KAVEENDRI DAMBURE HEWAGE

B.Sc (Hons)

Supervisors: Prof. Sally Cripps and Prof. Matthew Cleary

A thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

This research reported in this thesis was supported by the award of an Australian Research Council (ARC) scholarship through the ARC Training Centre in Data Analytics for Resources and Environments (DARE) – Project ID: IC190100031

School of Aerospace, Mechanical and Mechatronic Engineering
Faculty of Engineering
The University of Sydney
Australia

31 August 2025

Declaration

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes. I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Dilani Kaveendri Dambure Hewage
Date: 2025/03/28

Abstract

Rainfall variability significantly affects ecosystems, agriculture, and water management in eastern Australia. Influenced by global climate phenomena, these fluctuations challenge management practices at different locations and times of the year. Understanding these drivers is essential to improve resource management and mitigate the impacts of extreme weather events such as droughts and floods.

Advancements in physical and statistical climate models have enhanced our understanding of global climate phenomena in relation to daily rainfall extremes. However, these models often lack a probabilistic framework that can assess the relationship between climate drivers and the full distribution of monthly rainfall. A key challenge is the inadequate modeling of uncertainty and variability inherent in climate systems. Bayesian approaches with variable selection are valuable in this context, as they incorporate prior knowledge and provide robust parameter estimation. Nonetheless, applications of such models that comprehensively assess the relationship between climate indices and the full distribution of monthly rainfall—including extremes—remain limited. This gap highlights the need for further development to improve predictive capabilities for monthly rainfall patterns under varying climate conditions.

This thesis introduces a novel approach that employs Bayesian variable selection within a spatial quantile regression framework to examine the relationship between global climate indices and monthly rainfall distribution in New South Wales (NSW), Australia. By analyzing different quantiles of rainfall, this approach aims to offer a spatially varying inference of how these climate indices impact the entire spectrum of rainfall. This approach utilizes a hierarchical Bayesian quantile regression model to address distinct modeling requirements and complexities at each hierarchical level.

The second chapter introduces Bayesian variable selection in a quantile regression framework to infer the climate indices relationship to monthly rainfall distribution where we assume the independence among the locations considered for this study, as this is a common practice in climate science and as a basis to build the model that explicitly encodes the spatial dependence. The hierarchical methodology initially models the distribution of rainfall employing the asymmetric Laplace distribution (ALD) through a location-scale mixture (Kotz et al., 2001). Subsequently, variable selection is performed by modeling the regression coefficients as a two-component mixture using spike and slab prior (George and McCulloch, 1993).

In the third chapter, we extend the model in Chapter 2 to include spatial dependencies, where locations are analyzed jointly using one climate covariate, specifically the Southern Oscillation Index (SOI). We extend the hierarchical model to incorporate spatial dependence through a Gaussian process ($\mathcal{GP}$) prior on the regression coefficients. The first and second levels of the hierarchy remain unchanged. At the third level, spatial dependence is introduced via two mechanisms. First, the probability that the regression coefficient is non-zero is modeled as a probit regression where the argument of the link function varies spatially according to a Gaussian process prior. Second, the prior on regression coefficients, conditional on those coefficients being non-zero, is modeled as a normal distribution, where the mean of the distribution also varies spatially according to a Gaussian process prior (Wahba, 1990; Wood, 2013). To estimate the probability that the regression coefficient is non-zero, we follow the latent variable approach by Albert and Chib, 1993. An alternative approach to using the probit link is the use of the logit link function and using the data augmentation method by Polson et al., 2012.

The fourth chapter extends the model in Chapter 3 to include multiple global climate indices that incorporate spatial dependence. At the third level of the hierarchy, spatial dependence is incorporated via modeling joint inclusion probability using a multinomial logistic regression where the argument of the link function varies spatially according to a $\mathcal{GP}$

prior. A significant benefit of this method is that it provides spatially varying inferences on the joint impact of the climate indices on the full distribution of monthly rainfall.

Finally, this thesis concludes with a discussion of the key takeaways and potential directions for future research.

Acknowledgements

Firstly, I extend my deepest gratitude to my supervisors, Prof. Sally Cripps, Dr. Nandini Ramesh, and Prof. Matthew Cleary for their invaluable support and guidance throughout my PhD journey. Sally, your patience and mentorship have profoundly shaped my development as a Bayesian statistician. Thank you for your immense support and for continuously inspiring me. It is a privilege to be a student of such a well-renowned Bayesian statistician, both in Australia and internationally. I have learned so much from you, and I am truly grateful for everything you have done for me. Nandini, co-supervisor of my PhD in all but name, your encouragement and support have been pivotal to my growth. Thank you for believing in me—your kindness, especially during stressful times, has meant so much. Matt, your kindness and support have significantly eased my administrative challenges, and I greatly appreciate your steady presence and assistance.

I was privileged to be part of the DARE Center, a hub of exceptionally talented PhD students and collaborators. I am especially grateful to Prof. Willem Vervoort, the director of the center, for his unwavering support and dedication to all students. I also extend my thanks to Carolyn and Kelsey, whose administrative assistance was indispensable. I also would like to acknowledge the former administrative team, Jessica Cornock, Minh Ai Nguyen, Janice Zhang, and Dr. Solmaz Jahangir, for their support and assistance.

My sincere appreciation goes to Dr. Travis Stenborg, who was instrumental in helping me acquire my dataset, and to Dr. Joshua Simmons for his invaluable assistance with parallelization programming. I am also grateful to Assoc. Prof. Edward Cripps, Deputy Director of DARE and coauthor of our paper to be submitted to Environmetrics. His rigorous feedback and insightful contributions have significantly elevated the quality of our work, and I am immensely thankful for his expertise and guidance during the preparation of our

manuscript. I am also grateful to Dr. Michael Bertolacci for his valuable guidance and support in addressing questions related to his work, which greatly aided my understanding and strengthened my research.

A special mention is deserved for my friend and colleague, Dr. Rajitha Athukorala, a former DARE student, who was a pillar of support during the challenging times of my PhD. Rajitha, your advice and friendship have been invaluable throughout this journey.

I would also like to express my gratitude to the developers and contributors of the R packages as their work has greatly facilitated data analysis, modeling, and visualization. During the preparation of this thesis, ChatGPT was used for the purposes of text enhancement, including sentence structure, paraphrasing, etc. Where any text was modified by generative AI, the author then reviewed the resulting content for any errors, inaccuracies or biases, and modified it as required. The author takes full responsibility for the submitted thesis and ensures the work is their own and has used generative AI within the parameters of use, see University of Sydney generative AI guide for researchers.

None of this would have been possible without the love and support of my husband, Himanshu. You have been my stronghold, and your unwavering love and encouragement have carried me through the toughest times. I owe a profound debt of gratitude to my parents, whose sacrifices have shaped me into the person I am today. Although my mother could not be here to witness this milestone, she has always remained in my heart, and I truly hope I have made her proud. Dad, you always envisioned me earning the title of Doctor, and I hope I will be able to fulfill that dream. I am also immensely grateful to my brother, sister-in-law, and all my relatives for their unwavering support and encouragement throughout this journey. Special thanks to Pushpa aunty and Sarath uncle for their unwavering love and support during the COVID period, a time when I was alone and their care meant so much.

Furthermore, I must express my gratitude to the people of Sri Lanka, whose commitment to funding free education through their taxes has afforded me and many others incredible opportunities for learning and growth. Your support has played an essential role in my educational journey, and for this, I am eternally grateful.

Finally, I extend heartfelt thanks to my friends and teachers at Siridhamma College, my batchmates and lecturers at the University of Peradeniya, and my colleagues at the DARE Center. Special thanks to Vihanga and Linduni, whose friendship and steadfast support have provided immense comfort and encouragement throughout this journey.

I dedicate this work to the memory of my beloved mother, who passed away on May 20, 2023.

Authorship Declaration

Abstract Publications

- Kaveendri, D., Ramesh, N., and Cripps S. “A Comparative study of the Priors used in Bayesian Variable Selection for Quantile Regression”, AUSTRALASIAN APPLIED STATISTICS CONFERENCE (AASC 2022).

- Athukorala, R., Kaveendri, D., Ramesh, N., Chin, V., Vervoort, R.W. and Cripps S., “A Bayesian framework for rainfall estimation and uncertainty quantification using reanalysis data: A case study for the Sydney region using ERA5 climate data”, 24th International Congress on Modeling and Simulation (MODSIM 2021) - this is a group research project conducted as part of the coursework required by the ARC DARE Center.

Presentations

- Bayes on the Beach 2024 - Bayesian Variable Selection in Spatial Quantile Regression: An Application to Assess Global Climate Indices’ Influence on Rainfall Distributions in NSW - **Best Poster Presentation Prize**.

- The Australian Meteorological and Oceanographic Society Conference (AMOS 2024) - Presented the abstract titled "Bayesian Variable Selection for Spatial Quantile Regression: A case study for NSW in Australia using global climate indices".

Following publication is in preparation from this thesis

- Kaveendri, D., Ramesh, N., Cripps E., and Cripps S. (*in prep*) "*Climate Inference using Bayesian Variable Selection in Spatial Quantile Regression*". (to be submitted to the *Environmentrics Journal*, see Chapter 3)

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Dilani Kaveendri Dambure Hewage
Date: 2025/03/28

As the supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Professor Matthew Cleary
Date: 2025/03/28

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Motivation

Rainfall variability is a key factor shaping ecosystems, agricultural productivity, and water resource management in eastern Australia. The region is particularly vulnerable to fluctuations in rainfall driven by large-scale climate processes, which lead to significant variability across months and locations. Understanding the drivers of this variability is crucial for better management of natural systems and for mitigating the impacts of extreme weather events, such as droughts and floods, on human activities (see Fowler et al., 2022; CSIRO and Bureau of Meteorology, 2024; Waha et al., 2022).

Several major climatic phenomena, including the El Niño-Southern Oscillation (ENSO), the Indian Ocean Dipole (IOD), and the Southern Annular Mode (SAM), play a critical role in shaping rainfall variability across Eastern Australia. ENSO, in particular, is a dominant driver of variability, with El Niño typically bringing drier conditions and La Niña resulting in wetter-than-average periods (see Department of Agriculture, Water and the Environment, 2021). This variability is especially important for Eastern Australia, where La Niña events often result in substantial increases in rainfall, leading to flooding, while El Niño contributes to drought conditions that strain water resources, agriculture, and ecosystems (see Gergis et al., 2012; Delage and Power, 2020; Abram et al., 2020; Ma et al., 2022; Lane et al., 2023). Similarly, the IOD significantly influences rainfall patterns, particularly during winter and spring. The negative phase of the IOD tends to bring increased rainfall across parts of Eastern

Australia, while the positive phase is associated with drier conditions, exacerbating drought periods. The SAM also modulates rainfall in this region, influencing rainfall patterns in the southern and eastern parts of Australia, with stronger impacts during different seasons (see ARC Centre of Excellence for Climate Extremes, 2022).

While previous research has traditionally concentrated on the impact of these climate drivers on average rainfall patterns (see Opoku-Ankomah and Cordery, 1993; Cai et al., 2014; Duc et al., 2017; Holgate et al., 2022), recent studies have shifted focus toward understanding the variability in rainfall extremes. Investigating not only mean rainfall but also the full distribution-including extreme events-provides a more comprehensive understanding of how these climate phenomena shape rainfall in eastern Australia. This shift is particularly relevant for assessing risks associated with both flood and drought conditions, which have significant implications for the region's resilience to climate variability.

Recent studies have highlighted the critical roles of the SOI, the Dipole Mode Index (DMI), and the SAM in influencing extreme rainfall patterns in Australia. For instance, research by Min et al., 2013; Sun et al., 2015; Evan Hajani and Ishak, 2017; Bertolacci et al., 2019; Khastagir et al., 2022; Mbigi and Xiao, 2023 shows the importance of how these indices collectively shape the occurrence and intensity of daily extreme rainfall events. In practice, tools such as CSIRO's INDRA [1] provide simplified analyses of the influence of climate indices on extremes. Although relatively basic, INDRA demonstrates how these indices are applied in real-world contexts, particularly in agriculture and water management. .

Using Generalized Extreme Value (GEV) analysis, studies such as Min et al., 2013 and Sun et al., 2015 have demonstrated that the influence of large-scale climate drivers, including SOI, DMI and SAM, significantly impacts daily extreme rainfall patterns. These studies show that variations in these indices affect the intensity and frequency of extreme rainfall events, particularly in regions like eastern Australia. The research reveals that the effects of climate indices on extreme rainfall events tend to be more pronounced than their influence on

[1] https://www.csiro.au/en/research/technology-space/data/indra

the average rainfall. For instance, shifts in SOI, DMI, and SAM influence the likelihood of both extreme dry and wet conditions, making these climate drivers critical for understanding rainfall variability and associated risks.

The findings of Evan Hajani and Ishak, 2017 align with these observations, showing significant trends in annual and seasonal maximum rainfall influenced by key climate indices such as SOI, DMI, and SAM. In a similar frequentist framework, Feng and et al., 2018 employed regression models to assess the influence of extreme rainfall on agricultural systems in southeastern Australia, quantifying the impacts of ENSO and IOD on rainfall variability. Additionally, Haylock and Nicholls, 2000 analyzed long-term trends in extreme rainfall across Australia, using regression-based methods to explore the relationship between the SOI and extreme rainfall patterns. More recent studies have further investigated the relationship between climate indices and daily rainfall extremes in Australia, including work by Yildirim and Rahman, 2022; He et al., 2025; Dey et al., 2021; Nguyen et al., 2021, etc. Collectively, these studies enhance our understanding of how these climate indices impact extreme daily rainfall across Australia, emphasizing the importance of assessing the significance of climate indices' relation to daily rainfall patterns.

Despite significant advances in physical, process-based climate models, particularly in understanding large-scale atmospheric and oceanic interactions like ENSO, IOD, and SAM on daily extreme rainfall, there remains a scarcity of probabilistic models that capture the relationship between these climate drivers and the entire distribution of monthly rainfall. Many studies often overlook the inherent uncertainty and variability in climate systems. This gap has led to a growing interest in Bayesian models, which allow for the integration of prior knowledge and probabilistic estimation of parameters. A critical component of these models is variable selection, which helps identify the most influential climate drivers, ensuring that the model focuses on the most relevant predictors while accounting for uncertainty. However, the development and application of such Bayesian models with robust variable selection techniques remain limited, especially when applied to understanding the dependency of the

aforementioned climate indices on the full spectrum of monthly rainfall variability, including extreme events. More research is necessary to address this gap and enhance predictive capacities for rainfall distribution under varying climate conditions.

Moreover, the Australian Bureau of Meteorology (BoM) provides monthly and seasonal rainfall forecasts based on the chance of exceeding certain quantiles, rather than just focusing on average rainfall[2]. This is how rainfall forecasts are usually communicated to the Australian public. However, most of the research so far has focused on the mean, with little attention given to how climate drivers affect different parts of the rainfall distribution. There is a real gap in understanding the influence of large-scale climate indices on monthly rainfall quantiles, and this thesis takes an important initial step toward addressing that gap.

In this thesis, we apply Bayesian variable selection within a spatial quantile regression framework to model the entire distribution of monthly rainfall. This approach is crucial for understanding how key factors influence monthly rainfall, how these relationships vary across different locations (spatial variability), and how they affect different parts of the monthly rainfall distribution. The core research focus of this thesis is to enhance the understanding of how global climate indices impact monthly rainfall distribution capturing their spatial heterogeneity, through this novel contribution to Bayesian variable selection in spatial quantile regression.

Recent work in Bayesian quantile regression has focused on incorporating spatial dependencies and improving variable selection. Tareghian and Rasmussen, 2013 introduced Bayesian variable selection methods within the quantile regression framework, but without addressing spatial correlation. Reich, 2012 developed spatio-temporal quantile regression models that account for both temporal changes and spatial dependencies. Similarly, Duan et al., 2022 integrated spatial correlations specifically in the context of Australian temperature data. Furthermore, Bertolacci et al., 2019 proposed a hierarchical Bayesian mixture model as an alternative to quantile regression to make inferences about daily rainfall in Australia.

[2] http://www.bom.gov.au/climate/outlooks//rainfall/median/monthly/0

This model incorporates spatial dependencies through Gaussian process priors, allowing for flexible spatial relationships. However, these approaches do not include Bayesian variable selection for assessing the importance of predictors across quantiles, which remains a key limitation in analyzing variable effects within the rainfall distribution.

This thesis addresses these gaps in the literature by introducing a new method that combines Bayesian variable selection with spatial quantile regression. This approach helps to analyze how global climate indices affect the distribution of monthly rainfall across New South Wales (NSW), Australia. NSW was chosen as a representative case study, and we intend to extend the method to all eastern states in future work, given the important role of the Great Dividing Range in modulating the influence of ENSO and other climate drivers on rainfall.

A key advantage of this methodology is that it accounts for spatially varying inferences on the impact of climate indices on the monthly rainfall distribution at observed locations, while also providing out-of-sample predictions for these impacts. The specific contributions of this thesis are outlined in Section 1.2.

1.2 Thesis Contribution

Therefore, in this thesis, we develop a parsimonious methodology that incorporates Bayesian variable selection within spatial quantile regression framework to capture spatially varying dependencies between not only the central tendency but also the entire distribution of monthly rainfall and key climate indices. This is achieved through a hierarchical Bayesian quantile regression model, where each level of the hierarchy is tailored to specific modeling needs and complexities. Below, we outline the hierarchical structure based on the chapters.

(i) **Bayesian Variable Selection for Quantile Regression: Inference on the Climate Indices' Relation to Monthly Rainfall Distribution**

We develop a Bayesian variable selection model within a hierarchical framework to analyze monthly rainfall across multiple locations in New South Wales (NSW), Australia, treating each location independently (spatial independence among observational sites) which is often used in the applied literature (Jakob and Walland, 2016; Islam et al., 2017; Yilmaz et al., 2017; Wu and Leonard, 2019). Missing rainfall values are treated as unknown parameters and generated within the sampling scheme. The hierarchical approach first models rainfall distribution using the asymmetric Laplace distribution (ALD) through a location-scale mixture (Kotz et al., 2001). Variable selection is then performed by modeling the regression coefficients as a two-component mixture, where each coefficient has a probability of being drawn either from a point mass at zero, indicating exclusion, or from a normal distribution, indicating inclusion. This allows for flexible selection of relevant covariates (George and McCulloch, 1993). Subsequent chapters discuss the development and implementation of advanced statistical models that integrate spatial dependence.

(ii) **ENSO Inference on Monthly Rainfall using Bayesian Variable Selection in Spatial Quantile Regression**

In this chapter, we extend the model to include spatial dependencies, where locations are analyzed jointly using one climate covariate, specifically the Southern Oscillation Index (SOI). The decision to focus on a single climate index in this spatial analysis is motivated by the need to reduce the degrees of freedom in the model. Including multiple climate indices would significantly increase the model's complexity, particularly when accounting for spatial dependencies, which could lead to overfitting. By concentrating on SOI, we can more effectively manage this complexity while still capturing key climate-driven variations in rainfall. Moreover, SOI is known to have a dominant influence on rainfall patterns in the region, making it a robust choice for this

initial exploration of spatial dependencies in our model.

We extend the hierarchical model to incorporate spatial dependence through a Gaussian Process ($\mathcal{GP}$) prior on the regression coefficients. The first and second levels of the hierarchy remain unchanged. At the third level, spatial dependence is introduced via two mechanisms. First, the probability that the regression coefficient is non-zero is modeled as a probit regression where the argument of the link function varies spatially according to a Gaussian Process prior. Second, the prior on the regression coefficients, conditional on those coefficients being non-zero, is modelled as a multivariate normal distribution, where the mean of the distribution also varies spatially according to a Gaussian Process prior (Wahba, 1990; Wood, 2013).

(iii) **Extension of Variable Selection in Spatial Quantile Regression to Assess the Joint Impact of Climate Indices on Monthly Rainfall Distribution**

In this chapter, we extend this framework in (ii) to incorporate multiple climate indices, allowing us to capture more comprehensive climate-rainfall interactions. This gradual approach ensures that the complexity is introduced in a manageable way, starting with a single covariate and later expanding to include other significant indices such as the Dipole Mode Index (DMI).

At the third level, spatial dependence is modelled through the extension of a multinomial logistic regression for joint inclusion probability in the variable selection framework. This approach allows for the joint modeling of multiple climate indices, with a $\mathcal{GP}$ prior placed on the argument of the link function to capture the spatial variability.

1.3 Statistical Background

In this section, we provide a brief overview of the statistical techniques used throughout the thesis to construct the models.

1.3.1 Quantile Regression

Quantile regression, introduced by Koenker and Bassett, 1978, has become an essential tool in statistical modeling, offering a broader perspective than traditional mean regression. Unlike standard approaches that focus solely on the mean, quantile regression enables the estimation of conditional quantiles, providing insights into different points of the outcome distribution. This allows for a more comprehensive understanding of the relationship between variables across various quantiles, making it particularly useful in applications where extremes or distributional tails are of interest. With its flexibility and robustness, quantile regression has gained widespread use in diverse fields, from economics to climate science. For a detailed theoretical foundation and applications, see Koenker, 2005 and Yu et al., 2003.

We consider the quantile regression (QR) model for the general case,

$$\boldsymbol{y} = X\boldsymbol{\beta}^{(q)} + \boldsymbol{\epsilon}\,, \tag{1.1}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)'$ is the response variable, and $\boldsymbol{\beta}^{(q)} = (\beta_0^{(q)}, \ldots, \beta_K^{(q)})'$ is the vector of regression coefficients, consisting of the intercept $\beta_0^{(q)}$ and the coefficients for the K covariates, the design matrix $X = (\mathbf{1}_n, \mathbf{x}_1, \ldots, \mathbf{x}_K)$ and $\mathbf{x}_k = (x_{1,k}, \ldots, x_{n,k})'$ and the distribution of $\boldsymbol{\epsilon} = (\epsilon_1, ..., \epsilon_n)'$ is Asymmetric Laplace Distribution (ALD) and depend on the quantile q; $0 < q < 1$. Koenker, 2005 proved that the q^{th} quantile regression estimation for $\boldsymbol{\beta}^{(q)}$ can be obtained through the minimization the loss function given below,

$$\min_{\beta} \sum_{i} \rho_q(y_i - \boldsymbol{x}_i \boldsymbol{\beta}^{(q)}) \,, \tag{1.2}$$

where $\boldsymbol{x}_i$ denotes i^{th} row of X matrix and the check loss function $\rho_q(u)$ is defined as,

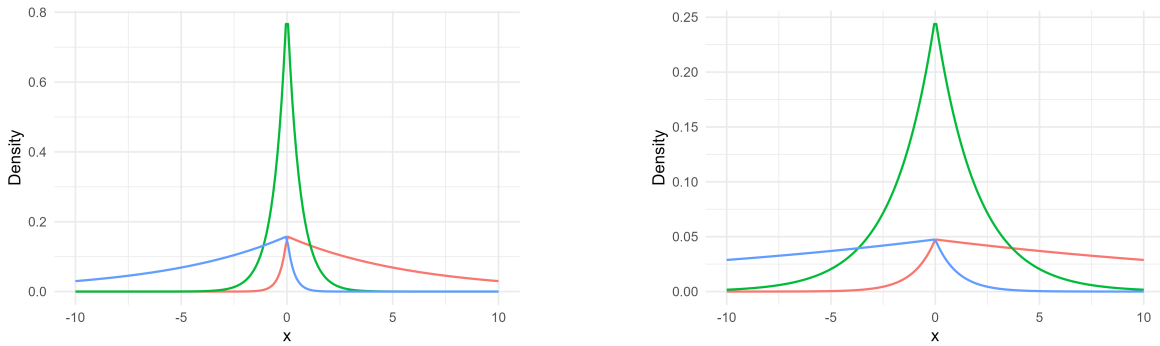
$$\rho_q(u) = u\{q - I(u < 0)\} \,, \tag{1.3}$$

where $I(\cdot)$ denotes the indicator function.

The Asymmetric Laplace distribution, where $0 < q < 1$ is the skew parameter, $\sigma > 0$ is the scale parameter, with location parameter equal to zero ($\mathcal{AL}(0, \sigma, q)$), and is given by,

$$f_q(\boldsymbol{\epsilon}|q, \sigma) \sim \left\{ \frac{q(1-q)}{\sigma} \right\}^n \exp \left\{ \frac{-\sum_{i=1}^{n} \rho_q(\epsilon_i)}{\sigma} \right\} . \tag{1.4}$$

Example density distributions are given in Figure 1.1 to give an understanding of how the ALD behaves in various quantiles under different settings using the ald package in R (Galarza and Lachos, 2021).



(a): $\sigma = 0.3$ (b): $\sigma = 1$

FIGURE 1.1: Asymmetric Laplace densities for the lower tail ($q = 0.05$, blue), the median ($q = 0.5$, green), and the upper tail ($q = 0.95$, red), using a location parameter $\mu = 0$ under different scale parameters σ. Panel (a) illustrates $\sigma = 0.3$, while panel (b) illustrates $\sigma = 1$.

Equations 1.1 and 1.4 imply that $y_i \sim \mathcal{AL}(\boldsymbol{x}_i\boldsymbol{\beta}^{(q)}, \sigma, q)$, with a likelihood function given by:

$$f_q(\boldsymbol{y}|\boldsymbol{\beta}^{(q)}, \sigma, q) \sim \left\{\frac{q(1-q)}{\sigma}\right\}^n \exp\left\{-\sum_{i=1}^{n} \frac{\rho_q(y_i - \boldsymbol{x}_i\boldsymbol{\beta}^{(q)})}{\sigma}\right\}. \tag{1.5}$$

1.3.2 Bayesian Estimation

In a Bayesian framework, where the data are denoted by $\boldsymbol{y}$ and the parameters by $\boldsymbol{\vartheta}$, inference typically focuses on the posterior distribution $p(\boldsymbol{\vartheta}|\boldsymbol{y})$. For simplicity, assume $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_d)' \in \mathbb{R}^d$. A common objective is to estimate expectations of functionals, $E[g(\boldsymbol{\vartheta})|\boldsymbol{y}]$, where $g : \mathbb{R}^d \to \mathbb{R}^{d'}$.

By Bayes' theorem, the posterior distribution is given by:

$$p(\boldsymbol{\vartheta}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\vartheta})p(\boldsymbol{\vartheta})}{p(\boldsymbol{y})},$$

where the numerator is typically straightforward to compute. However, the denominator, $p(\boldsymbol{y})$, known as the normalizing constant, is an integral:

$$p(\boldsymbol{y}) = \int_{\boldsymbol{\vartheta}} p(\boldsymbol{y}|\boldsymbol{\vartheta})p(\boldsymbol{\vartheta})d\boldsymbol{\vartheta},$$

which is often intractable. Even when this constant is known, obtaining analytical expressions for expectations can be challenging, particularly when d is large.

If it is possible to sample from the posterior distribution $p(\boldsymbol{\vartheta}|\boldsymbol{y})$, a Monte Carlo estimator can be employed to approximate the expectation:

$$\hat{E}[g(\boldsymbol{\vartheta})|\boldsymbol{y}] = \frac{1}{J}\sum_{j=1}^{J} g(\boldsymbol{\vartheta}^{[j]}), \tag{1.6}$$

where $\boldsymbol{\vartheta}^{[j]} \sim p(\boldsymbol{\vartheta}|\boldsymbol{y})$, for $j = 1, \ldots, J$, are samples drawn from the posterior distribution.

In the context of Bayesian inference, a common objective is to estimate expectations of functionals, $E[g(\boldsymbol{\vartheta})|\boldsymbol{y}]$, where $g : \mathbb{R}^d \to \mathbb{R}^{d'}$. Since the posterior distribution is often analytically intractable, especially when d is large, Markov Chain Monte Carlo (MCMC) methods provide an efficient solution for sampling from the posterior distribution, allowing the use of Monte Carlo estimators to approximate these expectations.

One prominent MCMC method is Gibbs sampling (Gelfand and Smith, 1990), a special case of the Metropolis-Hastings algorithm where the proposal distribution for each parameter is derived from its conditional posterior distribution, making the acceptance step redundant. Gibbs sampling involves iteratively sampling from the conditional distributions of each parameter, given the current values of all other parameters. This process is particularly advantageous when the conditional distributions are easier to sample from than the joint posterior.

By using Gibbs sampling, we can sequentially estimate the posterior distribution of each parameter while keeping the others fixed. This iterative process continues until the Markov chain reaches its stationary distribution, which corresponds to the desired posterior $p(\boldsymbol{\vartheta}|\boldsymbol{y})$. The samples from this stationary distribution are then used in Monte Carlo estimators to approximate expectations, such as in Equation (1.6).

When using conjugate priors on the parameters, the full conditional distributions often take a known form, simplifying the sampling process. This makes Gibbs sampling particularly effective, as it allows for direct sampling from these distributions without the need for more complex algorithms such as Metropolis-Hastings. Conjugate priors help ensure that the full conditionals are easy to compute, thereby improving computational efficiency.

Gibbs sampling, as part of MCMC, plays an essential role in Bayesian quantile regression, especially when dealing with complex distributions like the asymmetric Laplace distribution (ALD). The use of conjugate priors further enhances the practicality of the Gibbs sampling approach, providing an efficient and straightforward solution for posterior inference.

1.3.3 Bayesian variable selection

Considerable research has focused on variable selection methods in quantile regression (QR), with various approaches developed to improve inference accuracy in complex datasets. Among the most prominent are global-local shrinkage priors, such as the horseshoe prior Kohns and Szendrei, 2023, which adaptively shrinks irrelevant coefficients to zero while preserving the relevant ones. In addition to shrinkage-based methods, several penalized regularization techniques have been proposed in the recent literature Li et al., 2010; Tian and Song, 2019; Alhamzawi and Algamal, 2019; El Adlouni et al., 2017; Benoit et al., 2013, with the objective of improving variable selection by imposing penalties on the magnitude of coefficients. These methods can control overfitting and improve predictive accuracy by introducing constraints that favor simpler models. Still, they may suffer from the limitation of uniformly shrinking all coefficients, which can lead to bias in estimating non-zero coefficients.

A significant advancement in this domain is the spike-and-slab prior Yu et al., 2013; Xi et al., 2015, which addresses this limitation by combining two distributions to better differentiate between relevant and irrelevant variables. The spike-and-slab prior explicitly models the uncertainty in variable inclusion by using a "spike" distribution sharply concentrated around zero for irrelevant coefficients and a broader "slab" distribution for potentially influential ones.

Therefore, this thesis builds on the spike-and-slab framework in QR, initially treating the prior on the probability of variable inclusion, which is modeled independently of spatial covariates assuming spatial independence. Subsequently, it extends this by modeling inclusion probabilities as a function of geographical location through probit or logit regression. This novel approach enhances the model's ability to capture spatial variability in both single and multiple variable contexts, offering a more comprehensive understanding of spatial dependencies in climate-related data.

1.3.3.1 Spike and Slab Prior

A rigorous approach to Bayesian variable selection involves the introduction of latent binary indicator variables, denoted γ_j, for each covariate $j = 1, \ldots, p$. These binary indicators, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)$, determine whether a specific covariate is included in the model ($\gamma_j = 1$) or not ($\gamma_j = 0$). A natural choice for sparsity-inducing priors in this context is the "Spike and Slab" two-component mixture prior (Mitchell and Beauchamp, 1988; Leamer, 1978). Similarly to Castillo and Mismer, 2018, we describe a more general version of this prior.

The Spike and Slab prior can be mathematically defined as

$$\beta_j|\gamma_j \sim (1-\gamma_j)G_0(\beta_j) + \gamma_j G_1(\beta_j), \quad \gamma_j \sim \text{Be}(\pi)\,, \tag{1.7}$$

where, $G_0(\beta_j)$ represents the "spike," often modeled as a Dirac delta function at zero, indicating that the covariate is excluded from the model. $G_1(\beta_j)$ is the "slab," a diffuse prior, commonly a normal distribution, representing the inclusion of the covariate. The prior on the binary inclusion variable γ_j is typically a Bernoulli distribution with parameter $\pi \in [0, 1]$, which is the probability that the covariate is included in the model, i.e., the inclusion probability.

In the original formulation of Mitchell and Beauchamp, 1988, G_0 was a Dirac delta function at zero and G_1 was a uniform distribution. In this more general version, G_1 can also be a normal distribution, which allows more flexibility in modeling.

The posterior distribution of the inclusion variables $\boldsymbol{\gamma}$ is given by

$$p(\boldsymbol{\gamma}|D) = \frac{f(D|\boldsymbol{\gamma})p(\boldsymbol{\gamma})}{Z(D)}\,,$$

where, $f(D|\boldsymbol{\gamma})$ represents the likelihood of the data D given the inclusion variables $\boldsymbol{\gamma}$, while $p(\boldsymbol{\gamma})$ denotes the prior on the inclusion variables. The term $Z(D) = \sum_{\boldsymbol{\gamma}\in\{0,1\}^p} f(D|\boldsymbol{\gamma})p(\boldsymbol{\gamma})$ is the normalizing constant, which ensures that the posterior probabilities sum to one. This formulation is fundamental in Bayesian variable selection, allowing efficient computation of posterior distributions.

1.3.4 Incorporating Spatial Dependence in Variable Selection via Non-Parametric Approach

This thesis extends the spike and slab prior introduced in Section 1.3.3.1 to be spatially varying via the prior distribution of the regression coefficients. In this extension, the slab is modeled as a multivariate normal distribution with a spatially varying mean, while the spike is a point mass at zero (see Smith and Kohn, 1996). The probability that the regression coefficients are exactly zero is parameterized as a function of space, facilitating spatially adaptive variable selection. Smoothing splines are employed to capture both linear and non-linear dependencies between spatial coordinates and the inclusion probability, allowing for intricate pattern fitting without a pre-specified functional form.

1.3.4.1 Single Variable Approach

Let $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_S)$ be the vector of indicator variables for location, $s = 1, \ldots, S$ for the case of one covariate,

$$\Pr(\gamma_s = 1) = \pi_s = \Phi(g(x_{1,s}, x_{2,s})), \tag{1.8}$$

where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard normal distribution, and $g(x_{1,s}, x_{2,s})$ is a function that captures the spatial relationships where $x_{1,s}$ and $x_{2,s}$ are latitude and longitude in location s. The function $g(x_{1,s}, x_{2,s})$ can be expressed as:

$$g(x_{1,s}, x_{2,s}) = \alpha_0 + \sum_{m=1}^{m-1} \alpha_{m,1} x_{1,s}^m + \sum_{m=1}^{m-1} \alpha_{m,2} x_{2,s}^m + f(x_{1,s}, x_{2,s}),$$

where α_0 is the intercept, and the terms $\alpha_{m,1}x_{1,s}^m$ and $\alpha_{m,2}x_{2,s}^m$ represent the polynomial components, capturing linear spatial relationships up to degree $m-1$. The value of m, which represents the number of basis functions used to model the non-linear component $f(x_{1,s}, x_{2,s})$, determines the flexibility of the model. A larger m provides more flexibility in capturing complex spatial patterns but requires regularization to prevent overfitting. For more details see, Wood, 2013.

The smooth component $f(x_{1,s}, x_{2,s})$, modeled using smoothing splines, captures non-linear spatial dependencies. These basis functions are derived from the spatial coordinates through eigenvalue decomposition, and their selection is critical for achieving a balance between smoothness and model complexity.

Thus, smoothing splines allow the model to flexibly adapt to both linear and non-linear spatial patterns, while the spike and slab prior ensures that irrelevant covariates are excluded. This approach leads to a parsimonious and spatially adaptive model, allowing for variable inclusion probabilities to vary smoothly across space.

To efficiently implement this model in the case of binary outcomes, use of the *data augmentation approach* (see Albert and Chib, 1993; Wood and Kohn, 1998), introducing latent variables $\mathbf{w} = (w_1, \ldots, w_S)$ such that:

$$w_s = g(x_{1,s}, x_{2,s}) + e_s, \quad e_i \sim N(0,1)$$

The binary outcome γ_s is linked to w_s via a *probit*,

$$\Pr(\gamma_s = 1 | g(x_{1,s}, x_{2,s})) = \Pr(w_s > 0 | g(x_{1,s}, x_{2,s})) = \Phi(g(x_{1,s}, x_{2,s}))\,.$$

Alternatively, we use a *logit* model using the polya-gamma augmentation approach propose by Polson et al., 2012,

$$\Pr(\gamma_s = 1 | g(x_{1,s}, x_{2,s})) = \frac{1}{1 + \exp(-g(x_{1,s}, x_{2,s}))} .$$

This augmentation enables the use of conjugate priors for the regression coefficients, facilitating more straightforward computation.

1.3.4.2 Multiple Variable Approach

Let $\boldsymbol{\gamma}_s = \{0, 1\}^K$ denote the indicator variable set for K covariates at the location s and let $C = 2^K$ denote the possible combinations of binary values. As the number of combinations increases exponentially with the number of covariates, a large value of K can lead to significant computational complexity.

The joint probability for the c^{th} combination is given by

$$\pi_{s,c} = \Pr(\boldsymbol{\gamma}_s \in c) = \frac{\exp(g_c(x_{1,s}, x_{2,s}))}{1 + \sum_{r=1}^{C-1} \exp(g_r(x_{1,s}, x_{2,s}))}, \tag{1.9}$$

where $g_c(x_{1,s}, x_{2,s})$ represents the spatial function for the covariates at location s, accounting for the c^{th} combination of covariate outcomes. This captures the spatial correlation at location s of the covariates $x_{1,s}$ and $x_{2,s}$ to the joint probability for combination c.

1.4 Data : Rainfall and Climate Indices

In this thesis, we use rainfall data and key climate indices obtained from the Bureau of Meteorology, Australia (BOM) [3]. The dataset spans a period of 73 years, from 1950 to 2022, and includes detailed records of both monthly rainfall and climate indices such as the Southern

[3] http://www.bom.gov.au/climate/data/

Oscillation Index (SOI), Dipole Mode Index (DMI), and Southern Annular Mode (SAM). These data are fundamental to the analyses present in all chapters, allowing for a thorough investigation of the relationship between climate variability and monthly rainfall distribution.

1.4.1 Rainfall data

We use 711 observational sites in NSW, Australia, as indicated by the blue dots in Figure 1.2. These sites are selected because they have 20% or less missing values.

For analysis purposes, we select six stations (marked in red, shown in Figure 1.2) covering NSW; near the coast in three different locations in the middle, Sydney Botanic Gardens [site ID: 66006]; to the North, Yamba pilot Station [site ID: 58012]; and to the South, Moruya Heads Pilot Stations [site ID: 69018]. For inland stations, we chose Dunedoo Post Office [site ID: 64009] and Collarenebri (Albert St) [site ID: 48031] and, far inland, Pooncarie (Karpa Kora Station) [site ID: 47013].

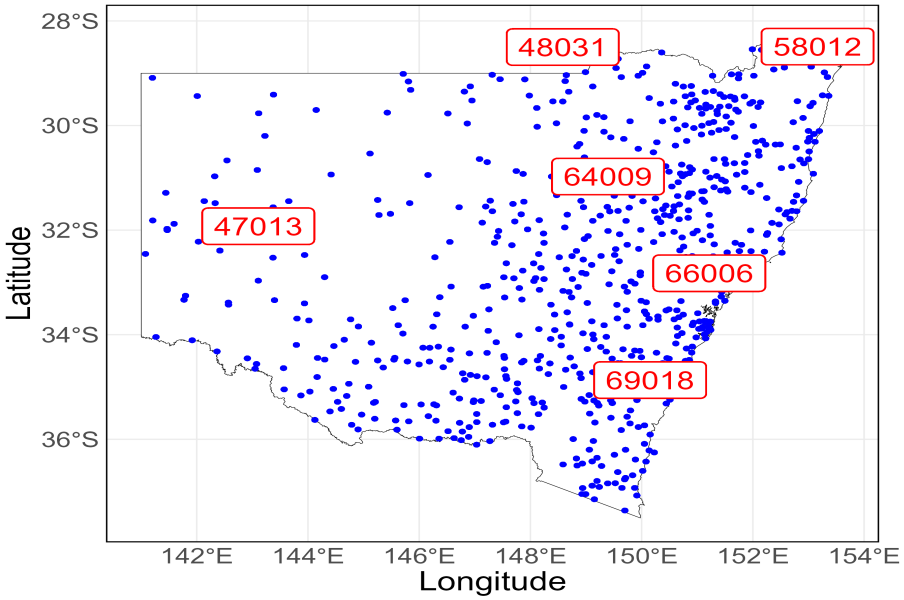


FIGURE 1.2: Map of 711 observational sites across NSW, with six key stations highlighted in red.

1.4.2 Large-Scale Climate Indices

The Southern Oscillation Index (SOI), Dipole Mode Index (DMI), and Southern Annular Mode (SAM) are crucial climate indices used to describe and predict atmospheric and oceanic patterns. Below is a brief overview of each,

- The **Southern Oscillation Index (SOI)** measures the atmospheric sea level pressure difference between Tahiti and Darwin (Troup, 1965). It is one of the principal measures used to indicate the strength and phase of the El Niño Southern Oscillation (ENSO) phenomenon, which affects weather patterns globally. ENSO consists of El Niño (warm) and La Niña (cool) events, each of which occurs every few years in the tropical Pacific Ocean. El Niño and La Niña events are associated with warm or cool sea surface temperatures in the central and eastern equatorial Pacific, respectively, and affect the overlying atmosphere to produce temperature, circulation, and moisture changes that impact weather on every continent.

- The **Dipole Mode Index (DMI)**, is a prominent interannual climate phenomenon in the tropical Indian Ocean, marked by a dipole pattern of sea surface temperature anomalies. A positive DMI event typically involves cooler sea surface temperatures near the southwestern coast of Sumatra, along with warmer conditions in the western equatorial Indian Ocean. In contrast, a negative DMI event is characterized by the opposite pattern: warmer waters off Sumatra and cooler conditions in the western Indian Ocean (Saji et al., 1999; Webster et al., 1999). This oscillation has significant impacts on weather and rainfall patterns across the Indian Ocean region.

- The **Southern Annular Mode (SAM)**, represents the latitudinal shift of the southern hemisphere's westerly wind belt, a zonally symmetric circulation pattern that circles Antarctica. This feature significantly influences rainfall and temperature

variability across the extra tropics of the Southern Hemisphere, with its effects extending to regions such as southern Australia and South America. Changes in SAM are closely linked to climate variability in these regions (Gillett et al., 2006; Garreaud, 2007; Hill and Raphael, 2009). Furthermore, SAM is considered an extratropical driver of variability, with its influence primarily confined to the southwest and southeast of the Australian continent (Hendon et al., 2007; Risbey et al., 2009b).

For clarity, throughout this thesis we refer to the ENSO, IOD, and SAM phenomena themselves as *climate drivers*. Their associated indicators — SOI, DMI, and the SAM index — are used as *climate indices* to characterise the variability of these drivers, and these indices are the covariates included in our study. This terminology is applied consistently throughout the thesis.

1.5 Thesis outline

The structure of this thesis is as follows. Chapter 2 introduces the development of a Bayesian variable selection framework for quantile regression, aimed at assessing the relationship between global climate indices and monthly rainfall across multiple locations in New South Wales, Australia. Each location is treated independently, as is often used in the climate and engineering literature, without accounting for spatial dependencies. Chapter 3 builds on this foundation by incorporating spatial dependence using a Gaussian process prior, with a particular focus on the Southern Oscillation Index (SOI) to model spatial variability in rainfall patterns. In Chapter 4, the methodology is further expanded to include multiple climate indices, offering a more comprehensive analysis of the interactions between climate drivers and rainfall across NSW in a spatial context. Finally, Chapter 5 concludes the thesis with a discussion of the main contributions, limitations, and potential directions for future research.

CHAPTER 2

Bayesian Variable Selection for Quantile Regression: Inference on the Climate Indices' Relation to Monthly Rainfall Distribution

---

In this chapter, we outline Bayesian variable selection using the spike and slab prior (Mitchell and Beauchamp, 1988), to assess the relationship between climate indices and monthly rainfall quantiles in New South Wales (NSW), Australia. The motivation for this analysis is to identify the influence of large-scale climate phenomena such as El Niño-Southern Oscillation (measured using the Southern Oscillation Index (SOI)), the Indian Ocean Dipole (measured using the Dipole Mode Index (DMI)), and the Southern Annular Mode (SAM) on the monthly rainfall distribution. In this chapter, we ignore the spatial dependence of monthly rainfall and this serves as the baseline model, which is a common practice in the climate and engineering literature (Yilmaz et al., 2017; Islam et al., 2017; Wu and Leonard, 2019; Jakob and Walland, 2016). In Chapter 3, we then incorporate spatial dependence and show that this improves the predictive performance of our model, as expected.

There has been extensive research done in the area of the impacts of large-scale climate phenomena on average rainfall (see Duc et al., 2017; Holgate et al., 2022 etc.). However, recent studies in climate science have shifted focus from understanding these relations to rainfall extremes, see for example (Yilmaz et al., 2017; Jakob and Walland, 2016; Westra et al., 2013; Evan Hajani and Ishak, 2017; Min et al., 2013; Sun et al., 2015 etc.). Studies such as Jakob and Walland, 2016, Westra et al., 2013, and Evan Hajani and Ishak, 2017 utilize the statistical trend analysis test to investigate the presence of trends in annual maximum daily precipitation under changing climatic conditions.

Methods like Generalized Extreme Value (GEV) models are commonly used in climate science to model the relationships between climate indices and daily rainfall extremes. Min et al., 2013 employ the non-stationary GEV approach as used in Zhang et al., 2010 and investigate the possible impact of ENSO, IOD, and SAM on seasonal extremes of rainfall and temperature over Australia. Moreover, Yilmaz et al., 2017 utilize both stationary and non-stationary GEV models to investigate the influence of climate oscillation indices on annual maximum daily rainfall data. A recent study by Castillo-Mateo et al., 2023 introduces a Bayesian variable selection approach within GEV models to study annual maximum daily temperature series, employing a stochastic search variable selection (SSVS) technique to assess the impact of atmospheric covariates. Despite the effectiveness of the GEV approach, it has limitations. GEV models focus on block maxima and may not effectively capture the entire rainfall distribution. Moreover, these models do not directly facilitate a quantile-based approach.

Despite extensive research on the influence of climate indices on rainfall variability, there remains a significant gap in understanding how these indices jointly affect rainfall distribution across different quantiles using a Bayesian variable selection framework. Some research, such as that by Tareghian and Rasmussen, 2013, focuses on applying Bayesian variable selection for daily precipitation downscaling, following the approach proposed by Yu et al., 2013. However, their study was limited to five stations in Canada, did not jointly model those locations, and did not explore the joint impact of multiple climate indices on rainfall quantiles.

To address these gaps, we contribute a Bayesian variable selection approach using a spike-and-slab prior to quantify the influence of climate indices on monthly rainfall quantiles. Although most of the aforementioned studies focus on daily rainfall, our study emphasizes monthly rainfall data. The use of monthly data is advantageous for reducing the short-term variability and noise often present in daily observations. By aggregating rainfall at the monthly scale, we can better analyze the relationships between climate indices, such as SOI,

DMI and SAM, and their influence on rainfall patterns while still capturing the variations in these relationships across seasons. This approach is novel in its application to climate data in New South Wales, Australia, where monthly rainfall quantiles are analyzed at multiple locations. The analysis is conducted on a site-by-site basis in New South Wales, Australia, providing insights into local rainfall patterns, as the foundation for building a model that explicitly encodes spatial variability, see Chapter 3.

We evaluate, via simulations, the performance of our method by comparing it to the Quantile Regression - Stochastic Search Variable Selection (QR-SSVS) approach proposed by Yu et al., 2013, that places a Cauchy prior over the marginal distribution of the regression coefficients. Our approach applies modified Zellner's g-prior in the context of quantile regression (similar to Alhamzawi and Yu, 2013) for the slab of the regression coefficients that are selected to be included in the model. This study ensures that our selected priors are consistent with the theoretical framework and empirical requirements of the research context.

In addition to simulation studies, we show the results of the impact of these climate indices on the distribution of monthly rainfall for selected sites in NSW. Finally, to evaluate the overall performance across all 711 sites considered in this study, we use a comprehensive performance metric proposed by Bertolacci et al., 2019, which allows for a detailed comparison of model fit and predictive accuracy.

This study demonstrates the utility of Bayesian variable selection methods for quantile regression in the context of climate data. It offers a flexible framework that can handle missing data and provide robust inference for large-scale climate indices at different monthly rainfall quantiles.

The structure of this chapter is as follows. Section 2.1 describes the model and the priors, Section 2.2 explains how the MCMC scheme is implemented, Section 2.3 briefly outlines the method proposed by Yu et al., 2013, Section 2.4 illustrates the results of the simulated

data and impact of climate indices on monthly rainfall quantiles, and Section 2.5 discusses the conclusions drawn from our findings.

2.1 Model and Priors

This chapter aims to build a Bayesian variable selection model that captures the dependencies between climate indices and the distribution of monthly rainfall. We do this by taking a hierarchical approach. The first level of the hierarchy defines the likelihood and relates monthly rainfall quantiles to large-scale climate indices. The second level of the hierarchy allows for variable selection by modeling the prior on the regression coefficients as a two-component mixture where the mixture components are a point mass at zero and a normal distribution, known as a spike and slab prior (Mitchell and Beauchamp, 1988).

2.1.1 Likelihood

Suppose we have observations on rainfall for a given month m, and location s, over years $t = 1, \ldots, T$ denoted by $\boldsymbol{y} = (y_1, \ldots, y_T)'$, and corresponding observations on K large-scale climate indicators, denoted by $\mathbf{x}_1, \ldots, \mathbf{x}_K$, where $\mathbf{x}_k = (x_{1,k}, \ldots, x_{T,k})'$, for $k = 1, \ldots, K$. For clarity, we omit explicit subscripts for s and m in the notation, and we note that the parameters will have different values for different locations s and months m.

Let $X = (\mathbf{1}_T, \mathbf{x}_1, \ldots, \mathbf{x}_K)$, where $\mathbf{1}_T$ is a $T \times 1$ vector of 1's. We follow Yu and Moyeed, 2001, and express the q^{th} quantile of rainfall, conditional on these large-scale climate indicators as, $X\boldsymbol{\beta}^{(q)}$, and connect these quantiles to the observations $\boldsymbol{y}$ via the model,

$$\boldsymbol{y} = X\boldsymbol{\beta}^{(q)} + \boldsymbol{\epsilon}\,, \tag{2.1}$$

where $\boldsymbol{\beta}^{(q)} = (\beta_0^{(q)}, \beta_1^{(q)}, \ldots, \beta_K^{(q)})'$ are the regression coefficients for the q^{th} quantile of rainfall conditional on the large-scale climate indicators. The error term $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_T)'$ has an Asymmetric Laplace distribution, where $0 < q < 1$ is the skew parameter, $\sigma > 0$ is the scale parameter, with location parameter equal to zero ($\mathcal{AL}(0, \sigma, q)$), and is given by,

$$f_q(\boldsymbol{\epsilon}) \sim \left\{ \frac{q(1-q)}{\sigma} \right\}^T \text{ exp } \left\{ \frac{-\sum_{t=1}^T \rho_q(\epsilon_t)}{\sigma} \right\} . \tag{2.2}$$

The check loss function is given by $\rho_q(u) = u(q - I(u < 0))$, $I(.)$ denotes the indicator function, s.t $I_{\mathcal{A}}(x) = 1$ if $x \in \mathcal{A}$ and $I_{\mathcal{A}}(x) = 0$ otherwise.

It follows that,

$$y_t \sim \mathcal{AL}(\boldsymbol{x}_t \boldsymbol{\beta}^{(q)}, \sigma, q) \text{ for } t = 1, \ldots, T.$$

So that the likelihood is,

$$f_q(\boldsymbol{y} | \boldsymbol{\beta}^{(q)}, \sigma, q) \sim \left\{ \frac{q(1-q)}{\sigma} \right\}^T \text{ exp } \left\{ \frac{-\sum_{t=1}^T \rho_q(y_t - \boldsymbol{x}_t \boldsymbol{\beta}^{(q)})}{\sigma} \right\} , \tag{2.3}$$

where $\boldsymbol{x}_t$ is the t^{th} row of X, i.e $(1, x_{t,k}, \ldots, x_{t,K})$. Note the corresponding error distribution and likelihood formulation are discussed in detail in Section 1.3.1.

2.1.2 Priors for Variable Selection

The goal of this chapter is to identify the large-scale climate indices which are useful in predicting monthly rainfall quantiles in NSW. We do this by placing a spike and slab prior (Mitchell and Beauchamp, 1988) on the vector of regression coefficients, $\boldsymbol{\beta}^{(q)}$. To indicate whether a particular large-scale climate index, k, is useful in predicting a particular monthly rainfall quantile, q, we define an indicator variable $\gamma_k^{(q)} = 1$, if index k is useful in predicting quantile q, and $\gamma_k^{(q)} = 0$, otherwise. The vector of all indicator variables is

$\boldsymbol{\gamma}^{(q)} = (\gamma_1^{(q)}, \ldots, \gamma_K^{(q)})$ and the set of useful indices by $A_1 = \{k; \gamma_k^{(q)} = 1, k = 1 \ldots, K\}$ and define A_0 similarly.

Our prior on $\boldsymbol{\beta}^{(q)}$ is then

$$\begin{aligned} \boldsymbol{\beta}_{A_1}^{(q)} | \boldsymbol{\gamma}^{(q)}, \sigma &\sim N\left(0, c\sigma (X'_{A_1} X_{A_1})^{-1}\right) , \\ \boldsymbol{\beta}_{A_0}^{(q)} | \boldsymbol{\gamma}^{(q)} &\sim \delta(0) , \\ \beta_0^{(q)} | \sigma &\sim N\left(0, c\sigma\right) , \end{aligned} \tag{2.4}$$

where c is a constant and $\delta(0)$ denotes Dirac delta function at zero. For the prior on $\boldsymbol{\beta}_{A_1}^{(q)}$ and $\beta_0^{(q)}$, we use a modified g prior in the context of quantile regression similar to Alhamzawi and Yu, 2013 (see papers; Dao et al., 2022; Dao and Nguyen, 2024; Soomro and Yu, 2025 followed or similar approach to this for computational convenience and conjugacy) and take $c = T$. The prior on σ is an Inverse Gamma, $\mathcal{IG}(a, b)$, with parameters $a = 1$ and $b = 0.5$.

We assume $\gamma_k^{(q)}$ are i.i.d. with,

$$\begin{aligned} \gamma_k^{(q)} | \pi &\sim \text{Be}(\pi) , \\ \pi &\sim \text{Beta}(1, 1), \end{aligned} \tag{2.5}$$

where Be(π) is the Bernoulli distribution with probability π and a hyperprior on π (jamesscott), that is, a Beta distribution.

2.2 Sampling Scheme

Let $\boldsymbol{\varphi}$ be the vector of parameters of interest. This vector of parameters includes the missing observations, so that $\boldsymbol{\varphi} = (\boldsymbol{\beta}^{(q)}, \boldsymbol{y}^{miss}, \boldsymbol{\gamma}, \sigma)$, where $\boldsymbol{y}^{miss}$ denotes missing values [1], and $\boldsymbol{y}^{obs}$ represents the vector of observed values. Inference regarding $\boldsymbol{\varphi}$ is via its posterior distribution $p(\boldsymbol{\varphi}|\boldsymbol{y}, X)$. We obtain a sampling based estimate for this joint distribution, as well as for the marginal distributions of the elements of $\boldsymbol{\varphi}$, using Markov chain Monte Carlo (MCMC). From hereon, we define $\boldsymbol{y}$ to be the collection of observed and missing values of monthly rainfall. The full conditional distributions for the elements of $\boldsymbol{\varphi}$ are available and therefore we use Gibbs sampling (Gelfand, 2000; Tierney, 1994) to obtain draws from the required posterior distribution.

To facilitate a sampling scheme, we follow Kozumi and Kobayashi, 2009 and augment the data with latent variables to represent the Asymmetric Laplace distribution as a mixture of a normal, and a standard exponential distribution, (see Kotz et al., 2001). Conditional on these latent variables (2.1), can be written as

$$\boldsymbol{y} = X\boldsymbol{\beta}^{(q)} + \theta\sigma\boldsymbol{z} + \zeta\sigma\boldsymbol{u} \cdot \sqrt{\boldsymbol{z}}\,, \tag{2.6}$$

where $\boldsymbol{u} = (u_1, \ldots, u_T)'$ with $\boldsymbol{u} \sim N(0, \boldsymbol{I}_T)'$ and $\boldsymbol{z} = (z_1, \ldots, z_T)'$ with $z_t \sim \exp(1)$ and i.i.d, for $t = 1, \ldots, T$.

The likelihood, conditional on $\boldsymbol{z}, \sigma, \boldsymbol{\beta}^{(q)}$ is given by,

$$f_q(\boldsymbol{y}|\boldsymbol{\beta}^{(q)}, \boldsymbol{z}, \sigma, q) \sim N(X\boldsymbol{\beta}^{(q)} + \theta\sigma\boldsymbol{z}, \zeta^2\sigma^2 \text{diag}_T(\boldsymbol{z})). \tag{2.7}$$

Letting $\mathbf{v} = \sigma\boldsymbol{z}$, we have

[1] we assume these values are Missing Completely at Random (MCAR)

$$f_q(\boldsymbol{y}|\boldsymbol{\beta}^{(q)}, \mathbf{v}, \sigma, q) \sim N(X\boldsymbol{\beta}^{(q)} + \theta\mathbf{v}, \zeta^2\sigma\text{diag}_T(\mathbf{v})). \tag{2.8}$$

See Section **??** for more details.

Algorithm 1 MCMC scheme for Bayesian Variable Selection in QR

Initialise $\boldsymbol{\beta}^{[1]}$, $\sigma^{[1]}$, $\boldsymbol{\gamma}^{[1]}$, $\boldsymbol{y}^{miss[1]}$ by drawing from their priors. For $j = 1, ..., J$, where J indicates the number of iterations of the sampling scheme.

(1) Generate $\mathbf{v}^{[j+1]}$ from $p(\mathbf{v}|\boldsymbol{y}, \sigma^{[j]}, \boldsymbol{\beta}^{[j]})$.
(2) Generate $\sigma^{[j+1]}$ from $p(\sigma|\boldsymbol{y}, \mathbf{v}^{[j+1]}, \boldsymbol{\beta}^{[j]})$.
(3) Generate $\boldsymbol{y}^{miss[j+1]}$ from $p(\boldsymbol{y}^{miss}|\boldsymbol{y}^{obs}, \mathbf{v}^{[j+1]}, \sigma^{[j+1]}, \boldsymbol{\beta}^{[j]})$.
(4) Generate $\boldsymbol{\gamma}^{[j+1]}$ and $\boldsymbol{\beta}^{[j+1]}$ jointly from $p(\boldsymbol{\gamma}, \boldsymbol{\beta}|\tilde{\boldsymbol{y}}, \mathbf{v}^{[j+1]}, \sigma^{[j+1]}, X)$ where $\tilde{\boldsymbol{y}} = \boldsymbol{y} - \theta\mathbf{v}$.
(5) Generate $\pi^{[j+1]}$ from $p(\pi|\sigma^{[j+1]}, \mathbf{v}^{[j+1]}, \tilde{\boldsymbol{y}}, X)$.

Details of the sampling scheme are given in Algorithm 1 and the full conditional posterior distributions are given in Appendix A2. Note that for clarity of notation, we drop superscript q. We use 10000 iterations with a 2000 burn-in period for the sampling scheme.

2.3 Model for QR-SSVS

In this section, we briefly outline the Bayesian variable selection model (QR-SSVS) used in Yu et al., 2013, which is implemented in the R function SSVSquantreg in MCMCpack (Martin et al., 2011). The authors use the likelihood in Equation 2.3 but integrate over the σ,

$$f_q(\boldsymbol{y}|\boldsymbol{\beta}^{(q)}, q) \sim \{q(1-q)\}^T \exp\left\{-\sum_{t=1}^{T} \rho_q(y_t - \boldsymbol{x}_t\boldsymbol{\beta}^{(q)})\right\} . \tag{2.9}$$

The $\boldsymbol{\beta}^{(q)}$'s are assumed to be independent apriori with the following mixture of the prior distribution,

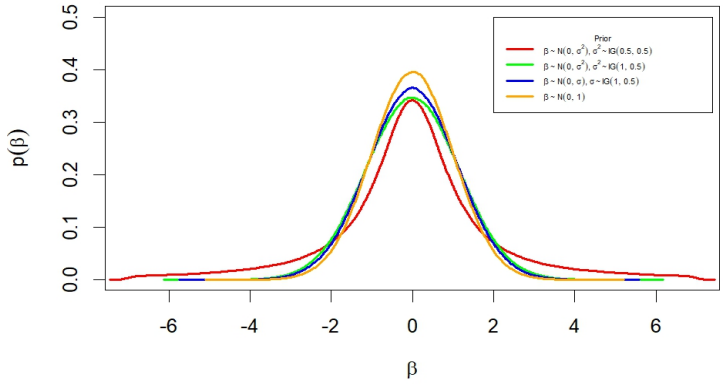
$$\begin{aligned} \beta_k^{(q)}|\gamma_k^{(q)} = 1 &\sim N\left(0, \lambda_k^{-1}\right) , \\ \beta_k^{(q)}|\gamma_k^{(q)} = 0 &\sim \delta(0) , \\ \beta_0^{(q)} &\sim N\left(\mu_0, \lambda_0^{-1}\right) , \end{aligned} \tag{2.10}$$

with μ_0 and λ_0 are both fixed.

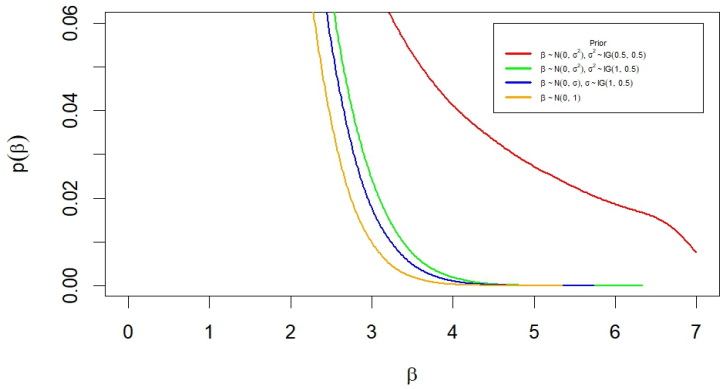
$$\begin{aligned} \lambda_k &\sim \text{Gamma}(\frac{1}{2}, \frac{1}{2}), \\ \gamma_k^{(q)} &\sim \text{Be}(\pi), \\ \pi &\sim \text{Beta}(1, 1). \end{aligned} \tag{2.11}$$

For details of the conditional posterior distribution of the parameters and the sampling scheme, refer to the QR-SSVS algorithm of the paper by Yu et al., 2013. For more details see the doctoral thesis by Reed, 2011.

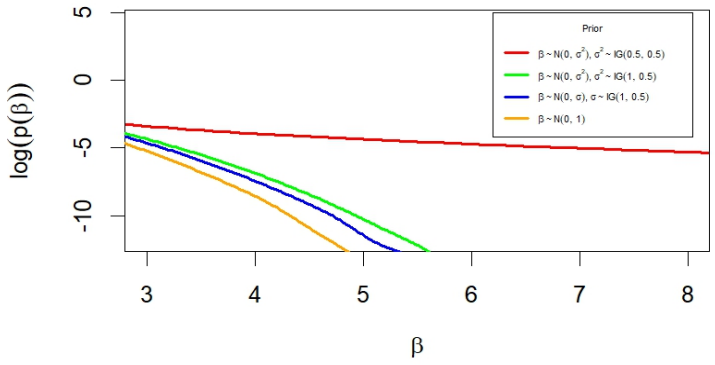
Our approach and that of Reed, 2011 exhibit notable similarities, but there are some differences between the two methods; Yu et al., 2013 integrates out the σ that induces a standard Cauchy prior marginally on the regression coefficients. Our approach treats σ as an unknown parameter, uses an Inverse Gamma ($\mathcal{IG}$) prior, and estimates its value within the sampling scheme. Our method takes into account the correlation among covariates, as it incorporates the $X'X$ matrix within the prior structure of the regression coefficients. Implementation of Yu et al., 2013, use the prior structure in Equation 2.10 for the regression coefficients in the SSVSquantreg() function in R, however, we note that their paper considers incorporating correlation among covariates to the prior structure in regression coefficients.



(a)



(b)



(c)

FIGURE 2.1: Panel (a) shows marginal distributions of β where the prior on β is, red: $\beta \sim N(0, \sigma^2)$ where $\sigma^2 \sim \mathcal{IG}(0.5, 0.5)$, yellow: standard normal, blue: $\beta \sim N(0, \sigma)$ where $\sigma \sim \mathcal{IG}(1, 0.5)$, and green: $\beta \sim N(0, \sigma^2)$ where $\sigma^2 \sim \mathcal{IG}(1, 0.5)$, while panel (b) shows the tail region in a linear scale, and panel (c) presents the tail region in a log scale.

Figure 2.1 illustrates the impact of different prior specifications on the marginal distribution of β which is given by, $p(\beta) = \int p(\beta|\sigma)p(\sigma)d\sigma$. Panel (a) compares different priors. The standard Cauchy prior (QR-SSVS method, red) exhibits heavy tails, allowing for large parameter values and accommodating potential outliers, but can be problematic due to its non-integrability (Gelman et al., 2003; Gelman et al., 2008). The proposed approach, where $\beta \sim \mathcal{N}(0, \sigma)$ with $\sigma \sim \mathcal{IG}(1, 0.5)$ (blue), has a higher peak and shorter tails compared to $\beta \sim \mathcal{N}(0, \sigma^2)$ with $\sigma^2 \sim \mathcal{IG}(1, 0.5)$ (green). Panels (b) and (c) highlight the tail behavior of these distributions. The QR-SSVS prior retains more probability mass in extreme values, allowing for larger β magnitudes, while hierarchical shrinkage priors, such as the normal-inverse gamma formulation, impose structured regularization, leading to a faster decay in tail probability and reducing the influence of extreme values (Ghosh, 2009; Polson and Scott, 2010). The logarithmic density plot in panel (c) further illustrates this effect, emphasizing the role of prior choice in controlling large coefficient values and improving posterior stability.

The specification of priors in Bayesian regression has been extensively studied, particularly in the context of balancing flexibility and regularization. Traditional choices, such as Gaussian priors, enforce strong shrinkage, promoting stable estimation but potentially over-penalizing large coefficients (Griffin and Brown, 2010). In contrast, heavy-tailed priors, such as the Cauchy prior, provide robustness against outliers and allow for large effect sizes and non-integrability in MCMC methods (Gelman et al., 2008; Ghosh et al., 2015). To overcome these limitations, Bayesian statistics has shifted toward hierarchical shrinkage priors which offer adaptive regularization, selectively shrinking small coefficients while allowing relevant signals to persist (Steel and ndez, 2000; Polson and Scott, 2010).

2.4 Results

2.4.1 Simulation Study

We compare our model with the QR-SSVS reported in Yu et al., 2013 and briefly outlined in 2.3 using 100 realizations of simulated monthly rainfall data. To simulate these values, our design matrix, X, for each month, consists of the standardized monthly values of the SOI, DMI, and SAM for the period 1950 to 2022 ($T = 73$), and their two-way interactions, so that, X is a $T \times 7$ matrix (which has as a first column, $\mathbf{1}_T$). For the study, we select the design matrix for the month of June and the heat map of the correlation between the covariates is illustrated in Figure 2.2. We assign three different regression effects: "2" for strong effects, "0.2" for weak effects, and "0" for zero effects, applied to the 5th quantile ($q = 0.05$). For the median quantile ($q = 0.5$), we multiply the regression effects of $q = 0.05$ by 2, and for the 95th quantile ($q = 0.95$), we multiply the regression effects of $q = 0.05$ by 3. The intercept is set to zero across all quantiles. We chose three different values for the scale parameter, σ, to be 0.3, 1, and 5 to generate data and run the proposed model using Algorithm 2.2 and QR-SSVS model using SSVSquantreg() function[2] in R. The error terms are generated using an Asymmetric Laplace distribution to simulate the data.

[2] https://search.r-project.org/CRAN/refmans/MCMCpack/html/SSVSquantreg.html

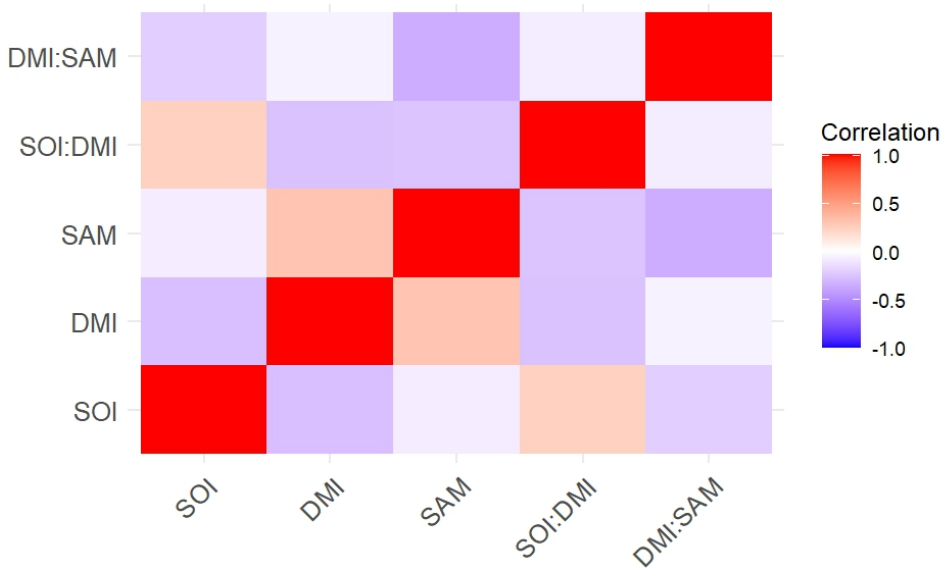


FIGURE 2.2: Correlation heat map for six covariates (climate indices and their interactions) in the simulated dataset.

For the performance metric, we assess the accuracy of the model using the mean absolute deviation (MAD) for each quantile q and realization r, which is given by:

$$\text{MAD}^{(q)r} = \frac{1}{T}\sum_{t=1}^{T} |\boldsymbol{x}_t\boldsymbol{\beta}^{(q)} - \boldsymbol{x}_t\hat{\boldsymbol{\beta}}^{(q)r}|,$$

where the superscript r denotes a specific realization, and $\hat{\boldsymbol{\beta}}^{(q)r}$ denotes the estimated posterior mean of the regression coefficients for realization r, $\boldsymbol{\beta}^{(q)r}$.

Additionally, we use the mean squared error (MSE) as in Yu et al., 2013 given by,

$$\text{MSE}^{(q)r} = \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{x}_t\boldsymbol{\beta}^{(q)} - \boldsymbol{x}_t\hat{\boldsymbol{\beta}}^{(q)r})^2.$$

Table A1.1 shows the mean differences and standard errors of the MAD values for three distinct quantiles, $q = 0.05, 0.5, 0.95$ at varying scale parameters, comparing the proposed method with the QR-SSVS method. Additionally, we report the p-values of the paired t-test, which is conducted to test the null hypothesis that there is no difference in the MAD values between the two methods for each quantile at each specified scale parameter. We assess these results at a significance level of $\alpha = 0.05$, to determine statistical differences. Similarly, Table A1.2 shows the values of the MSE metric.

The mean differences of both metrics, which represent the magnitude of the deviation between the two methods, show a varied behavior as the scale parameter increases. Initially, at $\sigma = 0.3$, the mean differences are relatively moderate but begin to decrease slightly when moving to $\sigma = 1$, suggesting a narrowing in performance differences between methods and also shows that there is no significant difference between the methods at $q = 0.5$. However, this trend reverses dramatically at $\sigma = 5$, where mean differences increase sharply, indicating a significant improvement in the performance of the proposed method under high variability conditions. Standard errors also increase with the scale parameter, indicating a reduction in the precision of the estimates as variability increases. Overall, the analysis of these results suggests that across all tested scale parameters ($\sigma = 0.3, 1, 5$), the proposed method exhibits statistically significant differences when compared to the QR-SSVS method, as denoted by p-values less than 0.001 in most instances.

To investigate the reason for these differences, an additional analysis was done using simulated data generated using the independent regressors drawn from a $\mathcal{N}(0, C)$, where $C = I$ and the results are stated in table 2.3 for MAD and table 2.4 for MSE respectively. Results show that the proposed method shows minimal differences in performance compared to

QR-SSVS at lower noise levels ($\sigma = 0.3$ and $\sigma = 1$), particularly at the median and upper quantiles. Moreover, we conduct this analysis using simulated data generated using one covariate (e.g., SOI) in the model and then using two covariates (SOI and DMI) in the model (results are in Appendix A1). The results in one covariate demonstrate similar significance with the results in table 2.3. However, the significance is changing with adding one more variable to the model and this changes markedly when the number of covariates increases to six, where the proposed method outperforms QR-SSVS across most of the conditions.

In summary, the observed performance differences may be attributed to the proposed method's handling of covariate correlations through the integration of $X'X$ in the prior of the non-zero regression coefficients, which handles the interactions and multicollinearity among predictors.

TABLE 2.1: Mean differences and standard errors of MAD for Proposed and QR-SSVS methods for the data generated using correlated regressors.

(a) $\sigma = 0.3$

Quantile	Method	Mean Difference	Standard Error	p-value
$q = 0.05$	Proposed vs QR-SSVS	-0.1821	0.0132	<0.001*
$q = 0.5$	Proposed vs QR-SSVS	-0.0113	0.0019	<0.001*
$q = 0.95$	Proposed vs QR-SSVS	-0.1098	0.0134	<0.001*

(b) $\sigma = 1$

Quantile	Method	Mean Difference	Standard Error	p-value
$q = 0.05$	Proposed vs QR-SSVS	-0.1495	0.0194	<0.001*
$q = 0.5$	Proposed vs QR-SSVS	-0.0012	0.0026	0.0631
$q = 0.95$	Proposed vs QR-SSVS	-0.0655	0.0248	0.0096*

(c) $\sigma = 5$

Quantile	Method	Mean Difference	Standard Error	p-value
$q = 0.05$	Proposed vs QR-SSVS	-4.5183	0.3069	<0.001*
$q = 0.5$	Proposed vs QR-SSVS	-0.3989	0.0352	<0.001*
$q = 0.95$	Proposed vs QR-SSVS	-4.3267	0.3470	<0.001*

Note: * denotes p-value < 0.05

TABLE 2.2: Mean differences and standard errors of MSE for Proposed and QR-SSVS methods for the data generated using correlated regressors.

(a) $\sigma = 0.3$

Quantile	Method	Mean Difference	Standard Error	p-value
$q = 0.05$	Proposed vs QR-SSVS	-0.2458	0.0211	<0.001*
$q = 0.5$	Proposed vs QR-SSVS	-0.0049	0.0014	<0.001*
$q = 0.95$	Proposed vs QR-SSVS	-0.1838	0.0281	<0.001*

(b) $\sigma = 1$

Quantile	Method	Mean Difference	Standard Error	p-value
$q = 0.05$	Proposed vs QR-SSVS	-0.1862	0.0174	<0.001*
$q = 0.5$	Proposed vs QR-SSVS	-0.0067	0.0095	0.0481
$q = 0.95$	Proposed vs QR-SSVS	-0.0491	0.1206	0.0067*

(c) $\sigma = 5$

Quantile	Method	Mean Difference	Standard Error	p-value
$q = 0.05$	Proposed vs QR-SSVS	-5.6172	0.3154	<0.001*
$q = 0.5$	Proposed vs QR-SSVS	-0.4583	0.0412	<0.001*
$q = 0.95$	Proposed vs QR-SSVS	-5.1356	0.3520	<0.001*

Note: * denotes p-value < 0.05

TABLE 2.3: Mean differences and standard errors of MAD for Proposed and QR-SSVS methods for the data generated using independent regressors.

(a) $\sigma = 0.3$

Quantile	Method	Mean Difference	Standard Error	p-value
$q = 0.05$	Proposed vs QR-SSVS	-0.1377	0.0124	<0.001*
$q = 0.5$	Proposed vs QR-SSVS	-0.0023	0.0015	0.1495
$q = 0.95$	Proposed vs QR-SSVS	-0.0108	0.0141	0.4461

(b) $\sigma = 1$

Quantile	Method	Mean Difference	Standard Error	p-value
$q = 0.05$	Proposed vs QR-SSVS	-0.1272	0.0198	<0.001*
$q = 0.5$	Proposed vs QR-SSVS	-0.0001	0.0026	0.9558
$q = 0.95$	Proposed vs QR-SSVS	-0.0379	0.0267	0.1952

(c) $\sigma = 5$

Quantile	Method	Mean Difference	Standard Error	p-value
$q = 0.05$	Proposed vs QR-SSVS	-4.8183	0.3420	<0.001*
$q = 0.5$	Proposed vs QR-SSVS	-0.4427	0.0391	<0.001*
$q = 0.95$	Proposed vs QR-SSVS	-3.9654	0.3341	<0.001*

Note: * denotes p-value < 0.05

TABLE 2.4: Mean differences and standard errors of MSE for Proposed and QR-SSVS methods for the data generated using independent regressors.

(a) $\sigma = 0.3$

Quantile	Method	Mean Difference	Standard Error	p-value
$q = 0.05$	Proposed vs QR-SSVS	-0.2051	0.0195	<0.001*
$q = 0.5$	Proposed vs QR-SSVS	-0.0027	0.0013	0.1123
$q = 0.95$	Proposed vs QR-SSVS	-0.0167	0.0308	0.5899

(b) $\sigma = 1$

Quantile	Method	Mean Difference	Standard Error	p-value
$q = 0.05$	Proposed vs QR-SSVS	-0.1972	0.0947	<0.001*
$q = 0.5$	Proposed vs QR-SSVS	-0.0305	0.0076	0.1356
$q = 0.95$	Proposed vs QR-SSVS	0.0978	0.1525	0.5227

(c) $\sigma = 5$

Quantile	Method	Mean Difference	Standard Error	p-value
$q = 0.05$	Proposed vs QR-SSVS	-5.2131	0.3221	<0.001*
$q = 0.5$	Proposed vs QR-SSVS	-0.3312	0.0292	<0.001*
$q = 0.95$	Proposed vs QR-SSVS	-4.5679	0.3156	<0.001*

Note: * denotes p-value < 0.05

2.4.2 Impact of Climate Indices on Monthly Rainfall Quantiles

We use the monthly rainfall data described in Section 1.4 as the response variable with the climate indices SOI, DMI, and SAM as covariates. Thus, the model for a given site and month at time t is given by,

$$y_t = \beta_0^{(q)} + \beta_1^{(q)}\text{SOI}_t + \beta_2^{(q)}\text{DMI}_t + \beta_3^{(q)}\text{SAM}_t + \epsilon_t. \tag{2.12}$$

We focus on the six selected stations in Figure 2.3 to illustrate the results, chosen because they represent a diverse range of geographical contexts.

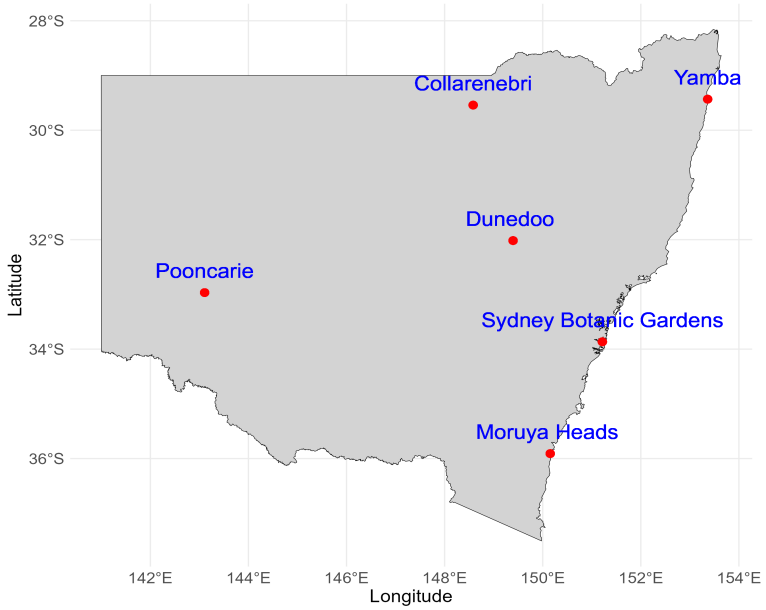


FIGURE 2.3: Map of six selected stations in NSW, Australia.

We use two metrics; the first metric is the estimated posterior mean of the probability that large-scale climate indices have an impact on rainfall, i.e., inclusion probability, given by

$$\hat{\pi}_k^{(q)} = \widehat{\mathbb{E}}(\pi_k^{(q)}|\boldsymbol{y}, X) = \frac{1}{J}\sum_{j=1}^{J} \pi_k^{(q)[j]},$$

where $\pi_k^{(q)[j]}$ are generated from $p(\pi_k^{(q)}|\boldsymbol{y}, X)$ via Algorithm 1. The second metric is the estimated posterior mean of the magnitude of the impact (size of the coefficients for the included climate indices), given that the impact exists, i.e., coefficient values, given by,

$$\hat{\beta}_k^{(q)} = \widehat{\mathbb{E}}(\beta_k^{(q)}|\gamma_k^{(q)}, \boldsymbol{y}, X) = \frac{1}{||J_1||}\sum_{j\in J_1} (\beta_k^{(q)[j]}|\gamma_k^{(q)[j]}),$$

where $J_1 = \{j; \gamma_k^{(q)[j]} = 1, j = 1, ..., J\}$ and $\beta_k^{(q)[j]}$ are generated from $p(\beta_k^{(q)}|\gamma_k^{(q)}, \boldsymbol{y}, X)$ via Algorithm 1, for each quantile q and covariate k.

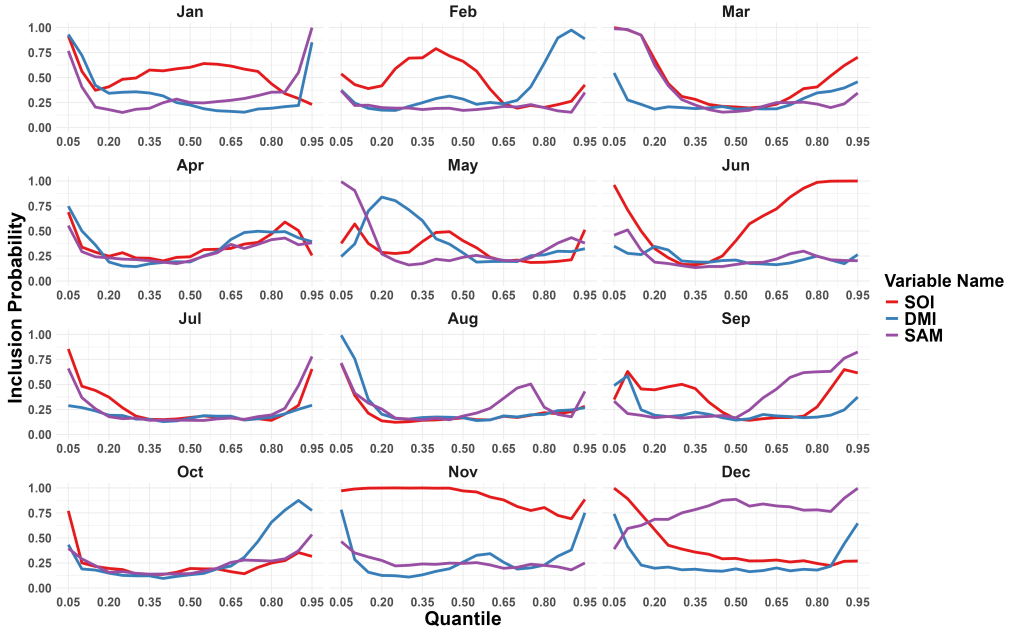
These metrics are plotted against the 19 quantile values for the six selected stations over 12 months and are illustrated in Figures 2.4 to 2.9. In the coefficient value plots, black stars indicate quantiles where the 95% credible intervals of the coefficient estimates do not include zero, denoting statistical significance.

For ease of presentation, the analysis and results for each station are described in separate subsections, followed by an overall summary subsection.

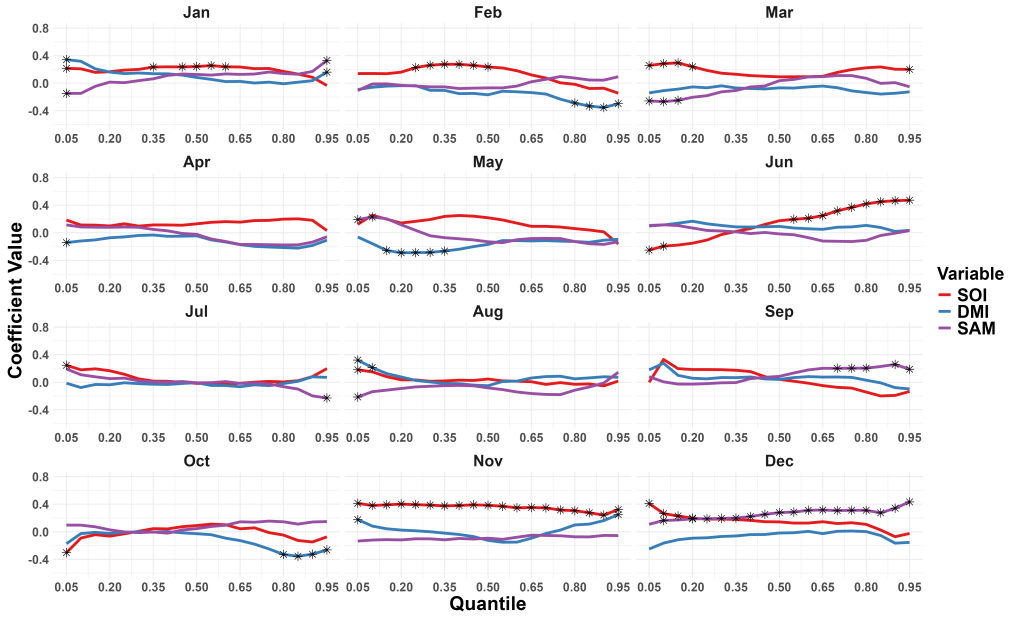
2.4.2.1 Sydney Botanic Gardens

Figure 2.4 presents the results for the Sydney Botanic Gardens, a coastal station. Panel 2.4(a) shows the dominant influences shift over the course of spring and summer, from SAM in September to DMI in October, which affects only the upper quantiles, to SOI in November, impacting the entire rainfall distribution, and then back to SAM in December. The rest of summer (January–February) experiences SOI's influence closer to the middle of the rainfall distribution. Moving into autumn (March–May), SOI and SAM contribute to drier conditions in March, though their effects are less pronounced in the wetter parts of the rainfall distribution. By May, SOI and DMI primarily affect the lower quantiles, reinforcing their association with reduced rainfall. During winter (June–August), SOI has a pronounced effect in June, where panel 2.4(b) shows that higher SOI values correspond to increased rainfall variability, leading to intensified rainfall in the upper quantiles, while lower SOI values indicate reduced variability. In particular, August shows a unique pattern in which SOI, DMI, and SAM collectively contribute to drier conditions.

The impact of DMI is generally not pronounced throughout the year, except for its influence on the upper and lower quantiles of the rainfall distribution in certain months, particularly February, May, and October. Notably, the coefficient for DMI is consistently negative in these quantiles, suggesting that positive IOD events are generally associated with a reduction in rainfall. Similarly, the impact of the SAM is less pronounced throughout the year except for notable impact in December for the majority of quantiles.



(a)



(b)

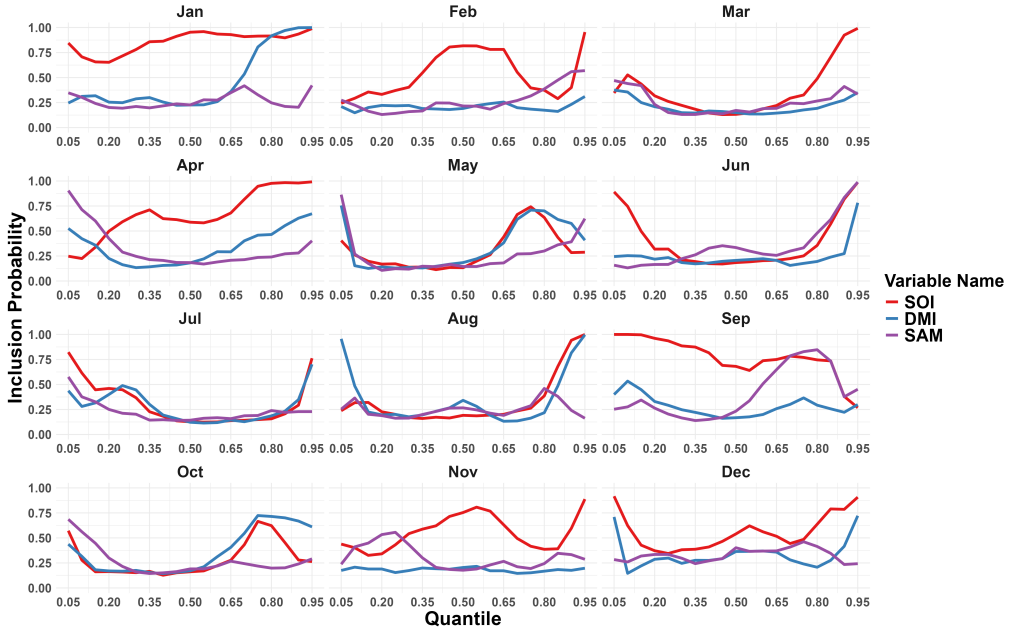
FIGURE 2.4: Panel (a) and (b) plots $\hat{\pi}_k^{(q)}$ and $\hat{\beta}_k^{(q)}$ (black star denotes the quantiles where 95% credible intervals that do not contain zero) respectively, across 12 months at different quantiles for Sydney Botanic Gardens [site 66006].

2.4.2.2 Yamba Pilot Station

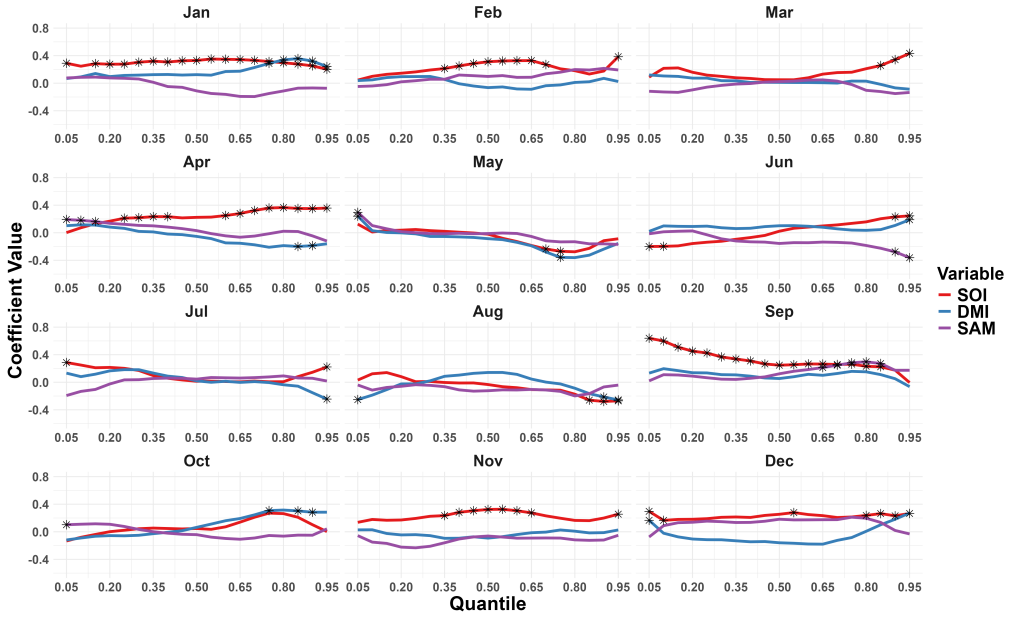
Figure 2.5 shows the results of the Yamba Pilot Station, a coastal location in northern NSW. The panel 2.5(a) shows that during the first half of the year, the impact of SOI is generally weak, except for a few notable anomalies. In January, SOI influences the entire rainfall distribution, whereas in February, its effect is limited to the middle quantiles. By April, SOI affects both the lower and upper parts of the distribution, a pattern that reemerges in June, suggesting an increased variability in extreme rainfall events during these months.

The panel 2.5(b) further illustrates that higher SOI values are generally associated with more intense rainfall, with the exception of June and August, where this relationship weakens. During the second half of the year, SOI's influence becomes less pronounced, except for distinct anomalies in September and November. In September, SOI exhibits a positive association in most quantiles, whereas in November this positive association is confined to the middle of the rainfall distribution.

The impact of DMI and SAM remains relatively weak throughout the year, with only a few exceptions. The most notable influence of SAM occurs in June, where it significantly affects the upper quantiles, indicating its role in shaping extreme rainfall events during this period.



(a)



(b)

FIGURE 2.5: Panel (a) and (b) plots $\hat{\pi}_k^{(q)}$ and $\hat{\beta}_k^{(q)}$ (black star denotes the quantiles where 95% credible intervals that do not contain zero) respectively, across 12 months at different quantiles for Yamba Pilot Station [site 58012].

2.4.2.3 Moruya Heads Pilot Station

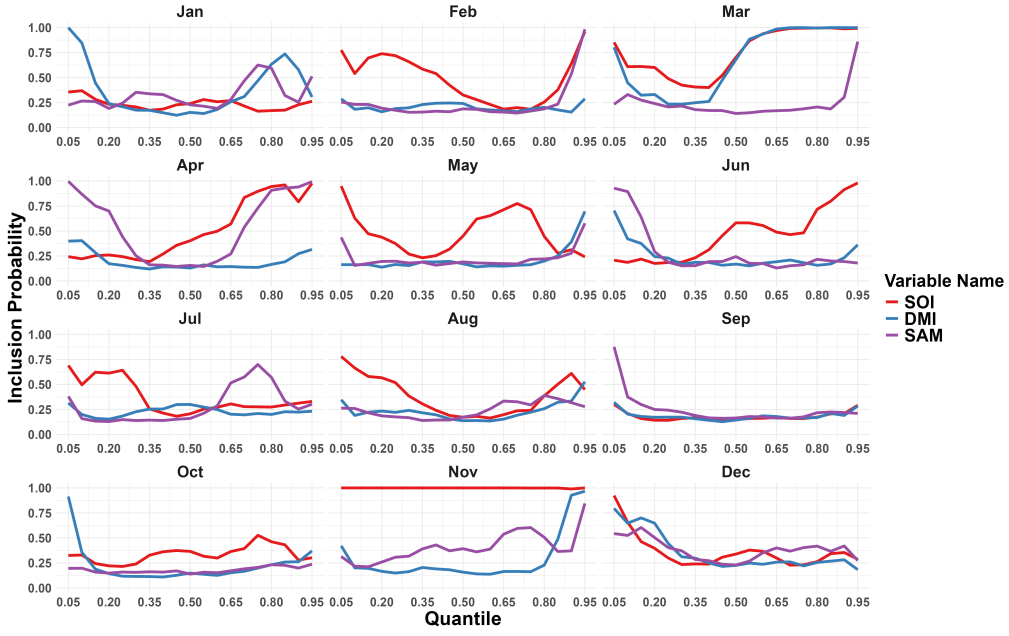
Figure 2.6 shows the results of the Moruya Heads Pilot Stations, a coastal location in southern NSW. Similarly to Sydney Botanic Gardens, the Panel 2.6(a) shows that the impact of SOI is more pronounced during the first half of the year. The panel 2.6(b) further highlights that higher SOI values are consistently associated with more intense rainfall during the first six months of the year.

In the second half of the year, SOI's influence weakens, except for a notable anomaly in November, similar to the pattern observed at Sydney Botanic Gardens. As shown in panel 2.6(b), SOI exhibits a positive association in most quantiles in November, suggesting its dominant role in shaping the rainfall distribution during this month.

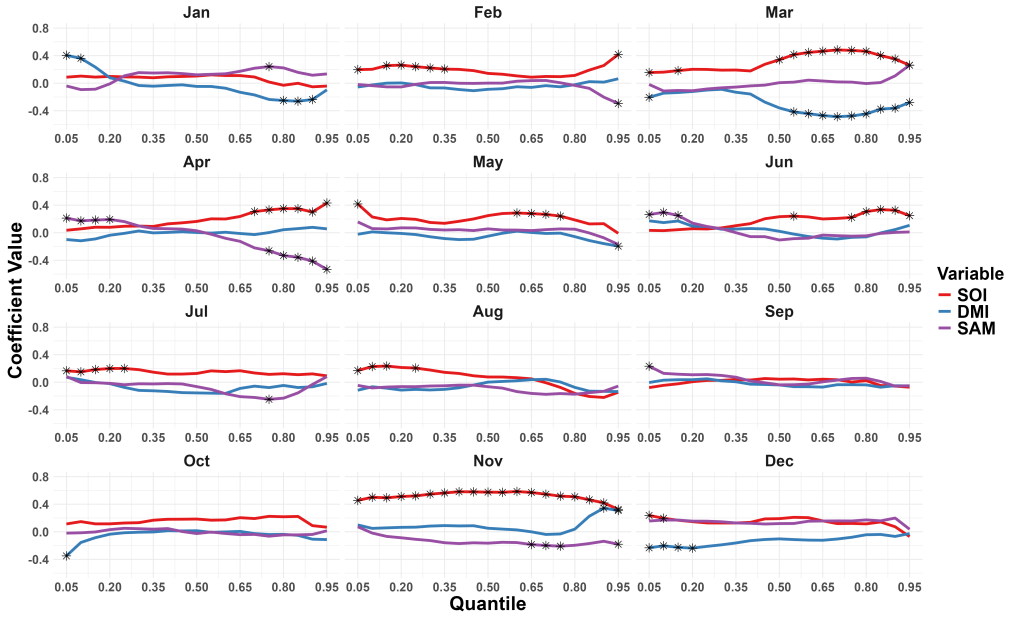
The impact of DMI remains weak throughout the year, except in March, where low DMI values are associated with reduced variability in rainfall. SAM also shows minimal influence for most months, with the exception of April, where it plays a significant role alongside SOI. In April, both positive SOI (La Niña) and negative SAM contribute to wet conditions, reinforcing their combined effect on increasing rainfall. However, SAM alone influences dry conditions, highlighting its asymmetric role in shaping rainfall extremes.

A particularly strong signal is observed in March, where an inclusion probability close to 1 indicates a significant influence of both SOI and DMI on the upper part of the rainfall distribution, suggesting that these climate drivers play a key role in extreme rainfall events during this month.

Additionally, as shown in Panel 2.6(b), high SAM values are associated with reduced rainfall variability, whereas low SAM values correspond to increased variability. This pattern suggests that SAM regulates rainfall consistency, with lower values amplifying rainfall extremes.



(a)



(b)

FIGURE 2.6: Panel (a) and (b) plots $\hat{\pi}_k^{(q)}$ and $\hat{\beta}_k^{(q)}$ (black star denotes the quantiles where 95% credible intervals that do not contain zero) respectively, across 12 months at different quantiles for Moruya Heads Pilot Station [site 69018].

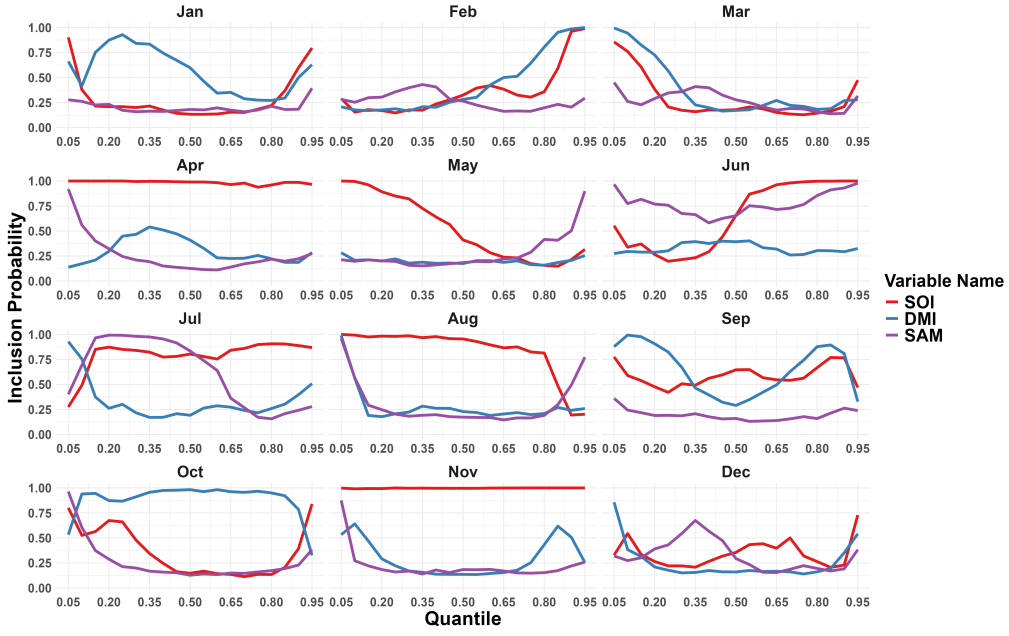
2.4.2.4 Dunedoo Post Office

Figure 2.7 shows the results for the inland station in the northern Murray-Darling Basin, Dunedoo Post Office. In the cool months (May-October), the impact of SOI is more pronounced except in October, as shown in the panel 2.7(a). Panel 2.7(b) shows that high SOI is associated with more rainfall for respective months where the impact of SOI is more pronounced. Furthermore, the influence of the SOI is much more likely to affect the full distribution of the rainfall for this location in four months of the year (April, July, August, and November).

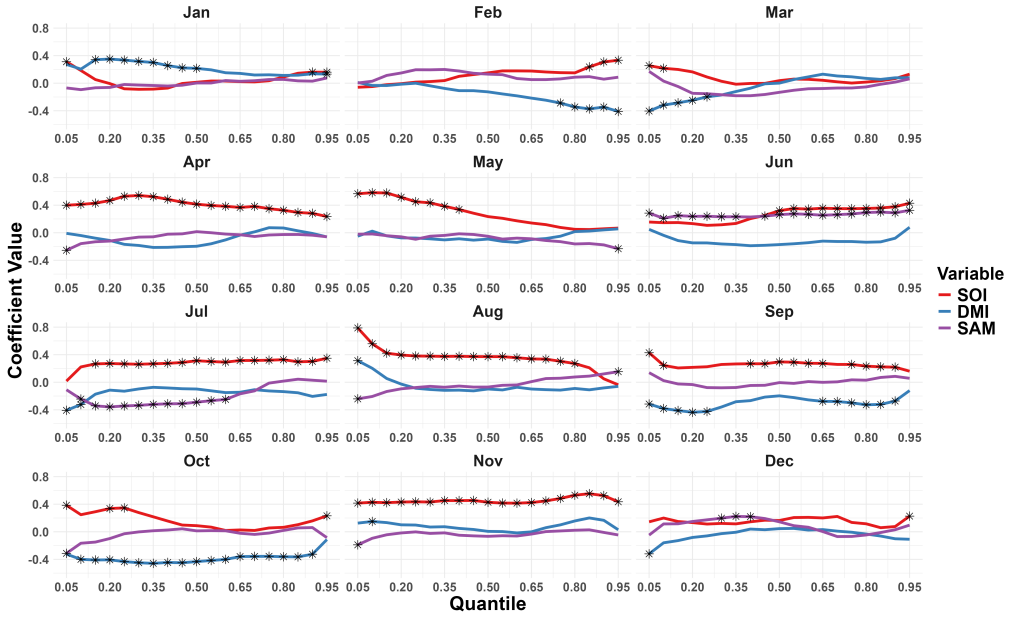
Impact of DMI is more pronounced in the spring months except for November where SOI impact majority of rainfall quantiles. Panel 2.7(b) show that the low values of DMI is associate with reduced rainfall variability in these months.

Interestingly, both SOI and DMI are more likely to influence the upper quantiles in February, showing a strong effect on wet conditions, while in March, their influence shifts towards the lower quantiles, indicating a notable impact in drier conditions.

SAM shows a less pronounced impact in most months except for June and July. The influence of SAM is more likely to affect the full distribution of rainfall for June. For July, low SAM values are associated with reduced rainfall variability as shown in panel 2.7(b).



(a)



(b)

FIGURE 2.7: Panel (a) and (b) plots $\hat{\pi}_k^{(q)}$ and $\hat{\beta}_k^{(q)}$ (black star denotes the quantiles where 95% credible intervals that do not contain zero) respectively, across 12 months at different quantiles for Dunedoo Post Office [site 64009].

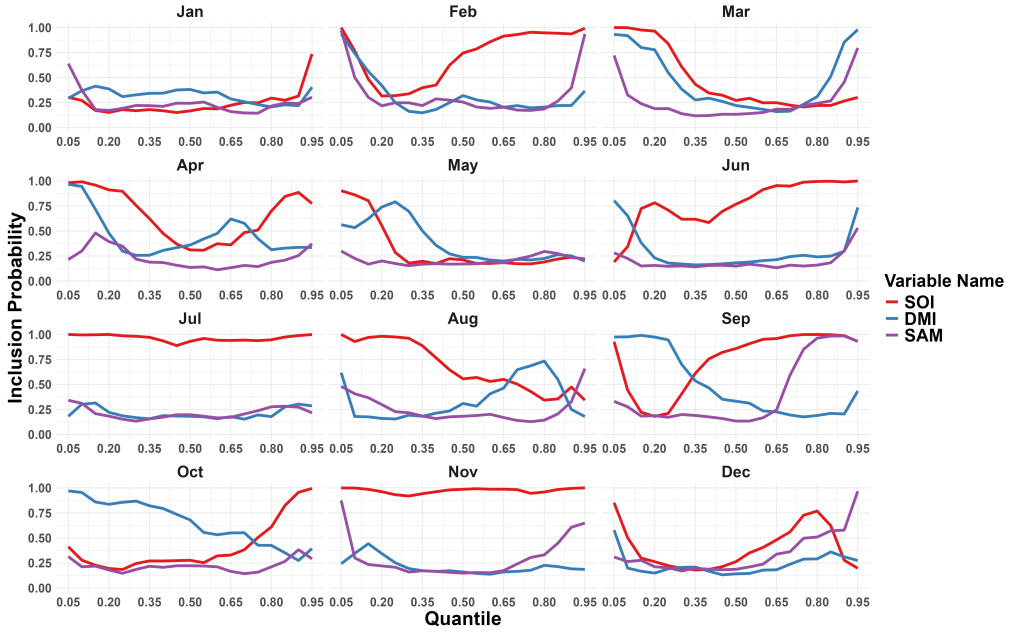
2.4.2.5 Collarenebri

In Collarenebri, an inland location in northern NSW, Figure 2.8(a) shows that the influence of SOI is more pronounced during winter and spring. In five months of the year, SOI is not only the dominant influence, but also affects the almost entire distribution of the rainfall. Panel also 2.8(b) shows that high values of SOI are associated with higher rainfall during this period.

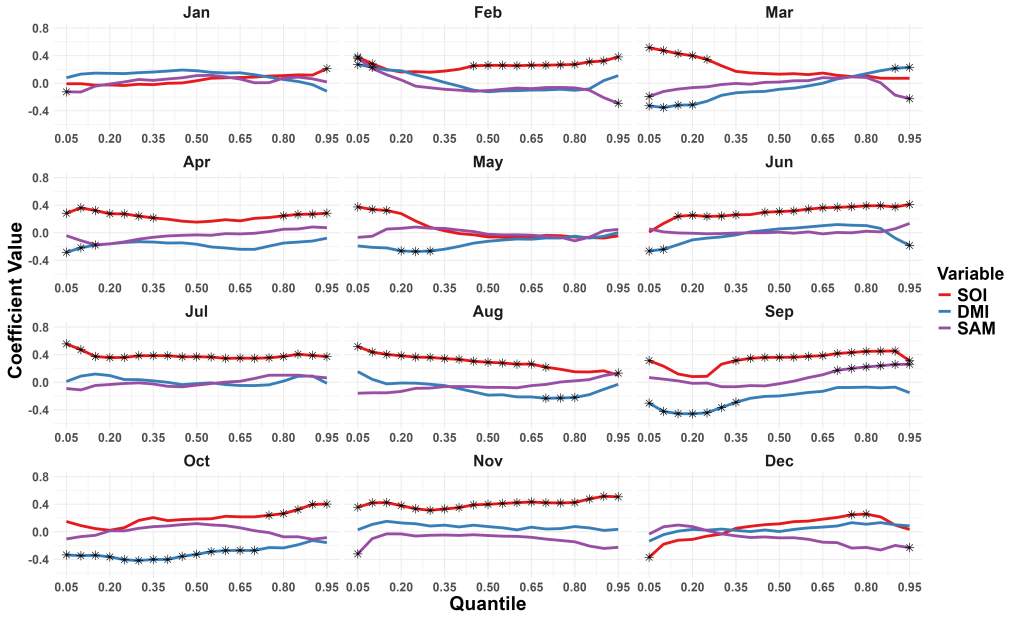
Similar to Dunedoo, the impact of DMI is more pronounced in spring months and the low values of DMI are associated with reduced rainfall variability during these months. However, an anomaly is shown in November where the influence of SOI is evident in majority of quantiles, aligning with patterns observed at other stations.

In March, the influence of both SOI and DMI is more pronounced in the lower quantiles of rainfall, aligning with the patterns observed at the Dunedoo station. Interestingly, a notable pattern emerges in September and October, where SOI primarily affects quantiles where the impact of DMI diminishes, and conversely, DMI exhibits stronger effects in quantiles where SOI's influence weakens. This inverse relationship suggests a dynamic interaction between these climate indices in modulating rainfall variability across different quantiles during these months.

The impact of SAM is less pronounced throughout the year for this particular station except for the notable influence at the upper quantiles in September.



(a)



(b)

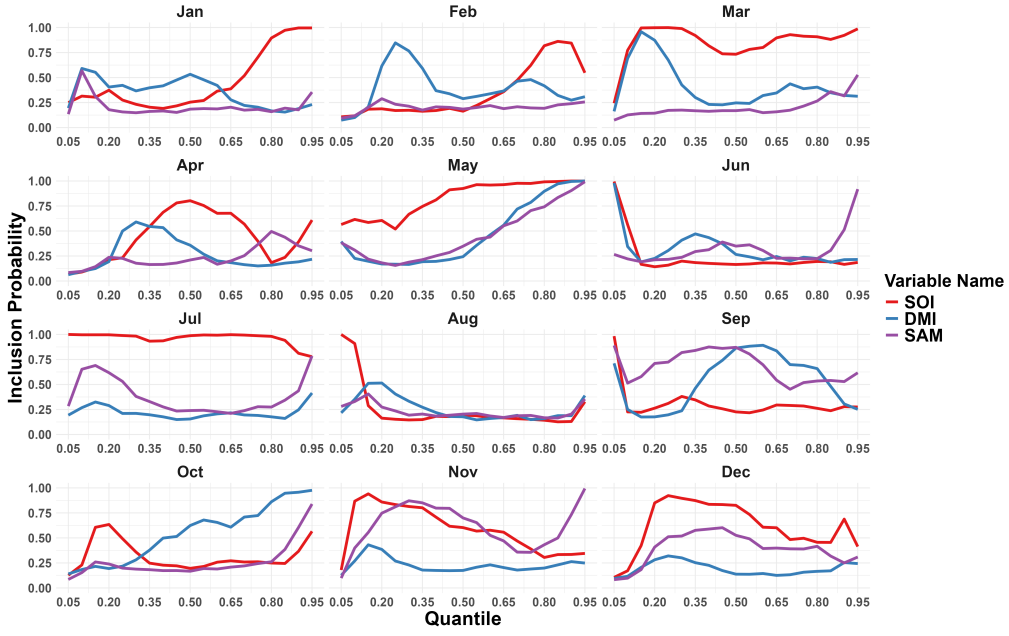
FIGURE 2.8: Panel (a) and (b) plots $\hat{\pi}_k^{(q)}$ and $\hat{\beta}_k^{(q)}$ (black star denotes the quantiles where 95% credible intervals that do not contain zero) respectively, across 12 months at different quantiles for Collarenebri (Albert St) [site 48031].

2.4.2.6 Pooncarie (Karpa Kora Station)

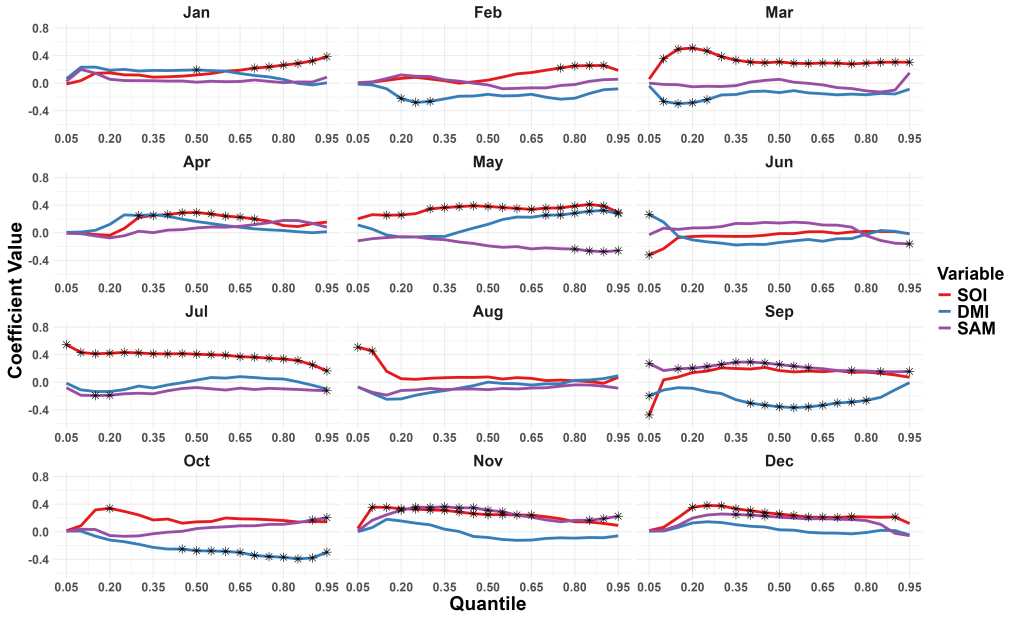
Figure 2.9 illustrates the results for Pooncarie, an inland station far from the coast. The panel 2.9(a) shows that the impact of SOI is evident during the first half of the year. However, this impact is more pronounced during the autumn months and significantly affects the entire rainfall distribution, particularly in March and May; the panel 2.9(b) also shows that high SOI values are associated with intense rainfall during these months. The impact of SOI is less pronounced in the second half of the year except for July, November, and December and tends to influence the full distribution of rainfall during these months.

As shown in the two inland stations mentioned above, the impact of DMI is more pronounced during the spring months, particularly in September and October, where panel 2.9(b) indicates that low DMI values are associated with reduced rainfall variability. Additionally, in February, the DMI impacts the lower quantiles of rainfall, suggesting a strong impact on the drier conditions.

Similarly to Collarenebri, the impact of SAM is generally less pronounced throughout the year at this station, except for the upper quantiles of rainfall in May.



(a)



(b)

FIGURE 2.9: Panel (a) and (b) plots $\hat{\pi}_k^{(q)}$ and $\hat{\beta}_k^{(q)}$ (black star denotes the quantiles where 95% credible intervals that do not contain zero) respectively, across 12 months at different quantiles for Pooncarie (Karpa Kora Station) [site 47013].

2.4.2.7 Overall Summary

Table 2.5 summarizes the dominant influence of SOI, DMI, and SAM on rainfall variability across six selected stations for different quantiles (lower, middle, and upper). The table highlights how the influence of each climate index varies across different months and quantiles at each station.

The results show distinct seasonal and spatial differences in the influence of SOI, DMI, and SAM on rainfall across coastal and inland regions. The SOI has a weaker influence on the median rainfall for coastal sites during the cooler months (May–October), except for anomalies observed in September at Yamba, the northern coastal site. This aligns with Tozer et al., 2023, who found that the impacts of SOI on mean seasonal rainfall are more pronounced in inland areas during this period. Pepler et al., 2014 attribute this to the influence of East Coast Lows (ECLs), which predominantly affect coastal regions of NSW, reducing the direct impact of SOI on rainfall in these areas. Our analysis supports these findings for median rainfall but reveals a different pattern for the lower and upper quantiles, where SOI exerts a strong influence on extreme rainfall in both inland and coastal regions. This suggests that while SOI's impact on median rainfall is less pronounced along the coast due to the influence of ECLs, its role in driving extreme rainfall variability remains significant across all sites.

In addition, Risbey et al., 2009b found a weak correlation between the mean autumn rainfall and SOI across most of NSW. Our findings reveal that SOI has a significant impact on both the lower and upper quantiles during these months, particularly at coastal sites. The relationship likely arises indirectly from the intensity and frequency of east coast lows (which are a rain-bearing phenomenon largely in autumn) being modulated by ENSO; and because ECLs bring such extreme rainfall, their presence or absence affects the tails of the distribution rather than the middle quantiles (see Favre et al., 2012; Pepler and Dowdy, 2021; Irving et al., 2024; Reid et al., 2025). This pattern differs for inland locations, such as Pooncarie, where SOI influences the entire rainfall distribution, highlighting a stronger and more consistent impact in autumn months.

Furthermore, our analysis reveals that in November, SOI consistently influences the full distribution of rainfall across nearly all sites, indicating a strong impact during this month. This aligns with previous research (Tozer et al., 2023; Risbey et al., 2009b; Pepler et al., 2014) and recent findings by Tozer et al., 2024, which highlight that La Niña-driven rainfall anomalies in November are among the strongest observed across Australia. During such events, warm SSTs around northern Australia enhance the meridional temperature gradient, strengthening the subtropical jet and promoting the development of short-wave troughs and cutoff lows, which facilitate widespread rainfall increases (Tozer et al., 2024).

The DMI exhibits a more localized and seasonally dependent influence, with its impact being stronger in the upper quantiles, particularly in the inland stations during the warmer months (November–April). This suggests that positive DMI phases, which are associated with reduced moisture transport from the Indian Ocean, may contribute to drier conditions during extreme events. In contrast, negative DMI phases, which enhance westerly moisture flux, may be responsible for increased rainfall in certain inland locations. However, its impact on the middle quantiles is less pronounced, indicating that its influence is most relevant for extreme wet and dry events rather than typical rainfall patterns.

SAM shows the weakest and most inconsistent influence across both inland and coastal stations, with its effects most notable in winter (June–July) and early summer (December). However, these impacts do not exhibit a clear seasonal or spatial pattern, suggesting that SAM's role in monthly rainfall variability in NSW is likely modulated by interactions with other climate drivers rather than acting as a dominant factor on its own (Hendon et al., 2007; Lim et al., 2016). In winter, positive SAM is often linked to drier conditions in southern and inland Australia, as it shifts the westerly wind belt poleward, reducing moisture availability (Meneghini et al., 2007; Raut et al., 2014). Additionally, SAM's influence on middle quantiles appears to be weaker at coastal sites, supporting previous findings that its effects on rainfall variability are more pronounced in inland regions (Hope et al., 2010; Montazerolghaem et al., 2016).

TABLE 2.5: Summary of the dominant effects of SOI, DMI, and SAM on rainfall across different quantiles (lower, middle, and upper) for six selected stations. Cells are light blue shaded when a climate index is dominant for that month.

(a) Lower quantiles

Station	Index	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Sydney Botanic Garden	SOI												
	DMI												
	SAM												
Yamba	SOI												
	DMI												
	SAM												
Moruya Heads	SOI												
	DMI												
	SAM												
Dunedoo	SOI												
	DMI												
	SAM												
Collarenebri	SOI												
	DMI												
	SAM												
Pooncarie	SOI												
	DMI												
	SAM												

(b) Middle quantiles

Station	Index	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Sydney Botanic Garden	SOI												
	DMI												
	SAM												
Yamba	SOI												
	DMI												
	SAM												
Moruya Heads	SOI												
	DMI												
	SAM												
Dunedoo	SOI												
	DMI												
	SAM												
Collarenebri	SOI												
	DMI												
	SAM												
Pooncarie	SOI												
	DMI												
	SAM												

(c) Upper quantiles

Station	Index	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Sydney Botanic Garden	SOI												
	DMI												
	SAM												
Yamba	SOI												
	DMI												
	SAM												
Moruya Heads	SOI												
	DMI												
	SAM												
Dunedoo	SOI												
	DMI												
	SAM												
Collarenebri	SOI												
	DMI												
	SAM												
Pooncarie	SOI												
	DMI												
	SAM												

2.4.3 Performance Measures

We measure the predictive performance of our model by comparing the empirically estimated quantile with the posterior predictive quantile. For a given month and site, let $\hat{y}^{(q)}$ and $y^{(q)*}$ be the empirically estimated quantile and the posterior predictive quantile, respectively, at quantile q. As a measure of agreement between $\hat{y}^{(q)}$ and $y^{(q)*}$, we follow Bertolacci et al., 2019 and define,

$$D^{(q)} = \hat{E}\left[\log \frac{y^{(q)*}}{\hat{y}^{(q)}} \mid \boldsymbol{y}\right].$$

Given the observed values, the predicted rainfall quantile, $y^{(q)*}$ is given by,

$$\begin{aligned} p(y^{(q)*}|\boldsymbol{y}) &= \int p(y^{(q)*}|\boldsymbol{y}, \boldsymbol{\beta}^{(q)}, \mathrm{v}, \sigma) p(\boldsymbol{\beta}^{(q)}|\sigma, \mathrm{v}, \boldsymbol{y}) p(\mathrm{v}|\sigma, \boldsymbol{y}) p(\sigma|\boldsymbol{y}) d\boldsymbol{\beta}^{(q)} d\mathrm{v} d\sigma, \\ &\approx \frac{1}{J}\sum_{j=1}^{J} N(X\boldsymbol{\beta}^{(q)[j]} + \theta^{(q)}\mathrm{v}^{[j]}, \zeta^{2(q)}\sigma^{[j]}\mathrm{v}^{[j]}), \end{aligned} \tag{2.13}$$

where $\boldsymbol{\beta}^{(q)[j]}$, $\mathrm{v}^{[j]}$ and $\sigma^{[j]}$ are draws from the joint posterior, $p(\boldsymbol{\beta}^{(q)}, \mathrm{v}, \sigma|\boldsymbol{y})$ using the sampling scheme described in Section 2.2.

When the value of $D^{(q)}$ is zero, the posterior predictive quantile is equal to the empirical quantile. Conversely, positive/ negative values of $D^{(q)}$ indicate overestimation/ underestimation of the predictive quantiles with the empirical quantiles, respectively. Higher absolute values of $D^{(q)}$ denote poorer performance in the model. Following Bertolacci et al., 2019, we define a model's performance as nominal if the interquartile range (IQR) of $D^{(q)}$ across sites contains zero. Figure 2.10 plots the median value of $D^{(q)}$ and the interquartile range (IQR) across all sites for the variable selection model, including the three covariates in Equation 2.12 where we analyze the sites independently. Here in this figure, a black point denotes

that the interquartile range (IQR) of $D^{(q)}$ across 711 sites contains zero, while a red triangle indicates that it does not.

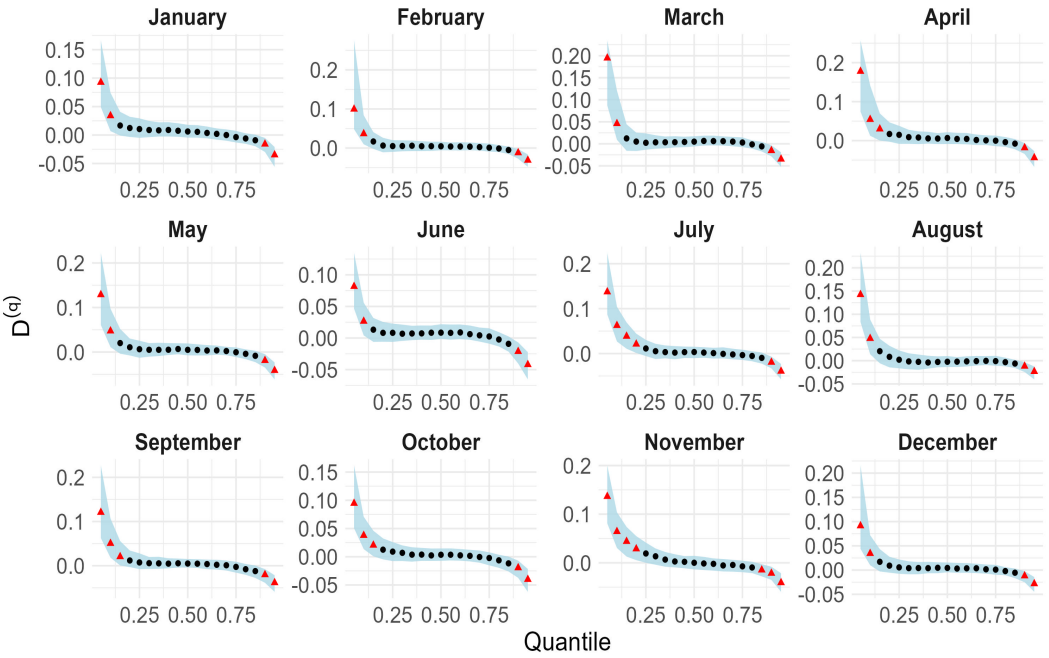


FIGURE 2.10: Median values (points) and interquartile range (shaded band) of $D^{(q)}$ across all sites, n_s, for each month and quantile. A black circle indicates that the corresponding interquartile range (i.e., the IQR of $D^{(q)}$ across sites, $s = 1, \ldots, n_s$) contains zero, while a red triangle indicates that it does not.

The performance of $D^{(q)}$ across the quantiles and months reveals clear patterns. For the middle quantiles between 0.25 and 0.75, the model shows nominal performance across all 12 months. However, in the lower ($q < 0.25$) and upper quantiles ($q > 0.75$), the model shows poorer performance, with consistent overestimation in the lower quantiles and underestimation in the upper quantiles, as denoted by the red triangles. This suggests a systematic underestimation of the overall variability in monthly rainfall. Moreover, the range of $D^{(q)}$, represented by the light blue shaded area (IQR), exhibits greater variation in the extreme quantiles of the distribution compared to the middle quantiles, indicating higher uncertainty in the models' estimates for the extremes.

This suggests that the influences detected in the extreme quantiles may, in some cases, be uncertain. This issue is particularly pronounced in coastal sites, where much of the detected influence is on extreme rainfall quantiles, rather than the well-estimated middle of the rainfall distribution.

2.5 Discussion

We developed a Bayesian variable selection model in the quantile regression framework to assess the impact of large-scale climate indices on monthly rainfall quantiles. The parameters associated with the variable selection model are estimated using a Gibbs sampling scheme. The missing values in monthly rainfall data are treated as unknown parameters and generated within the sampling scheme.

This study underscores the varying influence of SOI, DMI, and SAM on rainfall in NSW, highlighting key seasonal and regional differences. Although SOI plays a dominant role in extreme rainfall events, in both coastal and inland regions, the results show a weaker influence of SOI on median rainfall over coastal sites during the cooler months. This aligns with Pepler et al., 2014, which explains the limited ENSO influence along the eastern seaboard due to the presence of East Coast Lows (ECLs). Moreover, this does not explain the weak relationship over the northern seaboard, which has been attributed to the overriding influence of the IOD during winter (Holgate et al., 2022). Other studies (e.g., Hendon et al., 2007; Rakich et al., 2008) have found stronger relationships between eastern seaboard rainfall and SAM, particularly in summer. However, a notable exception is observed in November, when SOI consistently influences the full distribution of rainfall across nearly all sites, indicating a strong influence during this month. The DMI primarily affects extreme wet and dry conditions in inland regions during warmer months, while SAM exhibits minimal and inconsistent influence across stations, with its effects being more noticeable in winter and early summer but lacking a clear seasonal or spatial pattern.

Performance metrics show that the model demonstrates strong predictive performance in the middle quantiles, as measured by the metric proposed by Bertolacci et al., 2019. However, the model demonstrates poorer performance at both the lower and upper quantiles, indicating that its predictive ability is suboptimal in the extreme quantiles (lower and upper quantiles) of the distribution. This suggests that the model struggles to adequately capture the extreme values in the data, which are critical for accurately representing the behavior in the extreme quantiles.

In conclusion, although this model provides insights into the relationship between climate indices and the monthly rainfall distribution, performance metrics illustrate that the model fails to estimate the influence of climate drivers on extreme quantiles of rainfall accurately. Therefore, in the following chapter, we extend the model by incorporating spatial dependence to investigate whether the dependency between climate indices and rainfall quantiles exhibits spatial correlation and how this affects the predictive capabilities of the model.

CHAPTER 3

ENSO Inference on Monthly Rainfall using Bayesian Variable Selection in Spatial Quantile Regression

---

One of the dominant drivers of rainfall variability in eastern Australia is the El Niño-Southern Oscillation (ENSO), a large-scale climatic phenomenon that significantly influences the weather patterns of the region (Hauser et al., 2020; McBride and Nicholls, 1983). ENSO is a periodic climate event that involves variations in ocean temperatures and atmospheric pressures in the central and eastern tropical Pacific. It has two principal phases: El Niño, which occurs with higher-than-normal sea temperatures, and La Niña, marked by lower-than-normal sea temperatures. These phases alter global weather systems, affecting temperature, precipitation patterns, and the frequency of extreme weather events such as droughts and floods (see Verdon-Kidd and Kiem, 2009; Nicholls, 1989). The influence of ENSO extends to altering wind patterns, tropical rainfall, and ocean circulation, making it a critical factor in understanding the variability of rainfall in eastern Australia.

While extensive research has been conducted in this domain, a considerable proportion of existing studies have primarily centred their focus on understanding the impacts of these external influences on average rainfall (e.g., Holgate et al., 2022; Cai et al., 2014; Duc et al., 2017). This traditional emphasis on mean rainfall has provided valuable insights into general climate-rainfall dynamics. However, recent interest has shifted towards exploring other parts of the rainfall distribution, recognizing that extreme rainfall events and variability are critical for understanding the full spectrum of climate impacts. Accordingly, research by Evan Hajani and Ishak, 2017; Bertolacci et al., 2019; Min et al., 2013, and Sun et al.,

2015 collectively highlights the critical influence of climate phenomena such as ENSO and the Indian Ocean Dipole (IOD) on extreme rainfall patterns in Australia. These studies demonstrate that the impact of these climate indices can differ significantly across different quantiles of rainfall. These findings show the importance of using methods, such as quantile regression, to capture the varying dependency between climate drivers and monthly rainfall.

In this study, we integrate spatial dependence into the model presented in Chapter 2, thus facilitating a thorough examination of the influence of climate variability on various monthly rainfall quantiles. By incorporating Bayesian variable selection and inducing spatial dependence in quantile regression (QR), this approach allows us to understand how key drivers, such as ENSO, influence extreme rainfall events across different locations and quantiles, and how these relationships have evolved over time. Although some prior studies have utilized QR, they have often overlooked the spatial variability of climate influences, particularly in the context of Bayesian variable selection. For example, Tareghian and Rasmussen, 2013 proposes a Bayesian framework for variable selection in QR models, yet their work does not account for spatial heterogeneity in the covariates. On the other hand, Reich, 2012 introduces a spatiotemporal QR model that investigates linear trends in quantiles over time using temperature data from the United States, offering a joint modeling approach across the entire response distribution. Similarly, Duan et al., 2022 explores spatial correlations in QR, specifically in the context of Australian temperature data, but the focus remains on temperature rather than rainfall, and the analysis does not incorporate Bayesian variable selection.

Recent work by Bertolacci et al., 2019 uses a mixture model as an alternative to quantile regression. The authors use a hierarchical Bayesian mixture model to make climate inferences on daily rainfall in Australia. The effect of the covariates, along with short and long-term temporal dependencies, are modeled through a mixture of experts model. A Gaussian process prior captures the spatial dependencies of the mixture weights' parameters in the model, and its posterior distribution allows for spatially varying inferences. Following Holsclaw et al., 2017, they employ the latent variable approach developed by Polson et al., 2012. This

approach allows the conditional distributions of these latent variables and the mixture weight parameters to facilitate Gibbs sampling. However, their method does not incorporate variable selection to assess the importance of variables at different quantile levels.

Therefore, we develop a simple and more parsimonious method that incorporates Bayesian variable selection within a QR framework to capture the dependencies between not only the average but also the tails of the distribution of monthly rainfall and the Southern Oscillation Index (SOI). To achieve this, we use a hierarchical Bayesian quantile regression model; the first level of the hierarchy expresses the asymmetric Laplace distribution as a location-scale mixture (see Kotz et al., 2001; Kozumi and Kobayashi, 2009). The second level of the hierarchy expresses the prior of the regression coefficients as a two-component mixture, with components being a point mass at zero and a normal distribution. The third level incorporates spatial dependence via the prior on the regressions coefficients in two ways. First, the probability that the regression coefficient is non-zero is modeled as a probit regression where the argument of the link function varies spatially according to a Gaussian Process ($\mathcal{GP}$) prior. Second, the prior on the regression coefficients, conditional on those coefficients being non-zero, is modeled as a normal distribution, where the mean of that distribution is also varies spatially according to a Gaussian Process prior (Wahba, 1990; Wood, 2013). We utilize the Bayesian version, Wood and Kohn, 1998 of the latent variable approach developed by Albert and Chib, 1993, which allows the conditional distributions of these latent variables and the inclusion probability parameters to be efficiently sampled using Gibbs sampling.

This study validates the model using synthetic data, which is generated using the posterior estimates of the mean of the regression coefficients and the inclusion probabilities for chosen quantiles and for selected months across 711 locations in New South Wales (NSW), Australia and to further demonstrates the impact of jointly modeling all spatial sites, we compare our method with the method of analyzing sites individually. Despite significant advances in monthly rainfall modeling, there remains a scarcity of probabilistic models that provide a statistical understanding of the relation of climate phenomena to monthly rainfall quantiles

while accounting for spatial variability. Therefore, our study aims to identify the impact of SOI on various rainfall quantiles (specifically on the median, lower, and upper quantiles) at different locations and times of the year. We provide out-of-sample predictions for the impact of the SOI on monthly rainfall, as measured by the mean of the regression coefficients and the inclusion probabilities for various quantiles.

This chapter is structured as follows: section 3.1 describes the data and exploratory analysis of SOI vs Rainfall at different locations, section 3.2 outlines the model and priors, section 3.3 details the implementation of the MCMC scheme, section 3.4 presents the results for simulated data and the impact of SOI on monthly rainfall quantiles, and section 3.5 provides a summary and discussion of our conclusions.

3.1 The Data

The SOI is a metric for assessing ENSO phenomena, calculated based on the difference in sea level pressures between Tahiti and Darwin, as defined by Troup, 1965. This index is positive during La Niña events and negative during El Niño events. We utilize monthly SOI data from BOM, where the SOI for each month M is calculated as follows,

$$\text{SOI}_M = \frac{\Delta\overline{P}_M - \overline{\Delta P}_M}{sd(\Delta P_M)} .$$

Here, $\Delta\overline{P}_M$ represents the mean difference in Mean Sea Level Pressure (MSLP) between Tahiti and Darwin for month M, while $\overline{\Delta P}_M$ and $sd(\Delta P_M)$ refer to the long-term average and standard deviation of this difference for the same month, respectively. The long-term statistics are based on data from the 1933 to 1992 reference period. For more details, see Bertolacci et al., 2019 and website [1].

[1] https://www.bom.gov.au/climate/enso/soi/

From the data in section 1.4, six key stations were selected for further analysis, marked in red in Figure 3.1, and the names of the selected sites and the corresponding missing value percentage is provided in Table 3.1

TABLE 3.1: Names of Selected Stations

Station ID	Name	Missing Percentage (%)
66006	Sydney Botanic Gardens	3.53
58012	Yamba Pilot Station	0.34
69018	Moruya Heads Pilot Station	0.22
64009	Dunedoo Post Office	0.22
48031	Collarenebri (Albert St)	1.71
47013	Pooncarie (Karpa Kora Station)	0.22

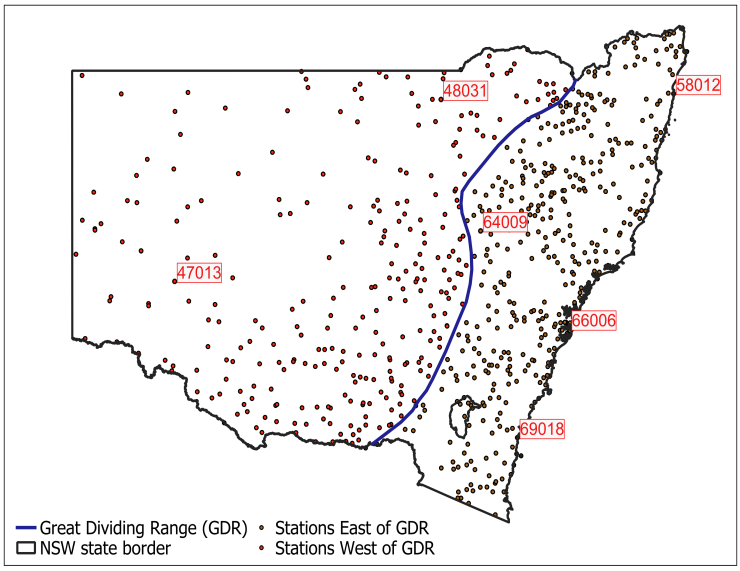
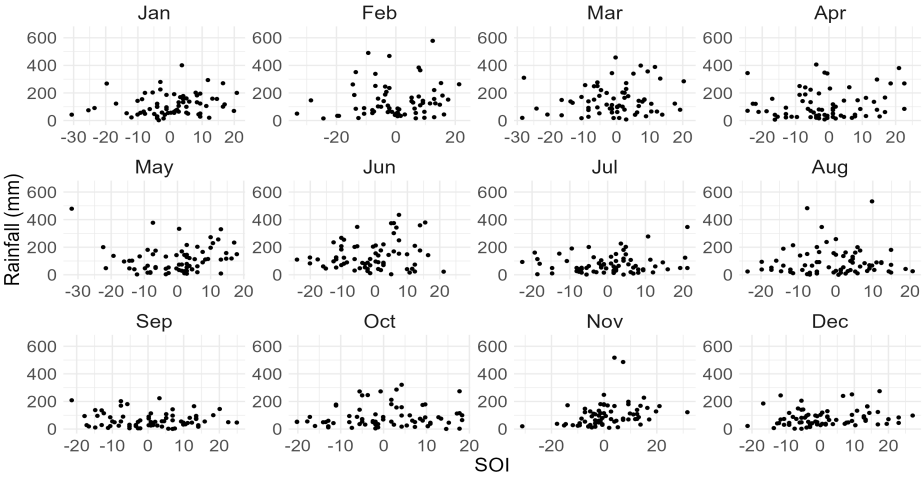


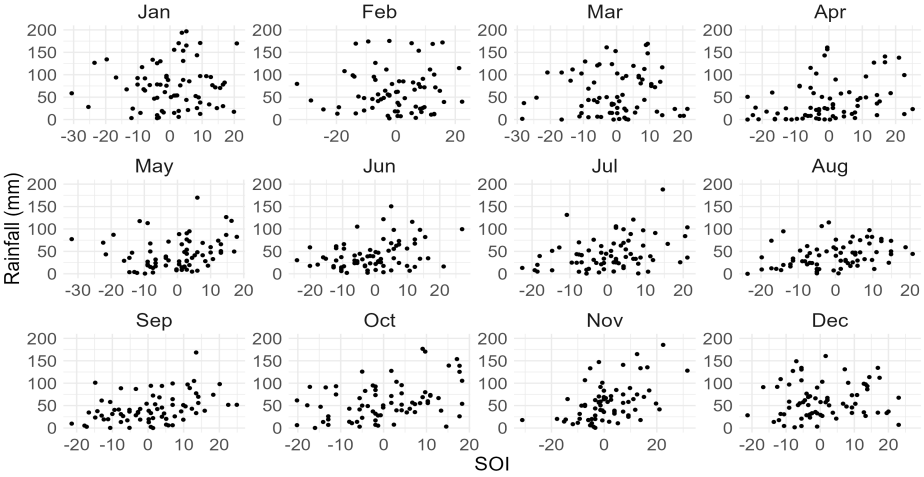
FIGURE 3.1: Locations of 711 Observational sites used in the study (Blue line: GDR boundary line, Red points: Stations west of the GDR, Brown points: Stations east of the GDR), six sites chosen to reflect different locations in NSW are marked in red.

The Figure (3.2) shows rainfall values against SOI for three stations; Sydney Botanic Garden (near the coast), Duneedo Post Office (inland), and Pooncarie (Karpa Kora Station) (far away from the coast). This figure illustrates how the distribution of rainfall varies over locations

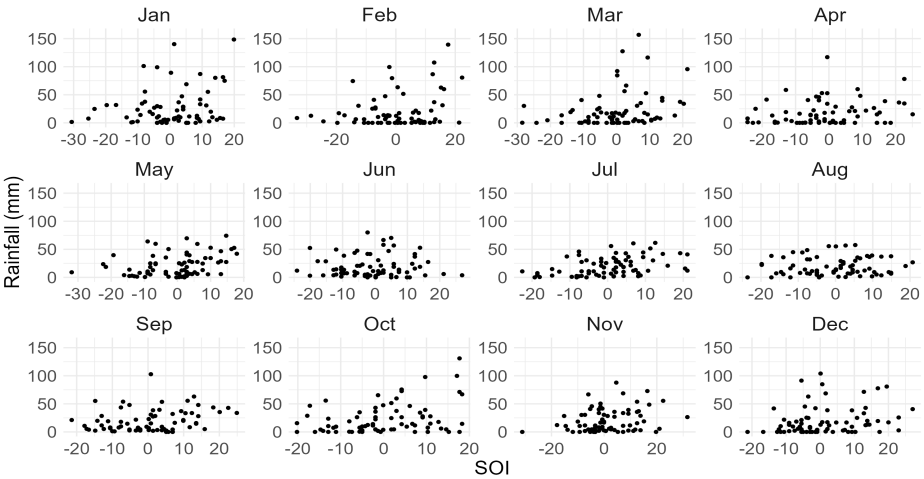
and across different months. At Sydney Botanic Garden, a coastal site, months such as January, March, and June show increased variability in rainfall with higher values of the SOI, while in months such as September, October, and December, monthly rainfall appears not be obviously impacted by the SOI. Rainfall at Dunedoo Post Office shows that the SOI's influence on rainfall is different from Sydney, a coastal location. For example, months such as October and November show the SOI has strong influence on rainfall. In contrast, months such as August show the SOI has a minimal impact on rainfall. At Pooncarie, which is in the far west of NSW, the impact of SOI on rainfall is stronger in months such as October and November, where positive SOI values correspond to higher rainfall, while winter months, such as June and July, show little impact on rainfall. This spatial and temporal variability of the dependence between rainfall and the SOI indicates the potential use of need for variable selection methods like spike and slab, to determine at which locations and in which months SOI is a relevant predictor. Additional figures for the other three stations are shown in figures B1.2, B1.3 and B1.4 in Appendix B2.



(a) Sydney Botanic Gardens [site 66006]



(b) Dunedoo Post Office [site 64009]



(c) Pooncarie (Karpa Kora Station) [site 47013]

FIGURE 3.2: Panel (a), (b) and (c) are scatter plots of SOI versus monthly rainfall for stations, Sydney Botanic Gardens, Dunedoo Post Office and Pooncarie (Karpa Kora Station) respectively .

3.2 Model and Priors

The goal of this chapter is to model the dependency between the SOI and the quantiles of monthly rainfall, and how this dependency varies across space. We do this by taking a hierarchical approach. The first level of the hierarchy, the likelihood, models the impact of the SOI on the quantiles of monthly rainfall, using observations on monthly rainfall and monthly SOI measurements. Subsequent levels of the hierarchy describes how this impact varies spatially.

3.2.1 Likelihood

Suppose we have observations of rainfall for a given month (e.g., January, February, March etc.) and given location i, measured over years $t = 1, \ldots, T$ denoted by $\boldsymbol{y}_i = (y_{1,i}, \ldots, y_{T,i})'$, with corresponding observations of SOI, $\mathbf{x} = (x_1, \ldots, x_T)'$, where x_t is the SOI in year t for a given month. Following Yu and Moyeed, 2001, we express the q^{th} quantile of monthly rainfall at time t, conditioned on x_t, as

$$\boldsymbol{y}_i = X\boldsymbol{\beta}_i^{(q)} + \boldsymbol{\epsilon}_i, \tag{3.1}$$

where $X = (\boldsymbol{x}_1', \ldots, \boldsymbol{x}_T')'$, with $\boldsymbol{x}_t = [1, x_t]$, and $\boldsymbol{\beta}_i^{(q)} = (\beta_{0,i}^{(q)}, \beta_{1,i}^{(q)})'$, are the regression coefficients for the q^{th} quantile of rainfall. Each term in the error vector, $\boldsymbol{\epsilon}_i = (\epsilon_{1,i}..., \epsilon_{T,i})'$, has an Asymmetric Laplace distribution,

$$f_q(\boldsymbol{\epsilon}_i | q, \sigma_i) \sim \left\{ \frac{q(1-q)}{\sigma_i} \right\}^T \exp \left\{ \frac{-\sum_{t=1}^T \rho_q(\epsilon_{t,i})}{\sigma_i} \right\}, \tag{3.2}$$

where $0 < q < 1$, and $\sigma > 0$, $\mathcal{AL}(0, \sigma, q)$, where $\mathcal{AL}(\mu, \sigma, q)$, denotes the asymmetric Laplace distribution, with location μ, scale σ, and skew parameter q. The check or loss function $\rho_q(\cdot)$, is given by $\rho_q(u) = u(q - I(u < 0))$, where $I(\cdot)$ denotes the indicator

function, s.t $I_{\mathcal{A}}(x) = 1$ if $x \in \mathcal{A}$ and $I_{\mathcal{A}}(x) = 0$ otherwise. Equations (3.1) and (3.2) imply $y_{t,i} \sim \mathcal{AL}(\boldsymbol{x}_t\boldsymbol{\beta}_i^{(q)}, \sigma_i, q)$ with a likelihood given by,

$$f_q(\boldsymbol{y}_i|\boldsymbol{\beta}_i^{(q)}, \sigma_i, q) \sim \left\{\frac{q(1-q)}{\sigma_i}\right\}^T \exp\left\{\frac{-\sum_{t=1}^T \rho_q(y_{t,i} - \boldsymbol{x}_t\boldsymbol{\beta}_i^{(q)})}{\sigma_i}\right\} . \tag{3.3}$$

3.2.2 Priors

We allow the impact of the SOI on monthly rainfall to vary across rainfall quantiles and space via the prior distribution of the regression coefficients. We place a "spike and slab" over these coefficients (see Mitchell and Beauchamp, 1988), where, for each quantile, q, the "slab" is a multivariate normal with a spatially varying mean, and the "spike" is a point mass at zero (see Smith and Kohn, 1996), with the probability that the regression coefficients are identically zero parameterized to be a function of space. For ease of exposition we drop the conditioning on the quantile q, but note that the parameters take different values for different quantiles.

For each quantile, q, and for locations, $i = 1, \ldots, n_s$, where n_s is the number of locations, let $\boldsymbol{\beta}_0 = (\beta_{0,1}, \ldots, \beta_{0,n_s})$ and $\boldsymbol{\beta}_1 = (\beta_{1,1}, \ldots, \beta_{1,n_s})$ be the vector of regression co-efficients corresponding to the intercepts and slopes respectively, and define $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{n_s})$ to be an indicator vector with $\gamma_i \in \{0, 1\}$ where $\beta_{1,i} = 1$ if $\gamma_i = 1$ and $\beta_{1,i} = 0$ if $\gamma_i = 0$. Let $\mathcal{A}_1 = \{i; \gamma_i = 1\}$ and define $\mathcal{A}_0$ similarly. Then the prior for $\boldsymbol{\beta}_1$ is,

$$\begin{aligned} p(\boldsymbol{\beta}_1|\mathcal{A}_1, \boldsymbol{\mu}_{\mathcal{A}_1}, \boldsymbol{\sigma}_{\mathcal{A}_1}) &= \prod_{i\in\mathcal{A}_1} p(\beta_{1,i}|\mu_i, \sigma_i) , \\ p(\beta_{1,i}|c, \mu_i, \sigma_i) &\sim \mathcal{N}\left(\mu_i, c\sigma_i(\mathbf{x}'\mathbf{x})^{-1}\right) \text{ for } i \in \mathcal{A}_1 , \\ \beta_{1,i}|\mathcal{A}_0 &\sim \delta(0), \text{ for } i \in \mathcal{A}_0 , \end{aligned} \tag{3.4}$$

while the prior for $\boldsymbol{\beta}_0$ is

$$\begin{aligned} p(\boldsymbol{\beta}_0|\boldsymbol{\sigma}) &= \prod_{i=1}^{n_s} p(\beta_{0,i}|\sigma_i) \,, \\ \beta_{0,i}|c,\sigma_i &\sim \mathcal{N}(0, c\sigma_i) \text{ for } i = 1,\ldots,n_s. \end{aligned}$$

The function $\delta(0)$ in Equation 3.4 denotes the Dirac delta function at zero. Note that we use the modified g-prior, similar to that in Zellner, 1986 and as proposed by Alhamzawi and Yu, 2013, in the context of quantile regression, on both $\beta_{1,i}$ and $\beta_{0,i}$, and take $c = T$.

The prior on σ_i is an inverse gamma prior with parameters a and b, $\mathcal{IG}(a,b)$, where $a = 1$ and $b = 0.5$, for $i = 1,\ldots,n_s$.

3.2.2.1 Prior for $\boldsymbol{\mu}$

We wish to model the mean, $\boldsymbol{\mu}$, as a flexible but smooth function of latitude, (lat) and longitude, (lon). Accordingly we express μ_i, for $i \in \mathcal{A}_1$ as a combination of linear and nonlinear components.

$$\mu_i = \mathbf{s}_i\boldsymbol{\eta} + f_\mu(\text{lat}_i, \text{lon}_i) \text{ for } i \in \mathcal{A}_1 \,, \tag{3.5}$$

where $\mathbf{s}_i = [1, \text{lat}_i, \text{lon}_i]$, $\boldsymbol{\eta} = (\eta_0, \eta_1, \eta_2)'$. We assume that apriori that the function f_μ is a Gaussian Process, $f_\mu \sim \mathcal{GP}(0, \tau^2\Omega)$ so that for every finite set of indices $i \in \mathcal{A}_1$, $\boldsymbol{f}_{\mathcal{A}_1,\mu} = \left(f_{\mathcal{A}_1(1),\mu}, \ldots, f_{\mathcal{A}_1(N),\mu}\right)$, with $N = ||\mathcal{A}_1||$, is multivariate normal,

$$\boldsymbol{f}_{\mathcal{A}_1,\mu} \sim \mathcal{N}\left(0, \tau^2\Omega_{\mathcal{A}_1}\right) \,, \tag{3.6}$$

where the cov $\left(f(\text{lat}_i, \text{lon}_i), f(\text{lat}_j, \text{lon}_j)\right) = \tau^2\omega_{ij}$ for i and $j \in \mathcal{A}_1$, is given by the reproducing kernel of a thin-plate spline Wahba, 1990. The parameter τ^2 controls the curvature of f and is called a smoothing parameter. We assume apriori that $\tau^2 \sim \mathcal{IG}(a,b)$ where $a = 1$ and $b = 0.5$.

For computational speed, we follow Wood, 2013 and approximate $\boldsymbol{f}_{\mathcal{A}_1,\mu}$ as a linear combination of truncated basis functions. This is done by taking an eigenvalue decomposition of $\Omega_{\mathcal{A}_1}$ so that $\Omega_{\mathcal{A}_1} = QDQ'$ and setting $\boldsymbol{f}_{\mathcal{A}_1,\mu} \approx H_{\mathcal{A}_1}\boldsymbol{\phi}$, where $H_{\mathcal{A}_1} = Q_{1:p}D_{1:p}^{1/2}$, and letting $\boldsymbol{\phi} \sim N(0, \tau^2 I_p)$, where $Q_{1:p}$ and $D_{1:p}$ correspond to the largest p eigenvectors and eigenvalues respectively.

3.2.2.2 Prior for $\boldsymbol{\gamma}$

We assume γ_i for $i = 1, \ldots, n_s$ are i.i.d. with,

$$\gamma_i | \pi_i \sim \text{Be}(\pi_i),$$

where $\text{Be}(\pi_i)$ is the Bernoulli distribution with probability $\Pr(\gamma_i = 1) = \pi_i$. We parameterise π to depend on latitude and longitude using a probit link function, $\Phi(\cdot)$, corresponding to the standard normal cdf as follows,

$$\pi_i \;=\; \Phi\left(\mathbf{s}_i\boldsymbol{\alpha} + f_\pi(\text{lat}_i, \text{lon}_i)\right) \text{ for } i = 1 : n_s, \tag{3.7}$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2)'$ and f is a Gaussian process, with smoothing parameter λ^2 i.e. $f_\pi \sim \mathcal{GP}(0, \lambda^2\Omega)$ so that

$$\boldsymbol{f}_{\mathcal{A},\pi} \sim \mathcal{N}\left(0, \lambda^2\Omega_{\mathcal{A}}\right), \tag{3.8}$$

where $\mathcal{A}$ is the set of indices $i = 1, \ldots, n_s$ with $\boldsymbol{f}$ approximated as a linear combination of truncated basis functions as before so that $\boldsymbol{f}_{\mathcal{A},\pi} \approx H_{\mathcal{A}}\boldsymbol{\delta}$, where $H_{\mathcal{A}} = Q_{1:p}D_{1:p}^{1/2}$, and $\boldsymbol{\delta} \sim N(0, \lambda^2 I_p)$. The smoothing parameter is $\lambda^2 \sim \mathcal{IG}(a, b)$, where $a = 1$ and $b = 0.5$. As an alternative to the probit link function, the logit link can be used, so that the Equation 3.7 can be changed as below,

$$\pi_i \;=\; \frac{\exp\left(\mathbf{s}_i\boldsymbol{\alpha} + f_\pi(\text{lat}_i, \text{lon}_i)\right)}{1+\exp\left(\mathbf{s}_i\boldsymbol{\alpha} + f_\pi(\text{lat}_i, \text{lon}_i)\right)} \text{ for } i = 1 : n_s. \tag{3.9}$$

3.3 Estimation and Inference

For each month and quantile, the complete set of parameters is $\boldsymbol{\Theta} = (\boldsymbol{B}, \boldsymbol{\sigma}, \boldsymbol{\delta}^*, \boldsymbol{\phi}^*, \lambda^2, \tau^2, \boldsymbol{Y}^{\text{miss}}, \boldsymbol{\gamma})$, where $\boldsymbol{B} = (\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)$, $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_{n_s})'$ and Y^{miss} is the collection of missing values across all sites, which we treat as unknown parameters. For each site i, let $\boldsymbol{y}_i^{\text{miss}}$ be the $k_i \times 1$ vector of missing values and let $\boldsymbol{Y}^{\text{miss}} = (\boldsymbol{y}_1^{\text{miss}}, \ldots, \boldsymbol{y}_{n_s}^{\text{miss}})'$, be the $\sum_{i=1}^{n_s} k_i \times 1$ vector of missing values across all sites and similarly $\boldsymbol{Y}^{\text{obs}} = (\boldsymbol{y}_1^{\text{obs}}, \ldots, \boldsymbol{y}_{n_s}^{\text{obs}})'$ be the vector of observed rainfall values. From hereon we define $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{n_s})'$, to be the collection of observed and missing values of monthly rainfall, so that inference regarding these parameters is made via the joint posterior distribution, $p(\boldsymbol{\Theta}|\boldsymbol{Y}, X)$. We use Markov chain Monte Carlo (MCMC) to perform the required multidimensional integration which we facilitate by using data augmentation in two ways.

(1) We follow Kozumi and Kobayashi, 2009 and represent the asymmetric Laplace distribution mixture of a normal, and a standard exponential distribution, (see Kotz et al., 2001). To do this we augment the data with latent variables, $\boldsymbol{u}_i = (u_{1,i}, \ldots, u_{T,i})'$, with $\boldsymbol{u}_i \sim N(0, I_T)$, and $\boldsymbol{z}_i = (z_{1,i}, \ldots, z_{T,i})'$ with $z_{t,i} \sim \exp(1)$ for $t = 1, \ldots, T$ and $i = 1, \ldots, n_s$.

Conditional on these latent variables (3.1), can be written as

$$\boldsymbol{y}_i = X\boldsymbol{\beta}_i + \theta\sigma_i\boldsymbol{z}_i + \zeta\sigma_i\boldsymbol{u}_i \cdot \sqrt{\boldsymbol{z}}_i \,. \tag{3.10}$$

The notation $\boldsymbol{u}_i \cdot \sqrt{\boldsymbol{z}}_i$, where $\boldsymbol{u}_i$ and $\sqrt{\boldsymbol{z}}_i$ are vectors of equal length, means the dot-product of $\boldsymbol{u}_i$ and $\sqrt{\boldsymbol{z}}_i$. Similarly, we use the notation $\sqrt{\boldsymbol{z}}_i$ to represent a vector whose elements are the square roots of the corresponding elements of the vector $\boldsymbol{z}_i$.

Letting $\mathbf{v}_i = \sigma_i \boldsymbol{z}_i$, the likelihood, conditional on $\mathbf{v}_i, \sigma_i, \boldsymbol{\beta}_i$ is given by,

$$f_q(\boldsymbol{y}_i|\boldsymbol{\beta}_i, \mathbf{v}_i, \sigma_i, q) \sim N\left(X\boldsymbol{\beta}_i + \theta\mathbf{v}_i, \zeta^2\sigma_i \text{diag}_T(\mathbf{v}_i)\right). \tag{3.11}$$

where the notation $\text{diag}_T(\mathbf{v}_i)$ refers to a $T \times T$ diagonal matrix with a $T \times 1$ vector $\mathbf{v}_i$ on the diagonal.

(2) To estimate $\pi_i = \Pr(\gamma_i = 1)$, we follow Albert and Chib, 1993; Wood and Kohn, 1998 and introduce the latent variables $\mathbf{w} = (\mathrm{w}_1, \ldots, \mathrm{w}_{n_s})$, where $\mathrm{w}_i = \boldsymbol{s}_i\boldsymbol{\alpha} + f_\pi(\text{lat}_i, \text{lon}_i) + e_i$. These continuous latent variables are connected to the indicator vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{n_s})$ by requiring that if $\gamma_i = 1$ then $\mathrm{w}_i > 0$ and if $\gamma_i = 0$, then $\mathrm{w}_i < 0$. For the alternative approach, i.e. logit link, the data augmentation method by Polson et al., 2012 can be used (see Appendix B5).

To draw $\boldsymbol{\mu}_{\mathcal{A}_1}$ it is convenient to express $\boldsymbol{f}_{\mathcal{A}_1,\mu}$ as $H_{\mathcal{A}_1}\boldsymbol{\phi}$, as in Section 3.2.2.1, and define $\boldsymbol{\phi}^* = (\boldsymbol{\eta}', \boldsymbol{\phi}')'$, $H^*_{\mathcal{A}_1} = [S_{\mathcal{A}_1}, H_{\mathcal{A}_1}]$, and $S = (\boldsymbol{s}'_1, \ldots, \boldsymbol{s}'_{n_s})'$, so that $\boldsymbol{\mu}_{\mathcal{A}_1} = H^*_{\mathcal{A}_1}\boldsymbol{\phi}^*$. Then it follows that $\boldsymbol{\phi}^* \sim \mathcal{N}(\mathbf{0}, \Psi)$, with $\Psi = \text{diag}(\boldsymbol{\tau}^{2*})$ and $\boldsymbol{\tau}^{2*} = (\boldsymbol{c}_\eta\mathbf{1}_3, \tau^2\mathbf{1}_p)$ where $\mathbf{1}_p$ is a row vectors of ones of length p. Similarly we express $\boldsymbol{\pi}$ in Equation 3.7, as $\boldsymbol{\pi} = \Phi(H^*_{\mathcal{A}}\boldsymbol{\delta}^*)$, with $\boldsymbol{\delta}^* = (\boldsymbol{\alpha}', \boldsymbol{\delta}')'$, and $\boldsymbol{\delta}^* \sim \mathcal{N}(\mathbf{0}, \Lambda)$, with $\Lambda = \text{diag}(\boldsymbol{\lambda}^{2*})$ and $\boldsymbol{\lambda}^{2*} = (\boldsymbol{c}_\alpha\mathbf{1}_3, \lambda^2\mathbf{1}_p)$. We set $c_\eta = c_\alpha = n_s$.

Algorithm 2 MCMC scheme

Input: Data $\mathcal{D}$, number of iterations J.
Output: Samples from posterior distributions i.e. $p(\boldsymbol{\Theta}|\boldsymbol{Y}, X)$.
Initialize parameters $(\boldsymbol{B}^{[1]}, \boldsymbol{\sigma}^{[1]}, \boldsymbol{\delta}^{*[1]}, \boldsymbol{\phi}^{*[1]}, \lambda^{2[1]}, \tau^{2[1]}, \boldsymbol{Y}^{\text{miss}[1]}, \boldsymbol{\gamma}^{[1]})$.
for $j = 1$ to J **do**
    Generate $\mathbf{V}^{[j+1]}$ from $p(\mathbf{V}|\boldsymbol{Y}, \boldsymbol{\sigma}^{[j]}, \boldsymbol{B}^{[j]})$ where $\mathbf{V} = (\mathbf{v}_1', \ldots, \mathbf{v}_{n_s}')'$.
    Generate $\boldsymbol{\sigma}^{[j+1]}$ from $p(\boldsymbol{\sigma}|\boldsymbol{Y}, \mathbf{V}^{[j+1]}, \boldsymbol{B}^{[j]})$.
    Generate missing values, $\boldsymbol{Y}^{\text{miss}[j+1]}$ from $p(\boldsymbol{Y}^{\text{miss}}|\boldsymbol{Y}^{\text{obs}}, \mathbf{V}^{[j+1]}, \boldsymbol{\sigma}^{[j+1]}, \boldsymbol{B}^{[j]})$.
    Generate $\mathbf{w}^{[j+1]}$ from $p(\mathbf{w}|\boldsymbol{\gamma}^{[j]}, \boldsymbol{\delta}^{*[j]})$ where $\mathbf{w}_{\mathcal{A}_1}^{[j+1]} \sim N_+(H_{\mathcal{A}_1}^* \boldsymbol{\delta}^*, 1)$ and $\mathbf{w}_{\mathcal{A}_0}^{[j+1]} \sim N_-(H_{\mathcal{A}_0}^* \boldsymbol{\delta}^*, 1)$ where the notation N_+ and N_- refer to normal distributions that are constrained to be positive and negative respectively.
    Generate $\boldsymbol{\pi} = \Phi(H_{\mathcal{A}}^* \boldsymbol{\delta}^*)$ by drawing $\boldsymbol{\delta}^{*[j+1]}$ from $p(\boldsymbol{\delta}^*|\mathbf{w}^{[j+1]}, \lambda^{2[j]})$.
    Generate $\lambda^{2[j+1]}$ from $p(\lambda^2|\boldsymbol{\delta}^{[j+1]})$.
    Generate $\boldsymbol{\mu}_{\mathcal{A}_1} = H_{\mathcal{A}_1}^* \boldsymbol{\phi}^*$ by drawing $\boldsymbol{\phi}^{*[j+1]}$ from $p(\boldsymbol{\phi}^*|\boldsymbol{\beta}_1^{[j]}, \tau^{2[j]})$ for $i \in \mathcal{A}_1$.
    Generate $\tau^{2[j+1]}$ from $p(\tau^2|\boldsymbol{\phi}^{[j+1]})$.
    Generate $\boldsymbol{\gamma}^{[j+1]}, \boldsymbol{B}^{[j+1]}$ jointly from $p(\boldsymbol{\gamma}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_0|\tilde{\boldsymbol{Y}}, \boldsymbol{\delta}^{*[j+1]}, \mathbf{V}^{[j+1]}, \boldsymbol{\sigma}^{[j+1]}, X)$ where $\tilde{\boldsymbol{Y}} = \boldsymbol{Y} - \theta\mathbf{V}$. The derivations are clearly stated in Appendix B3.
end for
return $\{\boldsymbol{B}^{[j+1]}, \boldsymbol{\sigma}^{[j+1]}, \boldsymbol{\delta}^{*[j+1]}, \boldsymbol{\phi}^{*[j+1]}, \lambda^{2[j+1]}, \tau^{2[j+1]}, \boldsymbol{Y}^{\text{miss}[j+1]}, \boldsymbol{\gamma}^{[j+1]}\}_{j=1}^{J}$.

All full conditional posteriors for Algorithm 2 are given in Appendix B3.

3.4 Results

3.4.1 Impact of SOI on Monthly Rainfall Quantiles

Results in Figure 3.3 and 3.4 show the impact of SOI on the quantiles $q = 0.05$, $q = 0.50$ and $q = 0.95$ for two metrics. The first metric is the estimated posterior mean probability that the SOI has an impact on quantiles of monthly rainfall. Figure 3.3 shows the spatial distribution of this metric given by $\hat{\pi}_k = \frac{1}{J}\sum_{j=1}^{J} \pi_k^{[j]}$, where the $\pi_k^{[j]}$ are generated as in Algorithm 2 and k is a location on a 100 by 100 grid of latitude and longitude. For details on out-of-sample prediction across this grid, see Equation B1.3 Appendix B4. This metric represents

the likelihood of *whether* SOI has a significant influence on the quantiles of the rainfall distribution. The second metric is the posterior mean of $\mu|\gamma = 1$. Figure 3.4 shows the spatial distribution of this metric given by, $\hat{\mu}_k|\gamma_{k\in\mathcal{G}_1} = \frac{1}{J}\sum_{j=1}^{J}\left(\mu_k^{[j]}|\gamma_{k\in\mathcal{G}_1}^{[j]}\right)$, where $\mathcal{G}_1 = \{k;\gamma_k = 1\}$ and the $\mu_k^{[j]}$ are generated as in Algorithm 2, see Equation B1.2 Appendix B4 for details on out-of-sample prediction. This metric represents the *mean magnitude* of the influence of the SOI on the quantiles of the rainfall distribution, if any.

Figure 3.3 shows that the SOI is most likely to influence the lowest 5th percentile of monthly rainfall- as measured by $\hat{\pi}_k$ - on those locations immediately west of the Great Dividing Range (GDR)[2], with a lesser impact on the coastal regions (east of the GDR) and for those regions in the far west of NSW. For the median monthly rainfall, Figure 3.3(b) shows a clear difference in the impact of the SOI for regions west/east of the GDR for the latter half of the year (mid-late winter, spring, and early summer). This difference is not so pronounced for the first six months. The SOI has the largest impact on rainfall west of the GDR, with a particularly pronounced difference between east and west of the GDR during the winter (June - August) and spring months (September - November), at median quantile, $q = 0.50$. This difference in impact on either side of the GDR is consistent with previous studies that have examined metrics close to the centre of the rainfall distribution such as the mean (e.g., Risbey et al., 2009b). Figure 3.3(c) also illustrates that the SOI has a significant impact on the 95th percentile of monthly rainfall during the spring months. Notably, in November, this impact is uniformly high across the entire state.

Figure 3.4 is the analogous heat map for $\hat{\mu}_k|\gamma_{k\in\mathcal{G}_1}$ across NSW. Similar to Figure 3.3, Figure 3.4 shows a significant difference in the mean magnitude of the impact of SOI on monthly rainfall between locations west and east of the GDR, with the SOI having a greater impact on monthly rainfall for locations west of the GDR than for locations east of the GDR at median quantile, $q = 0.50$. Additionally, during the months of May, July, and August, the

[2]The Great Dividing Range, commonly referred to as the Eastern Highlands, is a series of plateaus and mountain ranges that run parallel to the eastern coast of Australia, serving as a significant geographical barrier between the coast and the inland areas.

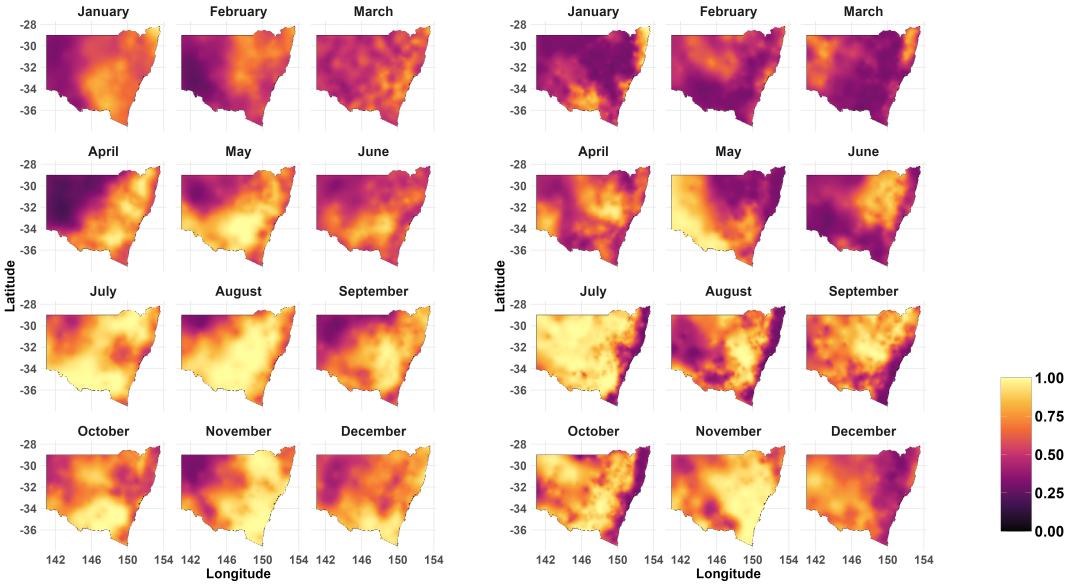
mean magnitude of SOI's impact is observed to be much higher in the lower tails ($q = 0.05$) compared to the upper tails ($q = 0.95$). This suggests that for high (La Niña-like) SOI values, the distribution of monthly rainfall becomes more concentrated around the median ($q = 0.5$) than low (El Niño-like) SOI values. Collectively Figures 3.3 and 3.4 show that the region west of the GDR experiences the largest influence from the SOI, and that the lower and upper quantile of rainfall are more sensitive to the SOI's influence rather than on the median rainfall.

To further examine the manner in which the SOI impacts the distribution of monthly rainfall on either side of the GDR, we divided the grid of locations into those locations that are west/east of the GDR and denote them by K_W and K_E respectively. We do this division by using the geographic boundary of the GDR[3], using spatial coordinates to classify locations into their respective regions as west (K_W) and east (K_E) of the GDR (see Figure 3.1). Figure 3.5(a) shows box plots of the values, $\overline{\pi}_W = \frac{1}{||K_W||}\sum_{k\in K_W}\hat{\pi}_k$, for the west (blue) and $\overline{\pi}_E = \frac{1}{||K_E||}\sum_{k\in K_E}\hat{\pi}_k$, for the east (orange), of the GDR across 12 months for quantiles $q = 0.05,\ 0.1,\ 0.2,\ 0.5,\ 0.8,\ 0.9,\ 0.95$. Figure 3.5(a) shows for the first four months of the year, January - April, the SOI is more likely to influence the tail quantiles of the rainfall distribution than the median quantile for locations both east and west of the GDR. This is an important finding given that most research, (see Risbey et al., 2009b; Chung and Power, 2017; Pepler et al., 2014) on the impact of the SOI focuses on the middle of the distribution rather than the tails, where, according to this analysis, the impact is more pronounced. Figure 3.5(a) also shows that from May to December, the impact of SOI is stronger to the west of the GDR than to the east across all quantiles, with a few exceptions. One exception is June, where the SOI east of the GDR exhibits a higher impact on upper tail quantiles compared to the west. Another exception is November where the impact of the SOI, is uniformly high for both east and west of the GDR across all quantiles. Moreover, in August, SOI impact on median rainfall is more probable in the east than on the west of GDR.

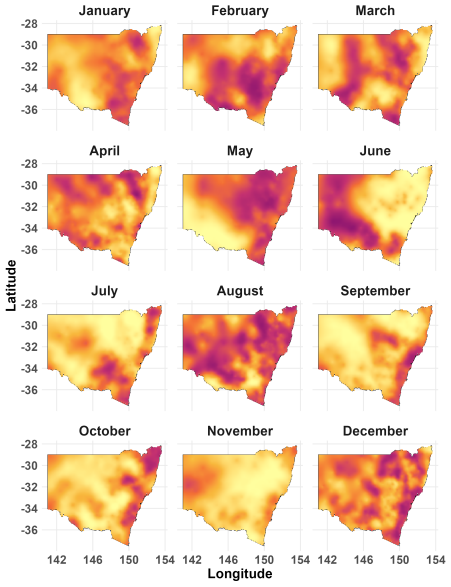
[3] https://services.ga.gov.au/gis/rest/services/NM$_Labelling_and_Boundaries/MapServer/10$

Figure 3.5(b) is an analogous plot for values, $\overline{\mu}_W = \frac{1}{||K_W||}\sum_{k\in K_W}\hat{\mu}_k$, for the west (blue) and $\overline{\mu}_E = \frac{1}{||K_E||}\sum_{k\in K_E}\hat{\mu}_k$, for the east (orange) of the GDR, and shows that from January to April this mean magnitude is constant across all quantiles and across the two regions. Figure 3.5(b) also shows that, with a notable exception, the SOI has a positive impact on rainfall, so that an increase in the SOI (La Niña-like) is associated with an increase in the values of all quantiles of the rainfall distribution. The notable exception is June, where the SOI negatively impacts the lowest quantiles of the distribution, for locations east of the GDR, positively impacts the upper quantiles of the distribution. This implies that the SOI increases the variability of monthly rainfall in June. To highlight this phenomenon, Figure 2.4(b) is a plot of the impact and magnitude of the SOI on monthly rainfall for a single station, Sydney Botanic Gardens. Figure 2.4(b) shows how the magnitude of the SOI varies sharply across the different quantiles for the month of June. Furthermore, in August, the opposite can be seen where positive impact decreases when the quantile level increases. Additional plots of the estimated posterior mean of the inclusion probabilities and the estimated posterior mean of the magnitude of the impact for selected sites are illustrated in Appendix B6.

Collectively, these results emphasize the complexity of the SOI's relationship with rainfall in NSW, highlighting the importance of detailed spatial and quantile-specific analyses to fully capture the influence of the SOI across regions and monthly rainfall quantiles. The findings reveal that the SOI has a more significant impact on lower and upper quantiles of monthly rainfall than on the median, and this influence is more pronounced in the west than in the east, particularly during the winter and spring months. These spatial differences underscore the critical role of topography, especially the GDR, in modulating atmospheric interactions associated with the SOI. Such distinctions highlight the importance of statistical analyses of climate phenomena that account for the non-uniform and seasonally variable impacts of the SOI across the rainfall distribution.

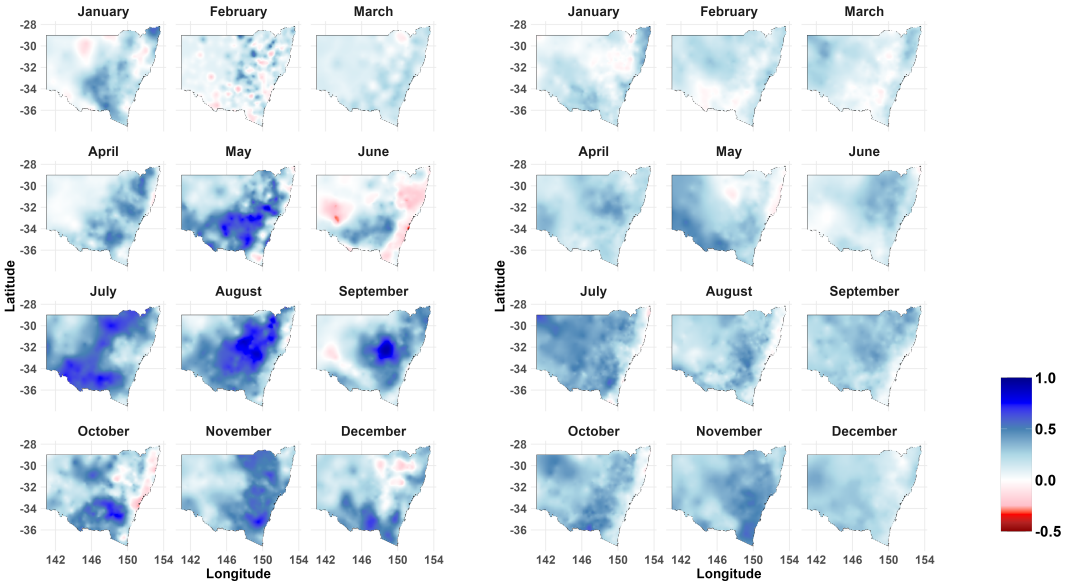


(a) $q = 0.05$ (b) $q = 0.5$

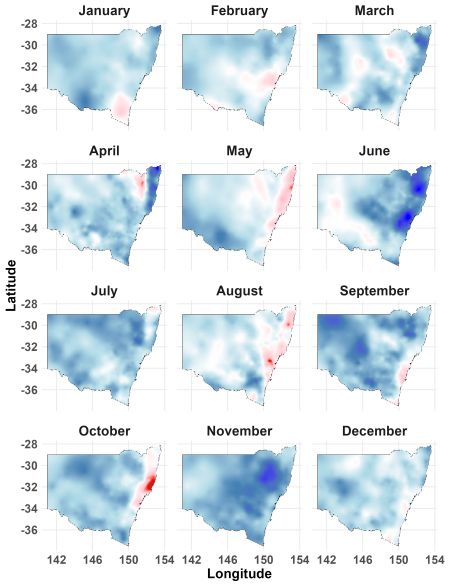


(c) $q = 0.95$

FIGURE 3.3: Panel (a), (b), and (c) show heat maps of $\hat{\pi}_k$ (representing the probability of inclusion of SOI as an influence on rainfall quantile q) across the NSW for 12 months at $q = 0.05$, $q = 0.5$, and $q = 0.95$, respectively.

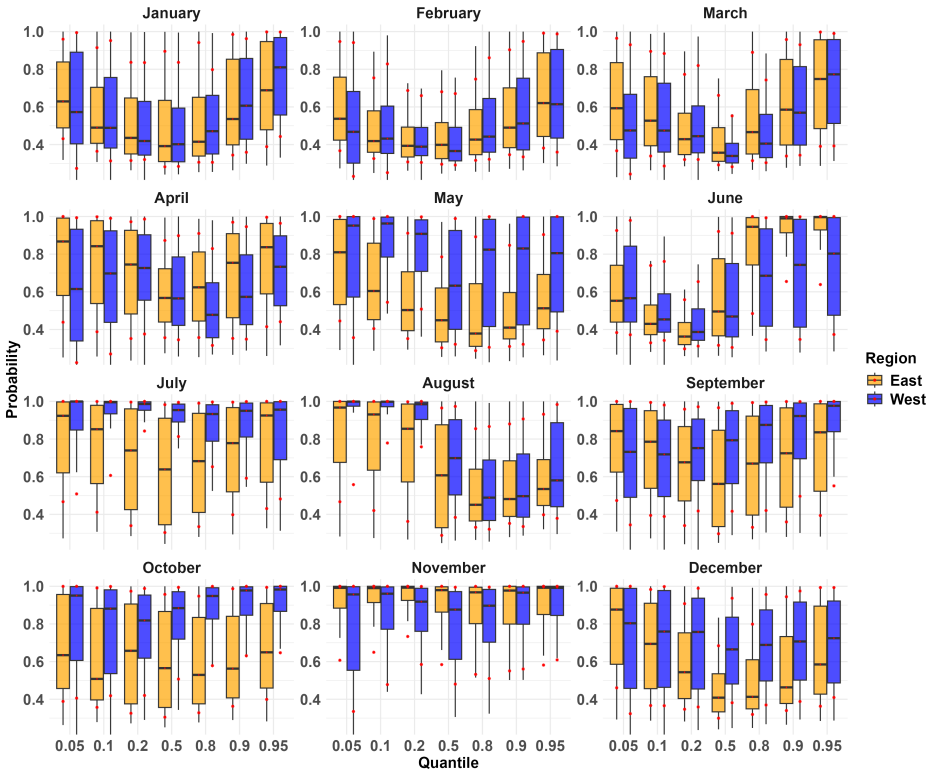


(a) $q = 0.05$ (b) $q = 0.5$

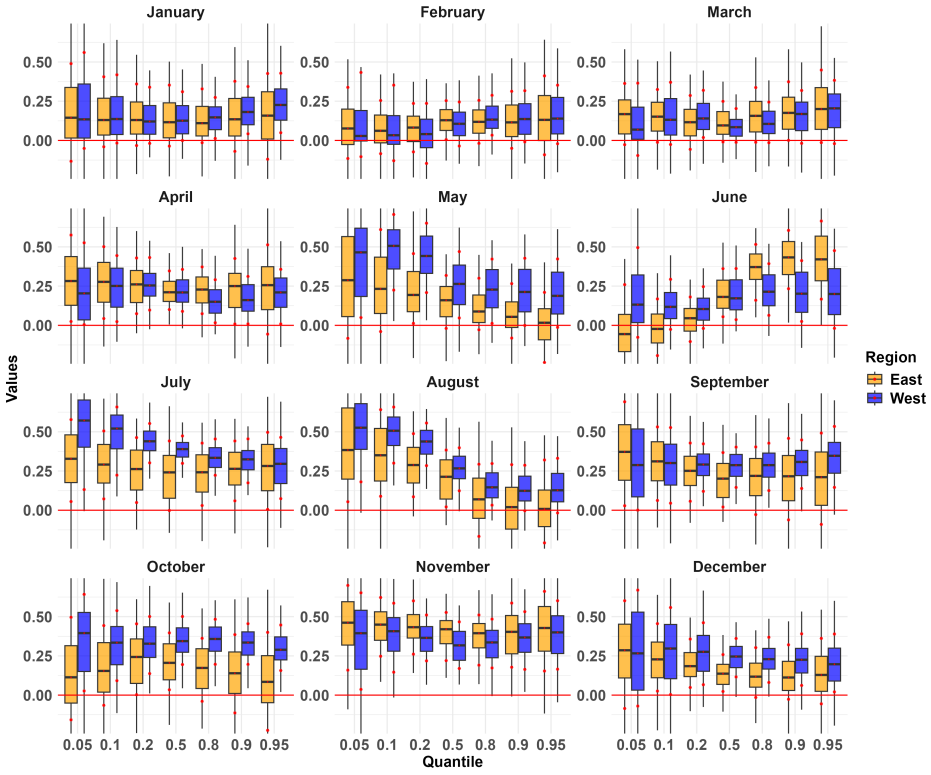


(c) $q = 0.95$

FIGURE 3.4: Panel (a), (b), and (c) show heat maps of $\hat{\mu}_k$ (representing the estimated strength of the influence of SOI on rainfall quantile q) across the NSW for 12 months at $q = 0.05$, $q = 0.5$, and $q = 0.95$, respectively.



(a)



(b)

FIGURE 3.5: Panel (a) presents box plot estimates of $\overline{\pi}_W$ and $\overline{\pi}_E$, representing the west and east regions of the GDR, respectively. Similarly, Panel (b) presents box plot estimates of $\overline{\mu}_W$ and $\overline{\mu}_E$. Red stars mark the 10th percentile, mean, and 90th percentile, and the red line in panel (b) marks zero.

3.4.2 Model Validation via Simulation Study

In this section, we show the advantages of allowing for spatial dependence in our model, versus analysing the sites separately, which would imply spatial independence, as is common practice in climate science (see Islam et al., 2017; Wu and Leonard, 2019; Jakob and Walland, 2016; Yilmaz et al., 2017). We simulate data from two data generating process (DGP), which we denote by DGP_0 and DGP_1. The data generating process DGP_0 assumes that there is no spatial dependence, which is achieved by fixing μ and π in Equations 3.5 and 3.7 to be zero and 0.5 respectively. The data generating process DGP_1 allows for spatial dependence. This is achieved by setting μ and π to be the estimated values of these parameters for the month of August, namely $\hat{\boldsymbol{\pi}}^q$ and $\hat{\boldsymbol{\mu}}^q$. We chose the month of August because the impact of SOI on August rainfall, as measured by the inclusion probability, π and the mean size of the coefficient, measured by μ varies significantly across quantiles as shown in Figure 3.5 panels (a) and (b). For example, these figures show the probability that SOI has an impact on the 5th quantile of monthly rainfall is close to 1 for both east and west regions. In contrast probability that SOI has an impact on 95th quantile of monthly rainfall is less than 0.5 for both east and west regions. A similar story is true for the size of the coefficient, where the impact of the SOI on the upper quantile in Figure 3.5 panel (b) is close to zero for August.

Using the data simulated from both DGP_0 and DGP_1 we estimate the impact of the SOI on monthly rainfall using two models. The first model we denote by $\mathcal{M}_0$, assumes that the data are spatially independent. This corresponds to fixing μ in Equation 3.5 to be zero, and fixing π in Equation 3.7 to be 0.5 for all locations. The second model $\mathcal{M}_1$ is our model as described in Section 3.2, without fixing μ and π.

For both models $\mathcal{M}_1$ and $\mathcal{M}_0$ we estimate the regression coefficients at all locations by the posterior mean, $\hat{\beta}_{1,i}^{(q)}$ given by,

$$\hat{\beta}_{1,i,l}^{(q)} = \frac{1}{J}\sum_{j=1}^{J}(\beta_{1,i,l}^{(q)[j]}|\gamma_{i,l}^{(q)[j]} = 1) \text{ for } l = 0, 1\ , \tag{3.12}$$

where $j = 1, \ldots, J$ denotes the posterior iterations after burn-in. Notations k and l denote the DGP and the $\mathcal{M}$ type respectively.

We measure performance of these two models using two metrics. First we measure the ability of each model to recover the probability of inclusion by the Average Symmetric Kullback-Leibler Divergence (ASKLD) which is given by,

$$\text{ASKLD}_{k,l} = \frac{1}{n_s}\sum_{i=1}^{n_s}\left\{\hat{\pi}_{i,l}^{(q)}\ \log\ \left(\frac{\hat{\pi}_{i,l}^{(q)}}{\pi_{i,k}^{(q)}}\right) + \pi_{i,k}^{(q)}\log\ \left(\frac{\pi_{i,k}^{(q)}}{\hat{\pi}_{i,l}^{(q)}}\right)\right\} \text{ for } k, l = 0, 1\ . \tag{3.13}$$

Note that in Equation 3.13, $\hat{\pi}^{(q)}$ is the estimated inclusion probability which depends on the model and $\boldsymbol{\pi}^{(q)}$ is the true value of inclusion probability for each DGP. For the second metric we define the distance between the regression coefficients, $\hat{\boldsymbol{\beta}}_1$ estimated by both models and the true value of $\boldsymbol{\beta}_1$ for each DGP, as the square root of the average, across all locations, of the squared differences between the estimated and true values which we compute as,

$$\text{RMSCD}_{k,l} = \sqrt{\frac{\sum_{i=1}^{n_s}(\hat{\beta}_{1,i,l}^{(q)} - \beta_{1,i,k}^{(q)})^2}{n_s}} \text{ for } k, l = 0, 1\ . \tag{3.14}$$

We label this metric as the Root Mean Squared Coefficient Deviation (RMSCD).

The boxplot estimates of RMSCD and ASKLD across the 30 realizations at the quantiles 0.05, 0.5, and 0.95, are illustrated in Figure 3.6, and show that the model, $\mathcal{M}_1$ (red) consistently outperforms the model, $\mathcal{M}_0$ (blue) when the data is generated from DGP_1, which incorporates spatial dependencies. Notably, there is a significant difference in model performance across most quantiles in this setting, with the spatial model exhibiting lower RMSCD and

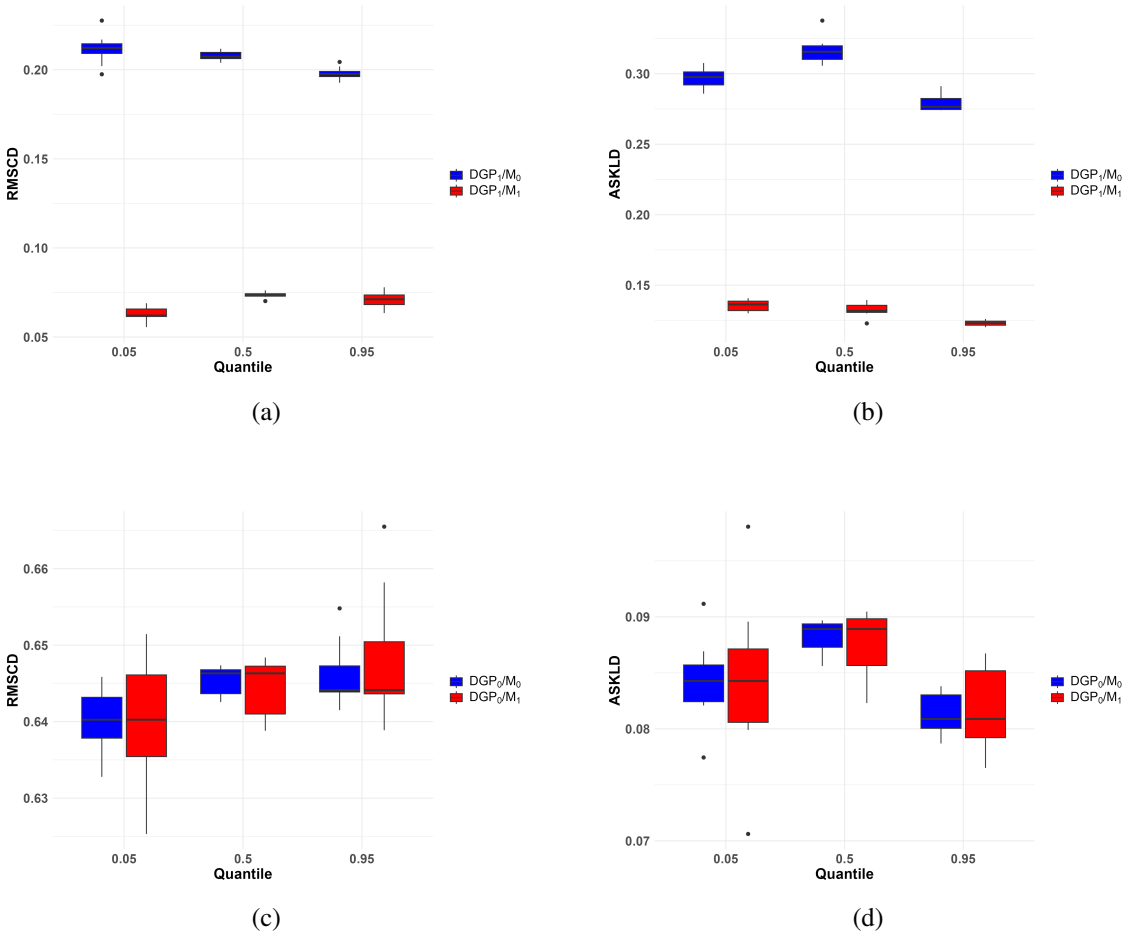


FIGURE 3.6: Panel (a) and (b) are boxplots for RMSCD and ASKLD measures for data generated under DGP_1. Similarly, panel (c) and (d) show the measures when generated from DGP_0 at $q = (0.05, 0.5, 0.95)$, color blue and red represent model $\mathcal{M}_0$ and $\mathcal{M}_1$ respectively.

ASKLD values, indicating better accuracy in estimating both the regression coefficient and the probability of inclusion. However, in DGP_0, where the data is generated without spatial dependence, panels (c) and (d) show that both models perform similarly across the quantiles. In some cases, the non-spatial model shows marginally lower RMSCD and ASKLD values for certain quantiles. In summary, model $\mathcal{M}_1$ performs significantly better than model $\mathcal{M}_0$ when the data exhibit spatial dependence, and performs on par with model $\mathcal{M}_0$ when there

is no spatial dependencies. This highlights the need to explicitly incorporate spatial dependencies in the estimation model. The additional simulation study using the estimates of the month November are stated in section B1.1 in Appendix B1.

3.4.3 Performance Measures

We use the same performance measure in chapter 2 for this section. For a given month, let $\hat{y}_i^{(q)}$ and $y_i^{(q)*}$ be the empirically estimated quantile and the posterior predictive quantile, respectively, at quantile q for site i. As a measure of agreement between $\hat{y}_i^{(q)}$ and $y_i^{(q)*}$, we follow Bertolacci et al., 2019 and define,

$$D_i^{(q)} = \hat{E}\left[\log \frac{y_i^{(q)*}}{\hat{y}_i^{(q)}} \mid \boldsymbol{y}_i\right].$$

To demonstrate the advantages of jointly modeling all spatial sites, as opposed to analyzing the sites individually, we evaluate the performance of these two statistical models. Following Bertolacci et al., 2019, we define a model's performance as nominal if the interquartile range (IQR) of $D_i^{(q)}$ across sites contains zero. Figure 3.7 illustrates the median and IQR of $D_i^{(q)}$ for two models: one which incorporates spatial dependence (red) and the other which does not (blue). The median of $D_i^{(q)}$ is represented by a circle when the model's performance is nominal, and by a triangle when it is not.

The results show that both models perform well for quantiles between the 25th and 75th percentiles across most months. However, in both models, performance is suboptimal for some of the middle quantiles in June. While the model which incorporates spatial dependence generally achieves nominal performance across all quantiles in most months, exceptions include January (90th percentile), February (95th percentile), March (75th and 80th percentiles), and August (95th percentile), where a few points in the upper tails are overestimated. Additionally, in June, the 5th percentile in the lower tails is underestimated. In contrast, the

model without spatial dependence tends to overestimate the lower tails (< 25%) and underestimate the upper tails (> 75%), with a broader range of $D_i^{(q)}$ values across most months. This increased variability highlights a lack of stability and consistency in predicting extreme quantiles. Overall, the spatially dependent model demonstrates clear advantages, offering more accurate and reliable predictions for extreme quantiles.

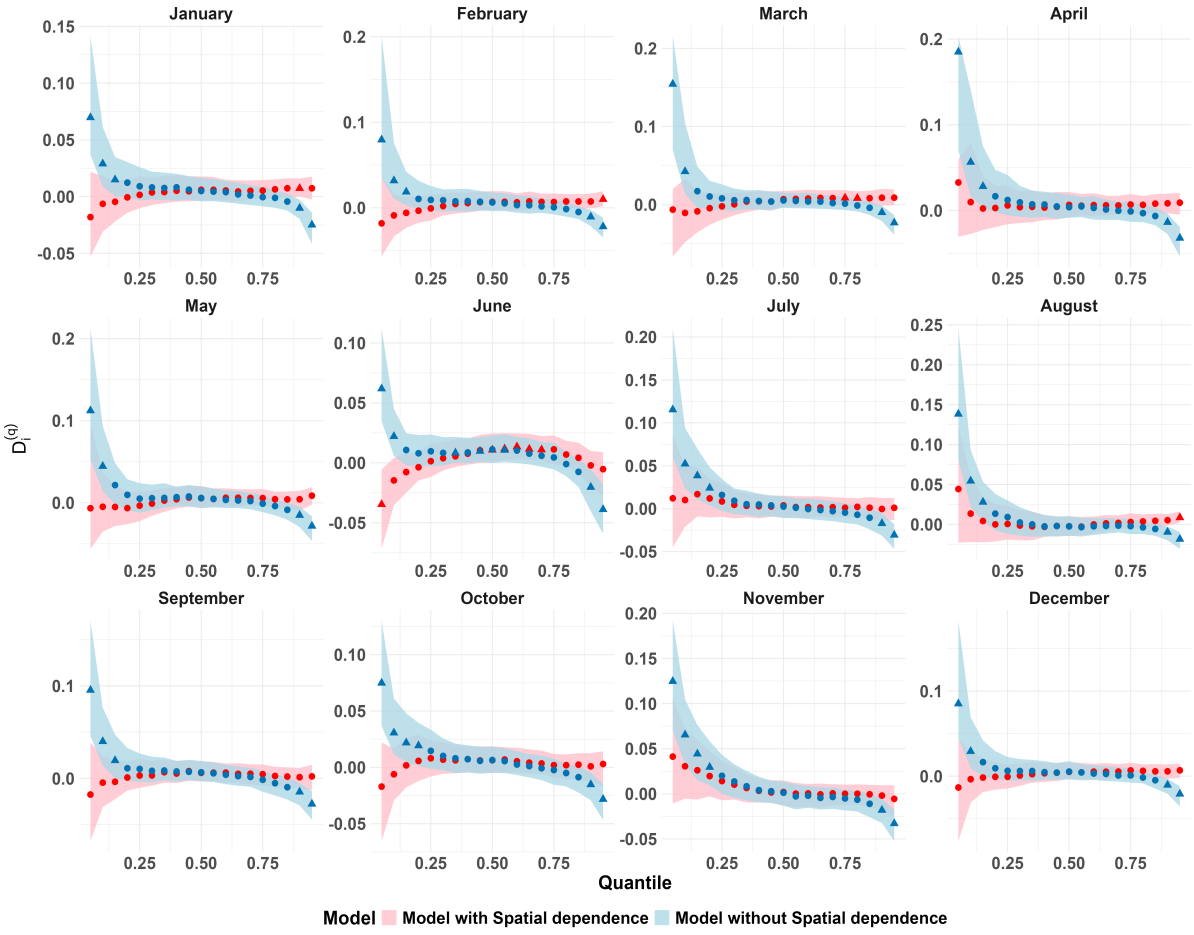
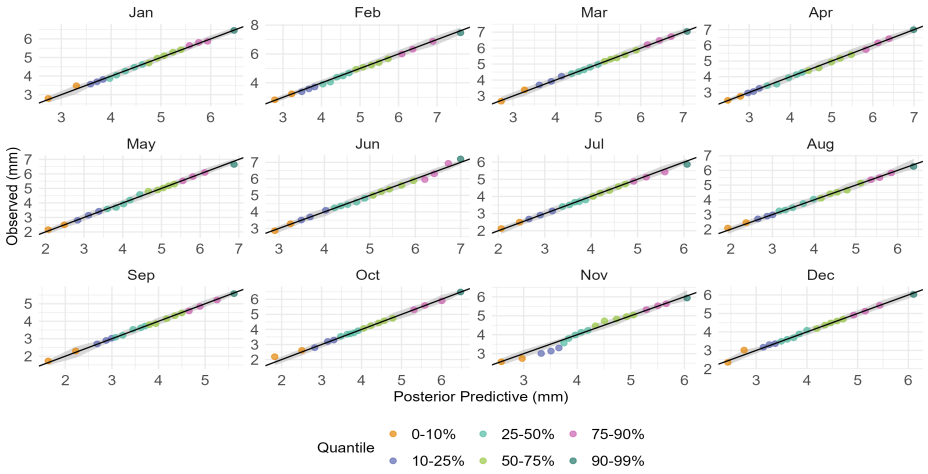
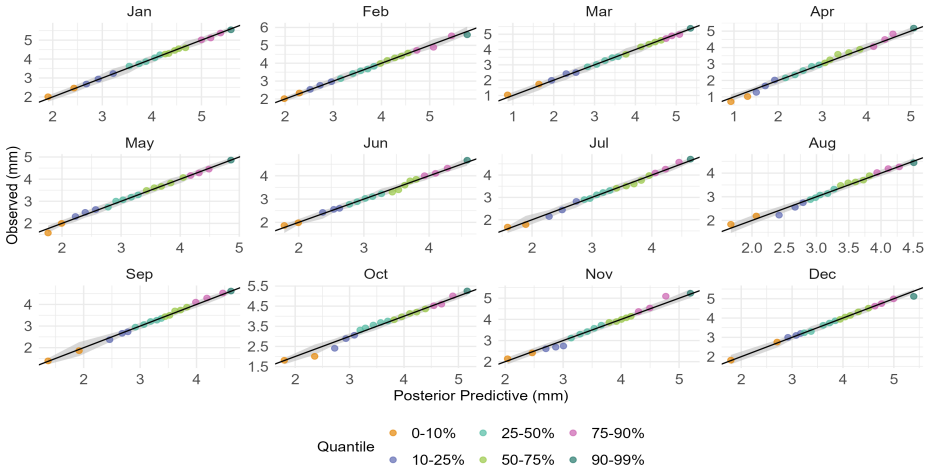


FIGURE 3.7: Median values (points) and interquartile range (shaded band) of $D_i^{(q)}$ across all sites, n_s, for each month and quantile. A red/ blue circle indicates that the corresponding quantile is nominal (i.e., the IQR of $D_i^{(q)}$, $i = 1.., n_s$, contains zero), while a red/ blue triangle indicates that it is not nominal. The results for the model with spatial dependence are shown in pink while results for the model without spatial dependence are shown in light blue.

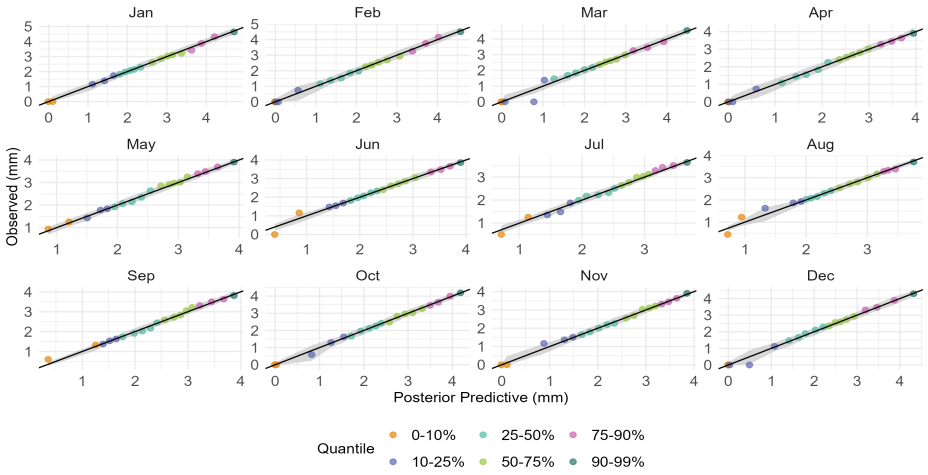
Additionally, to further analyze our model, we present Q-Q plots of empirical quantiles across the 12 months against the posterior mean of the predictive quantiles for the three stations discussed in section 3.1. These plots, shown in Figure 3.8, illustrate the alignment between the observed quantiles and those obtained from the posterior mean of the predictive quantiles at the selected sites. Furthermore, as shown in Figure 3.7, the $D_i^{(q)}$ values indicate that most sites demonstrate strong agreement across the majority of percentiles, with only a few exceptions. The Q-Q plots for Sydney Botanic Gardens (site 66006), Dunedoo Post Office (site 64009), and Pooncarie (Karpa Kora Station, site 47013) collectively demonstrate strong agreement between empirical and posterior predictive quantiles across 12 months, with most points closely following the diagonal line. The credible intervals generally encompass the observed values, highlighting the robustness of the predictive model. Minor deviations are observed at the extreme quantiles (e.g., 0–10% and 90–99%) at some sites and months, such as May, July, and November at Sydney Botanic Gardens, September and October at Dunedoo Post Office, and March and September at Pooncarie. These discrepancies indicate slight challenges in capturing extreme rainfall events, which are inherently variable. Overall, the model shows reliable performance across locations and temporal scales, effectively capturing rainfall quantile distributions with only small exceptions.



(a) Sydney Botanic Gardens [site 66006]



(b) Dunedoo Post Office [site 64009]



(c) Pooncarie (Karpa Kora Station) [site 47013]

FIGURE 3.8: Panel (a), (b), and (c) are Q–Q plots (quantiles from the posterior mean of the predictive densities vs observed) across 12 months for stations Sydney Botanic Gardens, Dunedoo Post Office, and Pooncarie (Karpa Kora Station), respectively. 80% posterior credible intervals for each quantile are shown by grey shaded bands.

3.5 Discussion and Conclusion

We have developed a Bayesian variable selection framework within a spatial quantile regression framework to analyze the relationship between the Southern Oscillation Index (SOI), and the quantiles of monthly rainfall for NSW, Australia. This framework accounts for spatial variability by incorporating spatially varying priors on regression coefficients of the model, by allowing the impact of SOI on monthly rainfall to vary across quantiles and locations. We employ a spike-and-slab prior structure to balance flexibility and sparsity in the model. The "slab" component is distributed as a multivariate normal with spatially varying mean, while the "spike" component assigns a point mass at zero with the probability of a regression coefficient being identically non-zero is parameterized to be a function of space. Given the computational demands of the model, we employ parallel computation techniques in R to enhance efficiency (Eugster et al., 2011).

The results show that the impact of the SOI on monthly rainfall is higher in areas to the west of the Great Dividing Range, particularly in the winter and spring months across the median, lower, and upper tails with a few exceptions, notably in June and November. Results of the impact on the median are consistent with prior analyses of the spatial pattern of ENSO influence on mean rainfall (Pepler et al., 2014; Risbey et al., 2009b) and rainfall terciles in Tozer et al., 2023. Our analysis extends this knowledge of ENSO's influence on rainfall to the different parts of the distribution, showing for example that from January to June, the influence of ENSO on rainfall is most likely to be felt in the extremes of the distribution rather than the median, a conclusion that is missed by conventional methods of detecting the relationship between ENSO and mean rainfall. We also find that SOI in June exhibits a higher impact to the east of GDR than on the west, and the impact on November shows a uniform significant impact of SOI across quantiles considered. Moreover, performance measures show our method performs better than traditional models that analyze locations independently in terms of predictive accuracy.

At a given location, the various quantiles of the rainfall distribution may be influenced by ENSO through different mechanisms. Since the physical phenomena associated with dry periods, average rainfall, and extreme rainfall are generally distinct, the analysis presented here can be used as a starting point to further our understanding of how ENSO interacts with precipitation over NSW by identifying the presence and strength of relationships between ENSO and these different parts of the rainfall distribution. For example, extreme rainfall in this region during April to June is often associated with East Coast Low systems (Dowdy et al., 2019), whose frequency and intensity could be impacted by large-scale climate drivers (Hopkins and Holland, 1997; Browning and Goodwin, 2013; Power and Callaghan, 2016; Barnes et al., 2023; **?**, **?**). Drought-like conditions, on the other hand, are more likely influenced by factors such as high-pressure systems, the subtropical ridge, surface temperatures, and tropical moisture transport (Ummenhofer et al., 2009; Timbal et al., 2010; Hope et al., 2010; Van Dijk et al., 2013; Holgate et al., 2025), which may also be influenced by climate drivers in different ways (see Meneghini et al., 2007; Risbey et al., 2009b; Ummenhofer et al., 2009; Verdon-Kidd and Kiem, 2009; Cai et al., 2009; Nicholls, 2010; Delage and Power, 2020) at different times of year through other mechanisms. The relationships we have detected between ENSO and the different parts of the monthly rainfall distribution across NSW can therefore be used to isolate particular mechanisms through which this influence is exerted.

However, the current model has limitations. It does not account for the joint impact of multiple large-scale climatic variables, and ongoing efforts are directed toward a joint modeling approach. In conclusion, our model offers significant insights into the influence of a climate index on the distribution of monthly rainfall. The next chapter aims to integrate multiple climate indices into the model, thereby improving its predictive precision and broadening its applicability in interpreting complex interactions between climate modes and monthly rainfall patterns.

CHAPTER 4

Extension of Variable Selection in Spatial Quantile Regression to Assess the Joint Impact of Climate Indices on Monthly Rainfall Distribution

---

This chapter builds upon the methodology presented in Chapter 3, extending it to incorporate multiple climate indices to better understand their joint effects on monthly rainfall quantiles while accounting for spatial variability. In addition to the SOI, we consider the Dipole Mode Index (DMI), which measures the Indian Ocean Dipole (IOD), both of which are known to play a critical role in shaping rainfall patterns in eastern Australia (Pui et al., 2012; Ummenhofer et al., 2009; Pepler et al., 2014).

Extensive research has explored the influence of the IOD and the El Niño-Southern Oscillation (ENSO) on rainfall patterns in southeast Australia (e.g., Meyers et al., 2007; Ummenhofer et al., 2009; Cai et al., 2011; Cai et al., 2012; Risbey et al., 2009a; Risbey et al., 2009b; Pepler et al., 2014). These studies provide valuable insights into how these climate indices drive variability in average and extreme rainfall events such as floods and droughts. However, much of this research has focused on understanding the impacts of ENSO and IOD independently, without addressing their joint spatial effects on monthly rainfall variability. Enhanced understanding of these spatial relationships could significantly improve the predictability of rainfall patterns in Eastern Australia (Risbey et al., 2009b).

Recent advances in spatio-temporal statistical models have gained significant attention for modeling extreme rainfall events (Thibaud et al., 2013; Zheng et al., 2015; De Fondeville and Davison, 2016; Saunders et al., 2017). These studies primarily employ generalized

extreme value (GEV) distributions or related extreme value theory approaches to capture the tail behavior of extreme daily precipitation events. However, such methods focus exclusively on the upper daily extremes, limiting their ability to analyze the full distribution of rainfall variability. In contrast, quantile regression provides a more flexible framework that allows for the modeling of rainfall across the entire distribution, making it particularly suitable for monthly rainfall data.

Moreover, a recent study by McGregor et al., 2024 uses a fractional attribute risk (FAR) to quantify the role of ENSO in modulating regional rainfall in Australia. This work is done by extending the work of Tozer et al., 2023 and quantifying how large the change in rainfall risk is for different phases of ENSO. They also stated the importance of assessing the impacts of multiple climate indices on rainfall risk. However, there remains a scarcity of probabilistic models in the literature that apply variable selection methods to quantify the joint impact of multiple climate indices on monthly rainfall distribution, specifically within a spatial quantile regression framework. For example, the study by Bertolacci et al., 2019 employs a mixture model to examine the joint impacts of climate indices on daily rainfall in Australia. However, this approach does not incorporate variable selection to identify the most influential predictors in daily rainfall.

The extended model employs a Bayesian hierarchical quantile regression framework. While the first and second levels of the hierarchy remain consistent with Chapter 3, the third level incorporates spatial dependence by modeling the joint inclusion probabilities of regression coefficients using a multinomial logistic regression where the argument of the link function is modeled spatially using a Gaussian Process ($\mathcal{GP}$) prior.

This chapter makes two contributions. First, we develop a novel methodology for Bayesian variable selection within a quantile regression framework to assess the joint influence of multiple climate indices on monthly rainfall distribution, incorporating spatial dependencies. Second, we contribute to climate science by providing spatially varying inferences on the joint impact of these indices on monthly rainfall quantiles.

This chapter is structured as follows: Section 4.1 describes the model and priors. Section 4.2 outlines the estimation and inference procedures. Section 4.3 presents results for the impact of ENSO and IOD on monthly rainfall quantiles. Finally, Section 4.4 provides a discussion of the findings.

4.1 Model and Priors

This chapter aims to examine the dependency between the joint impact of global climate indices, specifically the Southern Oscillation Index (SOI), Dipole Mode Index (DMI), and the quantiles of monthly rainfall. The focus is on how this relationship varies spatially across different months. To achieve this, we employ a hierarchical statistical framework that allows for both the integration of multiple climate drivers and the spatial variation in their impacts. This approach extends the methodology presented in Chapter 3 to a multiple-variable context, allowing us to infer the spatially varying joint impact of multiple climate indices.

4.1.1 Likelihood

Consider rainfall observations for a given month at location i, across years $t = 1, \ldots, T$, represented by $\boldsymbol{y}_i = (y_{1,i}, \ldots, y_{T,i})'$, and observations on K global climate indicators $\mathbf{x}_1, \ldots, \mathbf{x}_K$, where each $\mathbf{x}_k = (x_{1,k}, \ldots, x_{T,k})'$ for $k = 1, \ldots, K$. In this chapter, we are interested in the impact of the SOI and the DMI on monthly rainfall quantiles in New South Wales, so $K = 2$.

Following the methodology of Yu and Moyeed, 2001, we model the conditional q^{th} quantile of rainfall at location i as a linear combination of the climate indicators,

$$\boldsymbol{y}_i = X\boldsymbol{\beta}_i^{(q)} + \boldsymbol{\epsilon}_i \,, \tag{4.1}$$

where $X = (\mathbf{1}_T, \mathbf{x}_1, \mathbf{x}_2)$, and $\boldsymbol{\beta}_i^{(q)} = (\beta_{0,i}^{(q)}, \beta_{1,i}^{(q)}, \beta_{2,i}^{(q)})'$ are the regression coefficients for the q^{th} quantile of rainfall conditional on the global climate indicators. The error term $\boldsymbol{\epsilon}_i = (\epsilon_{1,i}, ..., \epsilon_{T,i})'$ has an Asymmetric Laplace Distribution (ALD) as in Chapter 3.

Therefore likelihood is given by,

$$f_q(\boldsymbol{y}_i|\boldsymbol{\beta}_i^{(q)}, \sigma_i, q) \sim \left\{\frac{q(1-q)}{\sigma_i}\right\}^T \exp\left\{\frac{-\sum_{t=1}^T \rho_q(y_{t,i} - \boldsymbol{x}_t\boldsymbol{\beta}_i^{(q)})}{\sigma_i}\right\}, \tag{4.2}$$

where $\boldsymbol{x}_t$ is the t^{th} row of X.

4.1.2 Priors

To capture the spatial variability in the effects of global climate indices on monthly rainfall quantiles as in Chapter 3, we employ a prior distribution on the regression coefficients that allows these impacts to vary by space using spike and slab prior. For simplicity, we omit explicit conditioning on the quantile q in this formulation but note that the parameters take different values for different quantiles.

Specifically for each quantile, q, and for locations, $i = 1, \ldots, n_s$, where n_s is the number of locations, let $\boldsymbol{\beta}_0 = (\beta_{0,i}, \ldots, \beta_{0,n_s})$ be the vector of regression coefficients corresponding to the intercepts, and define $\boldsymbol{\gamma}_i = (\gamma_{1,i}, \gamma_{2,i})$ to be an indicator vector with $\gamma_{k,i} = 1$ if $\beta_{k,i} \neq 0$ and $\gamma_{k,i} = 0$ if $\beta_{k,i} = 0$ for $k = 1, 2$. Let $\mathcal{S}_{1,i} = \{k; \gamma_{k,i} = 1; k = 1, 2\}$, so that $\mathcal{S}_{1,i}$ contains the indices of these covariates at site i for which the corresponding regression co-efficients are non-zero, and define $\mathcal{S}_{0,i}$ similarly. Then the prior specification is

$$
\begin{aligned}
p(\boldsymbol{\beta}_1^{(q)}, \ldots, \boldsymbol{\beta}_{n_s}^{(q)} | \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{n_s}, \boldsymbol{\sigma}) &= \prod_{i=1}^{n_s} p(\boldsymbol{\beta}_i^{(q)} | \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{n_s}, \sigma_i) \,, \\
\boldsymbol{\beta}_{\mathcal{S}_{1,i}} | c, \boldsymbol{\mu}_i, \sigma_i &\sim \mathcal{N}\left(\boldsymbol{\mu}_{\mathcal{S}_{1,i}}, c\sigma_i (X'_{\mathcal{S}_{1,i}} X_{\mathcal{S}_{1,i}})^{-1}\right), \\
\boldsymbol{\beta}_{\mathcal{S}_{0,i}} &\sim \delta(0), \\
p(\boldsymbol{\beta}_0 | c, \boldsymbol{\sigma}) &= \prod_{i=1}^{n_s} p(\beta_{0,i} | c, \sigma_i), \\
\beta_{0,i} &\sim \mathcal{N}(0, c\sigma_i) \text{ for } i = 1, \ldots, n_s. \qquad (4.3)
\end{aligned}
$$

The prior on σ_i is an inverse gamma prior with parameters a and b ($\mathcal{IG}(a, b)$), where $a = 1$ and $b = 0.5$, for $i = 1, \ldots, n_s$.

4.1.2.1 Prior for $\boldsymbol{\mu}$

As in Chapter 3, we define $\mathcal{A}_{k,1} = \{i; \gamma_{k,i} = 1\}$ to be the set of location indices for which the regression coefficient of the k^{th} covariate is non-zero, and define $\mathcal{A}_{k,0}$ similarly.

$$
\mu_{k,i} = \mathbf{s}_i \boldsymbol{\eta} + f_{k,\mu}(\text{lat}_i, \text{lon}_i) \text{ for } i \in \mathcal{A}_{k,1} \text{ and } k = 1, 2, \qquad (4.4)
$$

where $\mathbf{s}_i = [1, \text{lat}_i, \text{lon}_i]$ and $\boldsymbol{\eta} = (\eta_0, \eta_1, \eta_2)'$. We assume apriori that the function $f_{k,\mu}$ is a Gaussian Process, $f_{k,\mu} \sim \mathcal{GP}(0, \tau^2 \Omega)$, so that $\boldsymbol{f}_{k,\mathcal{A}_{k,1},\mu} = (f_{k,\mathcal{A}_{k,1}(1),\mu}, \ldots, f_{k,\mathcal{A}_{k,1}(N_k),\mu})$, is multivariate normal,

$$
\boldsymbol{f}_{k,\mathcal{A}_{k,1},\mu} \sim \mathcal{N}\left(0, \tau^2_{k,\mathcal{A}_{k,1}} \Omega_{k,\mathcal{A}_{k,1}}\right) \,, \qquad (4.5)
$$

with $N_k = ||\mathcal{A}_{k,1}||$ and the cov $\left(f_{k,\mathcal{A}_{k,1}}(\text{lat}_i, \text{lon}_i), f_{k,\mathcal{A}_{k,1}}(\text{lat}_j, \text{lon}_j)\right) = \tau^2_{k,\mathcal{A}_{k,1}} \omega_{ij}$ for i and $j \in \mathcal{A}_{k,1}$. We assume apriori that $\tau^2_{k,\mathcal{A}_{k,1}}$ has an inverse gamma distribution with parameters a and b ($\mathcal{IG}(a, b)$) where $a = 1$ and $b = 0.5$. Therefore, $\boldsymbol{f}_{k,\mathcal{A}_{k,1},\mu} \approx H_{k,\mathcal{A}_{k,1}} \boldsymbol{\phi}_{k,\mathcal{A}_{k,1}}$, where $H_{k,\mathcal{A}_{k,1}} = Q_{1:p} D^{1/2}_{1:p}$, and letting $\boldsymbol{\phi}_{k,\mathcal{A}_{k,1}} \sim N(0, \tau^2_{k,\mathcal{A}_{k,1}} I_p)$. Refer to Section 3.2.2.1 for more details.

4.1.2.2 Prior on γ

Let $\boldsymbol{\gamma}_i \in \{0,1\}^2$ represent a binary set for each location i and $C = 2^2$ denotes the total number of possible sets. We assume $\boldsymbol{\gamma}_i$ are independently and identically distributed (i.i.d.) with,

$$\boldsymbol{\gamma}_i \sim \text{Multinomial}(1, (\pi_{1,i}, \ldots, \pi_{C,i})) \,,$$

represents the Multinomial distribution with set of probabilities $(\pi_{1,i}, \ldots, \pi_{C,i})$ where $\pi_{c,i}$ denotes the joint probability for c^{th} possible set, satisfying the condition $\sum_{c=1}^{C} \pi_{c,i} = 1$.

We parameterise joint probability for each individual combination c, i.e $\pi_{c,i}^{(q)}$ as a multinomial logistic link function to depend upon spatial covariates (Kavelaars et al., 2022), latitude, (lat) and longitude, (lon), such that,

$$\pi_{c,i}^{(q)} = \frac{\exp(\mathbf{s}_i \boldsymbol{\alpha}_c + f_{c,\pi}(\text{lat}_i, \text{lon}_i))}{1 + \sum_{c=1}^{C-1} \exp(\mathbf{s}_i \boldsymbol{\alpha_c} + f_{c,\pi}(\text{lat}_i, \text{lon}_i))} \;\forall\; c, i \,, \tag{4.6}$$

where $\boldsymbol{\alpha}_c = (\alpha_{0,c}, \alpha_{1,c}, \alpha_{2,c})'$ and f is a Gaussian process, with smoothing parameter λ_c^2, so that,

$$\boldsymbol{f}_{c,\pi} \sim \mathcal{N}\left(0, \lambda_c^2 \Omega_c\right) , \tag{4.7}$$

where the set of indices $i \in c$ with $\boldsymbol{f}$ approximated as a linear combination of truncated basis functions as before so that $\boldsymbol{f}_{c,\pi} \approx H_c \boldsymbol{\delta}_c$, where $H_c = Q_{1:p} D_{1:p}^{1/2}$, and $\boldsymbol{\delta}_c \sim N(0, \lambda_c^2 I_p)$. The smoothing parameter is $\lambda_c^2 \sim \mathcal{IG}(a, b)$.

4.2 Estimation and Inference

For each month and quantile, the complete set of parameters is $\boldsymbol{\Theta} = (\boldsymbol{B}^*, \boldsymbol{\sigma}, \{\boldsymbol{\phi}^*_{k,\mathcal{A}_{k,1}}\}_{k=1}^K, \{\boldsymbol{\delta}^*_c\}_{c=1}^{C-1}, \{\lambda^2_c\}_{c=1}^{C-1}, \{\boldsymbol{\tau}^2_{k,\mathcal{A}_{k,1}}\}_{k=1}^K, \boldsymbol{Y}^{\text{miss}}, \boldsymbol{\Gamma})$, where $\boldsymbol{B}^* = (\boldsymbol{\beta}_0, \boldsymbol{\beta}_{\mathcal{A}_{\cdot,1}})$, $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_{n_s})'$ and $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{n_s})'$. As introduced in Chapter 3, the missing rainfall values are denoted as $\boldsymbol{Y}^{\text{miss}} = (\boldsymbol{y}_1^{\text{miss}}, \ldots, \boldsymbol{y}_{n_s}^{\text{miss}})'$, while the observed rainfall values are $\boldsymbol{Y}^{\text{obs}} = (\boldsymbol{y}_1^{\text{obs}}, \ldots, \boldsymbol{y}_{n_s}^{\text{obs}})'$. For consistency, we define $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{n_s})'$ as the complete set of observed and missing values of monthly rainfall. Inference is carried out through the joint posterior distribution $p(\boldsymbol{\Theta} \mid \boldsymbol{Y}, X)$. As in the earlier chapter, we employ MCMC to perform the necessary multidimensional integration, facilitated through data augmentation in two ways.

(1) We write the likelihood as in Chapter 3, the likelihood conditional on $\mathbf{v}_i, \sigma_i, \boldsymbol{\beta}_i$ is given by,

$$f_q(\boldsymbol{y}_i|\boldsymbol{\beta}_i, \mathbf{v}_i, \sigma_i, q) \sim N\left(X\boldsymbol{\beta}_i + \theta\mathbf{v}_i, \zeta^2\sigma_i \text{diag}_T(\mathbf{v}_i)\right). \tag{4.8}$$

(2) To estimate $\pi^{(q)}_{c,i}$, we follow Kavelaars et al., 2022; Polson et al., 2012 and introduce the latent variables for each combination c, $\mathbf{w}_c = (\mathrm{w}_{c,1}, \ldots, \mathrm{w}_{c,n_s})$, and these latent variables are connected to the indicator vector $\boldsymbol{\gamma}_i = (\gamma_{1,i}, \gamma_{2,i})$ by requiring that if $\boldsymbol{\gamma}_i \in c^{\text{th}}$ combination then $\mathrm{w}_{c,i} \sim \mathcal{PG}\{1, \mathbf{s}_i\boldsymbol{\alpha}_c + f_{c,\pi}(\text{lat}_i, \text{lon}_i) - \ln\sum_{d\neq c}\exp(\mathbf{s}_i\boldsymbol{\alpha}_d + f_{d,\pi}(\text{lat}_i, \text{lon}_i))\}$. Here, $\mathcal{PG}$ denotes the Pólya-Gamma distribution (see Polson et al., 2012, for details).

It is convenient to express $\boldsymbol{\mu}_{\mathcal{A}_{k,1}} = (\mu_{k(1)}, \ldots, \mu_{k(N_k)})$ and $\boldsymbol{f}_{k,\mathcal{A}_{k,1},\mu}$ as $H_{k,\mathcal{A}_{k,1}}\boldsymbol{\phi}_{k,\mathcal{A}_{k,1}}$, as in Section 4.1.2.1, and define $\boldsymbol{\phi}^*_{k,\mathcal{A}_{k,1}} = (\boldsymbol{\eta}', \boldsymbol{\phi}'_{k,\mathcal{A}_{k,1}})'$, $H^*_{k,\mathcal{A}_{k,1}} = [S_{k,\mathcal{A}_{k,1}}, H_{k,\mathcal{A}_{k,1}}]$, and $S = (\boldsymbol{s}'_1, \ldots, \boldsymbol{s}'_{n_s})'$, so that $\boldsymbol{\mu}_{\mathcal{A}_{k,1}} = H^*_{k,\mathcal{A}_{k,1}}\boldsymbol{\phi}^*_{k,\mathcal{A}_{k,1}}$. Then it follows that $\boldsymbol{\phi}^*_{k,\mathcal{A}_{k,1}} \sim \mathcal{N}(\mathbf{0}, \Psi)$, with $\Psi_{k,\mathcal{A}_{k,1}} = \text{diag}(\boldsymbol{\tau}^{2*}_{k,\mathcal{A}_{k,1}})$ and $\boldsymbol{\tau}^{2*}_{k,\mathcal{A}_{k,1}} = (\boldsymbol{c}_\eta\mathbf{1}_3, \tau^2_{k,\mathcal{A}_{k,1}}\mathbf{1}_p)$ where $\mathbf{1}_p$ is a row vectors of ones of length p. Similarly, we express $\boldsymbol{\pi}_c = (\pi_{c,1}, \ldots, \pi_{c,n_s})$ and define $\boldsymbol{\delta}^*_c = (\boldsymbol{\alpha}'_c, \boldsymbol{\delta}'_c)'$, and $\boldsymbol{\delta}^*_c \sim$

$\mathcal{N}(\mathbf{0}, \Lambda_c)$, with $\Lambda_c = \text{diag}(\boldsymbol{\lambda}_c^{2*})$ and $\boldsymbol{\lambda}_c^{2*} = (\boldsymbol{c}_\alpha \mathbf{1}_3, \lambda_c^2 \mathbf{1}_p)$, so that $\boldsymbol{\pi}_c = \frac{\exp(H_c^* \boldsymbol{\delta}_c^*)}{1+\sum_{c=1}^{C-1} \exp(H_c^* \boldsymbol{\delta}_c^*)}$. We fix $\boldsymbol{\delta}_C^* = 0$ for identifiability and set $c_\eta = c_\alpha = n_s$.

Algorithm 3 MCMC scheme

Input: Data $\mathcal{D}$, number of iterations J.
Output: Samples from posterior distributions i.e. $p(\boldsymbol{\Theta}|\boldsymbol{Y}, X)$.
Initialize parameters $(\boldsymbol{B}^{*[1]}, \boldsymbol{\sigma}^{[1]}, \{\boldsymbol{\phi}_{k,\mathcal{A}_{k,1}}^{*[1]}\}_{k=1}^K, \{\boldsymbol{\delta}_c^{*[1]}\}_{c=1}^{C-1}, \{\lambda_c^{2[1]}\}_{c=1}^{C-1}, \{\tau_{k,\mathcal{A}_{k,1}}^{2[1]}\}_{k=1}^K, \boldsymbol{Y}^{\text{miss}[1]}, \boldsymbol{\Gamma}^{[1]})$.
for $j = 1$ to J **do**
    Generate $\mathbf{V}^{[j+1]}$ from $p(\mathbf{V}|\boldsymbol{Y}, \boldsymbol{\sigma}^{[j]}, \boldsymbol{B}^{*[j]})$.
    Generate $\boldsymbol{\sigma}^{[j+1]}$ from $p(\boldsymbol{\sigma}|\boldsymbol{Y}, \mathbf{V}^{[j+1]}, \boldsymbol{B}^{*[j]})$.
    Generate $\boldsymbol{Y}^{\text{miss}[j+1]}$ from $p(\boldsymbol{Y}^{\text{miss}}|\boldsymbol{Y}^{\text{obs}}, \mathbf{V}^{[j+1]}, \boldsymbol{\sigma}^{[j+1]}, \boldsymbol{B}^{*[j]})$.
    for $c = 1$ to $C-1$ **do**
        $\mathbf{w}_c^{[j]}|\boldsymbol{\delta}_c^{*[j]}, \boldsymbol{\Gamma}^{[j]} \sim \mathcal{PG}(1, H_c^* \boldsymbol{\delta}_c^* - \ln \sum_{d \neq c} \exp(H_d^* \boldsymbol{\delta}_d^*))$.
        Generate $\boldsymbol{\pi}_c$ by drawing $\boldsymbol{\delta}_c^{*[j+1]}$ from $p(\boldsymbol{\delta}_c^*|\mathbf{w}_c^{[j+1]}, \lambda_c^{2[j]})$.
        Generate $\lambda_c^{2[j+1]}$ from $p(\lambda_c^2|\boldsymbol{\delta}_c^{[j+1]})$.
    end for
    for $k = 1$ to K **do**
        Generate $\boldsymbol{\mu}_{\mathcal{A}_{k,1}}$ by drawing $\boldsymbol{\phi}_{k,\mathcal{A}_{k,1}}^{*[j+1]}$ from $p(\boldsymbol{\phi}_{k,\mathcal{A}_{k,1}}^*|\boldsymbol{\beta}_{\mathcal{A}_{k,1}}^{[j]}, \tau_{k,\mathcal{A}_{k,1}}^{2[j]})$.
        Generate $\tau_{k,\mathcal{A}_{k,1}}^{2[j+1]}$ from $p(\tau_{k,\mathcal{A}_{k,1}}^2|\boldsymbol{\phi}_{k,\mathcal{A}_{k,1}}^{[j+1]})$.
    end for
    Generate $\boldsymbol{B}^{*[j+1]}$ and $\boldsymbol{\Gamma}^{[j+1]}$ jointly from $p(\boldsymbol{\Gamma}, \boldsymbol{\beta}_{\mathcal{A}_{\cdot,1}}, \boldsymbol{\beta}_0|\tilde{\boldsymbol{Y}}, \boldsymbol{\mu}_{\mathcal{A}_{\cdot,1}}, \mathbf{V}^{[j+1]}, \boldsymbol{\sigma}^{[j+1]})$.
end for
return $\{\boldsymbol{B}^{*[j+1]}, \boldsymbol{\sigma}^{[j+1]}, \{\boldsymbol{\phi}_{k,\mathcal{A}_{k,1}}^{*[j+1]}\}_{k=1}^K, \{\boldsymbol{\delta}_c^{*[j+1]}\}_{c=1}^{C-1}, \{\lambda_c^{2[j+1]}\}_{c=1}^{C-1}, \{\tau_{k,\mathcal{A}_{k,1}}^{2[j+1]}\}_{k=1}^K, \boldsymbol{Y}^{\text{miss}[j+1]}, \boldsymbol{\Gamma}^{[j+1]}\}_{j=1}^J$.

All full conditional posterior distributions for the algorithm 3, are stated in in Appendix C1.

4.3 Results

4.3.1 Impact of ENSO and IOD on Monthly Rainfall quantiles

The results in Figures 4.1, 4.2, 4.3 and 4.4 illustrate the impact of SOI and DMI on the monthly rainfall quantiles in NSW using the same metrics as in Chapter 3 for the rainfall quantiles $q = 0.05$, $q = 0.50$ and $q = 0.95$. These metrics are the marginal posterior mean probability that a climate index has an impact on the quantiles of monthly rainfall, $\hat{\pi}_{k,l} = \frac{1}{J}\sum_{j=1}^{J}\pi_{k,l}^{[j]}$, for each covariate $k = 1, 2$ and location l, and Figures 4.1 and 4.3 show the spatial distribution of this metric for each covariate. The second metric is the marginal posterior mean of $\mu_{k,l}|\gamma_{k,l} = 1$, given by $\hat{\mu}_{k,l} = \frac{1}{J}\sum_{j=1}^{J}(\mu_{k,l}^{[j]}|\gamma_{k,l\in\mathcal{G}_1}^{[j]})$, where $\mathcal{G}_1 = \{l; \gamma_{k,l} = 1\}$ and Figures 4.2 and 4.4 show the spatial distribution of this metric for each covariate. We further examine the two metrics mentioned above to the west and east of the GDR similar to Chapter 3. We denote $\overline{\pi}_{k,W} = \frac{1}{||L_W||}\sum_{l\in L_W}\hat{\pi}_{k,l}$, for the west (blue) and $\overline{\pi}_{k,E} = \frac{1}{||L_E||}\sum_{l\in L_E}\hat{\pi}_{k,l}$, for the east (orange), of the GDR for 12 months for the quantiles $q = 0.05,\ 0.1,\ 0.2,\ 0.5,\ 0.8,\ 0.9$ and 0.95 and each covariate k. Similarly, for the mean magnitude of the coefficient values, $\overline{\mu}_{k,W} = \frac{1}{||L_W||}\sum_{l\in L_W}\hat{\mu}_{k,l}$ and $\overline{\mu}_{k,E} = \frac{1}{||L_E||}\sum_{l\in L_E}\hat{\mu}_{k,l}$, for the west and east regions, respectively.

Figures 4.1- 4.6 reveal four key characteristics of the joint impact of SOI and DMI on monthly rainfall quantiles in NSW.

(1) The influence of SOI on rainfall is moderated by the presence of DMI during September and October. In these months a comparison of the results in Figures 4.1(a), 4.1(b) and 4.1(c) with those in 3.3(a), 3.3(b) and 3.3(c) shows that the presence of DMI in the model reduces the marginal posterior inclusion probability of SOI in the months of September and October, the months when the IOD is most active. This observation aligns with the results reported (Pepler et al. (2014); Ummenhofer et al. (2009); Meyers et al. (2007)), who found that the impact of SOI on

spring rainfall in eastern Australia depended on DMI.

The DMI moderates not only the dependency of the rainfall in September and October on SOI, but it also changes the spatial variability of this dependency. Figures 4.1 and 4.2 show, for the months September and October, that the differential impact of the SOI on monthly rainfall between locations east and west of the GDR is less pronounced once the impact of the DMI is taken into consideration. As shown in Figures 4.5(a) and 4.5(b), there is no notable difference in the impact of SOI on monthly rainfall between the two regions during these months.

(2) Compared to SOI, the DMI generally shows a less pronounced influence on all three rainfall quartiles ($q = 0.05$, $q = 0.5$ and $q = 0.95$). However, notable anomalies occur in September and October. During these months, the impact of SOI weakens, while the impact of DMI becomes more pronounced, as indicated by high inclusion probabilities (Figure 4.3), and exhibits strong negative magnitudes across a range of quantiles (Figure 4.4).

(3) The differential impact of DMI on monthly rainfall for locations east and west of the GDR is apparent in September and October. Similar to the impact of the SOI on monthly rainfall, Figure 4.6(a) shows that the impact of the DMI on monthly rainfall is higher in the west of the GDR than in the east. The corresponding graph of the mean magnitude of the impact (see Figure 4.6(b)) shows a significant negative magnitude in the west of the GDR in the quantiles considered, suggesting a notable reduction in rainfall associated with high DMI values in these areas.

Moreover, Figure 4.6(a) shows that the impact of the Dipole Mode Index (DMI) on rainfall is stronger at the lower and upper quantiles than at the middle quantiles for most months, forming a distinct U-shaped pattern. This pattern is evident across

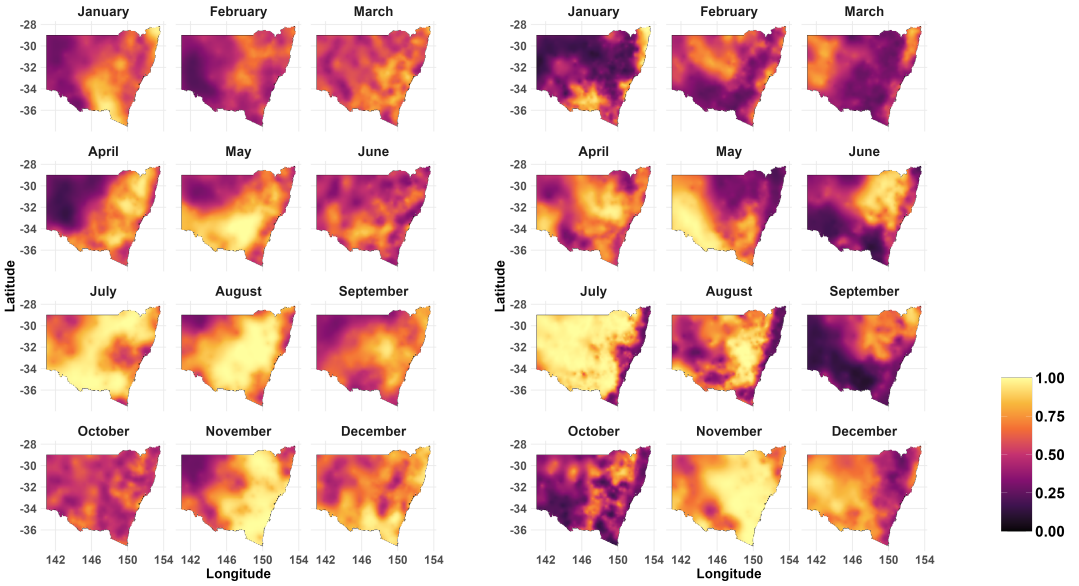
the majority of months, except for September and October, where this pattern does not hold. Figure 4.6(b) further illustrates that, in most months, the magnitude of the impact is relatively consistent across quantiles and regions; however, in September and October, the impact varies more noticeably across quantiles.

(4) In March, the impact of DMI on lower quantiles is more pronounced in most locations, suggesting that DMI plays a more significant role in driving dry anomalies during this month compared to its effect on typical or extreme wet conditions, as shown in Figures 4.3 and 4.4.

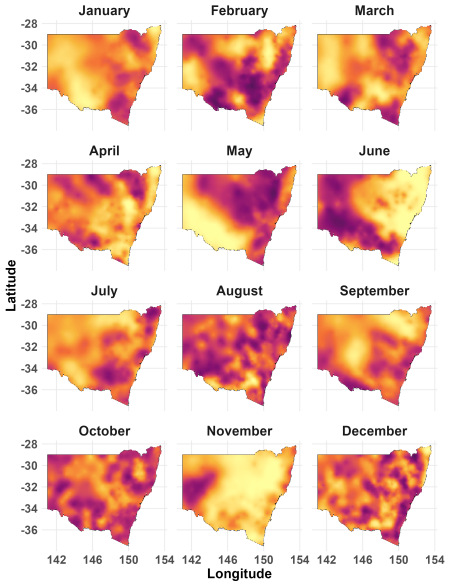
The results highlight that during the months September and October, the influence of the IOD tends to dominate over the effects of ENSO in all rainfall quantiles, including the lower, median, and upper extremes. However, the IOD has negligible impact on rainfall in November, and the strong influence of ENSO on November rainfall, noted in Chapter 3, is still apparent after accounting for the impact of the IOD. Pepler et al., 2014 also find a strong correlation between ENSO and IOD during the Austral spring [1], and note that it is challenging to disentangle their individual contributions. Our findings go some way to disentangling the impact of ENSO and IOD on spring rainfall by showing that ENSO dominates in November but IOD has a stronger influence in September and October.

The negative magnitude of the impact of DMI observed in our analysis in all rainfall quantiles suggests a reduction in rainfall associated with positive DMI. This finding is consistent with the established understanding that negative DMI events are associated with strong westerly moisture flux anomalies originating from the Indian Ocean, which are more likely to enhance rainfall in southeast Australia, particularly west of the GDR (Ummenhofer et al., 2009; Ummenhofer et al., 2011; Brown et al., 2009). In contrast, positive DMI events are typically linked to easterly moisture flux anomalies that tend to reduce rainfall in these regions.

[1]The term "Austral spring" refers to the spring months of September, October, and November, i.e., the spring season in the Southern Hemisphere.

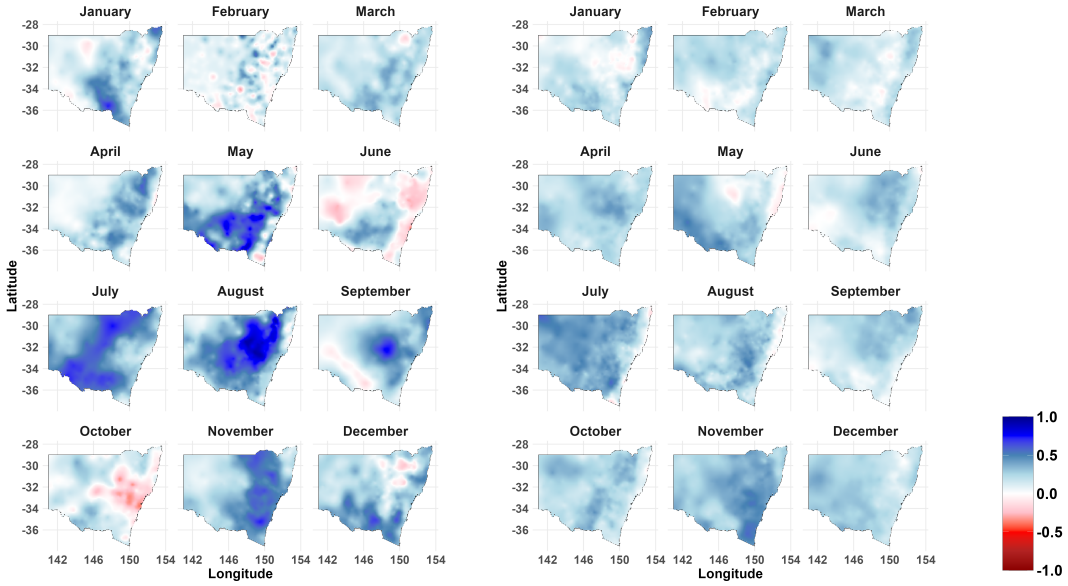


(a) $q = 0.05$ (b) $q = 0.5$

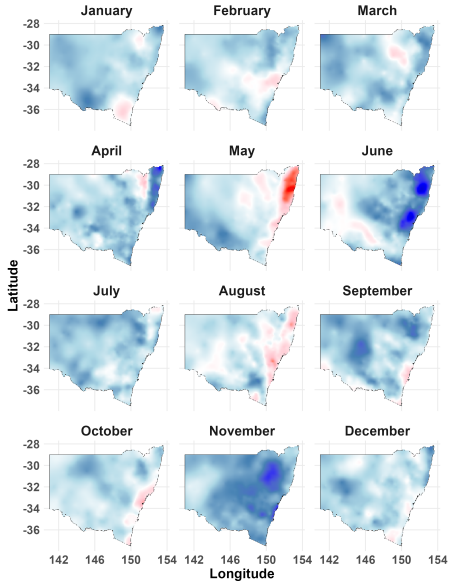


(c) $q = 0.95$

FIGURE 4.1: Panel (a), (b), and (c) show heat maps of $\hat{\pi}_{k,l}$ where k^{th} covariate is SOI (representing the marginal probability of inclusion of SOI as an influence on rainfall quantile q) across the NSW for 12 months at $q = 0.05$, $q = 0.5$, and $q = 0.95$, respectively.

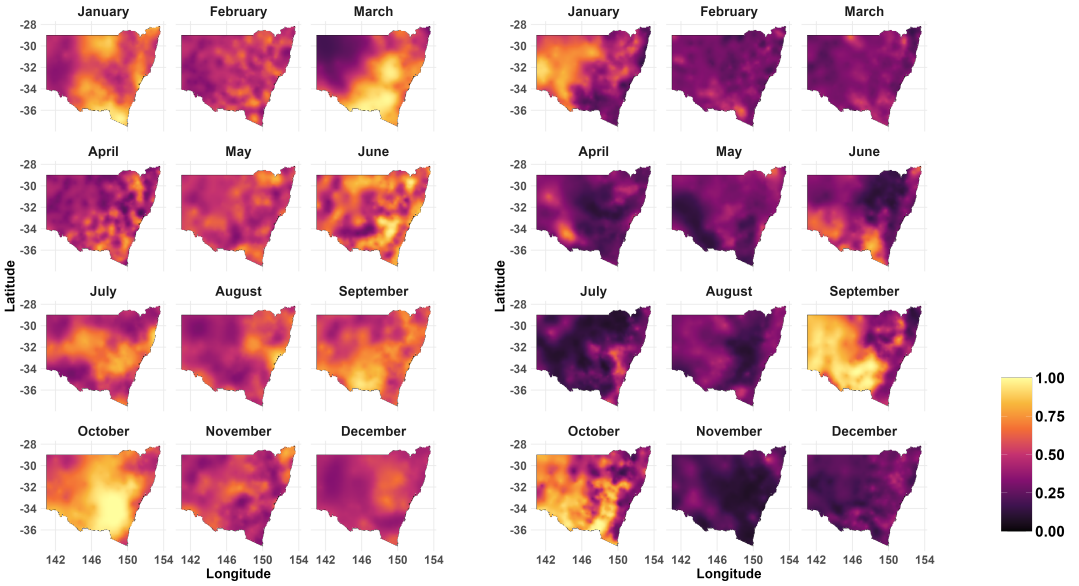


(a) $q = 0.05$ (b) $q = 0.5$

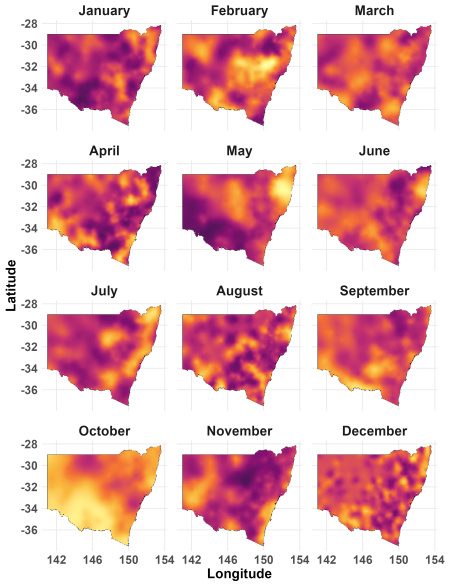


(c) $q = 0.95$

FIGURE 4.2: Panel (a), (b), and (c) show heat maps of $\hat{\mu}_{k,l}$ where k^{th} covariate is SOI (representing the estimated strength of the influence of SOI on rainfall quantile q) across the NSW for 12 months at $q = 0.05$, $q = 0.5$, and $q = 0.95$, respectively.

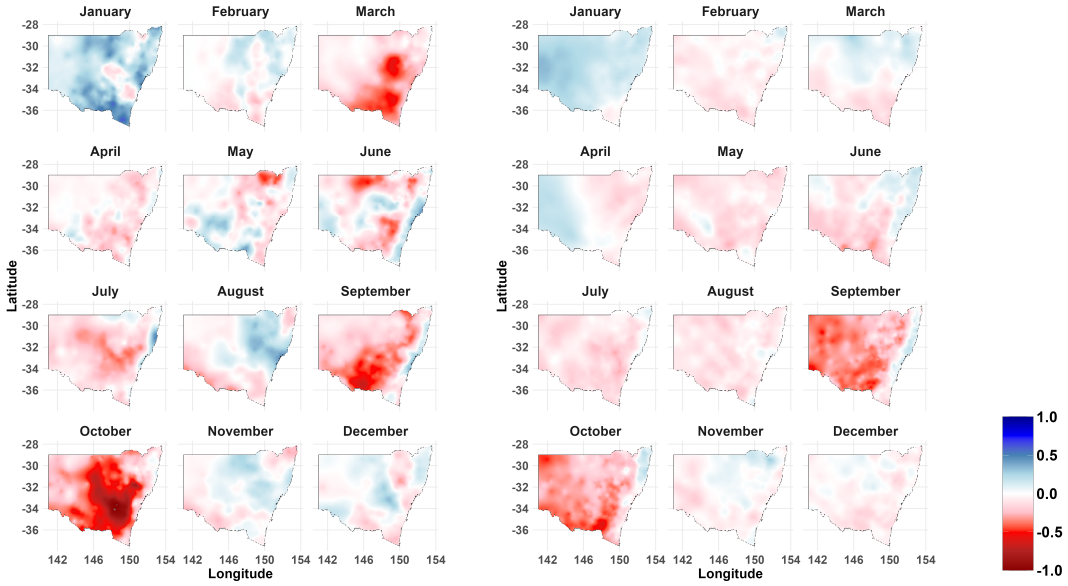


(a) $q = 0.05$ (b) $q = 0.5$

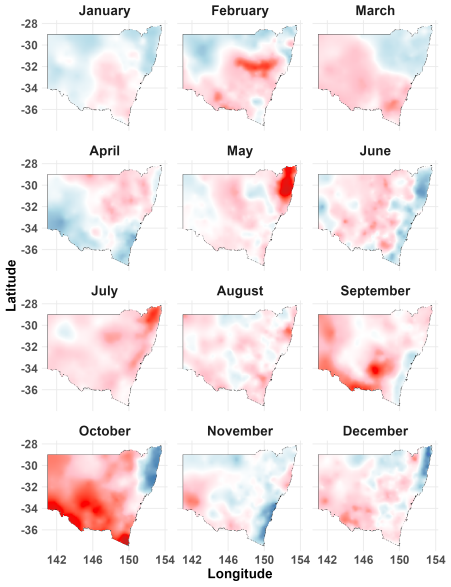


(c) $q = 0.95$

FIGURE 4.3: Panel (a), (b), and (c) show heat maps of $\hat{\pi}_{k,l}$ where k^{th} covariate is DMI (representing the marginal probability of inclusion of DMI as an influence on rainfall quantile q) across the NSW for 12 months at $q = 0.05$, $q = 0.5$, and $q = 0.95$, respectively.

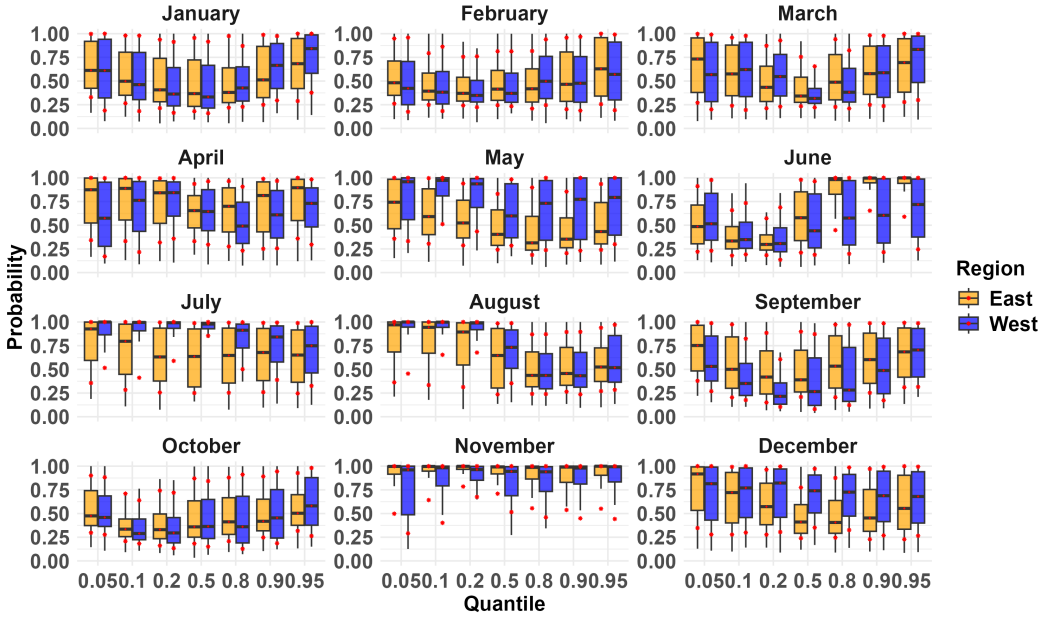


(a) $q = 0.05$ (b) $q = 0.5$

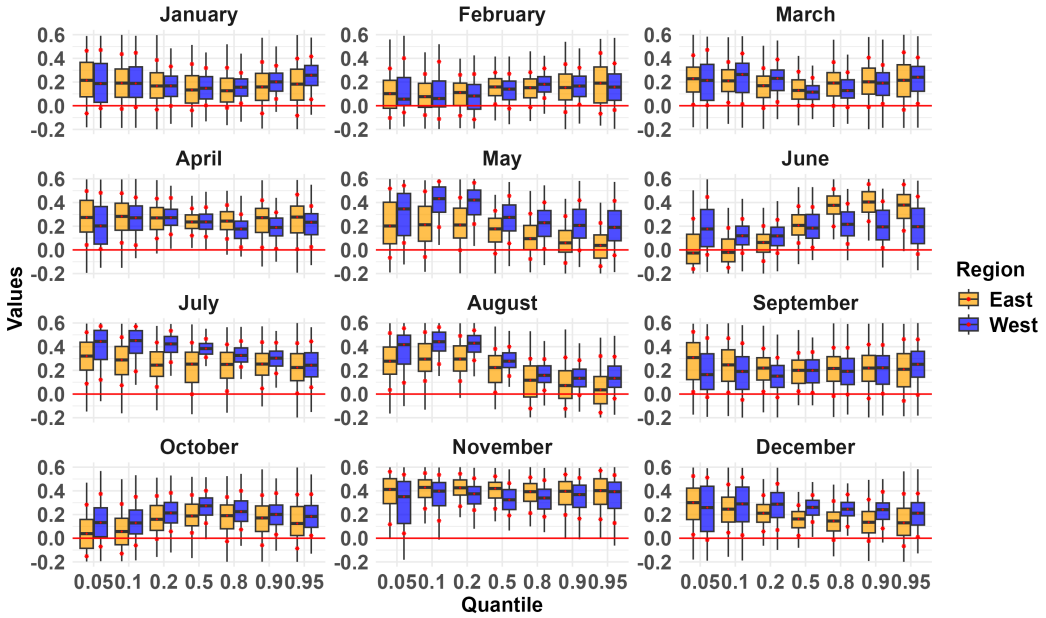


(c) $q = 0.95$

FIGURE 4.4: Panel (a), (b), and (c) show heat maps of $\hat{\mu}_{k,l}$ where k^{th} covariate is DMI (representing the estimated strength of the influence of DMI on rainfall quantile q) across the NSW for 12 months at $q = 0.05$, $q = 0.5$, and $q = 0.95$, respectively.

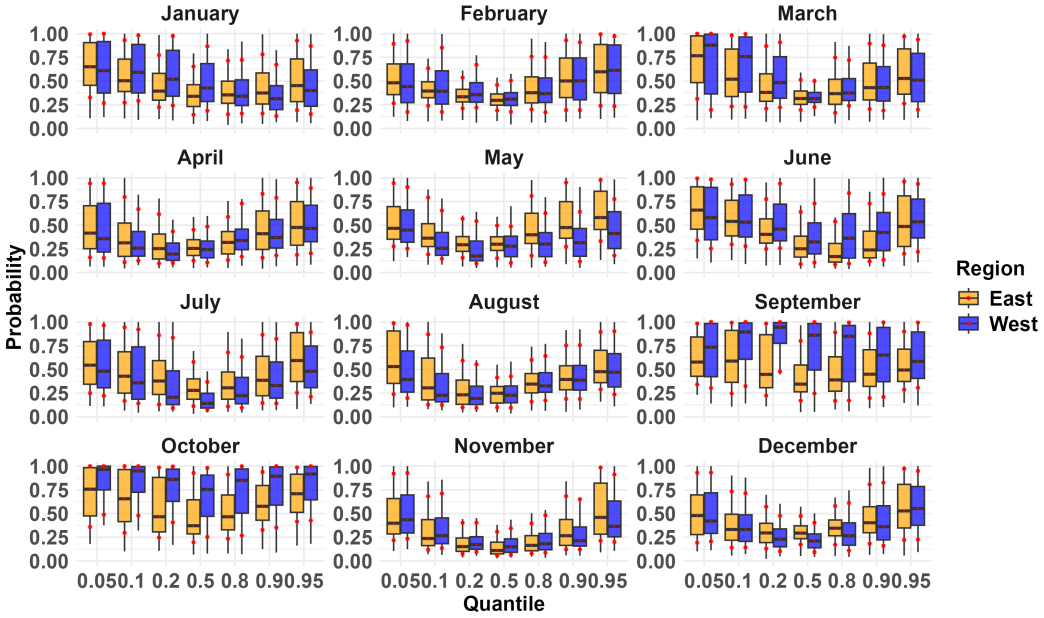


(a)

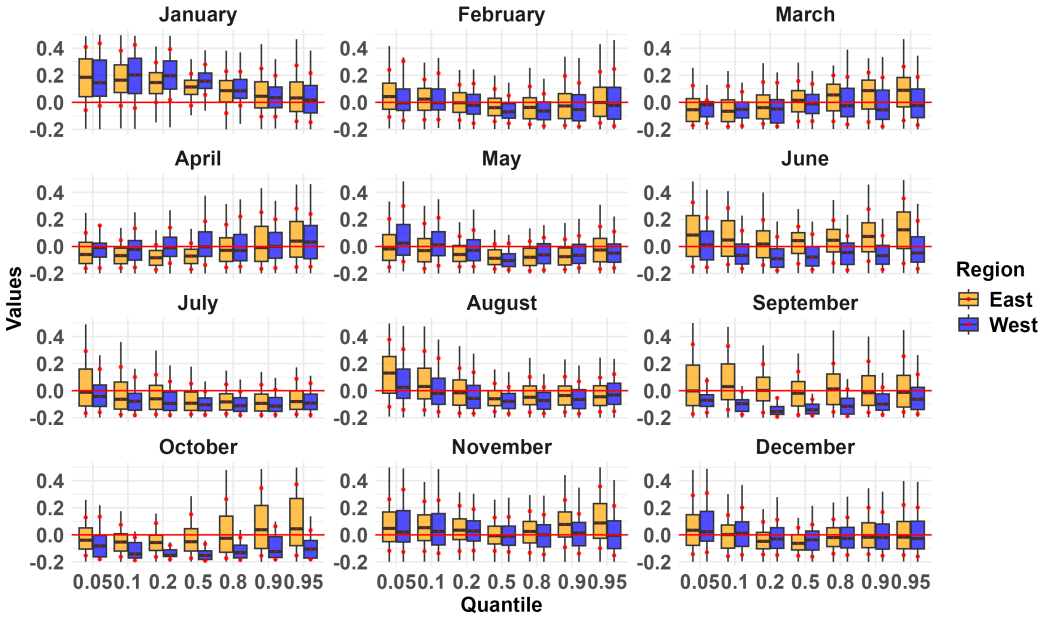


(b)

FIGURE 4.5: Panel (a) presents box plot estimates of $\overline{\pi}_{k,W}$ and $\overline{\pi}_{k,E}$, representing the west and east regions of the GDR, respectively. Similarly, panel (b) presents box plot estimates of $\overline{\mu}_{k,W}$ and $\overline{\mu}_{k,E}$ for k^{th} covariate i.e SOI. Red stars mark the 10th percentile, mean, and 90th percentile, and the red line in (b) marks zero.



(a)



(b)

FIGURE 4.6: Panel (a) presents box plot estimates of $\overline{\pi}_{k,W}$ and $\overline{\pi}_{k,E}$, representing the west and east regions of the GDR, respectively. Similarly, panel (b) presents box plot estimates of $\overline{\mu}_{k,W}$ and $\overline{\mu}_{k,E}$ for k^{th} covariate i.e DMI. Red stars mark the 10th percentile, mean, and 90th percentile, and the red line in (b) marks zero.

4.3.2 Performance Measures

We use the performance metric defined by Bertolacci et al., 2019 as described in Chapter 2 and 3,

$$D_i^{(q)} = \hat{E}\left[\log \frac{y_i^{(q)*}}{\hat{y}_i^{(q)}} \mid \boldsymbol{y}_i\right].$$

When $D_i^{(q)}$ equals zero, it indicates that the posterior predictive quantile is equal to the empirical quantile. In cases where the model overestimates or underestimates the empirical quantile, $D_i^{(q)}$ takes a positive or negative value, respectively, with larger absolute values reflecting poorer model performance.

Figure 4.7 presents the $D_i^{(q)}$ values for models with two covariates (SOI and DMI) and one covariate (SOI only), enabling a comparative analysis of their performance.The results indicate that the model incorporating two covariates provides slight improvements, achieving nominal performance for specific months and quantiles, notably the upper quantiles in January and February and the lower quantiles in June, where performance was previously suboptimal under the model with SOI only. In June, both models appear to overestimate rainfall for the middle quantiles. However, the inclusion of the DMI in the model does not appear to substantially improve predictive performance, potentially indicating that the impact of the DMI on monthly rainfall is statistically weaker compared to that of the SOI. Separate Figures for the model with one covariate and two covariates are attached in Appendix C2.

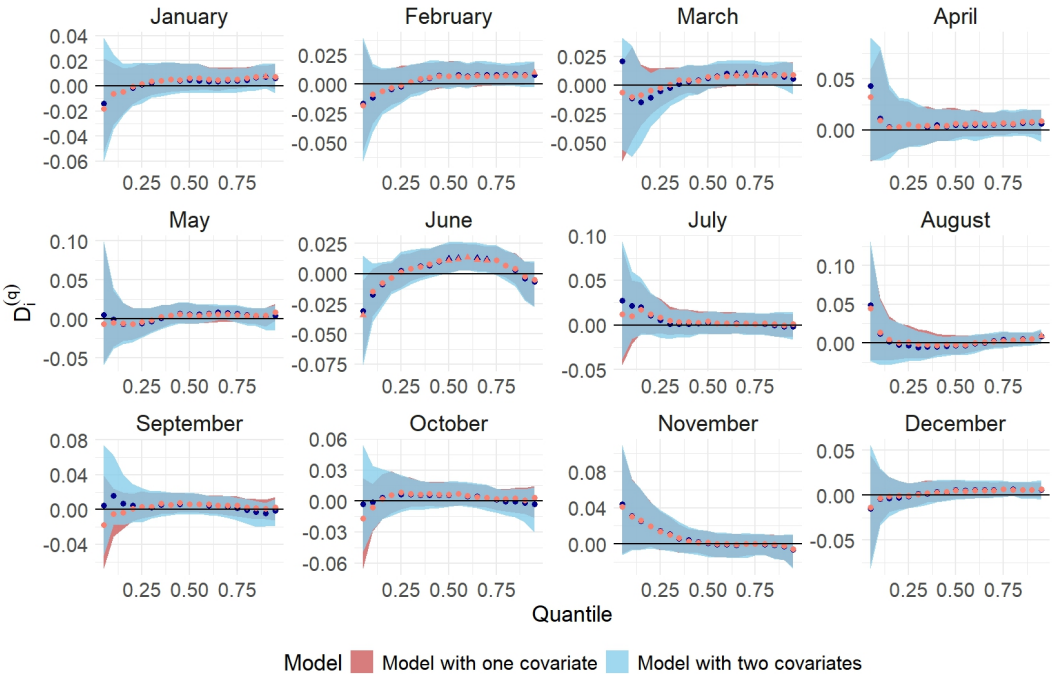


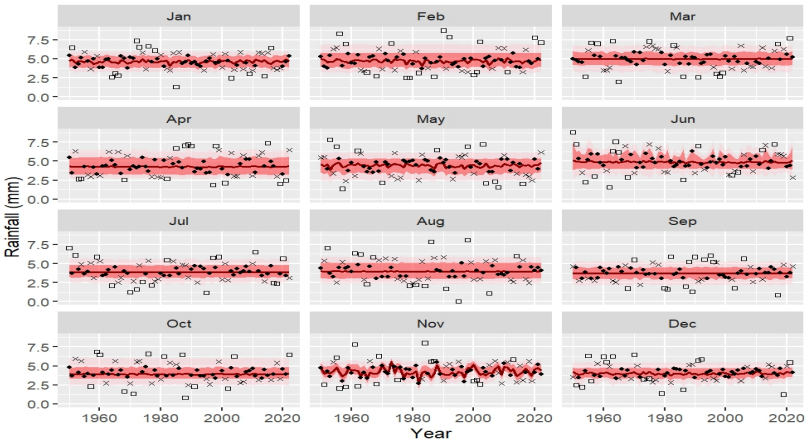
FIGURE 4.7: Median values of $D_i^{(q)}$, represented as points, and the interquartile ranges, shown as shaded bands in red and blue, across all sites (n_s) for each quantile and month. Points indicate nominal quantiles where the IQR of $D_i^{(q)}$, $i = 1, ..., n_s$, contains zero, while triangles in these colors denote that is not for models with one (SOI only) and two covariates (SOI and DMI) respectively (black line indicates zero).

4.3.3 Model diagnostics for spatial and temporal dependence

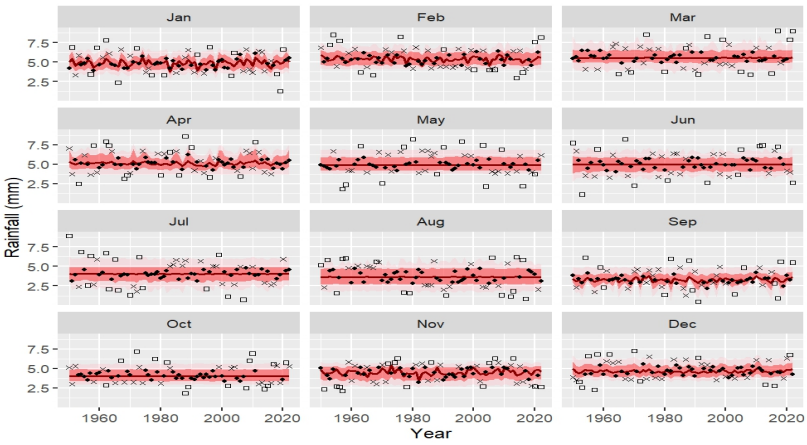
Using the posterior predictive distribution, we assess the model's ability to capture spatial and temporal dependencies. Figure 4.8 illustrates the model's ability to capture spatially varying rainfall patterns using posterior predictive coverage (PPC). This figure plots pointwise quantiles for the posterior predictive distribution for the six selected sites. Here, we plot $y^{(0.5)*}$ which is posterior predictive quantile for the median and shaded bands for $[y^{(0.25)*}, y^{(0.75)*}]$ corresponding to 50% PPC and $[y^{(0.1)*}, y^{(0.9)*}]$ which is 80% PPC, against years, along with the observed data.

These figures show that 80% of observed data falls inside the quantiles of 0.1 and 0.9 and 50% of observed data generally falls inside 0.25 and 0.75 rainfall quantiles for the 6 stations considered.

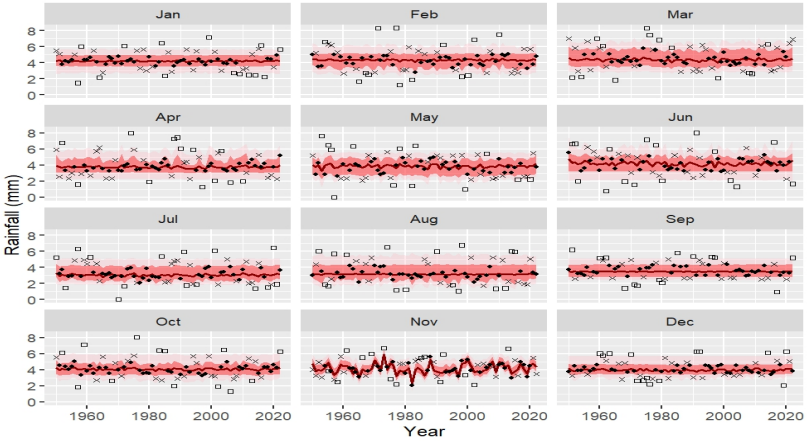
To evaluate the model's ability to capture temporal dependence, we examine autocorrelation plots of the residuals, defined as the posterior predictive median minus the median of the observed data, for six selected sites. The results in Figure 4.9 show that residual autocorrelations generally fall within confidence bounds 95%, indicating that the model adequately captures the temporal dependence in the rainfall series.



a) Sydney Botanic Gardens [66006]

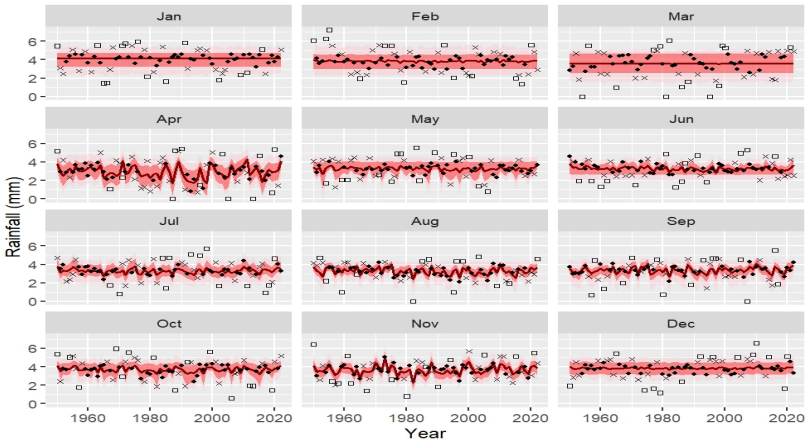


b) Yamba Pilot Station [58012]

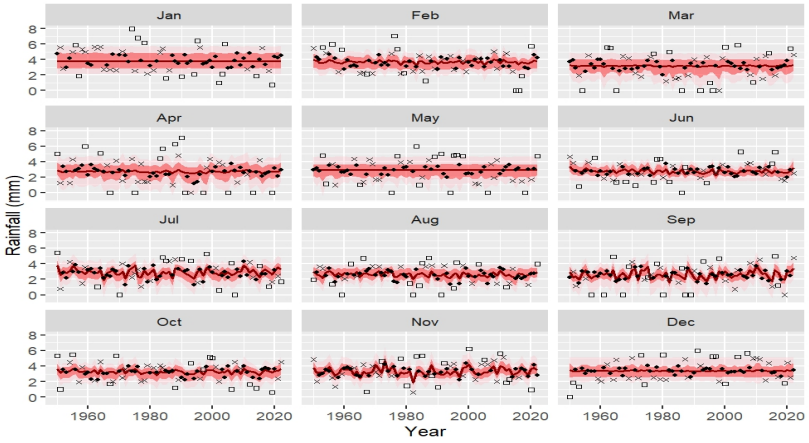


c) Moruya Heads [69018]

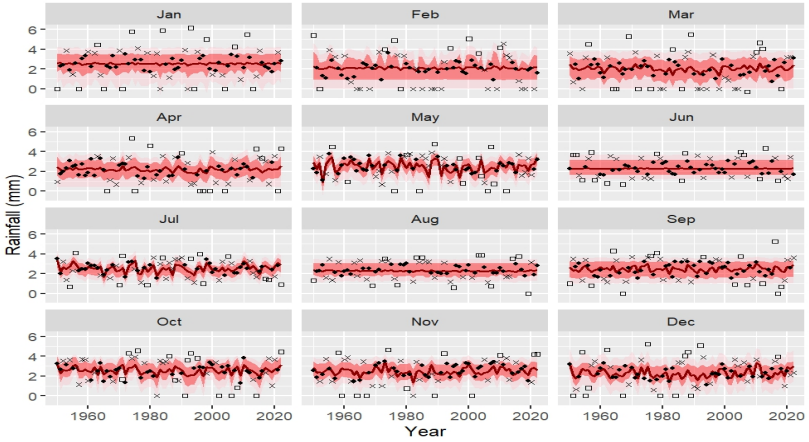
FIGURE 4.8: Posterior predicted median rainfall ($y^{(0.5)*}$, dark red line) against years, along with observed data for six stations, Open squares mark observed data values which lie outside the 80% PPC band ($[y^{(0.1)*}, y^{(0.9)*}]$, light pink), crosses mark those inside the 80% PPC band but outside the 50% band ($[y^{(0.25)*}, y^{(0.75)*}]$, dark pink), and filled circles mark those within the 50% band.



d) Dunedoo [64009]

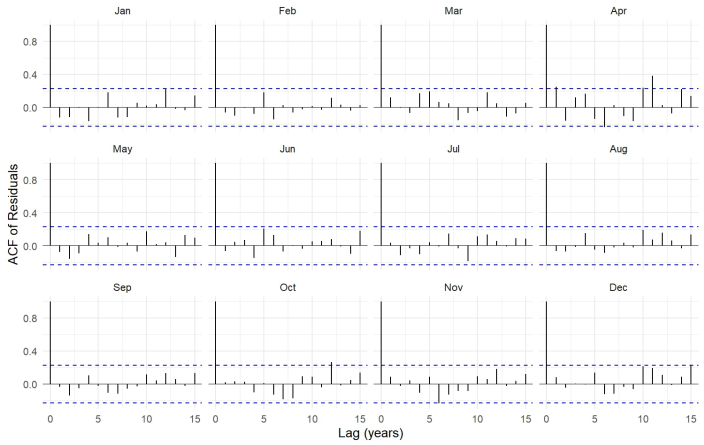


e) Collarenebri (Albert St) [48031]

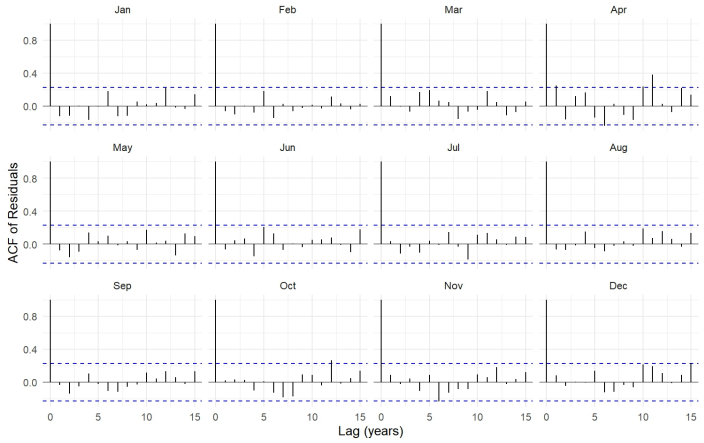


f) Pooncarie [47013]

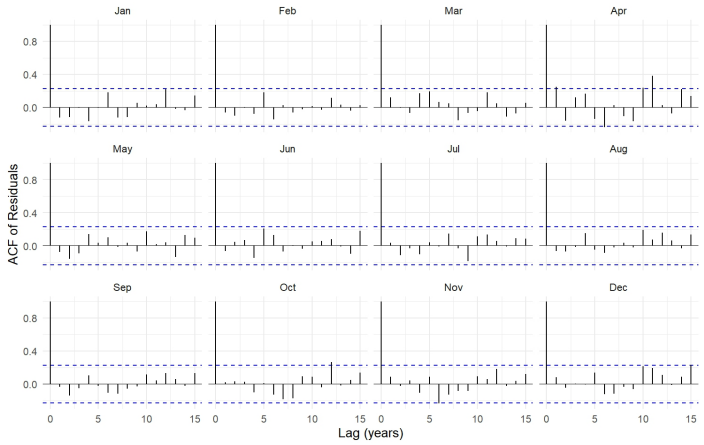
FIGURE 4.8: (continued from previous page)



a) Sydney Botanic Gardens [66006]

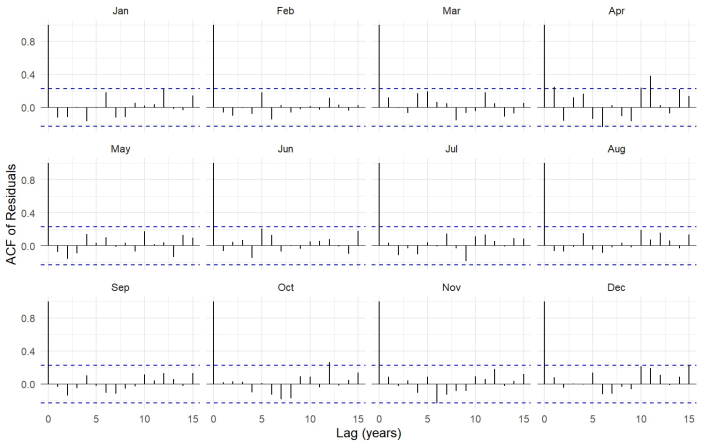


b) Yamba Pilot Station [58012]

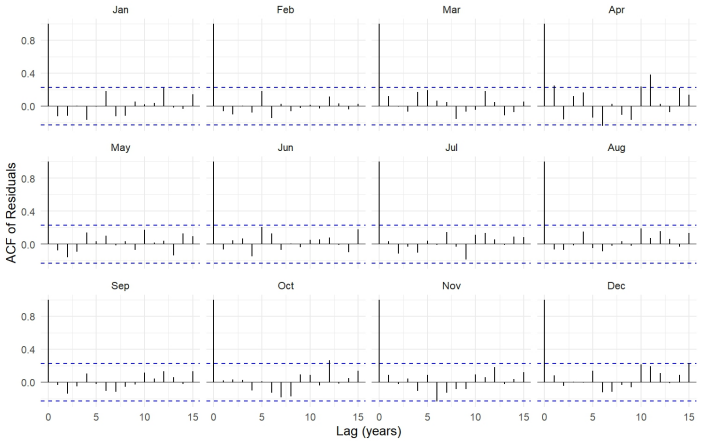


c) Moruya Heads [69018]

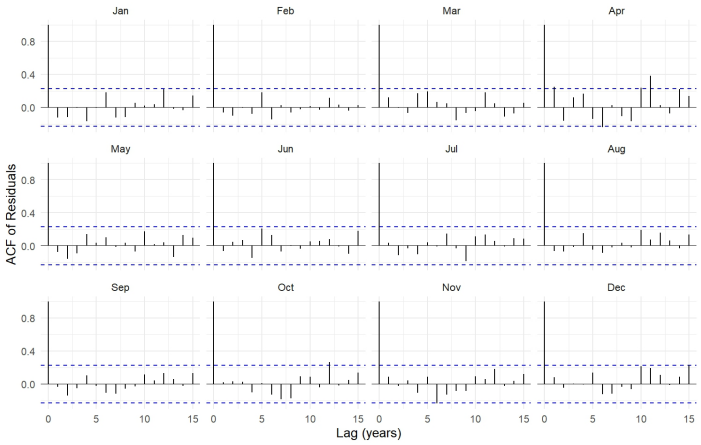
FIGURE 4.9: Autocorrelation plots of residuals ($y^{(0.5)*} - \hat{y}^{(0.5)}$) for six selected stations. Dashed lines indicate 95% confidence bounds for white noise.



d) Dunedoo [64009]



e) Collarenebri (Albert St) [48031]



f) Pooncarie [47013]

FIGURE 4.9: (continued from previous page)

4.4 Discussion and Conclusion

We develop a model that extends the approach in Chapter 3, which considered the spatially varying effect of a single climate index (SOI), incorporating multiple indices such as SOI and DMI and allowing spatially varying inference on their joint influence. The key contribution is combining spatial dependence with joint variable selection to identify which combinations of indices affect monthly rainfall quantiles, how their effects change over months, and at which specific locations in NSW these effects are important. This is achieved by multinomial logistic regression for the joint inclusion probabilities, where the argument link function varies spatially via a $\mathcal{GP}$ prior.

The incorporation of multiple variables and spatial dependency provides significant advantages. By modeling the joint probability structure, the proposed approach captures the interdependencies between climate indices that simpler models often overlook. This approach is critical for understanding how ENSO and IOD jointly influence rainfall patterns across regions, revealing spatial variations in their effects that are not apparent when assuming spatial independence.

Additionally, the multinomial logistic regression framework allows a probabilistic interpretation of the inclusion of SOI and DMI, identifying regions where their impacts are most pronounced. The results highlight the dominant influence of DMI in the spring months, particularly September and October, across multiple quantiles. Notably, SOI shows a significant impact across all quantiles in November, while the influence of DMI diminishes. The findings of these associations with the median rainfall align with prior studies of the mean rainfall, emphasizing the critical influence of ENSO and IOD on rainfall variability and the complex interplay between these phenomena during this specific season. The apparent influence of the DMI during summer and autumn, when the IOD is typically inactive, is most evident at the lower quantiles in January and March, suggesting an effect on drier conditions.

This may reflect indirect processes similar to those observed for ENSO, although such a relationship has not been documented in previous studies and warrants further investigation.

Performance measures indicate slight improvements in certain months and quantiles when adding the DMI alongside the SOI. However, in general, there is no substantial difference between the two models. This limited improvement may be due to the weaker statistical influence of the DMI on monthly rainfall distribution compared to the SOI.

In conclusion, this chapter highlights the importance of extending regression models to include multiple variables and spatial dependencies for analyzing the influence of climate indices on monthly rainfall distribution. The proposed approach provides a comprehensive framework for making spatially explicit inferences on the joint impact of SOI and DMI on different parts of the monthly rainfall distribution. Future research aims to explore the inclusion of additional climate indices or their interactions to further generalize the findings and refine our understanding of climate-driven rainfall variability.

CHAPTER 5

Discussion

This thesis presents the development of a Bayesian hierarchical framework for variable selection within spatial quantile regression with an application to modeling the relationship between global climate indices and monthly rainfall in NSW. Despite significant advances in physical and statistical modeling aimed at identifying the relationships between climate indices and daily rainfall extremes in Australia (see Yildirim and Rahman, 2022; He et al., 2025; Dey et al., 2021; Nguyen et al., 2021, etc.), there is a lack of probabilistic models that incorporate Bayesian variable selection within the quantile regression framework to assess the influence of global climate indices on the entire distribution of monthly rainfall.

Each chapter of this thesis builds model complexity, progressing from initial models that treat locations independently to more sophisticated models that incorporate spatial dependence within the variable selection framework. Chapter 2 introduces the foundational model which assumes spatial independence among observation sites (a common and still widely used approach in climate and engineering sciences), providing a baseline against which the benefits of incorporating spatial dependencies can be evaluated. Subsequent chapters discuss the development and implementation of advanced statistical models that integrate spatial dependence. This approach allows for the understanding of the spatial heterogeneity of the impact of climate indices on monthly rainfall distribution. A key advantage of our spatially dependent framework is that we provide out-of-sample predictions of these impacts, measured by the mean of the regression coefficients and the inclusion probabilities of climate indices for various quantiles of monthly rainfall in NSW.

5.1 Summary of Thesis: New Insights into the Impacts of ENSO and IOD on Monthly Rainfall Distribution in New South Wales

Chapter 1 introduces the aims of the thesis, from motivation to thesis contributions, and then describes the statistical background and inferential techniques related to the quantile regression framework. It briefly outlines a Bayesian hierarchical regression model, where the prior on the regression coefficients is assumed to follow a spike and slab prior, Mitchell and Beauchamp, 1988, to allow for variable selection as given by Equation 1.7. Notably, the probability of including global climate indices in the regression model is given a spatially varying prior, see Equations 1.8 and 1.9.

Chapter 2 applies this model to six weather stations in NSW (out of a total of 711) and analyzes monthly rainfall at these locations independently. These six stations were selected to represent different climate regions in NSW, three stations were coastal, while three were inland stations. Figure 1.2 shows the distribution of these stations. Chapter 2 reports on the inference on the relationship between SOI, DMI, and SAM and the monthly rainfall quantiles. The results show how the impact of these indices on monthly rainfall varies across quantiles, with some indices having a greater impact on the lower and upper quantiles of the distribution than on the median of the rainfall distribution. Performance metrics show that the predicted lower and upper quantiles of the distribution overestimate and underestimate the empirical quantile, respectively. This indicates that the model underestimates the overall variability of monthly rainfall, presumably due to the lack of other covariates in the model that might help explain the variability.

Chapter 3 builds on this by incorporating spatial dependence using a single global climate indicator, that is, SOI, to progressively develop the model for more complex scenarios. This contributes to climate science by assessing the varying spatial influence of ENSO on monthly

rainfall. The results show that the impact of SOI is generally stronger west of the GDR, but an anomaly is observed in June where SOI shows a more pronounced influence to the east of the GDR than on the west, while in November, SOI exerts a uniform effect across all regions. Furthermore, from January to June, SOI mainly affects the tails of the rainfall distribution rather than the median, confirming that ENSO's influence on rainfall is more pronounced in extremes, a key insight missed by conventional methods focused on mean rainfall.

Chapter 4 builds on the methodology in Chapter 3 by including the SOI and the DMI as covariates. The results reveal that during September and October, IOD exerts a stronger influence than ENSO in all quantiles, reducing the marginal inclusion probability of SOI while exhibiting a pronounced negative impact on rainfall west of the GDR. However, in November the strong influence of ENSO persists even after accounting for IOD's effects, as noted in Chapter 3. While previous studies have highlighted the influence of ENSO and IOD on mean rainfall during Austral spring (Pepler et al., 2014), our findings extend this understanding by analyzing the quantiles of monthly rainfall distribution, showing that IOD dominates early spring variability, whereas ENSO remains as the key driver in late spring and beyond. Model performance shows minor improvements with DMI included, but overall, SOI remains the stronger predictor of monthly rainfall, with DMI's influence being more seasonal and region-specific.

In all chapters, we use data augmentation to facilitate the MCMC scheme which was used to estimate and make inferences regarding the model.

5.2 Future directions

In this section, we discuss the limitations of our current model and propose avenues for future research.

A significant advancement would be to model the quantiles jointly. This approach addresses the issue of quantile crossing, which typically arises when quantiles are estimated independently. Recent work in joint quantile regression has focused on preventing quantile crossing, ensuring a consistent increase in quantile estimates across $q \in (0, 1)$. This advancement is crucial to maintaining the natural sequence among quantiles, improving the accuracy and consistency of statistical analyses. Notable efforts include non-crossing techniques for a specified set of quantiles by Bondell et al., 2010, and pioneering work using Gaussian processes by Tokdar and Kadane, 2012. Subsequent advances by Das and Ghosal, 2017 integrated spline into this framework. Reich, 2012 adapted these methodologies for spatial analysis using Bernstein polynomials and spatially varying coefficients. Further innovations by Yang and Tokdar, 2017 and Tokdar and Kadane, 2012 developed new parameterizations to model non-crossing quantile planes that can be applied over complex convex predictor domains in spatial analyses.

To our knowledge, none of the above analyses employs joint quantile modeling in a spatial context, coupled with Bayesian variable selection. To address this, work is underway to integrate a joint quantile approach into our final model. Jointly modeling the quantiles will constrain the parameter space of the variables in the inclusion probability, the π's in Equation 4.6, as well as the parameter space of the variables in the regression model, the β and the μ's in Equations 4.2 and 4.4 respectively, in a way that is difficult to express explicitly. However, it is possible to impose these constraints via a Metropolis-Hasting step in the Gibbs sampling scheme, which rejects those samples that violate this constraint. Ongoing work will look at how to achieve these constraints in a more fundamental and less computationally expensive manner.

In addition, to better understand how monthly rainfall is linked to global climate indices, we could add more indicators, like the Southern Annular Mode (SAM) and the Madden-Julian Oscillation (MJO), along with their interactions, and extend the analysis to account for any potential nonlinear relationships.

Furthermore, the framework can be extended to other regions of Australia. It can also incorporate a term to formally assess the east–west divide of the Great Dividing Range, including possible interactions with latitude and longitude. Moreover, the model can be combined with climate projections to simulate future rainfall for applications such as agriculture and water management.

All code for this work has been written in R (R Core Team, 2025), and in future it will be made available on GitHub to support reproducibility.

References

Abram, N. J., Wright, N. M., Ellis, B., et al. (2020). Coupling of indo-pacific climate variability over the last millennium. *Nature*, 579:385–392.

Albert, J. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of The American Statistical Association - J AMER STATIST ASSN*, 88:669–679.

Alhamzawi, R. and Algamal, Z. Y. (2019). Bayesian bridge quantile regression. *Communications in Statistics - Simulation and Computation*, 48(3):944–956.

Alhamzawi, R. and Yu, K. (2013). Conjugate priors and variable selection for Bayesian quantile regression. *Computational Statistics Data Analysis*, 64:209–219.

ARC Centre of Excellence for Climate Extremes (2022). Understanding Australia's rainfall. Technical report, ARC Centre of Excellence for Climate Extremes. Accessed: 2024-09-27.

Barndorff-Nielsen, O. E. and Shephard, N. (2002). Non-Gaussian Ornstein–Uhlenbeck-based Models and Some of Their Uses in Financial Economics. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 63(2):167–241.

Barnes, M., King, M., Reeder, M., and Jakob, C. (2023). The dynamics of slow-moving coherent cyclonic potential vorticity anomalies and their links to heavy rainfall over the eastern seaboard of australia. *Quarterly Journal of the Royal Meteorological Society*, 149.

Benoit, D., Alhamzawi, R., and Yu, K. (2013). Bayesian lasso binary quantile regression. *Computational Statistics*, 28.

Bertolacci, M., Cripps, E., Rosen, O., Lau, J., and Cripps, S. (2019). Climate inference on daily rainfall across the Australian continent, 1876–2015. *The Annals of Applied Statistics*, 13:683–712.

Bondell, H. D., Reich, B. J., and Wang, H. (2010). Noncrossing quantile regression curve estimation. *Biometrika*, 97(4):825–838.

Brown, J. N., McIntosh, P. C., Pook, M. J., and Risbey, J. S. (2009). An investigation of the links between ENSO flavors and rainfall processes in southeastern Australia. *Monthly Weather Review*, 137(11):3786 – 3795.

Browning, S. A. and Goodwin, I. D. (2013). Large-scale influences on the evolution of winter subtropical maritime cyclones affecting Australia's east coast. *Monthly Weather Review*, 141(7):2416 – 2431.

Cai, W., Cowan, T., and Sullivan, A. (2009). Recent unprecedented skewness towards positive indian ocean dipole occurrences and its impact on Australian rainfall. *Geophysical Research Letters*, 36.

Cai, W., Purich, A., Cowan, T., Rensch, P. V., and Weller, E. (2014). Did climate change–induced rainfall trends contribute to the Australian millennium drought? *Journal of Climate*, 27(9):3145 – 3168.

Cai, W., Rensch, P., Cowan, T., and Hendon, H. (2012). An asymmetry in the IOD and ENSO teleconnection pathway and its impact on Australian climate. *Journal of Climate*, 25:6318–6329.

Cai, W., Rensch, P. V., Cowan, T., and Hendon, H. H. (2011). Teleconnection pathways of ENSO and the IOD and the mechanisms for impacts on Australian rainfall. *Journal of Climate*, 24(15):3910 – 3923.

Castillo, I. and Mismer, R. (2018). Empirical Bayes analysis of spike and slab posterior distributions.

Castillo-Mateo, J., Asín, J., Cebrián, A. C., Mateo-Lázaro, J., and Abaurrea, J. (2023). Bayesian variable selection in generalized extreme value regression: Modeling annual maximum temperature. *Mathematics*, 11(3).

Chung, C. and Power, S. (2017). The non-linear impact of El Niño, La Niña and the Southern Oscillation on seasonal and regional Australian precipitation. *Journal of Southern Hemisphere Earth System Science*, 67:25–45.

CSIRO and Bureau of Meteorology (2024). Australia's changing climate: Australia's weather and climate including temperature, fire weather, rainfall, heavy rainfall, streamflow, tropical cyclones, snowfall. state of the climate 2024. Technical report, Commonwealth Scientific and Industrial Research Organisation (CSIRO) and Bureau of Meteorology (BoM). Accessed: 2025-08-16.

Dao, M. and Nguyen, L. (2024). Variable selection in macroeconomic stress test: a Bayesian quantile regression approach. *Empirical Economics*, pages 1–57.

Dao, M., Wang, M., Ghosh, S., and Ye, K. (2022). Bayesian variable selection and estimation in quantile regression using a quantile-specific prior. *Computational Statistics*.

Das, P. and Ghosal, S. (2017). Bayesian quantile regression using random B-spline series prior. *Computational Statistics Data Analysis*, 109:121–143.

De Fondeville, R. and Davison, A. C. (2016). High-dimensional peaks-over-threshold inference. *Biometrika*, 105(3):575–592.

Delage, F. and Power, S. (2020). The impact of global warming and the El Niño-Southern Oscillation on seasonal precipitation extremes in Australia. *Climate Dynamics*, 54.

Department of Agriculture, Water and the Environment (2021). Australia state of the environment 2021. Technical report, Department of Agriculture, Water and the Environment, Australian Government. Accessed: 2024-09-27.

Dey, R., Bador, M., Alexander, L., and Lewis, S. (2021). The drivers of extreme rainfall event timing in Australia. *International Journal of Climatology*, 41.

Dowdy, A., Pepler, A., Di Luca, A., Cavicchia, L., Mills, G., Evans, J., Louis, S., Mcinnes, K., and Walsh, K. (2019). Review of Australian east coast low pressure systems and associated extremes. *Climate Dynamics*, 53.

Duan, Q., McGrory, C., Brown, G., Mengersen, K., and Wang, Y.-G. (2022). Spatio-temporal quantile regression analysis revealing more nuanced patterns of climate change: A study of long-term daily temperature in Australia. *PLOS ONE*, 17:e0271457.

Duc, H. N., Rivett, K., MacSween, K., and Le-Anh, L. (2017). Association of climate drivers with rainfall in New South Wales, Australia, using Bayesian model averaging. *Theoretical*

and Applied Climatology, 127:169–185.

El Adlouni, S., Salaou, G., and St-Hilaire, A. (2017). Regularized Bayesian quantile regression. *Communications in Statistics - Simulation and Computation*, 47.

Eugster, M., Knaus, J., Porzelius, C., Schmidberger, M., and Vicedo, E. (2011). Hands-on tutorial for parallel computing with R. *Computational Statistics*, 26:219–239.

Evan Hajani, A. R. and Ishak, E. (2017). Trends in extreme rainfall in the state of New South Wales, Australia. *Hydrological Sciences Journal*, 62(13):2160–2174.

Favre, A., Hewitson, B., Tadross, M., Lennard, C., and Cerezo-Mota, R. (2012). Relationships between cut-off lows and the semiannual and southern oscillations. *Climate Dynamics*, 38:1473–1487.

Feng, P. and et al. (2018). Impacts of rainfall extremes on wheat yield in semi-arid cropping systems in eastern Australia. *Climatic Change*, 147(3-4):555–569.

Fowler, K., Peel, M., Saft, M., Peterson, T. J., Western, A., Band, L., Petheram, C., Dharmadi, S., Tan, K. S., Zhang, L., Lane, P., Kiem, A., Marshall, L., Griebel, A., Medlyn, B. E., Ryu, D., Bonotto, G., Wasko, C., Ukkola, A., Stephens, C., Frost, A., Gardiya Weligamage, H., Saco, P., Zheng, H., Chiew, F., Daly, E., Walker, G., Vervoort, R. W., Hughes, J., Trotter, L., Neal, B., Cartwright, I., and Nathan, R. (2022). Explaining changes in rainfall–runoff relationships during and after Australia's millennium drought: a community perspective. *Hydrology and Earth System Sciences*, 26(23):6073–6120.

Galarza, C. E. and Lachos, V. H. (2021). *ald: Asymmetric Laplace Distribution*. R package version 1.3.1.

Garreaud, R. D. (2007). Precipitation and circulation covariability in the extratropics. *Journal of Climate*, 20:4789–4797.

Gelfand, A. E. (2000). Gibbs sampling. *Journal of the American Statistical Association*, 95(452):1300–1304.

Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). *Bayesian Data Analysis*. Chapman and Hall/CRC, 2nd edition.

Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2(4):1360 – 1383.

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

Gergis, J., Gallant, A. J. E., Braganza, K., et al. (2012). On the long-term context of the 1997–2009 'Big Dry' in South-Eastern Australia: insights from a 206-year multi-proxy rainfall reconstruction. *Climatic Change*, 111:923–944.

Ghosh, J. (2009). Default prior distributions and efficient posterior computation in Bayesian factor analysis. *Journal of computational and graphical statistics : a joint publication of American Statistical Association, Institute of Mathematical Statistics, Interface Foundation of North America*, 18:306–320.

Ghosh, J., Li, Y., and Mitra, R. (2015). On the use of Cauchy prior distributions for Bayesian logistic regression. *Bayesian Analysis*, 13.

Gillett, N. P., Thompson, D. W. J., and Stott, P. A. (2006). External forcing of the southern annular mode and its influence on regional climate. *Journal of Climate*, 19:4489–4499.

Griffin, J. and Brown, P. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, 5:171–188.

Hauser, S., Grams, C. M., Reeder, M. J., McGregor, S., Fink, A. H., and Quinting, J. F. (2020). A weather system perspective on winter–spring rainfall variability in southeastern Australia during El Niño. *Quarterly Journal of the Royal Meteorological Society*, 146(731):2614–2633.

Haylock, M. and Nicholls, N. (2000). Trends in extreme rainfall indices for an updated high quality data set for Australia, 1910–1998. *International Journal of Climatology*, 20(13):1533–1541.

He, J., Li, S., Wang, B., Zhang, L., and Duan, K. (2025). Quantifying the impacts of ENSO on Australian summer rainfall extremes during 1960–2020. *Journal of Hydrology*, 654:132834.

Hendon, H. H., Thompson, D. W., and Wheeler, M. C. (2007). Australian rainfall and surface temperature variations associated with the southern hemisphere annular mode. *Journal of Climate*, 20(11):2452–2467.

Hill, K. J. and Raphael, M. N. (2009). Atmospheric circulation changes and their impact on climate over southern hemisphere extratropical regions. *Journal of Geophysical Research*, 114:D18116.

Holgate, C., Evans, J. P., Taschetto, A. S., Gupta, A. S., and Santoso, A. (2022). The impact of interacting climate modes on east Australian precipitation moisture sources. *Journal of Climate*, 35(10):3147 – 3159.

Holgate, C., Falster, G., Gillett, Z., Goswami, P., Grant, M., Hobeichi, S., Hoffmann, D., Jiang, X., Jin, C., Lu, X., Mu, M., Page, J., Parker, T., Vogel, E., Abram, N., Evans, J., Gallant, A., Henley, B., Kala, J., and Ukkola, A. (2025). Physical mechanisms of meteorological drought development, intensification and termination: an australian review. *Communications Earth Environment*, 6.

Holsclaw, T., Wolpert, R. L., and Calder, C. A. (2017). Bayesian nonparametric space-time analysis of multi-type criminal event occurrences. *Journal of the American Statistical Association*, 112(518):651–662.

Hope, P., Timbal, B., and Fawcett, R. (2010). Associations between rainfall variability in the southwest and southeast of Australia and their evolution through time. *International Journal of Climatology*, 30(9):1360–1371.

Hopkins, L. and Holland, G. (1997). Australian heavy-rain days and associated east coast cyclones: 1958-92. *Journal of Climate - J CLIMATE*, 10:621–635.

Irving, D., Risbey, J., Squire, D., Matear, R., Tozer, C., Monselesan, D., Ramesh, N., Reddy, P., and Freund, M. (2024). A multi-model likelihood analysis of unprecedented extreme rainfall along the east coast of Australia. *Meteorological Applications*, 31.

Islam, F., Imteaz, M., and Rasel, H. (2017). Analysing the effect of lagged climate indices on rainfall predictability for western Australia's north coast region. In *Proceedings of the 19th International Congress on Modelling and Simulation (MODSIM2017)*, pages 1353–1359. Modelling and Simulation Society of Australia and New Zealand.

Jakob, D. and Walland, D. (2016). Variability and long-term change in Australian temperature and precipitation extremes. *Weather and Climate Extremes*, 14:36–55.

Kavelaars, X., Mulder, J., and Kaptein, M. (2022). Bayesian multivariate logistic regression for superiority and inferiority decision-making under treatment heterogeneity.

Khastagir, A., Hossain, I., and Anwar, A. H. M. (2022). Efficacy of linear multiple regression and artificial neural network for long-term rainfall forecasting in western australia. *Meteorology and Atmospheric Physics*, 134.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press, New York.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50. Accessed 26 Aug. 2024.

Kohns, D. and Szendrei, T. (2023). Horseshoe prior Bayesian quantile regression. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 73(1):193–220.

Kotz, S., Kozubowski, T., and Podgorski, K. (2001). *The Laplace Distribution and Generalizations*.

Kozumi, H. and Kobayashi, G. (2009). Gibbs Sampling Methods for Bayesian Quantile Regression. Discussion Papers 2009-02, Kobe University, Graduate School of Business Administration.

Lane, T., King, A., Perkins-Kirkpatrick, S., Pitman, A., Alexander, L., Arblaster, J., Bindoff, N., Bishop, C., Black, M., Bradstock, R., Clarke, H., Gallant, A., Grose, M., Holbrook, N., Holland, G., Hope, P., Karoly, D., Raupach, T., and Ukkola, A. (2023). Attribution of extreme events to climate change in the Australian region – a review. *Weather and Climate Extremes*, 42:100622.

Leamer, E. E. (1978). *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. John Wiley & Sons, New York, NY.

Li, Q., Xi, R., and Lin, N. (2010). Bayesian regularized quantile regression. *Bayesian Analysis*, 1:1–26.

Lim, E.-P., Hendon, H., Arblaster, J., Delage, F., Nguyen, H., Min, S.-K., and Wheeler, M. (2016). The impact of the southern annular mode on future changes in southern hemisphere rainfall: The sam's impact on future SH rainfall. *Geophysical Research Letters*, 43.

Ma, Y., Sun, J., Dong, T., Yu, W., and Dong, W. (2022). More profound impact of CP ENSO on Australian spring rainfall in recent decades. *Climate Dynamics*, 60:1–15.

Martin, A. D., Quinn, K. M., and Park, J. H. (2011). *MCMCpack: Markov Chain Monte Carlo in R*. R package version 1.0.

Mbigi, D. and Xiao, Z. (2023). The southern annular mode: its influence on interannual variability of rainfall in north australia. *Climate Dynamics*, 62:1–14.

McBride, J. L. and Nicholls, N. (1983). Seasonal relationships between Australian rainfall and the Southern Oscillation. *Monthly Weather Review*, 111(10):1998–2004.

McGregor, S., Gallant, A., and Van Rensch, P. (2024). Quantifying ENSO's impact on Australia's regional monthly rainfall risk. *Geophysical Research Letters*, 51(6):e2023GL106298. e2023GL106298 2023GL106298.

Meneghini, B., Simmonds, I., and Smith, I. (2007). Association between Australian rainfall and the southern annular mode. *International Journal of Climatology*, 27:109 – 121.

Meyers, G., McIntosh, P., Pigot, L., and Pook, M. (2007). The years of El Niño, La Niña, and interactions with the Tropical Indian Ocean. *Journal of Climate*, 20(13):2872 – 2880.

Min, S.-K., Cai, W., and Whetton, P. (2013). Influence of climate variability on seasonal extremes over Australia. *Journal of Geophysical Research: Atmospheres*, 118(2):643–654.

Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404):1023–1032.

Montazerolghaem, M., Vervoort, R. W., Minasny, B., and Mcbratney, A. (2016). Long-term variability of the leading seasonal modes of rainfall in south-eastern Australia. *Weather*

and Climate Extremes, 13.

Nguyen, H., Wheeler, M. C., Hendon, H. H., Lim, E.-P., and Otkin, J. A. (2021). The 2019 flash droughts in subtropical eastern Australia and their association with large-scale climate drivers. *Weather and Climate Extremes*, 32:100321.

Nicholls, N. (1989). Sea surface temperatures and Australian winter rainfall. *Journal of Climate*, 2(9):965–973.

Nicholls, N. (2010). Local and remote causes of the southern Australian autumn-winter rainfall decline, 1958–2007. *Climate Dynamics*, 34:835–845.

Opoku-Ankomah, Y. and Cordery, I. (1993). Temporal variation of relations between New South Wales rainfall and the Southern Oscillation. *International Journal of Climatology*, 13(1):51–64.

Pepler, A. and Dowdy, A. (2021). Intense east coast lows and associated rainfall in eastern australia. *Journal of Southern Hemisphere Earth Systems Science*, 71.

Pepler, A., Timbal, B., Rakich, C., and Coutts-Smith, A. (2014). Indian Ocean Dipole overrides ENSO's influence on cool season rainfall across the eastern seaboard of Australia. *Journal of Climate*, 27:3816–3826.

Polson, N., Scott, J., and Windle, J. (2012). Bayesian inference for logistic models using polya-gamma latent variables. *Journal of the American Statistical Association*, 108.

Polson, N. G. and Scott, J. G. (2010). Local shrinkage rules, lévy processes and regularized regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74.

Power, S. B. and Callaghan, J. (2016). Variability in severe coastal flooding, associated storms, and death tolls in Southeastern Australia since the Mid–Nineteenth Century. *Journal of Applied Meteorology and Climatology*, 55(5):1139 – 1149.

Pui, A., Sharma, A., Santoso, A., and Westra, S. (2012). Impact of the El Niño–Southern Oscillation, Indian Ocean Dipole, and Southern Annular Mode on daily to subdaily rainfall characteristics in East Australia. *Monthly Weather Review*, 140(5):1665 – 1682.

R Core Team (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Accessed: 27 August 2025.

Rakich, C. S., Holbrook, N. J., and Timbal, B. (2008). A pressure gradient metric capturing planetary-scale influences on eastern australian rainfall. *Geophysical Research Letters*, 35(8).

Raut, B., Jakob, C., and Reeder, M. (2014). Rainfall changes over southwestern Australia and their relationship to the southern annular mode and ENSO. *Journal of Climate*, 27:5801–5814.

Reed, C. (2011). Bayesian parameter estimation and variable selection for quantile regression. PhD thesis, Brunel University, School of Information Systems, Computing and Mathematics.

Reich, B. (2012). Spatiotemporal quantile regression for detecting distributional changes in environmental processes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61.

Reid, K. J., Barnes, M. A., Gillett, Z. E., Parker, T., Udy, D. G., Ayat, H., Boschat, G., Bowden, A., Grosfeld, N. H., King, A. D., Richardson, D., Shao, Y., Teckentrup, L., Trewin, B., Hope, P., Zhou, L., Borowiak, A. R., Holgate, C. M., and Isphording, R. N. (2025). A multiscale evaluation of the wet 2022 in eastern australia. *Journal of Climate*, 38(4):909 – 929.

Risbey, J., Pook, M., McIntosh, P., Ummenhofer, C., and Meyers, G. (2009a). Characteristics and variability of synoptic features associated with cool season rainfall in southeastern Australia. *International Journal of Climatology - INT J CLIMATOL*, 29:1595–1613.

Risbey, J., Pook, M., McIntosh, P., Wheeler, M., and Hendon, H. (2009b). On the remote drivers of rainfall variability in Australia. *Monthly Weather Review*, 137:3233–3253.

Saji, N. H., Goswami, B. N., Vinayachandran, P. N., and Yamagata, T. (1999). A dipole mode in the tropical indian ocean. *Nature*, 401:360–363.

Saunders, K., Stephenson, A. G., Taylor, P. G., and Karoly, D. (2017). The spatial distribution of rainfall extremes and the influence of El Niño Southern Oscillation. *Weather and*

Climate Extremes, 18:17–28.

Smith, M. and Kohn, R. (1996). Nonparametric regression using Bayesian variable selection. *Journal of Econometrics*, 75(2):317–343.

Soomro, S. and Yu, K. (2025). Bayesian fractional polynomial approach to quantile regression and variable selection with application in the analysis of blood pressure among us adults. *Journal of Applied Statistics*, 52(1):97–118.

Steel, M. and ndez, C. (2000). Bayesian regression analysis with scale mixtures of normals. *Econometric Theory*, 16:80–101.

Sun, X., Renard, B., Thyer, M., Westra, S., and Lang, M. (2015). A global analysis of the asymmetric effect of ENSO on extreme precipitation. *Journal of Hydrology*, 530:51–65.

Tareghian, R. and Rasmussen, P. F. (2013). Statistical downscaling of precipitation using quantile regression. *Journal of Hydrology*, 487:122–135.

Thibaud, E., Mutzner, R., and Davison, A. C. (2013). Threshold modeling of extreme spatial rainfall. *Water resources research*, 49(8):4633–4644.

Tian, Y. and Song, X. (2019). Bayesian bridge-randomized penalized quantile regression. *Computational Statistics Data Analysis*, 144:106876.

Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions. *The Annals of Statistics*, 22(4):1701 – 1728.

Timbal, B. et al. (2010). Understanding the anthropogenic nature of the observed rainfall decline across south-eastern Australia. CAWCR Technical Report 26, Centre for Australian Weather and Climate Research, Melbourne.

Tokdar, S. T. and Kadane, J. B. (2012). Simultaneous Linear Quantile Regression: A Semiparametric Bayesian Approach. *Bayesian Analysis*, 7(1):51 – 72.

Tozer, C., Risbey, J., Pook, M., Monselesan, D., Irving, D., Ramesh, N., and Richardson, D. (2024). A tale of two novembers: Confounding influences on La Niña's relationship with rainfall in Australia. *Monthly Weather Review*, 152.

Tozer, C. R., Risbey, J. S., Monselesan, D. P., Pook, M. J., Irving, D., Ramesh, N., and Squire, D. T. (2023). Impacts of ENSO on Australian rainfall: What not to expect. *Journal*

of Southern Hemisphere Earth Systems Science, 73(1):77–81.

Troup, A. J. (1965). The "southern oscillation". *Quarterly Journal of the Royal Meteorological Society*, 91(390):490–506.

Ummenhofer, C. C., England, M. H., McIntosh, P. C., Meyers, G. A., Pook, M. J., Risbey, J. S., and Taschetto, A. S. (2009). What causes southeast Australia's worst droughts? *Geophysical Research Letters*, 36(4).

Ummenhofer, C. C., Gupta, A. S., Briggs, P. R., England, M. H., McIntosh, P. C., Meyers, G. A., Pook, M. J., Raupach, M. R., and Risbey, J. S. (2011). Indian and pacific ocean influences on southeast Australian drought and soil moisture. *Journal of Climate*, 24(5):1313 – 1336.

Van Dijk, A., Beck, H., Crosbie, R., Jeu, R., Liu, Y., Podger, G., Timbal, B., and Viney, N. (2013). The millennium drought in southeast Australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society. *Water Resources Research*, 49.

Verdon-Kidd, D. C. and Kiem, A. S. (2009). Nature and causes of protracted droughts in southeast Australia: Comparison between the Federation, WWII, and Big Dry droughts. *Geophysical Research Letters*, 36(22).

Waha, K., Clarke, J., Dayal, K., Freund, M., Heady, C., Parisi, I., and Vogel, E. (2022). Past and future rainfall changes in the Australian midlatitudes and implications for agriculture. *Climatic Change*, 170(3):1–21.

Wahba, G. (1990). *Spline models for observational data*. SIAM.

Webster, P. J., Moore, A. M., Loschnigg, J. P., and Leben, R. R. (1999). Coupled ocean-atmosphere dynamics in the indian ocean during 1997-98. *Nature*, 401:356–360.

Westra, S., Alexander, L. V., and Zwiers, F. W. (2013). Global increasing trends in annual maximum daily precipitation. *Journal of climate*, 26(11):3904–3918.

Wood, S. (2013). Applications of Bayesian smoothing splines. In *Bayesian Theory and Applications*. Oxford University Press.

Wood, S. and Kohn, R. (1998). A bayesian approach to robust binary nonparametric regression. *Journal of the American Statistical Association*, 93(441):203–213.

Wu, W. and Leonard, M. (2019). Impact of ENSO on dependence between extreme rainfall and storm surge. *Environmental Research Letters*, 14.

Xi, R., Li, Y., and Hu, Y. (2015). Bayesian quantile regression based on the empirical likelihood with spike and slab priors. *Bayesian Analysis*, 11.

Yang, Y. and Tokdar, S. T. (2017). Joint estimation of quantile planes over arbitrary predictor spaces. *Journal of the American Statistical Association*, 112(519):1107–1120.

Yildirim, G. and Rahman, A. (2022). Homogeneity and trend analysis of rainfall and droughts over Southeast Australia. *Natural Hazards*, 112:1–27.

Yilmaz, A., Imteaz, M., and Perera, B. (2017). Investigation of non-stationarity of extreme rainfalls and spatial variability of rainfall intensity–frequency–duration relationships: a case study of victoria, Australia. *International Journal of Climatology*, 37(1):430–442.

Yu, K., Chen, C., and Reed, C. (2013). Bayesian variable selection in quantile regression. *Statistics and its interface*, 6:261–274.

Yu, K., Lu, Z., and Stander, J. (2003). Quantile regression: Applications and current research areas. *Statistician*, 52(3):331–350.

Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics Probability Letters*, 54(4):437–447.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno De Finetti*, pages 233–243.

Zhang, X., Wang, J., Zwiers, F., and Groisman, P. (2010). The influence of large-scale climate variability on winter maximum daily precipitation over North America. *Journal of Climate - J CLIMATE*, 23:2902–2915.

Zheng, F., Thibaud, E., Leonard, M., and Westra, S. (2015). Assessing the performance of the independence method in modeling spatial extreme rainfall. *Water Resources Research*, 51(9):7744–7758.

Appendices

Appendix A

A1 Additional Simulation Study

A1.1 With one covariate

TABLE A1.1: Mean differences and standard errors of MAD for Proposed and QR-SSVS methods for the data generated using different scale parameters for the model with one covariate.

(a) $\sigma = 0.3$

Quantile	Method	Mean Difference	Standard Error	p-value
$q = 0.05$	Proposed vs QR-SSVS	-0.0508	0.0095	<0.001*
$q = 0.5$	Proposed vs QR-SSVS	-0.0011	0.0013	0.0590
$q = 0.95$	Proposed vs QR-SSVS	0.0172	0.0107	0.1110

(b) $\sigma = 1$

Quantile	Method	Mean Difference	Standard Error	p-value
$q = 0.05$	Proposed vs QR-SSVS	-0.0411	0.0112	<0.001*
$q = 0.5$	Proposed vs QR-SSVS	-0.0046	0.0032	0.1571
$q = 0.95$	Proposed vs QR-SSVS	0.0130	0.0176	0.4632

(c) $\sigma = 5$

Quantile	Method	Mean Difference	Standard Error	p-value
$q = 0.05$	Proposed vs QR-SSVS	-1.4845	0.1674	<0.001*
$q = 0.5$	Proposed vs QR-SSVS	-0.1443	0.0269	<0.001*
$q = 0.95$	Proposed vs QR-SSVS	-1.2866	0.1707	<0.001*

Note: * denotes p-value < 0.05

A1.2 With two covariates

TABLE A1.2: Mean differences and standard errors of MAD for Proposed and QR-SSVS methods for the data generated using different scale parameters for the model with two covariates.

(a) $\sigma = 0.3$

Quantile	Method	Mean Difference	Standard Error	p-value
$q = 0.05$	Proposed vs QR-SSVS	-0.0660	0.0093	<0.001*
$q = 0.5$	Proposed vs QR-SSVS	-0.0036	0.0013	0.0059*
$q = 0.95$	Proposed vs QR-SSVS	-0.0094	0.0119	0.4325

(b) $\sigma = 1$

Quantile	Method	Mean Difference	Standard Error	p-value
$q = 0.05$	Proposed vs QR-SSVS	-0.04813	0.0144	<0.001*
$q = 0.5$	Proposed vs QR-SSVS	0.0042	0.0023	0.0012*
$q = 0.95$	Proposed vs QR-SSVS	-0.0034	0.0205	0.8652

(c) $\sigma = 5$

Quantile	Method	Mean Difference	Standard Error	p-value
$q = 0.05$	Proposed vs QR-SSVS	-1.8213	0.1914	<0.001*
$q = 0.5$	Proposed vs QR-SSVS	-0.2408	0.0345	<0.001*
$q = 0.95$	Proposed vs QR-SSVS	-2.2134	0.2548	<0.001*

Note: * denotes p-value < 0.05

A2 Posterior Derivations of Bayesian Variable Selection in QR

(1) Generate $\mathrm{v}_t^{[j+1]}|\boldsymbol{y}, \sigma^{[j]}, \boldsymbol{\beta}^{[j]}$ from $\mathcal{GIG}(\frac{1}{2}, a_{\mathrm{V}_t}, b_{\mathrm{V}})$, where $a_{\mathrm{V}_t} = \sqrt{\frac{(y_t - \boldsymbol{x}_t\boldsymbol{\beta})^2}{\zeta^2\sigma}}$ and $b_{\mathrm{V}} = \sqrt{\frac{2}{\sigma} + \frac{\theta^2}{\zeta^2\sigma}}$.

(2) Generate $\sigma^{[j+1]}|\boldsymbol{y}, \mathbf{v}^{[j+1]}, \boldsymbol{\beta}^{[j]}$ from $\mathcal{IG}(a_\sigma, b_\sigma)$, where $a_\sigma = a + \frac{3T}{2} - 1$ and $b_\sigma = \frac{1}{2\zeta^2}(\tilde{\boldsymbol{y}} - X\boldsymbol{\beta})'V^{-1}(\tilde{\boldsymbol{y}} - X\boldsymbol{\beta}) + \sum_{t=1}^{T} \mathrm{v}_t + b$ where $V = \mathrm{diag}(\mathbf{v})$ and $\tilde{\boldsymbol{y}} = \boldsymbol{y} - \theta\mathbf{v}$.

(3) Generate $p(\boldsymbol{y}^{miss}|\boldsymbol{y}^{obs}, \mathbf{v}^{[j+1]}, \sigma^{[j+1]}) \sim N(\mu_y, \sigma_y^2)$ where $\mu_y = \theta\mathbf{v}$ and $\sigma_y^2 = \mathbf{v}\sigma\zeta^2$.

(4) Generate $\boldsymbol{\gamma}^{[j+1]}$ and $\boldsymbol{\beta}^{[j+1]}$ jointly from $p(\boldsymbol{\gamma}, \boldsymbol{\beta}|\boldsymbol{y}, \mathbf{v}^{[j+1]}, \sigma^{[j+1]})$.

The variable selection problem is to compute the posterior distribution of $\boldsymbol{\gamma}$

$$p(\boldsymbol{\gamma}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\gamma})p(\boldsymbol{\gamma})}{p(\boldsymbol{y})}. \tag{A1.1}$$

Let $\boldsymbol{\beta}_A = [\beta_0, \boldsymbol{\beta}_{A_1}]$, $X_A = [1, X_{A_1}]$ and the marginal density of $\boldsymbol{y}$ given the model for $\boldsymbol{\gamma} = 1$ is,

$$p(\boldsymbol{y}|\boldsymbol{\gamma},\ \mathrm{v}, X_A) = \int_{\mathbf{R}^K} p(\boldsymbol{y}|\boldsymbol{\gamma}, \sigma, \boldsymbol{\beta}_A,\ \mathrm{v}, X_A) p(\boldsymbol{\beta}_A|\boldsymbol{\gamma}, \sigma, X_A) d\boldsymbol{\beta}_A\,. \tag{A1.2}$$

Let $\mathbf{V}_y = \zeta^2\sigma\mathrm{diag}(\ \mathrm{v})$, and $\tilde{\boldsymbol{y}} = \boldsymbol{y} - \theta\ \mathrm{v}$,

$$\begin{aligned} p(\tilde{\boldsymbol{y}}|\boldsymbol{\gamma}, X_A,\ \mathrm{v}, \sigma) = \int_{-\infty}^{+\infty} & \frac{1}{(2\pi)^{\frac{T}{2}}|\mathbf{V}_y|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\tilde{\boldsymbol{y}} - X_A\boldsymbol{\beta}_A)'\mathbf{V}_y^{-1}(\tilde{\boldsymbol{y}} - X_A\boldsymbol{\beta}_A)\right) \\ & \frac{1}{(2\pi)^{\frac{K_\gamma+1}{2}}|T\sigma(X_A'X_A)^{-1}|^{\frac{1}{2}}} \exp\left(\frac{-1}{2}\frac{\boldsymbol{\beta}'(X_A'X_A)\boldsymbol{\beta}_A}{T\sigma}\right) d\boldsymbol{\beta}_A\,, \end{aligned} \tag{A1.3}$$

where $||A_1|| = K_\gamma$.

$$\begin{aligned} \propto & \frac{1}{|\mathbf{V}_y|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\tilde{\boldsymbol{y}}'\mathbf{V}_y^{-1}\tilde{\boldsymbol{y}})\right) \frac{1}{|T\sigma(X_A'X_A)^{-1}|^{\frac{1}{2}}} \\ & \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}(\boldsymbol{\beta}_A'(X_A'\mathbf{V}_y^{-1}X_A)\boldsymbol{\beta}_A - 2\boldsymbol{\beta}_A'X_A'\tilde{\boldsymbol{y}})\right) \exp\left(-\frac{1}{2}\frac{\boldsymbol{\beta}_A'(X_A'X_A)\boldsymbol{\beta}_A}{T\sigma}\right) d\boldsymbol{\beta}_A, \end{aligned} \tag{A1.4}$$

Let $\boldsymbol{\mu}_\beta = \mathbf{V}_\beta X_A'\mathbf{V}_y^{-1}\tilde{\boldsymbol{y}}$ and $\mathbf{V}_\beta = (X_A'\mathbf{V}_y^{-1}X_A + \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1})^{-1}$ where $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} = \frac{(X_A'X_A)}{T\sigma}$.

Then:

$$p(\tilde{\boldsymbol{y}}|\boldsymbol{\gamma}, X_A,\ \mathrm{v}, \sigma) \propto \frac{|X_A' X_A|^{\frac{1}{2}}}{|\mathbf{V}_y|^{\frac{1}{2}}} \frac{|\mathbf{V}_\beta|^{\frac{1}{2}}}{(T\sigma)^{\frac{1}{2}}} \exp\left(\frac{-1}{2}(\tilde{\boldsymbol{y}}' \mathbf{V}_y^{-1} \tilde{\boldsymbol{y}})\right) \exp\left(\frac{\boldsymbol{\mu}_\beta' \mathbf{V}_\beta^{-1} \boldsymbol{\mu}_\beta}{2}\right), \quad \text{(A1.5)}$$

$$p(\tilde{\boldsymbol{y}}|\boldsymbol{\gamma}, X_A,\ \mathrm{v}, \sigma) \propto \frac{|X_A' X_A|^{\frac{1}{2}}}{|\mathbf{V}_y|^{\frac{1}{2}}} \frac{|\mathbf{V}_\beta|^{\frac{1}{2}}}{(T\sigma)^{\frac{1}{2}}} \exp\left\{\frac{1}{2}(\tilde{\boldsymbol{y}}' [\mathbf{V}_y^{-1} - \mathbf{V}_y^{-1} X_A X_A' \mathbf{V}_y^{-1}] \tilde{\boldsymbol{y}} - \tilde{\boldsymbol{y}}' \mathbf{V}_y^{-1} \tilde{\boldsymbol{y}})\right\}. \quad \text{(A1.6)}$$

Generate posterior inclusion probability π_k,

(i)

$$\pi_k = p(\gamma_k = 1|\tilde{\boldsymbol{y}}, \mathbf{v}, \sigma, X_A) = \frac{p(\tilde{\boldsymbol{y}}|\gamma_k = 1, \mathbf{v}, \sigma, X_A) p(\gamma_k = 1)}{\sum_{l=0}^{1} p(\tilde{\boldsymbol{y}}|\gamma_k = l, \mathbf{v}, \sigma, X) p(\gamma_k = l)},$$

where $p(\gamma_k = 1) = \pi_0\ \forall k$.

Generate $\pi_0 \sim \text{Beta}(K_\gamma + 1, K - (K_\gamma + 1))$ where K is the number of covariates and K_γ is the number of non-zero covariates.

Generate $\boldsymbol{\beta}_A|\boldsymbol{y}, \mathbf{v}^{[j+1]}, \sigma^{[j+1]}, \boldsymbol{\gamma}^{[j+1]}$ from $N(\boldsymbol{\mu}_\beta, \boldsymbol{V}_\beta)$, where $\boldsymbol{V}_\beta = (X_A' \mathbf{V}_y^{-1} X_A + \{T\sigma(X_A' X_A)^{-1}\}^{-1})^{-1}$ and $\boldsymbol{\mu}_\beta = \boldsymbol{V}_\beta\ (X_A' \mathbf{V}_y^{-1} \tilde{\boldsymbol{y}})$ where $X_A = [1, X_{A_1}]$ and $\boldsymbol{\beta}_{A_0}|\boldsymbol{y}, \mathbf{v}^{[j+1]}, \sigma^{[j+1]}, \boldsymbol{\gamma}^{[j+1]} = 0$.

A2.1 Other Conditional posterior derivations

Using the likelihood function in Equation 1.5 and prior for parameters $\boldsymbol{\beta}$, v, σ, conditional posterior distribution can be derived as follows,

$$
\begin{aligned}
p(\sigma|\boldsymbol{y}, \mathrm{v}, X) &\propto p(\boldsymbol{y}|\sigma, \mathrm{v}, X)p(\mathrm{v}|\sigma)p(\sigma), \\
&\propto \frac{1}{(2\pi)^{\frac{T}{2}}|\mathbf{V}_y|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\tilde{\boldsymbol{y}} - X\boldsymbol{\beta})'\mathbf{V}_y^{-1}(\tilde{\boldsymbol{y}} - X\boldsymbol{\beta})\right) \\
&\quad \times \frac{1}{\sigma^T} \exp\left(-\frac{\sum \mathrm{v}_t}{\sigma}\right) \times \frac{1}{\sigma^{a+1}} \exp\left(-\frac{b}{\sigma}\right), \\
&\propto \frac{1}{\sigma^{\frac{T}{2}}|V|^{\frac{1}{2}}} \exp\left(-\frac{1}{2\zeta^2\sigma}(\tilde{\boldsymbol{y}} - X\boldsymbol{\beta})'V^{-1}(\tilde{\boldsymbol{y}} - X\boldsymbol{\beta})\right) \\
&\quad \times \frac{1}{\sigma^T} \exp\left(-\frac{\sum \mathrm{v}_t}{\sigma}\right) \times \frac{1}{\sigma^{a+1}} \exp\left(-\frac{b}{\sigma}\right).
\end{aligned}
$$

So the conditional posterior distribution of σ is,

$p(\sigma|\boldsymbol{y}, \mathbf{v}, \boldsymbol{\beta})$ from $\mathcal{IG}(a_\sigma, b_\sigma)$, where $a_\sigma = a + \frac{3T}{2} - 1$ and $b_\sigma = \frac{1}{2\zeta^2}(\tilde{\boldsymbol{y}} - X\boldsymbol{\beta})'V^{-1}(\tilde{\boldsymbol{y}} - X\boldsymbol{\beta}) + \sum_{t=1}^{T} \mathrm{v}_t + b$ where $V = \mathrm{diag}(\mathbf{v})$ and $\tilde{\boldsymbol{y}} = \boldsymbol{y} - \theta\mathbf{v}$.

The conditional posterior derivation of the v is as follows,

$$
\begin{aligned}
p(\mathrm{v}|\boldsymbol{y}, \sigma, X) &\propto p(\boldsymbol{y}|\mathrm{v}, X)p(\mathrm{v}|\sigma), \\
&\propto \frac{1}{(2\pi)^{\frac{T}{2}}|\mathbf{V}_y|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\tilde{\boldsymbol{y}} - X\boldsymbol{\beta})'\mathbf{V}_y^{-1}(\tilde{\boldsymbol{y}} - X\boldsymbol{\beta})\right) \times \frac{1}{\sigma^T} \exp\left(-\frac{\sum \mathrm{v}_t}{\sigma}\right).
\end{aligned}
$$

Thus, the quadratic form expands as:

$$
(\tilde{\boldsymbol{y}} - X\boldsymbol{\beta})'\mathbf{V}_y^{-1}(\tilde{\boldsymbol{y}} - X\boldsymbol{\beta}) = \sum_{t=1}^{T} \frac{(y_t - \boldsymbol{x}_t\boldsymbol{\beta} - \theta\mathrm{v}_t)^2}{\zeta^2\sigma\mathrm{v}_t}. \tag{A1.7}
$$

Substituting back:

$$p(\mathrm{v}|\boldsymbol{y},\sigma,X) \propto \mathrm{v}_t^{-1/2} \prod_{t=1}^{T} \exp\left(-\frac{1}{2}\frac{(y_t - \boldsymbol{x}_t\boldsymbol{\beta} - \theta\mathrm{v}_t)^2}{\zeta^2\sigma\mathrm{v}_t} - \frac{\mathrm{v}_t}{\sigma}\right). \tag{A1.8}$$

Rewriting in product form:

$$\mathrm{v}_t^{-1/2} \exp\left\{-\frac{1}{2}\left(a_{\mathrm{v}_t}^2\ \mathrm{v}_t^{-1} + b_{\mathrm{v}}^2\ \mathrm{v}_t\right)\right\}, \tag{A1.9}$$

where

$$a_{\mathrm{v}_t}^2 = \frac{(y_t - \boldsymbol{x}_t\boldsymbol{\beta})^2}{\zeta^2\sigma}, \quad b_{\mathrm{v}}^2 = \frac{2}{\sigma} + \frac{\theta^2}{\zeta^2\sigma}.$$

Since Equation A1.9 is the kernel of a generalized inverse Gaussian ($\mathcal{GIG}$) distribution, we have:

$$\mathrm{v}_t|y,\boldsymbol{\beta},\sigma \sim \mathcal{GIG}\left(\frac{1}{2}, a_{\mathrm{v}_t}, b_{\mathrm{v}}\right), \tag{A1.10}$$

where the probability density function of the $\mathcal{GIG}$ distribution, $\mathcal{GIG}(\nu,a,b)$, is given by:

$$f(x|\nu,a,b) = \frac{(b/a)^{\nu/2}}{2K_\nu(ab)} x^{\nu-1} \exp\left\{-\frac{1}{2}\left(a^2x^{-1} + b^2x\right)\right\}, \quad x > 0, \tag{A1.11}$$

with $-\infty < \nu < \infty$, and $a, b \geq 0$, and $K_\nu(\cdot)$ is a modified Bessel function of the third kind (Barndorff-Nielsen and Shephard (2002)).

A2.2 Missing Value Generation

To express the probability density $p(y_t)$ for the response variable y_t, following an Asymmetric Laplace Distribution (ALD), we need to start by writing the probability density function (PDF) of the ALD. y_t follows distribution, $\mathcal{AL}(0,\sigma,q)$

$$p(y_t) = \frac{q(1-q)}{\sigma} \exp\{-\frac{\rho|y_t|}{\sigma}\},$$

where $\rho|y_t| = q|y_t|$ if $y_t \geq 0$ and $\rho|y_t| = (1-q)|y_t|$ for $y_t < 0$.

Suppose y_t^{miss} denotes a missing value (Missing values completely at Random) at t, and this also follows a ALD ($y_t^{miss} \sim \mathcal{AL}(0, \sigma, q)$) and conditional on v_t, y_t^{miss} can be written as,

$$y_t^{miss}|\mathrm{v}_t, \sigma, q \sim N(\mu_y, \sigma_y^2),$$

where $\mu_y = \theta \mathrm{v}_t$ and $\sigma_y^2 = \sigma\zeta^2\mathrm{v}_t$.

Note that this has been done using the properties of the mixture representation of ALD (refer to Section **??**).

Suppose y_t^{miss} denotes a missing value at t and $\boldsymbol{y}_\nu$ where ν denotes set of indices for which y_t is missing.

Let $\Delta_t = 1$ if y_t observed and 0 otherwise. $\nu = \{j; \Delta_j = 0\}$,

Conditional on the the values of the posterior iterations in $\sigma^{[j+1]}$ and $\mathrm{v}^{[j+1]}$ in MCMC Scheme in section 2.2, generate the $\boldsymbol{y}^{miss[j+1]}$.

Appendix B

B1 Model Validation via Simulation Study

Additionally, We chose the month of November as Figure 3.5 panel (a) illustrates the significant impact of SOI on November rainfall, with the inclusion probability, π, and the mean of coefficient size, μ, showing high significance across all quantiles. For instance, the probability that SOI affects the 5th, 50th, and 95th quantiles of monthly rainfall is close to 1 for both the east and west regions. Unlike in August, where the upper quantile had a minimal impact, in November the inclusion probability remains high across all quantiles, indicating a strong influence of SOI. Similarly, Figure 3.5 panel (b) shows a substantial and positive SOI coefficient across all quantiles for November, reinforcing the overall significant relationship.

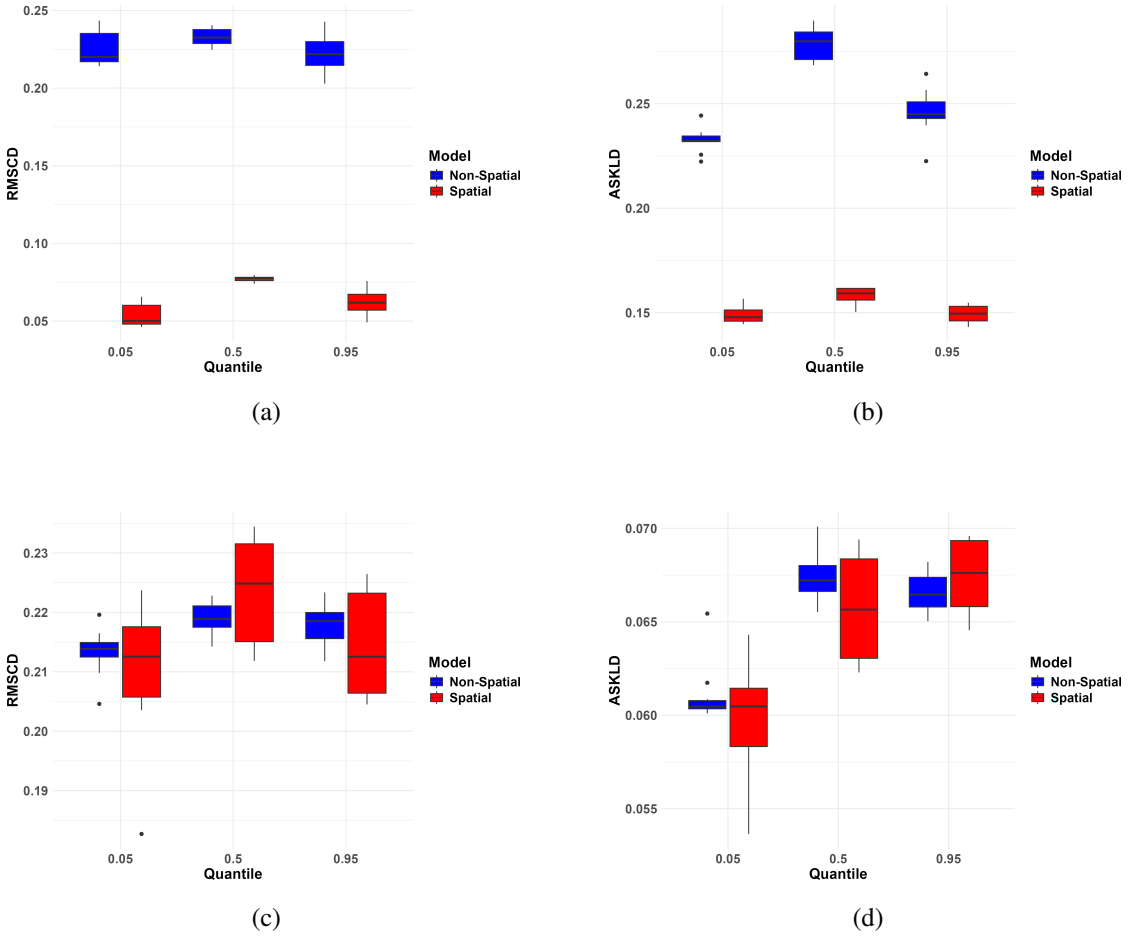


FIGURE B1.1: Panel (a) and (b) are boxplots for RMSCD and ASKLD measures for data generated under DGP_1. Similarly, panel (c) and (d) show the measures when generated from DGP_0 at $q = (0.05, 0.5, 0.95)$ for month November, color blue and red represent model $\mathcal{M}_0$ and $\mathcal{M}_1$ respectively.

B2 SOI vs Rainfall

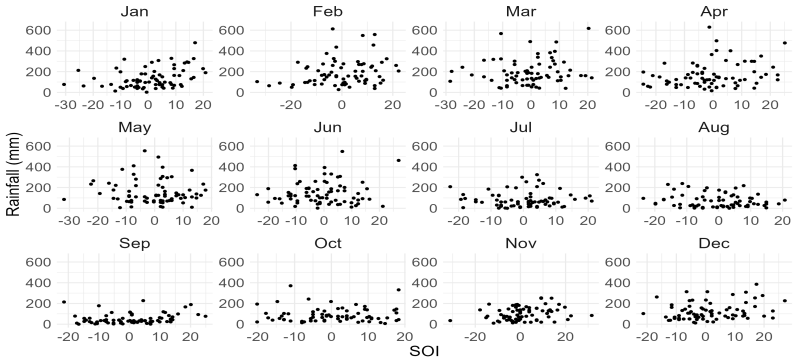


FIGURE B1.2: Scatter plot of SOI versus monthly rainfall for Yamba Pilot Station [58012] .

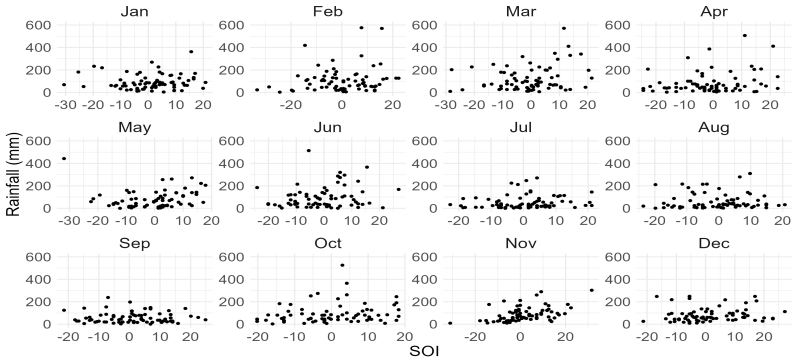


FIGURE B1.3: Scatter plot of SOI versus monthly rainfall for Collarenebri (Albert St) [48031].

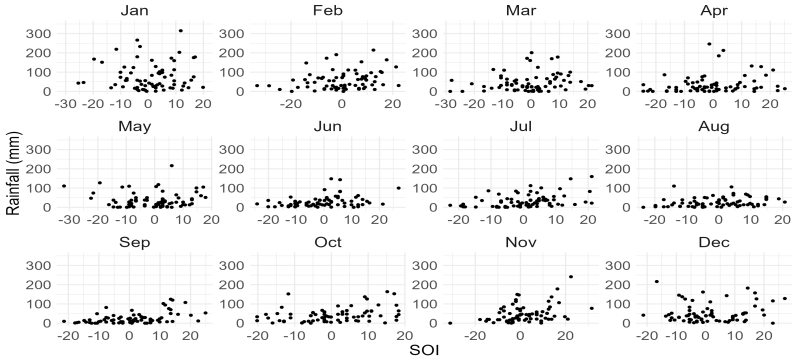


FIGURE B1.4: Scatter plot of SOI versus monthly rainfall for Moruya Heads [69018].

B3 Posterior Derivations for Algorithm

(1) $p(\mathbf{V}|\boldsymbol{Y},\boldsymbol{\sigma},\boldsymbol{B}) = \prod_{i=1}^{n_s}\prod_{t=1}^{T} p(\mathrm{v}_{t,i}|\boldsymbol{y}_i,\sigma_i,\boldsymbol{\beta}_i) \sim \mathcal{GIG}(\frac{1}{2}, a_{\mathrm{V}_{t,i}}, b_{\mathrm{V}_i})$, where $a_{\mathrm{V}_{t,i}} = \frac{(y_{t,i}-\boldsymbol{x}_t\boldsymbol{\beta}_i)^2}{\zeta^2\sigma_i}$ and $b_{\mathrm{V}_i} = \frac{2}{\sigma_i} + \frac{\theta^2}{\zeta^2\sigma_i}$ where $\mathcal{GIG}$ is the Generalized Inverse Gaussian distribution (see Kozumi and Kobayashi (2009)).

(2) $p(\boldsymbol{\sigma}|\boldsymbol{Y},\mathbf{V},\boldsymbol{B}) = \prod_{i=1}^{n_s} p(\sigma_i|\boldsymbol{y}_i,\mathbf{v}_i,\boldsymbol{\beta}_i) \sim \mathcal{IG}(a_{\sigma_i}, b_{\sigma_i})$, where $a_{\sigma_i} = a + \frac{3T}{2} - 1$ and $b_{\sigma_i} = \frac{1}{2\zeta^2}(\tilde{\boldsymbol{y}}_i - X\boldsymbol{\beta}_i)'V_i^{-1}(\tilde{\boldsymbol{y}}_i - X\boldsymbol{\beta}_i) + \sum_{t=1}^{T}\mathrm{v}_{t,i} + b$ where $V_i = \mathrm{diag}(\mathbf{v}_i)$ and $\tilde{\boldsymbol{y}}_i = \boldsymbol{y}_i - \theta\mathbf{v}_i$.

(3) $p(\boldsymbol{Y}^{\mathrm{miss}}|\boldsymbol{Y}^{\mathrm{obs}},\mathbf{V},\boldsymbol{\sigma}) = \prod_{i=1}^{n_s} p(\boldsymbol{y}_i^{\mathrm{miss}}|\boldsymbol{y}_i^{\mathrm{obs}},\mathbf{v}_i,\sigma_i) \sim \mathcal{N}(\mu_y,\sigma_y^2)$ where $\mu_y = \mathbf{v}_i\sigma_i\theta$ and $\sigma_y^2 = \mathbf{v}_i\sigma_i\zeta^2$.

(4) $p(\boldsymbol{\delta}^*|\mathbf{w},\lambda^2) \sim \mathcal{N}(\mu_{\delta^*}, V_{\delta^*})$ where $V_{\delta^*} = (H_{\mathcal{A}}^{*\prime}H_{\mathcal{A}}^{*} + \Lambda^{-1})^{-1}$, $\mu_{\delta^*} = V_{\delta^*}H_{\mathcal{A}}^{*}\mathbf{w}$.

(5) $p(\lambda^2|\boldsymbol{\delta}) \sim \mathcal{IG}(a_{\lambda^2}, b_{\lambda^2})$ where $b_{\lambda^2} = \frac{\boldsymbol{\delta}'\boldsymbol{\delta}}{2} + b$, $a_{\lambda^2} = a + \frac{p}{2} - 1$.

(6) $p(\boldsymbol{\phi}^*|\boldsymbol{\beta}_{\mathcal{A}_1,1},\tau^2) \sim \mathcal{N}(\mu_{\phi^*}, V_{\phi^*})$ where $V_{\phi^*} = (H_{\mathcal{A}_1}^{*\prime}H_{\mathcal{A}_1}^{*} + \Psi^{-1})^{-1}$, $\mu_{\phi^*} = V_{\phi^*}H_{\mathcal{A}_1}^{*}\boldsymbol{\beta}_{\mathcal{A}_1,1}$.

(7) $p(\tau^2|\boldsymbol{\phi}) \sim \mathcal{IG}(a_{\tau^2}, b_{\tau^2})$ where $b_{\tau^2} = \frac{\boldsymbol{\phi}'\boldsymbol{\phi}}{2} + b$, $a_{\tau^2} = a + \frac{p}{2} - 1$.

(8) $p(\boldsymbol{\gamma},\boldsymbol{\beta}_1,\boldsymbol{\beta}_0|\tilde{\boldsymbol{Y}},\boldsymbol{\delta}^*,\mathbf{V},\boldsymbol{\sigma},X) = p(\boldsymbol{\gamma}|\tilde{\boldsymbol{Y}},\boldsymbol{\delta}^*,\mathbf{V},\boldsymbol{\sigma},X)p(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0|\tilde{\boldsymbol{Y}},\boldsymbol{\mu},\mathbf{V},\boldsymbol{\sigma},\boldsymbol{\gamma})$.

$$p(\boldsymbol{\gamma}|\tilde{\boldsymbol{Y}},\boldsymbol{\delta}^*,\mathbf{V},\boldsymbol{\sigma},X) \propto p(\tilde{\boldsymbol{Y}}|\boldsymbol{\gamma},\boldsymbol{Y}^{\mathrm{miss}},\mathbf{V},\boldsymbol{\sigma},X)p(\boldsymbol{\gamma}|\boldsymbol{\delta}^*),$$

$$p(\tilde{\boldsymbol{Y}}|\boldsymbol{\gamma},\mathbf{V},\boldsymbol{\sigma},X) = \int p(\tilde{\boldsymbol{Y}}|\boldsymbol{\gamma},\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)p(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0|\boldsymbol{\gamma})d\boldsymbol{\beta}_1 d\boldsymbol{\beta}_0,$$

$$p(\boldsymbol{\beta}_1,\boldsymbol{\beta}_0|\boldsymbol{\gamma}) \sim N(\boldsymbol{\mu}_{\boldsymbol{B}},\Sigma_{\boldsymbol{B}}) \text{ for } i \in \mathcal{A}_1,$$

where $\boldsymbol{\mu}_{\boldsymbol{B}} = (0,\boldsymbol{\mu}_{\mathcal{A}_1})'$ and $\Sigma_{\boldsymbol{B}} = \begin{bmatrix} \Sigma_{\boldsymbol{\beta}_0} & 0 \\ 0 & \Sigma_{\boldsymbol{\beta}_{\mathcal{A}_1,1}} \end{bmatrix}$, Note that $\Sigma_{\boldsymbol{\beta}_0} = T\frac{\boldsymbol{\sigma}}{n_s}I_{n_s}$ and $\Sigma_{\boldsymbol{\beta}_{\mathcal{A}_1,1}} = T\sigma(\mathbf{x}'\mathbf{x})^{-1}$.

$$p(\tilde{\boldsymbol{Y}}|\boldsymbol{\gamma},\boldsymbol{\beta}_1,\boldsymbol{\beta}_0)=\prod_{i=1}^{n_s}p(\tilde{\boldsymbol{y}}_i|\beta_{1,i},\beta_{0,i}),i\in\mathcal{A}_1,$$

So that $p(\tilde{\boldsymbol{y}}_i|\beta_{1,i},\beta_{0,i})$ is normal in $\beta_{1,i},\beta_{0,i}$ with $N(\mu_{\beta_i},V_{\beta_i})$. This is clearly stated in 10.

(9) $p(\gamma_i=1|\tilde{\boldsymbol{y}}_i,\boldsymbol{\delta}^*,\mathbf{V},\boldsymbol{\sigma},X)=\frac{p(\tilde{\boldsymbol{y}}_i|\gamma_i=1,\mathbf{v}_i,\boldsymbol{\sigma}_i,X)p(\gamma_i=1|\boldsymbol{\delta}^*)}{\sum_{l=0}^{1}p(\tilde{\boldsymbol{y}}_i|\gamma_i=l,\mathbf{v}_i,\boldsymbol{\sigma}_i,X)p(\gamma_i=l|\boldsymbol{\delta}^*)}$.

(10) $p(\boldsymbol{\beta}_{\mathcal{A}_1,1},\boldsymbol{\beta}_0|\boldsymbol{\mu}_{\mathcal{A}_1},\tilde{\boldsymbol{Y}},\mathbf{V}^{[j+1]},\boldsymbol{\sigma},\boldsymbol{\gamma})=\prod_{i\in\mathcal{A}_1}p(\beta_{1,i},\beta_{0,i}|\mu_i,\mathbf{v}_i,\sigma_i^{[j+1]},\boldsymbol{\gamma})\sim\mathcal{N}(\mu_{\beta_i},V_{\beta_i})$,
where $V_{\beta_i}=(X'\mathbf{V}_\zeta^{-1}X+\{T\sigma_i(X'X)^{-1}\}^{-1})^{-1}$ and $\mu_{\beta_i}=V_{\beta_i}(X'\mathbf{V}_\zeta^{-1}\tilde{\boldsymbol{y}}_i+\{T\sigma_i(X'X)^{-1}\}^{-1}\boldsymbol{\mu}_{\boldsymbol{B}_i})$
where $\boldsymbol{\mu}_{\boldsymbol{B}_i}=(0,\mu_i)'$ and for $i\notin\mathcal{A}_1$, $p(\boldsymbol{\beta}_{0,\mathcal{A}_0}|\mathbf{V}^{[j+1]},\tilde{\boldsymbol{Y}},\boldsymbol{\sigma})=\prod_{i\notin\mathcal{A}_1}p(\beta_{0,i}|\mathbf{v}_i,\sigma_i)\sim$
$\mathcal{N}(\mu_{\beta_{0,i}},V_{\beta_{0,i}})$, where $V_{\beta_{0,i}}=(\mathbf{V}_\zeta^{-1}+(T\sigma_i)^{-1})^{-1}$ and $\mu_{\beta_{0,i}}=V_{\beta_{0,i}}\,\mathbf{V}_\zeta^{-1}\tilde{\boldsymbol{y}}_i$.

Furthermore, the marginal likelihood of $\tilde{\boldsymbol{y}}_i$ can be obtained by integrating out $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_0$. The resulting marginal likelihood for the case of $\gamma_i=1$ is given by,

$$p(\tilde{\boldsymbol{y}}_i|\gamma_i=1,\mathbf{v}_i,\boldsymbol{\sigma}_i,X)=\frac{|X'X|^{\frac{1}{2}}}{|2\pi|^{\frac{T}{2}}|\mathbf{V}_\zeta|^{\frac{1}{2}}}\frac{|\mathbf{V}_\beta|^{\frac{1}{2}}}{(T\sigma_i)^{\frac{1}{2}}}\exp\left\{\frac{1}{2}(\tilde{\boldsymbol{y}}_i'[\mathbf{V}_\zeta^{-1}-\mathbf{V}_\zeta^{-1}(XX')\mathbf{V}_\zeta^{-1}]\tilde{\boldsymbol{y}}_i-\tilde{\boldsymbol{y}}_i'\mathbf{V}_\zeta^{-1}\tilde{\boldsymbol{y}}_i)\right\},\tag{B1.1}$$

where,

$\mathbf{V}_\beta=(X'\mathbf{V}_\zeta^{-1}X+\{T\sigma_i(X'X)^{-1}\}^{-1})^{-1}$, $\mathbf{V}_\zeta=\zeta^2\sigma_i\,\mathrm{diag}(\mathbf{v}_i)$ and $\tilde{\boldsymbol{y}}_i=\boldsymbol{y}_i-\theta\mathbf{v}_i$.

B4 Prediction at Unobserved locations.

Let $\boldsymbol{g}_\mu(S^{\text{obs}})=H_{\mathcal{A}_1}\boldsymbol{\phi}^*$ and each iteration j we will have $\boldsymbol{g}_\mu^{[j]}(S^{\text{obs}})$ which is $H_{\mathcal{A}_1}\boldsymbol{\phi}^{*[j]}$.

To predict values at unobserved locations (grid of latitudes and longitudes) S^*, Using the properties of Multivariate Normal distribution, $E(\boldsymbol{g}_\mu(S^*)|\boldsymbol{g}_\mu^{[j]}(S^{\text{obs}}))$ can be written as follows,

$$E(\boldsymbol{g}_\mu(S^*)|\boldsymbol{g}_\mu^{[j]}(S^{\text{obs}})) = E(\boldsymbol{g}_\mu(S^*))+\text{cov}(\boldsymbol{g}_\mu(S^*),\boldsymbol{g}_\mu^{[j]}(S^{\text{obs}}))\text{var}(\boldsymbol{g}_\mu^{[j]}(S^{\text{obs}}))^{-1}(\boldsymbol{g}_\mu^{[j]}(S^{\text{obs}})-E(\boldsymbol{g}_\mu^{[j]}(S^{\text{obs}}))), \tag{B1.2}$$

where $E(\boldsymbol{g}_\mu(S^*)) = S^*\boldsymbol{\eta}^{[j]}$, $\text{var}(\boldsymbol{g}_\mu^{[j]}(S^{\text{obs}})) = \tau^{2[j]}\Omega_{\mathcal{A}_1}$ and $\text{cov}(\boldsymbol{g}_\mu(S^*),\boldsymbol{g}_\mu^{[j]}(S^{\text{obs}})) = \tau^{2[j]}\Omega_{S^*,S^{\text{obs}}}$, is given by the reproducing kernel of a thin-plate spline Wahba (1990).

Similarly, let $\boldsymbol{g}_\pi(S^{\text{obs}}) = H_{\mathcal{A}}\boldsymbol{\delta}^*$ and in each posterior iteration j we will have $\boldsymbol{g}_\pi^{[j]}(S^{\text{obs}})$ which is $H_{\mathcal{A}}\boldsymbol{\delta}^{*[j]}$.

$E(\boldsymbol{g}_\pi(S^*)|\boldsymbol{g}_\pi^{[j]}(S^{\text{obs}}))$ can be written as follows,

$$E(\boldsymbol{g}_\pi(S^*)|\boldsymbol{g}_\pi^{[j]}(S^{\text{obs}})) = E(\boldsymbol{g}_\pi(S^*))+\text{cov}(\boldsymbol{g}_\pi(S^*),\boldsymbol{g}_\pi^{[j]}(S^{\text{obs}}))\text{var}(\boldsymbol{g}_\pi^{[j]}(S^{\text{obs}}))^{-1}(\boldsymbol{g}_\pi^{[j]}(S^{\text{obs}})-E(\boldsymbol{g}_\pi^{[j]}(S^{\text{obs}}))), \tag{B1.3}$$

where $E(\boldsymbol{g}_\pi(S^*)) = S^*\boldsymbol{\alpha}^{[j]}$, $\text{var}(\boldsymbol{g}_\pi^{[j]}(S^{\text{obs}})) = \lambda^{2[j]}\Omega_{\mathcal{A}}$ and $\text{cov}(\boldsymbol{g}_\pi(S^*),\boldsymbol{g}_\pi^{[j]}(S^{\text{obs}})) = \lambda^{2[j]}\Omega_{S^*,S^{\text{obs}}}$.

B5 Alternative approach using Polya-gamma data augmentation

In alternative approach, step 8 in algorithm 2 will change to,

(1) $\mathbf{w}|\boldsymbol{\gamma},\boldsymbol{\delta}^* \sim \text{PG}\{1, H_{\mathcal{A}}^*\boldsymbol{\delta}^*\}$ where PG denotes Polya-gamma distribution (see Polson et al. (2012)) .
(2) Generate $\boldsymbol{\delta}^{*[j+1]}$ and $p(\boldsymbol{\delta}^*|\boldsymbol{\gamma},\mathbf{w},\lambda^{2[j]})$.
(3) Generate $\lambda^{2[j+1]}$ from $p(\boldsymbol{\lambda}^2|\boldsymbol{\delta}^{[j+1]})$.

And , $p(\gamma_i = 1|\boldsymbol{\delta}^*) = \frac{\exp(H^*\boldsymbol{\delta}^*)}{1+\exp(H^*\boldsymbol{\delta}^*)}$.

Note that, $p(\boldsymbol{\delta}^*|\mathbf{w}, \boldsymbol{\lambda}) \sim N(\mu_{\delta^*}, \mathbf{V}_{\delta^*})$ where $\mathbf{V}_{\delta^*} = ({H^*_{\mathcal{A}}}' W H^*_{\mathcal{A}} + \Lambda^{-1})^{-1}$, $\mu_{\delta^*} = \mathbf{V}_{\delta^*} {H^*_{\mathcal{A}}}' \boldsymbol{\kappa}$, where $\boldsymbol{\kappa} = (\gamma_1 - \frac{1}{2}.., \gamma_{n_s} - \frac{1}{2})'$, and W is the diagonal matrix of $\mathbf{w}$'s. and $p(\boldsymbol{\lambda}^2|\boldsymbol{\delta}) \sim$ IG$(a_{\lambda^2}, b_{\lambda^2})$ where $b_{\lambda^2} = b + \frac{\boldsymbol{\delta}'\boldsymbol{\delta}}{2}$, $a_{\lambda^2} = a + \frac{p}{2} - 1$.

All other derivations for other parameters remain the same as in B3.

B6 Anlaysis on selected stations

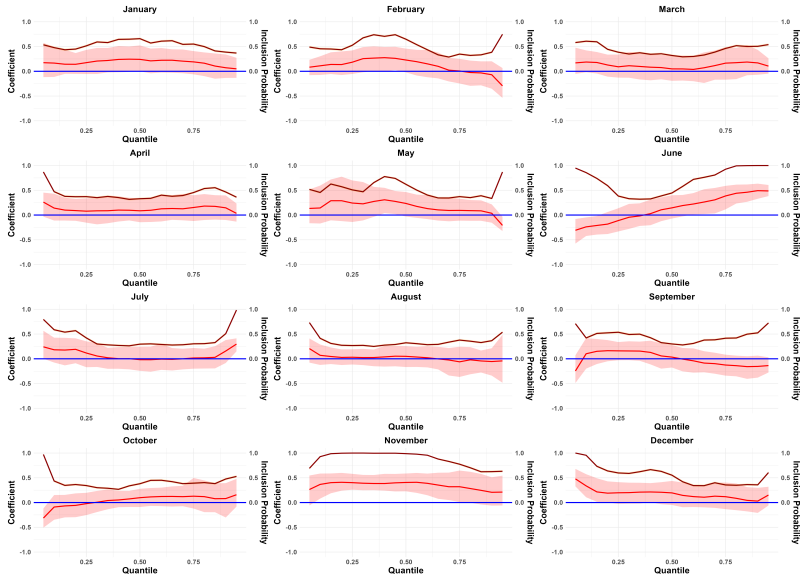


FIGURE B1.5: The estimated posterior mean of the inclusion probabilities (dark red) as well as the estimated posterior mean (red) and the 95% credible intervals (pink) for the magnitude of the impact, for Sydney Botanic Gardens [66006] (blue line indicates zero).

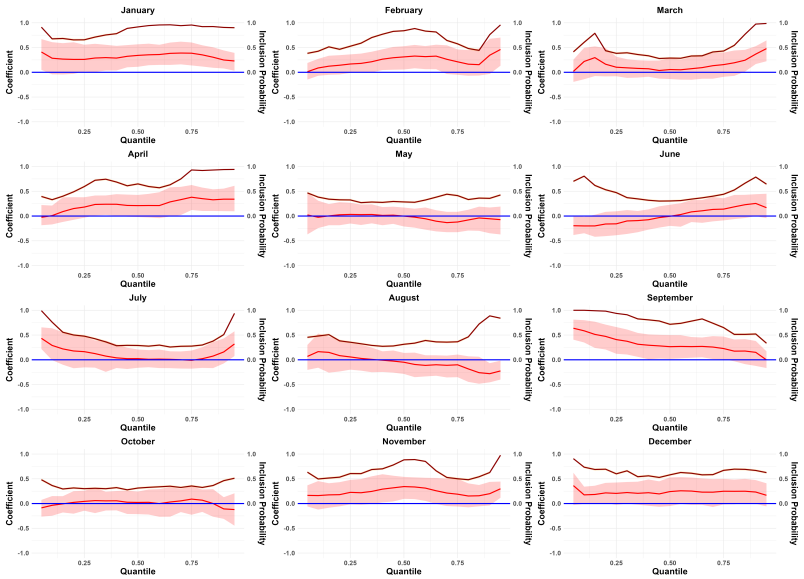


FIGURE B1.6: The estimated posterior mean of the inclusion probabilities (dark red) as well as the estimated posterior mean (red) and the 95% credible intervals (pink) for the magnitude of the impact, for Yamba Pilot Station [58012] (blue line indicates zero).

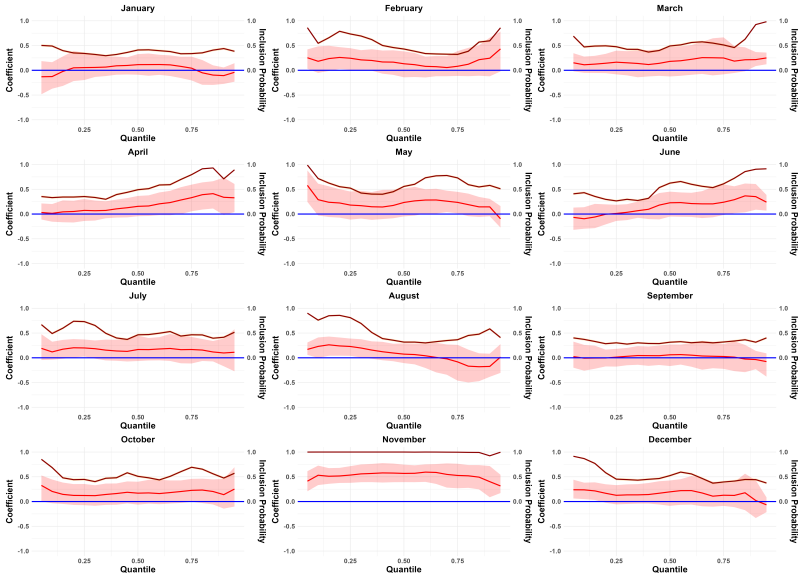


FIGURE B1.7: The estimated posterior mean of the inclusion probabilities (dark red) as well as the estimated posterior mean (red) and the 95% credible intervals (pink) for the magnitude of the impact, for Moruya Heads Pilot Station [69018] (blue line indicates zero).

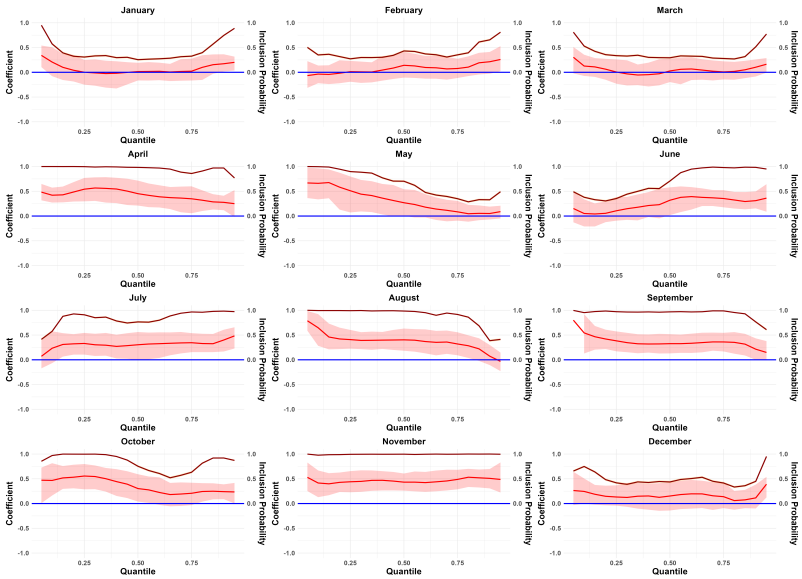


FIGURE B1.8: The estimated posterior mean of the inclusion probabilities (dark red) as well as the estimated posterior mean (red) and the 95% credible intervals (pink) for the magnitude of the impact, for Dunedoo Post Office [64009] (blue line indicates zero).

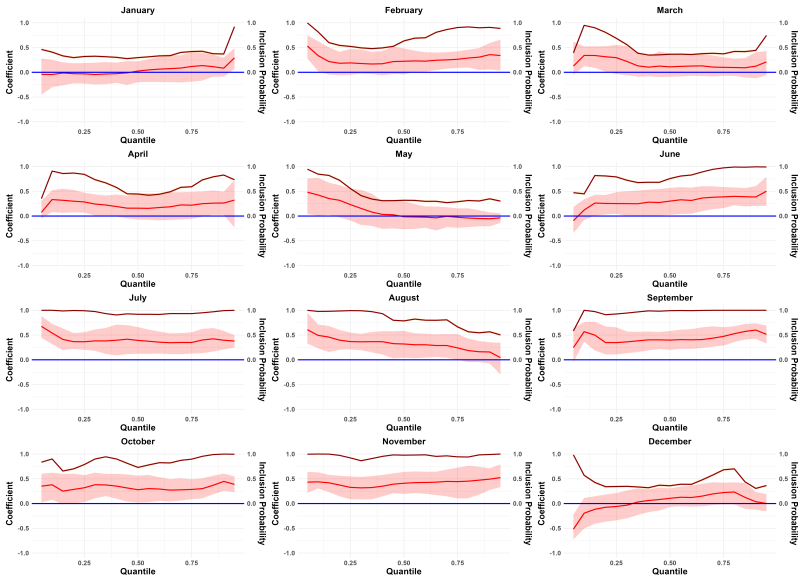


FIGURE B1.9: The estimated posterior mean of the inclusion probabilities (dark red) as well as the estimated posterior mean (red) and the 95% credible intervals (pink) for the magnitude of the impact, for Collarenebri (Albert St) [48031] (blue line indicates zero).

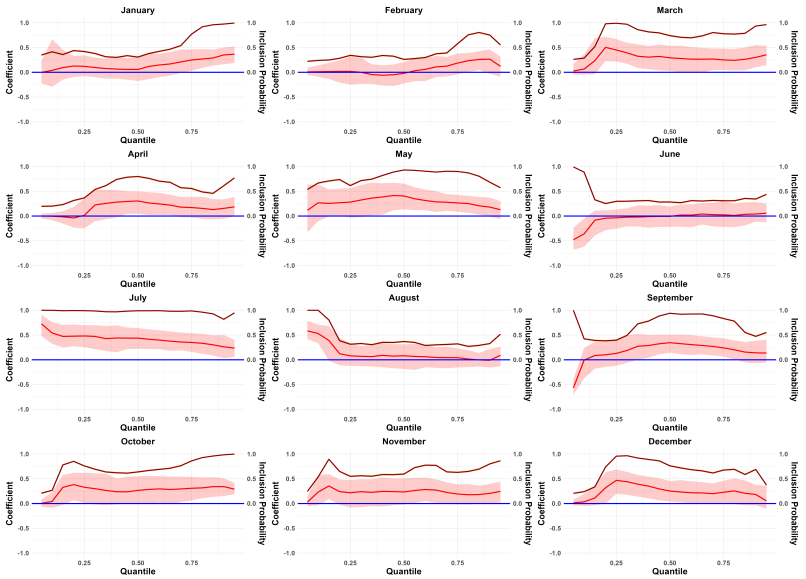


FIGURE B1.10: The estimated posterior mean of the inclusion probabilities (dark red) as well as the estimated posterior mean (red) and the 95% credible intervals (pink) for the magnitude of the impact, for Pooncarie (Karpa Kora Station) [47013] (blue line indicates zero).

Appendix C

C1 Posterior Derivations for Algorithm

(1) All other conditional posterior derivations remain the same as in Chapter 3 except for the following.

(2) $p(\boldsymbol{\delta}_c^*|\mathbf{w}_c, \lambda_c^2) \sim \mathcal{N}(\mu_{\delta_c^*}, V_{\delta_c^*})$ where $\mathbf{V}_{\delta_c^*} = (H_c^{*'} W_c H_c^* + \Lambda_c^{-1})^{-1}$, $\mu_{\delta_c^*} = \mathbf{V}_{\delta_c^*} H_c^* (\boldsymbol{\kappa}_c + W_c ln \sum_{d \neq c} \exp(H_d^* \boldsymbol{\delta}_d^*))$, where $\boldsymbol{\kappa}_c = \frac{I(\boldsymbol{\gamma}_i \in c) - \frac{1}{2}}{\mathbf{w}_c}$, and W_c is the diagonal matrix of $\mathbf{w}_c$'s.

(3) $p(\lambda_c^2|\boldsymbol{\delta}_c^{[j]}) \sim \mathcal{IG}(a_{\lambda_c^2}, b_{\lambda_c^2})$ where $b_{\lambda_c^2} = \frac{\boldsymbol{\delta}_c' \boldsymbol{\delta}_c}{2} + b$, $a_{\lambda_c^2} = a + \frac{p}{2} - 1$.

(4) $p(\boldsymbol{\Gamma}|\boldsymbol{\delta}_c^*, \tilde{\boldsymbol{Y}}, \mathbf{V}, \boldsymbol{\sigma}, X) = \prod_{i=1}^{n_s} \frac{p(\tilde{\boldsymbol{y}}_i|\boldsymbol{\gamma}_i, \mathbf{v}_i, \boldsymbol{\sigma}_i, X) p(\boldsymbol{\gamma}_i|\boldsymbol{\delta}_c^*)}{\sum_{c=1}^{C} p(\tilde{\boldsymbol{y}}_i|\boldsymbol{\gamma}_i, \mathbf{v}_i, \boldsymbol{\sigma}_i, X) p(\boldsymbol{\gamma}_i|\boldsymbol{\delta}_c^*)}$.

(5) $p(\boldsymbol{\phi}_{k,\mathcal{A}_{k,1}}^*|\boldsymbol{\beta}_{\mathcal{A}_{k,1}}, \tau_{k,\mathcal{A}_{k,1}}^2) \sim \mathcal{N}(\mu_{\phi_{k,\mathcal{A}_{k,1}}^*}, V_{\phi_{k,\mathcal{A}_{k,1}}^*})$ where $V_{\phi_{k,\mathcal{A}_{k,1}}^*} = (H_{k,\mathcal{A}_{k,1}}^{*'} H_{k,\mathcal{A}_{k,1}}^* + \Psi_{k,\mathcal{A}_{k,1}}^{-1})^{-1}$, $\mu_{\phi_{k,\mathcal{A}_{k,1}}^*} = V_{\phi_{k,\mathcal{A}_{k,1}}^*} H_{k,\mathcal{A}_{k,1}}^* \boldsymbol{\beta}_{\mathcal{A}_{k,1}}$.

(6) $p(\tau^2_{k,\mathcal{A}_{k,1}}|\boldsymbol{\phi}_{k,\mathcal{A}_{k,1}}) \sim \mathcal{IG}(a_{\tau^2_{k,\mathcal{A}_{k,1}}}, b_{\tau^2_{k,\mathcal{A}_{k,1}}})$ where $b_{\tau^2_{k,\mathcal{A}_{k,1}}} = \frac{\boldsymbol{\phi}'_{k,\mathcal{A}_{k,1}}\boldsymbol{\phi}_{k,\mathcal{A}_{k,1}}}{2} + b$, $a_{\tau^2_{k,\mathcal{A}_{k,1}}} = a + \frac{p}{2} - 1$.

Furthermore, the marginal likelihood of $\tilde{\boldsymbol{y}}_i$ can be obtained by integrating out $\boldsymbol{\beta}_{\mathcal{A}_{.,1}}$ and $\boldsymbol{\beta}_0$. The resulting marginal likelihood for the case of $\boldsymbol{\gamma}_i \in c$ is given by,

$$p(\tilde{\boldsymbol{y}}_i|\boldsymbol{\gamma}_i \in c, \mathbf{v}_i, \sigma_i, X_{\mathcal{A}}) = \frac{|X'_{\mathcal{A}}X_{\mathcal{A}}|^{\frac{1}{2}}}{|2\pi|^{\frac{T}{2}}|\mathbf{V}_\zeta|^{\frac{1}{2}}} \frac{|\mathbf{V}_\beta|^{\frac{1}{2}}}{(T\sigma_i)^{\frac{1}{2}}} \times \exp\left\{\frac{1}{2}\Big(\tilde{\boldsymbol{y}}'_i[\mathbf{V}_\zeta^{-1} - \mathbf{V}_\zeta^{-1}(X_{\mathcal{A}}X'_{\mathcal{A}})\mathbf{V}_\zeta^{-1}]\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{y}}'_i\mathbf{V}_\zeta^{-1}\tilde{\boldsymbol{y}}_i\Big)\right\},$$

$$\text{for} \quad \mathcal{A}_{.,1} \in c. \tag{C1.1}$$

where,

$\mathbf{V}_\beta = (X'_{\mathcal{A}}\mathbf{V}_\zeta^{-1}X_{\mathcal{A}} + \{T\sigma_i(X'_{\mathcal{A}}X_{\mathcal{A}})^{-1}\}^{-1})^{-1}$, $\mathbf{V}_\zeta = \zeta^2\sigma_i$ diag$(\mathbf{v}_i)$, $\tilde{\boldsymbol{y}}_i = \boldsymbol{y}_i - \theta\mathbf{v}_i$ and $X_{\mathcal{A}} = [1, X_{\mathcal{A}_{.,1}}]$.

C2 Performance Measures

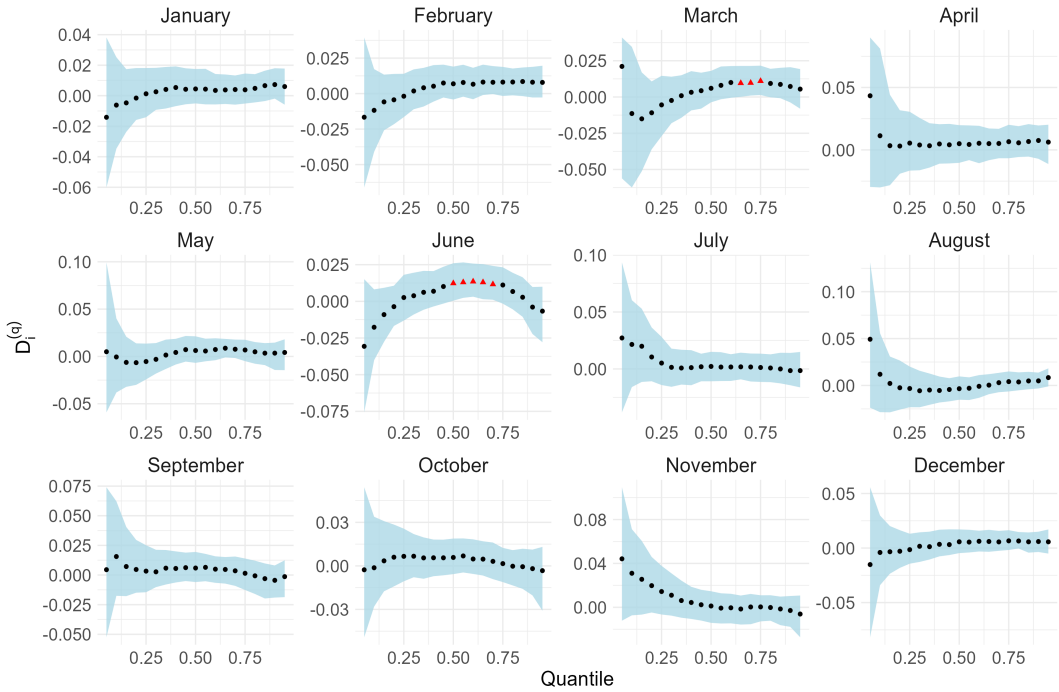


FIGURE C1.1: Median values (black points) and interquartile range (shaded band: light blue) of $D_i^{(q)}$ across all sites n_s for each month and quantile. A black circle indicates that the corresponding quantile is nominal (i.e., the IQR of $D_i^{(q)}$, $i = 1, ..., n_s$, contains zero), while a red triangle indicates that it is not, for the model with two covariates (both SOI and DMI).

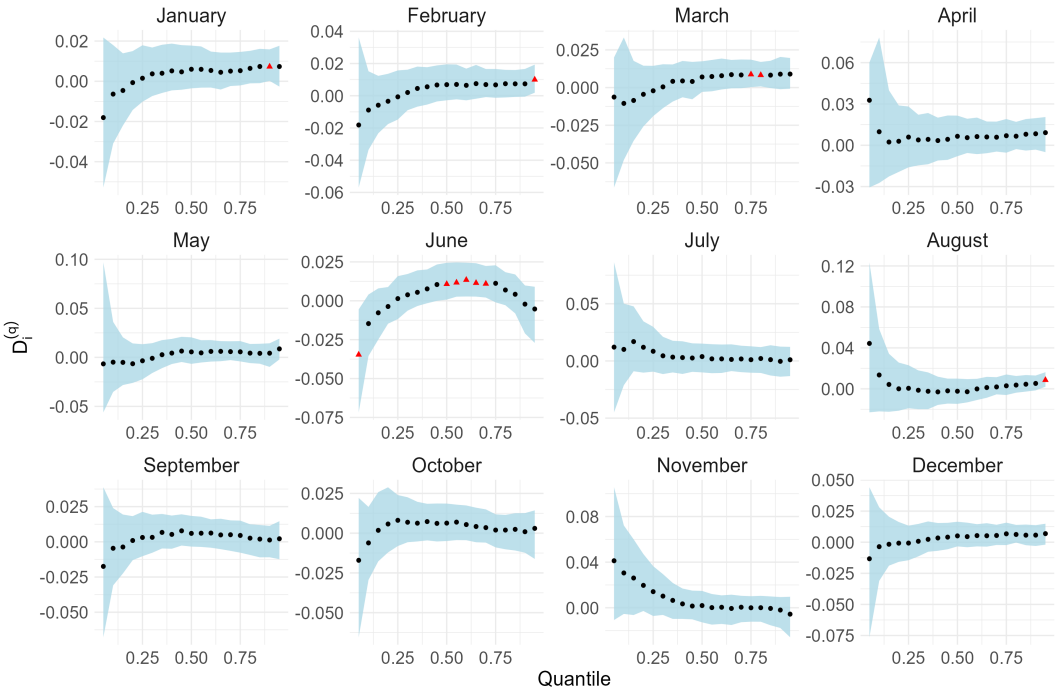


FIGURE C1.2: Median values (black points) and interquartile range (shaded band: light blue) of $D_i^{(q)}$ across all sites n_s for each month and quantile. A black circle indicates that the corresponding quantile is nominal (i.e., the IQR of $D_i^{(q)}$, $i = 1, ..., n_s$, contains zero), while a red triangle indicates that it is not, for the model with one covariate (SOI only).