Bayesian inference of coral bleaching dynamics

UPPSALA
UNIVERSITET

Uppsala University
Department of Information Technology

Master's Programme in Data Science

Chen Gu

17 June 2025

Abstract
Coral bleaching is a major threat to the survival of reefs, which are among our most important ecological systems. One way to estimate the dynamics of coral reefs is through mathematical modelling using Ordinary Differential Equation (ODE) models. In this project, we design a multi-layer Bayesian hierarchical model to capture the dynamics of coral reefs. We compare three pooling methods: complete pooling, no pooling, and partial pooling. However, due to the complex and nested structure of the dataset and model, we face significant challenges in sampling from the high-dimensional Bayesian posterior geometry. To address these issues, we employ the probabilistic programming framework Stan, which utilises Hamiltonian Monte Carlo (HMC) sampling. We adopt non-centered partial pooling as a reparameterisation strategy to further mitigate sampling difficulties. Additionally, we develop a comprehensive diagnostic pipeline to assess the robustness of this approach from multiple perspectives, enabling us to effectively overcome implementation challenges. This Bayesian reparameterisation approach outperforms traditional frequentist methods, such as non-linear least squares, in obtaining precise parameter estimates for complex ecological dynamics. It also provides more comprehensive uncertainty quantification. Our work establishes a data-validated framework for improved understanding of coral bleaching dynamics. Our modelling framework may inform more effective conservation strategies, and the methods developed in this work has potential applications for other complex dynamical systems beyond reef environments.

Acknowledgements
I want to express my gratitude to several important people who helped me with this thesis.

First, I want to sincerely thank my supervisor, Sara. At the beginning of this journey, I was extremely nervous, as it was my first time working on a full thesis project. However, Sara's clear and well-organized documentation, logical guidance, and deep understanding of the subject gave me complete confidence in every step she planned. She managed the pace of the thesis incredibly well, keeping the big picture in view while providing detailed advice and appropriate guidance at each stage. At the same time, she always took the time to listen carefully to my ideas and gave me the freedom to explore areas I was passionate about. I truly believe I received the best support anyone could hope for from a supervisor. Sara treated every student with sincerity and gave sharp, honest feedback that continuously pushed us to improve. I could truly feel the genuineness of her hope for our growth in every meeting and piece of feedback. She had a remarkable ability to see the unique potential in each person and offered them the right opportunities accordingly, tailoring a path for growth based on each individual's strengths and character, and I am deeply grateful for that. Her support allowed me to finish my thesis on time while still having the space to explore each component in depth. Without her encouragement, I would not have come this far or had the confidence to try so many things in my thesis. Looking back, I can see how much I've grown—not just in my understanding of Bayesian ideas, but as a student overall. What started as a simple interest has become a genuine passion that I hope to continue pursuing in the future.

I also want to thank my co-supervisor, Jodie. She patiently guided me through the details of coding and writing, showing me how to check and fix code step by step. At first, I really struggled with writing the thesis and using R, but I still remember one of our early meetings—Jodie was so patient and told me not to worry. She went through every line of my code with me, and that moment made me realise that I didn't need to rush through everything. From her, I learned that sometimes the slowest way is actually the most effective, and that solid learning comes from taking time to understand each step. She is incredibly good at conducting research in a very detailed way, and I drew a lot of inspiration from her working style.

I also want to thank my reviewer Per. He was very patient and asked important questions that helped me notice things I had missed in my thesis. His way of asking questions and carefully analysing the research changed how I think. He taught me that it's not just about understanding the main ideas or explaining them to others, but also about expressing everything as simply and clearly as possible. Once I tried that, I realised it's actually the best way to fully understand something. These are lessons I will take with me into my future studies.

I want to give very special thanks to my good friend, Nora. We spent almost every day together in the past 5 months in the thesis room on the fifth floor. We cheered each other on when things went well and encouraged each other when things were hard. I was always inspired by her never-give-up attitude and her commitment to doing her best work.

Special thanks go to Khang. He often says that my course notes helped him a lot, but the truth is that he helped me much more. He is such a talented and hardworking student, always warm-hearted, thoughtful, and precise in everything he does. He reminded me of important details in my studies, gave me valuable insights when I faced difficult math problems, and supported me with his experience whenever I felt confused. I will never forget the day we found out we had the same supervisor—it was such a fun and meaningful moment.

Finally, I want to thank my parents. Without their constant love and support, I would not have been able to reach this point. Thank you for everything.

Contents

1. Introduction

1.1 Background

Healthy coral reefs are some of the world's most biologically diverse and economically valuable ecosystems. Despite occupying less than 1% of the ocean floor, these complex habitats support an estimated 25% of marine life [25, 26] and provide critical services to coastal communities [38]. However, these important ecosystems are threatened by multiple challenges, including local pressures from fishing, coastal development, and pollution, as well as the widespread impact of global ocean warming, which causes coral bleaching [21].

Coral bleaching, primarily driven by global warming, stands as one of the major threats to coral reefs [43]. This phenomenon involves the whitening of corals due to the loss of their symbiotic algae and/or the pigments these algae contain. It was first described affecting extensive reef areas across the Pacific by Glynn in the early 1980s [36]. While healthy corals do continuously release algae, they do so in very low numbers [14]. However, corals are highly sensitive to unusual environmental conditions, such as abnormal water temperatures (both warm and cool), low salinity, and pollution. Under such stressful conditions, particularly elevated temperatures, the algal symbiont can become toxic to the coral animal. Consequently, the coral host expels its symbiotic partner from its tissues as a survival strategy, especially during events like heatwaves.

As the algal symbiont leaves the coral, the coral's thin tissues become transparent, revealing its white skeleton underneath. The typically dark brown colony (or other original coloration) becomes pale as the coral bleaches [4]. Field and laboratory evidence has shown that frequent or severe coral bleaching events can reduce reproductive capacity, growth, disease resistance, and ultimately the survival of affected corals across large geographic scales [21].

Aerial surveys during the 2024 mass coral bleaching event documented extensive impacts across the Great Barrier Reef [16]. Widespread bleaching was visible on 1,001 analyzeable reefs that were surveyed between the southern Marine Park and the Torres Strait. Some level of bleaching was observed on 79% (792 reefs) of shallow reef habitats. Concerningly, severity was high in many areas. For instance, 49% of the surveyed reefs experienced high to extreme bleaching (affecting >30% of their coral cover). Furthermore, 32% of all surveyed reefs showed very high or extreme bleaching (affecting >60% of their coral cover); this figure was 39% for reefs located specifically within the

Great Barrier Reef Marine Park. However, it is critical to understand that a bleached reef is not dead. Long-term monitoring by the Australian Institute of Marine Science (AIMS) confirms that Great Barrier Reef coral cover has historically fluctuated through cycles of disturbance and recovery [4]. Similar to recovery after cyclones, live coral can return after bleaching if given sufficient time, potentially 10-15 years, without further major disturbances, allowing diverse communities to re-establish [16, 4].

Given the significant ecological value of coral reefs, the severity of recent bleaching events, and the complex nature of potential recovery, a deeper understanding of coral reef dynamics is crucial. This study develops and compares hierarchical Bayesian models incorporating three distinct pooling approaches, complete pooling, partial pooling, and no pooling, to quantify the key ecological parameters driving coral cover dynamics. By explicitly modelling the transition between healthy and bleached states at both reef and site levels, we aim to improve our ability to predict reef trajectories under different environmental conditions. Our comparative analysis of these pooling approaches provides insights into the spatial variability of these ecological processes and delivers more robust parameter estimates with quantified uncertainty. This methodological advance enhances our capacity to forecast coral reef responses to environmental stressors, thereby contributing valuable information for reef conservation planning and management interventions, while also illustrating the potential for applying these advanced Bayesian methods to other complex dynamical systems.

1.2 Literature Review

Previous research on coral bleaching has utilised a variety of modelling approaches, each with inherent strengths and weaknesses. Early studies often employed statistical models to identify correlations between environmental stressors and bleaching prevalence. For example, Woesik [71] applied logistic regression to associate bleaching with solar radiation and water flow, but omitted crucial coral-symbiont feedback mechanisms, treating bleaching as a passive response rather than a dynamic process. Similarly, Donner [21] linked sea surface temperature anomalies to bleaching severity using linear regression, yet acknowledged the model's inability to explain threshold behaviours or physiological recovery. While these statistical approaches successfully identified thermal stress as a primary driver of bleaching, they tended to oversimplify the underlying dynamics.

Subsequent research has explored integrating machine learning methods to improve upon previous approaches to assess coral bleaching. Boonnam [11] investigated using the supervised machine learning approach methods such as Naive Bayes, support vector machines, and decision trees, to predict levels of coral damage. Unsupervised machine learning have also been used, in the

form of k-means clustering, in order to find similar characteristics of coral reefs. However, it still exhibited limited capacity to forecast future levels of coral bleaching. Lachs [48] fine-tuned the Degree Heating Weeks heat stress algorithm, focusing on optimising the HotSpot threshold and accumulation window, i.e., the temperature above which bleaching is triggered, and adjusting the accumulation window, this approach, while valuable for improving the accuracy of predictions based on satellite-derived sea surface temperature data, inherently depends on using coral bleaching observations to optimise the thermal stress that best predicts mass bleaching events. These data-driven approaches offer improvements in characterising coral bleaching, yet they are limited by static, non-mechanistic representations of the underlying processes.

Mechanistic models represent an advance over these previous approaches. An example is the Ordinary Differential Equations (ODEs) system presented by Brown et al. (2024) [13]. This system is illustrated conceptually in Figure 1.1. Such systems treat coral state transitions as a continuous process. This continuous treatment offers a rigorous framework. This framework aids in understanding change over time and the complex interactions between variables [82]. It allows for a more detailed depiction of the stress and recovery dynamics experienced by coral colonies. The Brown et al. model defines transitions between key coral conditions. It includes an assimilation state, termed C, whose intrinsic dynamics like growth or maintenance occur at a rate α. This healthy state C can transition to a bleaching state, termed B, through coral bleaching at rate β. Conversely, the bleached state B can recover to the healthy state C via coral assimilation at rate γ. Furthermore, the bleached state B faces a risk of bleaching induced mortality at rate μ. Understanding these distinct processes and their rates provides deeper insights into the resilience of coral communities.

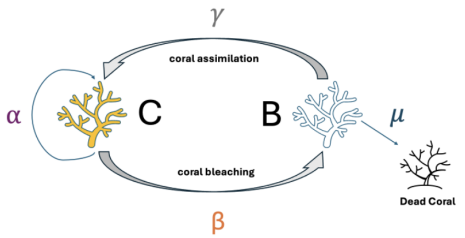


Figure 1.1. Schematic diagram of coral dynamics, representing the interactions between assimilating (healthy) coral cover (C) and bleached coral cover (B). The yellow coral represents C, the state of healthy, growing coral cover. The light blue coral represents B, the state of bleached coral cover, and the black skeletal structure represents dead coral. The model incorporates four key rate parameters: α (purple) represents the intrinsic growth rate of C; β (orange) represents the rate of bleaching, the transition from C to B; γ (grey) represents the recovery rate of bleached corals back to the healthy state (B to C); and μ (dark blue) represents the mortality rate of bleached corals leading to dead coral.

1.3 Project Objectives and Outline

Building upon this promising ODE framework and recognising the advantages of Bayesian methods, our project employs a hierarchical Bayesian approach to estimate model parameters using extensive monitoring data from the Great Barrier Reef. A key part of our project also involves designing comprehensive Bayesian experiment steps to test model stability. This statistical framework offers several key advantages for rigorously analysing complex ecological systems. Firstly, it provides full posterior distributions for model parameters, thereby explicitly quantifying uncertainty in parameter estimates and subsequent predictions. Secondly, the Bayesian framework allows for the incorporation of prior knowledge through informative prior distributions, helping to regularise the model and enhance estimate stability, particularly when dealing with noisy ecological data. Thirdly, and crucially for heterogeneous data like the AIMS dataset [4], it naturally supports hierarchical modelling at multiple levels (e.g., population, reef, and site), enabling partial pooling of information across units. Finally, the Bayesian workflow includes powerful tools for comprehensive model evaluation and comparison.

The computational demands of Bayesian inference, especially for complex models, often require sophisticated sampling techniques [49]. Markov Chain Monte Carlo (MCMC) methods are a class of algorithms designed to sample from the posterior distribution, which is typically analytically intractable [46]. A foundational MCMC algorithm is the Metropolis Hastings (MH) algorithm [56], which generates a sequence of samples by proposing new parameter states and accepting or rejecting them based on a probability ratio related to the posterior. However, MH can suffer from inefficiencies, particularly in high dimensional parameter spaces, due to its random walk behaviour [61]. Hamiltonian Monte Carlo (HMC) is a more advanced MCMC algorithm that builds upon MH by introducing auxiliary "momentum" variables and leveraging Hamiltonian dynamics. This approach allows HMC to explore the parameter space more efficiently, suppressing random walk behaviour and leading to faster convergence [42]. In this project, we utilise Stan, a probabilistic programming language, and its R interface, RStan [70]. Stan implements the No U Turn Sampler (NUTS), a highly optimised variant of HMC, providing a robust and user friendly platform for Bayesian inference [18].

We will also investigate different pooling strategies (no pooling, partial pooling, and complete pooling) within the Bayesian framework, which is particularly good for analysing data with the inherent structure of the AIMS dataset [4]. The concept of pooling is central to hierarchical modelling and defines how information is shared between sites and reefs. No pooling treats each reef as completely independent, estimating parameters separately without any sharing of information. Complete pooling, at the other extreme, assumes that all reefs are identical, estimating a single set of parameters for the entire Great Barrier Reef. Partial pooling represents a compromise, where reef

specific parameters are assumed to be drawn from a common population distribution. This method allows for individual reef variations while still leveraging information from the entire dataset, and it also accommodates variation at the site level. This approach typically leads to more robust and accurate parameter estimates.

In light of the motivation and existing knowledge, the gaps that this project intends to fill are: (1) designing diagnostic Bayesian experiments to provide clear, worked examples for more effective Stan implementation, improving accessibility for other researchers; (2) comparing Bayesian and frequentist approaches to evaluate their respective capabilities in quantifying uncertainty, particularly in complex ecological models; and (3) developing a deeper understanding of coral bleaching dynamics on the Great Barrier Reef through Bayesian hierarchical modelling, contributing to improved reef management through more accurate and uncertainty aware forecasts.

The structure of the remainder of this document is as follows. Chapter 1 introduces the research problem, providing context on coral bleaching dynamics and the need for Bayesian approaches. Chapter 2 details the data used for model calibration, including the AIMS [4] data collection methods, preprocessing steps, and inherent limitations. Chapter 3 describes the methodology related to the mathematical model, the Bayesian inference methods, the hierarchical model structure, the pooling method, the implementation in Stan, and the model calibration and evaluation procedures. Chapter 4 presents the findings from our analysis, comparing different pooling strategies and evaluating model performance. Finally, Chapter 5 summarizes the key contributions and implications of this work for coral reef management and Bayesian modelling applications in ecological contexts, and explores potential extensions and improvements to the current approach, outlining directions for future research.

2. Data Processing

2.1 Data Source

The data used in this project originates from the AIMS Long-Term Monitoring Programme (LTMP) [4], a comprehensive and long-running effort to monitor the condition of the Great Barrier Reef. The LTMP collects data on various aspects of reef health, with the primary variable of interest for this research being the percent cover of live hard coral. Coral cover is estimated through video transect surveys, where divers record footage along fixed lines on the reef, and subsequent analysis quantifies the proportion of live coral. The AIMS video transect surveys provide a time series of coral cover observations for each site, enabling the identification of bleaching events and, in many cases, subsequent recovery phases over the years.

2.2 Data Preparation and Selection

As we can see in Figure 2.1, we conducted data preparation and selection during several stages: initial data filtering, reef selection criteria and procedure, data formatting for Stan implementation, and preparation of an experimental dataset for sparsity analysis.

2.2.1 Initial Data Filtering

First, we filter the reef data to focus on observations from the 1995–2004 period. This timeframe is chosen to maximise potential overlap with the set of reefs analysed by Brown et al. [13] and to establish a basis for consistent sampling across any reefs subsequently selected. This ten-year timeframe also provides sufficient temporal coverage to analyse coral cover dynamics.

Initially, working with data confined to this established period and following the approach in the original paper done by Brown et al., we select 9 reefs for our analysis. However, our experience reveals that reef selection significantly impacts model predictions and parameter estimation. A limited selection of reefs may not fully capture the complexity of coral-bleaching dynamics. Some reefs might exhibit unusual patterns or respond differently to environmental factors, potentially biasing our parameter estimates. Expanding our dataset to include more reefs likely improves our ability to model the true dynamics of the system.

The limitations encountered with analysing a small selection of reefs independently underscore the need for a more integrated analytical strategy. To better capture the true system dynamics and account for variability effectively, our study adopts a Bayesian hierarchical modelling approach. This contrasts with the independent, frequentist parameter estimation for each reef in Brown et al. [13]. Such a hierarchical methodology is adept at sharing information across different reefs and sites, deriving more robust insights, particularly when applied to an expanded and diverse dataset. A full exposition of this framework, including the specific pooling strategies investigated, is provided in Chapter 3.

2.2.2 Reef Selection Criteria and Procedure

To leverage the strengths of this hierarchical Bayesian framework and ensure a comprehensive analysis, we expand our reef selection from 9 to 15 reefs across the Great Barrier Reef, standardising the observation period to 10 years (1995–2004). Each selected reef must contain exactly three observation sites with complete data for all 10 years, ensuring sufficient within-reef variability and within-sites variability while maintaining analytical consistency. This approach provides crucial site-level detail that allows our hierarchical models to distinguish between reef-level and site-level variability in coral dynamics.

Our reef selection procedure follows a two-step geographical grouping approach. First, we identify and automatically include reefs at the northernmost and southernmost points to capture the full north-south range of environmental conditions across the Great Barrier Reef. Second, we select additional reefs that are spaced evenly along the longitudinal range, ensuring balanced spatial representation.

2.2.3 Data Formatting for Stan Implementation

For model implementation in Stan, the coral cover data is structured as a matrix where rows typically represent observation time points and columns represent the different observation sites. This matrix, containing the time series for each site, is provided to Stan along with its dimensions, such as the total number of time points and the total number of sites. In our complete dataset, this involves observations from 45 sites distributed across 15 reefs.

To uniquely identify observations from different sites within the same year for the model, we generate distinct time indicators. This is achieved by combining the observation year with a small, unique offset for each site. These adjusted time indicators are then supplied to Stan. The modelling framework also uses a reference point, typically zero, to mark the beginning of the observation period.

A site to reef mapping is maintained to define the hierarchical structure. This is typically an array where each element indicates the parent reef for every observation site, allowing the model to understand which sites belong to the same reef and to establish appropriate hierarchical relationships. The total number of reefs is also provided to Stan, enabling correct dimensioning of reef level parameters within the model.

2.2.4 Preparation of an Experimental Dataset for Sparsity Analysis

To evaluate the relative advantages of different pooling strategies under varying data availability scenarios, we create a modified dataset with artificially induced sparsity. From our complete dataset of 15 reefs (each with three sites), we randomly select 5 reefs to retain all three sites, 5 reefs to retain only two sites, and 5 reefs to retain just one site. This experimental design allows us to quantify how different pooling approaches can leverage information from data-rich reefs to improve parameter estimates for data-poor reefs, a common challenge in ecological monitoring where sampling effort often varies across locations.

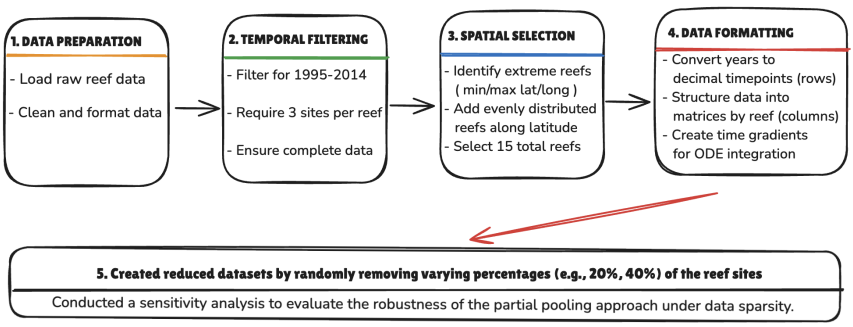


Figure 2.1. The data preprocessing workflow transforming raw monitoring data into model-ready inputs. Key stages include: (1) **Data Preparation**: Loading raw Australian reef data, calculating total hard coral cover, and normalising it to [0, 1]. (2) **Temporal Filtering**: Restricting data to the 1995-2004 period and retaining only reefs with complete annual observations within this timeframe. (3) **Spatial Selection**: Filtering for reefs with exactly 3 sites per year and no missing cover values, resulting in a final set of 15 geographically representative reefs. (4) **Data Formatting**: Converting calendar years to decimal timepoints (with site offsets), structuring data into matrices suitable for Stan, and preparing initial conditions. (5) **Sensitivity Analysis**: Generating sparser datasets (e.g., by randomly removing 20% or 40% of observations) and creating missing data indicators to evaluate model robustness, particularly the performance of partial pooling under data scarcity.

3. Methodology

3.0.1 Introduction to ODEs for Coral Dynamics

Ordinary Differential Equations (ODEs) are powerful tools for simulating dynamic processes in many scientific domains.

Definition 3.1 (Ordinary Differential Equation [52])**.** *A differential equation, in general, is an equation relating a function to its derivatives. An ODE is a differential equation where the unknown function y is a function of a single variable.*

In the context of coral reef dynamics, we use a specific ODE model provided by Brown et al. [13], which describes the transitions between healthy and bleached coral states over time. Before explicitly introducing the equations, we clearly define the state variables used in the model:

- $C(t) \geq 0$ represents the fraction of coral cover that is in a healthy (assimilating) state at time t.
- $B(t) \geq 0$ represents the fraction of coral cover that is in a bleached state at time t.

Together, these states form the total living coral cover at any given time. Additionally, there is an implicit fraction of dead coral, ensuring the total cover (healthy, bleached, and dead) always sums exactly to 1. Thus, the following constraint applies:

$$C(t) + B(t) \leq 1$$

With these definitions in place, the ODE system describing coral dynamics is explicitly given as:

$$\begin{aligned} \frac{dC}{dt} &= \alpha C(1 - (C + B)) - \beta C + \gamma B, \\ \frac{dB}{dt} &= \beta C - \gamma B - \mu B. \end{aligned} \tag{3.1}$$

In these equations, four key parameters model the coral dynamics. The parameter α represents the intrinsic growth rate of healthy coral, while β denotes the rate at which healthy coral transitions to a bleached state. Conversely, γ indicates the recovery rate at which bleached coral returns to a healthy state through assimilation. Finally, μ describes the mortality rate of bleached coral

induced by bleaching, all parameters are assumed to be strictly positive. The model incorporates both ecological dynamics and biological constraints, accurately capturing transitions between coral reef health states over time.

3.0.2 Bayesian inference

Bayes' Theorem

Our ODE model offers a structured framework for examining the dynamics of coral reefs. Given the complexity of ecological systems, a thorough understanding of the associated uncertainties is crucial. The Bayesian approach directly addresses this need. Unlike frequentist approaches, the Bayesian methods used in this study generate predictions that fully incorporate uncertainty from all perspectives. These include uncertainty in the parameter estimates themselves, which is captured by posterior distributions. It also incorporates uncertainty due to observation errors, an aspect accounted for in the likelihood function. Furthermore, it handles uncertainty arising from how parameters might vary systematically across different groups, a characteristic managed by hierarchical model components. This thorough approach to quantifying uncertainty is especially useful for complex ecological processes like coral dynamics, which vary over space and time. As detailed in Chapter 2, our coral-cover data involve unavoidable measurement errors and potential biases from the photographic-analysis methods used.

Given these challenges, understanding complex ecological systems requires an inherently statistical perspective. Bayesian methods are particularly well-suited for this, providing predictions that realistically reflect uncertainty, especially for dynamic processes with complex behaviour over space and time [47]. The mechanism for integrating new information and refining our knowledge within this Bayesian framework is Bayes' theorem, which offers a formal procedure for updating initial beliefs based on observed data.

Definition 3.2 (Bayes' theorem [31, 61])**.** *Let* $\mathbf{Y}_{\text{obs}}$ *be the observed data and let* θ *denote the vector of model parameters. Bayes' theorem states that the posterior probability density of* θ *given the data is*

$$p(\theta \mid \mathbf{Y}_{\text{obs}}) = \frac{p(\theta)\, p(\mathbf{Y}_{\text{obs}} \mid \theta)}{p(\mathbf{Y}_{\text{obs}})}, \qquad \textit{where} \quad p(\mathbf{Y}_{\text{obs}}) = \int p(\theta)\, p(\mathbf{Y}_{\text{obs}} \mid \theta)\, \mathrm{d}\theta.$$

In this expression the factor $p(\theta)$ *is the prior and* $p(\mathbf{Y}_{\text{obs}} \mid \theta)$ *is the likelihood. The quantity* $p(\mathbf{Y}_{\text{obs}})$*, called the marginal likelihood or evidence, ensures that the posterior distribution integrates to one.*

In our Bayesian framework, the prior distributions encode our existing knowledge and ecological assumptions about coral-cover dynamics before observing the data. The likelihood function represents the probability of observing our

coral-cover measurements given specific parameter values in our ODE model. The posterior distributions, which we obtain through Bayesian inference, provide comprehensive probability distributions for all four parameters (α, β, γ, and μ) in our ODE system, reflecting our updated knowledge after incorporating both prior information and observed data.

Bayesian hierarchical model likelihood

To specify our Bayesian hierarchical inference, we first describe the model for the observed data. We assume that individual coral-cover measurements, $y_{j,t}$, follow a normal (Gaussian) distribution. The model for a single observation at site j and time t is

$$y_{j,t} \sim \mathcal{N}(\hat{y}_{j,t}, \sigma^2_{\text{obs}}) \tag{3.2}$$

where $\mathcal{N}(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2. The mean, $\hat{y}_{j,t}$, is the deterministic prediction from the ODE system (3.1). It is calculated as $\hat{y}_{j,t} = f(\theta_j, t)$, which specifically represents the sum of the simulated healthy and bleached coral cover, $C_j(t) + B_j(t)$, as obtained by solving the ODE model for site j with its site-specific parameter vector $\theta_j = (\alpha_j, \beta_j, \gamma_j, \mu_j)$. The variance, σ^2_{obs}, represents Gaussian measurement noise.

Combining contributions across all J sites and T time points yields the full hierarchical likelihood for the complete data set, Y_{obs}:

$$p(Y_{\text{obs}} \mid \Theta, \sigma^2_{\text{obs}}) = \prod_{j=1}^{J} \prod_{t=1}^{T} \mathcal{N}(y_{j,t} \mid \hat{y}_{j,t}, \sigma^2_{\text{obs}}) \tag{3.3}$$

where $Y_{\text{obs}} = \{y_{j,t}\}_{j,t}$ is the complete data set, and $\Theta = \{\theta_j\}_{j=1}^{J}$ is the collection of all site-level parameter vectors.

This explicit likelihood formulation ensures that both the deterministic ODE dynamics and observational uncertainty are rigorously incorporated into the model.

Bayesian hierarchical model prior

Our model incorporates a three-level hierarchy reflecting the structure of coral reefs (sites within reefs, and reefs within a broader population). For clarity, the complete ODE parameter vector for any site j is denoted as $\theta_j = (\alpha_j, \beta_j, \gamma_j, \mu_j)$, with corresponding population-level means $\mu_{\text{pop}} = (\mu_\alpha, \mu_\beta, \mu_\gamma, \mu_\mu)$. The principle of how multilevel models balance information can be illustrated by considering estimation of a single ODE parameter, such as α_j, which is one component of the site-specific parameter vector.

The weighting factor ω determines how information is pooled between the higher-level mean and the individual site estimate, and is defined as:

$$\omega = 1 - \frac{\sigma^2_\alpha}{\sigma^2_\alpha + \sigma^2_{\text{site}}}. \tag{3.4}$$

where σ_α^2 represents the variance of the true values of the ODE parameter α across different groups (e.g., sites within a reef, or reefs themselves) around the higher-level mean μ_α. The variance term σ_{site}^2 denotes the sampling variance associated with the individual estimate $\hat{\alpha}_j$ for site j, reflecting the precision of estimating this parameter using only data from that site.

The multilevel estimate for this ODE parameter, denoted $\hat{\alpha}_j^{\text{multilevel}}$, can be conceptualised as a weighted average [65]:

$$\hat{\alpha}_j^{\text{multilevel}} = \omega\,\mu_\alpha + (1-\omega)\,\hat{\alpha}_j. \tag{3.5}$$

Here, μ_α represents the mean value of this specific ODE parameter α at a higher hierarchical level. For example, if α_j is a site-level parameter, μ_α could be the corresponding reef-level mean for α (i.e., the α-component of μ_{reef_i} from the hierarchical model), or it could represent the global population mean for α (the α-component of μ_{pop}). The term $\hat{\alpha}_j$ signifies the individual estimate of the ODE parameter α for site j, derived solely from the data of site j (i.e., $\hat{\alpha}_j^{\text{no-pool}}$).

The multilevel framework outlined above leads to different estimation strategies depending on the assumptions about these variances for each ODE parameter in θ_j.

Complete pooling

Under the complete-pooling assumption, we set $\omega = 1$ for the ODE parameter α. This scenario arises when we assume there is no true variation in α across groups, i.e., the heterogeneity component is zero, $\sigma_\alpha^2 = 0$ (see Figure 3.1). All groups j are therefore treated identically, and the estimate for each site becomes the population-level mean:

$$\hat{\alpha}_j^{\text{multilevel}} = \mu_\alpha. \tag{3.6}$$

In our coral-reef context, complete pooling means that every reef and site shares identical parameter values. If this assumption were applied to all four ODE parameters in θ_j, we would have $\theta_j \equiv \mu_{\text{pop}}$ for every site j, where we define the population-level mean vector as $\mu_{\text{pop}} = (\mu_\alpha, \mu_\beta, \mu_\gamma, \mu_\mu)$. The entire Great Barrier Reef would then be modelled as a single homogeneous system, ignoring any ecological differences between reefs.

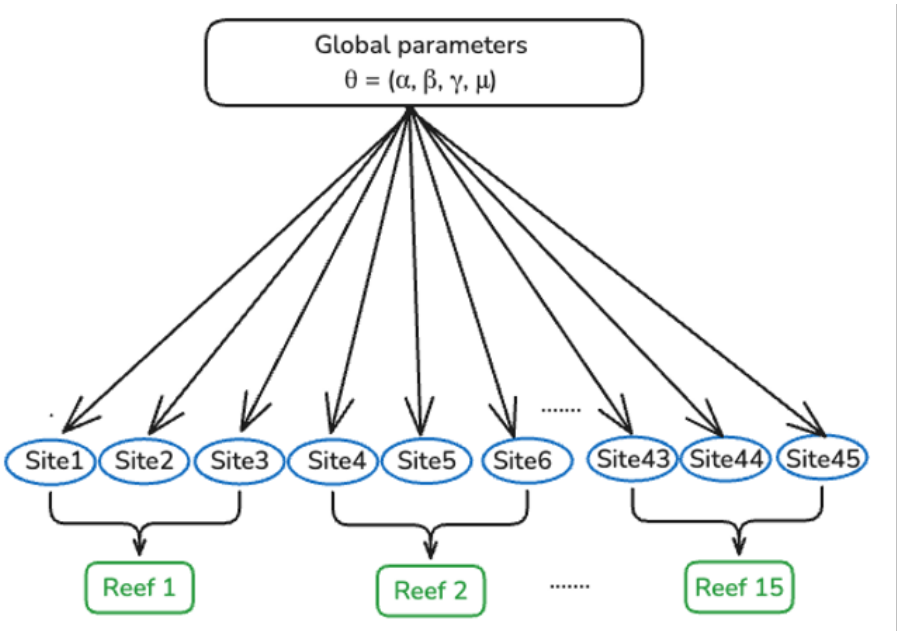


Figure 3.1. Complete pooling: all sites share the same global parameter vector $\theta = (\alpha, \beta, \gamma, \mu)$. Forty-five sites (blue ovals) inherit these parameters directly, and the sites are grouped into 15 reefs (green rectangles).

No pooling

Under the no-pooling assumption for an ODE parameter α, we effectively set $\omega = 0$. This occurs because we assume all sites and reefs are completely independent systems with no meaningful shared properties (see Figure 3.2).

To understand why $\omega = 0$ under this condition, consider the case where the reef-level variance σ_α^2 is effectively infinite (i.e., $\sigma_\alpha^2 \to \infty$). We can then evaluate the limit of the pooling fraction. Assuming σ_{site}^2 (the site-level sampling variance) is finite and positive, we divide both the numerator and denominator by σ_α^2:

$$\frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_{\text{site}}^2} = \frac{\frac{\sigma_\alpha^2}{\sigma_\alpha^2}}{\frac{\sigma_\alpha^2}{\sigma_\alpha^2} + \frac{\sigma_{\text{site}}^2}{\sigma_\alpha^2}} = \frac{1}{1 + \frac{\sigma_{\text{site}}^2}{\sigma_\alpha^2}}.$$

As $\sigma_\alpha^2 \to \infty$, the term $\frac{\sigma_{\text{site}}^2}{\sigma_\alpha^2}$ approaches 0. Thus, the limit of the fraction becomes $\frac{1}{1+0} = 1$. Substituting this result back into the definition of ω [see Eq. (3.4)],

$$\omega = 1 - 1 = 0.$$

The limiting calculation above confirms that ω is set to 0 when σ_α^2 is considered infinite.

Consequently, when $\omega = 0$, the higher-level mean μ_α provides no useful information for estimating any specific group's parameter. In this case, the estimate for the ODE parameter α_j for each group j is simply its individual estimate $\hat{\alpha}_j$:

$$\hat{\alpha}_j^{\text{multilevel}} = \hat{\alpha}_j. \tag{3.7}$$

In our coral reef context, this means each reef is analysed as a completely isolated system with its own unique parameter set θ_j, estimated solely from its own data without borrowing any information from other reefs or sites.

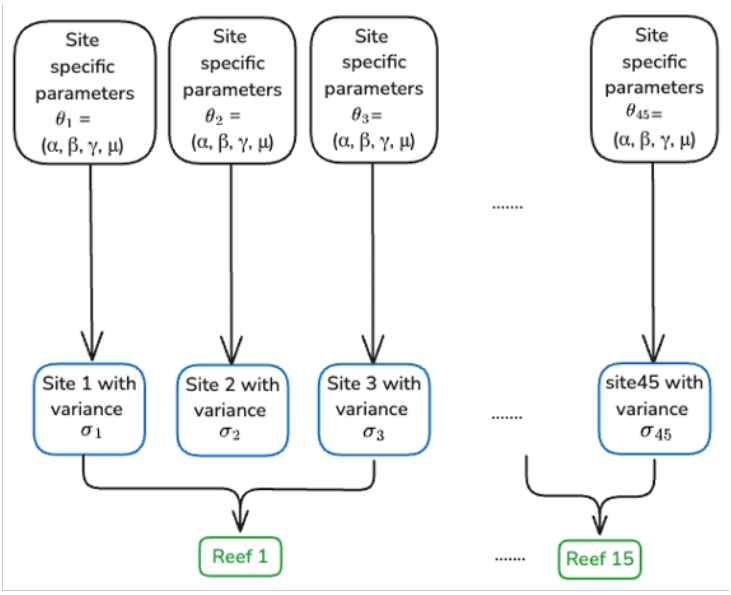


Figure 3.2. No pooling: each site j has its own parameter vector $\theta_j = (\alpha_j, \beta_j, \gamma_j, \mu_j)$, estimated independently from other sites and reefs. The diagram displays site-specific parameter boxes (top), blue ovals for sites, and green rectangles for reefs, indicating complete independence in parameter estimation across the system.

Partial pooling

Under the partial pooling assumption for an ODE parameter α, we effectively set ω to an intermediate value between 0 and 1. This setting reflects the assumption that sites and reefs share some common properties whilst retaining their individual characteristics, allowing information to be borrowed across the hierarchical structure (see Figure 3.3).

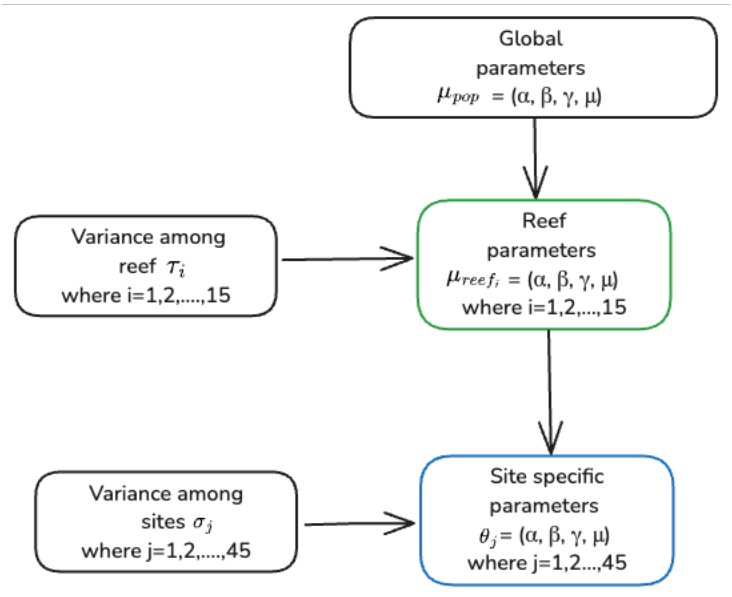


Figure 3.3. Illustrating the partial pooling approach across the reef system. The diagram displays, at the top, a rounded rectangle containing the global parameters $\mu_{\text{pop}} = (\alpha, \beta, \gamma, \mu)$ representing population-level effects. An arrow points to the green rounded rectangle containing reef parameters $\mu_{\text{reef}_i} = (\alpha, \beta, \gamma, \mu)$ for $i = 1, 2, \ldots, 15$, showing how reef-level parameters derive from global parameters. Additionally, a grey rounded rectangle on the left indicates "variance among reefs τ_i for $i = 1, 2, \ldots, 15$", with an arrow pointing to the reef parameters, demonstrating how reef-specific variation is incorporated. Similarly, the reef parameters flow down to site-specific parameters $\theta_j = (\alpha, \beta, \gamma, \mu)$ for $j = 1, 2, \ldots, 45$ (shown in blue), with another variance component "variance among sites σ_j for $j = 1, 2, \ldots, 45$" feeding into this level. This diagram effectively captures the complete three-level hierarchical structure with explicit variance components at each level.

At the base of the hierarchical model, the observed coral cover data are represented as $y_{j,t}$, with j indexing sites and t indexing time points. Each observation is modelled as a Normal random variable:

$$y_{j,t} \sim \mathcal{N}(\hat{y}_{j,t}, \sigma_y^2),$$

where $\mathcal{N}(\mu, \sigma^2)$ denotes a normal (Gaussian) distribution with mean μ and variance σ^2, $\hat{y}_{j,t} = f(\theta_j, t)$ is the ODE model prediction for site j at time t, and σ_y^2 quantifies the variance of the natural noise inherent in the dataset.

At the next level of the hierarchy, site-specific parameters are assigned:

$$\theta_j \sim \mathcal{N}(\mu_{\text{reef}_i}, \sigma_j^2),$$

where $\theta_j = (\alpha_j, \beta_j, \gamma_j, \mu_j)$ is the parameter vector for site j, $\mu_{\text{reef}_i} = (\alpha_i, \beta_i, \gamma_i, \mu_i)$ represents the parameters for reef i (with i indicating the reef to which site j belongs), and σ_j^2 is the covariance among sites. There are 15 reefs in total.

At the highest level of the model, reef parameters are defined as

$$\mu_{\text{reef}_i} \sim \mathcal{N}(\mu_{\text{pop}}, \tau_i^2),$$

where $\mathcal{N}(\mu, \sigma)$ again denotes a multivariate normal distribution, $\mu_{\text{pop}} = (\mu_\alpha, \mu_\beta, \mu_\gamma, \mu_\mu)$ is the global parameter vector, and τ_i^2 is the covariance among reefs, for $i = 1, 2, \ldots, 15$.

Bayesian hierarchical model posterior

Combining the hierarchical likelihood with the specified priors via Bayes' theorem yields the joint posterior distribution for all parameters in the model. In full, the posterior is proportional to the product of the likelihood and the successive levels of the prior:

$$\begin{aligned} p(\Theta, \mu_{\text{reef}}, \mu_{\text{pop}}, \ldots \mid Y_{\text{obs}}) \propto\ & \underbrace{p(Y_{\text{obs}} \mid \Theta, \sigma^2_{\text{obs}})}_{\text{Likelihood}} \times \underbrace{\prod_{j=1}^{J} p(\theta_j \mid \mu_{\text{reef}(i(j))}, \sigma^2_{\text{site}})}_{\text{Site-Level Prior}} \\ & \times \underbrace{\prod_{i=1}^{I} p(\mu_{\text{reef}i} \mid \mu_{\text{pop}}, \tau^2_{\text{reef}})}_{\text{Reef-Level Prior}} \\ & \times \underbrace{p(\mu_{\text{pop}})\, p(\sigma_{\text{site}})\, p(\tau_{\text{reef}})\, p(\sigma^2_{\text{obs}})}_{\text{Hyperpriors}}. \end{aligned} \tag{3.8}$$

where J is the total number of sites (which is 45 in this project), and I is the total number of reefs (which is 15 here).

The distribution $p(\Theta, \mu_{\text{reef}}, \mu_{\text{pop}}, \ldots \mid Y_{\text{obs}})$ embodies the complete and updated knowledge of the parameters once the observed data have been incorporated.

A central feature of the hierarchical model is that the posterior provides simultaneous estimates of coral dynamics across all three levels of the hierarchy. The posterior distribution for any site-specific parameter vector θ_j is informed not only by the data from site j, but is also regularised by the reef-level and population-level mean posteriors. This partial pooling behaviour shares information across the entire system, leading to more robust and credible inferences.

3.0.3 Markov Chain Monte Carlo for Bayesian inference

Computational Challenges in Bayesian Inference

Performing full Bayesian inference for our hierarchical ODE model, given by Equation (3.1) (denoted M), involves characterising the posterior distribution

of a comprehensive set of model parameters. This set of parameters, which we denote Θ, forms the overall parameter space of our model. As we mention in the partial pooling subsection, for our model M, these core parameters are organised into the following hierarchical levels:

$$\Theta = \{ \underbrace{\mu_{\text{pop}}}_{\text{1. Global Parameters}}, \underbrace{\tau_i,\ \{\mu_{\text{reef}_i}\}_{i=1}^{I}}_{\text{2. Reef-level Parameters \& Effects}}, \underbrace{\sigma_j,\ \{\theta_j\}_{j=1}^{J}}_{\text{3. Site-specific Parameters \& Effects}}, \underbrace{\sigma_y,\ y_{0\text{init,C,B}}}_{\text{4. Observation \& Initial State Parameters}} \}, \tag{3.9}$$

where $I = 15$ reefs, $j = 45$ sites.

In Bayesian inference, we calculate expectations to perform predictions or quantify uncertainty [49]. For example, the posterior mean of parameters Θ is an expectation, and variance, derived from expectations, measures the spread of parameter values. These expectations are mathematically represented as multi-dimensional integrals over the parameter space Θ. For a generic function $f(\Theta)$ of interest, we often need to compute:

$$\mathbb{E}_\pi[f(\Theta)] = \int f(\Theta)\pi(\Theta)\, d\Theta,$$

where $\pi(\Theta)$ is a probability density function (e.g., the prior $p(\Theta \mid M)$ or the posterior $p(\Theta \mid Y_{\text{obs}}, t, M)$).

The Law of Large Numbers provides the theoretical basis for approximating these integrals numerically using Monte Carlo methods. It states that the expectation $\mathbb{E}_\pi[f(\Theta)]$ can be approximated by drawing S independent and identically distributed (i.i.d.) samples, denoted Θ^s (where $s = 1, \ldots, S$), from the density $\pi(\Theta)$ and then calculating the sample average:

$$\hat{I}_S(f) = \frac{1}{S}\sum_{s=1}^{S} f(\Theta^s) \xrightarrow{S\to\infty} \mathbb{E}_\pi[f(\Theta)]. \tag{3.10}$$

However, effectively applying this principle to the specific integrals required for Bayesian inference with our complex hierarchical ODE model reveals significant practical difficulties.

Key inferential quantities, such as posterior predictions or parameter summaries, require integrating over the posterior distribution. For our specific partial pooling model, denoted M_{pp}, and its associated parameter set Θ_{pp}, this means working with $p(\Theta_{\text{pp}} \mid Y_{\text{obs}}, t, M_{\text{pp}})$. For example, the posterior predictive

density for replicated data Y_{rep} is:

$$\begin{aligned}\hat{p}(Y_{\text{rep}} \mid t, Y_{\text{obs}}, M_{\text{pp}}) &= \frac{1}{S}\sum_{s=1}^{S} p(Y_{\text{rep}} \mid t, \Theta_{\text{pp}}^{s}, M_{\text{pp}}), \\ &\approx \mathbb{E}_{p(\Theta_{\text{pp}} \mid Y_{\text{obs}}, t, M_{\text{pp}})}\left[p(Y_{\text{rep}} \mid t, \Theta_{\text{pp}}, M_{\text{pp}})\right], \\ &= \int p(Y_{\text{rep}} \mid t, \Theta_{\text{pp}}, M_{\text{pp}}) p(\Theta_{\text{pp}} \mid Y_{\text{obs}}, t, M_{\text{pp}})\, d\Theta_{\text{pp}}, \\ &\text{where } \Theta_{\text{pp}}^{s} \sim p(\Theta_{\text{pp}} \mid Y_{\text{obs}}, t, M_{\text{pp}}).\end{aligned} \tag{3.11}$$

The fundamental obstacle here is that the posterior distribution $p(\Theta_{\text{pp}} \mid Y_{\text{obs}}, t, M_{\text{pp}})$ is analytically intractable. Because the likelihood function involves the output of a numerical ODE solver, the resulting posterior distribution, formed by combining this complex likelihood with the priors, does not belong to any standard family from which independent and identically distributed (i.i.d.) samples can be readily generated. As Girolami (2008) notes, "Now the problem we face is that samples from a density which has no analytic form are required" [35].

Motivation for Markov Chain Monte Carlo

The computational difficulties we previously identified, especially the challenge of performing inference with our model's high-dimensional and non-standard posterior distribution, require advanced sampling techniques. It's often intuitive to introduce Markov chains by first considering finite state spaces.

Definition 3.3 (Markov Chain [3])**.** *Imagine a process where a variable $x^{(i)}$ at step i can only take one of s discrete values from a set $\mathscr{X} = \{x_1, x_2, \ldots, x_s\}$. This stochastic process, denoted $\{x^{(i)}\}$, is called a Markov chain if the probability of its current state, given all past states, depends only on its immediately preceding state:*

$$p(x^{(i)} \mid x^{(i-1)}, \ldots, x^{(1)}) = T(x^{(i)} \mid x^{(i-1)}). \tag{3.12}$$

Here, $T(x^{(i)} \mid x^{(i-1)})$ represents the probability of transitioning to state $x^{(i)}$ from the previous state $x^{(i-1)}$. The chain is described as ***homogeneous*** *if these transition probabilities $T(x^{(i)} \mid x^{(i-1)})$ remain the same for all steps i. A key property is that for any given previous state $x^{(i-1)}$, the sum of probabilities of transitioning to any possible next state $x^{(i)}$ is one: $\sum_{x^{(i)}} T(x^{(i)} \mid x^{(i-1)}) = 1$. In essence, the evolution of the chain within the space $\mathscr{X}$ depends solely on its current state and a fixed transition matrix (composed of these probabilities T).*

The key insight that connects Markov chains to our inference problem is the ability to construct a chain whose stationary distribution matches the target posterior distribution. This approach allows us to generate a sequence of

samples that, after an initial burn-in period which gives the chain time to reach its stationary behavior, can be treated as approximate draws from the posterior distribution, even when that distribution is known only up to a normalizing constant [20].

MCMC methods provide the framework to do exactly this. They construct a Markov chain carefully designed so that its stationary distribution is indeed the target posterior. The chain then generates samples that allow us to effectively explore the complex landscape of this posterior, even when it's known only up to a constant of proportionality [76]. The power of MCMC becomes particularly clear when dealing with complex, high-dimensional problems. In fact, for certain challenging tasks, such as estimating the volume of a convex body in d dimensions, MCMC simulation is the only general approach currently known to be feasible within polynomial time [77]. It is this demonstrated capability for handling such high-dimensional complexity that motivates applying MCMC to our Bayesian hierarchical ODE model.

Principles of the Hamiltonian Monte Carlo Method

In our project, we use the HMC method to efficiently sample from the posterior distribution of our ODE model parameters.

HMC draws inspiration from Hamiltonian dynamics in physics. In this framework, a system's state consists of both position $\theta \in \mathbb{R}^d$ and momentum $r \in \mathbb{R}^d$ of the same dimension. The Hamiltonian is a function of these two variables, written as $H(\theta, r)$. In HMC applications, we typically define the Hamiltonian as:

$$H(\theta, r) = U(\theta) + K(r). \tag{3.13}$$

HMC is an MCMC algorithm that avoids random walking behaviour and sensitivity to correlated parameters. It uses first-order gradients of the log-posterior to deterministically generate proposal trajectories in parameter space, steered by the momentum variables introduced in the Hamiltonian formulation. This momentum variable heavily influences the proposal distribution in HMC, and the algorithm updates both the target component and the momentum. Because of these characteristics, it can converge to high-dimensional target distributions considerably faster than other MCMC algorithms [42].

Our objective is to sample from a target probability distribution, denoted $\pi(\theta)$. HMC approaches this by defining a potential energy function proportional to the negative log-probability, $U(\theta) = -\log \pi(\theta)$. It also introduces an auxiliary momentum variable, r, which has an associated kinetic energy defined as $K(r) = \frac{1}{2} r^T Q^{-1} r$. The term Q represents a symmetric, positive-definite mass matrix. This matrix is typically chosen (e.g., as the identity matrix or estimated from the data) to tune the sampler's performance. While r is an artificial variable needed for the simulation dynamics, the ultimate aim is to generate samples of θ from the target distribution $\pi(\theta)$ [56].

As time t progresses, the position and momentum change according to Hamilton's equations [56]:

$$\frac{d\theta_i}{dt} = \frac{\partial H}{\partial r_i}, \quad \frac{dr_i}{dt} = -\frac{\partial H}{\partial \theta_i}. \tag{3.14}$$

There are two primary steps in the HMC algorithm (see Algorithm 1).

Algorithm 1 Hamiltonian Monte Carlo (HMC) sampling

Require: Initial position $\theta^{(1)}$, step size ε, number of leapfrog steps L, number of iterations Z, mass matrix $\mathbf{Q}$
Ensure: Posterior samples $\{\theta^{(t)}\}_{t=1}^{Z}$

Set initial state: $\theta \leftarrow \theta^{(1)}$
for $t = 1, \ldots, Z$ **do**
  Sample momentum: $r^{(0)} \sim \mathcal{N}(0, \mathbf{Q})$
  Set proposal: $\tilde{\theta} \leftarrow \theta$, $\tilde{r} \leftarrow r^{(0)}$
  for each leapfrog step $i = 1, \ldots, L$ **do**
    Update momentum (half-step): $\tilde{r} \leftarrow \tilde{r} - \frac{\varepsilon}{2}\nabla U(\tilde{\theta})$
    Update position (full-step): $\tilde{\theta} \leftarrow \tilde{\theta} + \varepsilon \mathbf{Q}^{-1}\tilde{r}$
    Update momentum (half-step): $\tilde{r} \leftarrow \tilde{r} - \frac{\varepsilon}{2}\nabla U(\tilde{\theta})$
  end for
  Compute acceptance probability:
    $\alpha = \min\left(1, \frac{\exp\{-U(\tilde{\theta}) - \frac{1}{2}\tilde{r}^T\mathbf{Q}^{-1}\tilde{r}\}}{\exp\{-U(\theta) - \frac{1}{2}r^{(0)T}\mathbf{Q}^{-1}r^{(0)}\}}\right)$
  Accept or reject proposal:
    With probability α, set $\theta \leftarrow \tilde{\theta}$
    Otherwise, retain θ
  Store $\theta^{(t)} \leftarrow \theta$
end for
Return posterior samples $\{\theta^{(t)}\}_{t=1}^{Z}$

First, regardless of the current θ, we sample r from $\mathcal{N}(0, Q)$, where Q is the mass matrix. Second, we solve the differential equations and propose a new state using Hamiltonian dynamics to sample θ. In practice, we employ the leapfrog method [56], a symplectic integrator that alternates half-steps of momentum and full-steps of position, as an approximation technique since we often cannot solve the Hamiltonian equations exactly. The leapfrog scheme is favored in HMC because it preserves volume and exhibits excellent long-term stability, even for complex, high-dimensional distributions. For a stepsize

$\varepsilon > 0$, half-step updates are carried out as follows:

$$\begin{aligned} r\left(t+\frac{\varepsilon}{2}\right) &= r(t) - \frac{\varepsilon}{2}\nabla U(\theta(t)), \\ \theta(t+\varepsilon) &= \theta(t) + \varepsilon M^{-1} r\left(t+\frac{\varepsilon}{2}\right), \\ r(t+\varepsilon) &= r\left(t+\frac{\varepsilon}{2}\right) - \frac{\varepsilon}{2}\nabla U(\theta(t+\varepsilon)). \end{aligned} \tag{3.15}$$

Bayesian Approach for Handling Missing Data

To address the common challenge of missing observations in ecological datasets such as ours, our Bayesian hierarchical model incorporates the following principled approach for handling incomplete data. We employ an indicator-based likelihood method that aligns with the ignorable missing-data mechanism framework described by Gelman et al. (2013) [30].

Following our standard notation, we define the complete data as $y = (y^{\text{obs}}, y^{\text{mis}})$, and introduce an inclusion indicator I where each element corresponds to a potential observation at site j, time t:

$$I_{j,t} = \begin{cases} 1 & \text{if } y_{j,t} \text{ is observed} \\ 0 & \text{if } y_{j,t} \text{ is missing} \end{cases} \tag{3.16}$$

Under the assumption that our data are missing at random (MAR), the posterior distribution of model parameters θ can be written as:

$$p(\theta \mid y^{\text{obs}}) \propto \pi(\theta)\, p(y^{\text{obs}} \mid \theta). \tag{3.17}$$

Here, $\pi(\theta)$ is the prior and $p(y^{\text{obs}} \mid \theta)$ is the observed-data likelihood. In practice, our Stan implementation calculates the log-likelihood using the indicator variable:

$$\log p(y^{\text{obs}} \mid \theta) = \sum_{j,t} I_{j,t} \cdot \log p(y_{j,t} \mid \theta_j, \sigma_j^2). \tag{3.18}$$

where

$$p(y_{j,t} \mid \theta_j, \sigma_j^2) = \mathcal{N}(y_{j,t} \mid \hat{y}_{j,t}, \sigma_j^2), \qquad \hat{y}_{j,t} = f(\theta_j, t). \tag{3.19}$$

This approach provides computational efficiency by focusing inference directly on model parameters rather than explicitly modelling missing values, while still yielding valid Bayesian inference under MAR assumptions [51, 33].

Bayesian Implementation with RStan

In this project, we use RStan [70] to implement our Bayesian hierarchical model; RStan provides a convenient R interface to the Stan probabilistic programming language. RStan utilises the No-U-Turn Sampler (NUTS) [42] as its default MCMC algorithm. Standard HMC requires manually tuning both the step size (ε) and the number of leapfrog steps (L) taken in each simulation.

NUTS, in contrast, eliminates the need to manually set L by detecting when the simulated trajectory begins to turn back on itself, which indicates that the simulation has proceeded sufficiently. During the initial warmup phase of sampling, NUTS adaptively tunes the step size ε to achieve an optimal acceptance rate, balancing efficient exploration of the parameter space with accurate simulation of the Hamiltonian dynamics. Additionally, NUTS initialises and adapts the mass matrix $\mathbf{Q}$ based on the posterior geometry, enhancing sampling efficiency in complex hierarchical models. By automatically optimising these parameters, NUTS facilitates reliable posterior sampling from complex probability distributions with minimal user intervention.

Given Stan's powerful sampling capabilities, the next crucial step is selecting the appropriate R interface to implement our specific Bayesian hierarchical model. While other R packages such as `rstanarm` and `brms` provide user-friendly interfaces to Stan, they are primarily designed for standard regression models (linear, generalised linear, and mixed-effects models). Our coral reef model, on the other hand, requires the differential equations, likelihood function, and parameter transformations to be implemented in a way that these higher-level interfaces cannot accommodate. `RStan` provides the necessary low-level access to Stan's modelling language, allowing us to precisely specify our ODE system and hierarchical structure. This flexibility is essential for implementing complex parameter relationships in our coral bleaching dynamics model.

Bayesian Approach for Handling Missing Data

To address the common challenge of missing observations in ecological datasets such as ours, our Bayesian hierarchical model incorporates the following principled approach for handling incomplete data. We employ an indicator-based likelihood method that aligns with the ignorable missing-data mechanism framework described by Gelman et al. (2013) [30].

Following our standard notation, we define the complete data as $y = (y^{\text{obs}}, y^{\text{mis}})$, and introduce an inclusion indicator I where each element corresponds to a potential observation at site j, time t:

$$I_{j,t} = \begin{cases} 1 & \text{if } y_{j,t} \text{ is observed} \\ 0 & \text{if } y_{j,t} \text{ is missing} \end{cases} \tag{3.20}$$

Under the assumption that our data are missing at random (MAR), the posterior distribution of model parameters θ can be written as:

$$p(\theta \mid y^{\text{obs}}) \propto \pi(\theta)\, p(y^{\text{obs}} \mid \theta). \tag{3.21}$$

Here, $\pi(\theta)$ is the prior and $p(y^{\text{obs}} \mid \theta)$ is the observed-data likelihood. In practice, our Stan implementation calculates the log-likelihood using the indicator variable:

$$\log p(y^{\text{obs}} \mid \theta) = \sum_{j,t} I_{j,t} \cdot \log p(y_{j,t} \mid \theta_j, \sigma_j^2). \tag{3.22}$$

where

$$p(y_{j,t} \mid \theta_j, \sigma_j^2) = \mathcal{N}(y_{j,t} \mid \hat{y}_{j,t}, \sigma_j^2), \qquad \hat{y}_{j,t} = f(\theta_j, t). \tag{3.23}$$

This approach provides computational efficiency by focusing inference directly on model parameters rather than explicitly modelling missing values, while still yielding valid Bayesian inference under MAR assumptions [51, 33].sufficiently. During the initial warmup phase of sampling, NUTS adaptively tunes the step size ε to achieve an optimal acceptance rate, balancing efficient exploration of the parameter space with accurate simulation of the Hamiltonian dynamics. Additionally, NUTS initialises and adapts the mass matrix M based on the posterior geometry, enhancing sampling efficiency in complex hierarchical models. By automatically optimising these parameters, NUTS facilitates reliable posterior sampling from complex probability distributions with minimal user intervention.

Given Stan's powerful sampling capabilities, the next crucial step is selecting the appropriate R interface to implement our specific Bayesian hierarchical model. While other R packages like "rstanarm" and "brms" provide user-friendly interfaces to Stan, they are primarily designed for standard regression models (linear, generalised linear, and mixed-effects models). Our coral reef model, on the other hand, needs the differential equations, likelihood function, and parameter transformations to be implemented in a way that these higher-level interfaces cannot. RStan provides the necessary low-level access to Stan's modelling language, allowing us to precisely specify our ODE system and hierarchical structure. This flexibility is essential for implementing complex parameter relationships in our coral bleaching dynamics model.

4. Experiments and Results

This chapter details the experimental setup and rigorous evaluation of various modeling approaches developed for the coral reef dynamics model.

The investigation begins with a focus on prior selection. This section outlines the experimental design for comparing different priors, the systematic workflow for prior predictive checks, and presents the results of this selection process.

Subsequently, the chapter extensively evaluates different pooling strategies. This comprehensive assessment includes model fit via posterior predictive checks, calculating Bayesian p-values, examining predictive accuracy and stability using leave-one-out cross-validation, and evaluating MCMC convergence and sampling quality. Furthermore, a comparison between centered and non-centered parameterizations for partial pooling models is presented.

The discussion then transitions to addressing the complexities of handling missing data within Bayesian hierarchical models, detailing its implementation and performance, particularly in the context of incomplete datasets. Following this, further analysis demonstrates how hierarchical modeling enhances predictive accuracy when applied to unseen data. Finally, the chapter concludes with a critical comparison of the developed Bayesian approaches against established non-Bayesian methods, aiming to highlight their relative advantages, differences, and overall performance characteristics.

4.1 Prior Selection

4.1.1 Experimental Design for Prior Comparison

Prior selection plays a crucial role in Bayesian hierarchical modelling, particularly for complex systems like coral reef dynamics. Appropriate priors help stabilise inference, prevent divergent transitions in the MCMC algorithm, and incorporate domain knowledge into the model [69]. In our analysis, we compare four distinct prior specifications for the partial pooling method, as detailed in Table 4.1. We begin with flat priors that impose minimal assumptions but can lead to inefficient sampling. The vague normal and small normal priors progressively add more structure while maintaining relatively weak constraints. Finally, our preferred Student-t priors incorporate domain knowledge with appropriate regularisation properties.

Parameter	Flat priors	Vague normal	Small normal	Student-t
Global parameters (μ_{pop})				
$\mu_{\text{pop},\alpha}$	Uniform(0, 10)	Normal(0, 2.5)	Normal(0, 1.5)	Student-t(4, 1.6, 0.8)
$\mu_{\text{pop},\beta}$	Uniform(0, 10)	Normal(0, 2.5)	Normal(0, 1.5)	Student-t(4, 0.9, 0.5)
$\mu_{\text{pop},\gamma}$	Uniform(0, 10)	Normal(0, 2.5)	Normal(0, 1.5)	Student-t(4, 0.2, 0.2)
$\mu_{\text{pop},\mu}$	Uniform(0, 10)	Normal(0, 2.5)	Normal(0, 1.5)	Student-t(4, 0.35, 0.25)
Variance components				
Reef variance (τ_i)	Uniform(0, 5)	Half-Normal(0, 1.0)	Half-Normal(0, 0.25)	Half-Student-t(4, 0, 0.15)
Site variance (σ_j)	Uniform(0, 5)	Half-Normal(0, 1.0)	Half-Normal(0, 0.25)	Half-Student-t(4, 0, 0.15)
Other parameters				
Observation error (σ)	Uniform(0, 3)	Half-Cauchy(0, 1.0)	Half-Cauchy(0, 0.2)	Half-Student-t(4, 0, 0.1)
Initial coral cover (N_0)	Uniform(0.01, 0.99)	Beta(1, 1)	Beta(2, 2)	Beta(2, 2)

Table 4.1. *Experiment of prior distributions for hierarchical model parameters. This table details the four different prior selections evaluated in our analysis: The flat priors with uniform distribution; Vague normal priors use large normal distributions; Small normal priors provide more constraints on the parameter space for better regularisation; Student-t priors incorporate domain knowledge while their heavier tails accommodate potential outliers. Each prior configuration is applied consistently across all model parameters, with appropriate adjustments for different parameter types.*

Prior Predictive Check Workflow

We design the workflow (see algorithm 2) of prior predictive checks in Stan by sampling parameters from their priors and generating synthetic data through the model's generative process using only the `data` and `generated quantities` blocks. This approach allows us to efficiently assess the plausibility of our priors by visualising simulated parameter values and data trajectories before fitting the model to any observed data.

Algorithm 2 Prior predictive checks in Stan

Require: Prior specifications $p(\theta)$, number of simulations S
Ensure: Prior predictive samples Y_{rep}
Define Stan model with only `data` and `generated quantities` blocks
In `generated quantities` block:
  Sample global parameters: $\mu_{\text{pop}} \sim p(\mu_{\text{pop}})$
for each reef $i = 1, \ldots, N_R$ **do**
    Sample reef parameters: $\mu_{\text{reef}_i} \sim p(\mu_{\text{reef}_i} | \mu_{\text{pop}}, \tau)$
    for each site j in reef i **do**
      Sample site parameters: $\theta_j \sim p(\theta_j | \mu_{\text{reef}_i}, \sigma^2)$
      Sample initial condition: $y_{0,j} \sim p(y_0)$
      Solve ODE: $\hat{y}_j = \text{solve\_coral\_model}(t, \theta_j, y_{0,j})$
      Generate observations with noise: $Y_{\text{rep},j} \sim \text{Normal}(\hat{y}_j, \sigma^2)$
    end for
end for
Execute model with `rstan::sampling`:
  Set warmup = 0
  Set algorithm = `"Fixed_param"`
  Set iter = S
Extract simulated parameters and data
Visualize:
  Plot parameter prior distributions
  Plot simulated data trajectories Y_{rep}
  Assess ecological plausibility of simulations
Return simulated parameters and Y_{rep}

Results of Prior Selection

We conduct prior predictive checks by simulating coral cover values under each of our prior specifications (Figure 4.1). The results reveal notable differences in the behaviour of the four prior configurations. Both uniform and vague normal (large normal) priors generate distributions with excessive probability mass at extreme values (near 0 and 1), which is ecologically implausible for coral cover. This observation is consistent with the Stan Development Team's caution that uniform priors can create an implicit preference for boundary values in constrained parameter spaces [70]. Similarly, vague normal priors with large variances place substantial probability mass in implausible regions of the parameter space, often resulting in unrealistic model behaviour during simulation [69].

In contrast, the small normal and Student-t priors yield more plausible simulations, with the Student-t prior producing the most realistic distribution of coral cover values. These prior predictive checks, performed before incorporating any observed data, confirm that the Student-t prior configuration best represents our domain knowledge regarding plausible coral cover dynamics.

Based on these results, we select the Student-t prior specification for our subsequent analyses, as it provides an appropriate balance between regularisation and flexibility to capture the true ecological patterns in the data.

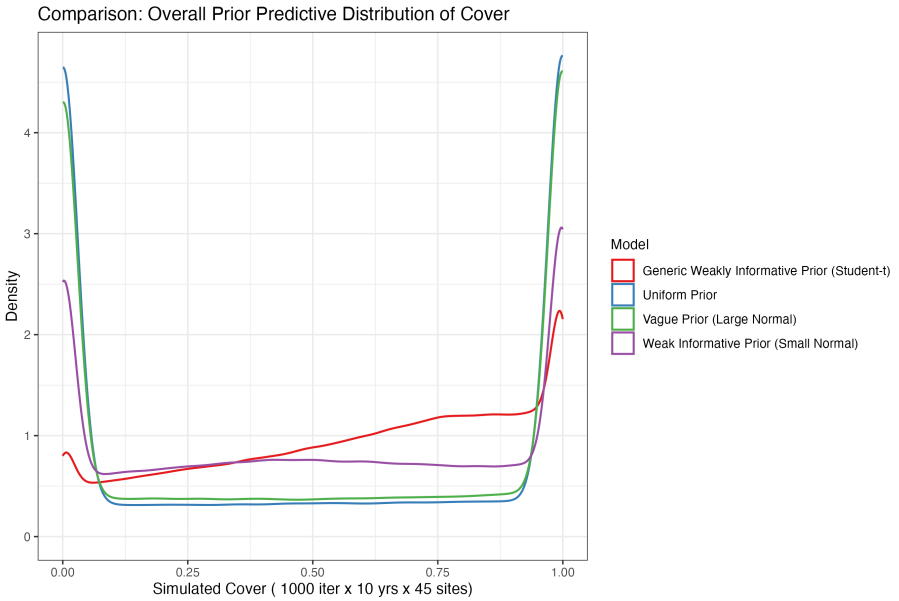


Figure 4.1. Prior predictive distributions of coral cover across four different prior specifications. Each curve shows the expected distribution of coral cover values based purely on simulations from the prior, before seeing any data. The uniform prior (blue) and vague normal prior (green) both exhibit unrealistic concentrations at extreme values (0 and 1). The small normal prior (purple) shows improved behaviour but retains some unrealistic features. The Student-t prior (red) yields the most ecologically plausible distribution, with a smoother profile across the range of possible coral cover values and appropriate density in the mid-range commonly observed in reef systems. Each distribution is based on 1000 simulations across all 45 sites and 10 years of the study period.

4.1.2 Evaluating Pooling Strategies for Coral Reef Dynamics Modelling

Our analysis compares four different parameter estimation approaches, each representing a different assumption about how data from different reefs are related. We investigate three types of pooling: complete pooling, in which all reefs share the same growth and decline parameters; no pooling, in which each reef has its own unique set of parameters with no relationship between reefs; and partial pooling, in which reef parameters are related using hierarchical models that balance individual reef characteristics with overall population trends. We test both centered and non-centered parameterisations in the par-

tial pooling models because this choice can affect computational stability. To determine the best model, we rely on three key diagnostic methods:PPC to assess the model fit to observed data, leave-one-out cross-validation to estimate and compare out-of-sample prediction performance, and MCMC diagnostics to verify the sampler reliability.

Model Fit: Posterior Predictive Checks

The posterior predictive check (PPC) serves as a model evaluation tool that allows us to assess whether the model can generate data consistent with observed patterns, thus providing a quantitative measure of model adequacy [81]. The procedure typically involves visualising posterior predictive distributions alongside the observed data and calculating a test statistic, $T(\cdot)$, which could be the mean, variance, or another relevant summary, applied to both the observed dataset Y_{obs} and each simulated dataset Y_{rep}. The posterior predictive p-value (p_{PPC}) is then defined as the proportion of simulated datasets in which the test statistic $T(Y_{\text{rep}})$ is as extreme or more extreme than the value calculated for the observed data, $T(Y_{\text{obs}})$.

To implement a PPC, we extract parameter values and posterior samples of predicted coral cover (y_{rep}) from the fitted Stan model and calculate quantiles (95%) to create credible intervals. These intervals represent the model's uncertainty in predictions at each timepoint. The plots show that the complete pooling approach fails to adequately capture the dynamics of coral cover (Figure 4.2), with observed data points frequently falling outside the predicted credible intervals. The no pooling method performs better (Figure 4.3), capturing more of the reef-specific patterns. However, both the centered and non-centered partial pooling methods (Figure 4.4 and Figure 4.5) demonstrate superior ability to capture the non-linear shape of the coral cover dynamics.

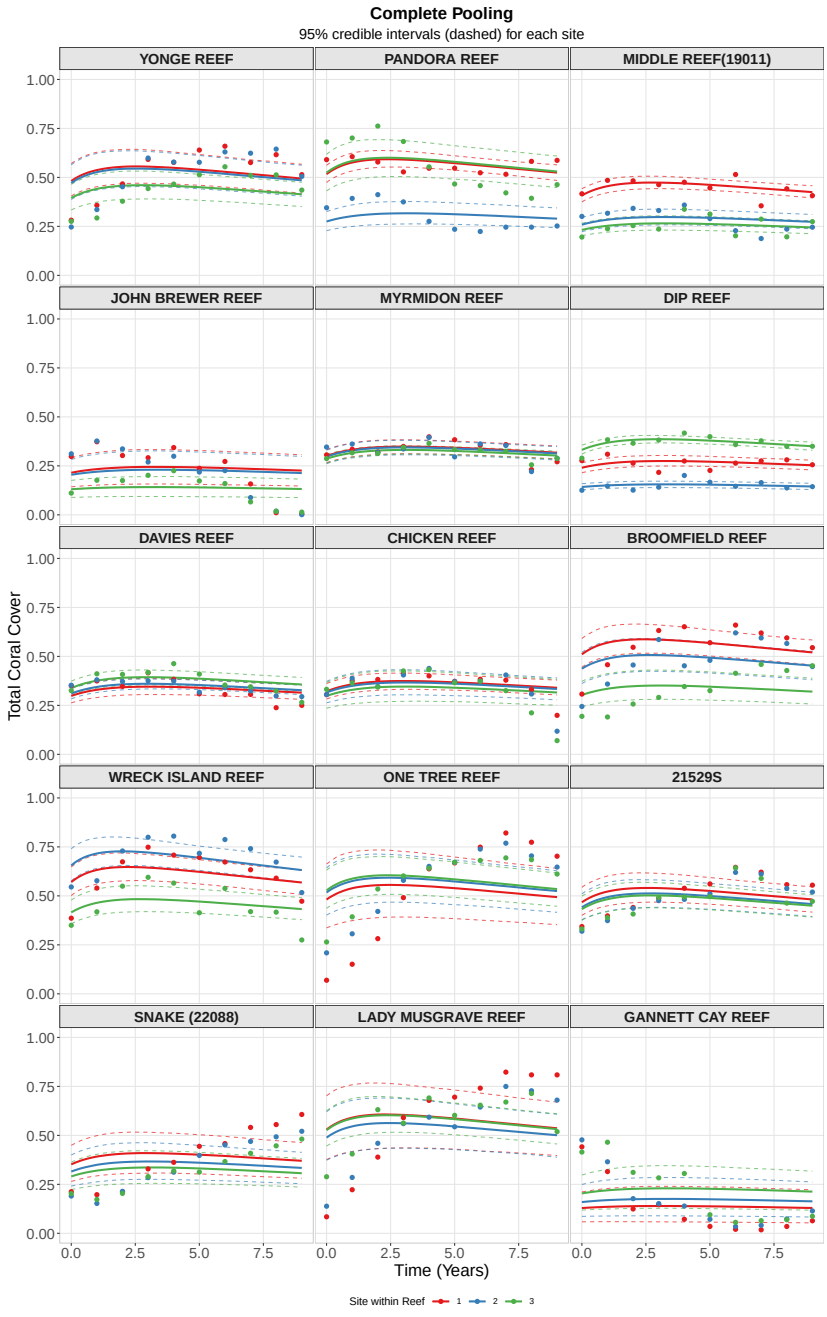


Figure 4.2. Posterior predictive checks for complete pooling approaches. Reefs are ordered by longitude. Different observation sites are distinguished by color: site 1 data is shown in red, site 2 in blue, and site 3 in green. For each site, observed proportional coral cover is displayed as points in its assigned color; a solid line of the same color indicates the median posterior prediction, and corresponding dashed lines represent that site's 95% credible intervals. The complete pooling model's intervals often miss the data points, especially for reefs like LADY MUSGRAVE REEF and ONE TREE REEF where coral cover changes in unique ways.

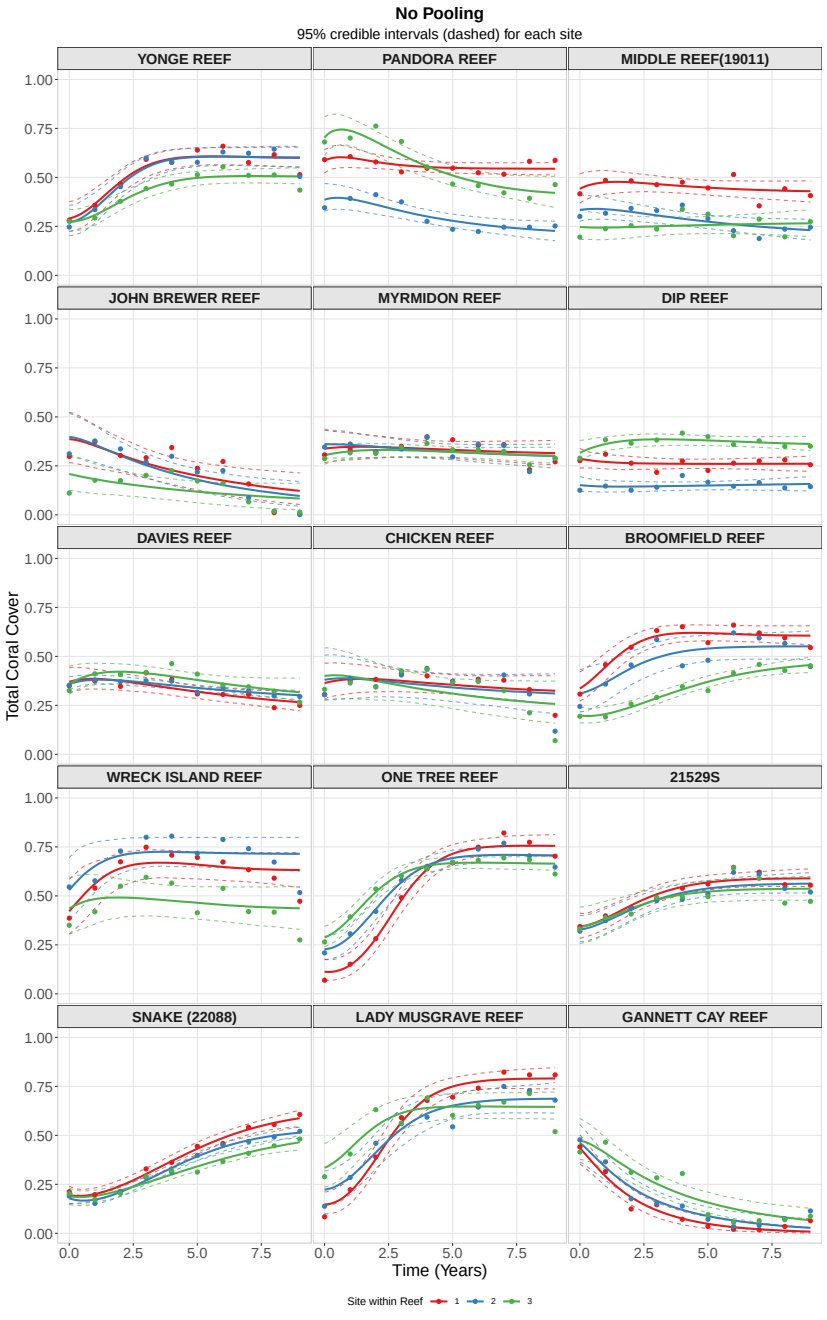


Figure 4.3. Posterior predictive checks for no pooling approaches. The no pooling model fits the data well across all reefs and sites compared to complete pooling, showing it can handle the different patterns seen in each reef.

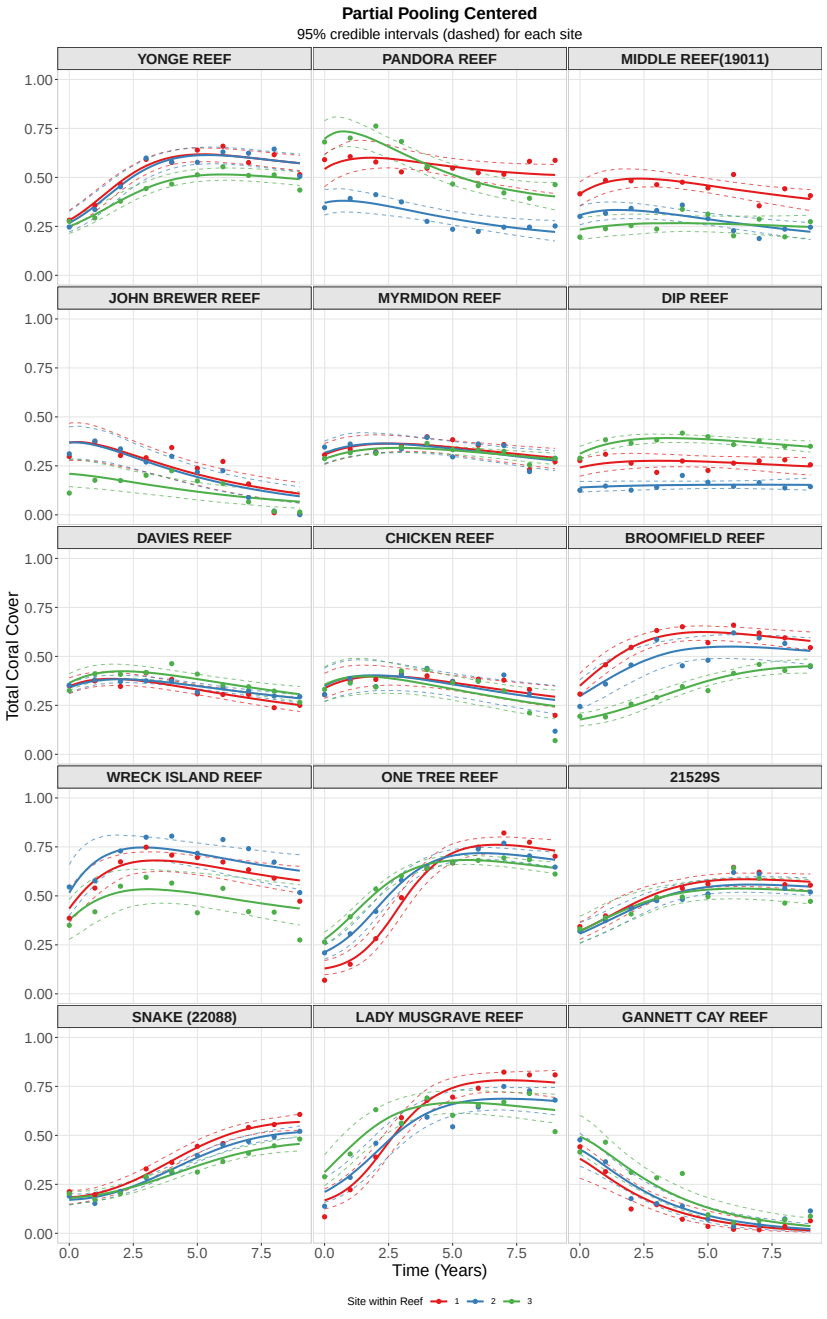


Figure 4.4. Posterior predictive checks for partial pooling centered approach. The partial pooling centered model effectively captures the diverse coral cover patterns across all reefs and sites, balancing between global reef trends and local site variations. The model's predictions closely follow observed data points with appropriate uncertainty bands, demonstrating its ability to account for various growth and decline patterns seen across different reef systems.

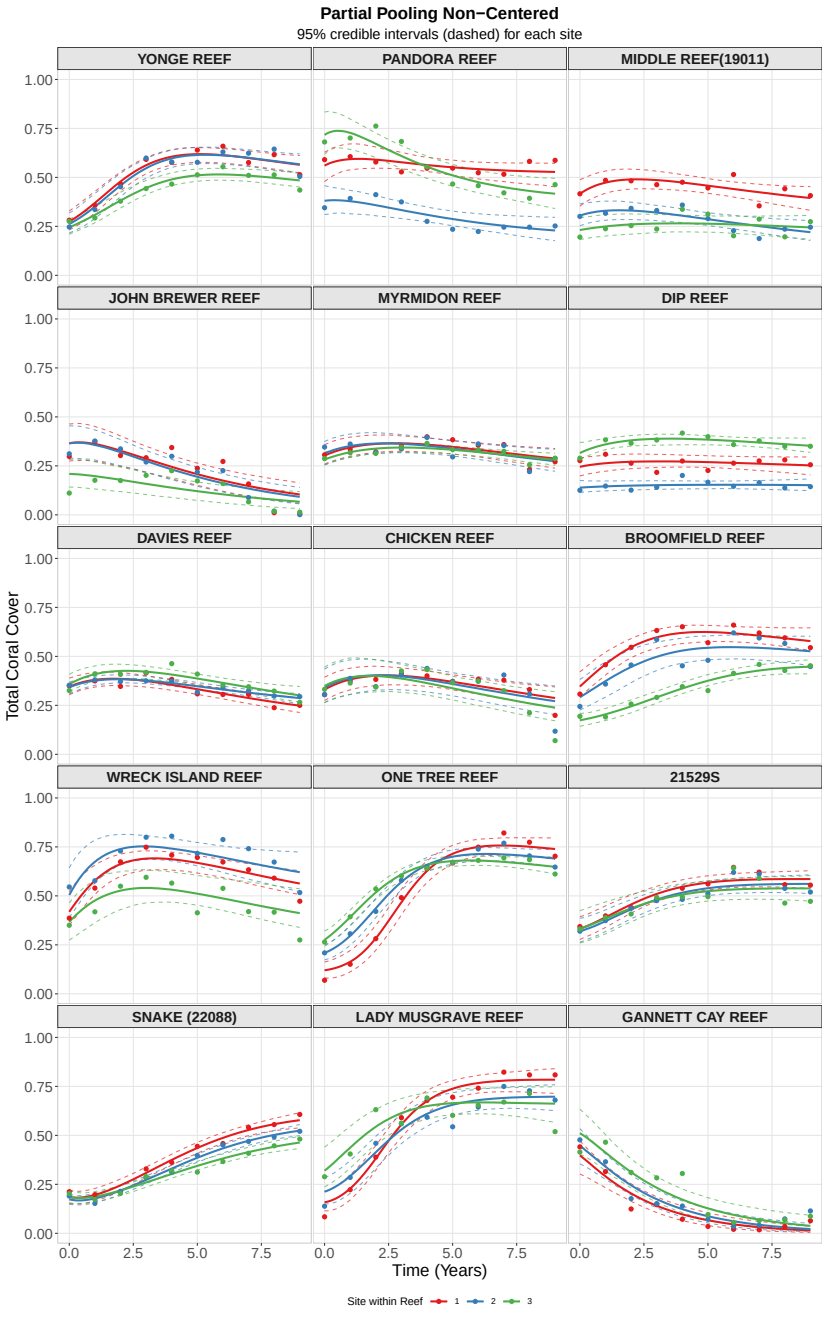


Figure 4.5. Posterior predictive checks for partial pooling non-centered approach. The partial pooling non-centered model shows robust performance in predicting coral cover trajectories across all reefs and sites with credible intervals effectively containing the observed data points.

Bayesian p-values

Quantitative analysis confirms what we see in the visual plots of PPC. To evaluate model performance systematically, we employ posterior predictive p-values, which represent the tail posterior probability for a test statistic generated by the model compared to the statistic observed in the data [70]. A posterior predictive p-value can be defined as

$$p\text{-value}(y) = \Pr(T(y_{\text{rep}}) > T(y) \mid y),$$

where $T(\cdot)$ is a test statistic or summary function, y is the observed data, and y_{rep} is replicated data generated from the model using posterior draws of the model parameters θ [28].

Importantly, when posterior uncertainty in θ is properly incorporated during the generation of y_{rep}, the p-value under a well-specified model is more likely to be near 0.5 than near 0 or 1 [30]. This reflects the fact that, for a good model, simulated data are expected to resemble the observed data, and thus the test statistic calculated from the replicated data is typically similar to that calculated from the observed data.

We calculate three metrics using p-values based on seven T statistics: the mean, minimum, maximum, standard deviation, 25th percentile, 75th percentile, and median, evaluated for each reef. According to Gelman (2003) [30], when uncertainty in θ is properly accounted for in the replicated data, the distribution of the p-value for a well-specified model is expected to be concentrated near 0.5. Based on this theoretical foundation, we develop the following metrics:

- **Average Distance from 0.5**: For each individual p-value calculated for a model, we find its absolute difference from the ideal value of 0.5 (i.e., $|p_{\text{value}} - 0.5|$). These individual differences are then averaged together for the model; lower values indicate better overall model calibration.
- **Percentage of Extreme p-values**: We check each individual p-value calculated for a model to see if it is less than 0.05 or greater than 0.95. We then calculate the percentage of all p-values for that model that fall into this extreme range.
- **Average p-value**: We compute the mean of all individual p-values for the model. Values closer to 0.5 demonstrate better overall model calibration.

Table 4.2 summarises the results of these metrics for each model. Looking at the average distance from 0.5, we see that partial pooling centereed method perform best (0.293), followed by complete pooling (0.306), with no pooling performing worst (0.323). For average p-values, partial pooling methods (0.509-0.517) are closest to the ideal 0.5, while complete pooling (0.481) and no pooling (0.553) deviate more. Most notably, the partial pooling methods have significantly fewer extreme p-values (14.3%) compared to complete pooling (42.9%) and no pooling (28.6%), demonstrating their superior stability and ability to capture the complex dynamic patterns across reefs.

Model	Avg Distance from 0.5	Extreme p-values (%)	Avg p-value
Partial pooling (centered)	**0.293**	**14.3**	**0.509**
Complete pooling	0.306	42.9	0.481
Partial pooling (non-centered)	0.307	**14.3**	0.517
No pooling	0.323	28.6	0.553

Table 4.2. *Posterior predictive p-value summary table comparing model performance across pooling methods. "Avg Distance from 0.5" and lower "Extreme p-values (%)" indicate better model fit. "Avg p-value" closer to 0.5 indicates better calibration. Results highlight the superior performance of partial pooling methods, with the centered approach showing the best overall calibration.*

Predictive Accuracy and Stability: Leave-One-Out Cross-Validation

From the posterior predictive checks, we observe similar performance between no pooling and partial pooling methods when applied to the complete dataset. To further differentiate model performance, we implement Leave-one-out cross-validation (LOO) using the Loo package for R [58, 73] with models fit using the Bayesian inference package Stan [70]. LOO estimates pointwise out-of-sample prediction accuracy using Pareto smoothed importance sampling (PSIS) [75], evaluating log-likelihood in posterior simulations without repeated model fitting. Since Stan works with joint density rather than separating likelihood components, our approach explicitly computes pointwise log-likelihood values $\log p(y_i|\theta^{(s)})$ for each observation i and posterior draw s as a vector. This explicit vectorisation allows the loo package to efficiently compute cross-validation metrics.

This approach provides several key indicators for model comparison: Pareto k, which assesses the reliability of the PSIS estimates used in LOO calculation [75], and pointwise expected log predictive density (ELPD) for predictive accuracy for each observation.

Figure 4.6 displays the Pareto k diagnostic values across all four models, with points colour-coded by stability level. When $k < 0.5$ (dark green points), the estimates are reliable; values between 0.5 and 0.7 (light green) indicate moderate stability; values between 0.7 and 1.0 (yellow) indicate influential observations; and values above 1.0 (red) are highly influential [73].

The complete pooling model exhibits the lowest percentage of problematic k values (2.7%, 12 observations with $k > 0.7$, including 2 with $k > 1.0$), but this stability comes at the cost of predictive accuracy due to excessive rigidity. The no pooling model shows the highest percentage (8.7%, 39 observations with $k > 0.7$, including 5 with $k > 1.0$), suggesting potential instability from independent parameter estimation with limited data per site. The partial pooling models demonstrate moderate percentages (centered: 5.6%, 25 observations with $k > 0.7$, including 4 with $k > 1.0$; non-centered: 7.6%, 34 observations with $k > 0.7$, including 4 with $k > 1.0$), which aligns with the-

oretical expectations for hierarchical models with small group sizes. Most observations across all models have k values below 0.5, indicating generally reliable estimates overall despite these influential points.

Despite the presence of some high Pareto k values, it is important to consider these diagnostics within the context of our complex Bayesian hierarchical model. As noted by one of the Stan developers, Vehtari [72], when $p_{\text{loo}} < p$ (the number of parameters) and p is relatively large compared to the number of observations ($p > n/5$), high k values may emerge even when the model is correctly specified. This occurs because the model's flexibility or weak population priors make it difficult to predict left-out observations using importance sampling techniques. We observe that for all our models, the effective number of parameters (p_{loo}) is lower than the actual number of parameters. With approximately 15 reefs and multiple parameters per reef in our hierarchical structure, the partial pooling models contain roughly 270 population-level, reef-level and site-specific parameters, which is substantial relative to our sample size. The presence of some high k values is expected and does not necessarily indicate model misspecification. Rather, it reflects the inherent complexity of modelling coral cover dynamics with reef-specific parameters and hierarchical structure.

Figure 4.6. Pareto k diagnostic values for all models. Points are colour-coded by stability level: stable ($k \leq 0.5$, dark green), moderately stable ($0.5 < k \leq 0.7$, light green), influential ($0.7 < k \leq 1.0$, yellow), and highly influential ($k > 1.0$, red). The red dashed line at 0.7 marks the threshold above which observations significantly influence model estimates. Complete pooling (top left) shows the fewest influential points, while no pooling (top right) shows the most. Both partial pooling approaches (bottom) show intermediate numbers of influential observations.

Figure 4.7 displays the pointwise ELPD values across models, where higher values indicate better predictive accuracy for individual data points. These values range from approximately -2 to 3, and sum to the total ELPD, which is one of our primary model selection criteria. The partial pooling models (both centered in green and non-centered in purple) show consistently higher ELPD values compared to no pooling (blue), with complete pooling (red) performing worst overall. Complete pooling shows greater variability, with a notable cluster of points dropping below 0 between data points 300-400. The partial pooling models maintain more points at higher ELPD values (above 1.5), particularly around data points 200-250 and 350-400, indicating superior predictive performance in these regions.

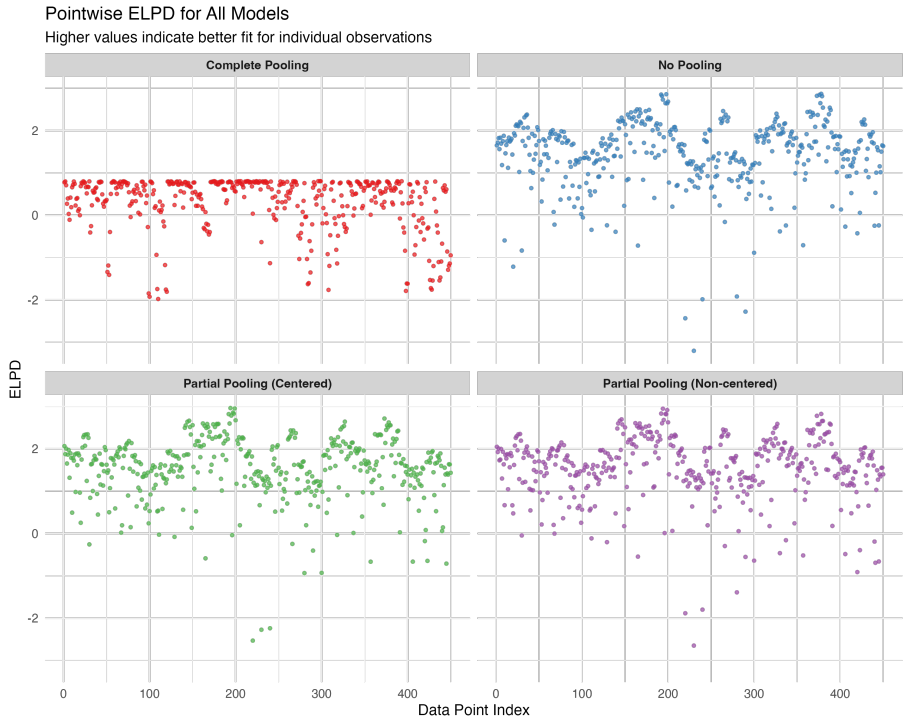


Figure 4.7. Pointwise ELPD for all models. Higher values indicate better predictive performance for individual data points. Complete pooling (red, top left) shows the poorest and most variable performance. No pooling (blue, top right) performs better but is outperformed by both partial pooling centered (green, bottom left) and non-centered (purple, bottom right) approaches, which show very similar patterns with consistently higher ELPD values. All plots share the same y-axis scale, with data points indexed by observation number.

Beyond these visual diagnostics, Table 4.3 provides a quantitative summary of model performance using metrics derived from PSIS-LOO. The table includes the estimated ELPD for each model, along with its associated Standard Error (SE). The SE quantifies the uncertainty in the ELPD calculation; smaller SE values suggest more reliable ELPD estimates. Evaluating the results, both partial pooling methods substantially outperform the other pooling strategies. The non-centered parameterisation achieves an ELPD of 694.14 (SE: 16.28), and the centered version reaches 689.14 (SE: 16.09). These values significantly exceed those of the no pooling (652.60, SE: 17.20) and complete pooling (134.93, SE: 13.78) approaches. The table also presents the p_{loo} values, which estimate the effective number of parameters in each model. Higher p_{loo} values indicate models with greater flexibility. The partial pooling and no pooling methods demonstrate substantial flexibility (127.05, 120.80, and 137.99, respectively) compared to the much simpler complete pooling model (3.33).

Model	ELPD	SE	p_{loo}	LOOIC
Partial pooling (non-centered)	**694.14**	16.28	127.05	**-1388.27**
Partial pooling (centered)	689.14	16.09	120.80	-1378.28
No pooling	652.60	17.20	**137.99**	-1305.20
Complete pooling	134.93	**13.78**	3.33	-269.85

Table 4.3. *Quantitative comparison of model performance using Leave-One-Out Cross-Validation metrics. ELPD (higher is better) measures predictive accuracy; SE (Standard Error) quantifies uncertainty in ELPD estimates; p_{loo} represents the effective number of parameters. The partial pooling models demonstrate substantially better predictive performance than either the no pooling or complete pooling alternatives, with nearly identical performance between centered and non-centered parameterisations.*

In conclusion, considering both predictive accuracy (ELPD) and diagnostic checks (p_{loo}, Pareto k), the partial pooling approaches represent the most reasonable compromise for the coral reef dataset. They provide the best predictive power whilst exhibiting a level of flexibility appropriate for the hierarchical nature of the data. The partial pooling (non-centered) model achieves the highest ELPD of 694.14, followed closely by the centered version at 689.14, both substantially outperforming the no pooling (652.60) and complete pooling (134.93) approaches. The moderate number of influential points identified by the Pareto k diagnostic is consistent with the theoretical understanding of hierarchical models with small group sizes (3 sites per reef) and reflects the models' ability to capture complex patterns in the data rather than fundamental misspecification. The partial pooling (non-centered) model, with slightly better ELPD (694.14 and 689.14), represents the best balance between flexibility and regularisation for this coral reef dynamics dataset, though both partial pooling approaches perform similarly well.

Evaluating MCMC Convergence and Sampling Quality

After checking model fit with PPC and LOO cross-validation, we now look at how well each model's sampling algorithm works. We use two key diagnostics: divergent transitions and R-hat values. Divergent transitions happen when the sampler encounters problems exploring the parameter space. Too many divergences mean we cannot trust the results. R-hat compares variation between chains to variation within chains. Values below 1.01 suggest the chains have converged properly [70, 73]. Table 4.4 shows that the complete pooling model performs best with zero divergent transitions and good R-hat (1.0035). The partial pooling model with non-centered parameterisation ranks second with only 11 divergences and excellent R-hat (1.0030). No-pooling has 12 divergences with the best R-hat (1.0004). The centered partial pooling model shows the most sampling problems with 517 divergences. While both partial pooling approaches performed nearly identically in our previous

tests, the significant difference in sampling quality (11 divergences versus 517) shows that non-centered parameterisation is computationally superior for hierarchical models. This demonstrates how parameterisation choices can have a significant impact on sampling reliability, even when predictive performance is similar.

Model	Divergent	Avg R-hat
Complete Pooling	0	1.0035
Partial Pooling (Non-Centered)	11	1.0030
No Pooling	12	1.0004
Partial Pooling (Centered)	**517**	1.0099

Table 4.4. *MCMC diagnostic metrics for the four coral reef models. Lower divergent transitions and R-hat values closer to 1.0 indicate better sampling reliability. All models were run with identical settings: 4 chains, 2000 iterations per chain (including 1000 for warmup), and consistent ODE solver tolerances.*

Comparing Centered and Non-centered Parameterisations for Partial Pooling

Our MCMC diagnostics revealed a striking contrast between centered and non-centered partial pooling models despite their similar predictive performance. To understand why the centered parameterisation produced 192 divergent transitions while the non-centered version had none, we need to examine the fundamental differences in how these models are structured.

As we mentioned in Chapter 2, our hierarchical models date structure parameters at multiple levels, with site-specific parameters nested under reef-specific parameters, which are nested under population-level parameters. This allows individual reefs to have their own unique parameter values while still sharing information through common population-level distributions [29]. However, how a hierarchical model is parameterised can have a significant impact on sampling efficiency.

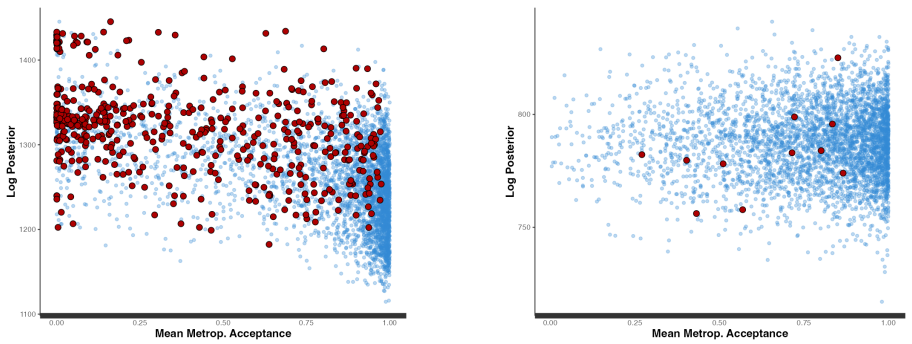
We initially used a centered parameterisation, where reef-specific parameters (θ_j, representing growth, transition, recovery, and mortality rates for reef j) were modelled as being directly drawn from a normal distribution with a reef-group mean (μ) and standard deviation (σ): $\theta_j \sim \text{Normal}(\mu, \sigma^2)$. Although conceptually simple, this approach can lead to sampling issues.

The centered method exhibits a characteristic curved "funnel" (see Figure 4.8a) shape in the posterior distribution [56]. This funnel arises because when the group-level standard deviation (σ) is small, the individual reef parameters (θ_j) are tightly clustered around the group mean (μ). This creates a region of high curvature and correlation in the posterior, which is difficult for HMC samplers, like the one used in Stan, to navigate. The leapfrog integrator, used by HMC to simulate Hamiltonian dynamics, can become unstable

in these regions, leading to divergent transitions (shown as red dots in the figures). Divergent transitions are a serious warning sign, indicating that the sampler is failing to accurately explore the posterior and that the results may be biased.

To overcome these sampling difficulties, we switched to a non-centered parameterisation [8]. Rather than directly sampling θ_j, we introduce standardised variables, $z_j \sim \text{Normal}(0,1)$, and transform them using: $\theta_j = \mu + \sigma z_j$. This approach separates the individual-reef variation (z_j) from the group-level location (μ) and scale (σ). This reparameterisation allows larger, more efficient movements in parameter space and reduces unwanted random walk behaviour. Figure 4.8b demonstrates the improvement. The funnel shape is largely gone, resulting in more reliable parameter estimates. We achieved greatly reduced divergent transitions (reducing from 85 to zero) and better convergence diagnostics. The $\hat{R}$ statistic, also known as the potential scale reduction factor, measures the ratio of the average variance of samples within each chain to the variance of the pooled samples across chains [1]. Our non-centered model produced $\hat{R}$ values very close to 1.0, indicating excellent convergence, alongside higher effective sample sizes. These computational improvements lead to more robust ecological inferences.

This geometric explanation of the posterior landscape clarifies why our diagnostic tests showed such different sampling behaviour between the two parameterisations despite their similar predictive performance. Both approaches model the same underlying relationships, but the non-centered version provides a computationally superior path to explore the parameter space.



(a) Centered method for hierarchical model

(b) Noncentered method for hierarchical model

Figure 4.8. Comparison of parameter estimation using partial pooling with centered (left) and non-centered (right) parameterisations. These results are based on a 10-year subset of coral cover data from the AIMS Long-Term Monitoring Program [4]. The diagnostic plots reveal a "funnel" shape in the posterior distribution for the centered approach (left), indicating sampling difficulties, while the non-centered approach (right) mitigates this issue, enabling more efficient and reliable sampling.

4.1.3 Handling Missing Data in Bayesian Hierarchical Models

By comparing model performance across different levels of data sparsity, we can determine whether the theoretical benefits of partial pooling, particularly its ability to share information across reefs, translate into practical benefits when dealing with incomplete observation records.

Implementation and Performance on Sparse Data

We use the method as we mentioned in Chapter 3 on how to use the Bayesian approach to deal with the incomplete dataset. We first implemented the indicator-based likelihood and tested its performance on a sparse dataset with approximately 50% of observations removed (5 reefs with original 3 sites, 5 reefs with 2 sites, and 5 reefs with 1 site).

Table 4.5 provides quantitative support for these findings through MCMC diagnostics. The partial pooling (non-centered) model achieves significantly fewer divergent transitions (1) compared to (9) for the no pooling model, indicating superior computational stability, particularly crucial for sparse data. This stability stems from the model's hierarchical structure, which allows sharing of information across reefs, regularising estimates even when individual reef data are limited. Both models show excellent convergence with average R-hat values very close to 1.0 (1.0037 vs 1.0003), indicating reliable parameter estimation.

Model	**Divergent (%)**	**Avg Rhat**
Partial pooling (non-centered)	**1 (0.05%)**	1.0037
No pooling	9 (0.45%)	**1.0003**

Table 4.5. *MCMC diagnostics comparison results. Lower divergent transitions and R-hat values closer to 1.0 indicate better model performance. The **partial pooling (non-centered)** model shows significantly fewer divergent transitions, demonstrating superior computational stability for sparse data scenarios.*

Further evidence of the partial pooling model's superiority is provided in Table 4.6, which shows the results of leave-one-out cross-validation. The partial pooling (non-centered) model achieves an ELPD of 444.51 compared to 405.0 for the no pooling model, with a substantial difference of 39.53 (SE: 19.93) strongly favouring the partial pooling approach. The LOOIC values, which represent the Leave-One-Out information criterion, also support this conclusion. The LOOIC is calculated as $-2 \times$ elpd (placing it on the deviance scale). The effective number of parameters, p_{loo}, is the difference between elpd and the non-cross-validated log posterior predictive density, quantifying how much more difficult it is to predict future data than the observed data [73]. In this case, the Partial Pooling (Non-centered) model's LOOIC of -889.02 (compared to -810.0 for the no-pooling model) further supports this conclusion. Both models have similar effective parameter counts (p_{loo} of 90.85 vs 90.89), indicating comparable model complexity.

Model	ELPD	SE	p_{loo}	LOOIC
Partial pooling (non-centered)	**444.51**	**13.62**	90.85	**-889.02**
No pooling	405.0	14.54	**90.89**	-810.0

Table 4.6. *Leave-One-Out Cross-Validation Results for reduced dataset models. Higher ELPD and lower LOOIC values indicate better predictive performance. Lower SE indicates more precise estimates. The **partial pooling (non-centered)** model demonstrates significantly better out-of-sample prediction accuracy compared to the no pooling approach with a 95% confidence interval for the ELPD difference of [0.47, 78.58].*

Overall, these findings confirm the theoretical advantage of partial pooling for handling limited data and support its use for analysing incomplete coral reef time series. The partial pooling model demonstrates better computational stability and superior predictive performance when dealing with sparse ecological data.

4.1.4 Enhancing Predictive Accuracy on Unseen Data Through Hierarchical Modelling

After establishing the advantages of partial pooling with non-centered parameterisation for model fit and handling sparse data, we are interested in assessing its predictive performance on future, unseen observations.

As a baseline, we conducted a time-based prediction test using our pretrained partial pooling (non-centered) and no pooling models. This approach tests genuine predictive skill, as the validation data from the years 2005–2007 were never used during model fitting, which was performed exclusively on data from the 1995–2004 period.

When examining predictive accuracy on the proportional scale (0–1), we consider two standard metrics [19]: Mean Absolute Error (MAE), which averages the absolute differences between predictions and observations, and Root Mean Squared Error (RMSE), which takes the square root of averaged squared differences, thereby giving greater weight to larger errors. As shown in Table 4.7, the initial test confirmed the partial pooling model makes better predictions overall, achieving an MAE of 0.125 and an RMSE of 0.169. This represents a 26% and 9% improvement over the no pooling approach, respectively.

Model	Validation Set (2005–2007) MAE	RMSE
Partial pooling (non-centered)	**0.125**	**0.169**
No pooling	0.168	0.185

Table 4.7. *Predictive accuracy comparison between No pooling and Partial pooling models. All models were trained on coral cover data from 1995–2004 and evaluated against future observations from 2005–2007. The Partial Pooling model has both the lowest MAE and RMSE, making it more accurate than the No-Pooling model, representing a 26% improvement in MAE and a 9% improvement in RMSE.*

While these baseline results are strong, we hypothesised that predictive accuracy could be further improved. In time-series analysis, it is often assumed that more recent data are more relevant for predicting the future than older data. The forgetting factor [17] is a technique used to formalise this intuition, allowing the model to balance the relative importance of past and new data through a control parameter, d. It systematically reduces the influence of older observations via exponential decay. The weight for each data point is calculated using the following formula:

$$\text{Weight} = d^{(t_{\max} - t)}$$

Here, d is the forgetting factor (a value between 0 and 1, in our case 0.9), $t_{\max}$ is the most recent timestamp in the training data (2004), and t is the timestamp of the specific data point being weighted.

This formula assigns the highest weight of 1 to the most recent data (where the exponent $t_{\max} - t$ is 0) and exponentially smaller weights to older data points. The `Years Ago` column in Table 4.8 shows the exponent used for each year's calculation, demonstrating this decay in influence over time.

Data Point Year (t)	Years Ago ($t_{\max} - t$)	Calculation	Resulting Weight	Influence
2004	0	0.9^0	1.000	100.0%
2003	1	0.9^1	0.900	90.0%
2002	2	0.9^2	0.810	81.0%
2001	3	0.9^3	0.729	72.9%
2000	4	0.9^4	0.656	65.6%
1999	5	0.9^5	0.590	59.0%
1998	6	0.9^6	0.531	53.1%
1997	7	0.9^7	0.478	47.8%
1996	8	0.9^8	0.430	43.0%
1995	9	0.9^9	0.387	38.7%

Table 4.8. *Weight calculations for training data using a forgetting factor of 0.9.*

After implementing this weighting scheme, we retrained both the partial pooling (non-centered) and no pooling models on the 1995–2004 data. Table 4.9 shows that while both models perform similarly on the training data, the partial pooling model with the forgetting factor makes significantly better predictions on the unseen validation data. It achieves both the lowest validation MAE (0.085) and RMSE (0.112), representing a 15% improvement in MAE and a 13% improvement in RMSE over the retrained no pooling approach. Furthermore, the refined model with the forgetting factor shows a 32% improvement in MAE (from 0.125 to 0.085) and a 33.7% improvement in RMSE (from 0.169 to 0.112) on the validation set when compared with the baseline pre-trained model without the forgetting factor (Table 4.7). These results confirm that combining hierarchical modelling with a forgetting factor further enhances predictive accuracy for this dataset.

	Training Set (1995–2004)		**Validation Set (2005–2007)**	
Model	**MAE**	**RMSE**	**MAE**	**RMSE**
Partial pooling	0.035	**0.046**	**0.085**	**0.112**
No pooling	**0.034**	0.046	0.100	0.128

Table 4.9. *Predictive accuracy comparison between the No-pooling and Partial-pooling models with forgetting factor across both training and validation periods. The table displays MAE and RMSE for the training set (1995–2004) and the validation set (2005–2007). Lower values, highlighted in bold green, indicate better performance. While the models perform similarly on the training data, the **Partial Pooling model** demonstrates superior predictive accuracy on the unseen validation data, making it the more robust and generalisable model. Specifically, it achieves a 15% improvement in MAE and a 13% improvement in RMSE on the validation set.*

The visualisations in Figure 4.9 and Figure 4.10 provide qualitative confirmation of these improved quantitative results, showing how the partial pooling model with the forgetting factor more faithfully captures the non-linear dynamics of coral cover across the validation period compared to the no pooling approach.

These findings demonstrate that partial pooling offers significant advantages over no pooling for ecological forecasting applications. The hierarchical structure allows the model to better capture the underlying dynamics by sharing information across reefs, which is especially valuable when predicting future states for conservation planning and management.

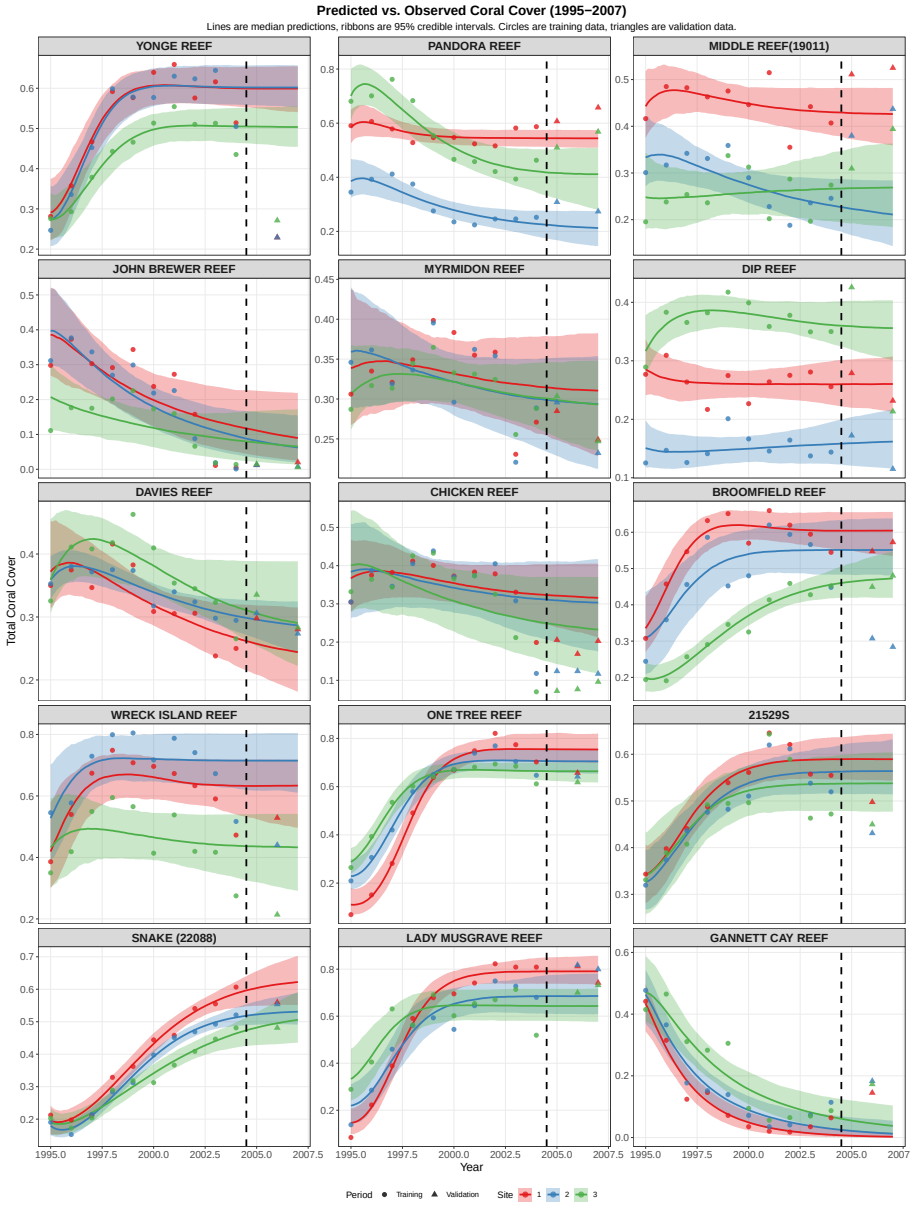


Figure 4.9. No pooling model predictions (solid lines for medians, shaded bands for 95% credible intervals) colored by site: red for Site 1, green for Site 2, and blue for Site 3. Actual observations are overlaid: circles for 1995–2004 training data, and triangles for 2005–2007 validation data. This model displays a more linear prediction curve compared to the partial pooling (non-centered) model.

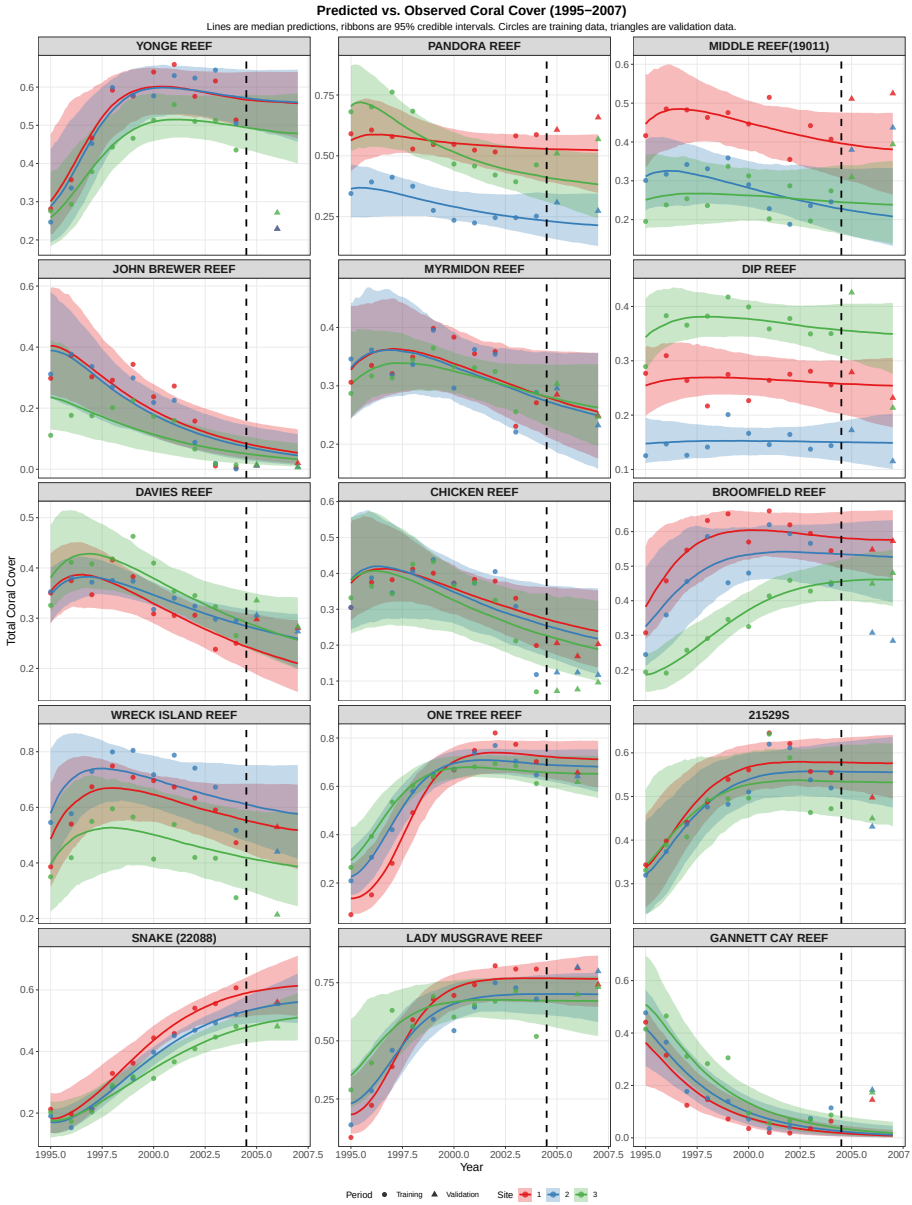


Figure 4.10. Partial pooling non centered prediction. This model exhibits a more non-linear prediction pattern and narrower credible intervals in most reefs compared to the no pooling model, particularly for John Brewer Reef and Myrmidon Reef.

4.1.5 Comparison with Non-Bayesian Approaches

After our comprehensive Bayesian model evaluation, we established that the partial pooling model with non-centered parameterisation provides the best

balance of predictive accuracy and computational stability. To further validate our approach, we now compare our hierarchical Bayesian model with traditional frequentist methods.

The frequentist approach used by Brown et al. [13] only captures dynamics at the reef level, which represents a significant limitation. In contrast, our hierarchical Bayesian model incorporates three levels of parameters: population-level, reef-specific, and site-specific. This multi-level structure allows our model to capture the ecological dynamics at a finer resolution, accounting for the considerable variation that exists not just between reefs but also between sites within the same reef.

To compare our approach with frequentist methods, we examine parameter estimates for One Tree Reef using both techniques. Figure 4.11 shows that both approaches produce similar median values for ODE parameters. However, the Bayesian model generates narrower credible intervals compared to the frequentist confidence intervals. The Bayesian approach with hierarchical structure demonstrates more precise uncertainty quantification. This precision provides clear parameter constraints for all four key parameters (α, β, γ, and μ) in the coral dynamics model.

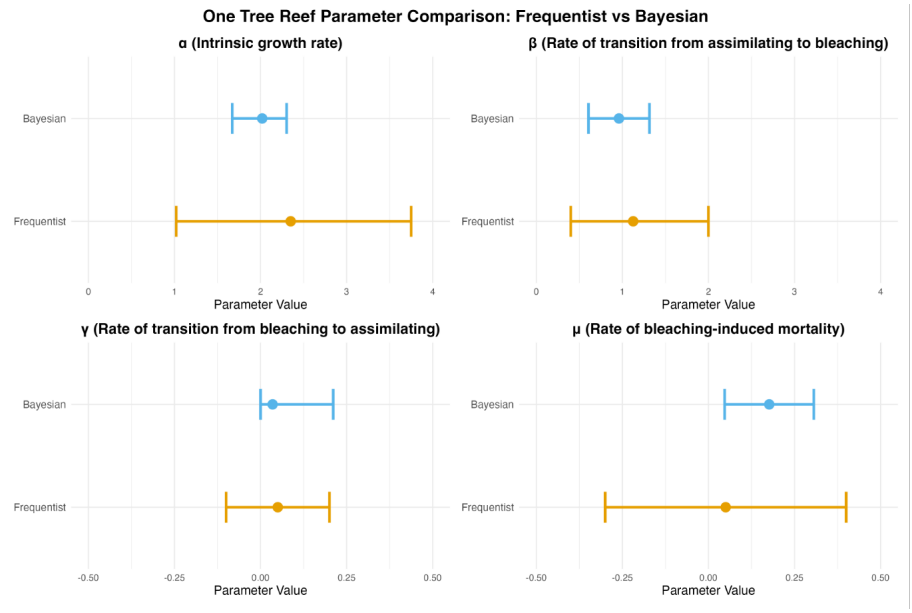


Figure 4.11. Parameter comparison between Bayesian (blue) and frequentist (orange) approaches for One Tree Reef. Each panel shows estimates for a key parameter in the coral dynamics model: α (intrinsic growth rate), β (rate of transition from assimilating to bleaching), γ (rate of transition from bleaching to assimilating), and μ (rate of bleaching-induced mortality). Points represent median/point estimates, while lines indicate 95% credible intervals (Bayesian) or confidence intervals (frequentist).

5. Conclusion and Future Work

5.1 Conclusion

This project introduces a novel three-layer Bayesian hierarchical model to describe coral reef dynamics, utilizing an ODE model and nested structured data. Our modeling approach enables enhanced understanding and predictions of coral reef dynamics, and is implemented using Stan, a probabilistic programming language that builds on HMC methods. Our model is validated through comprehensive pooling comparison experiments.

In doing so, this project offers several significant contributions to the field of Bayesian hierarchical modelling of coral reef dynamics:

1. **Three-Layer Hierarchical Bayesian Model**: We design and implement a novel three-layer hierarchical Bayesian model capable of capturing site-level dynamics with greater precision than traditional frequentist approaches, which typically only reach to the reef level. This model incorporates global-level, reef-level, site-specific level in a unified framework, allowing for better understanding of coral reef ecosystems and their complex dynamics.
2. **Prior Selection extensive analysis**: Our extensive analysis involves developing a robust prior selection analysis for hierarchical models that extends beyond the limited approaches described in official documentation for prior-posterior checks. Our methodology systematically evaluates prior influence on posterior distributions and implements sensitivity analyses to ensure model stability across a range of realistic ecological scenarios.
3. **Comprehensive Pooling Comparison Framework**: We design comprehensive experimental pipelines to evaluate different pooling strategies in Bayesian hierarchical models. Our systematic approach enables direct comparison of complete pooling, no pooling, and partial pooling methods with both centered and non-centered parameterisations.

Our assessment of model fit on complete data using posterior predictive checks reveals that both partial pooling and the no pooling model demonstrate better fit compared to the complete pooling approach, with the non-centered parameterisation showing particularly strong performance.

When evaluating computational stability and predictive accuracy on sparse data, the partial pooling model with non-centered parameterisation demonstrates clear advantages, while the no pooling model shows slightly better diagnostic metrics and faster runtime.

Our evaluation of predictive performance on sparse unseen future data demonstrates that hierarchical approaches offer substantial benefits for forecasting. The advanced three-layer model, which extends the partial pooling framework with an additional hyperprior layer, achieves the best MAE, representing significant improvements over both standard partial pooling and no pooling approaches. However, the standard partial pooling model exhibits the lowest RMSE, suggesting it may be more robust against occasional large prediction errors.

Taken together, our findings provide strong evidence that partial pooling with non-centered parameterisation offers the best overall performance for modelling coral reef dynamics. The three-layer hierarchical approach developed in this project shows particular promise for applications where data sparsity is a concern and accurate predictions at the site level are required.

To build upon these promising results and broaden the applicability of the developed framework, future research will focus on several key areas, as discussed in the next section.

5.2 Future Work

Future research will further explore alternative data processing techniques. In particular, data preprocessing will involve comparing different normalisation methods and approaches to outlier handling, with the aim of minimising bias, reducing noise in predictions, and extracting the maximum amount of information from ecological time series data. Furthermore, attention will be given to handling missing data. The research will systematically compare alternative Bayesian approaches, such as multiple imputation, to assess their impact on parameter inference within hierarchical models.

In addition to focusing on data handling, further efforts will also focus on the development of specialised tools. Specifically, future work will involve developing Stan packages that streamline the implementation of hierarchical models for complex ecological dynamics, along with providing templates and worked examples for different pooling approaches. The overall goal is to make Bayesian hierarchical modelling more accessible to researchers who may not have extensive experience in statistical programming. Additionally, we want our Bayesian hierarchical inference to be transferable, potentially helping us better understand other complex dynamical systems beyond ecology.

References

[1] Mohammed H AbuJarad and Athar Ali Khan. Exponential model: A bayesian study with stan. *International Journal of Recent Scientific Research*, 9(8):28495–28506, 2018.

[2] M Allenby Greg, E Rossi Peter, and E McCulloch Robert. Hierarchical bayes models: A practitioners guide. *SSRN Electronic Journal*, 2005.

[3] Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I Jordan. An introduction to mcmc for machine learning. *Machine learning*, 50:5–43, 2003.

[4] Australian Institute of Marine Science (AIMS). Aims long-term monitoring program: Video and photo transects (great barrier reef), 2015. Accessed 13-Feb-2025.

[5] Andrew H. Baird, Ranjeet Bhagooli, Peter J. Ralph, and Shunichi Takahashi. Coral bleaching: The role of the host. *Trends in Ecology & Evolution*, 24(1):16–20, 2009.

[6] Mark E. Baird, Mathieu Mongin, Farhan Rizwi, Line K. Bay, Neal E. Cantin, Monika Soja-Woźniak, and Jennifer Skerratt. A mechanistic model of coral bleaching due to temperature-mediated light-driven reactive oxygen build-up in zooxanthellae. *Ecological Modelling*, 386:20–37, 2018.

[7] David Barber and Yali Wang. Gaussian processes for bayesian estimation in ordinary differential equations. In *International conference on machine learning*, pages 1485–1493. PMLR, 2014.

[8] Michael Betancourt. Diagnosing suboptimal cotangent disintegrations in hamiltonian monte carlo. *arXiv preprint arXiv:1604.00695*, 2016.

[9] Julie C. Blackwood, Connie Okasaki, Andre Archer, Eliza W. Matt, Elizabeth Sherman, and Kathryn Montovan. Modeling alternative stable states in caribbean coral reefs. *Natural Resource Modeling*, 31(1):e12157, 2018.

[10] Julien Bodelet, Cecilia Potente, Guillaume Blanc, Justin Chumbley, Hira Imeri, Scott Hofer, Kathleen Mullan Harris, Graciela Muniz-Terrera, and Michael Shanahan. A bayesian functional approach to test models of life course epidemiology over continuous time. *International Journal of Epidemiology*, 53(1):dyad190, 01 2024.

[11] Nathaphon Boonnam, Tanatpong Udomchaipitak, Supattra Puttinaovarat, Thanapong Chaichana, Veera Boonjing, and Jirapond Muangprathub. Coral reef bleaching under climate change: Prediction modeling and machine learning. *Sustainability*, 14:6161, 05 2022.

[12] George EP Box. Sampling and bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 143(4):383–404, 1980.

[13] A. Brown, B. Smith, and C. Doe. Coral bleaching dynamics on the great barrier reef: New insights from a mathematical model. *Journal of Marine Science*, 58(4):123–145, 2024.

[14] Barbara E Brown and John C Ogden. Coral bleaching. *Scientific American*, 268(1):64–70, 1993.
[15] Paul-Christian Bürkner. brms: An r package for bayesian multilevel models using stan. *Journal of statistical software*, 80:1–28, 2017.
[16] Neal Cantin, Nicholas James, and Jessica Stella. Aerial surveys of the 2024 mass coral bleaching event on the great barrier reef. *Australian Institute of Marine Science, Townsville*, pages 2024–04, 2024.
[17] Weipeng Cao, Zhong Ming, Zhiwu Xu, Jiyong Zhang, and Qiang Wang. Online sequential extreme learning machine with dynamic forgetting factor. *IEEE access*, 7:179746–179757, 2019.
[18] Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1):1–32, 2017.
[19] Tianfeng Chai and Roland R Draxler. Root mean square error (rmse) or mean absolute error (mae)?–arguments against avoiding rmse in the literature. *Geoscientific model development*, 7(3):1247–1250, 2014.
[20] Petros Dellaportas and Gareth O Roberts. An introduction to mcmc. In *Spatial statistics and computational methods*, pages 1–41. Springer, 2003.
[21] Simon Donner, William Skirving, Christopher Little, Michael Oppenheimer, and Ove Hoegh-Guldberg. Global assessment of coral bleaching and required rates of adaptation under climate change. *Global Change Biology*, 11:2251–2265, 12 2005.
[22] A.E. Douglas. Coral bleaching—how and why? *Marine Pollution Bulletin*, 46(4):385–392, 2003.
[23] Martin Dyer, Alan Frieze, and Ravi Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *Journal of the ACM (JACM)*, 38(1):1–17, 1991.
[24] C Mark Eakin, Jessica A Morgan, Scott F Heron, Tyler B Smith, Gang Liu, Lorenzo Alvarez-Filip, Bart Baca, Erich Bartels, Carolina Bastidas, Claude Bouchon, et al. Caribbean corals in crisis: Record thermal stress, bleaching, and mortality in 2005. *PLoS One*, 5(11):e13969, 2010.
[25] HA El-Naggar, EE El-Gayar, ENE Mohamed, and MH Mona. Intertidal macro-benthos diversity and their relation with tourism activities at blue hole diving site, dahab, south sinai, egypt. *Sylwan*, 161(11):227–251, 2017.
[26] Hussein A El-Naggar. Human impacts on coral reef ecosystem. In *Natural resources management and biological sciences*. IntechOpen, 2020.
[27] Jonah Gabry, Daniel Simpson, Aki Vehtari, Michael Betancourt, and Andrew Gelman. Visualization in bayesian workflow. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 182(2):389–402, 2019.
[28] Andrew Gelman. Comment: fuzzy and bayesian p-values and u-values. 2005.
[29] Andrew Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3):515–534, 2006.
[30] Andrew Gelman. Two simple examples for understanding posterior p-values whose distributions are far from uniform. 2013.
[31] Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, London, 1995.

[32] Andrew Gelman, Xiao-Li Meng, and Hal Stern. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica sinica*, pages 733–760, 1996.
[33] Andrew Gelman and Iain Pardoe. Bayesian measures of explained variance and pooling in multilevel (hierarchical) models. *Technometrics*, 48(2):241–251, 2006.
[34] Andrew Gelman and Cosma Rohilla Shalizi. Philosophy and the practice of bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, 66(1):8–38, 2013.
[35] Mark Girolami. Bayesian inference for differential equations. *Theoretical Computer Science*, 408(1):4–16, 2008.
[36] Peter W Glynn. Extensive 'bleaching'and death of reef corals on the pacific coast of panama. *Environmental Conservation*, 10(2):149–154, 1983.
[37] Ghassan B Hamra, Richard F MacLehose, and Stephen R Cole. Sensitivity analyses for sparse-data problems—using weakly informative bayesian priors. *Epidemiology*, 24(2):233–239, 2013.
[38] Alastair R Harborne, Alice Rogers, Yves-Marie Bozec, and Peter J Mumby. Multiple stressors and the functioning of coral reefs. *Annual Review of Marine Science*, 9(1):445–468, 2017.
[39] Alan Hastings and Helen Edwards. Thresholds and the resilience of caribbean coral reefs. *Nature*, 450:98–101, 12 2007.
[40] Markus Heinonen, Cagatay Yildiz, Henrik Mannerström, Jukka Intosalmi, and Harri Lähdesmäki. Learning unknown ode models with gaussian processes. In *International conference on machine learning*, pages 1959–1968. PMLR, 2018.
[41] Ove Hoegh-Guldberg, Emma V. Kennedy, Hawthorne L. Beyer, Caleb McClennen, and Hugh P. Possingham. Securing a long-term future for coral reefs. *Trends in Ecology & Evolution*, 33(12):936–944, 2018.
[42] Matthew D Hoffman and Andrew Gelman. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(1):1593–1623, 2014.
[43] Terry P Hughes, Andrew H Baird, David R Bellwood, Margaret Card, Sean R Connolly, Carl Folke, Richard Grosberg, Ove Hoegh-Guldberg, Jeremy BC Jackson, Janice Kleypas, et al. Climate change, human impacts, and the resilience of coral reefs. *science*, 301(5635):929–933, 2003.
[44] John B Ingraham and Debora S Marks. Bayesian sparsity for intractable distributions. *arXiv preprint arXiv:1602.03807*, 2016.
[45] Mark Jerrum and Alistair Sinclair. The markov chain monte carlo method: an approach to approximate counting and integration. *Approximation algorithms for NP-hard problems*, pages 482–520, 1996.
[46] Christos Karras, Aristeidis Karras, Markos Avlonitis, and Spyros Sioutas. An overview of mcmc methods: From theory to applications. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pages 319–332. Springer, 2022.
[47] John K Kruschke and Wolf Vanpaemel. 13 bayesian estimation in hierarchical models. *The Oxford handbook of computational and mathematical psychology*, page 279, 2015.
[48] Liam Lachs, John Bythell, Holly East, Alasdair Edwards, William Skirving,

Blake Spady, and James Guest. Fine-tuning heat stress algorithms to optimise global predictions of mass coral bleaching. *Remote Sensing*, 13:2677, 07 2021.
[49] Ben Lambert. *A Student's Guide to Bayesian Statistics*. SAGE Publications Ltd, London, UK, 2018.
[50] Simon A Levin, Bryan Grenfell, Alan Hastings, and Alan S Perelson. Mathematical and computational challenges in population biology and ecosystems science. *Science*, 275(5298):334–343, 1997.
[51] Roderick JA Little and Donald B Rubin. *Statistical Analysis with Missing Data*. John Wiley & Sons, 3 edition, 2019.
[52] Henry Liu. Ordinary differential equations. Course Materials, Math 3027, Summer 2019. Lecture notes.
[53] J. Maynard, Kenneth Anthony, Paul Marshall, and I. Masiri. Major bleaching events can lead to increased thermal tolerance in corals. *Marine Biology*, 155:173–182, 08 2008.
[54] Hongyu Miao, Xiaohua Xia, Alan S Perelson, and Hulin Wu. On identifiability of nonlinear ode models and applications in viral dynamics. *SIAM review*, 53(1):3–39, 2011.
[55] Carl N Morris. Parametric empirical bayes inference: Theory and applications. *Journal of the American Statistical Association*, 78(381):47–55, 1983.
[56] Radford M Neal. Mcmc using hamiltonian dynamics. In Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng, editors, *Handbook of Markov Chain Monte Carlo*, chapter 5, pages 113–162. CRC Press, 2011.
[57] Omiros Papaspiliopoulos, Gareth O Roberts, and Martin Sköld. Non-centered parameterisations for hierarchical models and data augmentation. pages 307–326, 2003.
[58] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2024.
[59] Borja G. Reguero, Michael W. Beck, Vera N. Agostini, Philip Kramer, and Boze Hancock. Coral reefs for coastal protection: A new methodological approach and engineering case study in grenada. *Journal of Environmental Management*, 210:146–161, 2018.
[60] Bernhard M. Riegl and Samuel J. Purkis. Model of coral population response to accelerated bleaching and mass mortality in a changing climate. *Ecological Modelling*, 220(2):192–208, 2009.
[61] Christian P Robert and George Casella. The metropolis—hastings algorithm. In *Monte Carlo Statistical Methods*, pages 267–320. Springer, 2004.
[62] Christian P Robert et al. *The Bayesian choice: from decision-theoretic foundations to computational implementation*, volume 2. Springer, 2007.
[63] Małgorzata Roos, Thiago G. Martins, Leonhard Held, and Håvard Rue. Sensitivity analysis for bayesian hierarchical models. *arXiv: Methodology*, 2013.
[64] Donald B Rubin. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, pages 1151–1172, 1984.
[65] Yajuan Si, Natesh S Pillai, and Andrew Gelman. Bayesian nonparametric weighted sampling inference. 2015.
[66] Amy Sing Wong, Spyridon Vrontos, and Michelle L Taylor. An assessment of people living by coral reefs over space and time. *Global Change Biology*,

28(23):7139–7153, 2022.
[67] Sanne C Smid and Sonja D Winter. Dangers of the defaults: A tutorial on the impact of default priors when using bayesian sem with small samples. *Frontiers in Psychology*, 11:611963, 2020.
[68] David Souter, Serge Planes, Jérémy Wicquart, Murray Logan, David Obura, and Francis Staub. Status of coral reefs of the world: 2020, 2021.
[69] Stan Development Team. Prior choice recommendations. Wiki page, 2023. Accessed: 2025-05-23.
[70] Stan Development Team. *Stan Reference Manual*, 2024.
[71] Robert van Woesik and Terry J Done. Coral communities and reef growth in the southern great barrier reef. *Coral Reefs*, 16:103–115, 1997.
[72] Aki Vehtari. A quick note: What i infer from p_loo and pareto k values. Stan Discourse, 2018. Accessed: 2025-05-23.
[73] Aki Vehtari, Andrew Gelman, and Jonah Gabry. Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and computing*, 27:1413–1432, 2017.
[74] Aki Vehtari and Jouko Lampinen. Bayesian model assessment and comparison using cross-validation predictive densities. *Neural computation*, 14(10):2439–2468, 2002.
[75] Aki Vehtari, Daniel Simpson, Andrew Gelman, Yuling Yao, and Jonah Gabry. Pareto smoothed importance sampling. *Journal of Machine Learning Research*, 25(72):1–58, 2024.
[76] Eric-Jan Wagenmakers, Michael Lee, Tom Lodewyckx, and Geoffrey J Iverson. Bayesian versus frequentist inference. *Bayesian evaluation of informative hypotheses*, pages 181–207, 2008.
[77] Christopher K Wikle. Hierarchical bayesian models for predicting the spread of ecological processes. *Ecology*, 84(6):1382–1394, 2003.
[78] Christopher K Wikle, L Mark Berliner, and Noel Cressie. Hierarchical bayesian space-time models. *Environmental and ecological statistics*, 5:117–154, 1998.
[79] Clive Wilkinson and Jon Brodie. Catchment management and coral reef conservation. 2011.
[80] Samuel Wong, Shihao Yang, and Supeng Kou. Estimating and assessing differential equation models with time-course data. *arXiv preprint*, 12 2022.
[81] Yoshitaka Yano, Stuart L Beal, and Lewis B Sheiner. Evaluating pharmacokinetic/pharmacodynamic models using the posterior predictive check. *Journal of pharmacokinetics and pharmacodynamics*, 28:171–192, 2001.
[82] B Zheng. Ordinary differential equation and its application. *Highlights in Science, Engineering and Technology*, 72:645–651, 2023.