Adaptive monitoring of coral health at Scott Reef where data exhibit nonlinear and disturbed trends over time

Thilan AWLP[1,2,3,9], Fisher R[4,5,6], Thompson H[1,2,3], Menendez P[7,8], Gilmour J[4,6], and McGree JM[1,2,3,*]

[1]School of Mathematical Sciences, Faculty of Science, Queensland University of Technology (QUT), Australia

[2]Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS), Australia

[3]Centre for Data Science, QUT, Australia

[4]Australian Institute of Marine Science, Crawley, Australia

[5]Oceans Institute, University of Western Australia, Crawley, Australia

[6]Western Australian Marine Science Institution, Perth, Australia

[7]Department of Econometric and Business Statistics, Monash University, Australia

[8]Australian Institute of Marine Sciences, Townsville, Australia

[9]Department of Mathematics, University of Ruhuna, Sri Lanka

[*]Corresponding author:- Postal address: School of Mathematical Sciences, Faculty of Science, Queensland University of Technology, 2 George Street, Brisbane, QLD 4000; Tel: +61(0)410372540; Fax: +61 731382310; Email: james.mcgree@qut.edu.au

Abstract

Time series data are often observed in ecological monitoring. Frequently such data exhibit nonlinear trends over time potentially due to complex relationships between observed and auxiliary variables, and there may also be sudden declines over time due to major disturbances. This poses substantial challenges for modelling such data and also for model-based adaptive monitoring. We propose novel methods for finding adaptive designs for monitoring when historical data show such nonlinear patterns and sudden declines over time. This work is motivated by a coral reef monitoring program that has been established at Scott Reef; a coral reef off the Western coast of Australia.

Data collected for monitoring the health of Scott Reef are considered, and semiparametric and interrupted time series modelling approaches are adopted to describe how these data vary over time. New methods are then proposed that enable adaptive monitoring designs to be found based on such modelling approaches. These methods are then applied to find future monitoring designs at Scott Reef and form a set of recommendations for future monitoring.

Through applying the proposed methods, it was found that future information gain is expected to be similar across a variety of different sites, suggesting that no particular location needed to be prioritised at Scott Reef for the next monitoring phase. In addition, it was found that omitting some sampling sites/reef locations was possible without substantial loss in expected information gain, depending upon the disturbances that were observed.

The resulting adaptive designs are used to provide recommendations for future monitoring in this region, and for reefs where changes to the current monitoring practices are being sought. Furthermore, as the methods used and developed throughout this study are generic in nature, this research has the potential to improve ecological monitoring more broadly where complex data are being collected over time.

Keywords: Ecological monitoring, interrupted time series regression, mass bleaching events, semiparametric regression, sudden declines in trends.

1. Introduction

Coral reefs are one of the most beautiful and biologically diverse ecosystems globally. Unfortunately, environmental stressors such as severe cyclones and bleaching exposures have had a negative impact on coral reefs (Gilmour et al., 2019). As a result, the health of coral reefs are continually being monitored to estimate the impact of such disturbances and to identify additional vulnerabilities to decline.

In long-term coral reef monitoring, experimental design plays a vital role in creating survey designs to collect data for assessing coral health, trends over space and time, and to identify vulnerabilities of coral communities to different disturbances (Campbell et al., 2001). Broadly, there are two types of designs; static and adaptive. Static designs are those that do not change over time (e.g. the same sites/reefs are visited each year), and have been commonly used within monitoring programs. In contrast, adaptive designs can vary over time based on, for example, information from new data, and such methods have been proposed recently for determining when and where to sample within a coral reef to learn about coral health (Kang et al., 2016).

In the context of adaptive design, the adaptation is primarily informed by a statistical model. The purpose of this model is to extract information contained within the historical data to quantify uncertainties about, for example, the model itself, the model parameter values, and the response variable of interest, and then utilise this information to guide future surveys. For example, in Thilan et al. (2021), a spatial Beta regression model was developed for coral cover, and used to find future adaptive designs. When such designs were compared to those based on a linear model, the importance of appropriately capturing trends and variability within historical data was highlighted as this led to more informative and therefore efficient designs.

Ecosystems are subjected to a variety of observed and unobserved impacts which may interact in a variety of different ways (Newbold et al., 2020). For instance, coral reef ecosystems often exhibit nonlinear trends including sudden shifts due to mass coral bleaching, severe storms, and CoTS outbreaks (Done, 1992, McCook, 1999). These nonlinear trends poses significant challenges in modelling ecological data (Oddi et al., 2019), and this challenge is further exacerbated when there are sudden shifts in the overall trend due to disturbances (Scheffer et al., 2001).

Generally, semiparametric regression modelling approaches provide more flexibility than parametric models in describing a variety of relationships between (a function of) the mean response and given covariates (Crainiceanu et al., 2005). Thus, the development and use of semiparametric regression modelling approaches has received attention recently for modelling ecological data (Vercelloni et al., 2014, 2017). However, currently no methods exist for finding adaptive designs based on such models which limits how such information can be used to guide future reef monitoring. In addition, to account for sudden or sharp declines in the mean response due to disturbances such as a mass bleaching event, approaches from time-series regression modelling can be considered. Within a monitoring program, of further interest is then how the coral reef should be sampled to estimate the impact of such a disturbance.

In this paper, we propose new methods to find adaptive designs when the historical data exhibit nonlinear trends and sudden declines over time. The motivation for this research is the improvement of the Scott Reef Research Program (SRRP); a monitoring program of a coral reef system off the Western coast of Australia. We leverage information from the historical data though semiparametric and time series modelling approaches. Methods for finding adaptive designs based on such a modelling approach are then proposed, and designs are found under future monitoring scenarios at Scott Reef. These designs are then evaluated and used to provide recommendations for future surveys at Scott Reef and other reef monitoring programs where changes to the sampling practices are being contemplated.

2. Motivating example

Scott Reef is located 270 kilometres off the current coast of North-Western Australia (Gilmour and Smith, 2013) (Figure 1 (a)) and accordingly is isolated from many human impacts. However, these reefs are frequently exposed to cyclones and bleaching events. For example, due to elevated water temperatures over a few months in 1998, Scott Reef experienced a mass bleaching event, resulting in a decline of coral up to 80% (Gilmour et al., 2019, Gilmour and Smith, 2013) and thus, a complete change in coral cover trends was observed over time (Figure A.1). Furthermore, such disturbance exposure does not seem homogeneous across different survey locations. That is, there were survivors or relatively unharmed, moderately, and severely affected reef locations after this severe disturbance event (Figure A.2) (Gilmour and Smith, 2013). By adequately identifying the impacts of sudden disturbances, variations across the reef, and potential causes, it should be possible to develop efficient and appropriate monitoring practices that can change/evolve over time, and this is the aim of this paper.

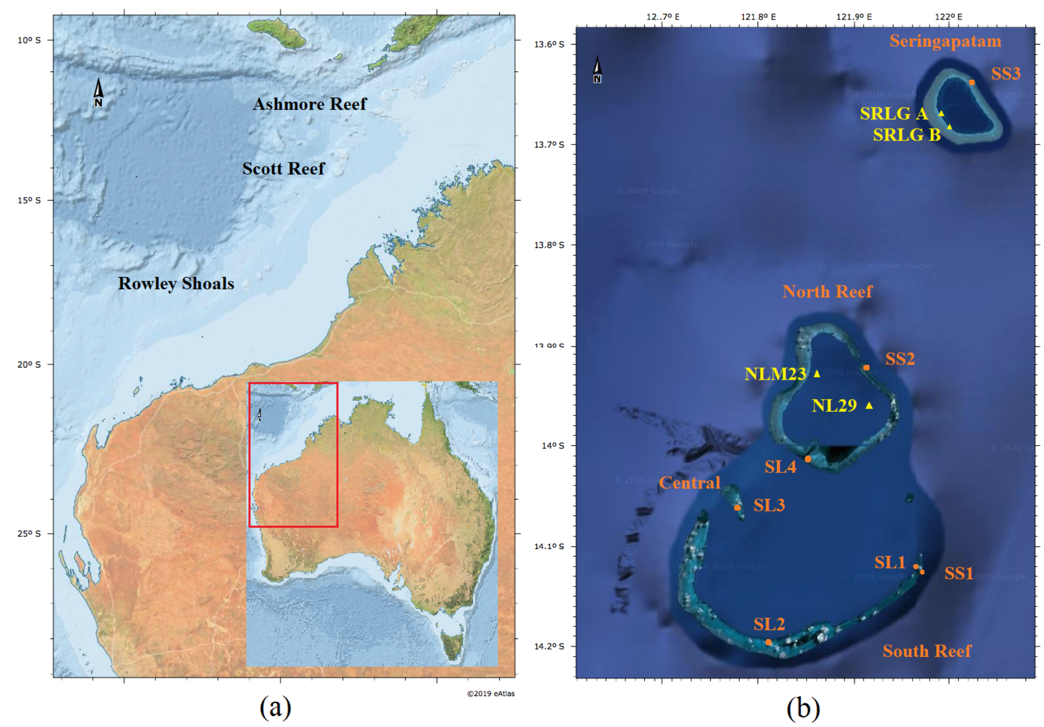

Figure 1: (a) The location of the system of Scott Reef and (b) the long-term monitoring sites located at South Reef, Central, North Reef, and Seringapatam (Google, n.d.). The orange points represent sites that have been surveyed since 1994 and the yellow triangles represent newly added sites after the 2016 bleaching event. "Bright Earth eAtlas basemap v1.0 (AIMS)".

3. Data

The system of Scott Reef comprises of four separate structures, namely North Reef, Central, South Reef, and Seringapatam (Figure 1 (b)). Under the SRRP, data have been collected over six habitats called slope, upper slope, crest, flat, lagoon, and outcrop from 1994.83 to 2017.92, where decimals represent survey times within a given year, i.e. .83 denotes the 10th of 12 months. Three core sites have been sampled to collect data, which are nested within each of seven reef locations (i.e. SL1, SL2, SL3, SL4, SS1, SS2, SS3) (Figure 1 (b)). In this study, data collected between 1994.83 and 2016.08 were considered in accordance with future monitoring objectives at Scott Reef (Section 4).

SRRP surveys are typically conducted in October but variations have been observed from year to year (Table A.1). For instance, when there was a severe disturbance, AIMS have collected data during, immediately after, and then later in the year depending upon the nature of the disturbance. In 1998, they conducted such pre- and post-bleaching surveys in January and October, respectively, where there was interest in quantifying coral loss during this time. When collecting disturbance data, AIMS has recorded bleaching exposure (i.e. $0 =$ No coral bleaching, $1 \geq 1\%$ coral bleached) and cyclone exposure in terms of the number of hours of reef system exposure to damaging waves (Puotinen et al., 2016). Additionally, reef location-specific cyclone and bleaching impacts have been documented.

4. Monitoring objectives

This study aims to develop recommendations for future monitoring at Scott Reef and other reef monitoring programs where changes are being considered. We hope to achieve this goal through considering the following two questions which form the basis for our two objectives:

(i) Are some reef locations (i.e. SL1, $\cdots$, SS3) within Scott Reef more important than others in providing information on coral health?

(ii) Which site at each reef location provides the most information about coral health?

5. Design framework

5.1. Modelling historical data

Capturing nonlinear trends

Semiparametric regression approaches can be used to capture nonlinear relationships within a regression model, and have been considered previously to describe data from coral communities on the Great Barrier Reef (GBR) (Vercelloni et al., 2014). To outline our adopted modelling approach, assume that $y_{srt} \sim \text{BIN}(n, p_{srt})$ where $n = 1250$ is the number of points in a 250m combined length of transects, $p_{srt} = 1/(1 + \exp(-\mu_{srt}))$, and $y_{srt}\,(s = 1, 2, 3;\ r = 1, \cdots, 7;\ t = 1994.83, \cdots, 2016.08)$ denotes the hard coral cover at the s-th site, in the r-th reef location at t-th survey time.

Then, a semiparametric regression modelling approach can be used to model the mean μ_{srt} as $\mu_{srt} = f(x_{srt})$, where x_{srt} are the survey times where data have been collected from sites that are nested within reef locations, and f is a smooth function.

There are different methods for modelling the smooth function, including cubic splines, B-splines, truncated polynomials, radial splines etc (Crainiceanu et al., 2005). We consider the low-rank thin-plate splines approach as it requires fewer parameters to estimate, and also it is insensitive to the choice of knots (Wood, 2003). Accordingly, the smooth function can be expressed as follows:

$$f(x_{srt}, \boldsymbol{\theta}_0) = \beta_0 + \beta_1 x_{srt} + \sum_{k=1}^{K} \delta_k |x_{srt} - \eta_k|^3, \quad (1)$$

where $\boldsymbol{\theta}_0 = (\beta_0, \beta_1, \delta_1, \cdots, \delta_K)^T$, β_0 is the intercept, β_1 is the regression coefficient for time, $\boldsymbol{\delta} = (\delta_1, \cdots, \delta_K)$ are random coefficients, η_k are knots, and K is the total number of knots. The values of $|x_{srt} - \eta_k|$ are calculated based on the sample quantile of x_{srt}'s (Crainiceanu et al., 2005).

Extensions to capture hierarchical structures within the data are straightforward via the inclusion of random effects, and such models can be defined by considering random effects for sites nested within reef locations as follows:

$$\log\left(p_{srt}/(1-p_{srt})\right) = \gamma_{sr} + \boldsymbol{\beta}_t \boldsymbol{z}_t + \boldsymbol{\beta}_d \boldsymbol{d}_r + f(x_{srt}, \boldsymbol{\theta}_0), \tag{2}$$

where γ_{sr} represents the corresponding nested random effect that follow a distribution $p(\gamma_{sr}|\lambda_r, \log\sigma_s)$ where $\lambda_r \sim p(\lambda_r|\log\sigma_r)$, where $\log\sigma_s$ and $\log\sigma_r$ are the corresponding site and reef random effects standard deviations in log scale, respectively. Wood (2003) describes the extension of such a model to accommodate the other potential covariates, and we follow this approach to incorporate bleaching exposure and cyclone hours. These covariates vary over time for the whole reef system, and thus are hereafter referred to as time-varying covariates. In Equation (2), $\boldsymbol{z}_t$ represents time-varying covariates and $\boldsymbol{\beta}_t$ is the corresponding vector of regression coefficients. Additionally, we incorporated three dummy variables to account for cyclone, severe cyclone, and bleaching exposures at different reef locations. The corresponding data matrix and the vector of regression coefficients are denoted as $\boldsymbol{d}_r$ and $\boldsymbol{\beta}_d$, respectively.

Accounting sudden declines in trends

Coral cover is often impacted by disturbances such as cyclones and bleaching events, and some major events will result in sudden declines in coral cover trends (De'ath et al., 2012, Osborne et al., 2011). We propose that estimating the impact of such major events can be achieved using an interrupted time series (ITS) regression approach (Bernal et al., 2017, Linden, 2015). In general, an ITS approach can account for sudden changes in the trend due to some intervention introduced or disturbance that has occurred (McDowall et al., 2019). When applying ITS, the type of impact due to the disturbance should be hypothesised. This may include a gradual change in slope or in both the intercept and slope within the model for the mean response (Bernal et al., 2017). In addition, some disturbances may cause an immediate change in the trend, but others may have a lag period before any effect appears. The reader is referred to Bernal et al. (2017) for more details about modelling different types of sudden changes in time-series data.

Based on coral cover trends over time (Figure A.1), we hypothesised that the 1998 mass bleaching event resulted in both changes to the intercept and slope when modelling coral cover trajectories. Furthermore, it was proposed that the impact existed for years as mortality does not happen completely during or a few months after bleaching (Gilmour and Smith, 2013, Baird and Marshall, 2002). The model defined previously using Equation (2) can now be extended to accommodate such an impact as follows:

$$\begin{aligned}\log\left(p_{srt}/(1-p_{srt})\right) =& \gamma_{sr} + \boldsymbol{\beta}_t \boldsymbol{z}_t + \boldsymbol{\beta}_d \boldsymbol{d}_r + f(x_{srt}, \boldsymbol{\theta}_0) + \beta_l \mathrm{BLE98}_{srt} \\ &+ \beta_s \mathrm{Time98}_{srt},\end{aligned} \tag{3}$$

where $\mathrm{BLE98}_{srt}$ represents the bleaching impact, i.e. $\mathrm{BLE98}_{srt} = 0$ before the bleaching event happened and otherwise it is equal to 1, β_l is the level change due to the bleaching impact. Here, $\mathrm{Time98}_{srt}$ represents the time before and after the bleaching event, i.e. $\text{Time98}_{srt} = 0$ before the bleaching event occurred, and after that, time increases with survey time, and β_s represents the corresponding slope change.

In the model, cyclone hours data were count values that varied over a large range; thus, the square-root transformation (Weber, 1990) was applied before including this covariate into the model. This transformation was also applied to ensure a linear relationship was appropriate between cyclone hours and $\log(p_{srt}/(1 - p_{srt}))$ (O'Hara and Kotze, 2010). Previous studies have considered centring covariates to avoid numerical issues when fitting a given model, and we follow this approach for the time-varying covariates (Selig et al., 2012, Vercelloni et al., 2014). Furthermore, we calculated $|x_{srt} - \eta_k|$ by considering the centered time (Crainiceanu et al., 2005).

Approximating the posterior distribution

Within a Bayesian framework, we are interested in estimating the joint posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{\xi} \mid \boldsymbol{y}_p, \boldsymbol{z}_p, \boldsymbol{v}_p, \boldsymbol{d}_p)$ of model parameters and random effects, where $\boldsymbol{\theta} = (\beta_0,\ \beta_1,\ \boldsymbol{\beta}_t,\ \boldsymbol{\beta}_d,\ \beta_l,\ \beta_s,\ \boldsymbol{\delta},\ \log\sigma_\delta,\ \log\sigma_r,\ \log\sigma_s)$ denotes all parameters in the model (Equation (3)), $\boldsymbol{\xi}$ is a matrix representation for the nested random effects, $\boldsymbol{d}_p$ denotes previously used static survey designs at Scott Reef, $\boldsymbol{v}_p$ represents data matrices related to $|x_{srt} - \eta_k|$ and the interrupted component (i.e. BLE98_{srt} and Time98_{srt}), $\boldsymbol{y}_p$ denotes the collected coral cover data, and $\boldsymbol{z}_p$ is the collected time-varying covariates, where we have shifted notation such that all previously collected data will be indexed by p. This will be convenient when considering future monitoring scenarios, see later. To estimate the posterior distribution (see Appendix A.3 for more details), Markov Chain Monte Carlo (MCMC) methods can be used. For this purpose, WinBUGS was used (Lunn et al., 2000).

Model selection

To find the most appropriate model to describe the previously collected data at Scott Reef, we considered the $\mathcal{M}-$closed perspective of Bernardo and Smith (2009). Accordingly, the most appropriate model for the data is assumed to be contained within a finite set of L candidate models $\{m \in 1, 2, \cdots, L\}$. We defined the class of models by considering the following components: the nested random effects for sites within reef locations (NRE); a smooth component (SC); all the available covariates (ALL COV), i.e. Time, Bleaching, Cyclone hours, Interrupted 98 (i.e. BLE98$_{srt}$ and Time98$_{srt}$), location-specific covariates impacts, i.e. Cyclone Loc2 (i.e. Cyclone Loc and Severe cyclone Loc) and Bleaching Loc, and the interaction between Bleaching and Cyclone. The most appropriate model within this class was then determined via the deviance information criterion (DIC) with a preference for the model with the smallest of these values (Spiegelhalter et al., 2014). Prior information was specified to be vague on likely range of values of each parameter (Table A.2). In addition, to appropriately capture the nonlinear features of the data, a specific number of knots needs to be determined. For this, we followed the approach of Ruppert (2002) where the number of knots was increased until there was little to no improvement in model fit. This resulted in the use of three knots.

5.2. Adaptive design

We consider a Bayesian inference framework for undertaking adaptive design as it supports the incorporation of knowledge gained through previously collected data to guide future sampling via a prior distribution. The prior that was used for this study was the posterior distribution for the model that was found to be most appropriate for the collected data. Based on this prior information, the value of a design with respect to a given monitoring objective can be evaluated as described below.

Quantifying designs

Define a design as $\boldsymbol{d} = (\boldsymbol{d}_1, \boldsymbol{d}_2, \cdots, \boldsymbol{d}_{n_s})$, where n_s is the number of sites appearing in a new design out of all sites (i.e. 7 reef locations $\times$ 3 sites = 21 sites). The usefulness of such a design $\boldsymbol{d}$ can be quantified via what is called a utility function which evaluates how much information will be provided from data $\boldsymbol{y}$ to address a specific monitoring objective. As it is unknown what data will be observed, the expectation of the utility function is taken with respect to this and other unknowns as follows:

$$E[u(\boldsymbol{d}, \boldsymbol{z}, \boldsymbol{y}|\boldsymbol{y}_p, \boldsymbol{z}_p, \boldsymbol{v}_p, \boldsymbol{d}_p)] = \int_{\boldsymbol{y}} u(\boldsymbol{d}, \boldsymbol{z}, \boldsymbol{y}|\boldsymbol{y}_p, \boldsymbol{z}_p, \boldsymbol{v}_p, \boldsymbol{d}_p) \\ p(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{d}, \boldsymbol{y}_p, \boldsymbol{z}_p, \boldsymbol{v}_p, \boldsymbol{d}_p) d\boldsymbol{y}, \tag{4}$$

where $\boldsymbol{z}$ represents specific values of the time-varying covariates which define particular future monitoring scenarios. Further details about these scenarios will be provided later in this section.

The choice of a utility function depends on the monitoring objective. Here, our goal is to determine the relative importance of survey locations for providing information about coral health based on a statistical model. This suggests we are interested in the precise estimation of parameters in the adopted model so that, for example, we can precisely quantify the impact of each disturbance. Thus, we chose the Kullback-Leibler divergence (KLD) (Kullback and Leibler, 1951), which is a specific utility for the parameter estimation. The KLD utility function can be expressed as follows (Friel and Pettitt, 2008):

$$\begin{aligned} u(\boldsymbol{d}, \boldsymbol{y}, \boldsymbol{z}|\boldsymbol{y}_p, \boldsymbol{z}_p, \boldsymbol{v}_p, \boldsymbol{d}_p) = \int_{\boldsymbol{\theta}} \int_{\boldsymbol{\xi}} & p(\boldsymbol{\theta}, \boldsymbol{\xi} \mid \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{y}_p, \boldsymbol{z}_p, \boldsymbol{v}_p, \boldsymbol{d}, \boldsymbol{d}_p) \\ & \times \log p(\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{z}, \boldsymbol{y}_p, \boldsymbol{z}_p, \boldsymbol{v}_p, \boldsymbol{d}, \boldsymbol{d}_p) d\boldsymbol{\xi} d\boldsymbol{\theta} \\ & - \log p(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}_p, \boldsymbol{y}_p, \boldsymbol{v}_p, \boldsymbol{d}, \boldsymbol{d}_p). \end{aligned} \quad (5)$$

Evaluating this utility thus measures how much the posterior distribution diverges from the prior. In terms of designs, a larger deviation for a given design $\boldsymbol{d}$ indicates more has been learned from data collected according to that design. Thus, we seek a design $\boldsymbol{d}$ that maximises the expectation given in Equation (4) where the utility is defined in Equation (5).

To find an optimal design, we need to evaluate the expected utility. However, typically this expression does not have a closed-form solution. Thus, a numerical approximation is required. One can use the Monte Carlo integration (Ryan, 2003) for this purpose which can be defined as follows:

$$E[u(\boldsymbol{d}, \boldsymbol{y}, \boldsymbol{z} | \boldsymbol{z}_p, \boldsymbol{y}_p, \boldsymbol{v}_p, \boldsymbol{d}_p)] \approx \frac{1}{J} \sum_{j=1}^{J} u(\boldsymbol{d}, \boldsymbol{y}^{(j)}, \boldsymbol{z} | \boldsymbol{z}_p, \boldsymbol{y}_p, \boldsymbol{v}_p, \boldsymbol{d}_p), \tag{6}$$

where $J\,(\geq 100)$ is the controlling parameter for Monte Carlo integration.

Algorithm 1: Approximate the expected utility

1. Initialise $\boldsymbol{d}, \boldsymbol{y}_p, \boldsymbol{z}_p, \boldsymbol{v}_p, \boldsymbol{d}_p, \boldsymbol{z}, t, J$
2. For $j = 1$ to J do
3. Simulate $\boldsymbol{\theta}^{(j)}, \boldsymbol{\xi}^{(j)} \sim p(\boldsymbol{\theta}, \boldsymbol{\xi} \mid \boldsymbol{y}_p, \boldsymbol{z}_p, \boldsymbol{v}_p, \boldsymbol{d}_p)$
4. Simulate $\boldsymbol{y}^{(j)} \sim p(\boldsymbol{y} \mid \boldsymbol{\theta}^{(j)}, \boldsymbol{\xi}^{(j)}, \boldsymbol{z}, \boldsymbol{y}_p, \boldsymbol{z}_p, \boldsymbol{v}_p, \boldsymbol{d}_p)$ at the next survey time t based on $\boldsymbol{d}$ via a Taylor series approximation to the mean response
5. Estimate $p(\boldsymbol{\theta}, \boldsymbol{\xi} \mid \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{y}_p, \boldsymbol{z}_p, \boldsymbol{v}_p, \boldsymbol{d}, \boldsymbol{d}_p)$ via Laplace approximation
6. Evaluate KLD utility $u(\boldsymbol{d}, \boldsymbol{z}, \boldsymbol{y}^{(j)} | \boldsymbol{z}_p, \boldsymbol{y}_p, \boldsymbol{v}_p, \boldsymbol{d}_p)$
7. Store $u(\boldsymbol{d}, \boldsymbol{z}, \boldsymbol{y}^{(j)} | \boldsymbol{z}_p, \boldsymbol{y}_p, \boldsymbol{v}_p, \boldsymbol{d}_p)$
8. End for
9. Output $E[u(\boldsymbol{d}, \boldsymbol{z}, \boldsymbol{y} | \boldsymbol{z}_p, \boldsymbol{y}_p, \boldsymbol{v}_p, \boldsymbol{d}_p)] \approx \frac{1}{J} \sum_{j=1}^{J} u(\boldsymbol{d}, \boldsymbol{z}, \boldsymbol{y}^{(j)} | \boldsymbol{z}_p, \boldsymbol{y}_p, \boldsymbol{v}_p, \boldsymbol{d}_p)$

Approximating the expected utility

The approach to approximate the expected utility is outlined in Algorithm 1, which begins by initialising some parameters (line 1). To approximate the expected utility for a given design $\boldsymbol{d}$, many data sets need to be simulated based on the given design (line 2). For this purpose, we simulate parameter and random effect values from the prior distribution (line 3). To simulate coral cover for the next survey time point t (line 4), we propose the Taylor series expansion around the mean at the last survey point, i.e. 2016.08 (Table A.6). In such a case, a bivariate Taylor series expansion needs to be applied as the regression model includes two time variables following the incorporation of the interrupted regression component. For the given model, the Taylor series approximation can be described as follows: $f(x_{srt}, \text{Time98}_{srt}) \approx f(a, b) + f_{x_{srt}}(a, b)(x_{srt} - a) + f_{\text{Time98}_{srt}}(a, b)(\text{Time98}_{srt} - b)$, where $f(x_{srt}, \text{Time98}_{srt})$ is the value of the function at the next survey point (i.e. $t = 2016.33$) and (a, b) are values which with the Taylor series is centered (see Appendix A.7 for extrapolation results). The posterior distribution needs to be approximated for each simulated data set (line 5). Given this needs to be performed a large number of times, this is a computationally demanding step in the algorithm. To address this, a Laplace approximation (Overstall et al., 2018) was adopted. Given this approximation to the posterior distribution, the KLD utility can be evaluated (see Appendix A.9 for more details) (lines 6-7). Finally, an average of KLD utility values is used to approximate the expected utility value (line 9).

Finding optimal designs

We are now able to approximate the expected utility of a given design $\boldsymbol{d}$. The next step is to find the optimal design $\boldsymbol{d}^*$ out of the set of candidate designs which maximises the expected utility, i.e. $\boldsymbol{d}^* = \arg\max_{\boldsymbol{d}} E[u(\boldsymbol{d}, \boldsymbol{z}, \boldsymbol{y}|\boldsymbol{z}_p, \boldsymbol{y}_p, \boldsymbol{v}_p, \boldsymbol{d}_p)]$. Here, candidate designs need to be formulated in accordance with monitoring objectives. As far as Objective (i) is concerned, we will investigate whether some reef locations within Scott Reef have greater utility than others. Accordingly, seven candidate designs were formulated considering all possible combinations where six out of seven reef locations will be sampled. The corresponding designs were labelled as SL1, $\cdots$, SS3 where for example, design SL1 denotes that no data will be collected from the reef location SL1 for the next survey time. As there will be a fixed number of potential candidate designs (i.e. seven) under this objective, we will enumerate all possible designs to determine the optimal. Next, under Objective (ii), we determine the optimal design consisting of the most informative site (out of the three) at each reef location. To locate these optimal designs, the coordinate-exchange algorithm was used based on five randomly selected initial designs (Meyer and Nachtsheim, 1995).

To evaluate our adaptive designs, two disturbance scenarios (i.e. two different values for $\boldsymbol{z}$) were considered. These were: (a) actual covariate data collected at the next survey time (i.e. 2016.33) and (b) bleaching and cyclone impacts including an interaction between them at each reef location (see Appendix A.8 for more details). This means that Scenario (a) withdraws the prevailing cyclone exposure while Scenario (b) includes cyclone location disturbances and cyclone-bleaching interactions for each reef location. Each objective defined previously will be assessed under these two disturbance scenarios.

6. Results

6.1. Modelling historical data from Scott Reef

Model selection

To select the most appropriate model for Scott Reef hard coral cover data, we defined the class of models by considering all components described in Section 5.1. The corresponding model comparison results based on DIC are provided in Table A.3. As the model with a smaller value of DIC is preferred, the most appropriate model found for coral cover can be described as follows:

$$\begin{aligned}\log\left(p_{srt}/(1-p_{srt})\right) =& \beta_0 + \beta_1 \text{Time} + \beta_2 \text{Bleaching}_t + \beta_3 \text{Cyclone hours}_t \\ &+ \beta_4 \text{Bleaching} \times \text{Cyclone}_{rt} + \beta_5 \text{BLE98}_t + \beta_6 \text{Time98}_t \\ &+ \beta_7 \text{Cyclone Loc}_{rt} + \beta_8 \text{Severe cyclone Loc}_{rt} \\ &+ \beta_9 \text{Bleaching Loc}_{rt} + \delta_1 v_{1,rst} + \delta_2 v_{2,rst} + \delta_3 v_{3,rst} + \gamma_{sr},\end{aligned} \tag{7}$$

where $\beta_i, i = 1, \cdots, 9$ are the regression coefficients and δ_1, δ_2, and δ_3 are random coefficients. The goodness-of-fit of this model was assessed and found to be appropriate, see Appendix A.6 for further details.

6.2. Adaptive design

Importance of reef locations

Under Objective (i), we aim to determine the relative importance of seven reef locations at Scott Reef. Firstly, the disturbance Scenario (a) is considered. For this evaluation, we formulated seven designs (i.e. SL1, $\cdots$, SS3), as described in Section 5.2. To evaluate these designs, the KLD mean expected utility values were evaluated. These results are shown in Figure 2 (a), where the y-axis represents the efficiency of each design with respect to the design where all reef locations were included for survey at the next survey time point. These design efficiencies were found with respect to sampling all seven reef locations by evaluating the expected utility of each design 20 times, and taking the average (see Appendix A.10 for more details). A summary of utility evaluations is given in Table 1 to aid in interpretation.

According to Table 1, SL3 is the optimal design as it has the highest mean efficiency. As the missing reef location within this design is SL3, this reef location can be considered as the least informative reef location under Scenario (a). This suggests that less information is expected to be lost by omitting reef location SL3 compared to omitting any other reef locations. Similarly, the reef location SL2 can be reported as the most informative reef location as omitting this reef location resulted in the largest efficiency reduction (Table 1). However, it should be noted that there is very little difference

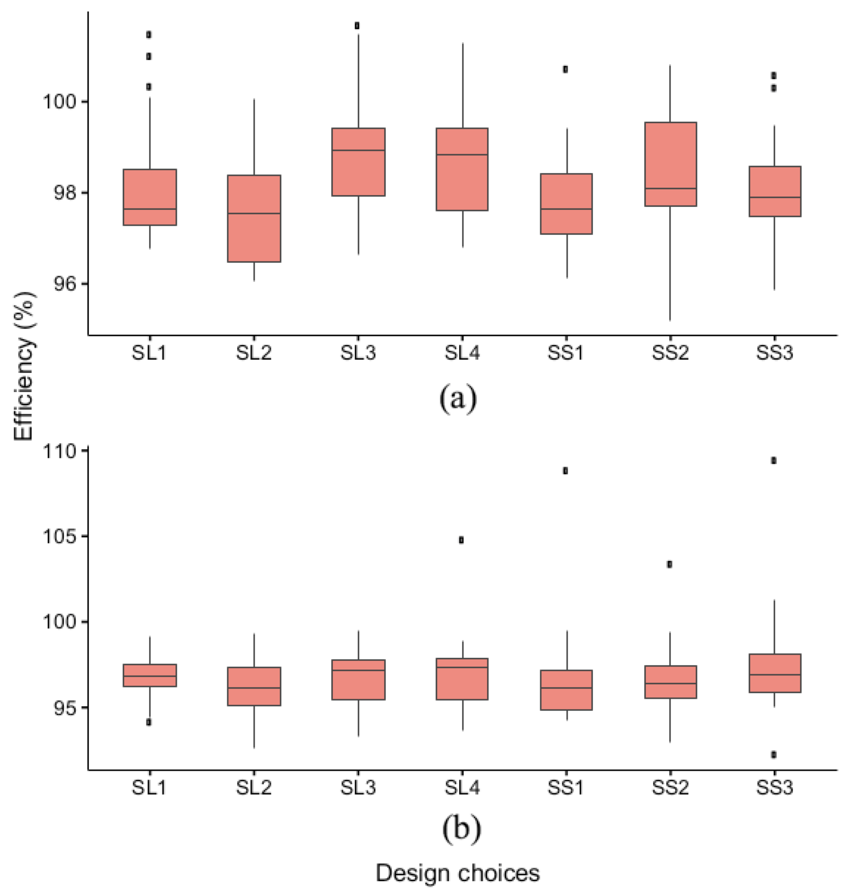

Figure 2: Utility evaluations results under disturbance Scenario (a) and (b) for seven designs (x-axis). y-axis represents efficiency calculated against the design that included all seven reef locations.

in the efficiency values between reef locations (Figure 2 (a)). Indeed, some of the differences observed could potentially be due to Monte Carlo error.

Secondly, we evaluate Objective (i) under the disturbance Scenario (b). The corresponding utility evaluation results and summary are provided in Figure 2 (b) and Table 2, respectively. It is evident from Figure 2 (b) that all the reef locations have similar median efficiency values under Scenario

Table 1: Summary of utility evaluations for seven designs under Scenario (a) in Objective (i).

Design	Mean efficiency (%)	Standard deviation
SL1	98.213	1.399
SL2	97.579	1.195
SL3	98.876	1.395
SL4	98.774	1.290
SS1	97.818	1.082
SS2	98.420	1.354
SS3	98.054	1.176

(b). The design SS3 has the highest efficiency; thus, the reef location SS3 can be reported as the least informative reef location under Scenario (b). Furthermore, comparing Figure 2 (a) and (b) shows that there are similar efficiency for designs under the two disturbance scenarios considered. This suggests that the optimal design is robust to the two scenarios considered under Objective (i).

To explore these design selections, consider that the posterior means $\boldsymbol{\theta}^*_{m_1}$ will be similar to the prior means $\boldsymbol{\theta}^*_{m_0}$, so this should not particularly contribute to the optimal design selection under KLD utility function, see Equation (A.1). Accordingly, these design selections could be driven by the posterior variance-covariance of the parameters. Upon investigating this, the larger utility values appeared to be related to the estimation of the reef random effect standard deviation parameter (i.e. $\log \sigma_r$). This is typically where

Table 2: Summary of utility evaluations for seven designs under Scenario (b) in Objective (i).

Design	Mean efficiency (%)	Standard deviation
SL1	96.708	1.302
SL2	96.205	1.682
SL3	96.775	1.612
SL4	97.064	2.360
SS1	96.699	3.145
SS2	96.663	2.108
SS3	97.474	3.369

the largest change from the prior was observed, and thus could potentially be driving the design selection i.e. designs that provide more information on the variability of reef locations are being preferred.

Table 3: Summary of utility evaluations under each disturbance scenario in Objective (ii).

Design	Mean efficiency (%)	Standard Deviation
Scenario (a)	86.285	5.492
Scenario (b)	71.184	2.616

Informative sites at each reef location

Under Objective (ii), we determine the optimal design that consists of the most informative site at each reef location subject to two disturbance scenarios. The mean efficiency values of each optimal design are provided in Table 3. The selected sites from each reef location into the optimal designs under the two scenarios are reported in Table A.7. It can be seen from the table that the optimal designs contain different site combinations under the two scenarios. It was found through investigating these optimal designs that the estimation of log σ_r was a main contributor to these optimal design selections. This indicates that the optimal site combination under a given disturbance scenario provides more information about the between reef variability.

The optimal designs under the two scenarios have the mean efficiency values of 86.285% and 71.184% respectively (Table 3). It is evident from these a considerable reduction in mean efficiency under Scenario (b). This indicates that if disturbances are similar to previous years, then minimal sampling (i.e. one site per reef location) will capture a substantial proportion of information. However, when a variety of disturbance combinations are observed, then there appears to be more information lost, so there might be value to undertaking additional sampling in such cases.

7. Discussion

The present study was designed to develop adaptive design methods using semiparametric and ITS models and utilised prior knowledge captured through such methods to guide future surveys at Scott Reef. We demonstrated the use of such a modelling approach in finding adaptive design when data show nonlinear trends with some sudden shifts over time. For this purpose, it was shown that changes around major environmental disturbances in ecological monitoring could be accounted for using an ITS regression modelling approach. This enabled prior information from historical data to be appropriately formed when such data potentially exhibit complex ecological relationships.

We assessed the importance of reef locations under Objective (i) subject to two disturbance scenarios. The results showed that there was very little difference between the selection of which reef location to omit under any of the scenarios. This indicates that the design choice is relatively inconsequential. Also, dropping one reef location resulted in very little information loss, allowing the survey effort to be reduced without losing a substantial amount of information about the parameters in the developed model.

Under Objective (ii), we found the optimal designs consisting of one site at a given reef location based on two disturbance scenarios. This provides insight into the most appropriate site to sample from a given reef location depending on prevailing disturbance conditions. The differences between the optimal designs between these two scenarios suggests that site selection depends on the disturbances that have been observed, and our methods provide a framework with which to make this decision.

In terms of modelling monitoring data, the Gompertz model has been considered recently (MacNeil et al., 2019) for capturing nonlinear relationships in population growth. However, such a model proved to not be flexible enough to capture nonlinear trends where observations have been collected with unequal time gaps. In such circumstances, semiparametric modelling approaches can be utilized to capture nonlinear trends, as demonstrated in this study. The consideration of such a modelling approach meant that new design methods needed to be proposed such that adaptive designs could be found in this context. This was demonstrated by considering two future disturbance scenarios, and assessing the performance of these designs against more resource intensive sampling.

In terms of future research directions, we did not consider uncertainty about future disturbances but instead considered how the optimal design choice changed depending on the given disturbance conditions. If one were to do so, it might yield robust designs over various disturbance patterns. The use of this will most likely be dependent on a given year. Indeed, it is likely that it will be known if a severe cyclone and mass bleaching has occurred at Scott Reef, but data on the occurrence of other disturbances may not be readily available.

Recommendations for future reef monitoring based on findings from Scott Reef

(i) Pre-assessment of the expected information gain by location (e.g. site or reef location) can be used to determine if any locations can be prioritised for data collection. Here, it was found that information gain from sites was similar, so no particular location needed to be prioritised over another.

(ii) After an extensive monitoring period, explore reduced sampling practices as there is potential to reduce sampling effort (e.g. drop site or reef location) without experiencing significant information loss about coral health.

(iii) Evaluate disturbance patterns at monitoring locations as these can influence information gain e.g. here it was shown that more information about coral health was obtained when new disturbance patterns were experienced when compared to historical disturbance patterns.

(iv) On-going review of monitoring practices is recommended to assess effectiveness of adaptive designs

8. Acknowledgements

Thilan AWLP was supported by the Australian Technology Network of Universities Industry Doctoral Training Centre (ATN IDTC) Scholarship. McGree JM was supported by an Australian Research Council Discovery Project (DP200101263). We would like to thank the ACEMS, the AIMS and QUT Centre for Data Science. Historical data was collected by AIMS with funding from Woodside as operator for and on behalf of the Browse Joint Venture. Also, we would like to acknowledge computer resources and services such as HPC provided by QUT. And finally, thanks to Nicole Ryan for her help regarding Scott Reef data. Author contributions: Thilan AWLP, McGree JM, and Fisher R designed the research; Thilan AWLP performed the research and wrote the paper; Fisher R, Thompson H, Menendez P, and McGree JM provided research conception and critical review of manuscript drafts. Gilmour J contributed through conceptualisation and the acquisition of the SRRP data. Further, we have no conflicts of interest to disclose.

9. Additional information

9.1. Data availability

The access to data that support the findings of this study are available through the below link from the corresponding author, Thilan AWLP, upon reasonable request. Giving such access requires prior approval from the Australian Institute of Marine Science who owns these data:
https://doi.org/10.5281/zenodo.4574062

9.2. Computer code

All code to reproduce the analysis can be found through the following DOI:
https://doi.org/10.5281/zenodo.4586187

References

J. P. Gilmour, K. L. Cook, N. M. Ryan, M. L. Puotinen, R. H. Green, G. Shedrawi, J.-P. A. Hobbs, D. P. Thomson, R. C. Babcock, J. Buckee, et al., The state of western australia's coral reefs, Coral Reefs 38 (2019) 651–667.

R. A. Campbell, B. D. Mapstone, A. D. Smith, Evaluating large-scale experimental designs for management of coral trout on the Great Barrier Reef, Ecological Applications 11 (2001) 1763–1777.

S. Y. Kang, J. M. McGree, C. C. Drovandi, M. J. Caley, K. L. Mengersen, Bayesian adaptive design: improving the effectiveness of monitoring of the Great Barrier Reef, Ecological applications 26 (2016) 2637–2648.

A. Thilan, E. E. Peterson, P. Menendez, J. Caley, C. Drovandi, C. Mellin, J. McGree, Bayesian design methods for improving the effectiveness of ecosystem monitoring, Working paper (2021).

T. Newbold, D. P. Tittensor, M. B. Harfoot, J. P. Scharlemann, D. W. Purves, Non-linear changes in modelled terrestrial ecosystems subjected to perturbations, Scientific reports 10 (2020) 1–10.

T. J. Done, Phase shifts in coral reef communities and their ecological significance, Hydrobiologia 247 (1992) 121–132.

L. J. McCook, Macroalgae, nutrients and phase shifts on coral reefs: scientific issues and management consequences for the Great Barrier Reef, Coral reefs 18 (1999) 357–367.

F. J. Oddi, F. E. Miguez, L. Ghermandi, L. O. Bianchi, L. A. Garibaldi, A nonlinear mixed-effects modeling approach for ecological data: Using temporal dynamics of vegetation moisture as an example, Ecology and Evolution 9 (2019) 10225–10240.

M. Scheffer, S. Carpenter, J. A. Foley, C. Folke, B. Walker, Catastrophic shifts in ecosystems, Nature 413 (2001) 591–596.

C. Crainiceanu, D. Ruppert, M. P. Wand, Bayesian analysis for penalized spline regression using WinBUGS (2005). Available from https://ro.uow.edu.au/cgi/viewcontent.cgi?article=3526context=eispapers.

J. Vercelloni, M. J. Caley, M. Kayal, S. Low-Choy, K. Mengersen, Understanding uncertainties in non-linear population trajectories: A Bayesian semi-parametric hierarchical approach to large-scale surveys of coral cover, PloS one 9 (2014) e110968.

J. Vercelloni, K. Mengersen, F. Ruggeri, M. J. Caley, Improved coral population estimation reveals trends at multiple scales on Australia's Great Barrier Reef, Ecosystems 20 (2017) 1337–1350.

J. Gilmour, L. Smith, Discovering Scott Reef: 20 years of exploration, 2013.

M. Puotinen, J. A. Maynard, R. Beeden, B. Radford, G. J. Williams, A robust operational model for predicting where tropical cyclone waves damage coral reefs, Scientific reports 6 (2016) 1–12.

S. N. Wood, Thin plate regression splines, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65 (2003) 95–114.

G. De'ath, K. E. Fabricius, H. Sweatman, M. Puotinen, The 27–year decline of coral cover on the great barrier reef and its causes, Proceedings of the National Academy of Sciences 109 (2012) 17995–17999.

K. Osborne, A. M. Dolman, S. C. Burgess, K. A. Johns, Disturbance and the dynamics of coral cover on the great barrier reef (1995–2009), PloS one 6 (2011) e17516.

J. L. Bernal, S. Cummins, A. Gasparrini, Interrupted time series regression for the evaluation of public health interventions: a tutorial, International journal of epidemiology 46 (2017) 348–355.

A. Linden, Conducting interrupted time-series analysis for single-and multiple-group comparisons, The Stata Journal 15 (2015) 480–500.

D. McDowall, R. McCleary, B. J. Bartos, Interrupted time series analysis, Oxford University Press, 2019.

A. Baird, P. Marshall, Mortality, growth and reproduction in scleractinian corals following bleaching on the Great Barrier Reef, Marine Ecology Progress Series 237 (2002) 133–141.

R. P. Weber, Basic content analysis, 49, Sage, 1990.

R. O'Hara, J. Kotze, Do not log-transform count data, Nature Precedings (2010) 1–1.

E. R. Selig, K. S. Casey, J. F. Bruno, Temperature-driven coral decline: the role of marine protected areas, Global Change Biology 18 (2012) 1561–1570.

D. J. Lunn, A. Thomas, N. Best, D. Spiegelhalter, WinBUGS-A Bayesian modelling framework: concepts, structure, and extensibility, Statistics and computing 10 (2000) 325–337.

J. M. Bernardo, A. F. Smith, Bayesian theory, volume 405, John Wiley & Sons, 2009.

D. J. Spiegelhalter, N. G. Best, B. P. Carlin, A. Van der Linde, The deviance information criterion: 12 years on, Journal of the Royal Statistical Society: Series B: Statistical Methodology (2014) 485–493.

D. Ruppert, Selecting the number of knots for penalized splines, Journal of computational and graphical statistics 11 (2002) 735–757.

S. Kullback, R. A. Leibler, On information and sufficiency, The annals of mathematical statistics 22 (1951) 79–86.

N. Friel, A. N. Pettitt, Marginal likelihood estimation via power posteriors, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70 (2008) 589–607.

K. J. Ryan, Estimating expected information gains for experimental designs with application to the random fatigue-limit model, Journal of Computational and Graphical Statistics 12 (2003) 585–603.

A. M. Overstall, J. M. McGree, C. C. Drovandi, An approach for finding fully Bayesian optimal designs using normal-based approximations to loss functions, Statistics and Computing 28 (2018) 343–358.

R. K. Meyer, C. J. Nachtsheim, The coordinate-exchange algorithm for constructing exact optimal experimental designs, Technometrics 37 (1995) 60–69.

M. A. MacNeil, C. Mellin, S. Matthews, N. H. Wolff, T. R. McClanahan, M. Devlin, C. Drovandi, K. Mengersen, N. A. Graham, Water quality mediates resilience on the Great Barrier Reef, Nature ecology & evolution 3 (2019) 620–627.

A. Gelman, J. Hill, Data analysis using regression and multilevel hierarchical models, volume 1, Cambridge University Press New York, NY, USA, 2007.