Model-Based Adaptive Design Methods for Improving the Effectiveness of Reef Monitoring

Pubudu Thilan

James McGree[1], Erin Peterson[1], Christopher Drovandi[1], Julian Caley[1], Patricia Menendez[2]

[1]School of Mathematical Sciences, Queensland University of Technology

[2]Australian Institute of Marine Sciences

Outline

- Motivation
- Background
- Objectives
- Proposed Bayesian design framework
  - Prior information
  - Designs for Ecological Monitoring
  - Evaluation
- References
- Questions

Motivation

Great Barrier Reef (GBR)

- The Great Barrier Reef (GBR) is the world's largest coral reef system spreading over 348,000 Km^2.
- It has a great historical and cultural value.
- It attracts around 5.2 billion dollars each year to the Australian economy through tourism.

Motivation

Why Monitoring is Required?

- Coral reef are under many environmental threats.
- The world's first major coral bleaching occurred in 1998.
- It was an eye opener for scientists to establish monitoring programmes.
- Monitoring programs play a key role in identifying patterns, trends and threats to coral reef systems.

Motivation

GBR Monitoring

- Australian Institute of Marine Science (AIMS) has been surveying for the health of GBR over 20 years.
- Their Long-term Monitoring Program (LTMP) represents the longest and continuous record of change in reef communities.
- In LTMP, samples are collected **every two years** from the **selected reefs (sites)**.

Questions:

Why every two years? Why not every year or every 3 years?

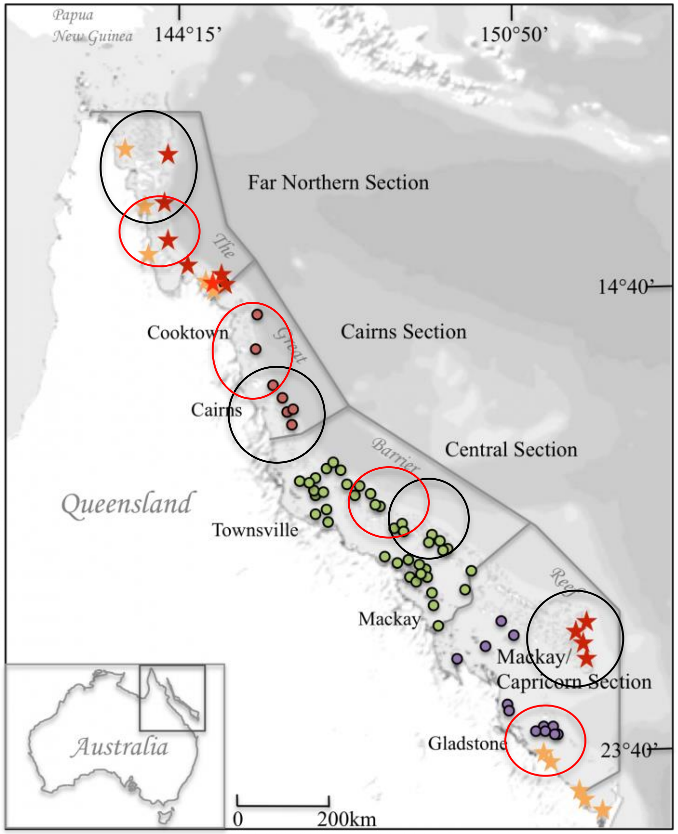
Are they collecting data from optimal set of reefs/sites?

Background
Optimal design

Experimental Design involves finding optimal sampling locations in **space** or/and **time** (a design).

An **optimal design** maximises the amount of "information" that can be obtained for a given amount of data collection effort.

We focus on finding optimal sampling locations in **space** for a future time period.



Background

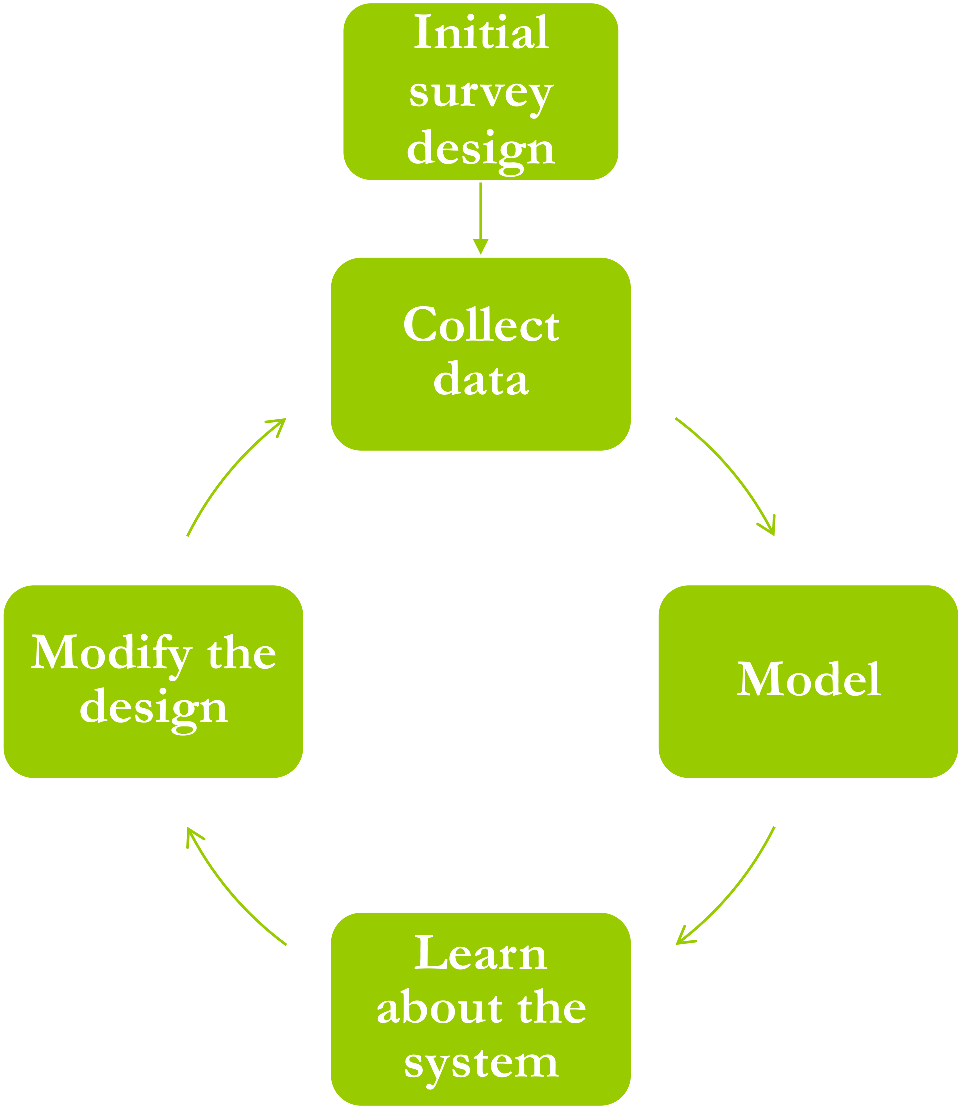
Types of designs

There are two basic types of designs:

- Static design – Design which remains fixed over time.
- Adaptive design – Design which changes over time.

Background

Adaptive design approach



Background

Current LTMP design

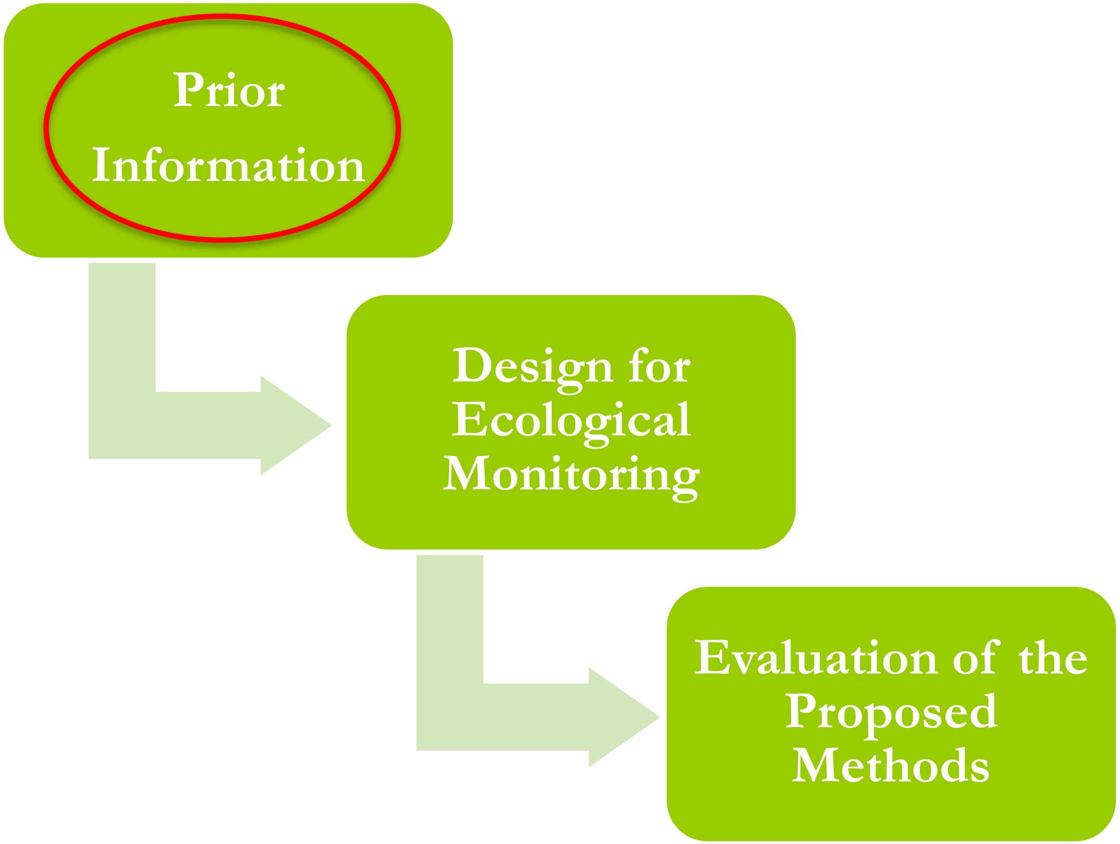
- In LTMP, data are gathered from predetermined reefs and time points that do not change over time.
- Potentially, more effective monitoring could be achieved if sampling designs were developed
  - ✓ through analyzing previously collected data and,
  - ✓ incorporating disturbance information such as cyclone impacts, occurrence of bleaching events, and crown-of-thorns starfish outbreaks.

Objectives

We evaluate adaptive design through considering:

1. The effect of **visiting less LTMP sites** within different regions of the GBR.
2. Comparing the ability to accurately estimate parameters of the fixed LTMP design to **designs that change over time** depending on the reef condition and disturbances.

Proposed Bayesian design framework



- Comparison
- Objective 1
- Objective 2

Prior information

Why Whitsunday region?

Relatively large amount of data is available for this region.

Diverse range of anthropogenic impacts that have occurred in this region over time.

Prior information

Potential covariates

Covariate	Description	Resolution	Type
Shelf position	inshore, middle, outer	0.005°	Site-specific
No-take	Not allowed fishing	0.005°	
Bathymetry	Depth below sea level	0.0003°	
Chlorophyll	Mean concentration of chlorophyll A pigments	0.01°	
CRS_T_AV	Mean temperature	0.01°	
Cyclone	0 = No cyclone effects, 1 = Some cyclone effects	0.01°	Time-varying
Bleaching	0 = No coral bleaching, 1 ≥ 1% coral bleached	0.01°	
CoTS	Mean A. solaris densities per manta tow	0.01°	
Time	Sampling year	2002-2015	

Prior information

A statistical model for coral cover

We assume geostatistical mixed Beta regression model;

$$\eta_{ijk} \coloneqq \boldsymbol{G}(\mu_{ijk}) = \boldsymbol{G}\left(E(y_{ijk} \,|\boldsymbol{\theta})\right) = \boldsymbol{l}_{ijk}^{T}\boldsymbol{\beta}_{l} + \boldsymbol{z}_{ijk}^{T}\boldsymbol{\beta}_{z} + \beta_{t}\mathrm{Time}_{k} + r_{ijk}\,,$$

$$y_{ijk} \sim \mathrm{Beta}(\mu_{ijk}\,, \psi_{ijk}\,)$$

$\boldsymbol{G}(\cdot)$ −logit linking function
$\boldsymbol{l}_{ijk}$ −site-specific covariates
$\boldsymbol{\beta}_{l}$ −regression coefficients for the site specific-covariates
$\boldsymbol{z}_{ijk}$ −time-varying covariates
$\boldsymbol{\beta}_{z}$ −regression coefficients of time-varying covariates
β_{t} −regression coefficient of Time

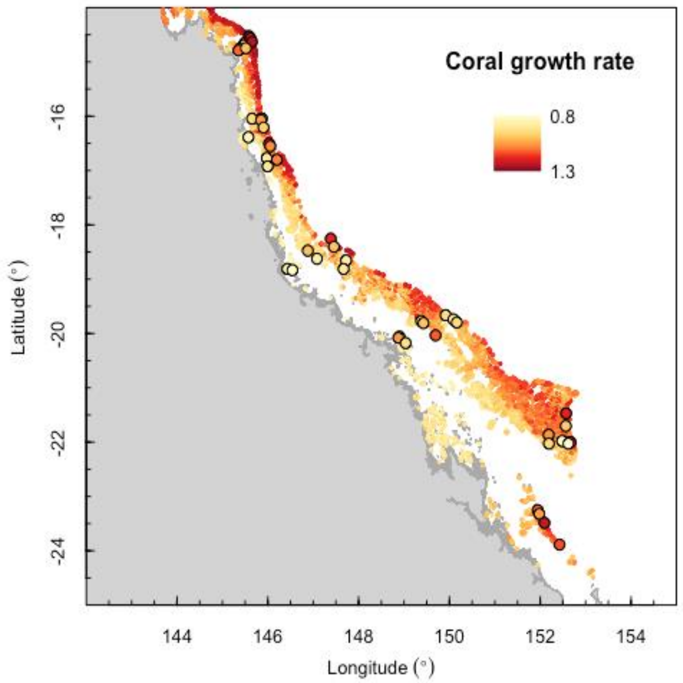
Prior information

Why spatial variability needs to be considered?

Understanding how coral cover varies through space is essential for deriving sampling strategies.

Informs how close sites need to be to capture the heterogeneity.

Improves parameter estimates and model predictions in areas where you didn't sample when data are spatially dependent.



Source: Mellin c, et al. (in review).

Prior information

Covariance model

We chose to use the Gaussian covariance model parameterized as

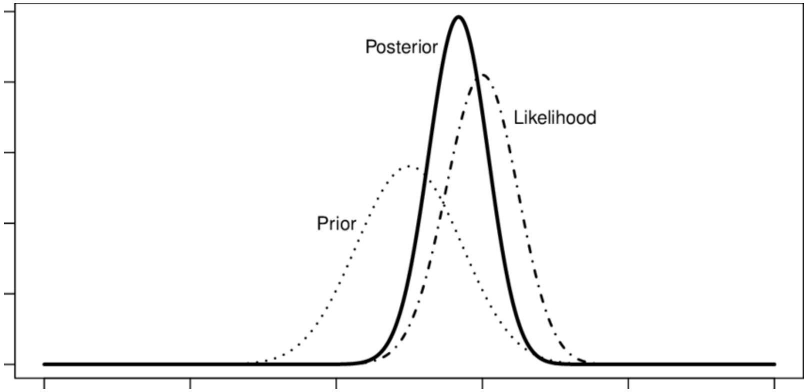
$$\boldsymbol{\Sigma_r} = \sigma_r^2 \exp\left(-\left(\frac{a h_{s_{i_1} s_{i_2}}}{\phi}\right)^2\right), i_1, i_2 = 1, \ldots, n$$

where $h_{s_{i_1} s_{i_2}}$ is the distance between sites s_{i_1} and s_{i_2}, σ_r^2 is the variance of the spatial process (partial sill) and ϕ is the range parameter.

Prior information

Bayesian framework

- We use a Bayesian modelling framework as we can use expert elicited and previously collected data to inform priors on the model and model parameters.
- We chose a weakly informative multivariate normal prior distribution for the parameter and update using the data.

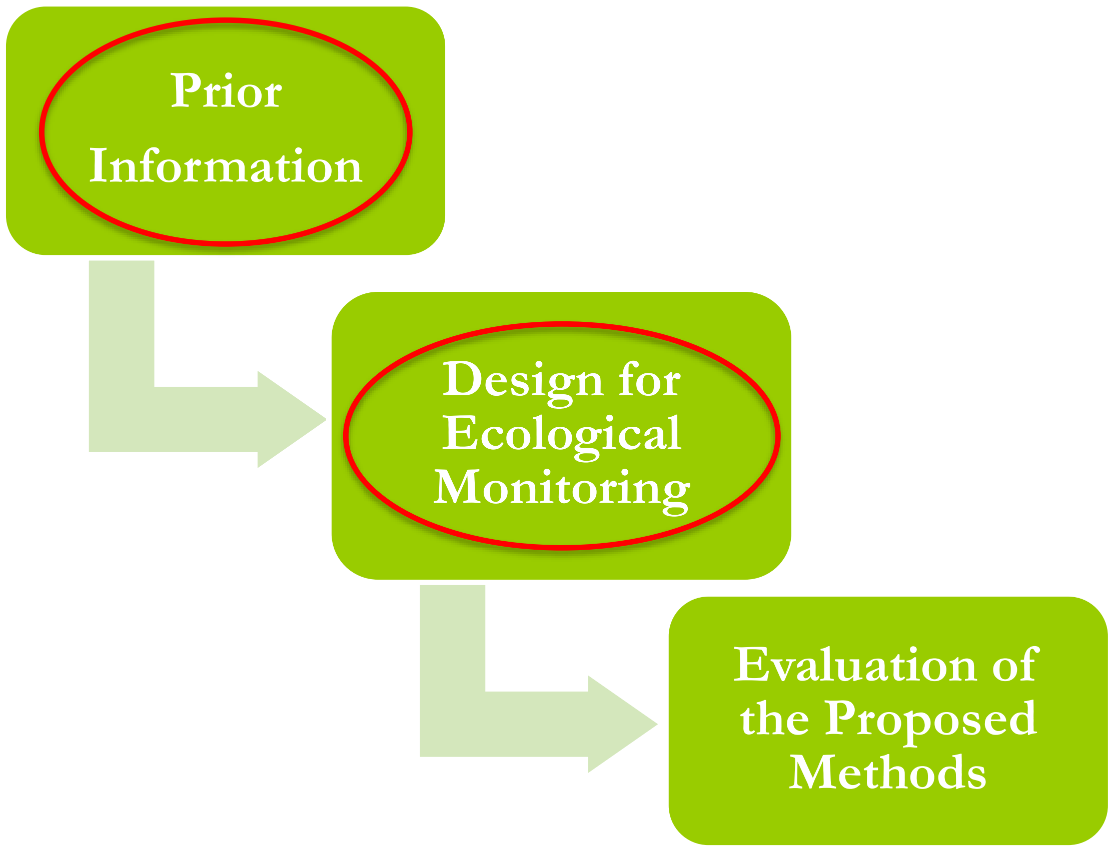


Prior information

Prior for the design

	Mean	Standard deviation	Lower bound of 95% credible interval	Upper bound of 95% credible interval
Intercept	-1.27	0.08	-1.43	-1.12
Time	-0.04	0.03	-0.10	0.01
Middle-shelf	0.15	0.08	-0.01	0.32
Outer-shelf	0.91	0.21	0.50	1.31
log CoTS	-0.01	0.01	-0.02	0.00
No-take	0.28	0.09	0.11	0.45
Cyclone	-0.45	0.05	-0.55	-0.35
Bleaching	-0.22	0.07	-0.35	-0.08
Bathymetry	-0.11	0.02	-0.15	-0.06
Chlorophyll	-0.80	0.10	-0.99	-0.61
CRS_T_AV	-0.23	0.05	-0.33	-0.13
Log variance	-2.52	0.04	-2.61	-2.44
Log partial sill	-5.98	0.48	-6.93	-5.03
Log range	-1.12	0.06	-1.24	-1.00

Proposed Bayesian design framework



- Comparison
- Objective 1
- Objective 2

Design for ecological monitoring

How can designs be evaluated?

- We collect data based on designs.
- To evaluate designs, we need to quantify how much information is in data.
- A utility function $\boldsymbol{u(d, y, \theta)}$ quantifies the worth of observing data $\boldsymbol{y}$ from design $\boldsymbol{d}$ in terms of achieving the specified monitoring objective/s.
- The expected utility function can be defined as follows:

$$u(\boldsymbol{d}) = \int_{\boldsymbol{y}} \int_{\boldsymbol{\theta}} u(\boldsymbol{d}, \boldsymbol{y}, \boldsymbol{\theta}) p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{d}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}\, d\boldsymbol{y}.$$

Design for ecological monitoring

Ecological monitoring

- There are additional uncertainties associated with where and when the time-varying disturbances will occur.
- To account for these additional uncertainties, the above expected utility can be extended as follows:

$$u(\boldsymbol{d}) = \int_{\boldsymbol{y}} \int_{\boldsymbol{\theta}} \int_{\boldsymbol{z}} u(\boldsymbol{d}, \boldsymbol{z}, \boldsymbol{y}, \boldsymbol{\theta}) p(\boldsymbol{y} | \boldsymbol{\theta}, \boldsymbol{d}, \boldsymbol{z}) p(\boldsymbol{\theta}) p(\boldsymbol{z} | \boldsymbol{d}, \boldsymbol{\kappa}) d\boldsymbol{z}\, d\boldsymbol{\theta}\, d\boldsymbol{y},$$

where the expectation is now taken over the distribution of the time-varying covariates (as well as $\boldsymbol{\theta}$ and $\boldsymbol{y}$).

Design for ecological monitoring

Which utility to choose?

- The precise estimation of model parameters that describe the impact of disturbances.
- So, we need a parameter estimation utility (e.g. Kullback-Leibler divergence (KLD)).
- KLD measures the distance between the prior and posterior distributions:

$$u(\boldsymbol{d}, \boldsymbol{y}, \boldsymbol{z}) = \int_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{d}, \boldsymbol{z}) \log p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{d}, \boldsymbol{z}) d\boldsymbol{\theta} - \log p(\boldsymbol{y}|\boldsymbol{d}, \boldsymbol{z})$$

where $\log p(\boldsymbol{y}|\boldsymbol{d}, \boldsymbol{z}) = \int_{\boldsymbol{\theta}} \log p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{d}, \boldsymbol{z}) d\boldsymbol{\theta}$ is the marginal likelihood.

Design for ecological monitoring

Approximating the expected utility

Unfortunately, in general, the above expectation does not have a closed form solution, and therefore needs to be approximated.

One common approach is via Monte Carlo integration as follows:

$$\hat{u}(\boldsymbol{d}) = \frac{1}{T}\sum_{t=1}^{T} u\left(\boldsymbol{d}, \boldsymbol{z}^{(t)}, \boldsymbol{\theta}^{(t)}, \boldsymbol{y}^{(t)}\right),$$

where T is the controlling parameter of Monte Carlo approximation, $\boldsymbol{\theta}^{(t)} \sim p(\boldsymbol{\theta})$, $\boldsymbol{z}^{(t)} \sim p(\boldsymbol{z}|\boldsymbol{d}, \boldsymbol{\kappa})$, $\boldsymbol{y}^{(t)} \sim p\left(\boldsymbol{y}\middle|\,\boldsymbol{\theta}^{(t)}, \boldsymbol{d}, \boldsymbol{z}^{(t)}\right)$.

Design for ecological monitoring

Simulate time-varying covariates

Site number	Bleaching proportion	Cyclone proportion	CoTS proportion	Log CoTS mean	Log CoTS standard deviation
1	0.12	0.25	0.62	-4.44	2.80
2	0.12	0.25	0.62	-4.60	2.99
3	0.12	0.25	0.62	-4.83	3.27
4	0.12	0.12	0.62	-3.84	2.88
5	0.12	0.12	0.62	-3.82	2.86
6	0.12	0.12	0.62	-3.80	2.85
7	0.12	0.12	0.62	-4.31	2.85
8	0.12	0.12	0.62	-4.29	2.81
9	0.12	0.12	0.62	-4.30	2.83
10	0.12	0.37	0.75	-5.45	3.90
11	0.13	0.38	0.77	-5.28	3.73
12	0.12	0.37	0.75	-5.28	3.73
13	0.12	0.37	0.75	-9.70	1.41
14	0.12	0.37	0.75	-9.62	1.40
15	0.12	0.37	0.75	-9.39	1.40
16	0.12	0.50	0.75	-6.90	2.43
17	0.13	0.49	0.77	-6.65	2.27
18	0.12	0.50	0.75	-6.43	2.12
19	0.12	0.25	0.75	-8.45	1.47
20	0.12	0.25	0.75	-8.14	1.47
21	0.12	0.25	0.75	-7.96	1.47
22	0.12	0.37	0.75	-7.17	2.18
23	0.12	0.37	0.75	-6.86	2.05
24	0.12	0.37	0.75	-6.62	1.96
25	0.12	0.37	0.75	-8.26	2.28
26	0.12	0.37	0.75	-8.26	2.28
27	0.13	0.36	0.77	-8.26	2.28

Design for ecological monitoring

Simulate coral cover data

Once the time-varying covariates were generated, the following model was used to simulate the coral cover proportion data:

$$\text{logit}(\mu_{ijk}) = \beta_0 + \beta_1 \text{Middle-shelf}_{ijk} + \beta_2 \text{Outer-shelf}_{ijk} + \beta_3 \text{No-take}_{ijk}$$

$$+\beta_4 \text{Cyclone}_{ijk} + \beta_5 \text{Bleaching}_{ijk} + \beta_6 \text{log CoTS}_{ijk} + \beta_7 \text{Bathymetry}_{ijk} +$$
$$\beta_8 \text{Chlorophyll}_{ijk} + \beta_9 \text{CRS\_T\_AV}_{ijk} + \beta_{10} \text{Time}_{ijk} + r_{ijk}\,.$$

Design for ecological monitoring

Optimising the design

$$\hat{u}(\boldsymbol{d}) = \frac{1}{T}\sum_{t=1}^{T} u\left(\boldsymbol{d}, \boldsymbol{z}^{(t)}, \boldsymbol{\theta}^{(t)}, \boldsymbol{y}^{(t)}\right),$$

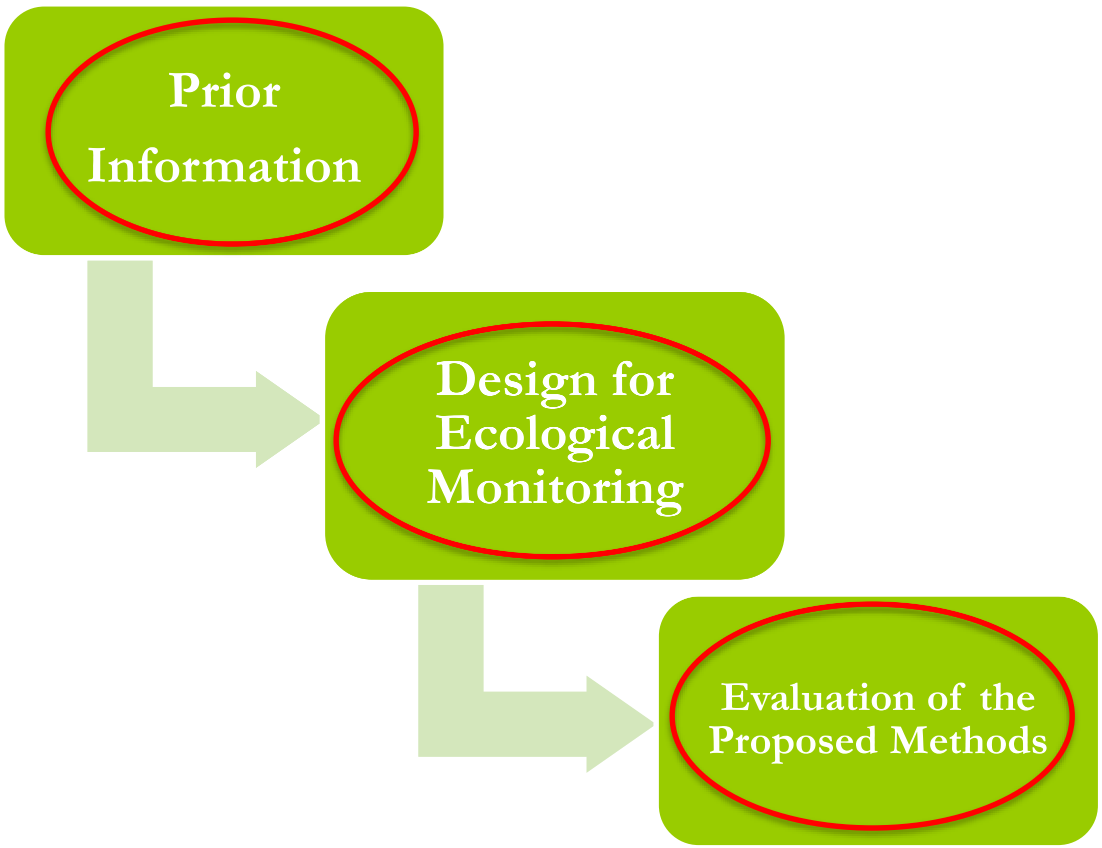
$$\boldsymbol{d}^{*} = \arg\max_{\boldsymbol{d}\in\boldsymbol{D}} \hat{u}(\boldsymbol{d})$$

In our reef monitoring scenarios, an optimal design will define which sites should be visited in the next year.

As such, the design will be discrete with a finite number of potential solutions.

In such situations, the coordinate-exchange algorithm can be used to maximise the expected utility function.

Proposed Bayesian design framework



- Comparison
- Objective 1
- Objective 2

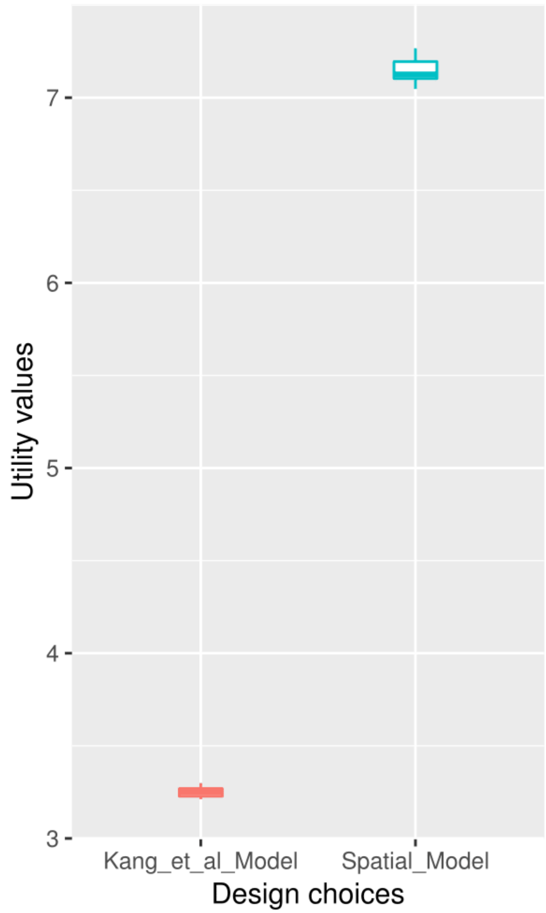
Evaluating the framework

Comparison between Kang et al. (2016) and spatial model

We compared designs based on Kang et al. (2016) model and our spatial model.

Corresponding efficiency is approximately 47%.

Information gained from designs based on spatial model is almost twice the amount compared to designs based on Kang et al. (2016) model.

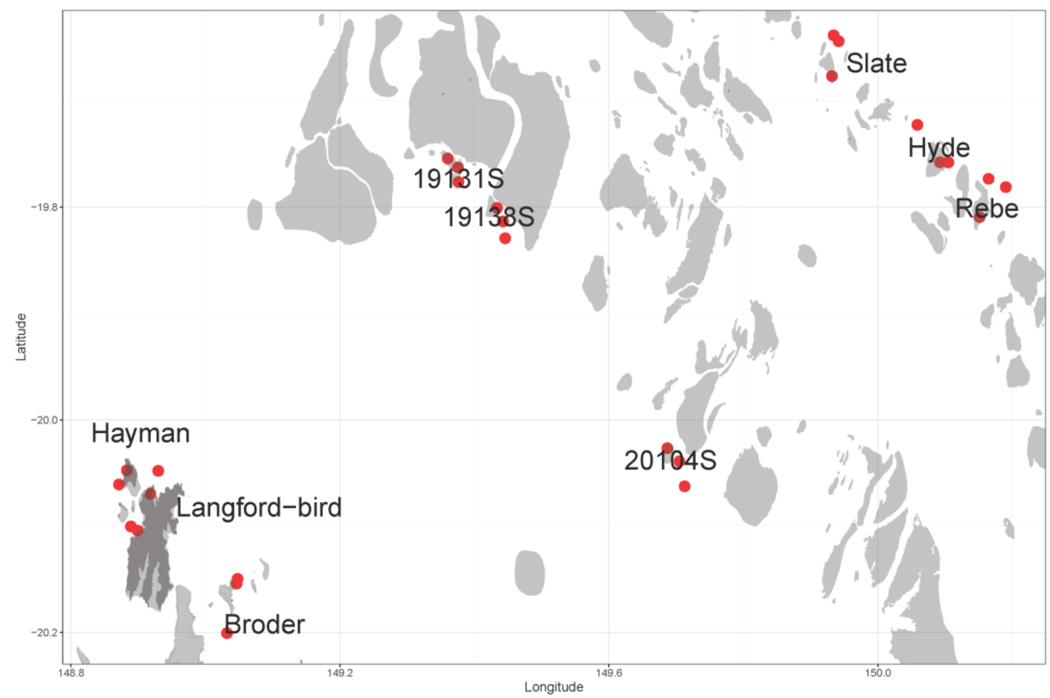


Evaluating the framework

Objective 1: Effect of having less LTMP sites

We follow two approaches;

- Drop reefs one by one;
- Drop one site from each reef.



Effect of having less LTMP sites

Dropping reefs

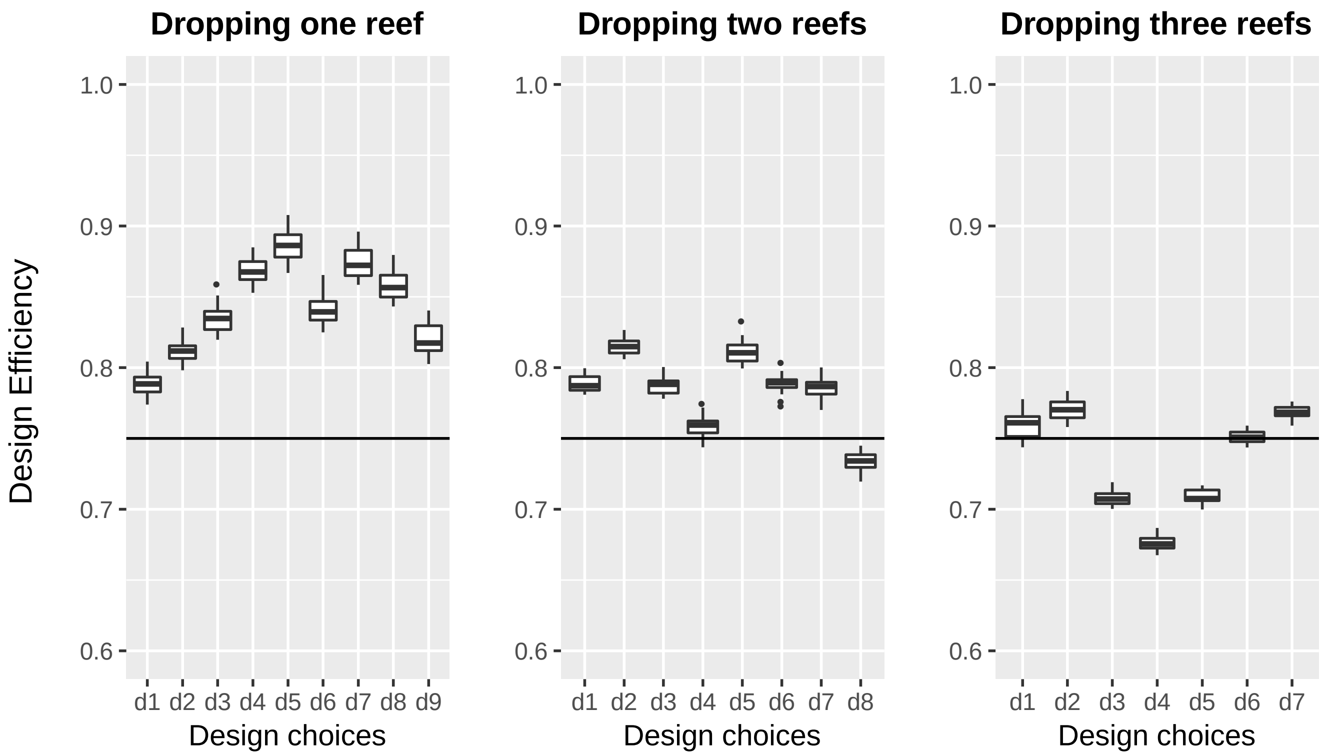


Figure 1: Efficiencies of designs after dropping one (left), two (middle), and three (right) reef/reefs in the Whitsunday region. The black horizontal line represents 75% efficiency level.

The design without Hayman Island reef (left) and the design without both Hayman Island reef and Rebe reef (middle) still retain around 89% and 81% efficiency, respectively.

Even after dropping three reefs, there are designs with more than 75% efficiency (right).

Effect of having less LTMP sites

Dropping reefs (Cont.)

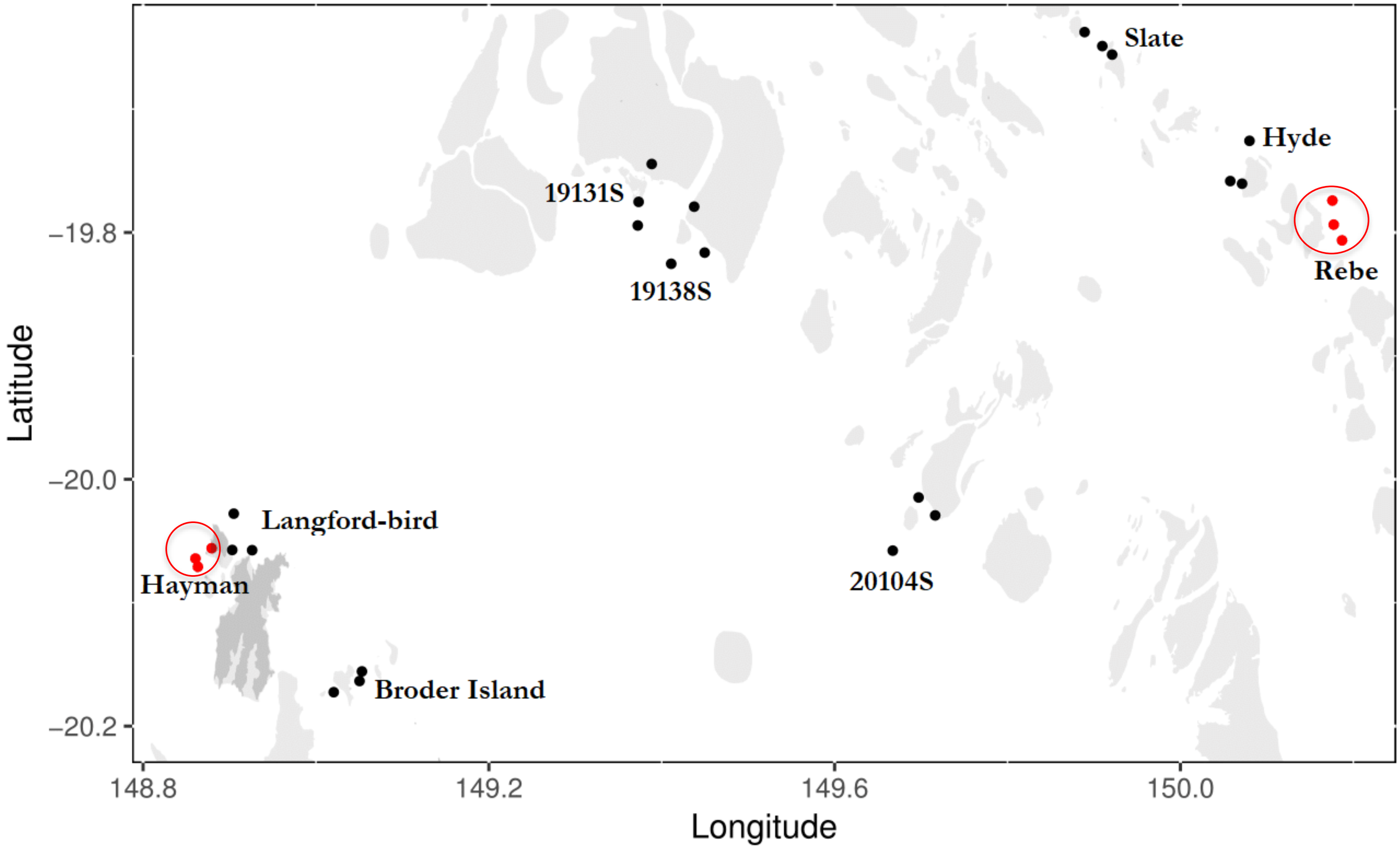


Figure 2: Spatial locations of sites on the reefs in the Whitsunday region. Sites on the two least informative reefs are displayed using red colour.

Effect of having less LTMP sites

Dropping sites

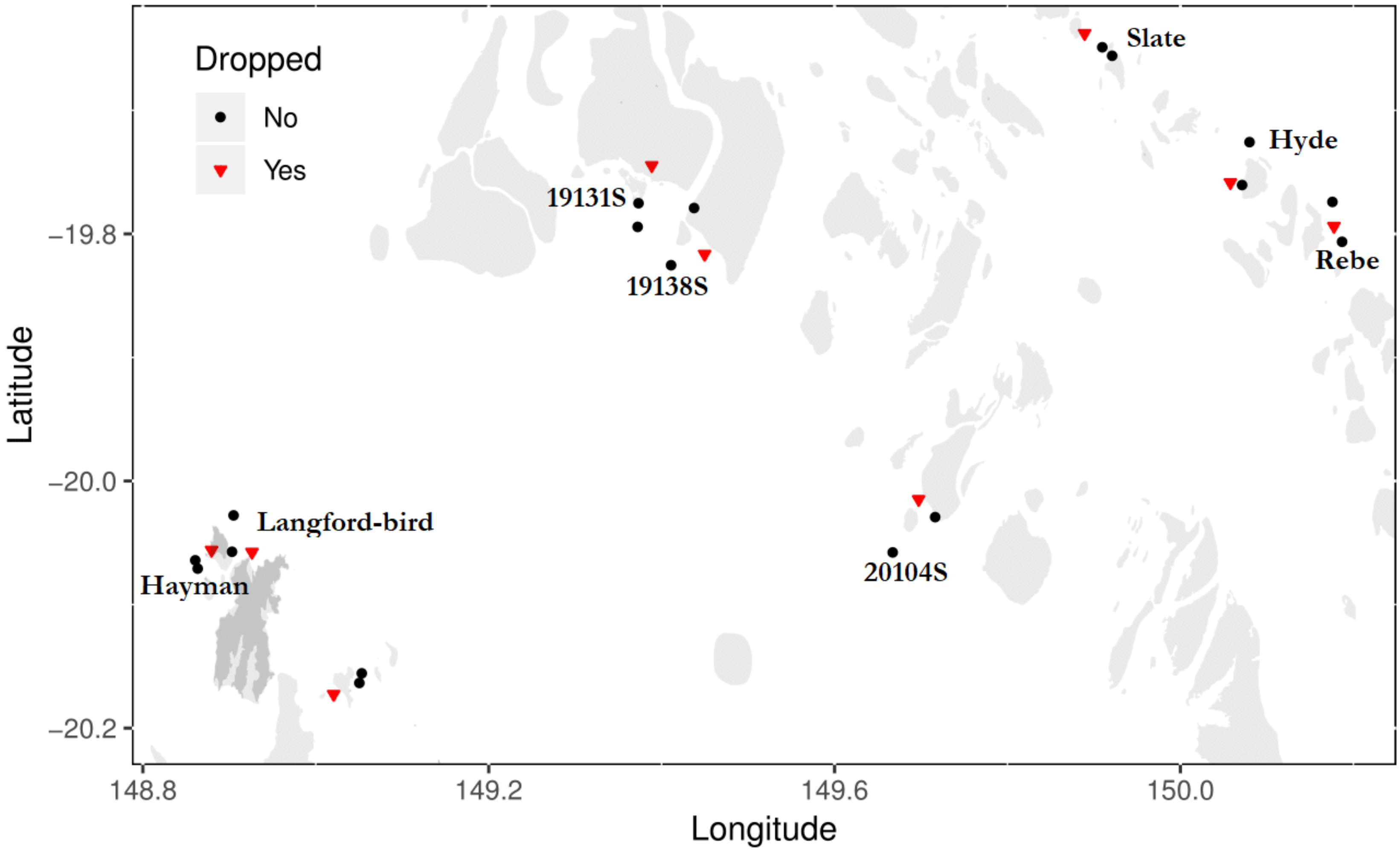


Figure 3: Spatial locations of retained/dropped sites on each reef.

Resulting the best design has approximately 85% efficiency.

Evaluating the framework

Objective 2: Designs depending on the disturbances

We will examine the ways in which an optimal design can be found based on the reef condition.

To evaluate this sampling framework, a range of disturbance patterns were simulated and designs were found based on these patterns.

- Scenario 1: A pattern consistent with historical disturbance data.
- Scenario 2: CoTS disturbances are simulated on selected sites.

Evaluating the framework

Why certain reefs/sites were selected?

Given the optimal designs we have found, it is important to consider why certain reefs/sites were selected over others.

There can be one or more contributing factors towards it:

- Distance between reefs/sites (spatial effect in the model);
- Differences in covariate values between reefs/sites;
- Prior uncertainty about estimated coefficients.

Here, we will try to present some potential reasons for selecting certain reefs/sites based on these factors.

Designs depending on the disturbances

Scenario 1

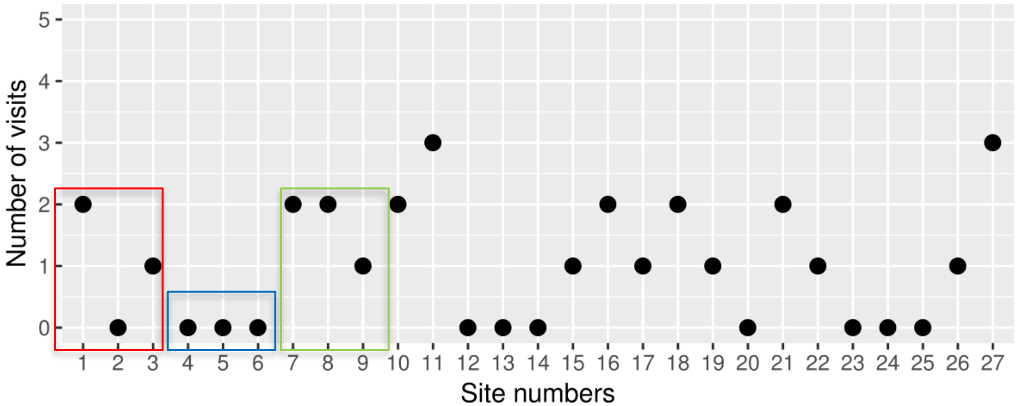


Figure 4: The optimal design sites and the number of visits to each site.

Reef names	Shore	Site numbers
Broder Island reef	I	1,2,3
Langford-bird reef	I	4,5,6
Hayman Island reef	I	7,8,9
20104S	M	10,11,12
19138S	M	13,14,15
Rebe reef	O	16,17,18
19131S	M	19,20,21
Hyde reef	O	22,23,24
Slate reef	O	25,26,27

The design we found does not visit all the LTMP sites in the region, but it collects more data from some selected sites.

Designs depending on the disturbances

Scenario 1 (Cont.)

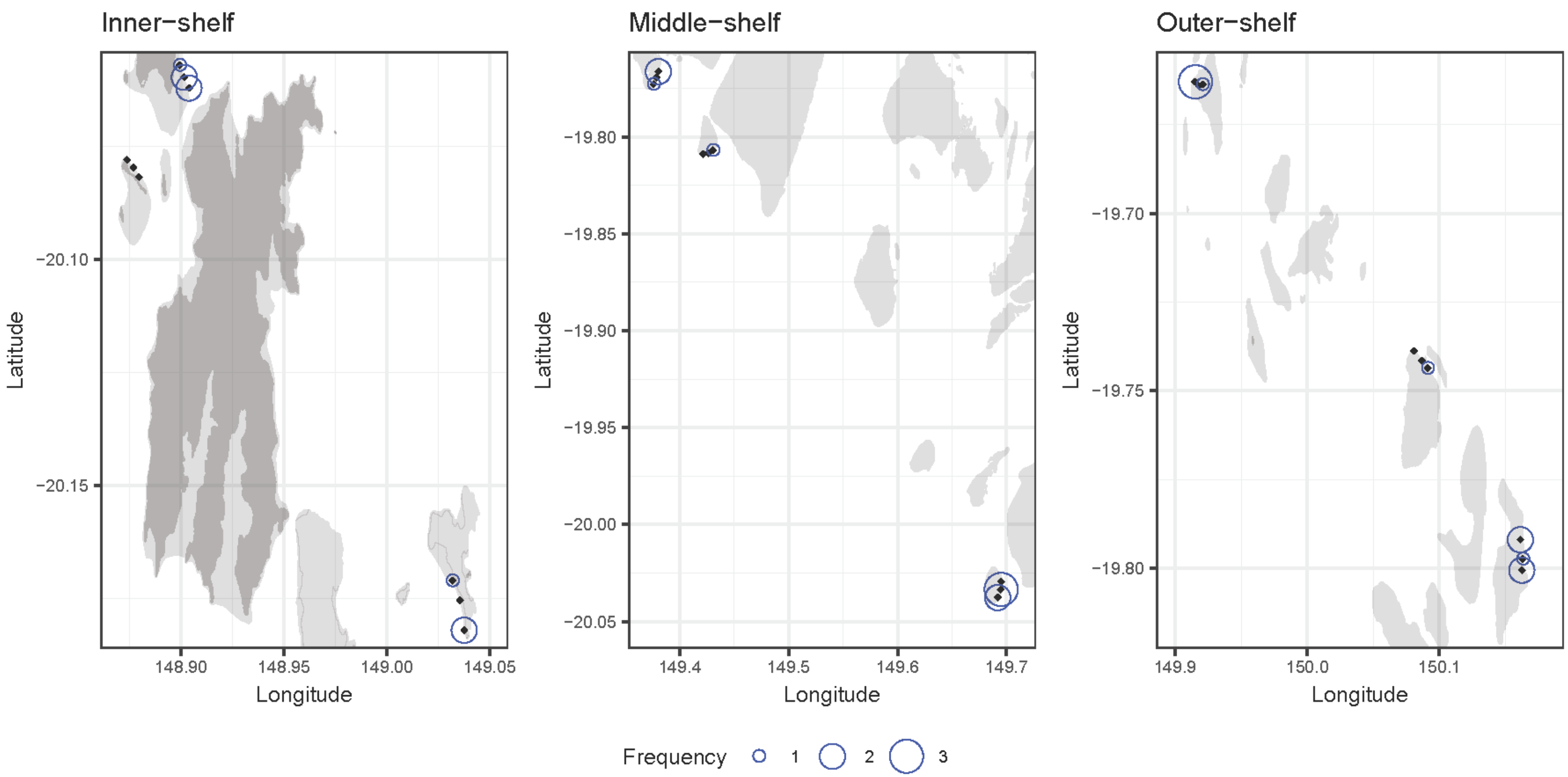


Figure 6: Spatial locations of the optimal design's (Scenario 1) sites in the Whitsunday region on the GBR.

Significance

- ✓ Demonstrate the use of time-varying covariates information in deciding sampling locations for the coming year.
- ✓ Demonstrate the ability to identify less informative reefs/sites:
    - Two reefs can be disregarded without a substantial loss in information about coral cover.
    - One site can be neglected from each reef, while still retaining 85% of the information.
- ✓ Could provide highly informative data compared to the current LTMP design in order to achieving specified monitoring objectives.

References

1. Diggle, P., and Lophaven, S. (2006). Bayesian Geostatistical Design. *Scandinavian Journal of Statistics*, 33(1), 53-64. doi:10.1111/j.1467-9469.2005.00469.x.

2. Kang, S. Y., McGree, J. M., Drovandi, C. C., Caley, M. J., and Mengersen, K. L. (2016). Bayesian adaptive design: improving the effectiveness of monitoring of the Great Barrier Reef. *Ecological Applications*, 26(8), 2635-2646. doi:10.1002/eap.1409.

3. Lagos-Alvarez, B. M., Fustos-Toribio, R., Figueroa-Zúñiga, J., and Mateu, J. (2017). Geostatistical mixed beta regression: a Bayesian approach. *Stochastic Environmental Research and Risk Assessment*, 31(2), 571-584.

4. Loredo, T. J., Berger, J. O., Chernoff, D. F., Clyde, M. A., and Liu, B. (2012). Bayesian methods for analysis and adaptive scheduling of exoplanet observations. *Statistical Methodology, 9*(1-2), 101-114. doi:10.1016/j.stamet.2011.07.005.

5. Lovett, G. M., Burns, D. A., Driscoll, C. T., Jenkins, J. C., Mitchell, M. J., Rustad, L., . . . Haeuber, R. (2007). Who Needs Environmental Monitoring? *Frontiers in Ecology and the Environment, 5*(5), 253-260. doi:10.1890/1540-9295(2007)5[253:WNEM]2.0.CO;2.

6. McGree, J. M. (2017). Developments of the total entropy utility function for the dual purpose of model discrimination and parameter estimation in Bayesian design. *Computational Statistics and Data Analysis* , 113, 207-225. doi:10.1016/j.csda.2016.05.020.

7. McGree, J., Drovandi, C. C., and Pettitt, A. N. (2012). A sequential Monte Carlo approach to the sequential design for discriminating between rival continuous data models.

8. Morgan, C. C., Huyck, S., Jenkins, M., Chen, L., Bedding, A., Coffey, C. S., Wathen, J. K. (2014). Adaptive Design: Results of 2012 Survey on Perception and Use. *Therapeutic Innovation & Regulatory Science, 48*(4), 473-481. doi:10.1177/2168479014522468.

9. Price, D. J., Bean, N. G., Ross, J. V., and Tuke, J. (2016). On the efficient determination of optimal Bayesian experimental designs using ABC: A case study in optimal observation of epidemics. *Journal of Statistical Planning and Inference, 172*, 1-15. doi:10.1016/j.jspi.2015.12.008.

10. Ryan, E. G., Drovandi, C. C., and Pettitt, A. N. (2015). Fully Bayesian Experimental Design for Pharmacokinetic Studies. *Entrophy, 17*(3), 1063-1089. doi:10.3390/e17031063.

11. Vercelloni, J., Caley, M. J., Kayal, M., Low-Choy, S., and Mengersen, K. (2014). Understanding Uncertainties in Non-Linear Population Trajectories: A Bayesian Semi-Parametric Hierarchical Approach to Large-Scale Surveys of Coral Cover. *PLOS ONE*, 9(11), e110968. doi:10.1371/journal.pone.0110968.

Thank You!