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

Pubudu Thilan, James McGree, Erin Peterson, Christopher Drovandi

School of Mathematical Sciences

Queensland University of Technology

Outline

- Motivation
- Objectives
- Background
- Methodology
- Results
- Conclusions and Future Works
- References

Motivation

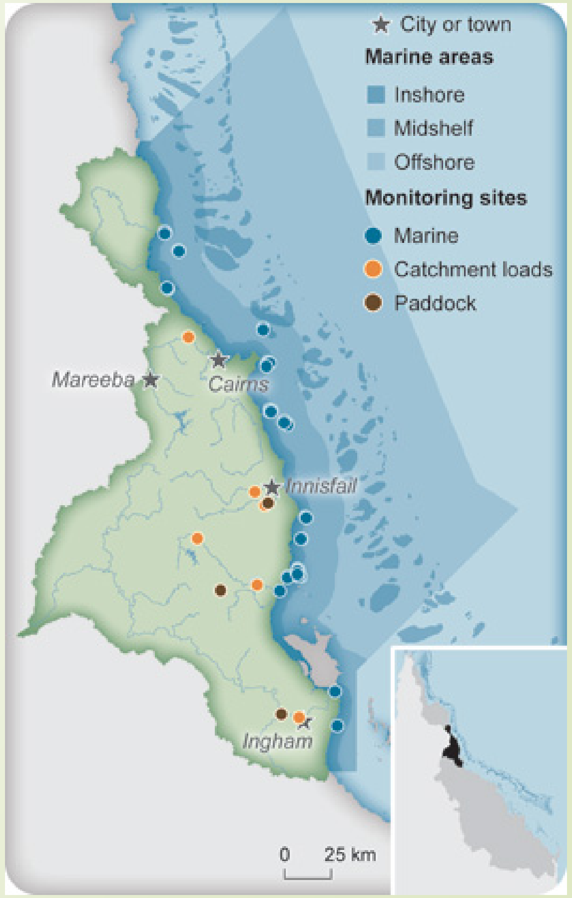
- Reef monitoring programs are a key instrument in identifying patterns, trends, and threats to coral reef systems.
- However, implementation and maintenance of such programs are highly expensive.
- Much of the current work on developing cost effective monitoring programs pays particular attention on using adaptive design methods.

Objectives

- Our objective is to develop a model-based adaptive design to improve the effectiveness of reef monitoring. This work is in progress and I will show some initial results.

What is Design?

- The choice of sampling locations in space and time.
- There are two basic types of designs.
- Static designs – Designs which remain fixed over time.
- Adaptive designs – Designs which change over time.



Finding the Optimal Design

- The goal of the design phase is to form a new design (e.g., locations) that will provide new data to inform our objectives.
- Optimal experimental designs may be used to achieve the experimental goals more rapidly and hence reduce experimental costs.
- We choose the optimal design based on some utility functions.

Utility functions

- A utility function *u(d, y, θ)* represents the expected worth of the experimental data obtained under the design ***d*.**
- The aim is to find the optimal design which maximises the expected value of the utility function:

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

where *θ*-model parameter, *p*(*θ*)-prior distribution, *p(y/θ,d)*– likelihood of unobserved data given that the design *d* is applied.

- For example, maximizing inverse of the average prediction variance.

Model

- The impact of some environmental variables on coral reef has a temporal and spatial variability.
- Coral cover model should capture this variability well.
- We assume geostatistical mixed beta regression model;

$$\eta_i := G\left(E(y_i|\Theta)\right) = x_i^T \beta_\mu + z_i, i = 1, \dots, n,$$

$$\xi_i = H(\psi_i) = s_i^T \beta_\psi.$$

where η_i-linear mixed model for mean, ξ_i-linear model for precision, z_i-random effect terms, $z_i|\Sigma_{z_i} \sim \mathrm{MVN}(0,\Sigma_{z_i})$,where Σ_{z_i}comes from a covariance model.

Approximate the Utility

- Often utilities are functions of posterior and they do not have a closed form solution.
- A utility can be numerically approximated using MC integration as

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

where $\boldsymbol{\theta}^{(t)} \sim p(\boldsymbol{\theta})$ and $\boldsymbol{y}^{(t)}$ is drawn from $p(\boldsymbol{y}|\boldsymbol{\theta}^{(t)},\boldsymbol{d})$.

- A large number of posterior distributions need to be sampled to evaluate the expected utility of a given design.
- This renders the process much more computationally expensive than inference.

Approximate the Posterior

- Laplace approximation can quickly produce an approximation to the posterior using multivariate normal distribution.
- This can be expressed mathematically as

$$P(\theta|y,d) \approx N(\theta|\hat{f}, A^{-1})$$

where $\hat{f}$ denotes the mode of the posterior distribution and A denotes Hessian of the negative log posterior at $\hat{f}$.

Approximate the Likelihood

- Random effects are part of the model but cannot be part of the likelihood because they are not real data.
- The log likelihood of the model is

$$l(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{d}) = Log\int p\,(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{d},\boldsymbol{z})p(\boldsymbol{z})d\boldsymbol{z},$$

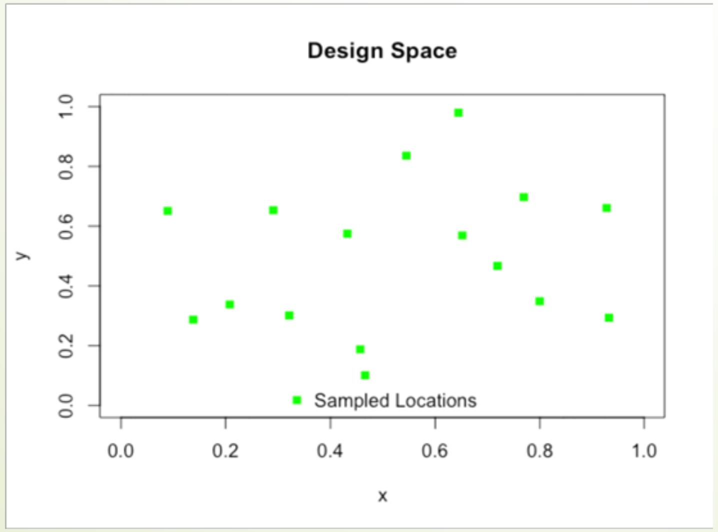
where $p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{d},\boldsymbol{z})$ is the distribution of responses given the random effects, $p(\boldsymbol{z})$ is the distribution of random effects.

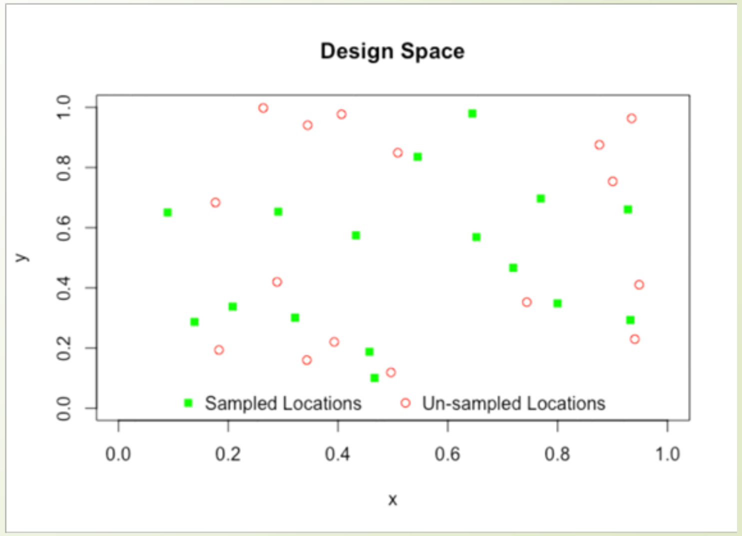
- Then the Monte Carlo log likelihood approximation is

$$l(\boldsymbol{\theta}) = Log\left(\frac{1}{m}\sum_{k=1}^{m} p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{d},z_k)\right).$$

Methodology

Prior → Data → Posterior → Posterior Predictive Distribution → Utility





Our Utility Approximation

Our utility is the inverse of the average prediction variance.

1. Simulate data from the posterior predictive distribution, $Y_k: k = 1, \dots, L$.
2. Calculate the variance at each site.

$$a_1 = VAR[\tilde{y}_1^{(1,1)}, \tilde{y}_1^{(1,2)}, \dots, \tilde{y}_1^{(1,L)}]$$

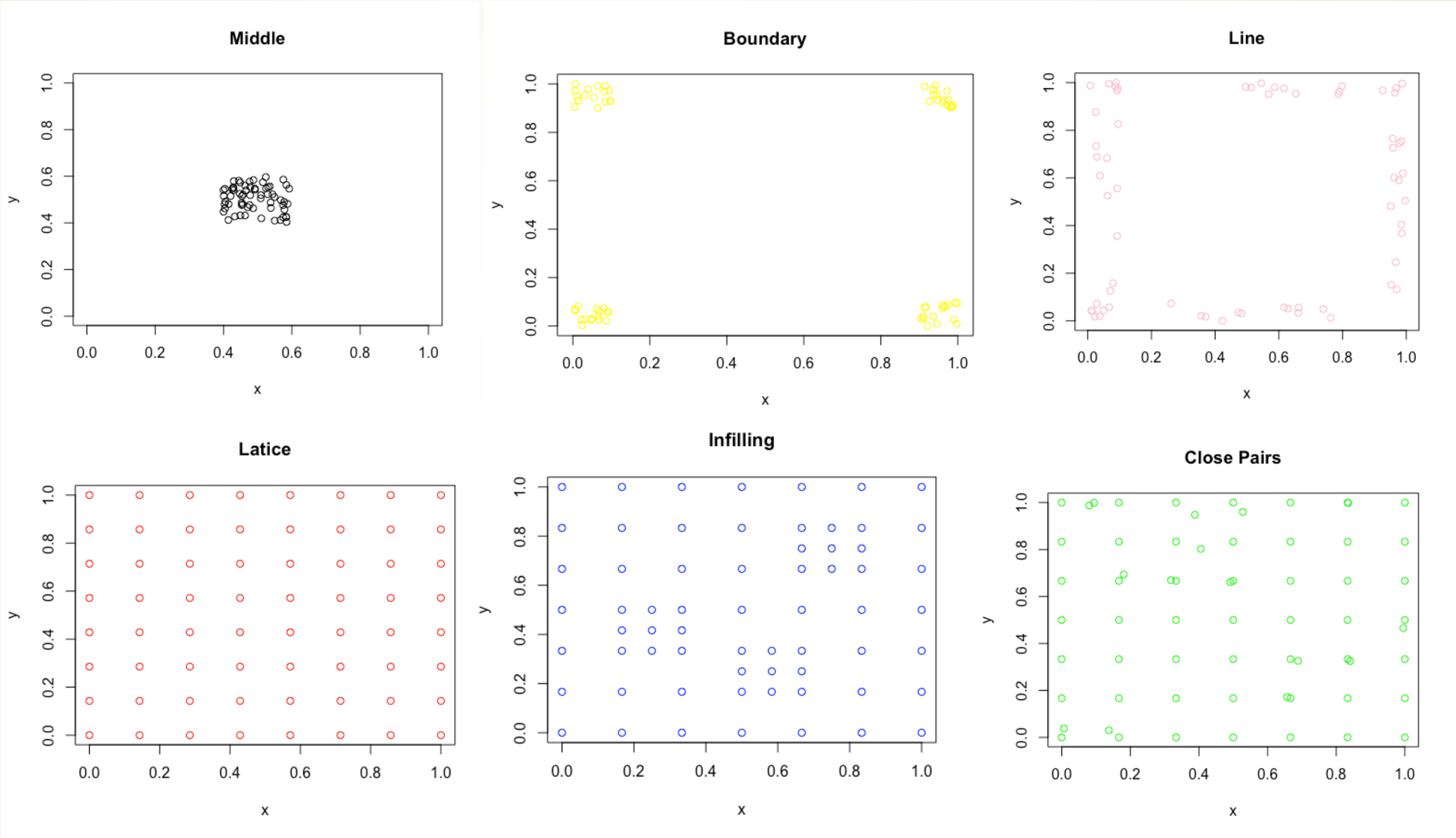
$$a_2 = VAR[\tilde{y}_2^{(1,1)}, \tilde{y}_2^{(1,2)}, \dots, \tilde{y}_2^{(1,L)}]$$

$$a_{n+n0} = VAR[\tilde{y}_{n+n0}^{(1,1)}, \tilde{y}_{n+n0}^{(1,2)}, \dots, \tilde{y}_{n+n0}^{(1,L)}]$$

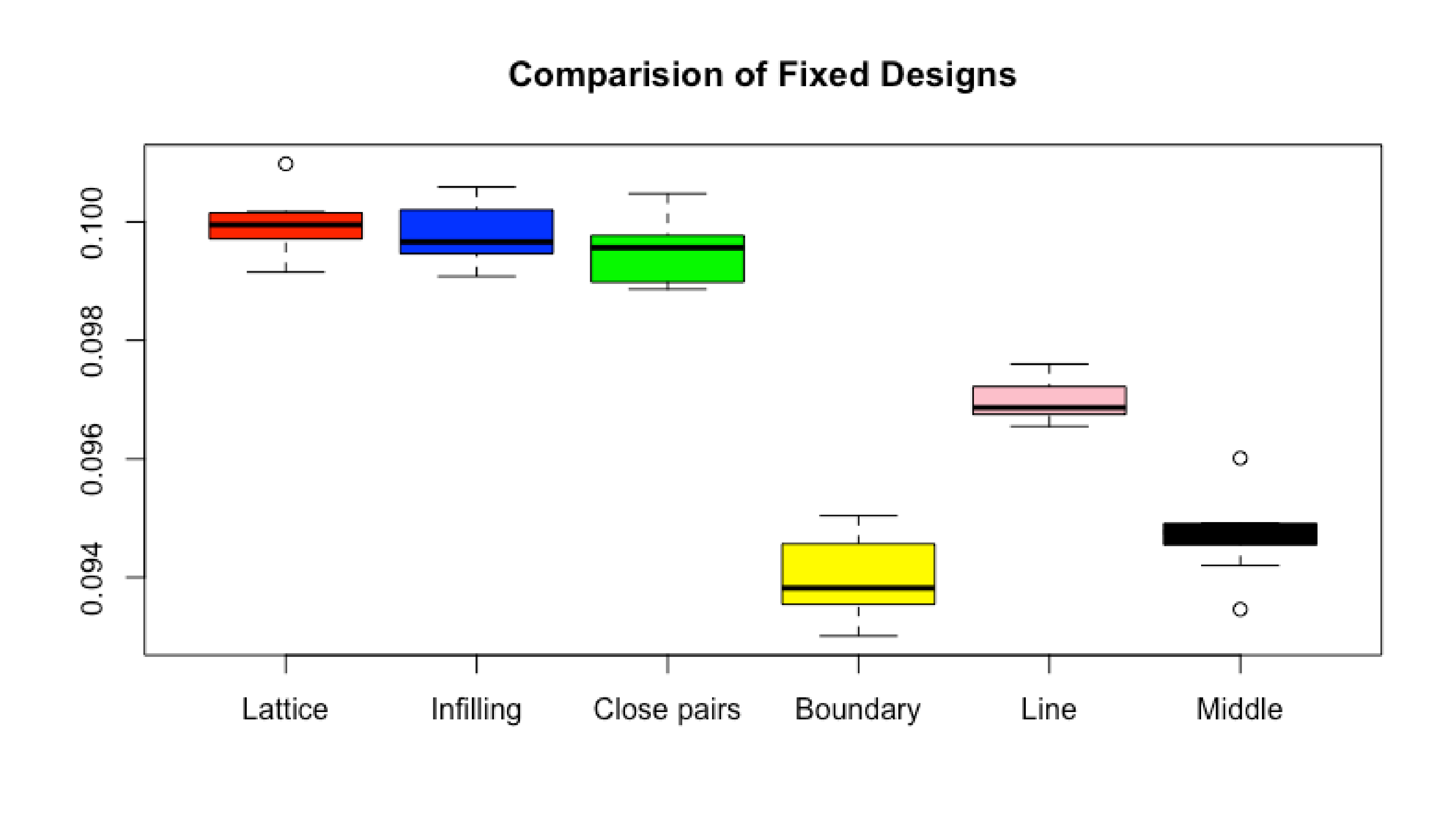
n-# of sampled sites
n_0-# of un-sampled sites

3. The utility is $u(d, y) = \frac{1}{a_1 + a_2 + \cdots + a_{n+n_0}}$.

Comparisons of Fixed Designs



Results


Comparision of Fixed Designs
0.100
0.098
0.096
0.094
Lattice
Infilling
Close pairs
Boundary
Line
Middle

Discussion

- We compared a limited number of geometrically simple classes of design.
- These type of simple designs are easily explained to practitioners.

Conclusions and Future Works

- The general spatial structure of a design is more important than the precise location of each point within it.
- Therefore, this work can be extended to find the general spatial structure of a design using an optimization algorithm.

References

1. Coker, D. J., V. Chaidez and M. L. Berumen. 2017. "Habitat Use and Spatial Variability of Hawkfishes with a Focus on Colour Polymorphism in Paracirrhites forsteri." PLOS ONE 12 (1): e0169079. doi: 10.1371/journal.pone.0169079.
2. Diggle, P. J., & Ribeiro, P. J. (2007). Model-based Geostatistics (1. Aufl. ed.): Springer Science + Business Media.
3. Figueroa-Zuniga, J. I., Arellano-Valle, R. B., & Ferrari, S. L. P. (2013). Mixed beta regression: A Bayesian perspective. COMPUTATIONAL STATISTICS & DATA ANALYSIS, 61, 137-147. doi:10.1016/j.csda.2012.12.002.
4. Jensen, B. S., Nielsen, J. B., & Larsen, J. (2013, 2013). Bounded Gaussian process regression.
5. Kang, S. Y., McGree, J. M., Drovandi, C. C., Caley, M. J., & 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.
6. Lagos-Álvarez, B. M., Fustos-Toribio, R., Figueroa-Zúñiga, J., & Mateu, J. (2017). Geostatistical mixed beta regression: a Bayesian approach. Stochastic Environmental Research and Risk Assessment, 31(2), 571-584. doi:10.1007/s00477-016-1308-5.

Thank You!