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.
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.
A utility function $u(d, y, \theta)$ 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:

$$U(d) = \int \int u(d, y, \theta) p(y|\theta, d)p(\theta) d\theta dy$$

where $\theta$-model parameter, $p(\theta)$-prior distribution, $p(y|\theta, d)$-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;

\[
\eta_i = G(E(y_i|\Theta)) = x_i^T \beta_\mu + z_i, i = 1, \ldots, n,
\]

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

where \( \eta_i \) - linear mixed model for mean, \( \xi_i \) - linear model for precision, \( z_i \) - random effect terms, \( z_i | \Sigma_{z_i} \sim MVN(0, \Sigma_{z_i}) \), where \( \Sigma_{z_i} \) 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
  \[
  U(d) = \frac{1}{T} \sum_{t=1}^{T} U(d, y^{(t)}, \theta^{(t)}),
  \]
  where \( \theta^{(t)} \sim p(\theta) \) and \( y^{(t)} \) is drawn from \( p(y | \theta^{(t)}, d) \).
- 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

\[ l(\theta|y, d) = \log \int p(\theta|y, d, z)p(z)dz \],

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

- Then the Monte Carlo log likelihood approximation is

\[ l(\theta) = \log \left( \frac{1}{m} \sum_{k=1}^{m} p(y|\theta, d, z_k) \right) \].
Methodology

1. Prior
2. Data
3. Posterior
4. Posterior Predictive Distribution
5. Utility

![Design Space with Sampled Locations](image1)

![Design Space with Sampled and Un-sampled Locations](image2)
Our Utility Approximation

Our utility is the inverse of the average prediction variance.

1. Simulate data from the posterior predictive distribution, $Y_k: k = 1, \ldots, L$.

2. Calculate the variance at each site.

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

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

$$a_{n+n_0} = VAR[\tilde{y}_{n+n_0}^{(1,1)}, \tilde{y}_{n+n_0}^{(1,2)}, \ldots, \tilde{y}_{n+n_0}^{(1,L)}]$$

$n$-# of sampled sites

$n_0$-# of un-sampled sites

3. The utility is $u(d, y) = \frac{1}{a_1 + a_2 + \ldots + a_{n+n_0}}$. 
Comparisons of Fixed Designs
Results

Comparison of Fixed Designs
Discussion

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


Thank You!