Robust optimal observation of a metapopulation

Pagendam, D.E. [1] and **P.K. Pollett** [1]

[1] *Department of Mathematics, University of Queensland, St Lucia, Qld 4072, Australia*
Email: pagendam@maths.uq.edu.au

Abstract: A metapopulation consists of interacting populations, each occupying distinct spatially separated patches of habitat. Modelling these populations has become increasingly important because anthropogenic impacts on spatially homogeneous populations have led to increased habitat fragmentation and accidental introduction of invasive species. We employ a two-parameter continuous-time Markovian model for patch occupancy, which has been used previously to study the spread of disease in closed populations. This model takes into account proximity of patches, rather than any detailed spatial structure. Our model will be particularly suited to studying species occupying marine environments that are subject to complex ocean currents, where the connectivity between habitat patches is dynamic, and most applicable to invertebrate and plant species that are dispersed passively by these currents.

We address the problem of designing an optimal sampling scheme that specifies when to observe the population in order to obtain the most accurate and precise estimates of the parameters. Our approach is based on approximating the true likelihood of observing the numbers of occupied patches at a particular sequence of times by a much simpler Gaussian likelihood, obtained from a diffusion approximation of the underlying discrete-state Markov process. This approximation is known to be highly accurate when the number of patches is sufficiently large. Furthermore, when coupled with an appropriate optimization technique such as the Cross-Entropy Method (used here), the approximation gives rise to a robust procedure for determining the optimal observation schedule, one which is considerably simpler than would otherwise be possible.

We investigate the performance of two design criteria: ED-optimality and a particular Maximin-optimality criterion. ED optimal design stems from the well known D-optimal design, but where the design is chosen to maximize the expected value of the D-optimality criterion (the determinant of the Fisher Information matrix), given priors on the parameters. The Maximin-optimal design maximizes the minimum value of the D-optimality criteria given the priors. These criteria can yield robust designs, which allow one to incorporate prior belief about the parameter values before any data is collected. We investigate 'informative' and 'less informative' priors for each parameter in combination with both design criteria. By assuming a true parameter set for a given metapopulation, we are able to compare kernel density estimates of the maximum likelihood estimator under the various combinations of optimality criteria and prior distributions. We illustrate our methods with reference to a model for the spread of crown of thorns starfish (*Acanthaster planci*) amongst the 55 islands comprising the Ryukyu group in Japan.

Our results suggest that ED-optimality criteria can sometimes lead to bimodality in distribution of the maximum likelihood estimator. However, estimation appears to be greatly improved if the ED criteria is used in conjunction with less informative priors for the parameter that governs the rate of drift towards equilibrium. Interestingly, our Maximin-optimality criteria does not give rise to such bimodality, but does result in a considerable loss in precision.

Keywords: Metapopulation; Markov process; Optimal observation

1. INTRODUCTION

In recent decades metapopulations have become a focal point of ecological research, as scientists grapple with understanding the consequences for species survival of habitat fragmentation and the spread of invasive species (Dobson 2003, Gilpin and Hanski 1991, Hanski 1999). Here we employ a simple two-parameter metapopulation model, which has been used widely in epidemiology, known as the *SIS Epidemic* (Susceptible to Infected to Susceptible) (Weiss and Dishon 1971), and more recently in ecology, being a stochastic analogue (Pollett 2001, Ross et al. 2006, Ross et al. 2009) of the standard Levins model (Levins 1969). One of the challenges practitioners face is obtaining precise estimates of the parameters from observational data. We present methods for obtaining an optimal observation schedule that are robust to our initial uncertainty about the true parameter values, through the use of appropriate prior probability distributions. We employ informative and relatively diffuse priors for the two parameters of our model to create four combinations of jointly independent priors. The priors are examined in conjunction with two optimality criteria used in experimental design: (i) the ED-optimality criterion (Pronzato and Walter 1985) and (ii) a Maximin-optimality criterion (Pronzato and Walter 1988).

The most common optimality criteria used in experimental design are based on the Fisher Information Matrix, which is defined in terms of the likelihood of a set of observations under a given model. Ordinarily, for models such as ours, the likelihood is difficult to obtain analytically. We therefore rely on approximations, utilizing a Gaussian diffusion approximation of the underlying discrete-state Markov process. This method has been used successfully for parameter estimation for a number models commonly used in ecology and population biology (Ross et al. 2006, Ross et al. 2009), and which we exploit here in concert with the Cross-Entropy (CE) Method to find the optimal observation times.

We use simulation to demonstrate how incorporating prior information can change attributes of the maximum likelihood estimator. The spread of the crown of thorns starfish (*Acanthaster planci*) in the Ryukyu islands of Japan is used as an illustrative example of the methods. Recommendations are made, that apply specifically to our model, regarding the best choice of optimality criteria and prior probability densities.

2. MODEL

Let N denote the number of patches of habitat. Each patch is identified as either *occupied* or *unoccupied*. There are two parameters, λ and μ (both > 0), being, respectively, the colonization rate and the local extinction rate. Our model is a birth-death process whose state n represents the number of occupied patches. Its state space is $S = \{0, 1, 2, \ldots, N\}$ and its non-zero transition rates are $q(n, n+1) = \lambda n(1 - n/N)$ and $q(n, n-1) = \mu n$.

Since our model assumes that the rate of colonization is independent of distances between patches, we advocate that it is particularly appropriate for studying species of invertebrates and plants in marine environments, where dispersal is passive (via ocean currents) and can follow complex dynamic patterns. Yamaguchi's (1986) description of *A. planci* infestations in the 55 islands of the Ryukyu group in Japan (where outbreaks have led to serious destruction of coral reef communities), suggests that our model with $N = 55$ (islands/patches) would be appropriate for studying the spread of the infestation.

It has been observed (Ross et al. 2006) that estimates of the parameters λ and μ are strongly correlated. This has two undesirable consequences: (i) it can be difficult to visualize confidence regions for (λ, μ), and (ii) it is difficult to identify appropriate prior distributions that incorporate the dependence between the parameters. Fortunately we can avoid these problems by re-parameterizing our model using $\alpha = \lambda - \mu$ ($\alpha > 0$) and $\rho = \mu/\lambda$ ($0 < \rho \leq 1$). The process typically exhibits two phases over the time-scale of interest: (i) an early growth phase, and (ii) an equilibrium phase (strictly speaking, *quasi*-equilibrium because extinction will happen eventually). These phases are illustrated in Figure 1. As will become apparent from (1) below, ρ determines the position of the equilibrium n_{eq} ($\approx N(1 - \rho)$) and α governs the rate of drift towards it.

3. METHODS

Approximating the likelihood. Maximum likelihood is the most widely used statistical method of parameter estimation. It involves constructing a joint probability density function L, being the likelihood of observing a particular set of data, which here are the numbers of occupied patches observed at a particular sequence of times, given fixed parameter values, and then maximizing this function over all parameter values $\boldsymbol{\theta}$ in the parameter space $\boldsymbol{\Theta}$ (the set of all allowable parameter values). For our model, and indeed for any Markov process, the likelihood is given by $L(\boldsymbol{\theta}; n_1, \ldots, n_d | n_0) = \prod_{i=1}^{d} p(n_i, t_i | n_{i-1}, t_{i-1})$, where for us $\boldsymbol{\theta} = (\alpha, \rho)$, and where $p(n_i, t_i | n_{i-1}, t_{i-1})$ is the probability of the process making a transition from state n_{i-1} (occupied patches) at time t_{i-1} to n_i at time t_i, d is the number of observations, and n_0 is the initial state (at time $t_0 = 0$).
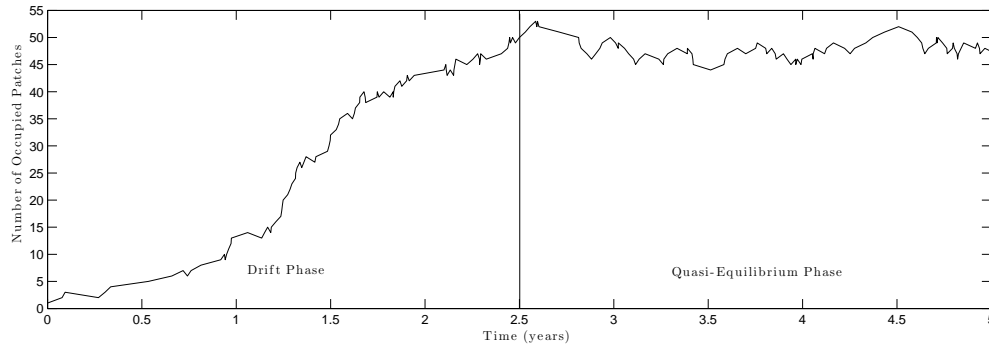
Figure 1. Simulated colonization of the Ryukyu islands by *A. planci* over a five year period using $N = 55$, $\lambda = 4$ yr^{-1} and $\mu = 0.5$ yr^{-1}.

Even for simple Markov processes such as birth-death processes, it is usually impossible to obtain analytical expressions for the transition probabilities. Furthermore, even when it is possible, these expressions may not allow the likelihood to be maximized easily. This is certainly true of our model. We therefore rely on a Gaussian diffusion approximation to obtain an approximating likelihood, which is valid in the limit as N becomes large. These approximations have been used successfully for a family of population models known as *density dependent* Markov chains, and we refer the reader to Pollett (2001), Ross et al. (2006) and Ross et al. (2009), and the references therein, for more information on their application in population biology. The likelihood from the resulting diffusion approximation is given by

$$L(\alpha, \rho; n_1, \ldots, n_d | n_0) = (2\pi)^{-d/2} |\mathbf{\Sigma}|^{-1/2} \exp\left(-\tfrac{1}{2}(\mathbf{n} - \mathbf{m})\mathbf{\Sigma}^{-1}(\mathbf{n} - \mathbf{m})'\right),$$

where $\mathbf{n} = (n_1, \ldots, n_d)$ is the vector of observed numbers of occupied patches at times $t_1, \ldots, t_d$ and $\mathbf{m} = (m(t_1), \ldots, m(t_d))$ is the corresponding vector of expected numbers obtained from the approximation:

$$m(t) := N \frac{x_0 x_{\text{eq}}}{x_0 + (x_{\text{eq}} - x_0)e^{-\alpha t}}, \tag{1}$$

where $x_0 := n_0/N$ is the initial proportion of occupied patches and $x_{\text{eq}} := 1 - \rho$ is the (quasi-) equilibrium proportion of occupied patches. The elements of the $d \times d$ covariance matrix $\mathbf{\Sigma}$ are

$$\Sigma_{i,j} = \text{Cov}(n(t_i), n(t_j)) = \begin{cases} Nv(t_i) & (i = j) \\ Nv(t_{\min(i,j)}) \left(\frac{x_0 + (x_{\text{eq}} - x_0)e^{-\alpha t_{\max(i,j)}}}{x_0 + (x_{\text{eq}} - x_0)e^{-\alpha t_{\min(i,j)}}}\right)^2 e^{-\alpha |t_i - t_j|} & (i \neq j) \end{cases}$$

where

$$\begin{aligned} v(t) = x_0 x_{\text{eq}} \Big[(1 + \rho) \Big(x_0^3(e^{2\alpha t} - 1) + 2(x_{\text{eq}} - x_0)^3(1 - e^{-\alpha t}) \\ - 6(x_{\text{eq}} - x_0)x_0^2(1 - e^{\alpha t}) + 6(x_{\text{eq}} - x_0)^2 x_0 \alpha t \Big) - x_0 x_{\text{eq}} \Big(x_0^2(e^{2\alpha t} - 1) - 4(x_{\text{eq}} - x_0)x_0(1 - e^{\alpha t}) \\ + 2(x_{\text{eq}} - x_0)^2 t\alpha \Big) \Big] / \Big[2(1 - \rho)e^{2\alpha t}(x_0 + (x_{\text{eq}} - x_0)e^{-\alpha t})^4 \Big]. \end{aligned}$$

Optimal observation. Carefully designed experiments promise the practitioner significant rewards over other designs having the same constraints: typically a greater experimental efficiency resulting from an increase in the precision of point estimates with fewer samples or observations. Optimal experimental design is a branch of statistics where the aim is to find a design that maximizes some objective function related to the efficiency of the experiment. Traditionally, optimal design has focussed on linear models and generalized linear models, with little attention given to optimal observation of stochastic processes. This is to the detriment of fields such as epidemiology and ecology where stochastic processes provide much more suitable models than many of the classical linear and non-linear models routinely adopted in statistics. The optimal observation of Markov processes has been considered by a small number of authors (see Cook et al. (2008), Pagendam and Pollett (2008), and the references therein), and has application in both natural and controlled experiments.

Traditional (frequentist) optimal designs generally focus on optimizing some function of the Fisher Information matrix, $\mathcal{I}(\boldsymbol{\theta}; \mathbf{y})$, which is the inverse covariance matrix of the maximum likelihood estimator. For us it is

approximated by way of the diffusion approximation and has entries

$$\mathcal{I}_{i,j} = \left(\frac{\partial \mathbf{m}}{\partial \theta_i}\right) \mathbf{\Sigma}^{-1} \left(\frac{\partial \mathbf{m}}{\partial \theta_j}\right)' + \tfrac{1}{2}\mathrm{trace}\left(\mathbf{\Sigma}^{-1} \left(\frac{\partial \mathbf{\Sigma}}{\partial \theta_i}\right) \mathbf{\Sigma}^{-1} \left(\frac{\partial \mathbf{\Sigma}}{\partial \theta_j}\right)\right).$$

We will assume that we have observed the state n_0 of our metapopulation at time $t_0 = 0$ and that we wish to schedule d further observations over the subsequent $t_{\max}$ years (for the numerical examples presented below, $d = 3$ and $t_{\max} = 5$). This translates to finding the optimal set of observation times $\mathbf{t}^\star = (t_1^\star, \ldots, t_d^\star)$ (where $0 < t_1 < t_2 < \cdots < t_d < t_{\max}$) in the design space $\mathbf{T}$ (the set of possible schedules) that minimizes the volume of the confidence ellipse for the maximum likelihood estimate of $\boldsymbol{\theta} = (\alpha, \rho)$. The most popular optimality criterion used in the literature is D-optimality:

$$\mathbf{t}^\star = \underset{\mathbf{t} \in \mathbf{T}}{\operatorname{argmax}} \det(\mathcal{I}(\mathbf{t}, \boldsymbol{\theta})).$$

The main problem for models such as ours is that the optimal design depends on the parameters. This means that in order to obtain the schedule which is optimal for estimating α and ρ, we already need to know their true values. The most obvious solution is to simply use a 'best guess'. However, if this initial guess is poor it can have potentially disastrous consequences. A much better option is to incorporate prior information into the experimental design to determine an optimal schedule that is robust to a wide range of true parameter values. Whilst the idea of using prior probability distributions stems from a Bayesian statistical paradigm, Pronzato and Walter (1985) and Pronzato and Walter (1988) introduced criteria for using priors in the design phase, but where the design criterion remains typically frequentist and where the data are collected with a view to analysis using traditional frequentist methods. Bayesian optimal design is another option; one might choose to maximize the Kullback-Leibler divergence between prior and posterior distributions for example (see Chaloner and Verdinelli 1995), but such designs will not be considered here.

We first consider the ED-optimality criterion proposed by Pronzato and Walter (1985), where the expected determinant of $\mathcal{I}$ under the prior distribution is maximized:

$$\mathbf{t}^\star = \underset{\mathbf{t} \in \mathbf{T}}{\operatorname{argmax}} \int \det(\mathcal{I}(\mathbf{t}, \boldsymbol{\theta})) p(\boldsymbol{\theta}) d\boldsymbol{\theta}.$$

Another robust optimal design criterion proposed by Pronzato and Walter (1988) is Maximin optimality. This acts to maximize the minimum value of $\mathcal{I}$ under the prior distribution:

$$\mathbf{t}^\star = \underset{\mathbf{t} \in \mathbf{T}}{\operatorname{argmax}} \, \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\min} \, \det(\mathcal{I}(\mathbf{t}, \boldsymbol{\theta})).$$

It should be noted that the Maximin-optimal design depends only on those parameter values with non-zero prior probability density and not on the actual magnitudes of the densities themselves.

For the purposes of illustration, we shall examine the effects of two priors (one more informative than the other) for each of α and ρ. We will take the true value of α, governing the net rate of spread of *A. planci*, as being 3.5 yr^{-1} and our priors for α are given two different functional forms. The informative prior is a gamma density $f_\alpha(\alpha; q, r) = \alpha^{q-1} r^q \exp(-r\alpha)/\Gamma(q)$ (where $\Gamma(\cdot)$ is the gamma function), with a mean of 3.5 and variance of 0.72. The less informative prior is a piecewise-continuous density comprising a uniform density over the interval $[0, u]$ and an exponential density over the interval (u, ∞):

$$f_\alpha(\alpha; p, d, u) = \begin{cases} p/u & (0 \le \alpha \le u) \\ (1-p)de^{-d(\alpha-u)} & (\alpha > u) \end{cases} \tag{2}$$

where

$$d = \frac{p}{u(1-p)}. \tag{3}$$

For our example, we constructed the less-informative prior for α by taking $u = 10$ and $p = 0.9$. The resulting prior has a mean and variance of approximately 5.61 and 10.98 respectively.

In light of very little information about α we advocate constructing a prior as in (2) by following three simple steps:

1. Define an interval $[0, u]$ ($u < \infty$) which is considered to be the most likely range of values of α.

2. Define a probability p ($0 < p < 1$) quantifying the level of certainty that α lies within the interval $[0, u]$.
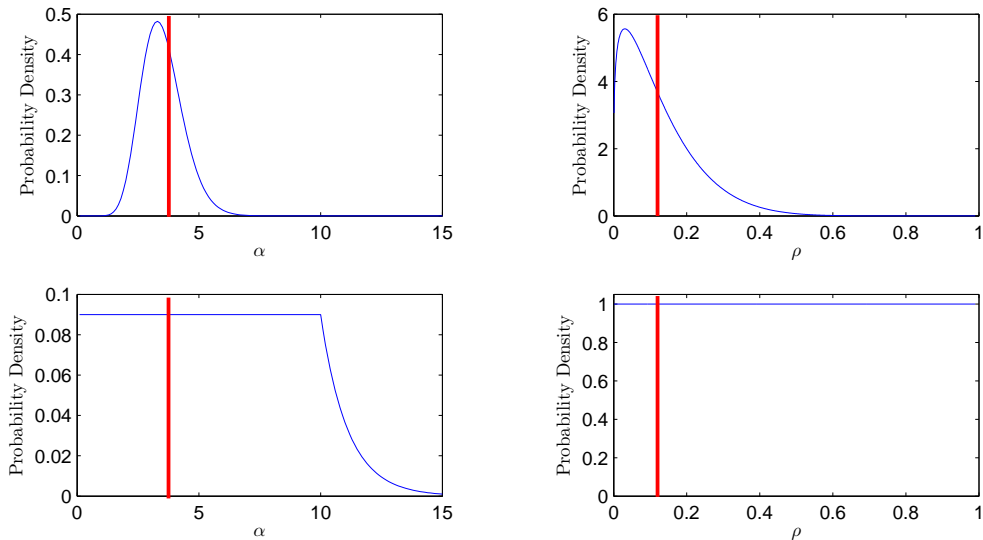
Figure 2. Prior probability density functions for α and ρ: (a) informative priors and (b) less informative priors. Location of true parameter values represented by vertical lines.

α Prior	ρ Prior	Criterion	Time of Observations (years)
Informative	Informative	ED-Optimal	0, 1.8684, 3.2874, 5
Informative	Less Informative	ED-Optimal	0, 1.5771, 2.8004, 5
Less Informative	Informative	ED-Optimal	0, 1.1559, 2.8838, 5
Less Informative	Less Informative	ED-Optimal	0, 1.0906, 2.7323, 5
Informative	Informative	Maximin	0, 0.0385, 4.9615, 5
Informative	Less Informative	Maximin	0, 0.0385, 4.9615, 5
Less Informative	Informative	Maximin	0, 0.0385, 4.9615, 5
Less Informative	Less Informative	Maximin	0, 0.0385, 4.9615, 5

Table 1. Optimal days for observations under various priors and optimality criteria (day zero corresponds to the observation n_0 prior to designing the experiment).

3. Identify the appropriate parameter of the exponential distribution using (3), which ensures the prior probability density is continuous at u.

The true value of ρ for the spread of *A. planci* was taken as $1/8$ and both informative and less-informative priors were taken to have a beta density, $f_\rho(\rho; a, b) = \rho^{a-1}(1 - \rho)^{b-1}/B(a, b)$ (where $B(\cdot, \cdot)$ is the beta function). For the informative prior we set the mean equal to $1/8$ and the variance equal to 0.01, while for the non-informative case we set the mean to 0.5 and the variance to $1/12$, so that the density is uniform over $[0, 1]$ (see Figure 2).

The optimal observation times under our different priors and objective functions were obtained using an easily implemented Cross-Entropy method similar to that presented in Pagendam and Pollett (2008).
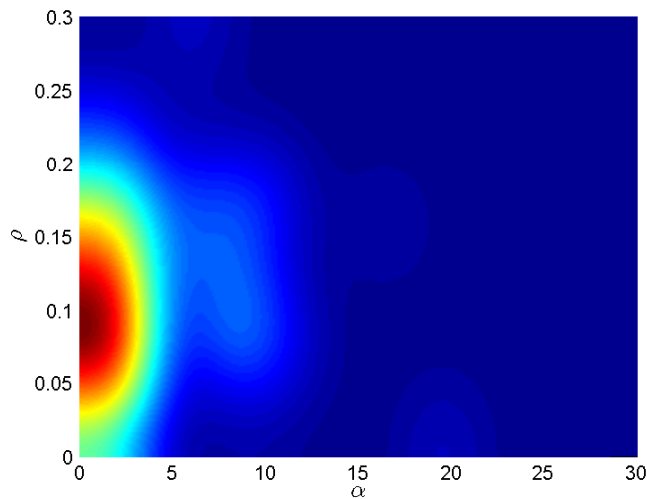
4. RESULTS

Table 1 provides the optimal times for observation of the metapopulation for the various combinations of the priors with the two optimality criteria. Each observation schedule was computed in Matlab on a 2.4 GHz Intel Core Duo laptop with 2GB of RAM in roughly 5 minutes. We note that all of the observation schedules make use of the full five years and that this is a *feature* of the optimal design rather than a constraint.

Figures 3(a) and 3(b) show kernel density estimates of the maximum likelihood estimator under the various observation schedules given that the true parameter values are $\alpha = 3.5$ yr^{-1} and $\rho = 1/8$ (corresponding to $\lambda = 4$ yr^{-1} and $\mu = 0.5$ yr^{-1}). Each figure was created by simulating 100 trajectories of the model using the true parameters and then estimating the parameters under the corresponding observation schedule using maximum likelihood.
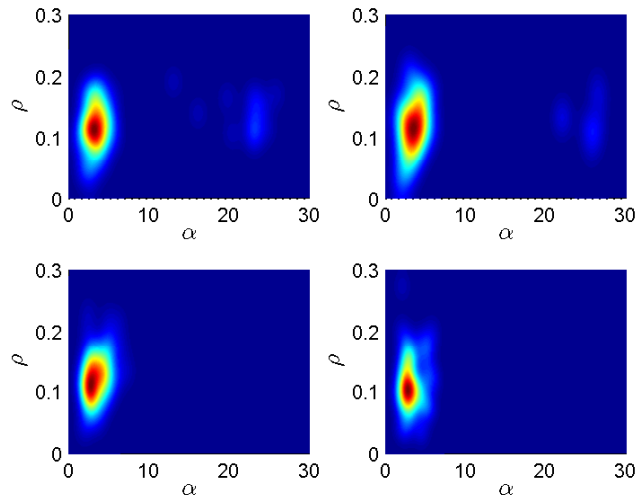
(a) ED-optimal schedule: informative priors on both parameters (top left), α informative and ρ less informative (top right), α less informative and ρ informative (bottom left), and, less informative priors on both parameters (bottom right).

(b) Maximin-optimal schedule: since all four observation schedules were identical, only a single kernel density estimate is presented.

Figure 3. Kernel density estimates of the maximum likelihood estimators.

5. DISCUSSION

Our results show that different priors and optimality criteria can have significant effects on the distribution of the maximum likelihood estimator. For the example of *A. planci* in the Ryukyu islands, we favour the use of the ED-optimality criteria in conjunction with a less-informative prior on α. The ED-optimal observation times show that in two of the four plots there is a second mode in the distribution which appears near $\alpha = 25$. The precise reason for its presence is unclear. However, it seems that the design is compromised by averaging over our priors. We can see from Table 1 that the one important difference between the ED-optimal observation schedules is the timing of the first observation, which is earlier when the prior for α is less informative. It appears that for the two bimodal plots, the first observation is sufficiently late that much of the information about the drift can be lost. The result appears to be a systematic positive bias in the estimator of α. It is likely that a prior that assigned less probability mass to lower values of α would avoid such problems, but this would require the practitioner to have better prior knowledge and this is not always the case. Interestingly, when less informative priors are used for α the ED-optimal scheme performs very well and there is no bimodality.

The Maximin-optimal observation schedules were identical for each of the priors, since the design criteria is only dependent on the domain of the priors and not on the magnitude of the probability densities. Under the Maximin-

optimal design, t_1 was to be taken as soon as practicably possible after t_0, whilst t_2 was to be taken as close to $t_3 = 5$ as possible. For practicality, we imposed the constraint that observation times were a minimum of 2 weeks apart. In simple terms, the Maximin-optimal design attempts to maximize the precision of the maximum likelihood estimator under the worst possible set of parameters, given our initial state. For our model, this is a design that maximizes the precision given that we have started in quasi-equilibrium (i.e. no initial drift is observed). For studies such as the spread of *A. planci*, this design criteria is likely to be far too pessimistic to be of practical use. Comparing Figure 3(b) to Figure 3(a) illustrates just how poorly the design performs given that our metapopulation did not begin in quasi-equilibrium, with the α being grossly under-estimated the majority of the time.

For colonisation of the Ryukyu islands by *A. planci*, we have shown that the use of ED-optimal designs with informative priors for α can lead to estimates of α that are seriously biased. In addition, we have shown that the Maximin-optimality criteria of Pronzato and Walter (1988) considered here is of little practical use for our model. Future avenues for research could consider the use of other more appropriate Maximin-optimality criteria that incorporate the magnitude of the prior probability densities.

ACKNOWLEDGEMENT

REFERENCES

Chaloner, K. and I. Verdinelli (1995), Bayesian experimental design: a review, *Statistical Science*, 10, 273–304.

Cook, A. Gibson, G. and C. Gilligan (2008), Optimal observation times in experimental epidemic processes, *Biometrics*, 64, 860–868.

Dobson, A. (2003), Metalife! *Science*, 301, 1488–1490.

Gilpin, M.E. and I. Hanski (1991), *Metapopulation Dynamics*, Academic Press, New York.

Levins, R. (1969), Some demographic and genetic consequences of environmental heterogeneity for biological control, *Bulletin of the Entomological Society of America*, 15, 237–240.

Hanski, I. (1999), *Metapopulation Ecology*, Oxford University Press, Oxford UK.

Pagendam, D.E. and P.K. Pollett (2008), Optimal sampling and problematic likelihood functions in a simple population model, *Environmental Modeling and Assessment*, doi:10.1007/s10666-008-9159-1.

Pollett, P.K. (2001), Diffusion approximations for ecological models, In (Ed. Fred Ghasssemi) *Proceedings of the International Congress on Modelling and Simulation*, Vol. 2, Modelling and Simulation Society of Australia and New Zealand, Australia, pp. 843–848.

Pronzato, L. and E. Walter (1985), Robust experimental design via stochastic approximation, *Mathematical Biosciences*, 75, 103–120.

Pronzato, L. and E. Walter (1988), Robust experiment design via maximin optimization, *Mathematical Biosciences*, 89, 161–176.

Ross, J.V., Taimre, T. and P.K. Pollett (2006), On parameter estimation in population models, *Theoretical Population Biology*, 70, 498–510.

Ross, J.V., Pagendam, D.E. and P.K. Pollett (2008), On parameter estimation in population models II: Multidimensional processes and transient dynamics, *Theoretical Population Biology*, doi:10.1016/j.tpb.2008.12.002.

Rubinstein, R.Y. and D.P. Kroese (2008), The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning, Springer-Verlag, New York, USA.

Weiss, G. and M. Dishon (1971), On the asymptotic behaviour of the stochastic and deterministic models of an epidemic, *Mathematical Biosciences*, 11, 261–265.

Yamaguchi, M. (1986), *Acanthaster planci* infestations of reefs and coral assemblages in Japan: a retrospective analysis of control events, *Coral Reefs*, 5, 23–30.