Statistical Report

Ecology, 103(2), 2022, e03573

Recursive Bayesian computation facilitates adaptive optimal design in ecological studies

Clinton B. Leach [1,8], Perry J. Williams [2], Joseph M. Eisaguirre [2,3], Jamie N. Womble [4,5], Michael R. Bower,[4] and Mevin B. Hooten[1,6,7]

[1]*Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, Colorado 80523 USA*
[2]*Department of Natural Resources and Environmental Science, University of Nevada, Reno, Nevada 89557 USA*
[3]*U.S. Fish and Wildlife Service, Marine Mammals Management, Anchorage, Alaska 99503 USA*
[4]*Southeast Alaska Inventory and Monitoring Network, National Park Service, Juneau, Alaska 99801 USA*
[5]*Glacier Bay Field Station, National Park Service, Juneau, Alaska 99801 USA*
[6]*U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, Fort Collins, Colorado 80523 USA*
[7]*Department of Statistics, Colorado State University, Fort Collins, Colorado 80523 USA*

Citation: Leach, C. B., P. J. Williams, J. M. Eisaguirre, J. N. Womble, M. R. Bower, and M. B. Hooten. 2022. Recursive Bayesian computation facilitates adaptive optimal design in ecological studies. Ecology 103(2):e03573. 10.1002/ecy.3573

Abstract. Optimal design procedures provide a framework to leverage the learning generated by ecological models to flexibly and efficiently deploy future monitoring efforts. At the same time, Bayesian hierarchical models have become widespread in ecology and offer a rich set of tools for ecological learning and inference. However, coupling these methods with an optimal design framework can become computationally intractable. Recursive Bayesian computation offers a way to substantially reduce this computational burden, making optimal design accessible for modern Bayesian ecological models. We demonstrate the application of so-called prior-proposal recursive Bayes to optimal design using a simulated data binary regression and the real-world example of monitoring and modeling sea otters in Glacier Bay, Alaska. These examples highlight the computational gains offered by recursive Bayesian methods and the tighter fusion of monitoring and science that those computational gains enable.

Key words: *Bayesian hierarchical modeling; computational efficiency; monitoring; survey design.*

Introduction

Ecological science involves both data collection and statistical modeling, but these two fundamental elements are often developed separately and sequentially in practice. Studies are commonly structured based on static random or space-filling designs. These designs have useful properties in some inferential settings, but may not represent the most efficient use of limited field resources, especially in complex, dynamic ecological systems. In fact, dynamically evolving processes may be monitored more efficiently with dynamically evolving designs (Hooten et al. 2009). Such designs can reduce redundancy in data collection (Wikle and Royle 1999) and produce higher quality data and lower prediction uncertainty (Hooten et al. 2009). Moreover, static surveillance monitoring may not make use of existing ecological knowledge that can lead to improved study designs and inference (Nichols and Williams 2006). In contrast, optimal adaptive survey design recognizes that existing data (e.g., from a pilot study or previous monitoring work) provide ecological information that can be leveraged to ensure that future data collection efforts are set up to be efficient and informative (Wikle and Royle 2005, Hooten et al. 2009).

The optimal design process is iterative, and proceeds through the following steps: collection of data, development and fitting of a statistical model, generation of predictions, evaluation and selection of a new design based on the model and its predictions, collection of new data using that design, and so on (Williams et al. 2018, Hooten et al. 2019). Throughout this process, practitioners are required to make a number of choices. Among these is the choice of model framework and structure.

Manuscript received 3 February 2021; revised 7 July 2021; accepted 3 August 2021. Corresponding Editor: José Miguel Ponciano.
[8] E-mail: clint.leach@gmail.com

Hierarchical Bayesian modeling has become widespread in ecology and is particularly well suited to integrating ecological processes with unknown parameters and noisy data (Berliner 1996, Wikle and Hooten 2010). This integration is achieved by specifying three levels of the statistical model (Berliner 1996): the data model that connects observations to the latent ecological process, the process model that describes that ecological process and its associated uncertainty, and parameter models that use prior information to constrain and inform the parameters of the data and process models.

Fitting these hierarchical models is computationally intensive and time consuming, especially for large spatiotemporal models (e.g., requiring more than 10 h in Williams et al. [2018]). Furthermore, evaluating a given design often requires generating predictions of the observations that design might produce in a future data collection effort, augmenting the original data with the predicted data, and fitting the model to the augmented data set. This fitted model then provides a means to evaluate how those new data would affect our understanding of the ecological process and its uncertainty. Finding an optimal design requires repeating this process (i.e., fitting the Bayesian hierarchical model) for every potential design. When the number of potential designs is large, the computational burden of each individual model fit renders this task computationally infeasible, requiring substantial cloud-based or cluster resources (e.g., Williams et al. 2018), or completely intractable.

The computational burden of the hierarchical Bayesian treatment has limited its application in optimal design settings, instead forcing practitioners to rely on other methods (e.g., Kalman filters; Wikle and Royle 2005, Hooten et al. 2009), or explore a relatively limited subset of designs (Williams et al. 2018). Thus, we currently lack the ability to carry the inference offered by modern Bayesian statistical models forward into the design phase without having to make compromises about the designs and models we consider. The main computational bottleneck involves updating an existing posterior distribution with predicted future data, which is the crux of Bayesian optimal adaptive design.

Recursive Bayesian inference provides methods that are well-suited to addressing this bottleneck. In particular, recursive methods enable a statistical model to be fit in a series of steps (e.g., to different groups of data, or to new data as it becomes available; Hooten et al. 2019). This partitioning of the statistical fitting procedure can offer large computational gains over fitting the full model every time new data need to be assimilated (Hooten et al. 2019). Recursive Bayesian methods have recently been used to facilitate computation in complex ecological models (e.g., Hooten et al. 2016, Gerber et al. 2018), but they have yet to be applied in the optimal design setting.

In what follows, we demonstrate how recursive Bayesian methods can be integrated into the optimal design workflow to substantially reduce the computational cost of assimilating new data from each potential design. We first provide an overview of the optimal design process in a Bayesian hierarchical setting, and identify the crucial role that recursive Bayesian computing can play in reducing the computational burden. Then we demonstrate the recursive Bayes optimal design approach in an application using simulated data and a binary regression framework. Finally, we apply the recursive Bayes optimal design framework to a complex Bayesian hierarchical model of sea otter spatiotemporal dynamics, demonstrating the substantial computational gains that recursive Bayesian methods offer. These examples highlight the important role that recursive Bayesian methods can play in formally coupling optimal adaptive design and modern hierarchical Bayesian modeling, leading to improved ecological inference and closing the feedback loop between modeling and data collection.

Methods

Evaluating a design

We represent all possible observations of an ecological process of interest as an $N \times 1$ vector $\mathbf{y}$ (e.g., containing abundance or presence/absence at a complete set of sites in a study area). Then, we collect an initial sample of n_1 observations, $\mathbf{y}_1$, produced by an $n_1 \times N$ design matrix $\mathbf{K}_1$ that maps the full domain to the initial observations such that $\mathbf{y}_1 = \mathbf{K}_1\mathbf{y}$. The design matrix, $\mathbf{K}_1$, is usually a (sparse) matrix composed of zeros and ones that selects the subset of $\mathbf{y}$ that is observed. Given a Bayesian model with parameters $\boldsymbol{\theta}$, we obtain a sample from the first-stage posterior distribution $[\boldsymbol{\theta}|\mathbf{y}_1]$ using an appropriate stochastic sampling algorithm (Gelfand and Smith 1990). Importantly, the recursive Bayes procedure outlined here is compatible with any valid first-stage sampling algorithm (Hooten et al. 2019), including Markov chain Monte Carlo (MCMC) and Hamiltonian Monte Carlo (HMC), and any software implementation thereof (e.g., NIMBLE [de Valpine et al. 2017] or Stan [Carpenter et al. 2016]).

In the optimal design framework, we use the existing data $\mathbf{y}_1$ and the fitted model to compare how different designs for future data collection affect our estimate of some target quantity (often a measure of uncertainty). Formally, we let $l = 1, \ldots, L$ index the set of all possible designs, defined by $n_2 \times N$ design matrices $\mathbf{K}_2^{(l)}$, that would produce a new set of n_2 observations given by $\mathbf{y}_2^{(l)} = \mathbf{K}_2^{(l)}\mathbf{y}$. The first step of optimal design is to define a design criterion, $d^{(l)}\left(\mathbf{y}_1, \mathbf{y}_2^{(l)}\right)$, that summarizes some aspect of our understanding of the process given both the original and the new data. Choices of design criterion often include prediction variance (Hooten et al. 2009) or the variance of model parameters (Hooten et al. 2012) or derived quantities (e.g., abundance; Williams et al. 2018), in which case the goal of optimal design is to find the design that will yield the smallest variance (i.e., the least uncertainty).

Generally, the design criterion will depend either on the posterior predictive distribution of the full process

$$\left[\mathbf{y}\middle|\mathbf{y}_1,\mathbf{y}_2^{(l)}\right[=\int]\mathbf{y}\middle|\mathbf{y}_1,\mathbf{y}_2^{(l)},\boldsymbol{\theta}\Big][\boldsymbol{\theta}|\mathbf{y}_1,\mathbf{y}_2^{(l)}]d\boldsymbol{\theta}, \tag{1}$$

or directly on the posterior distribution of the model parameters $\left[\boldsymbol{\theta}\middle|\mathbf{y}_1,\mathbf{y}_2^{(l)}\right]$. In principle, to evaluate the design criterion for a given $\mathbf{K}_2^{(l)}$, we need to measure the process (i.e., observe $\mathbf{y}_2^{(l)}$), augment the existing data with that measurement, and fit the statistical model to characterize $\left[\boldsymbol{\theta}\middle|\mathbf{y}_1,\mathbf{y}_2^{(l)}\right]$ and hence the posterior predictive distribution $\left[\mathbf{y}\middle|\mathbf{y}_1,\mathbf{y}_2^{(l)}\right]$. Thus, computing $d^{(l)}\left(\mathbf{y}_1,\mathbf{y}_2^{(l)}\right)$ given new (predicted) data requires fitting the model to the augmented data, which, in the case of modern Bayesian hierarchical models, may be computationally demanding. If the number of potential designs, L, is large, the standard Bayesian optimal design procedure becomes intractable.

However, we can further decompose the posterior distribution as

$$\begin{aligned}\left[\boldsymbol{\theta}\middle|\mathbf{y}_1,\mathbf{y}_2^{(l)}\right] &\propto \left[\mathbf{y}_2^{(l)}\middle|\mathbf{y}_1,\boldsymbol{\theta}\right]\left[\boldsymbol{\theta}\middle|\mathbf{y}_1\right],\\ &\propto \left[\mathbf{y}_2^{(l)}\middle|\mathbf{y}_1,\boldsymbol{\theta}\right]\left[\mathbf{y}_1\middle|\boldsymbol{\theta}\right]\left[\boldsymbol{\theta}\right],\end{aligned} \tag{2}$$

where $[\boldsymbol{\theta}|\mathbf{y}_1] \propto [\mathbf{y}_1|\boldsymbol{\theta}][\boldsymbol{\theta}]$ is available from the original model fit to $\mathbf{y}_1$. This natural decomposition of the posterior distribution of $\boldsymbol{\theta}$ makes clear that the first-stage posterior distribution, $[\boldsymbol{\theta}|\mathbf{y}_1]$, serves as a prior on $\boldsymbol{\theta}$ in the second-stage analysis of the augmented data (Hooten et al. 2019). In prior-proposal recursive Bayes (PPRB), we also use $[\boldsymbol{\theta}|\mathbf{y}_1]$ as the proposal distribution in a Metropolis-Hastings MCMC algorithm to update the posterior distribution of $\boldsymbol{\theta}$ given the new data produced by a given design (Hooten et al. 2019).

At step k of the PPRB MCMC algorithm, we sample a proposal $\boldsymbol{\theta}^{(*)} \sim [\boldsymbol{\theta}|\mathbf{y}_1]$. We then accept that proposal and set $\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(*)}$ with probability $\min(1,\ r)$, where

$$\begin{aligned} r &= \frac{\left[\mathbf{y}_2^{(l)}\middle|\boldsymbol{\theta}^{(*)},\mathbf{y}_1\right]\left[\boldsymbol{\theta}^{(*)}\middle|\mathbf{y}_1\right]}{\left[\mathbf{y}_2^{(l)}\middle|\boldsymbol{\theta}^{(k-1)},\mathbf{y}_1\right]\left[\boldsymbol{\theta}^{(k-1)}\middle|\mathbf{y}_1\right]}\frac{\left[\boldsymbol{\theta}^{(k-1)}|\mathbf{y}_1\right]}{\left[\boldsymbol{\theta}^{(*)}\middle|\mathbf{y}_1\right]},\\ &= \frac{\left[\mathbf{y}_2^{(l)}\middle|\boldsymbol{\theta}^{(*)},\mathbf{y}_1\right]}{\left[\mathbf{y}_2^{(l)}\middle|\boldsymbol{\theta}^{(k-1)},\mathbf{y}_1\right]}, \end{aligned} \tag{3}$$

where allowing the original posterior distribution to serve as both prior and proposal enables the cancellation and results in a ratio that depends only on the conditional likelihood of the new data. Note that, because we often have a finite MCMC sample from the first model fit (and thus a finite set of proposals, $\boldsymbol{\theta}^{(*)}$, for this stage), we can pre-compute $\left[\mathbf{y}_2^{(l)}\middle|\boldsymbol{\theta}^{(*)},\mathbf{y}_1\right]$ for each design and proposal before the second stage and in parallel (Hooten et al. 2019). This pre-computation, together with the relative simplicity of the above Metropolis-Hastings ratio, can lead to a substantial decrease in computation time compared to running the full sampling algorithm for every design. Code implementing PPRB for the following examples is available in Data S1 (archived in Leach 2021) and Data S2 (archived in Eisaguirre 2021).

Generating potential future data

The above discussion assumes that $\mathbf{y}_2^{(l)}$ is known, which, of course, it is not. In the case of Gaussian models (e.g., Wikle and Royle 1999), the design criterion $d^{(l)}$ depends only on $\mathbf{K}_2^{(l)}$ and the modeled dependence structure, and thus new data are not required. In more complex models, evaluating the design requires predictions of the new data $\mathbf{y}_2^{(l)}$ (Wikle and Royle 2005). These predictions can be readily generated by draws from the posterior predictive (or imputation) distribution of $\mathbf{y}_2^{(l)}$ produced from the first stage model

$$\left[\mathbf{y}_2^{(l)}\middle|\mathbf{y}_1\right] = \int\left[\mathbf{y}_2^{(l)}\middle|\mathbf{y}_1,\boldsymbol{\theta}\right][\boldsymbol{\theta}|\mathbf{y}_1]d\boldsymbol{\theta}. \tag{4}$$

We can then use a multiple imputation approach (Rubin 1996, Scharf et al. 2017) to average the design criterion over the imputation distribution such that

$$\begin{aligned} d^{(l)}(\mathbf{y}_1) &= \int d^{(l)}\left(\mathbf{y}_1,\mathbf{y}_2^{(l)}\right)[\mathbf{y}_2^{(l)}|\mathbf{y}_1]d\mathbf{y}_2^{(l)},\\ &= E\left(d^{(l)}\left(\mathbf{y}_1,\mathbf{y}_2^{(l)}\right)|\mathbf{y}_1\right) \end{aligned} \tag{5}$$

In practice, we can compute this expectation by first obtaining samples, $\mathbf{y}_2^{(l)(m)} \sim [\mathbf{y}_2^{(l)}|\mathbf{y}_1]$ for $m = 1,\ \ldots,\ M$, from the first stage posterior predictive distribution. We then fit the model to each of the M augmented data sets $\left(\mathbf{y},\mathbf{y}_2^{(l)(m)}\right)$, obtain samples from each of the posterior distributions $\left[\boldsymbol{\theta}\middle|\mathbf{y}_1,\mathbf{y}_2^{(l)(m)}\right]$ using PPRB, and compute the mean of the corresponding design criteria:

$$d^{(l)}(\mathbf{y}_1) = \frac{1}{M}\sum_m d^{(l)}\left(\mathbf{y}_1,\mathbf{y}_2^{(l)(m)}\right). \tag{6}$$

The accuracy of the expectation will improve as M grows larger, but often a relatively small M (on the order of 10) will be sufficient (Rubin 1996).

The multiple imputation procedure enables us to account for the uncertainty in the future data in the evaluation of the design criterion. Alternatively, if accounting for such uncertainty is not desired or necessary, we could generate a single point estimate of future data by computing the posterior predictive mean (or median or mode, as appropriate) of $\left[\mathbf{y}_2^{(l)}\middle|\mathbf{y}_1\right]$ or assigning a single fixed forecast of $\mathbf{y}_2^{(l)}$ from another source (e.g., expert opinion).

Optimization

Given an ability to rapidly compute the design criterion $d^{(l)}$, the optimal design can be obtained by finding the design that minimizes (or maximizes) $d^{(l)}$. In cases where the design space is relatively small, each design can be evaluated and the global optimum selected. If the design space is too large to evaluate every design, an optimization routine may be required (e.g., an exchange algorithm; Royle and Nychka 1998).

Example: Simulated Binary Regression

Consider a situation in which we seek to predict the occupancy of a particular species across a spatial domain comprising 100 discrete units (e.g., plots or transects) over which we measure a covariate (e.g., through remote sensing), x_i for $i = 1, \ldots, 100$ (Fig. 1b). Let $\mathbf{y}$ be a vector comprising binary occupancy at all sites. An initial data collection effort randomly samples 10 of these sites, producing initial data set $\mathbf{y}_1 = \mathbf{K}_1\mathbf{y}$ with covariates $\mathbf{x}_1 = \mathbf{K}_1\mathbf{x}$, where $\mathbf{K}_1$ is a 10×100 matrix of zeros and ones, with a single one in each row identifying the sampled plot.

We model these data using binary regression, with a Bernoulli likelihood and a probit link function (Φ^{-1}, the inverse CDF of a standard normal distribution). The resulting full Bayesian model is as follows:

$$\begin{aligned} y_{1i} &\sim \text{Bernoulli}(p_i), \\ \Phi^{-1}(p_i) &= \beta_0 + \beta_1 x_{1i}, \\ \beta_0 &\sim \text{Normal}\left(0,\ \sigma_0^2\right), \\ \beta_1 &\sim \text{Normal}\left(0,\ \sigma_1^2\right). \end{aligned} \tag{7}$$

We use the data augmentation and Gibbs sampling approach of Albert and Chib (1993) to draw an MCMC sample from the posterior distribution of the regression coefficients ($[\boldsymbol{\beta}|\mathbf{y}_1]$) and the posterior predictive distribution of occupancy across the study domain ($[\mathbf{y}|\mathbf{y}_1]$). This posterior sample could alternatively be generated using other algorithms (e.g., HMC) or software (e.g., brms; Bürkner 2017) without changing the following workflow. The goal of the optimal design framework is to use this initial data set and model output to select the next site (of the remaining 90) to be sampled. That is, each of the remaining 90 sites represent a potential design, indexed by l, that corresponds to a 1×100 design vector $\mathbf{K}_2^{(l)}$ that has a single 1 in the position of the sampled site and produces new data $y_2^{(l)} = \mathbf{K}_2^{(l)}\mathbf{y}$.

To select the optimal design, we first specify a design criterion. Our goal is to predict occupancy. Thus we seek the design that minimizes the total posterior predictive variance given both the initial and new data

$$d^{(l)}\left(\mathbf{y}_1, y_2^{(l)}\right) = \sum_{i=1}^{100} \text{var}(y_i|\mathbf{y}_1, y_2^{(l)}), \tag{8}$$

where $\text{var}(y_i|\mathbf{y}_1, y_2^{(l)})$ is the pointwise posterior predictive variance calculated using the MCMC sample from $\left[y_i\middle|\mathbf{y}_1, y_2^{(l)}\right]$ produced by PPRB. To account for the uncertainty in predictions of the future $y_2^{(l)}$, we average this design criterion over the imputation distribution $\left[y_2^{(l)}\middle|\mathbf{y}_1\right]$ to obtain $d^{(l)}(\mathbf{y}_1)$. In the binary case, imputed realizations $y_2^{(l)(m)}$ can only take on values of 0 or 1, enabling efficient computation of this expectation (see Appendix S1).

Evaluating the design criterion requires sampling from $\left[\boldsymbol{\beta}\middle|\mathbf{y}_1, y_2^{(l)(m)}\right]$ for all 90 designs and imputed future data sets. Rather than fit the model to the entire data set $\left(y_2^{(l)(m)}, \mathbf{y}_1\right)$, we apply PPRB to use the existing output from the initial MCMC algorithm (i.e., the $\boldsymbol{\beta}^{(k)}$ drawn from $[\boldsymbol{\beta}|\mathbf{y}_1]$). At step k of the second-stage MCMC algorithm for design l and imputed data m, we implement PPRB as follows:

1) Sample a proposal $\boldsymbol{\beta}^{(*)} \sim [\boldsymbol{\beta}|\mathbf{y}_1]$ (i.e., selected randomly with replacement from the first stage MCMC sample).
2) Compute the PPRB Metropolis-Hastings ratio

$$r = \frac{\left[y_2^{(l)(m)}\middle|\boldsymbol{\beta}^{(*)}\right]}{\left[y_2^{(l)(m)}\middle|\boldsymbol{\beta}^{(k)}\right]}. \tag{9}$$

3) Accept the proposal $\boldsymbol{\beta}^{(*)}$ with probability $\min(r, 1)$.

Using the second-stage samples $\boldsymbol{\beta}^{(k)}$, we compute the design criteria for each potential design (averaging over the imputation distribution) and identify the optimal site for the next sample (Fig. 1c, d). The expected design criterion is largest (i.e., with the largest predictive variance) for sites with more extreme covariate values (Fig. 1c). Despite the fact that the inflection point (i.e., the x value where $p(x) = 0.5$) represents the largest Bernoulli variance, sampling locations with covariate values slightly larger than this inflection point produce the smallest total expected prediction variance. The positioning of the optimal design just off the inflection point highlights the fact that, even in simple models, the optimal design is often not intuitive and justifies the need for rigorous optimal design in ongoing ecological monitoring. In more complex spatiotemporal models, often with larger design spaces, it becomes even more difficult to identify effective designs a priori (Wikle and Royle 1999, 2005), further emphasizing the need for PPRB and the speed with which it enables us to evaluate a potentially large number of designs.

Example: Spatiotemporal Dynamics

Sea otters (*Enhydra lutris*) are an apex predator of the nearshore marine community of the North Pacific Ocean and nearly went extinct at the turn of the 20th century. Reintroductions, translocations, and legal

protections allowed sea otters to recolonize much of their former range (Williams et al. 2019). The return of sea otters has influenced marine food webs, both directly and indirectly, with impacts on commercially important fisheries. Thus, information regarding the continued growth and expansion of sea otters is critical for predicting future expansion, understanding their role as a keystone species, and for informing natural resource management (Tinker et al. 2019). Sea otters are surveyed using aircraft over large spatial domains and flight time is typically restricted by range and fuel capacity of aircraft and operating conditions (Williams et al. 2017*a*). Sampling designs must balance requisite data collection with human safety, aircraft availability, and cost.

To estimate growth and colonization dynamics of sea otters in Glacier Bay, Alaska, Williams et al. (2017*b*) developed a mechanistically motivated reaction-diffusion model known as ecological diffusion, embedded within a Bayesian hierarchical framework with data, process, and parameter levels (sensu Berliner 1996). This model was fit to sea otter aerial survey counts (Esslinger 2019), with data and process models specified as follows

$$\begin{aligned} &\text{Data Model}: y_t(\mathbf{s}_i) \quad \sim \text{Binomial}(n_t(\mathbf{s}_i), \phi), \\ &\text{Process Model}: n_t(\mathbf{s}_i) \quad \sim \text{Poisson}(u(\mathbf{s}_i, t)), \\ &\frac{\partial u(\mathbf{s},t)}{\partial t} = \left(\frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2}\right)[\mu(\mathbf{s},t)u(\mathbf{s},t)] + \gamma u(\mathbf{s},t). \end{aligned} \tag{10}$$

In Eq. 10, $y_t(\mathbf{s}_i)$ represents sea otter count data at locations $\mathbf{s}_i$ for $i = 1, \ldots, J$ during time t, $n_t(\mathbf{s}_i)$ is the

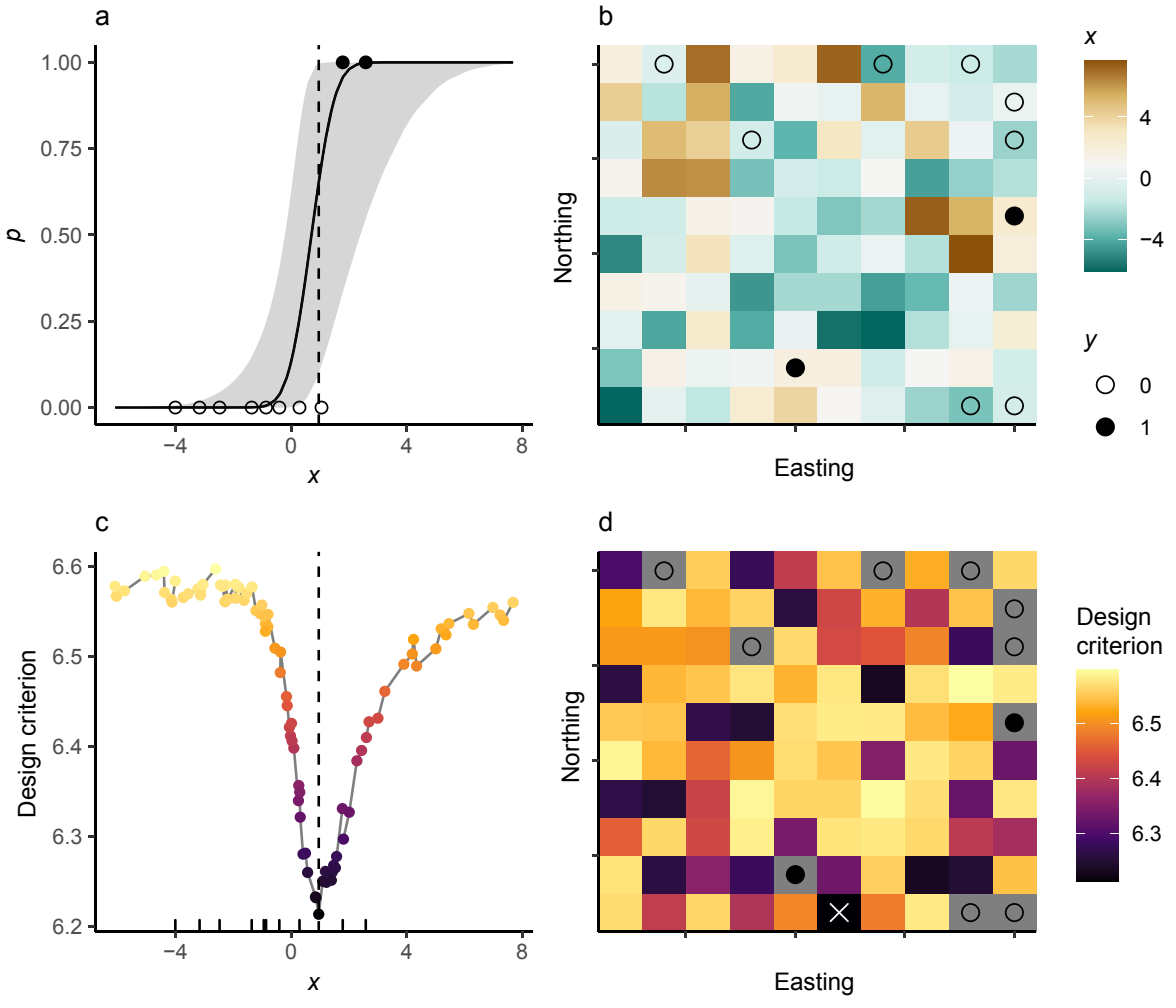

FIG. 1. Initial data collection effort and evaluation of the optimal design for a binary regression model. (a) The initial estimated occupancy probability p as a function of x, with the black line showing the posterior median and the gray ribbon showing the 95% credible interval. The points show the data obtained from (b) an initial random survey of the study domain. In panel b, each square is a sampling plot with measured covariate x indicated by the color of the fill. The points indicate the initial sites surveyed, with filled circles indicating the organism was present and open circles indicating absence. Each of the remaining plots represent a potential design for the next survey. (c, d) The computed design criteria for each of the 90 potential designs as a function of the x value at each plot (c) and mapped on the study domain (d). The dashed line in panels a and c indicates the x value of the optimal plot and the white × identifies the optimal plot in panel d.

true latent abundance of sea otters, and ϕ is the individual sea otter detection probability. The dynamic abundance intensity process, $u(\mathbf{s}, t)$, was governed by an ecological diffusion PDE, with a constant instantaneous Malthusian growth γ and motility $\mu(\mathbf{s}, t)$ modeled as a function of spatially varying covariates. See Appendix S1 for the full model specification. The model described in Eq. 10 was fit to baseline data, $\mathbf{y}_1$ collected for $t = 1993, \ldots, 2012$, using a custom MCMC algorithm (Williams et al. 2017*b*). We used our algorithm to obtain three MCMC chains in parallel, with 50,000 draws per chain plus 10,000 for burn in. This required approximately 5 h per chain, for a total of 15 CPU hours.

Following the optimal design framework, our goal was to use the initial data $\mathbf{y}_1$ and corresponding model output to inform the collection of transects to be surveyed in a hypothetical 2013, producing new data $\mathbf{y}_2^{(l)}$. Rather than choosing a design criterion based on the posterior predictive density as in the binary regression example, in this example, we focused on the latent total abundance intensity in 2013, $u_{2013} = \int_S u_{2013}(\mathbf{s})d\mathbf{s}$. Our objective was to minimize the variance of the total intensity $d^{(l)}\left(\mathbf{y}_1, \mathbf{y}_2^{(l)}\right) = \text{var}\left(u_{2013}\middle|\mathbf{y}_1, \mathbf{y}_2^{(l)}\right)$, where $\text{var}(u_{2013}|\mathbf{y}_1, y_2^{(l)})$ is the predictive process variance calculated using the MCMC sample from $[u_{2013}|\mathbf{y}_1, \mathbf{y}_2^{(l)}]$ produced by PPRB. We used the multiple imputation approach described above to average this design criterion over the posterior predictive distribution of $\mathbf{y}_2^{(l)}$ using $M = 100$ draws of $\mathbf{y}_2^{(l)(m)} \sim \left[\mathbf{y}_2^{(l)}\middle|\mathbf{y}_1\right]$, such that

$$d^{(l)}(\mathbf{y}_1) \approx \frac{1}{M}\sum_{m=1}^{M} \text{var}(u_{2013}|\mathbf{y}_1, \mathbf{y}_2^{(l)(m)}). \quad (11)$$

For each design l and imputed future data set m, we implemented the PPRB MCMC algorithm as follows:

1) Sample a proposal $\left(\mathbf{n}_{2013}^{(*)}, \mathbf{u}_{2013}^{(*)}\right) \sim [\mathbf{n}_{2013}, \mathbf{u}_{2013}|\mathbf{y}_1]$.
2) Compute the PPRB Metropolis-Hastings ratio

$$r = \frac{\left[\mathbf{y}_2^{(l)(m)}\middle|\mathbf{n}_{2013}^{(*)}\right]}{\left[\mathbf{y}_2^{(l)(m)}\middle|\mathbf{n}_{2013}^{(k)}\right]}. \quad (12)$$

3) Accept the proposal $\left(\mathbf{n}_{2013}^{(*)}, \mathbf{u}_{2013}^{(*)}\right)$ with probability $\min(r, 1)$.

From the resulting MCMC chains, we computed $d^{(l)}$ for each candidate design. Given constraints associated with aircraft range and availability, approximately 20 transects can be flown per day in Glacier Bay, resulting in $\binom{170}{20}$ possible survey designs given the dimensions of the survey area and the resolution of the data collection methods (Williams et al. 2018). While it was not feasible to assess all possible designs, the PPRB procedure allowed us to compare many more than was previously practical. Given that the ecological diffusion model required approximately 15 CPU hours to estimate parameters using an MCMC algorithm, previous optimization routines that fit the model for each design using MCMC would require 1500 CPU hours to assess just 1 design over 100 imputed data sets (Williams et al. 2018). In contrast, the PPRB approach permitted us to compute the design criterion of 1,000 designs (Fig. 2a) in about 480 CPU hours, which we reduced to < 5 h of run time by parallelizing evaluation over multiple CPUs. The survey design that optimized our design criterion (Fig. 2b) reduced the variance of the hypothetical 2013 sea otter abundance estimate by 38% over the average random design.

Discussion

Applying the principles of optimal design to make efficient use of field resources requires methods that make efficient use of computational resources. This is especially true for ecological studies in which Bayesian hierarchical models are deployed. These models can capture rich mechanistic information (Wikle and Hooten 2010) but are often computationally demanding to fit. We proposed the use of PPRB (Hooten et al. 2019) to alleviate the computational burden and make evaluating a large number of designs feasible. We demonstrated this method using a binary regression model and highlighted that optimal designs may not always be intuitive without a comprehensive search of the design space (Wikle and Royle 1999, 2005). Furthermore, we applied the procedure to a complex Bayesian hierarchical model of sea otter spatiotemporal dynamics and demonstrated the substantial computational gains that PPRB produces relative to fitting the entire model for every considered design and possible data set.

The rapid and relatively extensive search of the design space allowed us to identify a collection of transects for a hypothetical 2013 sea otter aerial survey that would produce a more precise estimate of the total sea otter abundance in Glacier Bay than the average random design. Given that sea otters are a keystone species (Estes and Palmisano 1974) with a rapidly expanding range and abundance in Glacier Bay (Williams et al. 2019), accurately estimating their abundance is crucial for monitoring and conserving the nearshore ecosystem in the face of environmental and anthropogenic changes (Coletti et al. 2016, Tinker et al. 2019). We demonstrated that the optimal design framework, by leveraging existing knowledge of sea otter dynamics, learned through the combination of existing survey data and the mechanistic principles embedded in the reaction-diffusion PDE, can help make monitoring data as useful

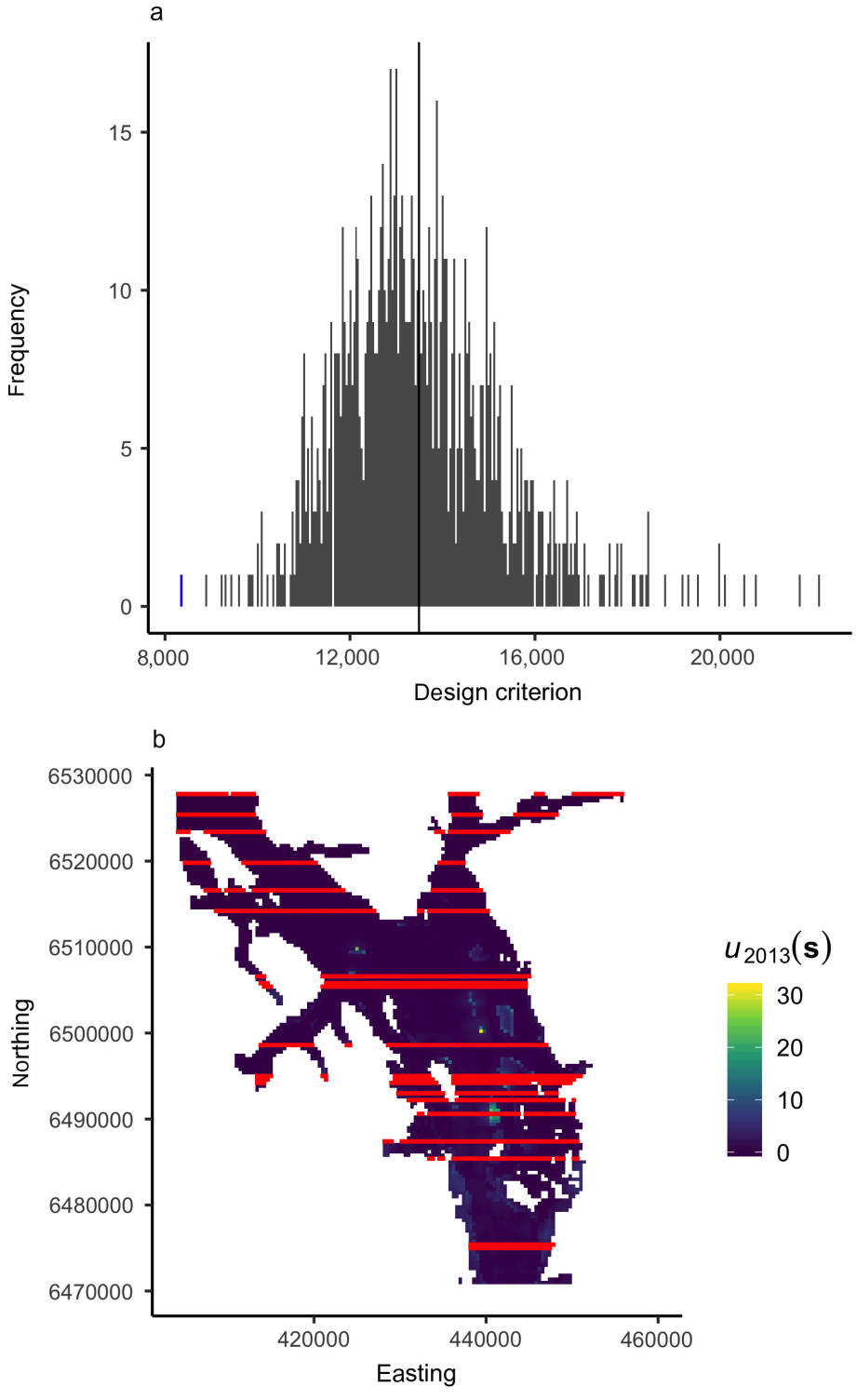

FIG. 2. (a) Histogram of design criteria computed for 1,000 randomly selected sea otter survey designs. The mean is given by the black line and the blue bar indicates the optimal survey design. Each design consisted of 20 transects to be flown over Glacier Bay in southeastern Alaska. (b) Forecasted sea otter abundance intensity across Glacier Bay in 2013, $u_{2013}(\boldsymbol{s})$, and optimal hypothetical survey design. Red lines correspond to the 20 transects representing the optimal design.

and informative as possible (Nichols and Williams 2006).

The challenges inherent in monitoring sea otters in Glacier Bay and across the North Pacific—costly data collection, dynamic spatiotemporal processes, and the need for quality data for conservation, management, and inference—are emblematic of the challenges faced throughout much of ecology. Both optimal design and Bayesian hierarchical modeling offer potential solutions to some of these challenges, and the use of PPRB allows them to be coupled more easily. This coupling will help to make the optimal design framework accessible to larger monitoring efforts across broader spatial domains and, in particular, may assist in the targeting of monitoring efforts across the sea otter range in the North Pacific (Eisaguirre et al. 2021).

Further, by reducing the computational burden of evaluating a given design, PPRB allows for greater flexibility in implementing the other components of the iterative optimal design framework, including the types of the data collected, the choice of design criteria, and the optimization framework. We focused on applications where the goal was to choose the locations at which to collect a single type of data, but the PPRB optimal design framework could be extended to target sampling across multiple data types (e.g., in integrated population modeling [Schaub et al. 2007] or multispecies studies). Further, we demonstrated two choices of design criterion based on predictive variance (in the binary regression example) and the variance of latent derived quantities (in the sea otter example). The choice of design criterion will depend on the goals of a particular study and could include other components, such as the costs of implementing a given design (Williams and Brown 2020), the benefits of any connected management actions (Williams and Brown 2020), or a measure of the strength of preferential sampling implied by a design (Diggle et al. 2010, Gelfand and Shirota 2019). Last, the approach we described can be enhanced by additional optimization strategies. The speed gains offered by PPRB may make the application of optimization frameworks (e.g., exchange algorithms, Royle and Nychka 1998) more feasible, and the identification of a global optimum more likely.

As we have demonstrated in our examples and discussion, recursive Bayesian methods offer to substantially ease the computational burden of coupling optimal design procedures with Bayesian hierarchical modeling. By facilitating this coupling, recursive Bayesian methods help close the feedback loop between data collection and data analysis, allowing the knowledge produced by Bayesian hierarchical modeling to inform monitoring efforts that improve and accelerate ecological learning and inference.

Acknowledgments

Funding for this research was provided by the National Park Service Inventory and Monitoring Program and NSF DEB 1927177. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. The findings and conclusions of the U.S. Fish and Wildlife Service authors in this article are their own and do not necessarily represent the views of the U.S. Fish and Wildlife Service. We thank the U.S. Geological Survey Advanced Research Computing team for use of the Yeti Supercomputer (https://doi.org/10.5066/F7D798MJ).

Literature Cited

Albert, J. H., and S. Chib. 1993. Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association 88:669–679.

Berliner, L. M. 1996. Hierarchical Bayesian time series models. Pages 15–22 *in* K. M. Hanson, and R. N. Silver, editors. Maximum entropy and Bayesian methods. Springer Netherlands, Dordrecht, the Netherlands.

Bürkner, P.-C. 2017. brms: an R package for Bayesian multilevel models using Stan. Journal of Statistical Software 80:1–28.

Carpenter, B., A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. A. Brubaker, P. Li, and A. Riddell. 2016. Stan: a probabilistic programming language. Journal of Statistical Software 76:1–32.

Coletti, H. A., J. L. Bodkin, D. H. Monson, B. E. Ballachey, and T. A. Dean. 2016. Detecting and inferring cause of change in an Alaska nearshore marine ecosystem. Ecosphere 7:e01489.

de Valpine, P., D. Turek, C. J. Paciorek, C. Anderson-Bergman, D. T. Lang, and R. Bodik. 2017. Programming with models: Writing statistical algorithms for general model structures with nimble. Journal of Computational and Graphical Statistics 26:403–413.

Diggle, P. J., R. Menezes, and T. Su. 2010. Geostatistical inference under preferential sampling. Journal of the Royal Statistical Society: Series C (Applied Statistics) 59:191–232.

Eisaguirre, J. M. 2021. jmeisaguirre/glba optimal design (v1.0.0). Zenodo. https://doi.org/10.5281/zenodo.5172817

Eisaguirre, J. M., P. J. Williams, X. Lu, M. L. Kissling, W. S. Beatty, G. G. Esslinger, J. N. Womble, and M. B. Hooten. 2021. Diffusion modeling reveals effects of multiple release sites and human activity on a recolonizing apex predator. Movement Ecology 9:1–14.

Esslinger, G. 2019. Sea otter aerial survey data from Glacier Bay National Park and Preserve, 1999–2012. U.S. Geological Survey data release. https://doi.org/10.5066/P9SBAFF6

Estes, J. A., and J. F. Palmisano. 1974. Sea otters: their role in structure nearshore communities. Science 185:1058–1060.

Gelfand, A. E., and S. Shirota. 2019. Preferential sampling for presence/absence data and for fusion of presence/absence data with presence-only data. Ecological Monographs 89:1–17.

Gelfand, A. E., and A. F. M. Smith. 1990. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association 85:398–409.

Gerber, B. D., M. B. Hooten, C. P. Peck, M. B. Rice, J. H. Gammonley, A. D. Apa, and A. J. Davis. 2018. Accounting for location uncertainty in azimuthal telemetry data improves ecological inference. Movement Ecology 6:14.

Hooten, M. B., F. E. Buderman, B. M. Brost, E. M. Hanks, and J. S. Ivan. 2016. Hierarchical animal movement models for population-level inference. Environmetrics 27:322–333.

Hooten, M. B., D. S. Johnson, and B. M. Brost. 2019. Making recursive Bayesian inference accessible. American Statistician 75:185–194

Hooten, M. B., B. E. Ross, and C. K. Wikle. 2012. Optimal spatio-temporal monitoring designs for characterizing population trends. Pages 443–459 *in* R. A. Gitzen, J. J. Millspaugh, A. B. Cooper, and D. S. Licht, editors. Design and analysis of

long-term ecological monitoring studies. Cambridge University Press, Cambridge, Massachusetts, USA.
Hooten, M. B., C. K. Wikle, S. L. Sheriff, and J. W. Rushin. 2009. Optimal spatio-temporal hybrid sampling designs for ecological monitoring. Journal of Vegetation Science 20:639–649.
Leach, C. B. 2021. Simulated data example of recursive Bayesian computing for optimal design (v1.0.0). Zenodo. https://doi.org/10.5281/zenodo.5172768
Nichols, J. D., and B. K. Williams. 2006. Monitoring for conservation. Trends in Ecology and Evolution 21:668–673.
Royle, J. A., and D. Nychka. 1998. An algorithm for the construction of spatial coverage designs with implementation in SPLUS. Computers and Geosciences 24:479–488.
Rubin, D. B. 1996. Multiple Imputation after 18+ Years. Journal of the American Statistical Association 91:473–489.
Scharf, H., M. B. Hooten, and D. S. Johnson. 2017. Imputation approaches for animal movement modeling. Journal of Agricultural, Biological, and Environmental Statistics 22:335–352.
Schaub, M., O. Gimenez, A. Sierro, and R. Arlettaz. 2007. Use of integrated modeling to enhance estimates of population dynamics obtained from limited data. Conservation Biology 21:945–955.
Tinker, M. T., V. A. Gill, G. G. Esslinger, J. Bodkin, M. Monk, M. Mangel, D. H. Monson, W. W. Raymond, and M. L. Kissling. 2019. Trends and carrying capacity of sea otters in southeast Alaska. Journal of Wildlife Management 83:1073–1089.
Wikle, C. K., and M. B. Hooten. 2010. A general science-based framework for dynamical spatio-temporal models. Test 19:417–451.
Wikle, C. K., and J. A. Royle. 1999. Space-time dynamic design of environmental monitoring networks. Journal of Agricultural, Biological, and Environmental Statistics 4:489–507.
Wikle, C. K., and J. A. Royle. 2005. Dynamic design of ecological monitoring networks for non-Gaussian spatio-temporal data. Environmetrics 16:507–522.
Williams, B. K., and E. D. Brown. 2020. Scenarios for valuing sample information in natural resources. Methods in Ecology and Evolution 11:1534–1549.
Williams, P. J., M. B. Hooten, G. G. Esslinger, J. N. Womble, J. L. Bodkin, and M. R. Bower. 2019. The rise of an apex predator following deglaciation. Diversity and Distributions 25:895–908.
Williams, P. J., M. B. Hooten, J. N. Womble, and M. R. Bower. 2017*a*. Estimating occupancy and abundance using aerial images with imperfect detection. Methods in Ecology and Evolution 8:1679–1689.
Williams, P. J., M. B. Hooten, J. N. Womble, G. G. Esslinger, and M. R. Bower. 2018. Monitoring dynamic spatio-temporal ecological processes optimally. Ecology 99:524–535.
Williams, P. J., M. B. Hooten, J. N. Womble, G. G. Esslinger, M. R. Bower, and T. J. Heey. 2017*b*. An integrated data model to estimate spatiotemporal occupancy, abundance, and colonization dynamics. Ecology 98:328–336.

Supporting Information

Additional supporting information may be found in the online version of this article at http://onlinelibrary.wiley.com/doi/10.1002/ecy.3573/suppinfo

Open Research

Novel code is provided as Supporting Information and archived on Zenodo. The code for the simulated data example is archived with https://doi.org/10.5281/zenodo.5172768 and the code from the sea otter example is archived with https://doi.org/10.5281/zenodo.5172817. The data used for the sea otter example are available from Esslinger (2019).