Computational Statistics and Data Analysis 179 (2023) 107643

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Adaptive spatial designs minimizing the integrated Bernoulli variance in spatial logistic regression models - with an application to benthic habitat mapping

Susan Anyosa [a,*], Jo Eidsvik [a], Oscar Pizarro [b]

[a] Department of Mathematical Sciences, Norwegian University of Science and Technology, Norway
[b] Australian Centre for Field Robotics, University of Sydney, Australia

ARTICLE INFO

Article history:
Received 29 September 2021
Received in revised form 10 October 2022
Accepted 11 October 2022
Available online 14 October 2022

Repository link: https://github.com/anyosa/adaptive_spatial_designs

Keywords:
Adaptive sampling
Spatial design
Integrated Bernoulli variance
Spatial GLM
Underwater robotics
Habitat mapping

ABSTRACT

Constructing efficient spatial sampling designs is important for monitoring, understanding and characterizing environmental processes acting over large-scale spatial domain. Multiple types of devices can accomplish informative sampling in various ways, and time and computational limitations should be considered too. In-situ sampling is often accurate, but it tends to be very sparse in the vast spatial areas considered for mapping purposes, and careful constructions of experimental sampling designs are needed. A criterion for mapping spatial presence-absence variables is presented. The expected integrated Bernoulli variance criterion is used in order to explore regions with more uncertainty about the binary outcomes. This approach uses a hierarchical Bayesian logistic regression model for the binary variable and develops approximate closed form expressions for the expected integrated Bernoulli variance. The approximations are fast to compute for many designs. The expressions are extended to find adaptive designs in the setting of sequential spatial exploration, where there are limited computation resources and operational time restrictions. In simulation studies the approximations are compared with tedious Monte Carlo sampling methods. Approximations are shown to be accurate and to work for a large spatial grid. An application in benthic habitat mapping is presented. It considers a coral dataset from Australia to demonstrate the suggested sampling design approach.

1. Introduction

Motivated by applications in habitat mapping, we present an approach for constructing efficient spatial sampling designs to characterize presence-absence variables in a spatial domain of interest. Efficiency here refers to fast calculations of sampling designs that enables one to wisely choose which locations to sample. Only then can design comparison be applied in field deployments which are often characterized by limited computing resources and (semi) autonomous operations. Moreover, approaches must be efficient and scale well to the commonly-encountered setting of a large spatial grid.

Reliable habitat maps are important for environmental management and conservation. On-shore wetland characterization (Adamo et al., 2016) and forest vegetation structure mapping (Abbas et al., 2020) are important for biodiversity and conservation as well as for land use policies with associated commitments to satisfy habitat directives, and such habitat

* Corresponding author.
E-mail address: susan.anyosa@ntnu.no (S. Anyosa).

https://doi.org/10.1016/j.csda.2022.107643

maps are quite accurately assessed by a combination of field tests, satellite data, aerial drone photographs and LIDAR data. Near-shore kelp mapping from remote sensing or aerial observations (Schroeder et al., 2019) is critical to monitor the productivity in the sea-life food chain. In the ocean, habitat maps (Brown et al., 2011) tend to be more uncertain and they are relatively difficult to observe. The current standard is to use a towed camera system or to mount imaging tools on an underwater vehicle (Williams et al., 2010), and the processing of these data indicates the habitat classes at the locations where the images are gathered. Such benthic habitat mapping operations are rather costly, and the vehicle deployments must be planned carefully to get the best out of limited resources. Hence, there is clearly a need for realistic statistical modeling and development of ideas from experimental design that can work in this context.

For this challenge of spatial sampling designs to improve a presence-absence map with limited time and resources, there is little point in sampling regions where the habitats are well known (Shields et al., 2020). Instead, one should spend efforts on improving the habitat map in poorly understood parts of the domain. By using a statistical model for the habitat variables and coupling this with a design criterion, one can quantitatively justify that certain sampling designs are more valuable than others. For presence-absence mapping, we propose to use a design criterion building on the integrated Bernoulli variance (IBV). In this setting, the design that minimizes the expected spatially IBV, see e.g. Bect et al. (2019) and Fossum et al. (2021), is preferred among a number of candidate designs. We prefer the design with the smallest expected IBV because it provides the largest expected uncertainty reduction (Bect et al., 2019). We outline an effective computational approximation of the expected IBV (EIBV) criterion in the setting of a Bayesian logistic regression model. Here, the class probabilities depend on covariates via a latent variable regression model including a smooth spatial dependence structure. The results developed in this paper rely on Gaussian approximations for the latent variables as well as results on incomplete Gaussian integrals of the logistic function (Demidenko, 2013).

There has lately been growing interest in sampling designs for generalized linear models (GLMs), see e.g. Ryan et al. (2016), Overstall and Woods (2017) and Woods et al. (2017). In this paper, we study the EIBV criterion for GLMs, which in some sense is more target-oriented than commonly used A and D designs, see e.g. Ryan et al. (2016). Hence, our proposal aligns more with that of active learning approaches to sampling designs (Schein and Ungar, 2007; Yang and Loog, 2018).

As in many other situations with spatial designs, the underwater vehicle deployments for habitat mapping tend to be done sequentially or operated semi-autonomously. In some cases, the vehicles have computing units on-board and the design can be modified during the operation. In other cases, the deployments are done in batches, with the processing done on a mother-ship between the sequential deployments. This sequential way of sampling naturally leads to opportunities for adaptive designs. Compared with earlier efforts emphasizing spatially balanced sampling (Foster et al., 2020) or habitat class coverage (Bender et al., 2013; Shields et al., 2020), we formalize the statistical modeling parts via the hierarchical spatial GLM and put this in a framework of adaptive sampling design focused on minimizing the EIBV.

The main contributions are deriving closed form approximate expressions for the EIBV criterion under Gaussian hierarchical GLM assumptions, embedding the closed form EIBV expressions in an adaptive framework which eases operations with (semi) autonomous vehicles and presenting a scalable implementation of design criterion and model updating expressions using Fourier domain calculations working for large-size high-resolution spatial domains assuming relatively sparse sampling.

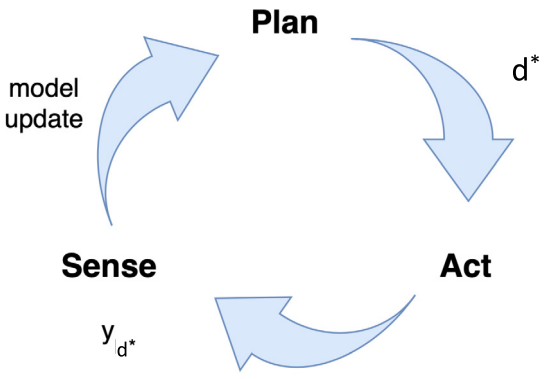
In Section 2 we introduce notions of adaptive sampling designs and of IBV as sampling criterion. In Section 3 we present the hierarchical Bayesian logistic model and the Gaussian approximations, and build on these to develop new approximate expressions for the EIBV. In Section 4 we show results of the suggested approach via simulation studies. In Section 5 we apply the methods to a coral field dataset from Lizard Island in Australia. We conclude with closing remarks in Section 6. The Appendix contains detail descriptions of the mathematical results that make our approach scale for large-size models. The link for the source codes is in the Code availability statement.

2. Problem specification

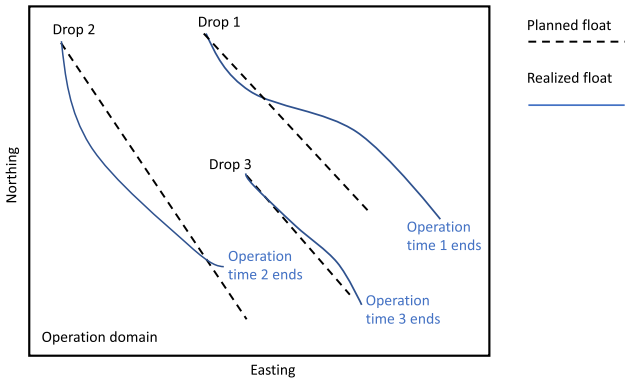
We consider a geographical domain $\mathcal{M} \subset \mathcal{R}^2$. A spatial location is denoted by $\boldsymbol{s} \in \mathcal{M}$. Albeit mostly using continuous space notation, we will denote N grid locations in the domain by $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N$, and use this whenever appropriate. At each location, the indicator variable $y_{\boldsymbol{s}} \in \{0, 1\}$ represents absence or presence data. The probability of presence is denoted by $P(y_{\boldsymbol{s}} = 1)$, $\boldsymbol{s} \in \mathcal{M}$. Features are denoted by a length p covariate vector $\boldsymbol{x}_{\boldsymbol{s}}$, $\boldsymbol{s} \in \mathcal{M}$. The covariates are assumed known, say from acoustic or remote sensing data sources, while the indicator variables must be measured in-situ, and this leads to sparse sampling. As the presence-absence probabilities are often rather uncertain (far from 0 and 1), one aims to improve the prediction of a categorical map by collecting more in-situ data at select locations or survey lines. Such spatial data will be informative away from measurement locations because of spatial correlation. Moreover, the data gathering will lead to an improved understanding of how the covariates relate to the presence-absence data.

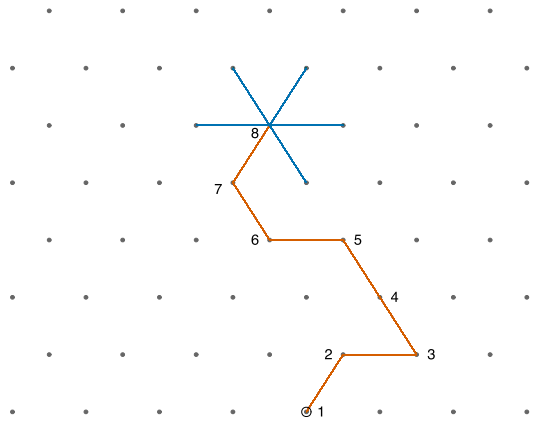
2.1. Design context

With limited capacity for data gathering in large-size spatial domains, one must prioritize the available sampling resources. This means that in-situ measurements must be collected along survey lines or trajectories that carry the most information. We consider the common situation of adaptive sampling designs, where one gathers data in batches and plans the next design depending on the results from the previous batches.



(a)





(b) (c)

Fig. 1. (a) The loop shows the adaptive design d^* chosen based on the updated model, and $\boldsymbol{y}_{d^*}$ is the data collected in this design. (b) Illustration of three sequential deployments of marine floats where the design choice involves drop point and operation time for each deployment. (c) An adaptive path example on a waypoint graph, with starting point marked with a circle. There are five planning choices at the current point (8th batch) since one will not return to a location with known habitat.

A design at batch or stage $t = 1, \ldots$ is denoted $d_t \subset \mathcal{D}_t$, where $\mathcal{D}_t$ is the set of all possible designs at this stage. The presence-absence data collected in a design is denoted by $\boldsymbol{y}_{d_t} = (y_{d_t,1}, \ldots, y_{d_t,|d_t|})$. We further let $\boldsymbol{Y}_{t-1} = (\boldsymbol{y}_{d_1}, \ldots, \boldsymbol{y}_{d_{t-1}})$ be the concatenation of data gathered up to stage $t-1$, with the associated designs $\boldsymbol{D}_{t-1} = (d_1, \ldots, d_{t-1})$. At the initial stage, before the first data batch, we have $\boldsymbol{D}_0 = \emptyset$.

The setting with adaptive sampling designs is compactly illustrated by the sense-plan-act workflow in Fig. 1(a), which is commonly used in control engineering applications. At each stage of the adaptive sampling scheme, the planning returns the best design which is denoted d^*. When committing to this design, the data $\boldsymbol{y}_{d^*}$ are gathered. These data are used to update the model, and this again forms the basis for planning the best design at the next stage.

Batch sequential designs cover a range of possibilities going from a single stage (static) design to a batch size of one where the planning is done after only a single observation. The batch size and the number of stages will depend on the situation: There is a balance between the exploration gain one gets from frequent model updating and planning, and the flexibility to do so in the operation. We next provide examples of two cases that are relevant for our application of benthic habitat mapping. These are illustrated in Fig. 1(b) and (c).

Fig. 1(b) illustrates the situation where a benthic imaging float is dropped at a certain location and then drifts with the ocean current for a set operation time. During this deployment it gathers image data indicating presence-absence of a specific bottom type / biotope (for instance coral habitat). Deployments are done in a sequential setting, with several drops and operation times selected in an order. In this illustration, there is some difference between the planned paths and the realized paths which will be the case in practice because there is uncertainty in the currents. This kind of float has limited flexibility to modify its trajectory over the seafloor. Moreover, with its simplest setting, the float has no capabilities for processing or analyzing data onboard. This is only done when it is recovered. Computing units at a support vessel

will then load the float image data, process and analyze this, update the model, and lay out a new plan for the next drop location. Fig. 1(c) illustrates an autonomous underwater vehicle (AUV) operating on a waypoint graph composed of spatial grid nodes. In this instance there are six possible sampling directions at a waypoint. AUVs generally have some onboard computing resources, and this allows adaptation to occur as the AUV acts by moving to the most promising design waypoint. With the onboard resources, frequent model updates means that only some data is collected in each batch, i.e. between two waypoints in the display.

2.2. Expected integrated Bernoulli variance criterion for mapping

In categorical mapping problems, it is natural to use the IBV to represent the prediction uncertainty. We have

$$\text{IBV} = \int_{\boldsymbol{s}\in\mathcal{M}} P(y_{\boldsymbol{s}}=1)[1-P(y_{\boldsymbol{s}}=1)]d\boldsymbol{s}, \tag{1}$$

which will be dominated by locations where the probability is close to 0.5.

We now consider the situation where one has interest in gathering information to reduce the IBV. The evaluation of different sampling designs involves the expectation over data $\boldsymbol{y}_{d_t}$ in the possible designs at each stage $t = 1, \ldots$. In the sequential setting, the choice of a design d_t will be influenced by the observations made in earlier designs $\boldsymbol{D}_{t-1}$. The optimal design (in theory) requires looking into the future, beyond just the next single time step (as is done in this paper). There is an immediate reduction in the EIBV at the current stage and a continuation value of committing to a design for expected reduction in the IBV at future stages. As the solution to this kind of a dynamic program is only possible to compute for small-size problems with one or two stages, one ends up using approximate roll-out heuristic strategies in practice. In roll-out strategies one looks ahead at the next stage(s) and evaluates the outcomes for each possible future state to approximate the optimal decision at the current stage (Goodson et al., 2017). To simplify the optimization process, a myopic strategy is used in the current paper. This heuristic approach finds the optimal design at each stage, under a working assumption that it will be the last design step. The adaptive designs are then selected according to minimum EIBV at the current stage:

$$d_t^* = \text{argmin}_{d_t\in\mathcal{D}_t}\{\text{EIBV}(d_t|\boldsymbol{D}_{t-1})\}. \tag{2}$$

The calculations for each possible design are based on

$$\text{EIBV}(d_t|\boldsymbol{D}_{t-1}) = \sum_{\boldsymbol{y}_{d_t}} \int_{\boldsymbol{s}\in\mathcal{M}/\boldsymbol{D}_t} p_{\boldsymbol{s}}(\boldsymbol{y}_{d_t})[1-p_{\boldsymbol{s}}(\boldsymbol{y}_{d_t})]d\boldsymbol{s}P(\boldsymbol{y}_{d_t}|\boldsymbol{Y}_{t-1}), \qquad p_{\boldsymbol{s}}(\boldsymbol{y}_{d_t}) = P(y_{\boldsymbol{s}}=1|\boldsymbol{y}_{d_t},\boldsymbol{Y}_{t-1}), \tag{3}$$

where $\mathcal{M}/\boldsymbol{D}_t$ denotes the spatial domain with all data locations in designs $(d_t, \boldsymbol{D}_{t-1})$ removed, and where $P(\boldsymbol{y}_{d_t}|\boldsymbol{Y}_{t-1})$ is the probability mass function of the data $\boldsymbol{y}_{d_t}$, conditional on the currently available data. The expectation over these data in (3) involves the sum over $2^{|d_t|}$ terms in the sample space of $\boldsymbol{y}_{d_t}$. Going back to the sense-plan-act loop in Fig. 1(a), the optimal design selected in (2) augments the set $\boldsymbol{D}_t = (d_t^*, \boldsymbol{D}_{t-1})$ and data in the design is collected so that $\boldsymbol{Y}_t = (\boldsymbol{y}_{d_t^*}, \boldsymbol{Y}_{t-1})$. The data is used to update the model, and this forms the basis for the next stage design.

Note that by interchanging the order of sum and integration in (3), it can be interpreted as an integrated expected Bernoulli variance (EBV). At each location $\boldsymbol{s}$, we then have

$$\text{EBV}_{\boldsymbol{s}}(d_t|\boldsymbol{D}_{t-1}) = \sum_{\boldsymbol{y}_{d_t}} p_{\boldsymbol{s}}(\boldsymbol{y}_{d_t})[1-p_{\boldsymbol{s}}(\boldsymbol{y}_{d_t})]P(\boldsymbol{y}_{d_t}|\boldsymbol{Y}_{t-1}), \tag{4}$$

and $\text{EIBV}(d_t|\boldsymbol{D}_{t-1}) = \int_{\boldsymbol{s}\in\mathcal{M}/\boldsymbol{D}_t} \text{EBV}_{\boldsymbol{s}}(d_t|\boldsymbol{D}_{t-1})d\boldsymbol{s}$. This EBV is useful as a diagnostic map to see which parts of the domain that are expected to have a large reduction in the IBV (Fossum et al., 2021). It further simplifies our statistical derivations below, which are indicated for one location only.

Note that in an operational setting one can modify the EIBV expression, for instance by including a compensation for the batch operation and data processing time. It is then possible to solve for the best design based on how much information is gained per time unit and according to the resource allocation.

3. Design calculations for a Bayesian latent Gaussian logistic model

Evaluations of the IBV and EIBV require a model formulation. In the following, we describe a hierarchical Bayesian logistic model with Gaussian modeling assumptions for the latent variables. In Section 3.1 and 3.2 we outline the core elements of the suggested approximation. To avoid overly complex notation this is presented for the first stage only, denoting the design by $d \in \mathcal{D}$. In Section 3.3 we put this in the context of multiple stages, explaining the adaptive approach and the resulting algorithm.

3.1. Bayesian latent Gaussian logistic model

The logistic model for the presence-absence variables at location $\boldsymbol{s} \in \mathcal{M}$ entails probabilities

$$P(y_{\boldsymbol{s}}|\eta_{\boldsymbol{s}}) = \frac{\exp(\eta_{\boldsymbol{s}})^{y_{\boldsymbol{s}}}}{1+\exp(\eta_{\boldsymbol{s}})}, \quad y_{\boldsymbol{s}} \in \{0,1\}, \tag{5}$$

where the linear predictor $\eta_{\boldsymbol{s}} = \boldsymbol{x}_{\boldsymbol{s}}^T\boldsymbol{\beta} + w_{\boldsymbol{s}}$. This contains a regression effect incorporating the features $\boldsymbol{x}_{\boldsymbol{s}}$ at location $\boldsymbol{s}$ and a spatially structured effect $w_{\boldsymbol{s}}$. We assume conditional independence between presence-absence variables at different locations, given the linear predictors. The regression parameters are modeled by a Gaussian prior distribution; $\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$, where the mean and covariance matrix are specified from previous knowledge or preliminary data. The latent spatial variable $w_{\boldsymbol{s}}$, $\boldsymbol{s} \in \mathcal{M}$, is modeled by a zero mean Gaussian process prior with a Matern covariance kernel with roughness parameter 3/2, i.e. $\Sigma_w(\boldsymbol{s},\boldsymbol{s}') = \zeta^2(1+\phi h_{\boldsymbol{s},\boldsymbol{s}'})e^{-\phi h_{\boldsymbol{s},\boldsymbol{s}'}}$, where $\zeta > 0$ is the standard deviation parameter, $\phi > 0$ the correlation decay parameter and $h_{\boldsymbol{s},\boldsymbol{s}'}$ is the Euclidean distance between locations $\boldsymbol{s}$ and $\boldsymbol{s}'$. (We fit this covariance model to data in Section 5.) In vector form, assuming the discretization of the spatial domain, we have $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_N)'$ with mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)'$ and $N \times N$ covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{X}\boldsymbol{\Sigma}_\beta\boldsymbol{X}' + \boldsymbol{\Sigma}_w$, where $\boldsymbol{X}$ is the collection of covariates with $\boldsymbol{x}_{\boldsymbol{s}}^T$ in the rows and $\boldsymbol{\Sigma}_w$ is the covariance matrix of all spatial effects at the N grid locations. The marginal density $p(\eta_{\boldsymbol{s}})$ of the linear predictor is then Gaussian with mean $\mu_{\boldsymbol{s}} = \boldsymbol{x}_{\boldsymbol{s}}^T\boldsymbol{\mu}_\beta$ and variance $\sigma_{\boldsymbol{s}}^2 = \boldsymbol{x}_{\boldsymbol{s}}^T\boldsymbol{\Sigma}_\beta\boldsymbol{x}_{\boldsymbol{s}} + \zeta^2$. By using these modeling assumptions and expression (5), the marginal presence probability is

$$P(y_{\boldsymbol{s}} = 1) = \int_{\eta_{\boldsymbol{s}}} P(y_{\boldsymbol{s}} = 1|\eta_{\boldsymbol{s}})p(\eta_{\boldsymbol{s}})d\eta_{\boldsymbol{s}}. \tag{6}$$

The EIBV design criterion relies on computing conditional expressions for the linear predictors; $p(\eta_{\boldsymbol{s}}|\boldsymbol{y}_d)$, the presence-absence probabilities; $P(y_{\boldsymbol{s}}|\boldsymbol{y}_d)$, and marginalizing over the probability mass function $P(\boldsymbol{y}_d)$ of the data in design d. To efficiently compute the EIBV design criteria and hence enable quick evaluation of a large set of possible designs, we next derive approximate closed form solutions based on the hierarchical latent Gaussian modeling assumptions. In doing so, we let the length $|d|$ vector $\boldsymbol{\mu}_d$ denote the mean value elements of the linear predictor at the design locations, and $\boldsymbol{\Sigma}_d$ is the associated $|d| \times |d|$ covariance matrix. Moreover, $\boldsymbol{\Sigma}_{\cdot,d}$ is the $N \times |d|$ matrix with cross-covariance terms between the linear predictors at all grid locations and those in the design.

3.2. Main elements of a closed form approximation for EIBV

We summarize the main building blocks required for the closed form approximation of the EIBV. Recall that this is presented for one stage only. The adaptive elements are outlined in Section 3.3.

Logistic approximation by Gaussian cumulative distribution function

The integral in expression (6), without any data gathered yet, can be approximated by using $\frac{e^x}{1+e^x} \approx \Phi(\alpha x)$ with coefficient $\alpha = 0.58$ for the logistic function in (5), see e.g. Demidenko (2013). Here, Φ represents the cumulative distribution function (cdf) of a standard normal variable. Inserting this into equation (6), we get

$$P(y_{\boldsymbol{s}} = 1) \approx \int_{-\infty}^{\infty} \Phi(\alpha\eta_{\boldsymbol{s}}) \frac{1}{\sqrt{2\pi\sigma_{\boldsymbol{s}}^2}} \exp\left(-\frac{(\eta_{\boldsymbol{s}} - \mu_{\boldsymbol{s}})^2}{2\sigma_{\boldsymbol{s}}^2}\right) d\eta_{\boldsymbol{s}} = P(Z < \alpha\eta_s) = \Phi\left(\frac{\alpha\mu_{\boldsymbol{s}}}{\sqrt{1+\alpha^2\sigma_{\boldsymbol{s}}^2}}\right), \tag{7}$$

where $Z \sim N(0,1)$, and the interest is in the probability of this variable being smaller than another independent Gaussian variable $\alpha\eta_{\boldsymbol{s}}$. Standardization of the linear combination of the two gives the result.

The Bernoulli variance (BV) at a location $\boldsymbol{s}$ is then $\Phi(-m_{\boldsymbol{s}})\Phi(m_{\boldsymbol{s}}) = \Phi_2(m_{\boldsymbol{s}}, -m_{\boldsymbol{s}}, 0)$, where $m_{\boldsymbol{s}} = \alpha\mu_s/\sqrt{1+\alpha^2\sigma_{\boldsymbol{s}}^2}$ and where $\Phi_2(u_1, u_2; q)$ denotes the bivariate standard normal cdf with correlation q, evaluated at point (u_1, u_2). Fig. 2 (left) shows the BV for two cases with $\mu_{\boldsymbol{s}} = 0$ (top) and $\mu_{\boldsymbol{s}} = 1$ (bottom). In this illustration we set $\sigma_{\boldsymbol{s}}^2 = 1$. Contours of the bivariate Gaussian density with 0 correlation are shown in colored lines. The dashed vertical and horizontal lines are at $\pm m_{\boldsymbol{s}}$. Naturally, the BV is largest when the mean is 0 giving a 0.5 chance of presence-absence and BV$= 0.25$.

A probit link would allow exact marginalization, but our suggested approach is then a bit more numerically challenging in the Gaussian approximations that are explained next.

Quadratic expansions and Gaussian approximations

The logistic link function in the probabilistic model for the presence-absence in (5) can be linearized. Setting linearization point $\eta_{\boldsymbol{s}}^*$, a quadratic expansion gives

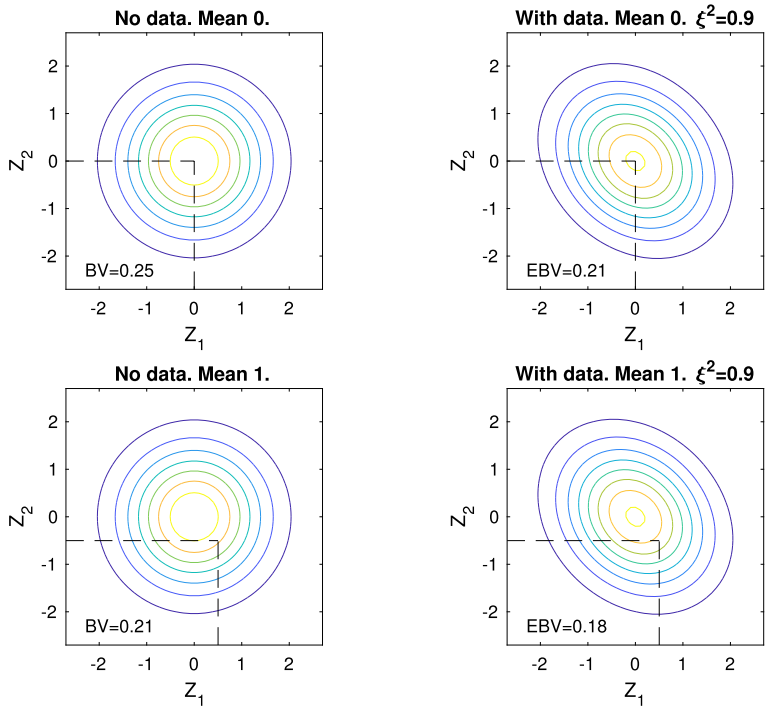


Fig. 2. Illustration of approximate Bernoulli variance (BV) calculations (left) for a case with mean 0 (top) and 1 (bottom). The approximate expected Bernoulli variance (EBV) calculations (right) from data show how the expected uncertainty is reduced by the design. Dashed lines indicate regions where cumulative probabilities are calculated.

$$p(y_{\boldsymbol{s}}|\eta_{\boldsymbol{s}}) \propto \exp\left(y_{\boldsymbol{s}}\eta_{\boldsymbol{s}} - \log(1+e^{\eta^*_{\boldsymbol{s}}}) - \frac{e^{\eta^*_{\boldsymbol{s}}}}{1+e^{\eta^*_{\boldsymbol{s}}}}(\eta_{\boldsymbol{s}}-\eta^*_{\boldsymbol{s}}) - \frac{e^{\eta^*_{\boldsymbol{s}}}}{2(1+e^{\eta^*_{\boldsymbol{s}}})^2}(\eta_{\boldsymbol{s}}-\eta^*_{\boldsymbol{s}})^2 - \dots\right),$$
$$\approx \exp\left(-\frac{(z_{\boldsymbol{s}}-\eta_{\boldsymbol{s}})^2}{2\kappa^2_{\boldsymbol{s}}}\right), \tag{8}$$

where $z_{\boldsymbol{s}}$ is a function of the data $y_{\boldsymbol{s}}$ and linearization point $\eta^*_{\boldsymbol{s}}$, and we have

$$z_{\boldsymbol{s}} = z_{\boldsymbol{s}}(y_{\boldsymbol{s}}, \eta^*_{\boldsymbol{s}}) = \frac{y_{\boldsymbol{s}} - \frac{e^{\eta^*_{\boldsymbol{s}}}}{1+e^{\eta^*_{\boldsymbol{s}}}} + \eta^*_{\boldsymbol{s}}\frac{e^{\eta^*_{\boldsymbol{s}}}}{(1+e^{\eta^*_{\boldsymbol{s}}})^2}}{\frac{e^{\eta^*_{\boldsymbol{s}}}}{(1+e^{\eta^*_{\boldsymbol{s}}})^2}}, \qquad \kappa^2_{\boldsymbol{s}} = \kappa^2_{\boldsymbol{s}}(\eta^*_{\boldsymbol{s}}) = \frac{(1+e^{\eta^*_{\boldsymbol{s}}})^2}{e^{\eta^*_{\boldsymbol{s}}}}. \tag{9}$$

By collecting the transformed measurements in a length $|d|$ vector $\boldsymbol{z}_d(\boldsymbol{y}_d, \boldsymbol{\eta}^*)$, and forming a diagonal $|d| \times |d|$ matrix $\boldsymbol{K}_d = \text{diag}\left\{\kappa^2_{\boldsymbol{s}}(\eta^*_{\boldsymbol{s}}); \boldsymbol{s} \in d\right\}$ for the part that enters in the quadratic component in (8), it is straightforward to form a Gaussian approximation for the posterior distribution of $\eta_{\boldsymbol{s}}$ given $\boldsymbol{y}_d$. In vector form, the linear predictor $\boldsymbol{\eta}$ has approximate conditional mean

$$\hat{\boldsymbol{\mu}} = \boldsymbol{\mu} + \boldsymbol{\Sigma}_{\cdot,d}\left[\boldsymbol{\Sigma}_d + \boldsymbol{K}_d\right]^{-1}(\boldsymbol{z}_d(\boldsymbol{y}_d, \boldsymbol{\eta}^*) - \boldsymbol{\mu}_d), \tag{10}$$

and approximate conditional covariance matrix

$$\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{\cdot,d}\left[\boldsymbol{\Sigma}_d + \boldsymbol{K}_d\right]^{-1}\boldsymbol{\Sigma}^T_{\cdot,d}, \qquad (\hat{\sigma}^2_{\boldsymbol{s}_1}, \dots, \hat{\sigma}^2_{\boldsymbol{s}_N}) = \text{diag}(\hat{\boldsymbol{\Sigma}}). \tag{11}$$

Conveniently, the covariance does not depend on the data outcomes. The most natural linearization point is $\eta^*_{\boldsymbol{s}} = \mu_{\boldsymbol{s}}$. When the scheme in (10) is iterated until convergence, with new linearization points from the approximate conditional mean, one can fit a Gaussian approximation at the posterior mode (Eidsvik et al., 2009). In this case with iteration, the approximate conditional covariance in (11) depends on the data via the modified linearization point. However, for design purposes, and in settings with relatively large sampling lines, we expect a first step analysis to be sufficiently accurate on expectation over the transformed data. One can also get a similar approximation by taking the expected values in the quadratic expansion of $e^{\eta_{\boldsymbol{s}}}$, see e.g. Evangelou and Eidsvik (2017). We check the properties of the approximation in the simulation study in Section 4.

EIBV approximation I

At location $\boldsymbol{s} \in \mathcal{M}$, an approximate Gaussian posterior with mean $\hat{\mu}_{\boldsymbol{s}} = \hat{\mu}_{\boldsymbol{s}}(\boldsymbol{y}_d)$ computed by iteration for the linearization point in (10) and variance $\hat{\sigma}^2_{\boldsymbol{s}}$ in (11) can be used to approximate the logistic expectation in (7). Similar to the earlier prior expressions, after observing data $\boldsymbol{y}_d$ we get

$$P(y_{\boldsymbol{s}} = 1|\boldsymbol{y}_d) \approx \int_{-\infty}^{\infty} \Phi(\alpha\eta_{\boldsymbol{s}}) \frac{1}{\sqrt{2\pi\hat{\sigma}^2_{\boldsymbol{s}}}} \exp\left(-\frac{(\eta_{\boldsymbol{s}} - \hat{\mu}_{\boldsymbol{s}})^2}{2\hat{\sigma}^2_{\boldsymbol{s}}}\right) d\eta_{\boldsymbol{s}} = \Phi\left(\frac{\alpha\hat{\mu}_{\boldsymbol{s}}}{\sqrt{1+\alpha^2\hat{\sigma}^2_{\boldsymbol{s}}}}\right) = \hat{p}_{\boldsymbol{s}}(\boldsymbol{y}_d). \tag{12}$$

By either using direct summation over the sample space of $\boldsymbol{y}_d$ or Monte Carlo sampling for large-size designs, we get the following EIBV approximation:

$$\text{EIBV}^I(d) = \sum_{\boldsymbol{y}_d \in \{0,1\}^{|d|}} \sum_{\boldsymbol{s} \in \mathcal{M}/d} \hat{p}_{\boldsymbol{s}}(\boldsymbol{y}_d)\left[1 - \hat{p}_{\boldsymbol{s}}(\boldsymbol{y}_d)\right] P(\boldsymbol{y}_d). \tag{13}$$

EIBV approximation II

Recall that the approximate posterior variances $\hat{\sigma}^2_{\boldsymbol{s}}$ in (11) do not depend on the data when the linear approximation in (9) is used without iteration. This means that the probabilities $\hat{p}_{\boldsymbol{s}}(\boldsymbol{y}_d)$ in (12) as a function over the data samples in (13) only depend on the approximate posterior mean $\hat{\mu}_{\boldsymbol{s}} = \hat{\mu}_{\boldsymbol{s}}(\boldsymbol{y}_d)$.

By building on the Gaussian approximation to the posterior with the transformed measurement $\boldsymbol{z}_d$ in (10) and (11), we have that

$$\begin{aligned} &\hat{\boldsymbol{\mu}} \sim N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\cdot,d}\left[\boldsymbol{\Sigma}_d + \boldsymbol{K}_d\right]^{-1} \boldsymbol{\Sigma}_{\cdot,d}^T\right), \\ &\hat{\mu}_{\boldsymbol{s}} \sim N(\mu_{\boldsymbol{s}}, \xi^2_{\boldsymbol{s}}), \qquad \xi^2_{\boldsymbol{s}} = \boldsymbol{\Sigma}_{\boldsymbol{s},d}\left[\boldsymbol{\Sigma}_d + \boldsymbol{K}_d\right]^{-1} \boldsymbol{\Sigma}_{\boldsymbol{s},d}^T, \end{aligned} \tag{14}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{s},d}$ is the cross-covariance between location $\boldsymbol{s}$ and design locations. The latter expression for $\xi^2_{\boldsymbol{s}}$ represents the approximate reduction in variance resulting from the conditioning to data, and this part plays an important role in the closed form expressions.

As the integrand expression for the EIBV in (3) is now a function of the conditional mean value $\hat{\mu}_{\boldsymbol{s}}(\boldsymbol{y}_d)$, while other elements are fixed, the expectation is done with respect to this conditional mean value. Since it is approximately Gaussian (expression (14)), we have

$$\text{EBV}_{\boldsymbol{s}}(d) \approx \int_{-\infty}^{\infty} \Phi\left(\frac{\alpha\hat{\mu}_{\boldsymbol{s}}}{\sqrt{1+\alpha^2\hat{\sigma}^2_{\boldsymbol{s}}}}\right) \Phi\left(-\frac{\alpha\hat{\mu}_{\boldsymbol{s}}}{\sqrt{1+\alpha^2\hat{\sigma}^2_{\boldsymbol{s}}}}\right) p(\hat{\mu}_{\boldsymbol{s}}) d\hat{\mu}_{\boldsymbol{s}} = P(Z_1 \le M, Z_2 \le -M), \tag{15}$$

where $p(\hat{\mu}_{\boldsymbol{s}})$ is the Gaussian probability density function defined by (14), and where $M = r_{\boldsymbol{s}}\hat{\mu}_{\boldsymbol{s}}$ with $r_{\boldsymbol{s}} = \frac{\alpha}{\sqrt{1+\alpha^2\hat{\sigma}^2_{\boldsymbol{s}}}}$, and Z_1 and Z_2 are two independent standard normal variables. The expression involves probability calculations for two linear combinations of Gaussian variables, and the result is a bivariate cdf over standardized variables $U_1 = (Z_1 - M + r_{\boldsymbol{s}}\mu_{\boldsymbol{s}})/\sqrt{1+r^2_{\boldsymbol{s}}\xi^2_{\boldsymbol{s}}}$ and $U_2 = (Z_2 + M - r_{\boldsymbol{s}}\mu_{\boldsymbol{s}})/\sqrt{1+r^2_{\boldsymbol{s}}\xi^2_{\boldsymbol{s}}}$ with $\text{Corr}(U_1, U_2) = -\xi^2_{\boldsymbol{s}} r^2_{\boldsymbol{s}}/(1+r^2_{\boldsymbol{s}}\xi^2_{\boldsymbol{s}})$. We then get the following approximation:

$$\text{EIBV}^{II}(d) = \int_{\boldsymbol{s} \in \mathcal{M}/d} \text{EBV}^{II}_{\boldsymbol{s}}(d) d\boldsymbol{s}, \qquad \text{EBV}^{II}_{\boldsymbol{s}}(d) = \Phi_2\left(r_{\boldsymbol{s}}\mu_{\boldsymbol{s}}/\sqrt{1+r^2_{\boldsymbol{s}}\xi^2_{\boldsymbol{s}}}, -r_{\boldsymbol{s}}\mu_{\boldsymbol{s}}/\sqrt{1+r^2_{\boldsymbol{s}}\xi^2_{\boldsymbol{s}}}; -r^2_{\boldsymbol{s}}\xi^2_{\boldsymbol{s}}/(1+r^2_{\boldsymbol{s}}\xi^2_{\boldsymbol{s}})\right). \tag{16}$$

We expand on the illustrative case from Fig. 2 to gain insight in the new EIBV approximation. Fig. 2 (right) shows the basis underlying the EBV calculation for two cases with $\mu_{\boldsymbol{s}} = 0$ (top) and $\mu_{\boldsymbol{s}} = 1$ (bottom). Here, we set $\sigma_{\boldsymbol{s}} = 1$ and $\xi^2_{\boldsymbol{s}} = 0.9$. The EBV is smaller in the case with $\mu = 1$ (bottom), but the reduction from the BV (left displays) is largest for $\mu = 0$ (top) where the BV is bigger. The dashed vertical and horizontal lines are at $\pm r_{\boldsymbol{s}}\mu_{\boldsymbol{s}}/\sqrt{1+r^2_{\boldsymbol{s}}\xi^2_{\boldsymbol{s}}}$. Note that a variance reduction of $\xi^2_{\boldsymbol{s}} = 0.9$ is rather large. Such a reduction would occur near design locations or at locations that have very similar covariates as in the design, or both of these combined (see Equation (14)). Even with this level of variance reduction, the EBV is quite large. This tells us that there is limited information in the binary observation, which makes it even more important to care about the experimental design of binary data. To achieve a smaller EBV the contours of the Gaussian pdf in Fig. 2 (right) must be tighter to indicate more substantial negative correlation in the bivariate Gaussian density. The correlation is $-r^2_{\boldsymbol{s}}\xi^2_{\boldsymbol{s}}/(1+r^2_{\boldsymbol{s}}\xi^2_{\boldsymbol{s}})$, where $r^2_{\boldsymbol{s}}\xi^2_{\boldsymbol{s}} = \alpha^2\xi^2_{\boldsymbol{s}}/(1+\alpha^2(\sigma^2_{\boldsymbol{s}} - \xi^2_{\boldsymbol{s}}))$. This correlation goes to -1 when the variance reduction $\xi^2_{\boldsymbol{s}}$ goes to infinity faster than $(\sigma^2_{\boldsymbol{s}} - \xi^2_{\boldsymbol{s}}) > 0$.

Scalable computations using Fourier domain matrix calculations

The calculations of $\text{EBV}^{II}_{\boldsymbol{s}}(d)$ in (16) must be done for all $\boldsymbol{s} \in \mathcal{M}/d$ and for several possible designs d. For each design, we need to factorize a matrix $\boldsymbol{\Sigma}_d + \boldsymbol{K}_d$ and then propagate the effect from the observation locations via length $1 \times |d|$ vector $\boldsymbol{\Sigma}_{\boldsymbol{s},d}$ or do so in a matrix operation with $\boldsymbol{\Sigma}_{\cdot,d}$. We next outline an efficient approach splitting the spatial and regression elements to achieve a scalable implementation. (The details involving Fourier domain calculations are in Appendix A.)

We combine effects in a size $N + p$ vector $\boldsymbol{v} = (\boldsymbol{w}', \boldsymbol{\beta}')'$, and then use the appropriate linear combination of this to form the predictor. In the discretized model formulation, $\boldsymbol{v}$ is Gaussian distributed with mean $(0, \mathbf{X}\boldsymbol{\mu}_\beta)$ and covariance matrix

$$\begin{pmatrix} \boldsymbol{\Sigma}_w & 0 \\ 0 & \boldsymbol{\Sigma}_\beta \end{pmatrix}.$$

The selection of design points and its linear predictors can be specified by $\boldsymbol{\eta}_d = [\boldsymbol{A}_d, \boldsymbol{X}_d]\boldsymbol{v}$, where the matrix $\boldsymbol{A}_d$ has $|d|$ rows and N columns, and $A_d(i, j) = 1$ if j is the location of the ith measurement, and otherwise $A_d(i, j) = 0$. Further, the $|d| \times p$ matrix $\boldsymbol{X}_d$ collects the covariates in these design locations. To form the diagonal of the matrix in (14), we do Fourier domain computations for the spatial part $w_{\boldsymbol{s}}$, while the regression parts are done explicitly. The spatial variable is in this way embedded on a torus and the associated covariance matrix representation for $\boldsymbol{\Sigma}_w$ is circulant. The use of Fourier domain calculations means that the design problems can be studied for large-scale spatial problems. An underlying assumption is that of a stationary process variable $w_{\boldsymbol{s}}$ and that the design size $|d|$ is much smaller than the number of spatial grid nodes N.

3.3. Sequential setup and adaptive algorithm

In the derivations of the closed form solution in Section 3.2 we considered designs d at the initial time only, ignoring the stages $t = 1, 2, \ldots$ of sequential designs d_t. We will now bring this back into our suggested methods for adaptive sampling under the hierarchical modeling assumptions with a latent Gaussian structure.

Relying on the myopic strategy in (2)-(3), the optimal design at stage t is now based on the updated probabilistic model $\hat{p}_{\boldsymbol{s}}(\boldsymbol{y}_{d_t}|\boldsymbol{Y}_{t-1}) \approx P(y_{\boldsymbol{s}} = 1|\boldsymbol{y}_{d_t}, \boldsymbol{Y}_{t-1})$. This probability is refined along with the Gaussian approximation to the linear predictor over the sampling stages. The closed form results derived in Section 3.2 still hold under the Bayesian logistic regression model, but now these are done conditional on the currently available data $\boldsymbol{Y}_{t-1}$. The EIBV is computed over the data $\boldsymbol{y}_{d_t}$ in design $d_t \in \mathcal{D}_t$, conditional on $\boldsymbol{Y}_{t-1}$.

Conditional on data $\boldsymbol{Y}_{t-1}$, the updated mean, see (14), is refined by iterating the linearization point, given $\boldsymbol{Y}_{t-1}$. This means that a Gaussian approximation is fitted at the posterior mode, given all the data so far. This mode is also used as a linearization point before the next stage data $\boldsymbol{y}_{d_t}$ is evaluated. This means that the preferred design criterion EIBV^{II} is still fast to compute, as the evaluation for new design d_t uses no additional linearization for the predictive mean as a function of data $\boldsymbol{y}_{d_t}$. The main parts of the adaptive sampling strategy are summarized in Algorithm 1. The details around the matrix factorizations used for the closed form solution in a model for $\boldsymbol{y}_{d_t}$, given $\boldsymbol{Y}_{t-1}$ are outlined in Appendix B. The Fourier domain calculations are used for the spatial parts involved in this EIBV evaluation (see Appendix A). A working assumption for these computations is that the number of design points is very small compared with the grid size.

Algorithm 1 Adaptive design approach.

$\boldsymbol{D}_0 = \emptyset$. $\boldsymbol{Y}_0 = \emptyset$
Specify initial Gaussian model with $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.
for $t \leftarrow 1$ **to** $\mathcal{T}$ **do**
  for $d_t \leftarrow 1$ **to** $|\mathcal{D}_t|$ **do**
    Compute the approximation $\text{EIBV}^{II}(d_t|\boldsymbol{D}_{t-1})$. ▷ Equation (16)
  end for
  Choose next design $d_t^* = \text{argmin}_{d_t} \text{EIBV}^{II}(d_t|\boldsymbol{D}_{t-1})$. ▷ Equation (2)
  Set $\boldsymbol{D}_t = \{\boldsymbol{D}_{t-1}, d_t^*\}$.
  Observe the data $\boldsymbol{y}_{d_t^*}$. Set $\boldsymbol{Y}_t = (\boldsymbol{Y}_{t-1}, \boldsymbol{y}_{d_t^*})$
  Update Gaussian approximation (iteratively solve for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$), given all data $\boldsymbol{Y}_t$. ▷ Equation (10) and (11)
end for

4. Simulation studies

We present two synthetic examples. In the first example we compare three different approximations of the EIBV on a small-size spatial domain. This will let us assess properties of the approximated values and see if the closed-form EIBV approximation works well. We further use this first example to compare adaptive sampling designs with pre-scripted designs. In the second example we use the closed-form EIBV approximation on a large-size spatial domain, and we compare it with alternative approaches for designs.

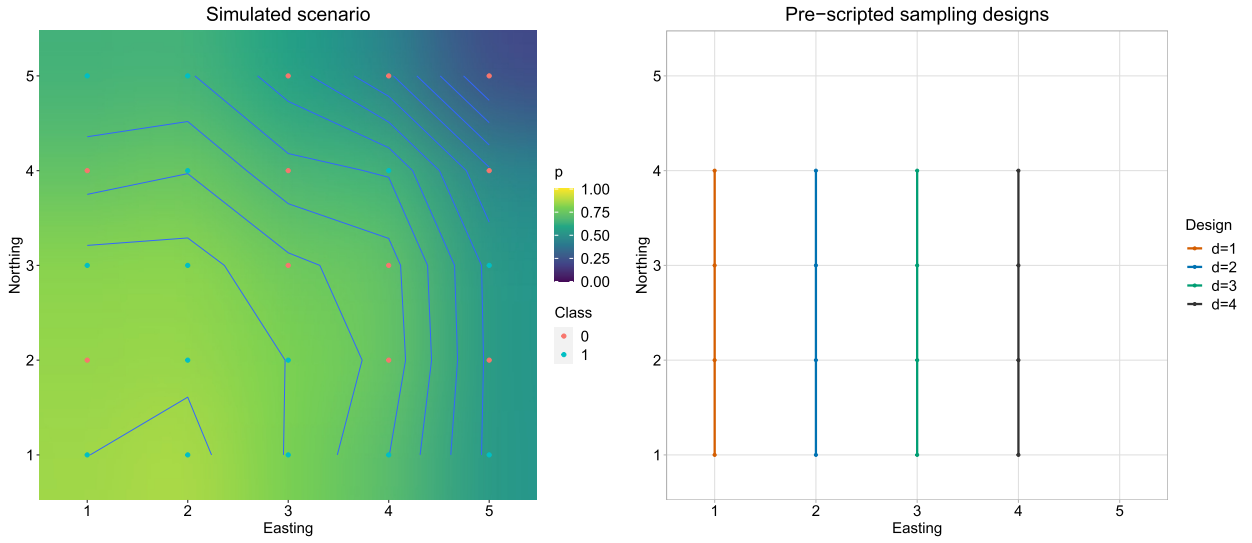


Fig. 3. (a) One realization of a simulated ground truth where class 1 (blue) is presence while class 0 (red) is absence. (b) Four batch sampling designs of size 4.

4.1. Comparison of approximations

We compare approximations EIBV I and EIBV II with a full fledged Monte Carlo approximation, considered the 'gold standard' for the remainder of this section. The objective of this simulation is to compare how close the approximations are to each other. Here, the Monte Carlo results will be exact when the sample size goes to infinity since it avoids the use of logistic and Gaussian approximations. But in practice, such time-consuming runs involving Markov chain Monte Carlo (MCMC) runs would not be feasible to deploy onboard a robotic vehicle.

In our modeling framework, this EIBV approximation is based on running MCMC algorithms for each of the multiple data outcomes of $\boldsymbol{y}_d$ and obtaining $P(\boldsymbol{y}_d)$ by Monte Carlo averaging over data. Hence, for the Monte Carlo implementation we generate MCMC realizations $\boldsymbol{\eta}^1, \ldots, \boldsymbol{\eta}^B$ from the posterior $p(\boldsymbol{\eta}|\boldsymbol{y}_d)$ and then approximate the EIBV according to

$$\text{EIBV}^{MCMC}(d) = \sum_{\boldsymbol{y}_d \in \{0,1\}^{|d|}} \text{IBV}(\boldsymbol{y}_d) P(\boldsymbol{y}_d), \qquad \text{IBV}(\boldsymbol{y}_d) = \sum_{s \in \mathcal{M}/d} \hat{p}_{\boldsymbol{s}}(1-\hat{p}_{\boldsymbol{s}}), \qquad \hat{p}_{\boldsymbol{s}} = \frac{1}{B}\sum_{b=1}^{B} p_{\boldsymbol{s}}^b, \qquad p_{\boldsymbol{s}}^b = \frac{\exp(\eta_{\boldsymbol{s}}^b)}{1+\exp(\eta_{\boldsymbol{s}}^b)}. \tag{17}$$

In this simulation study we consider a spatial domain $\mathcal{M}$ discretized to a grid of dimension 5×5 so $N = 25$. The design matrix has size-2 rows $\boldsymbol{x}_{\boldsymbol{s}}$ with a constant term and a covariate that increases quadratically away from a location (x^*, y^*) in the grid. This simulation study consists of 100 replicate runs, with each replicate being generated from the combination of nine inputs that varies in each run. The first inputs are the (x^*, y^*) coordinates, generated from discrete uniform distributions $U(1, 5)$. The other inputs are hyperparameters for $\boldsymbol{\beta}$ and $\boldsymbol{w}$, and they follow continuous uniform distributions: $\mu_{\beta_1} \sim U(-2, 2)$, $\mu_{\beta_2} \sim U(-5, 5)$, $\sigma_{\beta_1} \sim U(0.2, 0.8)$, $\sigma_{\beta_2} \sim U(0.5, 1.5)$, $\rho \sim U(-0.5, 0.5)$, $\zeta \sim U(0.1, 0.5)$ and $\phi \sim U(7, 11)$.

These hyperparameters for each replicate define a Gaussian prior distribution for the linear predictor $\boldsymbol{\eta}$, with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Then we have a total of 100 Gaussian prior distributions. Based on a prior distribution and the model for the absence-presence data, we simulate ground truths by i) drawing $\boldsymbol{\eta}$, ii) calculating $\boldsymbol{p}$ and iii) drawing $\boldsymbol{y}$.

Fig. 3(a) illustrates one simulated ground truth from a prior distribution, with a priori presence probabilities $\boldsymbol{p}$ (color coded) and one realized set of classes $\boldsymbol{y} \in \{0, 1\}$ for all the 25 locations (indicated by red or blue dots). Here, the probability of presence is highest near $(2, 1)$. For this prior scenario, the initial IBV is 4.88 (Eq. (1)). We study the EIBV approximations for four different designs ($d = 1, 2, 3, 4$), each of size $|d| = 4$. These can be seen in Fig. 3(b).

The comparison of the EIBV values is done by the three different proposed solutions:

- EIBV I from (13): We use the Gaussian approximation and the probit substitution for the logit for each combination of $\boldsymbol{y}_d$ of all the data $2^{|d|}$ to obtain the predictive probabilities and EBV at all grid locations. We sum these over the grid nodes to get the final value of the EIBV.
- EIBV II from (16): We directly use the closed form solution based on the transformed data $\boldsymbol{z}_{\boldsymbol{s}}$.

Table 1
Mean and standard error (in parentheses) of the different metrics evaluated for pre-scripted designs of $|d| = 4$. The computation is based on 100 replicates.

Comparison		Spearman	Ranking	Bias	RMSE
MCMC	EIBV I	0.75 (0.05)	0.82 (0.04)	0.06 (0.01)	0.11 (0.01)
MCMC	EIBV II	0.65 (0.06)	0.74 (0.04)	0.04 (0.01)	0.10 (0.01)
EIBV I	EIBV II	0.71 (0.06)	0.86 (0.03)	-0.02 (0.00)	0.03 (0.00)

- EIBV MCMC from (17): We use a Langevin MCMC algorithm where we sample $\boldsymbol{\eta}$ from the posterior $p(\boldsymbol{\eta}|\boldsymbol{y}_d)$. We tuned the Langevin sampler parameter to get an acceptance rate of about 0.57. This is run for $1,000,000$ iterations and we thinned the samples to have acceptable autocorrelation values.

We compute the three EIBV approximations using the priors defined by the 100 replicates, and for the four pre-scripted designs, i.e. each approximation is computed 400 times in total. We use dedicated cores of a computer cluster (Själander et al., 2019) to perform the task. The mean and standard error of the computer time in seconds per approximation is 0.37 (0.01) for EIBV I, 0.04 (0.01) for EIBV II and 1670.61 (6.93) for EIBV MCMC.

The MCMC approximation is considered as 'gold standard' when we compare to EIBV I and II. We report the comparison between the EIBV I and II too. To assess the performance of approximations we report the following metrics:

- Spearman: This is the Spearman rank correlation estimate that we use to compare the ranking of the four designs. Each estimate was computed using four EIBV values from each approximation at each replicate.
- Ranking: This is the proportion of runs where the two competing methods choose the same design as being 'best'.
- Bias: This is the value of the difference between the EIBV approximations.
- Root mean square error (RMSE): This is $\sqrt{\sum_{r=1}^{R} \mathrm{v}_r^2/R}$, where v_r is the difference between the values that are compared, for replicate $r = 1, \ldots, R$, $R = 100$.

In Table 1 we show pair comparison performance of the approximations.

We see evidence of similarity between the approximation results. The Spearman's rank correlations have high values, the matching in the rankings where the two approximations choose the same design as 'best' has high values too. Moreover the bias has acceptable values, considering the rather large variability in the replicate runs. The results of EIBV I and II are more similar than that of MCMC. The upwards bias in the analytical EIBV approximations is likely due to using $\Phi(\alpha x)$ instead of the logistic function.

We next use approximation EIBV II in Algorithm 1 to define different adaptive sampling designs of size $|d| = 5$. The adaptive design choices are made after every single observation, and they are limited to four options, turn south, north, east or west (except at the boundaries). At every stage t we hence evaluate the EIBV of the four directions, and the design d_t with minimum EIBV is chosen. Data from the replicate's simulated ground truth are used to update the mean and covariance of the Gaussian approximation after every stage. Each path is hence determined from the data at the selected design locations, which are specified according to this ground truth, and vary over the replicate runs.

In Fig. 4 (a) we see one realization of the ground truth. There is here more uncertainty around the contour of the center of the region. In 4 (b) we see two pre-scripted designs ($d = 5$ and $d = 6$) and one adaptive design. Design $a = 1$ has starting point (5,1). The adaptive design shows higher flexibility when it explores the region, choosing to move to the part with more uncertainty chose to the center.

The objective of this simulation is to compare the realized IBV after using the adaptive and the pre-scripted approaches. We generate sets of ground truths by sampling from the 100 Gaussian prior distributions. In each scenario we evaluate one pre-scripted design and one adaptive design. Each adaptive design has a starting point randomly chosen over the grid. We compare the following cases:

1. Adaptive designs with random initial point and pre-scripted design $d = 5$,
2. Adaptive designs with random initial point and pre-scripted design $d = 6$.

The comparison of realized IBV after collecting data with pre-scripted designs and adaptively generated ones is done through hypothesis testing on pair differences. The null hypothesis tests have mean of the IBV differences (diff) between adaptive and pre-scripted approaches is equal to zero, i.e. $\mu_{\text{diff}} = 0$ (see Table 2).

For both comparisons we obtain a low p-value, and it rejects the null hypothesis at a 1% significance level. The adaptive approach has significantly smaller realized IBV. We see the adaptive approach offers more flexibility to explore regions with more uncertainty.

4.2. Large scale example

We now consider a large-size example with 150×225 grid nodes. We set 13 possible sampling designs which are defined to be straight-line transects of 150 grid nodes length, as illustrated in Fig. 5 (a). This is mimicking a float gathering data

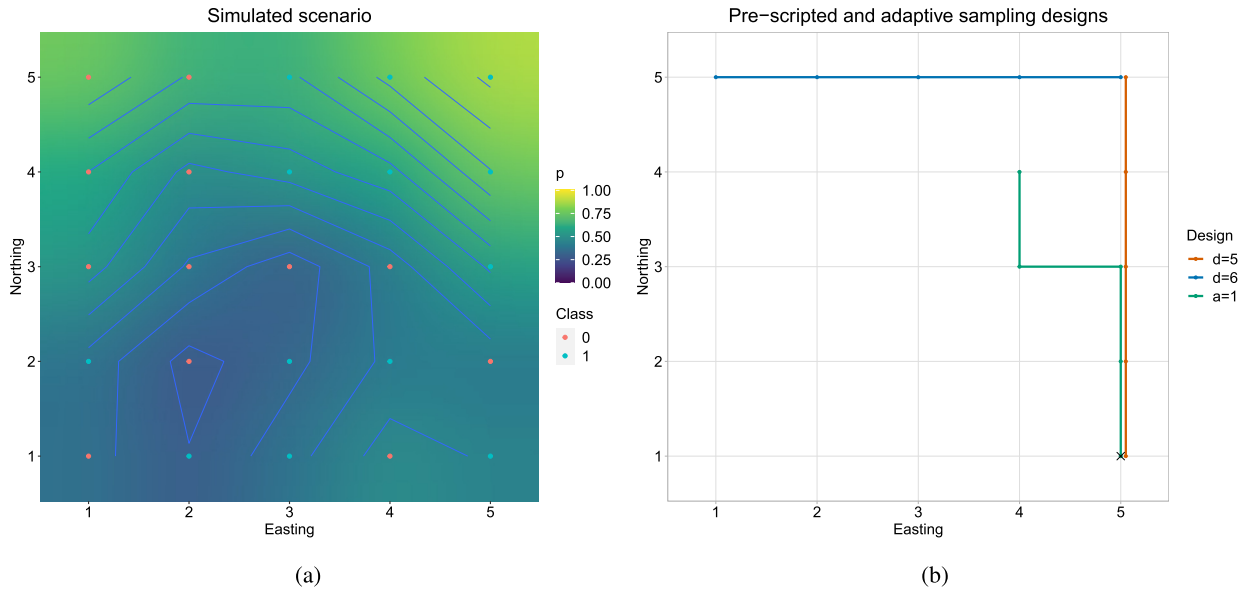


Fig. 4. (a) One realization of the simulated scenario where class 1 (blue) is presence while class 0 is absence. (b) Adaptive and pre-scripted sampling paths.

Table 2
Comparison of adaptive and pre-scripted approaches. For the IBV and the IBV difference we report the mean and standard error. We report the t-statistic and the p-value, 0.0· means significant in at least the 3rd decimal. Boldface fonts indicate significance at the 0.01% level. Each comparison is based on 100 replicates.

Comparison	IBV		Difference	t-statistic
	Adaptive	Pre-scripted		
1	3.20 (0.13)	3.49 (0.11)	-0.29 (0.05)	**-5.98 (0.0·)**
2	3.25 (0.13)	3.51 (0.11)	-0.26 (0.05)	**-5.62 (0.0·)**

along its south-north trajectory.

We study results of an adaptive strategy using EIBV to sample three of these 13 transects. The adaptive EIBV II approximation strategy is compared with other approaches inspired by what has been presented recently in the literature of habitat mapping. Motivated by Foster et al. (2020) we run a strategy that focuses on spatial balance in the designs. We show the results of using a design with the maximum minimal spatial distance between sampling transects. This means that lines 1, 7, 13 in Fig. 5 are used for this *Spatial Bal* design. Inspired by Shields et al. (2020) we run a strategy which focus on minimizing the uncertainty (variance). For the *Prediction Var* strategy, we compute the BV using the current model uncertainty only (without spatial integration, and not using the expectation over the next stage data outcomes). The sampling transect with largest average BV at the current stage is selected. This is an adaptive strategy, where the model is updated before selecting the next transect. The computing time is much larger for the EIBV criterion as it takes into account the entire field. On our laptop implementation, it is about 20 times slower per stage evaluation than the Prediction variance criterion which only looks at probabilities in the transect lines. The Spatial balance criterion is even faster than this as there is no adaptation. In practice this might not be relevant, as the EIBV computing time is still within seconds, and this is negligible compared with the time for annotating float data.

We generate 100 replicate datasets from a hierarchical logistic model. Here, the mean is $-2+4x_i$, where the x_i covariate is the normalized easting coordinate going from 0 (western boundary) to 1 (eastern boundary). This favors absence in the western part and presence in the eastern parts of the domain. We randomize over the other model parameters in each replicate. This is done uniformly around reference values: The spatial variance component ζ^2 is uniformly distributed within 1 ± 0.9. The intercept and slope variance parameters are similarly uniform in 1 ± 0.9, and their correlation is -0.5 ± 0.4. The spatial correlation range is uniform 1500 ± 500 m. Results of one of the synthetic replicate realizations of the linear predictor and true presence-absence data are shown in Fig. 5 (b-c).

The EIBV strategy tends to select a line in the middle and one line on either side. But they are rather spread out, more so than for the *Prediction Var* strategy, which is more focused on exploiting locations with high BV, rather than exploring the model uncertainty in the entire domain.

After the sampling has occurred for each of the replicate runs, we compute the realized IBV for the three sampling lines. We also study other performance metrics: the integrated mis-classification probabilities (MCP) which we define by $\sum_{i=1}^{N} P(y_{\boldsymbol{s}_i} \neq y_{\boldsymbol{s}_i}^{\text{true}} | \boldsymbol{Y}_t)$, and the negative log score (NLS) of the observations: $-\sum_{i=1}^{N} \log P(y_{\boldsymbol{s}_i} = y_{\boldsymbol{s}_i}^{\text{true}} | \boldsymbol{Y}_t)$. Lower values in

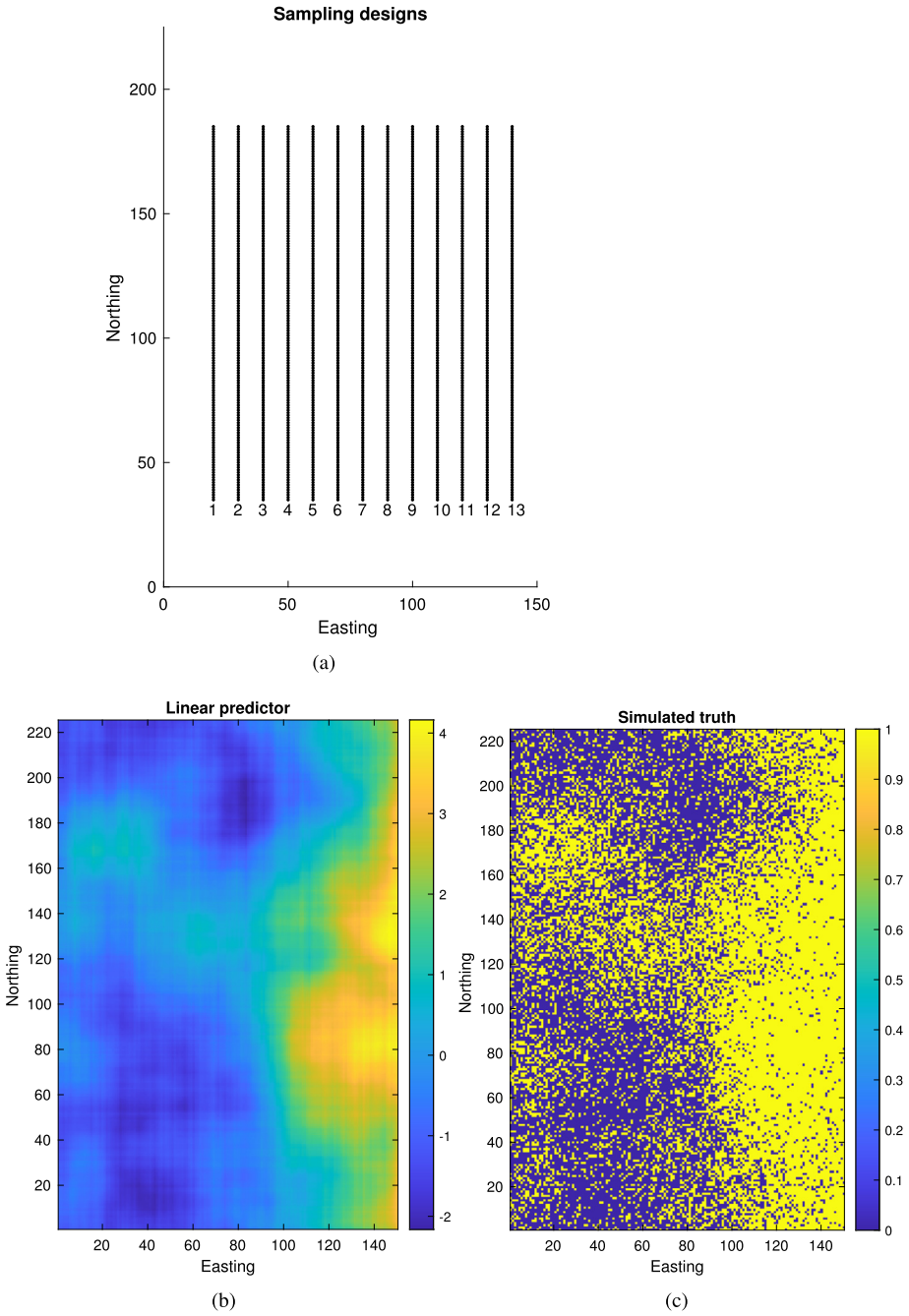


Fig. 5. (a) The synthetic example has 13 potential sampling design transects. (b) A realization from the linear predictor containing the regression trend and the Gaussian random field model. (c) Realization of presence-absence data from the hierarchical logistic model.

these metrics are preferred. The approach in Algorithm 1 emphasizes low EIBV, but one is hoping to achieve reasonably good performance on the other metrics too. In the comparison, we use differences in metrics for each replicate. We compute a t statistic from the pair differences. The associated p-value of a two-sided pair t-test allows us to study if there are significant improvements by using one sampling strategy rather than the other. There are three comparisons: EIBV versus *Prediction Var*, EIBV versus *Spatial Bal*, and *Prediction Var* versus *Spatial Bal*. Each one is evaluated for the realized IBV, integrated MCP and NLS.

In Table 3 (left columns) we summarize the results of this study. In this example, most metrics are clearly lower for the EIBV strategy than for the Prediction variance and Spatial balance designs. The Spatial balance sampling appears to be better than the Prediction variance strategy, possibly because of the relatively large model uncertainty which makes it important

Table 3
T-statistic from differences in replicates, with p-values of the two-sided test in parentheses. Results are for realized IBV, integrated Mis-classification probability (MCP) and negative log score (NLS). The computation is based on 100 replicate studies. *EIBV* means three sequentially selected transects using the adaptive EIBV criterion, *Spatial Bal* means three pre-selected transects with maximum-minimal distancing, *Prediction Var* means three adaptively selected designs based on minimizing the transect line Bernoulli variance, given the current data. The bottom row is comparing the EIBV criterion for the correct model versus the mis-specified one. Here, 0.0· means significant in at least the 3rd decimal. Boldface fonts indicate significance at the 0.05% level.

	Correct model			Mis-specified model		
	IBV	MCP	NLS	IBV	MCP	NLS
EIBV-Pred Var	**-2.89 (0.0·)**	**-4.21 (0.0·)**	**-4.58 (0.0·)**	-1.89 (0.06)	**-3.24 (0.0·)**	**-3.80 (0.0·)**
EIBV-Spat Bal	**-2.93 (0.0·)**	**-4.11 (0.0·)**	**-2.74 (0.0·)**	**-3.44 (0.0·)**	**-4.51 (0.0·)**	-1.63 (0.10)
Pred Var-Spat Bal	0.51 (0.61)	0.95 (0.34)	**2.99 (0.0·)**	-1.16 (0.25)	-0.99 (0.32)	**2.60 (0.01)**
correct - mis-specified	–	–	–	-0.31 (0.72)	-1.46 (0.15)	**-3.52 (0.0·)**

to span the space in the design. The Prediction variance strategy focuses on the transects with probabilities closest to 0.5, and this appears to limit exploration, at least in some of the replicates.

In the 100 replicate runs we used the same model in the data generation and the adaptive sampling and model updating. We now check the sensitivity to model mis-specification. In Table 3 (right columns) we show results of doing the analysis on a model with mis-specified parameters. In particular the data are generated as before, but in the model for the sampling line selection we use the fixed reference values for the model parameters (stated before the $\pm$ in the parameter specification). In the results we notice similar tendencies as in the correctly specified model, where the EIBV strategy seems to be better. There differences between the Spatial balance and Prediction variance strategies are now more unclear. When comparing the correctly specified model with the mis-specified model (Table 3, bottom row), and both are using the EIBV approach for transect design selection, we see that a mis-specified model is missing some in the adaptation. Overall, the NLS metric seems to penalize the most (larger T-statistics).

5. Application: Lizard Island

We analyze a real-world dataset from Trimodal, a section of coral reef on Lizard Island in Australia. For this coral reef, one knows the ground truth of presence-absence variables via extensive benthic imaging and annotation efforts. It is hence useful to test models and methods.

5.1. Pre-processing and data collection

The domain of interest is indicated in Fig. 6(a) and in Fig. 6(b). In this display the color code is representative of the reef depth. In the map-view dark blue colors are not part of the reef. Since this reef has been studied so much, there are not only excellent digital elevation maps (DEM) of the reef, but also ground truth coral presence-absence data at dense 1×1 m^2 resolution. We match various data files in a pre-processing step, so that we can align the DEM data with coral information files. For our purposes, we allocate covariates $\boldsymbol{x}_{\boldsymbol{s}}$ and presence-absence variables $y_{\boldsymbol{s}}$ to a 158×107 grid of 1×1 m^2 resolution. This includes some edge zone around the reef (Fig. 6(b)). From the DEM files, we use the `R` package `raster` to compute the slope and aspect measures (Horn, 1981), and we consider depth, slope and aspect as covariates, so $\boldsymbol{x}_{\boldsymbol{s}} = (1, x_{1,\boldsymbol{s}}, x_{2,\boldsymbol{s}}, x_{3,\boldsymbol{s}})$. Coral presence ($y_{\boldsymbol{s}} = 1$) is indicated by black circles in Fig. 6(b).

With improved sensor qualities and onboard facilities, underwater vehicles like floats drifting along the seabed or AUVs equipped with cameras can collect benthic information. These robotic vehicles have various speed and operation limitations, so they must be guided properly. They also have different flexibility in the way that autonomy can be implemented onboard, and this calls for efficient statistical sampling design strategies that can be operated in a semi-autonomous setting. Here, we consider sequential deployments of a float, giving batch data along sampling lines as it drifts with the current. A benthic imaging float is a low-cost underwater vehicle that can be used efficiently in such operations. When the float ends the dive, the data is loaded on a computer, and the statistical model is updated with the acquired presence-absence data from the sampling line. The refined model is then used to plan the next sampling line for the float, see Fig. 1(b). We play the game of collecting data from such a float, in a retrospective setting meaning that the ground truth data have already been gathered. In actual operations, the habitat classes are extracted in an auto-encoder from image scans (Shields et al., 2020), but we simply use the processed category data along the sampling lines in this study. The potential sampling designs are straight lines going from east-south-east to west-north-west, assuming a fixed sea current in this direction.

5.2. Statistical parameter specification

To define a prior distribution for the regression component and the spatial structured variables we partition the geographical domain in two subregions. The first is a small subset of the data along the reef ridge to estimate the parameters for the prior distribution. The second subregion is used in the subsequent sampling design evaluation. From the subset of data, we first specify the prior mean and the variance of $\boldsymbol{\beta}$ by using a standard Fisher-scoring algorithm for a non-spatial

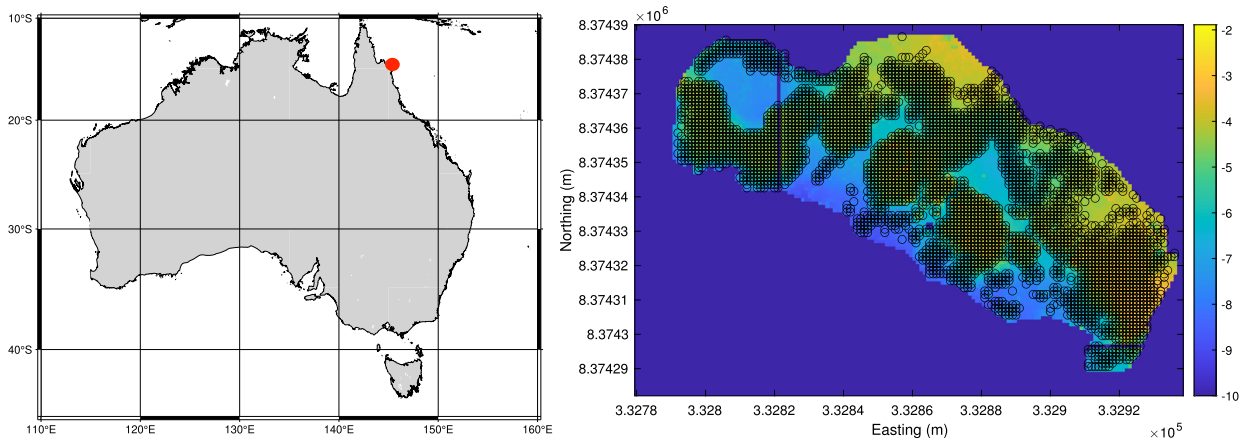


Fig. 6. (a) Australia and the location of Lizard Island. (b) Lizard Island's Trimodal site depth map and circles for coral locations. (Outside reef locations have simply been set to 10 m depth here.)

GLM. For this case, the most important covariate is depth, while aspect ratio and slope are less significant in the regression. Based on this we set

$$\boldsymbol{\mu}_\beta = \begin{bmatrix} 3 \\ 0.6 \\ 0.03 \\ 0.002 \end{bmatrix} \text{ and } \boldsymbol{\Sigma}_\beta = \begin{bmatrix} 0.2 & 0.03 & -0.001 & -0.0002 \\ 0.03 & 0.006 & 0.0001 & 0 \\ -0.001 & 0.0001 & 0.0001 & 0 \\ -0.0002 & 0 & 0 & 0.0001 \end{bmatrix}.$$

To specify the covariance model of the spatial effect $\boldsymbol{w}$ we use maximum marginal likelihood estimation for the standard deviation ζ and correlation decay parameter ϕ:

$$(\hat{\zeta}, \hat{\phi}) = \text{argmax}_{\zeta,\phi} \left\{ \int p(\boldsymbol{y}|\boldsymbol{\eta}) p_{\zeta,\phi}(\boldsymbol{\eta}) d\boldsymbol{\eta} \right\} \tag{18}$$

Here, we use the first-order Laplace approximation, see e.g. Evangelou et al. (2011), to evaluate the integrand at every iteration of a numerical optimization scheme. The specified parameter values are $\hat{\zeta} = 2.2$ and $\hat{\phi} = 0.25$, indicating an effective correlation range of about 15 grid cells in the discretization of the spatial domain.

5.3. Analysis and results

We define 14 possible drop locations for the float. These lines are set with random starting points, while ensuring that the entire transect stays within the domain of interest. Each of these drops has an operation time of about 30 seconds. Assuming the float drifts with the west-north-west current at 1 m/s it then makes a sampling line covering 29 grid cells in a drop. At each step in the sequential strategy (Algorithm 1), we evaluate the EIBV criterion for all the remaining drop locations, and the design with smallest EIBV is selected at each stage. The ground truth data along the sampling line is then used to update the statistical model. This strategy continues until all 14 lines have been sampled.

Fig. 7 shows the coral presence probability results at the initial time (left) and after 14 float operations (right).

The sampling lines are recognized by dark colors. Compared with the initial probabilities of coral presence, there is much change in the predictions at unsampled locations towards highly confident values close to 0 and 1. This occurs because we have increased our knowledge on both the regression parameters and the spatial random effects which, combined, enable better interpolations of the linear predictor.

Based on the prior model, the IBV is 3067. At the first stage, the sampling line starting at grid cell $(136, 51)$ is selected. This has an EIBV of 2810. The biggest EIBV at the first stage is for the sampling line starting at $(95, 35)$. During the sequential deployment, this line is not selected last, but as number 5 in the stage order. Thus, it is clear that the relative ranking of the 14 drop locations changes as new presence-absence data becomes available.

Fig. 8 (right, black circles) shows the results of the adaptive EIBV strategy over the 14 stages. At most stages it is decreasing substantially, though maybe less at the end. We again compare the performance of this strategy with that of a pre-scripted strategy of transects (red squares) and the adaptive prediction variance strategy picking the transect which has the largest average BV (blue stars) at each stage. The pre-scripted transects are covering the domain step by step and gradually filling in the spatial domain, but there is no adaptation in the selection here. The comparison is done with respect to the realized IBV (left), integrated mis-classification probabilities (middle) and the negative log score (right). For

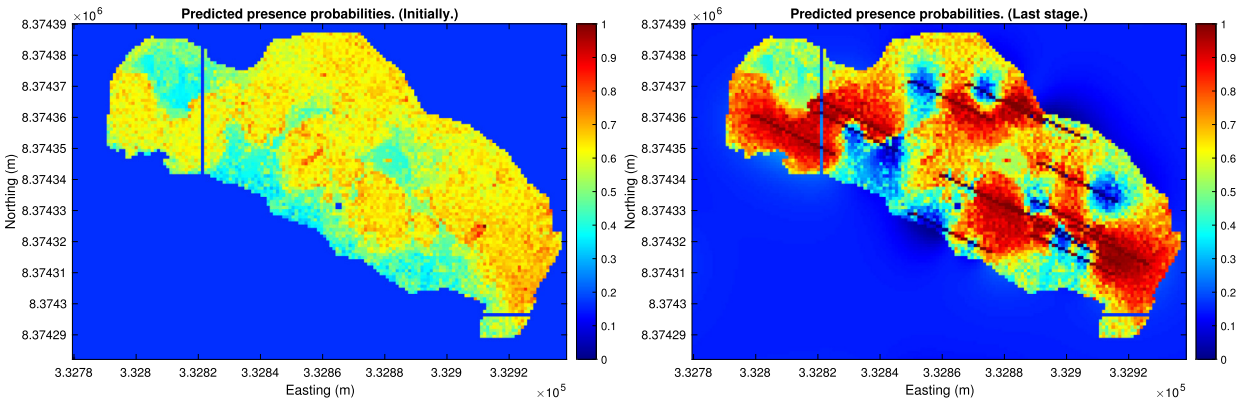


Fig. 7. Lizard Island case. (a) Initial prediction. (b) Stage 14 prediction (right).

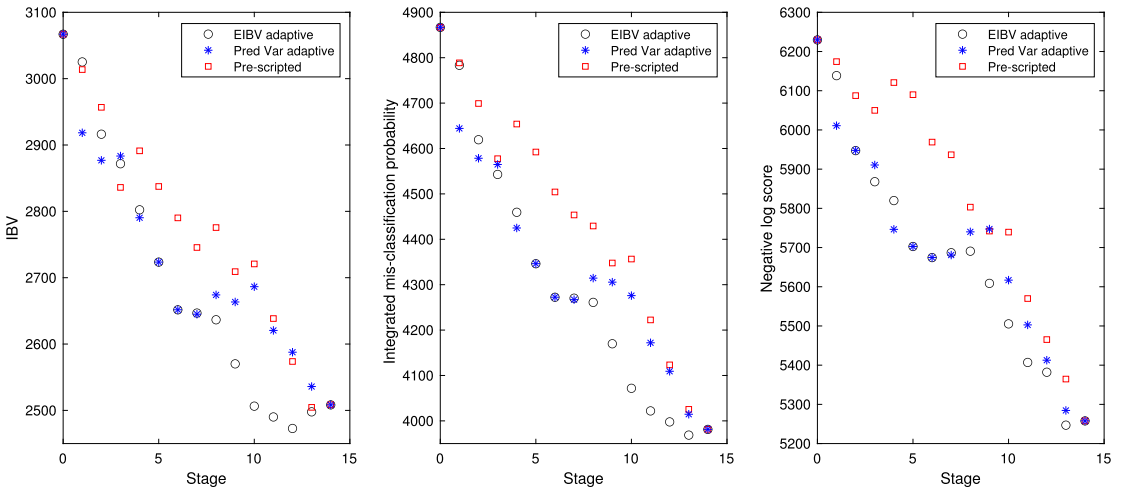


Fig. 8. Realized IBV, integrated mis-classification probability and negative log score metrics for the Lizard Island case. There are $k = 1, \ldots, 14$ realized floats (first axis) in the sequential strategy. Adaptive EIBV Strategy (black circle), adaptive largest BV transect selection (blue star) and pre-scripted spatially balanced design (red square) are compared.

all strategies the metrics tend to go down over stages, with some variability due to the coral presence-absence observations that are included in the sequential statistical model description. Here, the regression parameters are rather well known, and the adaptive EIBV strategy and the Prediction variance strategy are performing better than the Spatially balanced design. The EIBV-based strategy (black circles) appears to have steadier decline over stages, while the Prediction variance strategy seems to have a more fluctuating decline curve. Even though the pre-scripted strategy performs a bit worse in this example, we noticed that carefully pre-planned transects that consider uncertainty regions in addition to some spatial coverage perform on-par with the adaptive strategies, but such designs are not always easy to plan wisely in advance.

The EIBV strategy takes longer to compute than just computing the average BV for each transect, as is done in the Prediction variance strategy. Every EIBV calculation in this example takes about 1 sec on a laptop computer. This is not much time compared with that required for encoding float data. The efficient closed-form EIBV calculations outlined here enable a new way of studying the performance of different designs.

6. Closing remarks

We have developed new approximate closed form solutions for the expected integrated Bernoulli variance for spatial generalized linear models. The results demonstrate that one can very quickly compare several experimental designs based on this criterion, and build on this to choose informative sampling schemes for improving presence-absence maps. The closed form solutions are based on Gaussian approximations and functional approximations for binary regression models, and the end results involve bivariate Gaussian cumulative distribution functions calculations. The approximations are validated in simulation studies with respect to gold-standard, but computationally-expensive, Monte Carlo methods.

We implement and show results of an adaptive sampling design algorithm for expected integrated Bernoulli variance reduction. This is based on a myopic assumption, where only one stage of optimization is done and then the spatial Gaussian model approximation is updated at every stage. We compare our approach with other strategies for sampling (pre-scripted and adaptive). In a synthetic example with much prior uncertainty, our approach is better than the other two, and a spatially covering design is better than the more aggressive approach of choosing designs with locally high Bernoulli variance. In an application with corals in Australia, there is a lot of prior knowledge on the covariate effects, and the design choosing design candidates with high Bernoulli variance is always on par with the suggested strategy, while a pre-scripted spatial coverage design performs worse here. The Australian case study is motivated by applications in benthic habitat mapping, where robotic vehicles are currently used for mapping the presence-absence of environmental features, including corals. This can be used with various degrees of autonomy in the underwater vehicle. We demonstrate this for the operation of an imaging float, applied to a well-characterized coral reef site in Australia. Our approach shows the importance of transferring knowledge from the statistics field to solve habitat mapping problems. Once we have knowledge of the probabilistic relations between bathymetric data and the distribution of types of benthic cover, we can specify better regression and spatial effect priors. These priors can then be used to improve the mapping of another sampling region and help to solve the 'cold start' problem by collecting these initial in-situ datasets. Furthermore, the statistical criterion and adaptive sampling approach presented in this paper can easily be used in the mapping of other environment domains in need of scalable sampling strategies.

Rather than having a greedy strategy that looks only at the current and next step, there might be some to gain by implementing more complex strategies like rapidly exploring random trees (RRT), see e.g. Shields et al. (2021) or adapted Monte Carlo Tree Search (MCTS) that provide longer horizon informative paths (Arora et al., 2019). A major advantage of the computational speed gain of the presented methodology is that it enables comparison of a large number of designs in applied problems without the need for sub-sampling. The speed gain makes non-myopic strategies more feasible to implement too, since evaluating outcomes at possible future states could be done quickly. This could open the development of research in non-myopic sampling design strategies.

Further work includes having imperfect observations of the habitat, as we assume the sensor measures the ground truth category here. Assuming the observations are not necessarily perfect at the locations where they are made, it becomes relevant to include different kinds of sensor data. In this context one can also study which data type is more informative, not only where to conduct the observations. It would also be interesting to include more habitat classes. Our current Bernoulli variance criterion and the developed closed form solutions are based on logistic regression models for two classes. More work is required to generalize both to a model with multiple classes. Exploring the effects of non-myopic (non-greedy) strategies is of interest too. Then a comparison between different location-search strategies can be done, considering the energy, time and computational resources and the accuracy of the mapping.

We only looked at spatial variability in the current paper. In many biological settings it is of course also important to monitor changes over time. One can then design sampling strategies over space and time, and it would be natural to get higher sampling intensity at spatial regions which are expected to change faster in time. In an operational setting, it would be very interesting to couple the statistical results with operations involving multiple vehicles relying on different levels of autonomy and adaptation. These are difficult optimization problems involving operational constraints, but statistical developments must be a part of this development to reliably assimilate different kinds of data and to help guide the sampling efforts of various sensors and vehicles.

Data availability

The source codes are available at the link: https://github.com/anyosa/adaptive_spatial_designs.

Acknowledgements

SA and JE thank the Norwegian Research Council for funding the MASCOT project (grant no 305445). The Lizard Island data was gathered, in part, with support from Australian Research Council (grants DP1093448 and FT110100609) and the University of Sydney. A special thanks to Jackson Shields for sharing his habitat labels for the Trimodal site.

Appendix A. Scalable calculations in the Fourier domain

For fast calculations in our context with large grids and sparse designs, we embed the matrix $\boldsymbol{\Sigma}_w$ on a torus structure to get at circulant matrix $\boldsymbol{C}$. We can then use established Fourier domain tricks (Gray, 2006; Paciorek, 2007) for large-scale computations. Eidsvik et al. (2009) use similar Fourier calculations for GLMs.

The zero-mean spatially structured latent variable $\boldsymbol{w} = (w_1, \ldots, w_N)'$ is represented on a regular $n_1 \times n_2$ grid, $N = n_1 n_2$, and we refer to the $n_1 \times n_2$ matrix $\boldsymbol{w}^m = (w^m_{0,0}, w^m_{0,1}, \ldots, w^m_{n_1-1,n_2-1})$ as the matrix associate of length-N vector $\boldsymbol{w}$. The matrix $\boldsymbol{C}$ is defined by the covariance between $w^m_{0,0}$ and all other variables as they are positioned on the gridded torus. It is a block circulant covariance matrix, and we arrange the N relevant covariance entries in an $n_1 \times n_2$ matrix which we denote $\boldsymbol{C}^m$.

In what follows, dft2 refers to the two dimensional discrete Fourier transform, i.e.

$$\text{dft2}(C^m)_{j_1',j_2'} = \sum_{j_1=0}^{n_1-1}\sum_{j_2=0}^{n_2-1} C^m_{j_1',j_2'} \exp[-2\pi\iota(\frac{j_1 j_1'}{n_1} + \frac{j_2 j_2'}{n_2})], \tag{A.1}$$

for $j_1' = 1,\ldots,n_1$, $j_2' = 1,\ldots,n_2$, with $\iota = \sqrt{-1}$. We let idft2($\boldsymbol{d}^m$) denote the two dimensional inverse discrete Fourier transform of $n_1 \times n_2$ matrix $\boldsymbol{d}^m$.

Multiplication of matrix $\boldsymbol{C}$ with length N vector $\boldsymbol{w}$ is done efficiently in the Fourier domain. The $n_1 \times n_2$ matrix associate of vector $\boldsymbol{b} = \boldsymbol{C}\boldsymbol{w}$ is

$$\boldsymbol{b}^m = \text{Re}\{\text{dft2}[\text{dft2}(\boldsymbol{C}^m) \odot \text{idft2}(\boldsymbol{w}^m)]\}, \tag{A.2}$$

where $\odot$ refers to elementwise multiplication.

The Fourier calculation is used in two places in the suggested algorithm. First, the conditional mean calculation that is updated (iteratively) at every stage is using the Fourier calculations to compute $\boldsymbol{\Sigma}_{\cdot,d}[\boldsymbol{\Sigma}_d + \boldsymbol{K}_d]^{-1}(\boldsymbol{y}_d - \boldsymbol{\mu}_d)$. We have that $\boldsymbol{\Sigma}_{\cdot,d} = [\boldsymbol{\Sigma}_w \boldsymbol{A}^T + \boldsymbol{X}\boldsymbol{\Sigma}_\beta \boldsymbol{X}_d]$, where $\boldsymbol{A}$ is a selection matrix for the design locations in the grid and where $\boldsymbol{X}_d$ consists of the covariates at design points. Defining length $|d|$ vector $\boldsymbol{a} = [\boldsymbol{\Sigma}_d + \boldsymbol{K}_d]^{-1}(\boldsymbol{y}_d - \boldsymbol{\mu}_d)$, we then have

$$\boldsymbol{\Sigma}_{\cdot,d}\boldsymbol{a} = \boldsymbol{\Sigma}_w \boldsymbol{A}^T \boldsymbol{a} + \boldsymbol{X}\boldsymbol{\Sigma}_\beta \boldsymbol{X}_d \boldsymbol{a}. \tag{A.3}$$

The first part is solved in the Fourier domain with $\boldsymbol{C}$ for $\boldsymbol{\Sigma}_w$. The second part is only of size $|d| << N$ which is fast to solve explicitly. With iterations to reach the mode of the posterior approximation, the part with $\boldsymbol{K}_d$ is linearized at every step. There are then several solves like this at every stage with updating of the Gaussian model for the linear predictor.

Second, the calculation of $\xi_{\boldsymbol{s}}^2$ in (14) also proceeds by evaluating Fourier domain calculations for the spatially structured covariance part. Using the same form of $\boldsymbol{\Sigma}_{\cdot,d}$ as above, the variance part in (14) equals

$$[\boldsymbol{\Sigma}_w \boldsymbol{A}^T + \boldsymbol{X}\boldsymbol{\Sigma}_\beta \boldsymbol{X}_d][\boldsymbol{\Sigma}_d + \boldsymbol{K}_d]^{-1}[\boldsymbol{\Sigma}_w \boldsymbol{A}^T + \boldsymbol{X}\boldsymbol{\Sigma}_\beta \boldsymbol{X}_d]^T = [\boldsymbol{\Sigma}_w \boldsymbol{L}_u + \boldsymbol{X}\boldsymbol{\Sigma}_\beta \boldsymbol{X}_d \boldsymbol{L}][\boldsymbol{\Sigma}_w \boldsymbol{L}_u + \boldsymbol{X}\boldsymbol{\Sigma}_\beta \boldsymbol{X}_d \boldsymbol{L}]^T, \tag{A.4}$$

where $\boldsymbol{L}\boldsymbol{L}^T = [\boldsymbol{\Sigma}_d + \boldsymbol{K}_d]^{-1}$ and $\boldsymbol{L}_u = \boldsymbol{A}^T\boldsymbol{L}$. The $\xi_{\boldsymbol{s}}^2$ terms are defined via the diagonal of this matrix, and they are calculated line by line from $\boldsymbol{\Sigma}_w \boldsymbol{L}_u(:,\boldsymbol{s}) + \boldsymbol{X}\boldsymbol{\Sigma}_\beta \boldsymbol{X}_d \boldsymbol{L}(:,\boldsymbol{s})$, where the $\boldsymbol{L}(:,\boldsymbol{s})$ notation means column s of the matrix $\boldsymbol{L}$. We then multiply this vector with itself and sum the components. The Fourier domain trick with $\boldsymbol{C}$ for $\boldsymbol{\Sigma}_w$ is used to speed up the first term in equation (A.4). The latter term is only of size $|d| << N$, so again explicit lower dimensional matrix expressions are fast for this part.

Appendix B. Sequential calculations

At each stage $t = 1,\ldots,$ the data design size is $|d_t| << N$. It is then efficient to factorize matrices of the data dimension in the updating, rather than storing and working with the large-size $N \times N$ covariance matrix of the latent state. We next describe efficient computations for this situation.

The updated Gaussian approximation at stage $t-1$ is based on solving expression (10) iteratively, given the current data $\boldsymbol{Y}_{t-1}$ from design set $\boldsymbol{D}_{t-1}$. With initial linearization point $\boldsymbol{\eta}^{*,0}$, this is modified according to

$$\boldsymbol{\eta}^{*,i+1} = \boldsymbol{\mu} + \boldsymbol{\Sigma}_{\cdot,\boldsymbol{D}_{t-1}}\left[\boldsymbol{\Sigma}_{\boldsymbol{D}_{t-1}} + \boldsymbol{K}_{\boldsymbol{D}_{t-1}}(\boldsymbol{\eta}^{*,i})\right]^{-1}\left(\boldsymbol{z}_{\boldsymbol{D}_{t-1}}(\boldsymbol{Y}_{t-1},\boldsymbol{\eta}^{*,i}) - \boldsymbol{\mu}_{\boldsymbol{D}_{t-1}}\right), \quad i = 0,1,\ldots, \tag{B.1}$$

until convergence to the conditional mode denoted $\hat{\boldsymbol{\mu}}_{t-1}$. Here, $\boldsymbol{z}_{\boldsymbol{D}_{t-1}}(\boldsymbol{Y}_{t-1},\boldsymbol{\eta}^{*,i})$ is the transformed data at this stage. The initial linearization point can be selected from the previous stage; $\boldsymbol{\eta}^{*,0} = \hat{\boldsymbol{\mu}}_{t-2}$. Solving equation (B.1) involves a mix of Fourier domain evaluations and explicit matrix-vector products as described in Appendix A. The updated approximation of the covariance is

$$\hat{\boldsymbol{\Sigma}}_{t-1} = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{\cdot,\boldsymbol{D}_{t-1}}\left[\boldsymbol{\Sigma}_{\boldsymbol{D}_{t-1}} + \boldsymbol{K}_{\boldsymbol{D}_{t-1}}\right]^{-1}\boldsymbol{\Sigma}_{\cdot,\boldsymbol{D}_{t-1}}^T, \tag{B.2}$$

where $\boldsymbol{K}_{\boldsymbol{D}_{t-1}}$ is evaluated at $\hat{\boldsymbol{\mu}}_{t-1}$.

For the calculations required for the EIBV design criterion at stage t, the expectation is done with respect to $\boldsymbol{y}_{d_t}$ in this updated model, given the realized data $\boldsymbol{Y}_{t-1}$ so far. We use

$$\tilde{\boldsymbol{\Sigma}}_t = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{\cdot,\boldsymbol{D}_t}\left[\boldsymbol{\Sigma}_{\boldsymbol{D}_t} + \boldsymbol{K}_{\boldsymbol{D}_t}\right]^{-1}\boldsymbol{\Sigma}_{\cdot,\boldsymbol{D}_t}^T, \qquad \xi_{\boldsymbol{s}}^2 = \tilde{\boldsymbol{\Sigma}}_t(\boldsymbol{s},\boldsymbol{s}) - \hat{\boldsymbol{\Sigma}}_{t-1}(\boldsymbol{s},\boldsymbol{s}), \qquad \boldsymbol{s} \in \mathcal{M}, \tag{B.3}$$

where $\boldsymbol{\Sigma}(\boldsymbol{s},\boldsymbol{s})$ are the diagonal elements of matrix $\boldsymbol{\Sigma}$ and $\boldsymbol{K}_{\boldsymbol{D}_t}$ is evaluated at the mode $\hat{\boldsymbol{\mu}}_{t-1}$ from the last stage. There are no additional iterations used for the design calculations because there are no new data yet. Just like in equation (A.4), Fourier domain tricks can be used to compute $\tilde{\boldsymbol{\Sigma}}_t$. Further, let $\boldsymbol{R} = \left[\boldsymbol{\Sigma}_{\boldsymbol{D}_t} + \boldsymbol{K}_{\boldsymbol{D}_t}\right]$ denote the $(|\boldsymbol{D}_{t-1}| + |d_t|) \times (|\boldsymbol{D}_{t-1}| + |d_t|)$ matrix that must be factorized in equation (B.3). This holds the covariances of all previously gathered data and the current (prospective) data at time t. We split this matrix in blocks, and its inverse $\boldsymbol{Q} = \boldsymbol{R}^{-1}$ is then defined via:

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{R}_{1,1}^{-1} + \boldsymbol{R}_{1,1}^{-1}\boldsymbol{R}_{1,2}\boldsymbol{Q}_{2,2}\boldsymbol{R}_{2,1}\boldsymbol{R}_{1,1}^{-1} & -\boldsymbol{R}_{1,1}^{-1}\boldsymbol{R}_{1,2}\boldsymbol{Q}_{2,2} \\ -\boldsymbol{Q}_{2,2}\boldsymbol{R}_{2,1}\boldsymbol{R}_{1,1}^{-1} & \boldsymbol{Q}_{2,2} \end{bmatrix}, \qquad \boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_{1,1} & \boldsymbol{R}_{1,2} \\ \boldsymbol{R}_{2,1} & \boldsymbol{R}_{2,2} \end{bmatrix}, \tag{B.4}$$
$$\boldsymbol{Q}_{2,2} = [\boldsymbol{R}_{2,2} - \boldsymbol{R}_{2,1}\boldsymbol{R}_{1,1}^{-1}\boldsymbol{R}_{1,2}]^{-1}.$$

Here, the inverse $\boldsymbol{R}_{1,1}^{-1}$ is available from the previous calculations at stage $t-1$, which means that the new matrix calculations are of rather small size $|d_t|$, facilitating effective design evaluations at each stage.

References

Abbas, S., Nichol, J.E., Wong, M.S., 2020. Object-based, multi-sensor habitat mapping of successional age classes for effective management of a 70-year secondary forest succession. Land Use Policy 99, 103360.

Adamo, M., Tarantino, C., Tomaselli, V., Veronico, G., Nagendra, H., Blonda, P., 2016. Habitat mapping of coastal wetlands using expert knowledge and earth observation data. J. Appl. Ecol. 53, 1521–1532.

Arora, A., Furlong, P.M., Fitch, R., Sukkarieh, S., Fong, T., 2019. Multi-modal active perception for information gathering in science missions. Auton. Robots 43, 1827–1853.

Bect, J., Bachoc, F., Ginsbourger, D., 2019. A supermartingale approach to gaussian process based sequential design of experiments. Bernoulli 25, 2883–2919.

Bender, A., Williams, S.B., Pizarro, O., 2013. Autonomous exploration of large-scale benthic environments. In: 2013 IEEE International Conference on Robotics and Automation. IEEE, pp. 390–396.

Brown, C.J., Smith, S.J., Lawton, P., Anderson, J.T., 2011. Benthic habitat mapping: a review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. Estuar. Coast. Shelf Sci. 92, 502–520.

Demidenko, E., 2013. Mixed Models: Theory and Applications with R. John Wiley & Sons.

Eidsvik, J., Martino, S., Rue, H., 2009. Approximate bayesian inference in spatial generalized linear mixed models. Scand. J. Stat. 36, 1–22.

Evangelou, E., Eidsvik, J., 2017. The value of information for correlated GLMs. J. Stat. Plan. Inference 180, 30–48.

Evangelou, E., Zhu, Z., Smith, R.L., 2011. Estimation and prediction for spatial generalized linear mixed models using high order Laplace approximation. J. Stat. Plan. Inference 141, 3564–3577.

Fossum, T.O., Travelletti, C., Eidsvik, J., Ginsbourger, D., Rajan, K., 2021. Learning excursion sets of vector-valued gaussian random fields for autonomous ocean sampling. Ann. Appl. Stat. 15, 597–618.

Foster, S.D., Hosack, G.R., Monk, J., Lawrence, E., Barrett, N.S., Williams, A., Przeslawski, R., 2020. Spatially balanced designs for transect-based surveys. Methods Ecol. Evol. 11, 95–105.

Goodson, J.C., Thomas, B.W., Ohlmann, J.W., 2017. A rollout algorithm framework for heuristic solutions to finite-horizon stochastic dynamic programs. Eur. J. Oper. Res. 258, 216–229.

Gray, R.M., 2006. Toeplitz and Circulant Matrices: A Review. Now Publishers Inc.

Horn, B.K.P., 1981. Hill shading and the reflectance map. Proc. IEEE 69, 14–47.

Overstall, A.M., Woods, D.C., 2017. Bayesian design of experiments using approximate coordinate exchange. Technometrics 59, 458–470.

Paciorek, C.J., 2007. Bayesian smoothing with gaussian processes using Fourier basis functions in the spectralgp package. J. Stat. Softw. 19. nihpa22751.

Ryan, E.G., Drovandi, C.C., McGree, J.M., Pettitt, A.N., 2016. A review of modern computational algorithms for bayesian optimal design. Int. Stat. Rev. 84, 128–154.

Schein, A.I., Ungar, L.H., 2007. Active learning for logistic regression: an evaluation. Mach. Learn. 68, 235–265.

Schroeder, S.B., Dupont, C., Boyer, L., Juanes, F., Costa, M., 2019. Passive remote sensing technology for mapping bull kelp (nereocystis luetkeana): a review of techniques and regional case study. Glob. Ecol. Conserv. 19, e00683.

Shields, J., Pizarro, O., Williams, S.B., 2020. Towards adaptive benthic habitat mapping. In: 2020 IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp. 9263–9270.

Shields, J., Pizarro, O., Williams, S.B., 2021. Feature space exploration for planning initial benthic auv surveys. arXiv preprint. arXiv:2105.11598.

Själander, M., Jahre, M., Tufte, G., Reissmann, N., 2019. EPIC: an energy-efficient, high-performance GPGPU computing research infrastructure. arXiv:1912.05848.

Williams, S.B., Pizarro, O., Jakuba, M., Barrett, N., 2010. Auv benthic habitat mapping in south eastern Tasmania. In: Field and Service Robotics. Springer, pp. 275–284.

Woods, D.C., Overstall, A.M., Adamou, M., Waite, T.W., 2017. Bayesian design of experiments for generalized linear models and dimensional analysis with industrial and scientific application. Qual. Eng. 29, 91–103.

Yang, Y., Loog, M., 2018. A benchmark and comparison of active learning for logistic regression. Pattern Recognit. 83, 401–415.