# Choosing between strategies for designing surveys: autonomous underwater vehicles

Scott D. Foster[1]*, Geoffrey R. Hosack[1], Nicole A. Hill[2], Neville S. Barrett[2] and Vanessa L. Lucieer[2]

[1]CSIRO's Wealth from Oceans Flagship, GPO Box 1538, Hobart, Tas.,7001, Australia; and [2]The Institute of Marine and Antarctic Science, The University of Tasmania, Private Bag 129, Hobart, Tas.,7001, Australia

## Summary

**1.** Autonomous underwater vehicles (AUV), which collect images of marine habitats, are now an established sampling tool. The use of AUVs is becoming more widespread as they offer a non-destructive method to survey substantial spatial areas. The design of AUV surveys has historically been based on expert knowledge and AUV-specific considerations, such as reducing geolocation error. The expert knowledge encompasses intuition, previous surveying experiences and holistic knowledge of the study region.

**2.** We investigate the statistical aspects to AUV survey design for estimation of percentage cover of key benthic biota. We investigate the presence of spatial autocorrelation in AUV data using model-based geostatistics and examine the effect of autocorrelation on survey design by examining different *design strategies* – methods for placing AUV transects. The design strategies are assessed by inspecting the expected bias and the expected standard deviation of model predictions, where the model depends on the choice of design.

**3.** The AUV data exhibited a wide range of autocorrelation, from non-existent to substantial. The design strategies varied in their statistical performance and nearly all strategies had shortcomings. Design strategies that were consistently poor performers had (i) transects placed in parallel in a single spatial dimension and (ii) made no attempt to spread out the transects in space. The superior design types had more transect-to-transect separation (but not too much) and effectively spanned important covariates.

**4.** The results give guidelines to researchers designing AUV surveys for biological mapping and for monitoring. In particular, we demonstrate that any spatial design should seek spatial balance, such as would be introduced by a systematic or stratified component within a randomized design. Knowledge of the system under study should be incorporated and, if possible, should be done so in a formal manner that is objective and repeatable.

**Key-words:** autocorrelation, autonomous underwater vehicles, geostatistics, GRTS, integrated nested laplace approximation, matérn, model-based design, temperate reef

## Introduction

A fundamental goal for all ecological surveys is that the resulting data should be able to answer the questions for which it was collected. Ideally, the survey should also aim to be *optimal*, *multi-purposed*, *robust* and *feasible* (e.g. Atkinson 1996; Stevens & Olsen 2004; Robertson *et al.* 2013). This means that the survey will provide the best value for money in terms of information content, provide sufficient information for a variety of disparate reasons/taxa (if required) and provide sufficient information if the survey is compromised due to fieldwork logistics, and be practical.

The survey design for a spatial region is the choice of the number of samples and the locations of those samples within the region. It may also include the choice of ecological components to measure (e.g. abundance of a species) as well as how these quantities get measured (e.g. laboratory replications and subsampling). It is our experience that many ecological researchers tend to be pragmatic about survey design. They base the design on their previous surveying experiences and holistic knowledge of the study region. Often, the resulting surveys generate data that provide an adequate, or even good, representation of the study region. Nevertheless, these surveys could be streamlined for both cost and researcher effort, with a minimal drop in statistical efficiency.

In this article, we focus on survey design in the marine realm where autonomous underwater vehicles (AUVs) are a powerful tool for obtaining non-extrusive ecological samples. We focus on design implications for inference about biota, which is our primary interest and is sampled by images, rather than using images to inform about the AUV's location – methods for accurately estimating the AUV location presented in Kim & Eustice (2009) and Williams *et al.* (2010). Images, such as those obtained from an AUV, have many advantages over traditional sampling methods such as dredges, trawls and

*Correspondence author. E-mail: scott.foster@csiro.au

grabs as they are non-extractive, and it is relatively easy and cheap to acquire substantial amounts of data (Williams *et al.* 2012). Images from an AUV are typically higher quality than those from other image-based platforms as they maintain a constant height above the seafloor, provide even illumination for photography and track pre-planned routes so that the images can be accurately geolocated. The data are permanently stored and can be revisited for new research directions and for quality control. AUVs are recognized as an ideal tool for image acquisition (see Williams *et al.* 2012), and in the future, underwater images will provide one of the major tools for continental and regional sampling for a variety of different objectives. These objectives include the studies of species and communities (Grasmueck *et al.* 2006; Clarke, Tolimieri, & Singh 2009; Sherman & Smith 2009; Bridge *et al.* 2011), extreme and sensitive environments (Singh *et al.* 2004; Armstrong *et al.* 2006; Sherman & Smith 2009), invasives (Barrett *et al.* 2010) and monitoring for fisheries management (Smale *et al.* 2012).

Despite the obvious appeal of AUVs, there remain statistical issues with how the resulting data are used. Spatial dependence (autocorrelation), if present in the data, is likely to be important as it changes important qualities of the data. This change, when coupled with a model or method that ignores autocorrelation, will lead to optimistic standard errors of the parameter estimates. However, the autocorrelation can be exploited, using geostatistical methods, to produce predictions that have lower variance than those from a comparable form. Unfortunately, autocorrelation is typically ignored in the analysis of AUV data (e.g. Bridge *et al.* 2011; Smale *et al.* 2012). Similar platforms such as towed video and remotely operated vehicles will exhibit similar qualities to AUVs in that autocorrelation is hard to avoid and that it is frequently ignored. Autocorrelation in data from these platforms is likely as data are acquired with small spatial separation, as the images are necessarily spatially aggregated into transects. Consideration of these inherent data qualities should have direct implications for how to design future data collection efforts. Careful survey planning will streamline, or even optimize, the effort and cost required to obtain the data resource.

Autonomous underwater vehicle surveys are generally conducted for the purpose of providing information on the distribution and abundance of species and/or communities of interest. In a statistical sense, these aims can be incorporated into *prediction*; either an areal quantity is predicted, such as the average prevalence in a region, or a location-specific quantity is predicted, such as the probability of presence at that site. Combined over many locations, the location-specific predictions form a predicted distribution map.

The statistical issues surrounding survey design for AUVs share similarities with other well-established and emerging ecological sampling techniques. Researchers have been sampling by measuring quadrats spaced along a transect for decades (example applications are Stohlgren, Bull & Otsuki 1998; Barrett, Buxton & Edgar 2009). However, unlike this sampling strategy, transects for an AUV survey are often nonlinear and crossover, which facilitates higher geolocation accuracy (Williams *et al.* 2010). Unmanned aerial vehicles are a similar technological sampling platform to AUVs and have been used to estimate terrestrial and marine animal abundances (e.g. Hodgson, Kelly & Peel 2013; Vermeulen *et al.* 2013). Aerial vehicles are often used to conduct strip transect, which is a simplification of a line-transect survey (Buckland *et al.* 1993). While the *technology* is similar to AUVs, the *data* produced are often qualitatively different – the images from the AUV transect are typically not all scored for biological content and so the data do not represent a census of the sampling area.

The aim of this article is to provide advice to researchers about strategies for designing AUV surveys for regions with no prior biological information, but for which prior environmental data exist. Despite the AUV focus, these results should be applicable to other image-yielding tools such as towed video and remotely operated vehicles, as well as to similar traditional survey designs, such as belt transects. To this end, we assess variants of popular randomized and systematic designs under the presence of autocorrelation for their prediction ability. These types of designs are easy to implement, produce data that can be multipurposed and are likely to be robust against missed transects that arise due to extraneous circumstances in the field. These are not always attributes of model-based designs, which are not probabilistic and require a correct, or at least adequate, known model on which to base a design. We assess a series of design strategies, under the presence of a variety of levels of autocorrelation, by conducting Monte Carlo studies (see Section Performance of Different Designs). The amount of autocorrelation is based on that observed in a number of different taxonomic groupings, so the results of the simulation study are firmly grounded in reality. Across this diverse set of design strategies, we demonstrate that successful designs have common elements, which could easily be incorporated into any AUV field survey.

## Materials and methods

### IMAGE DATA AND ITS COLLECTION

An AUV is a submersible robot that is deployed from a research vessel. It traverses a pre-programmed transect before surfacing for redeployment and/or maintenance. The type of AUV considered in this study carries a camera that captures images on a short time scale, with sequential images having minimal spatial separation (Williams *et al.* 2012). The motivating data for this study were obtained using the AUV 'Sirius', see http://www.acfr.usyd.edu.au/research/projects/subsea/auvSIRIUS. shtml. The AUV Sirius can maintain constant height above the seabed and thus achieve constant scale and optimal lighting for the acquired imagery. A combination of on-board navigation tools allows precise underwater navigation, and post-processing of imagery (Williams *et al.* 2012) allows further spatial precision to be given to the acquired imagery (metre-scale precision over kilometre-scale transects). When processing the images collected in the field, each image can be enumerated for a number of different attributes, typically these are the presence/absence or the abundance of a species, morphotaxa or higher-level taxonomic groups. The choice of taxonomic detail depends on the quality of the imagery, the habitat being targeted, the visual similarity of the taxa sampled and the level of detail required by the research question.

In this study, we utilize data obtained from the AUV Sirius while surveying an extensive temperate reef feature surrounding the Hippolyte Rocks, an offshore island group in SE Tasmania, Australia (see Fig. 1). The reef system is composed of highly fractured dolerite and granite, which is steeply sloping with its base in water depths of around 90 m, and is distinct from many of the coastal reefs in the region. The Hippolyte Rocks was surveyed in 2008–2009 using multibeam sonar to develop a detailed habitat map, which could be used for subsequent mission planning. This was followed by an AUV deployment to acquire detailed imagery to describe benthic faunal cover and habitat distribution (Nichol *et al.* 2009). Six AUV transects were conducted on the Hippolyte Rocks to cover the range of depth and other environmental gradients that may occur across this complex topographic feature. AUV transects were approximately 2·5 km in total length and followed a 'clustered sparse grid' (CSG) pattern (see Fig. 1, which has six transects). In the CSG, there is one long central 'subtransect', which was oriented down the depth gradient, and four shorter subtransects oriented perpendicular to the first. This design was used to capture the down-depth gradient, expected to affect the distribution of biota, but also to achieve replication across depth bands. The 'crossing over' in the AUV transect path facilitated the accurate ($\sim$ 1 m) georeferencing of images on the seafloor (Williams *et al.* 2010).

Autonomous underwater vehicle images were scored for the biota with a frequency of every 100th image to reduce gross autocorrelation from scoring the same patch of sea floor. This corresponded to an approximate spatial separation of 30–40 m between images. Subsampling images also provide a means to manage the substantial manual effort required to score images. Preliminary analysis indicate that this frequency was sufficient to capture the full range of habitats and environmental gradients present in the region (N. Hill, unpublished data). Within each image, 50 random points were scored and biota identified

to the species or morphospecies level where possible. Biota was subsequently aggregated to eight coarser morphotypes that reflect a spectrum of types. The types of morphotypes considered include dominant habitat-forming taxa, different environmental responses and several morphological complexities.

The physical environmental variables were derived from 2 m grids of bathymetry and backscatter from a multibeam echo-sounder system (Nichol *et al.* 2009). The covariates chosen for this analysis were selected from the spatial products generated to describe the topographic and textural characteristics of the seafloor as determined by Lucieer *et al.* (2013). These variables included depth (bathymetry), and the mean and standard deviation of the backscatter texture as derived from the novel segmentation procedure described in Lucieer *et al.* (2013). The mean and standard deviation of the backscatter represents the textural 'hardness' and 'roughness' of the seafloor, respectively. We further limited our sampling frame to only consider reef substratum using the habitat classification for this region from Lucieer *et al.* (2013).

### GEOSTATISTICAL ANALYSIS

Many researchers working with AUV data ignore spatial autocorrelation, or only account for it in the systematic component of the model (by adding spatial covariates in the model). This approach is likely to lead to deflated standard errors of the model's parameters. Further, a comparable geostatistical analysis can decrease the prediction variance by incorporating the spatial dependence *and* the proximity of the prediction to the observations. A statistical model for towed video (a device with many of the same properties as AUV) that accounts for autocorrelation has been presented by Foster & Bravington (2009). This model treated the transects as *linear* only and assumed that different transects were independent. Both of these assumptions are likely to be unsatisfactory for AUV data as individual transects are typically not straight lines, crossover paths and are often in spatially similar regions.

To understand the range of spatial autocorrelation and the dependence on the physical covariates in the Hippolyte data, we performed a model-based geostatistical analysis (Diggle & Ribeiro 2007) for each of the 8 morphotypes. The nomenclature used throughout this manuscript is applicable to any of the morphotypes, and this removes unnecessary and confusing notation. At spatial location $i = 1 \ldots n$, the presence/absence of a morphotype, $y_i$ say, was modelled using a latent spatial random variable, $S_i$, and potentially as a function of the $v = 1 \ldots$ 3 environmental covariates, $\{x_{ki}\}$. That is

$$\text{logit}(\mathbf{E}(y_i|S_i)) = \sum_{v=1}^{3} f_v(x_{vi}) + S_i \qquad \text{eqn 1}$$

where logit($\cdot$) is the standard log-odds link function. Preliminary data exploration suggested that the parametric form of the function, $f_v(\cdot)$, is likely to approximated by a linear or quadratic polynomial. The latent spatial variable $S_i$ was assumed to be Gaussian with correlation, between locations, given by a Matérn function with smoothness parameter set to $v = 1$ (for details of Matén correlation, see page 51 of Diggle & Ribeiro 2007). The covariance between two sites is the correlation between sites multiplied by a variance that is not dependent on spatial location. The spatial variance partially quantifies the potential impact, of the latent spatial variable $S_i$, on the predictions of the quantity of interest ($y_i$).

Note that a substantial spatial dependence and/or variance could be indicative of three things: (i) the morphotype naturally varies at a large spatial scale (after covariate effects have been removed), (ii) that there
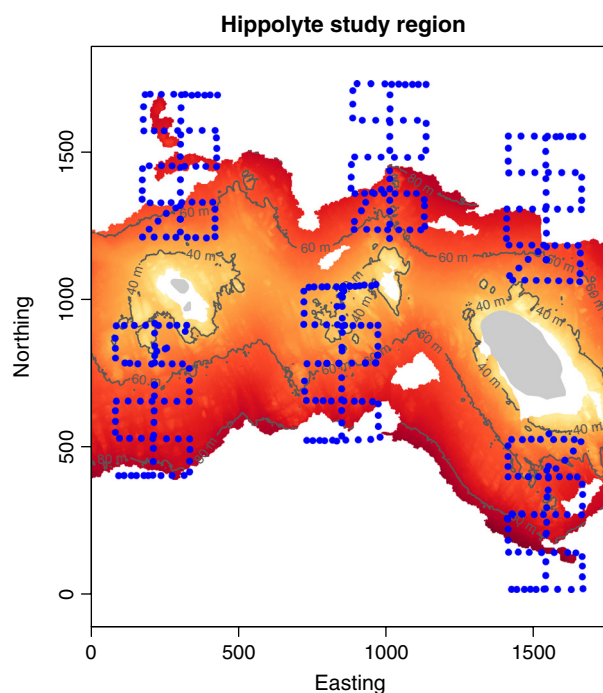


**Hippolyte study region**

**Fig. 1.** Hippolyte Island study area showing: islands (grey), image locations (blue dots) and depth in colour and contours. The depth ranged from 13 to 90 m throughout the region. Only the locations with hard substrate (the study area) are plotted, white areas are not in the study area (soft substrate).

are important covariates omitted from the analysis or (iii) the covariates are predicted from data that are not colocated with the AUV data (see Foster, Shimadzu & Darnell 2012).

We performed model selection for the covariates by fitting a model with every covariate combination where the polynomials were either linear or quadratic. The form of the model chosen was that which had the highest marginal log-likelihood, the log probability of the data given only the model form. This is a well-known quantity and forms the basis of selection statistics such as BIC (Schwarz 1978) and Bayes factors (Kass & Raftery 1995).

Estimation of model parameters was performed by posing the geostatistical model's latent spatial variable ($S_i$) as a Gaussian Markov random field (Lindgren, Rue & Lindström 2011), which enables estimation through the *integrated nested Laplace approximation* (INLA; Rue, Martino & Chopin 2009 implemented in the INLA R-package, version 0.0-1376414568). Importantly, this approach allows for posterior interpretation of model parameters and predictions, just as Markov chain Monte Carlo methods do, but with a drastically reduced computational load. We refer to Cameletti *et al.* (2013) and Blangiaro *et al.* (2013) for introductions to the INLA approach. We use the default priors on all parameters.

## PERFORMANCE OF DIFFERENT DESIGNS

Our focus in this article is to provide advice to field researchers about how to survey areas that have no available biological data. We consider variants of well-known and common designs, namely simple random sampling, stratified (across space) random sampling, systematic sampling and GRTS sampling (for generalized random tesselation stratified; Stevens & Olsen 2004). The GRTS sampling strategy provides a spatially balanced sample – one that has samples well spread through the study region and does not have many samples in close proximity to each other. GRTS sampling is not the only design strategy to provide spatial balance (Robertson *et al.* 2013 for another recent approach), but it is now well established in natural resource research. Our study designs should all be feasible, up to minor departures for known geographical features. For this study region, there is a preference for transects that run north to south, like those performed in the original survey (see Fig. 1). This is due to the expectation that transects in this direction will cover more variation in the physical environment and therefore also in the distribution of morphospecies. In total, we investigated 13 survey designs, of which an example is given in Fig. 2. The design types are as follows:

*Random 1D*: Six transects run north to south, and their locations (eastings) are completely random. This is the most basic design type and is very easy to implement.

*Random 2D*: Three transects run north to south, and three run east to west. Locations are completely random. This design is a modification of Random 1D that could space images more effectively.

*Radial Rad*: The starting position of transects is randomly placed along a line that approximates the ridge between the main islands. The direction of the transect is proportional to the distance along the ridgeline. This gives a set of transects that run almost radially from the ridgeline. The resulting set of transects are all approximately down slope, the preferred deployment direction for the AUV. There are twelve of these transects, as each transect is approximately half the length of a full north–south transect. The radial nature may enable sampling of the shallower regions more effectively.

*Stratified 1D*: Six transects run north to south, and each transect is randomly placed in a separate easting 'strata' so that transects are likely to be well separated. The strata divide the eastings into six

non-overlapping intervals with equal area. The stratification will spread the images throughout space more effectively than the simple random approach.

*Stratified 2D*: Three transects run in both directions. The strata divide the eastings and northings into non-overlapping intervals with equal area. Combines the strengths of the Random 2D and the Stratified 1D in that images should be more effectively spaced throughout the study region.

*Stratified Rad*: As Radial Rad., except that the start locations of the transects are placed in strata along the ridgeline. The strata divide the angles into non-overlapping intervals. The stratification will place images more evenly throughout the study area than the Random Rad. design.

*Grid 1D*: Six transects run north to south and are equidistant from each other. The first transect is randomly placed. Images are very evenly spaced throughout the study region at the expense of randomization.

*Grid 2D*: Three transects run in both directions. The locations of the first transects, in both directions, is generated by random. Images are very evenly spaced, and there are crossovers between transects, which may help autocorrelation estimation.

*Grid Rad*: As Radial Rad., except that the start locations of all transects are systematically placed to be equidistant. Space images evenly throughout the study region, within the constraint of using radial transects.

*GRTS 1D*: Six transects run north to south. The locations of the transects are generated using the GRTS framework. Transects should be spatially balanced, but the images will not be (they are constrained to be within a transect).

*GRTS 2D*: Three transects run in both directions. Locations of transects, in each direction, are generated using the GRTS framework. The placement of transects in one direction does not use information about the locations of the transects in the other direction. Transects in each direction will be spatially balanced, but the images will not be.

*Clustered Sparse Grid (CSG)*: Four transects, with pattern taken from the original Hippolyte Rocks data, are placed in one of four quarters of the survey area. The quarters are delineated by the midpoints of the easting and northing of the study area. Retains the crossover pattern that is common for AUV surveys and spreads the transects out in space.

*Status Quo*: The implemented design (displayed in Fig. 1) with the middle two transects removed. The removal is to even the number of images in the survey region, and the middle two were chosen to leave a set of spatially distinct transects. This decision was made in consultation with the experts who originally designed the survey. The status quo design is similar to a particular realization of the CSG design, but it is a design with almost extreme intertransect separation. It is not randomly chosen and represents a design that encompasses the experts' knowledge about the study area. Since it is not random, the use of design-based estimates of areal quantities, such as sample means, is not formally justified, and calculation of standard errors is problematic at best. This is inconsequential if estimation is performed using a model.

A desirable attribute of the simulation study is that it should reflect the Hippolyte data as much as possible and only alter specific aspects of it to test the different designs. For this reason, we restricted the number of transects to 6 for most designs as performing additional transects requires a substantial amount of extra expensive fieldwork. The exceptions were (i) the radial designs, which are on average only half length and therefore had double the number of transects and (ii) the CSG and status quo designs that had a reduced number of transects to maintain parity with respect to the number of images in the other designs. We
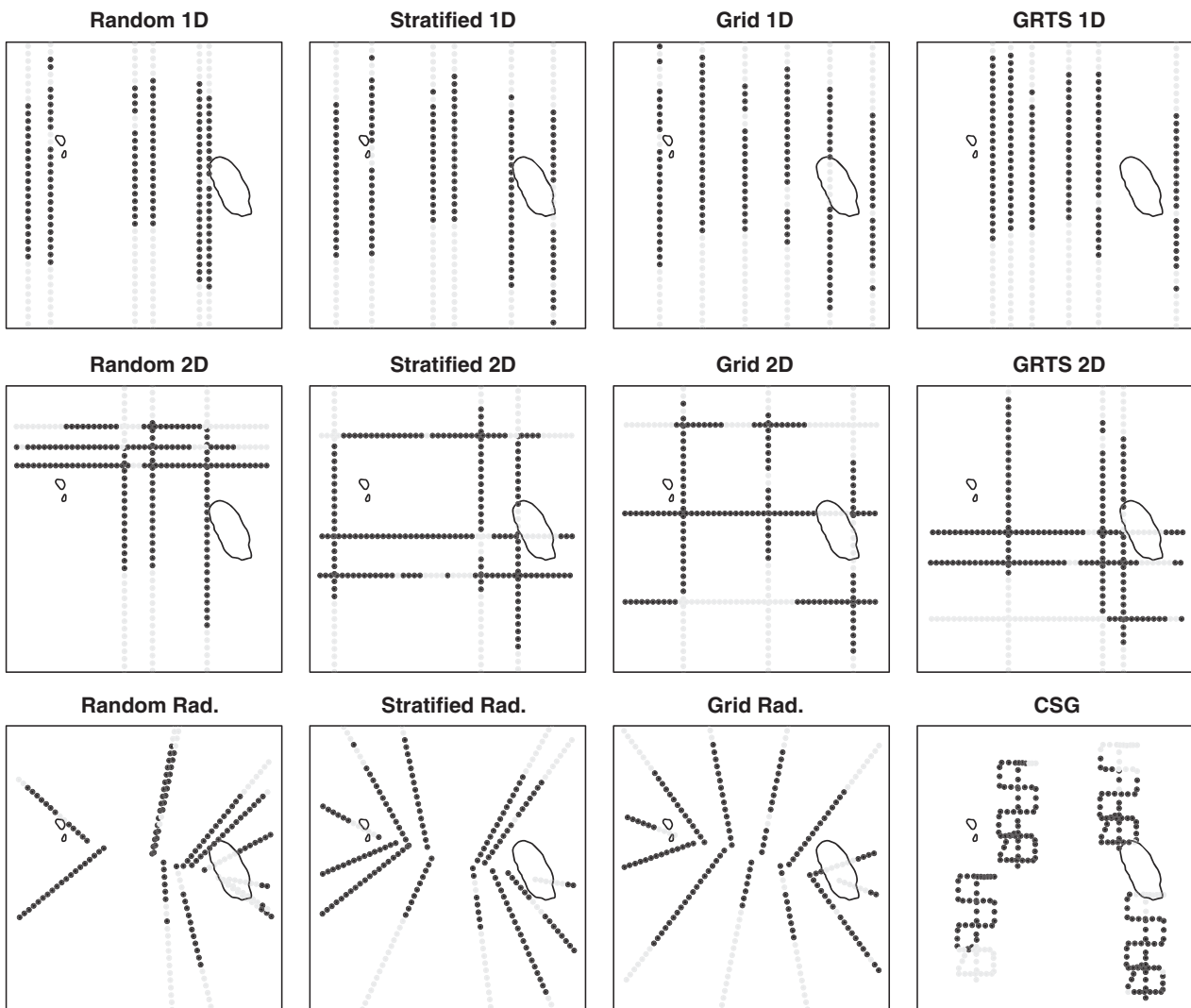
**Fig. 2.** Example designs for the study region. Black dots are images over hard substrate and will be included in the analysis. Grey dots are those images over soft substrate or in shallow water; they are not included in the analysis. The status quo design is omitted from this plot, but it consists of the four outer (east–west) transects in Fig. 1.

envisage that two radial transects could be completed in a single AUV deployment, which is possible since one end of all transects is spatially clustered in the middle of the sample region. Hence, the nominal number of transects for radial designs remains six. In all designs, the spacing of the images along the transect was 37·45 m, which matches the spacing in the observed Hippolyte data.

All designs, except status quo, are obtained by randomly placing transects in the study region, and, for simplicity's sake, we ignore the geographical features of the study area when generating the initial design. To incorporate the geographical features, we remove generated locations that are placed on the islands or are not over reef, see Fig. 2. This design process introduces a variable number of images within different realizations of each design strategy and between design strategies. This is an unfortunate but necessary source of variation in the simulation study.

We include the GRTS designs as researchers are starting to use them for these types of experiments, and we want to examine the usefulness of such an approach. With small sample sizes, like the number of transects used here, we would not expect large benefits over more traditional designs. Also, the GRTS 2D design only balances over each

dimension – it does not balance over two-dimensional space. We are currently unaware of any design algorithm that can perform this task for anything other than point samples.

For the CSG and the status quo designs, the individual transect pattern could have been chosen differently. One choice would be a pattern with more downslope subtransects and less crossovers. We believe that these are likely to be minor variations in the theme: images in a transect are clustered, and transects are well separated.

### Evaluation of designs

The goal in evaluating the designs is to find the design type that efficiently and reliably contains the information needed to make good predictions (see Müller 2007 for design principles in a spatial setting). In the situation considered in this article, we assume that the design is for a *previously unsampled location* that has the required covariates measured. This is the typical situation in many, but not all, imaged-based marine surveys. This design problem has been termed a *prospective* design by Diggle & Lophaven (2006) and Diggle & Ribeiro (2007). Note that we carry information from the previously collected data, see

Section Image Data and Its Collection, into the comparative study only through its parameter's posterior distribution, $p(\theta|y)$, and drop any information about the spatial random variable $S$. This allows us to investigate different realizations of the spatial patterns that morphotypes may take. Our motivation for doing this is that a multipurposed design strategy must be able to accommodate all these realizations.

Intuitively, a good design will enable the reproduction of the underlying spatial random variable, see Section Geostatistical Analysis. Note that the spatial surface is dependent on the random value of the latent spatial variable, $S_i$, *and* dependent on the model's systematic component $\sum_{k=1}^{3} f_k(x_{ki})$. All of these terms, in addition to the parameters defining the distribution of $S$, must be estimated from the hypothetical survey data. This process mimics the analysis phase for data from a survey of an unsampled site.

The notation used from herein is $y_n^*$ is the morphotype presence/absence surface that will be predicted, $y$ the set of observations from the original Hippolyte data, $y_d^*$ the set of hypothetical observations from survey design $d$; $S_n^*$ the spatial random variable for the hypothetical sampling frame (Monte Carlo realization). In practice, $y_n^*$ and $S_n^*$ are defined on a dense grid of almost 332 000 locations, and $y_d^*$ are a subset of $y_n^*$. To evaluate a design for a particular set of hypothetical survey data, we compare the set of posterior predictions, $\mathbf{E}(y_n^*|y_d^*)$, against the simulated spatial surface. For computation reasons, we predict on a thinned grid with $K = 3334$ locations. The thinning process is not performed for $y_d^*$.

We evaluate the posterior predictions by summarizing each predicted surface through a small number of functions, called *loss* functions, which quantify the discrepancy between model and simulated data. These loss functions are similar, in spirit, to those used in Diggle & Lophaven (2006), although we note that any loss function could be used (see Warren, Perez-Heydricha & Yunus 2013 for a risk-based loss function).

**1.** *Mean bias*: Average difference between the *predicted* probability of presence and the *actual* (simulated) probability of presence, $\mathcal{L}_1(d, y_d^*, \theta, S_n^*|y) = 1/K\sum_{k=1}^{K} (\hat{\mu}_{n,k} - \mu_{n,k})$, where $\hat{\mu}_n = \hat{\mathbf{E}}(y_n^*|y_d^*)$ and $\mu_n = \mathbf{E}(y_n^*|S_n^*)$ from (1). This measures the design's ability to estimate an areal quantity. Systematic deviations from zero imply bias and are undesirable property.

**2.** *Mean prediction uncertainty*: Average posterior standard deviation, $\mathcal{L}_2(d, y_d^*, \theta, S_n^*|y) = 1/K\sum_{k=1}^{K} (\hat{\sigma}_k)$, where $\hat{\sigma}_k$ is the estimated posterior standard deviation of the location's prediction. This measures the modelled uncertainty in the set of predictions. Low values of $\mathcal{L}_2(d, y_d^*, \theta, S_n^*|y)$ indicate less uncertainty and better predictions.

**3.** *Worst-case bias*: Maximum absolute difference between the *predicted* probability of presence and the *actual* probability of presence, $\mathcal{L}_3(d, y_d^*, \theta, S_n^*|y) = \max_k\{|\hat{\mu}_{n,k} - \mu_{n,k}|\}$. This measures the scale of departure of predictions for a set of spatial locations. If the surface is estimated well, then $\mathcal{L}_3(d, y_d^*, \theta, S_n^*|y)$ will be small.

**4.** *Worst-case prediction uncertainty*: Maximum posterior standard deviation, $\mathcal{L}_4(d, y_d^*, \theta, S_n^*|y) = \max_k\{\hat{\sigma}_k\}$. This measures the maximal uncertainty for a design. Small values of $\mathcal{L}_4(d, y_d^*, \theta, S_n^*|y)$ indicate that all locations' predictions have low uncertainty.

All the summaries are functions of the random observations for the $d^{th}$ design ($y_d^*$), the random spatial effects ($S_n^*$) and the model's parameters ($\theta$). Hence, the value of the four loss functions is also random, and a summary of its distribution needs to be inspected for inference. Here, like elsewhere in the Bayesian design literature (Chaloner & Verdinelli 1995; Müller 1999; Diggle & Ribeiro 2007), we summarize using the expectation of the loss functions. The expectation is found by Monte Carlo methods with by averaging random draws from the joint distribution, viz.

$$
\begin{aligned}
\mathcal{L}_h(d|y) &= \mathbf{E}(\mathcal{L}_h(d, y_d^*, \theta, S_n^*|y)) \\
&= \int \mathcal{L}_h(d, y_d^*, \theta, S_n^*|y)p(y_d^*|S_n^*, \theta)p(S_n^*|\theta)p(\theta|y) \\
&\quad d\theta dS_n^* dy_d^* \approx \sum_{b=1}^{B} \mathcal{L}_h\left(d, y_d^{*(b)}, \theta^{(b)}, S_n^{*(b)}|y\right) \\
&\quad p\left(y_d^{*(b)}|S_n^{*(b)}, \theta^{(b)}\right)p\left(S_n^{*(b)}|\theta^{(b)}\right)p\left(\theta^{(b)}|y\right).
\end{aligned}
\tag{eqn 2}
$$

where $p(\cdot)$ is the probability density (or distribution) defined by its arguments. Each of the $B$ sets of samples, $\{\theta^{(b)}\}$, $\{S_n^{*(b)}\}$ and $\{y_d^{*(b)}\}$, as well as the corresponding value of the loss function, is obtained using the Algorithm 1. We use $B = 100$ in this study. The set of designs in Fig. 2 is an example a single realization of the randomized designs.

Algorithm 1.

### Lemma

Simulation and summary of a single Monte Carlo realization.

**1.** Draw model parameters $\theta^{(b)}$ from their posterior, $p(\theta|y)$, where $y$ is the Hippolyte data,

**2.** Draw spatial random variable $S_n^{*(b)}$ from $p\left(S_n^{*(b)}|\theta^{(b)}\right)$,

**3.** Draw the observation data $y_d^{*(b)}$ from $p\left(y_d^{*(b)}|S_n^{*(b)}, \theta^{(b)}\right)$. The locations are themselves randomized for each design (except Status Quo).

**4.** Calculate the value of $\mathcal{L}_h(d, y_d^{*(b)}, \theta^{(b)}, S_n^{*(b)}|y)$ and save. Calculation of this function requires estimation of the posterior distributions, $p\left(y_n^*|y_d^{*(b)}\right)$. This is performed using the INLA method (Rue, Martino & Chopin 2009).

Important attributes of this simulation approach are: (i) the uncertainty in the model's parameter values is included in the simulation model, prediction and the loss function – we do not use plug-in values; (ii) the predictions are made from the (posterior) predictive distribution conditional on the new data only – if the new data do not support good estimation of model properties, then predictive abilities will suffer; and 3) the model is non-Gaussian. Most of these issues have been visited before for Bayesian experimental design (see Müller 1999; Zidek, Sun & Le 2000; Diggle & Lophaven 2006; Diggle & Ribeiro 2007, but its use is not yet widespread. Importantly, avoiding plug-in values better represents the uncertainty of the whole system and an analysis on data arising from it. The gain is the ability to assess designs using a variety of plausible model formulations. This makes the inferences more robust to model misspecification as the model's uncertainty is incorporated. The extra cost is computational, which is partially mitigated by recent advances in computing (Rue, Martino & Chopin 2009; Lindgren, Rue & Lindström 2011).

The first step in the Monte Carlo routine is the only place where information from the real biological data (from the Hippolytes) are used. This is done to cement the simulation of data sets into reality. Without this, the parameters for the model would have to be chosen from prior belief without knowledge from previous data. Also, only incorporating the real biological data to inform the choice of parameters is consistent with our research objectives, which are to provide advice about surveying areas with no available biological data. So, when making predictions, we have to ignore the previous information on the spatial random variables. Of course, there are situations when the previous data *should not* be ignored, such as revisiting previously surveyed areas to obtain data that *augments* the current data. This could be incorporated into (2) by conditioning all quantities on the observed data *and* the latent spatial variables, as well as the new data.

# Results

## GEOSTATISTICAL ANALYSIS

The geostatistical analyses showed that there is substantial variation in the way that the different morphotypes respond to the environment, both in terms of the covariate effects and in the properties of the latent spatial variable (Fig. 3 and Table S1 in the Supporting Information). For example, all morphotypes responded to depth with most morphotypes favouring water with moderate depths (approximately 55–65 m), although there is uncertainty around this preference. The exceptions to the general depth pattern are *Ecklonia radiata* (a kelp) and Bryozoa, which both preferred shallow water. *Ecklonia radiata* is dominant in shallow water, which is both light (enabling photosynthesis) and more exposed to wave action (Hill *et al*. 2010). No morphotype had any relationship with the mean backscatter value, at the scale that the multibeam echo sounder data were processed. Recall that all of the images used in this analysis are on hard substrate, so this result implies that the distinction between the *level* of hardness within the study region is immaterial. Two morphotypes – Gorgonians and Soft Corals – were related to the standard deviation in backscatter. Both had a higher probability of presence in locations with variable backscatter, that is, those locations with at least a moderate amount of seabed textual complexity (rugosity).

The amount of spatial dependence varied across morphotypes: Crinoidea (high posterior spatial correlation with expected correlation of ~0·3 at 1000 m separation) and cup sponges (low posterior spatial correlation with expected correlation of ~0·1 at 80 m) were the two extremes. The posterior dependence, especially for Crinoidea, has substantial uncertainty as does the posterior distributions for most of the other parameters in the model. In spite of the uncertainty, there is clearly evidence to suggest that assuming that the images are independent is likely to be misleading, as the average interimage distance is only 33·45 m. The amount of spatial variance also varied from substantial (*Ecklonia radiata*) to inconsequential (cup sponges).

## SIMULATION STUDY RESULTS

We present only the results for the simulation based on the Soft Coral morphotype and for the first two loss functions, for the sake of brevity, given in Fig. 4. The remaining simulations based on the other seven morphotypes follow the same general patterns, with minor differences. The complete results are given in the Supporting information.

The mean bias summaries ($\mathcal{L}_1(d|\boldsymbol{y})$) indicate that all design strategies will over predict (predictions are higher than they should be). The amount of bias is not constant over the different design types. The bias ranges from
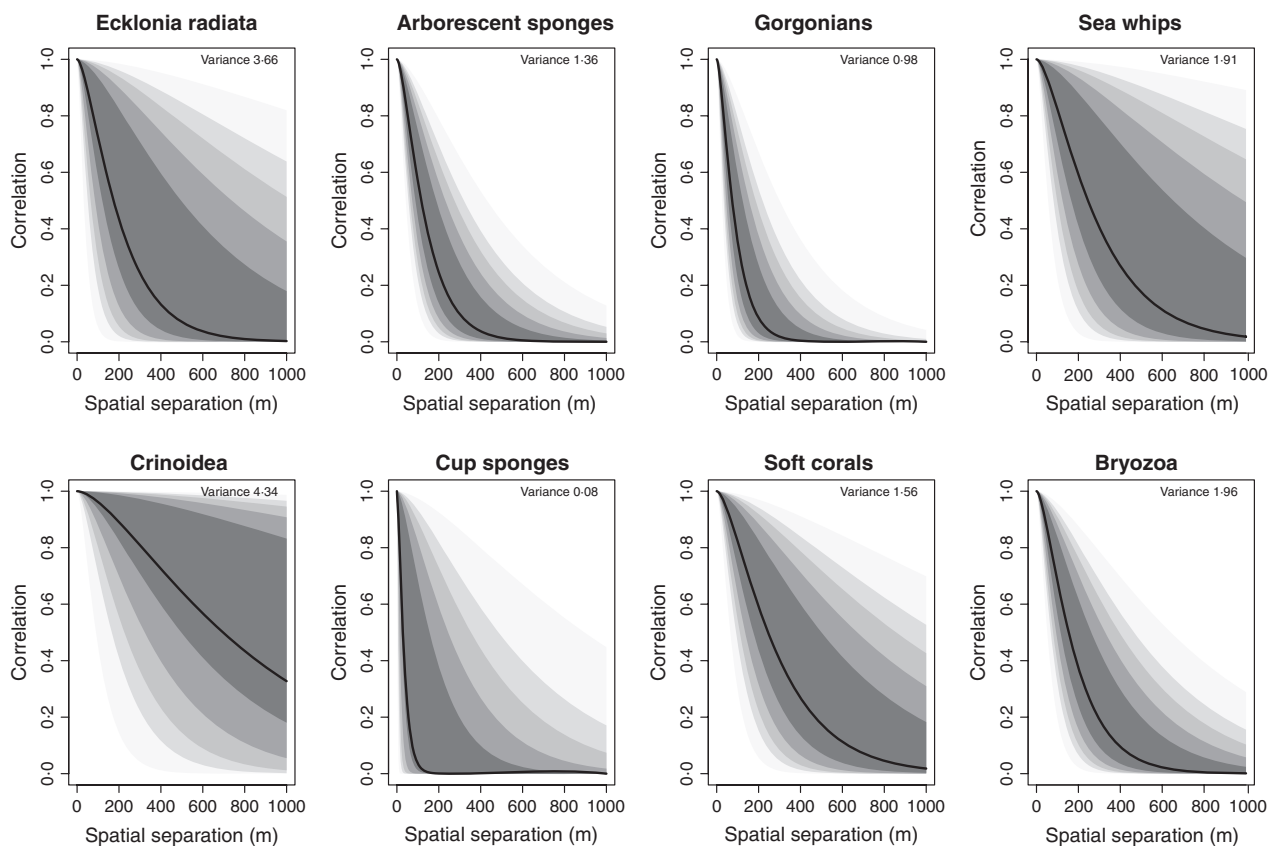


**Fig. 3.** Posterior summaries of autocorrelation (Matérn) function. Solid line is the marginal posterior expectation, and shading corresponds to the 60th, 80th, 90th, 95th and 99th quantiles of the posterior distribution. The spatial variance estimates are the mean of the marginal posteriors.
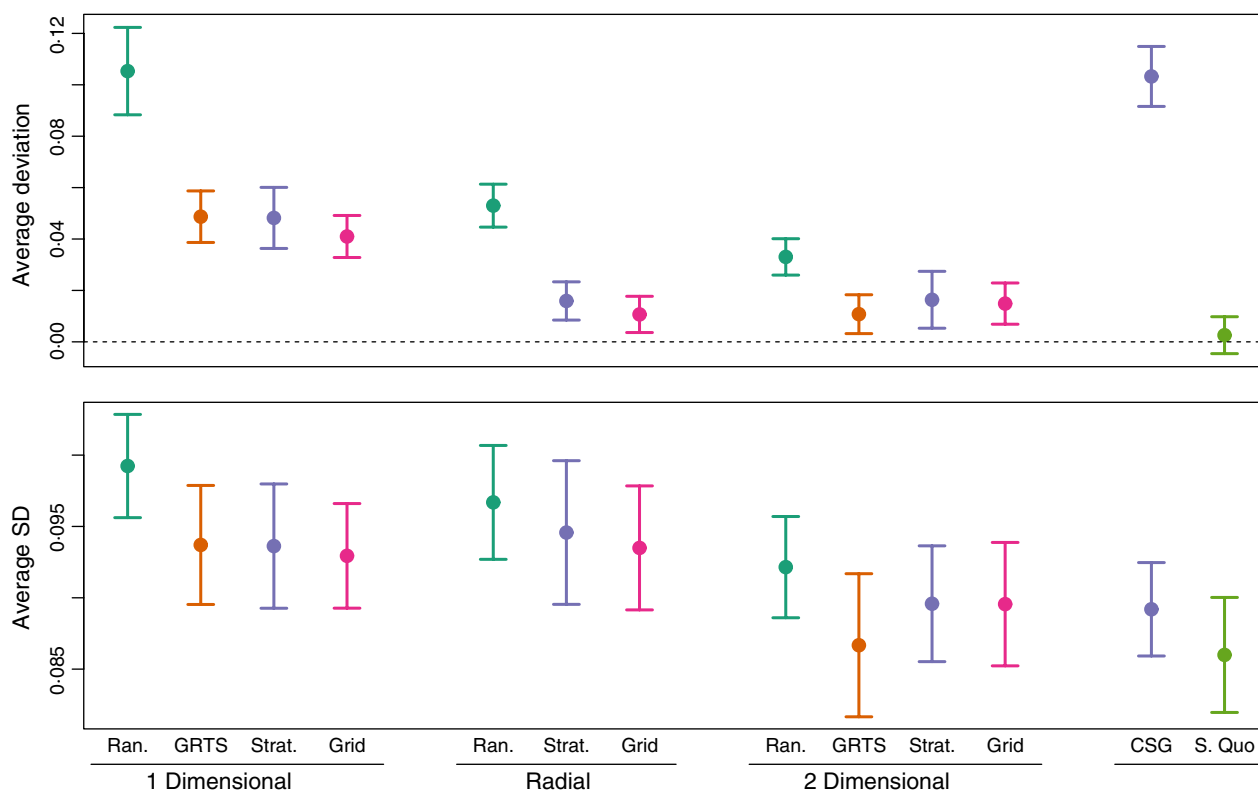
**Fig. 4.** Results for the simulation based on the Soft Coral morphotype. The points are the expected values of the loss functions for average bias ($\mathcal{L}_1(d|\mathbf{y})$, top panel) and average prediction standard deviation ($\mathcal{L}_2(d|\mathbf{y})$, bottom panel). The intervals are the 95% confidence intervals for the expected values. The expectations and confidence intervals are for the 100 simulation replicates. Values near zero for $\mathcal{L}_1(d|\mathbf{y})$ indicate low bias. Smaller values of $\mathcal{L}_2(d|\mathbf{y})$ indicate lower uncertainty.

substantial (CSG and Random 1D) to negligible (status quo). The worst-case bias summaries ($\mathcal{L}_3(d|\mathbf{y})$) show a similar pattern to the mean bias. Note that the absolute bias is bounded to [0,1] as it is a prediction of a probability. Given this, the worst-case bias is large – it occurs when the prediction is almost the opposite of truth. In general, the one-dimensional designs give high mean bias and high worst-case bias, while the designs with more systematic structure (two-dimensional, radial and Status Quo) show less bias. The exception is the CSG design, which performed similarly to Random 1D.

The designs also delineate with respect to mean prediction uncertainty, $\mathcal{L}_2(d|\mathbf{y})$, although the practical difference may be immaterial. In general, the two-dimensional designs performed better than the one-dimensional designs. The worst-case prediction uncertainty, $\mathcal{L}_4(d|\mathbf{y})$, provides some further discrimination to the mean prediction uncertainty. When simulations for all morphotypes are considered, $\mathcal{L}_4(d|\mathbf{y})$ additionally suggests that the more systematically structured randomizations perform better – in particular, the grid and stratified designs perform better than completely random. The GRTS 2D randomization produces an equivalently low-level maximal uncertainty (measured by $\mathcal{L}_4(d|\mathbf{y})$) to stratified and gridded designs.

The method to generate designs had the unfortunate but necessary side effect of varying the number of images for different realizations of designs and for different designs. The

average number of images per design and per species is given in the Supporting information. The designs with the highest numbers of images were status quo, CSG and the two-dimensional designs. To investigate the possibility that the differences we saw were from this side effect rather than from the different design strategies, we looked to see whether there was any relationship between the loss functions and the number of images. This was done by regressing the loss functions against the design strategies average number of images. For the average bias, $\mathcal{L}_1(d|\mathbf{y})$, and the average prediction uncertainty, $\mathcal{L}_2(d|\mathbf{y})$, we found there was a very slight negative trend – the designs with more images had less bias and predicted better. However, none of the relationships had a significant slope. The slight negative trend was also seen in the worst-case loss functions. However, for these functions, the relationship was significant. If the uncertainty in the average number of predictions and the average loss function were accounted for in the regression, then the evidence for the relationship would be further diminished.

## Discussion

The analyses of the data from the Hippolyte Rocks suggest that for some morphotypes, at least, there is a strong spatial effect (see Section Geostatistical Analysis). Intuitively, this autocorrelation will have implications about how future designs are conducted as it matters *where* the sampling locations are, with respect to each other. The simulation study in

Section Performance of Different Designs confirms this intuition by showing that different design strategies contain quite different levels of information in terms of the quality of areal and point predictions. Our results suggest that those designs with greater spatial balance, as might be introduced by including stratification or other systematic components, tend to provide more information for estimation.

It is commonly considered that designs for prediction (not estimation) should be space filling giving more or less uniform spatial distributions (Royle & Nychka 1998 give some justification). On the other hand, designs that account for the uncertainty from not knowing parameter values, which have to be estimated from the data, should be less regular (Diggle & Lophaven 2006; Diggle & Ribeiro 2007) and contain some locations that are spatially near others. In fact, design-optimality criteria are antithetic for these two objectives (Zimmerman 2006), which has lead to compound criteria (Zimmerman 2006; Müller & Stehlı́k 2010). The addition of the spatially close locations is thought to aid estimation of the spatial dependence, especially for morphotypes with short spatial dependence (see Lark 2002). In the context of AUV design, it is not immediately clear what implications these principles will have. In an AUV survey, the images within an AUV transect are spatially similar while between transects they are typically spatially separate.

Clarity around this AUV-specific quandary is given by the simulation study, Section Performance of Different Designs. Designs with more transect-to-transect spatial separation perform better, on the whole, than designs with little or no structure. Care should be taken to remember that images within these transects are generally close relative to the distance of images between transects. The CSG design type is, in many ways, the extreme of this principle and shows that it is possible to 'over-cluster' within transect images. CSG transects, separated using stratification, could still be too compact to provide adequate information for model estimation and prediction. However, even this design type can produce good designs – the status quo is one such design and its success could be attributable to a good design for the systematic part of the model. It is possible that particular designs from other design strategies could perform even better.

A key point of the simulation study in Section Performance of Different Designs is that the status quo design is successful with respect to the other design strategies. The status quo design is based on experts synthesis of prior knowledge (formal and informal) of the region, which incorporated information on how environmental gradients affect biological assemblages in the study region. The status quo design exhibited the following qualities:

**1.** It has good coverage of important covariates (such as depth). The distribution of depth is different for the status quo design compared to the other designs, on average (Fig. 5). In particular, there are more deep images, although the difference is modest.

**2.** It has (almost) maximal separation between transects with clustered observations within transects. The distribution of image-to-image separations is different for the status quo design (Fig. 5). The status quo design has more large separations, fewer moderate separations and a reasonable number of short separations (the CSG design has more short separations).

While these are factual observations, it is important to note that the good performance of the status quo design could have occurred by *chance*. A good design can still be generated from a design strategy that is, on average, poor. The process used to generate the status quo design required *a priori* knowledge of which environmental gradients influence assemblages, and the transects were arranged to achieve this. Even sampling of envi-
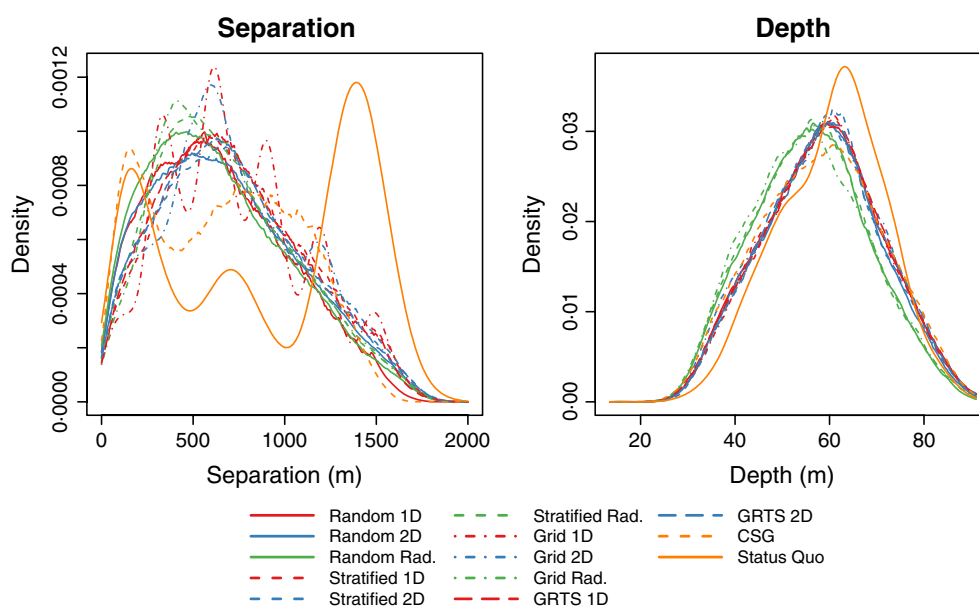


**Fig. 5.** Left panel: distribution of separation distances between sample locations. Right panel: distribution of depths at sample locations. The values represent the median (over 100 design realizations).

ronmental gradients has previously been shown to be a beneficial strategy (Hirzel & Guisan 2002). For the Hippolyte Islands region, this strategy requires knowledge of relatively few environmental gradients and primarily requires knowledge of depth, which is strongly associated with other important gradients.

Our study only considers a single data set, and immediate generalizations to all data sets should not be made, although conclusions are robust across the five morphospecies considered. However, it is possible to hypothesize about the consequence of altering the designs considered and the amount of information available for the design process. If the researchers were able to conduct more transects, giving more images, then we would expect that the amount of bias and the prediction uncertainty would decrease. This reduction is a natural consequence of incorporating more information into the analysis and making predictions given set of observations with better spatial coverage. If more images were scored, on the same number of transects, then it is likely that the autocorrelation will be better estimated, at least for small separations. Better estimates of autocorrelation will reduce the uncertainty in the model and so a deduction in prediction variance will also be obtained. However, we strongly suspect that performing more transects will reduce uncertainty by a much greater amount than scoring more images. This is because many prediction locations will still be far from the observed locations, and the benefit of having a more precise autocorrelation function is subsequently not important.

The model-based estimators that we use in this paper allow flexibility in design as it was immaterial how the sites were chosen, provided they contain enough information about the environmental gradients and other terms in the model and design. This should be contrasted with sampling-based estimators, which do require explicit knowledge of the design. In fact, design-based estimators for AUV surveys are not straightforward to derive, as randomization is restricted through the necessity to perform transects from which images are sampled. Importantly, simple averages of the data (and their standard errors) will have undesirable properties as these data are not independent, and the average will not reflect important covariates.

There are some aspects of the results from the simulation study that are concerning. Chief amongst these is the propensity for all design types (excepting the single status quo design) to overestimate the probability of observing the morphotypes (positive expected bias term $\mathcal{L}_1(d|\mathbf{y})$). We suspect that this bias is largely due to very slight overprediction for numerous locations spatially removed from an observed location, especially those that are in deep water. There are two remedies for this problem: (i) increase the spatial coverage of the design so that prediction points are close to image locations and (ii) try to ensure that the design provides a good estimate for the systematic part of the model. The evidence for the first point is given by the slight reduction in the severity in the loss functions with increasing numbers of images (see Supporting information). The status quo design has a large number of images and has a higher proportion of them in deeper water (Fig. 5). These qual-

ities imply that the locations of the prediction points are likely to be relatively close and that coverage of depth is more even – both are likely contributors to status quo's success. Note that just increasing the number of points will not guarantee a good design, this is the case with the CSG strategy which had a similar number of images to the status quo strategy.

The method for performing the assessment of the different design strategies is enabled by recent advances in statistical computing for geostatistical models (Rue, Martino & Chopin 2009; Lindgren, Rue & Lindström 2011). The approach enables a design to be tested for *all* aspects of the analysis phase (estimation and prediction), not just the prediction step. Traditionally, this has not been possible due to the excessive computational load. Our approach also accounts for parameter uncertainty in the model by simulating data sets from their posterior distribution. This avoids the requirement that data sets are to be generated using the same set of parameter values. The results, from our approach, will be more robust to a wide variety of plausible parameter values (e.g. Müller 1999).

We conclude by providing some concrete guidance to researchers about which design to use in future. If there is little or no prior information about the distribution of environmental covariates, then we feel that researchers should use a structured two-dimensional design (gridded, stratified, or GRTS but not completely random). In these situations, it is advisable to avoid any one-dimensional design or a CSG design. If there is previous information about the physical environment of the study area, then it should be used in the design if possible. The question is how? We believe that it should be done formally using well-defined principles such as randomization and stratification, or model-based design approaches. Without this, a design is constructed by intuition or *ad-hoc* methods (at best), which can be a risky strategy.

## Acknowledgements

## References

Armstrong, R.A., Singh, H., Torres, J., Nemeth, R.S., Can, A., Roman, C., Eustice, R., Riggsd, L. & Garcia-Moliner, G. (2006) Characterizing the deep insular shelf coral reef habitat of the Hind Bank marine conservation district (US Virgin Islands) using the Seabed autonomous underwater vehicle. *Continental Shelf Research*, **26**, 194–205.

Atkinson, A.C. (1996) The Usefulness of Otimum Experimental Designs. *Journal of the Royal Statistical Society – Series B*, **58**, 59–79.

Barrett, N.S., Buxton, C.D. & Edgar, G.J. (2009) Changes in invertebrate and macroalgal populations in Tasmanian marine reserves in the decade following protection. *Journal of Experimental Marine Biology and Ecology*, **370**, 104–119.

Barrett, N., Seiler, J., Anderson, T., Williams, S., Nichol, S. & Hill, N. (2010) Autonomous Underwater Vehicle (AUV) for mapping marine biodiversity in coastal and shelf waters: implications for marine management. pp. 1–6. *OCEANS 2010 IEEE Conference.*

Blangiaro, M., Cameletti, M., Baio, G. & Rue, H. (2013) Spatial and spatio-temporal models with RINLA. *Spatial and Spatio-temporal Epidemiology*, **4**, 33–49.

Bridge, T.C.L., Done, T.J., Friedman, A., Beaman, R.J., Williams, S.B., Pizzaro, O. & Webster, J.M. (2011) Variability in mesophotic coral reef communities along the Great Barrier Reef, Australia. *Marine Ecology Progress Series*, **428**, 63–75.

Buckland, S.T., Anderson, D.R., Burnham, K.P. & Laake, J.L. (1993) *Distance Sampling: Estimating Abundance of Biological Populations*. Chapman and Hall, London.

Cameletti, M., Lindgren, F., Simpson, D. & Rue, H. (2013) Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Advances in Statistical Analysis*, **97**, 109–131.

Chaloner, K. & Verdinelli, I. (1995) Bayesian Experimental Design: A Review. *Statistical Science*, **10**, 273–304.

Clarke, M.E., Tolimieri, N. & Singh, H. (2009) Using the seabed AUV to assess populations of ground fish in untrawlable areas. *The Future of Fisheries Science in North America* (eds R.J. Beamish & B.J. Rothschild), pp. 357–372. Springer, Netherlands.

Diggle, P. & Lophaven, S. (2006) Bayesian Geostatistical Design. *Scandinavian Journal of Statistics*, **33**, 53–64.

Diggle, P.J. & Ribeiro, P.J., Jr. (2007) *Model-based Geostatistics*. Springer, New York.

Foster, S.D. & Bravington, M.V. (2009) Analysis and prediction of faunal distributions from video and multi-beam sonar data using Markov models. *Environmetrics*, **20**, 541–560.

Foster, S.D., Shimadzu, H. & Darnell, R. (2012) *Uncertainty in Spatially Predicted Covariates: Is it Ignorable?* : Series C (Applied Statistics), **61**, 637–652.

Grasmueck, M., Eberli, G.P., Viggiano, D.A., Correa, T., Rathwell, G. & Luo, J. (2006) Autonomous underwater vehicle (AUV) mapping reveals coral mound distribution, morphology, and oceanography in deep water of the Straits of Florida. *Geophysical Research Letters*, **33**, L23616.

Hill, N.A., Pepper, A.R., Puotinen, M.L., Hughes, M.G., Edgar, G.J., Barrett, N.S., Stuart-Smith, R.D. & Leaper, R. (2010) Quantifying wave exposure in shallow temperate reef systems: Applicability of fetch models for predicting algal biodiversity. *Marine Ecology Progress Series*, **417**, 83–95.

Hirzel, A. & Guisan, A. (2002) Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling*, **157**, 331–341.

Hodgson, A., Kelly, N. & Peel, D. (2013) Unmanned Aerial Vehicles (UAVs) for Surveying Marine Fauna: A Dugon Case Study. *PLoS ONE*, **8**, e79556.

Kass, R.E. & Raftery, A.E. (1995) Bayes Factors. *Journal of the American Statistical Association*, **90**, 773–795.

Kim, A. & Eustice, R.M. (2009) Toward AUV survey design for optimal coverage and localization using the Cramer Rao lower bound. pp. 1–7. *Proceedings of the IEEE/MTS OCEANS Conference and Exhibition*.

Lark, R.M. (2002) Optimized spatial sampling of soil for estimation of the variogram by maximum likelihood. *Geoderma*, **105**, 49–80.

Lindgren, F., Rue, H. & Lindström, J. (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society – Series B*, **73**, 423–498.

Lucieer, V., Hill, N.A., Barrett, N.S. & Nichol, S. (2013) Do marine substrates 'look' and 'sound' the same? Supervised classification of multibeam acoustic data using autonomous underwater vehicle images. *Estuarine, Coastal and Shelf Science*, **117**, 94–106.

Müller, W.G. (2007) *Collecting Spatial Data: Optimum Design of Experiments for Random Fields*, 3rd edn. Springer Verlag, Berlin.

Müller, P. (1999) Simulation-based optimal design. *Bayesian Statistics 6: Proceedings of the Sixth Valencia International Meeting* (eds A.P. Bernando, J.O. Berger, A.P. Dawid & F.M. Smith), pp. 459–474. Oxford University Press, Oxford.

Müller, W.G. & Stehlík, M. (2010) Compound optimal spatial designs. *Environmetrics*, **21**, 354–364.

Nichol, S.L., Anderson, T.A., McArthur, M., Heap, A.D., Siwabessy, P.J.W. & Brooke, B.P. (2009) *Southeast Tasmania Temperate Reef Survey*. Post Survey Report Record 2009/43. Geoscience Australia, Canberra.

Robertson, B.L., Brown, J.A., McDonald, T. & Jaksons, P. (2013) BAS: Balanced acceptance sampling of natural resources. *Biometrics*, **69**, 776–784.

Royle, A.J. & Nychka, D. (1998) An algorithm for the construction of spatial coverage designs with implementation in SPLUS. *Computers & Geosciences*, **24**, 479–488.

Rue, H., Martino, S. & Chopin, N. (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society – Series B*, **71**, 319–392.

Schwarz, G. (1978) Estimating the Dimension of a Model. *The Annals of Statistics*, **6**, 461–464.

Sherman, A.D. & Smith, K.L. Jr. (2009) Deep-sea benthic boundary layer communities and food supply: a long-term monitoring strategy. *Deep Sea Research Part II: Topical Studies in Oceanography*, **56**, 1754–1762.

Singh, H., Armstrong, R., Gilbes, F., Eustice, R., Roman, C., Pizarro, O. & Torres, J. (2004) Imaging coral I: imaging coral habitats with the SeaBED AUV. *Subsurface Sensing Technologies and Applications*, **5**, 25–42.

Smale, D.A., Kendrick, G.A., Harvey, E.S., Langlois, T.J., Hovey, R.K., Van Niel, K.P. *et al.* (2012) Regional-scale benthic monitoring for ecosystem-based fisheries management (EBFM) using an autonomous underwater vehicle (AUV). *ICES Journal of Marine Science*, **69**, 1108–1118.

Stevens, D.L., & Olsen, A.R. (2004) Spatially balanced sampling of natural resources. *Journal of the American Statistical Association*, **99**, 262–278.

Stohlgren, T.J., Bull, K.A. & Otsuki, Y. (1998) Comparison of rangeland vegetation sampling techniques in the Central Grasslands. *Journal of Range Management*, **51**, 164–172.

Vermeulen, C., Lejeune, P., Lisein, J., Sawadogo, P. & Bouché, P. 2013) Unmanned Aerial Survey of Elephants. *PLoS ONE*, **8**, e54700.

Warren, J.L., Perez-Heydricha, C. & Yunus, M. (2013) Bayesian spatial design of optimal deep tube well locations in Matlab, Bangladesh. *Environmetrics*, **24**, 377–386.

Williams, S.B., Pizarro, O., Webster, J.M., Beaman, R.J., Mahon, I., Johnson-Roberson, M. & Bridge, T.C.L. (2010) Autonomous underwater vehicle–assisted surveying of drowned reefs on the shelf edge of the Great Barrier Reef, Australia. *Journal of Field Robotics*, **27**, 675–697.

Williams, S.B., Pizarro, O., Jakuba, M.V., Johnson, C.R., Barrett, N., Babcock, R.C. *et al.* (2012) Autonomous Underwater Vehicle monitoring of benthic reference sites. *IEEE Robotics and Automation Magazine*, **19**, 73–84.

Zidek, J.V., Sun, W. & Le, N.D. (2000) Designing and integrating composite networks for monitoring multivariate Gaussian pollution fields. *Applied Statistics*, **11**, 24–49.

Zimmerman, D.L. (2006) Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. *Environmetrics*, **17**, 635–652.

## Supporting Information

Additional Supporting Information may be found in the online version of this article.

**Appendix S1.** Additional results from geostatistical analyses and from simulation studies