



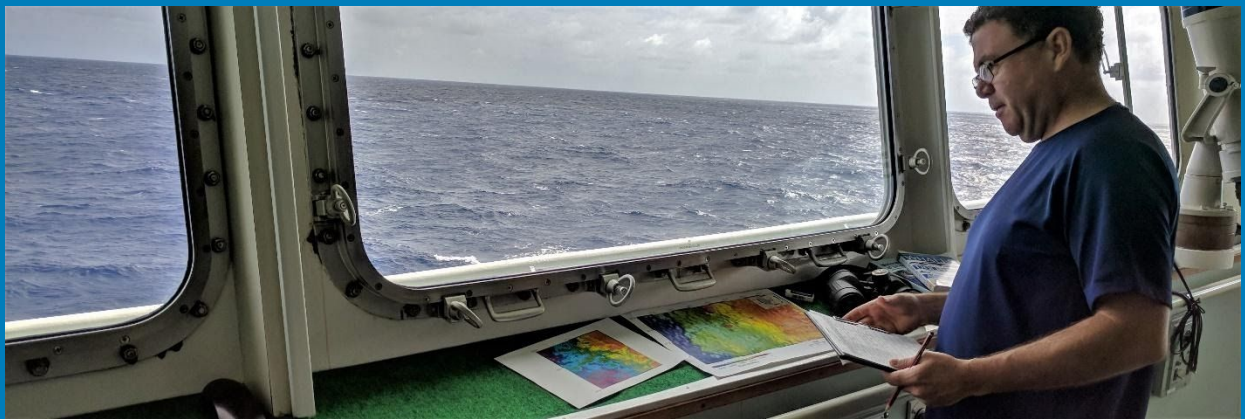
Marine
Biodiversity
Hub

National Environmental Science Programme

2. Statistical Considerations for Monitoring and Sampling

Scott D. Foster*, Jacquomo Monk, Emma Lawrence, Keith R. Hayes, Geoffrey R. Hosack, Tim Langlois, Garnet Hooper & Rachel Przeslawski

* scott.foster@csiro.au



Chapter citation:

Foster SD, Monk J, Lawrence E, Hayes KR, Hosack GR, T. Langlois, Hooper G & Przeslawski R. 2020. Statistical considerations for monitoring and sampling. In *Field Manuals for Marine Sampling to Monitor Australian Waters, Version 2*. Przeslawski R, Foster S (Eds). National Environmental Science Program (NESP).

Introduction

A rigorous scientific process is essential to forming sound conclusions that can inform evidence-based decision-making. This process starts with defining a research question, assessing what level of information is needed and then critically assessing how that information should be obtained (see Table 2.1 and Hayes et al., 2019). Evidence can be obtained from a variety of sources, ranging from expert opinion, through ad-hoc data collection, then well-designed observational surveys, and finally to randomised controlled experiments. Well-designed experiments/surveys that are targeted to the research question are often more expensive than other options, and in certain circumstances (e.g. an inability to randomly allocate sample units to treatment/control groups), may be unavailable. The other sources of information, however, may be adequate depending on the research question and situation (see Leek and Peng, 2015). Table 2.1 for example provides a brief overview of a hierarchy of research questions and the types of data that are appropriate to answer them.

Table 2.1 Different types of research questions (adapted from Leek and Peng, 2015 and Hayes et al 2019)

Research Type	Description	Example Question	Examples of adequate data sources	Complexity
Descriptive associations	Summaries of observed data	What is happening within our <i>sample</i> ?	Expert judgement*, and all forms of controlled/uncontrolled trials, and observational studies with or without a representative sample	Simple
Exploratory	Identify trends and relationships <i>within the sample</i> ¹	What correlates with reef die-back in the sample?		
Inferential	Extending the patterns in the sample to the <i>population</i> from which the sample was taken	What is the status of species X in a marine protected area?		
Predictive	Predict the values at unsampled locations based on sampled data	What assemblage is likely to be found in this location?		

Causal	Identify the reason for a particular association	Are the management actions having an effect?	Expert judgement*, all forms of randomised controlled trials and observational studies with a representative sample so long as the effects of potentially confounding variables can be controlled.	Complex
--------	--	--	--	---------

There is no way to tell if the *sample's* associations are the same as the *population's*

*Expert judgement will likely be influenced by a variety of well known heuristics biases. Attempts should be made to control for these during any elicitation exercise.

Observational studies using data from well-designed surveys (e.g. surveys that ensure samples are representative of the population of interest) are able to answer all types of research questions, and are sometimes the only source of adequate information (Table 2.1). These research questions include those concerning status and trends of biological populations and ecological metrics.

Observational data are generated by scientists observing the current state of the system, whether it be an altered system (e.g. after the establishment of a reserve or an industry) or not (e.g. a baseline survey). The key attribute of an observational study is that there is no attempt to intentionally alter parts of the system for the sole purpose of quantifying effects. Rather, the data is gathered and analysed in a manner that provides information on what the system is like (its status), how it is evolving (its trends) and what may be responsible for these trends (its causative factors). As an example, if baseline/ground-truth surveys are conducted inside and outside previously established no-take marine park boundaries, then they would generate observational data. This chapter discusses the requirements for appropriate statistical design for this observational data.

Causal research questions (attributing observed changes to specific causes) are the most difficult questions to answer. In this case (ideally randomised) controlled experiments are typically recommended, but in this context there are usually limiting factors whose discussion is beyond the scope of this manual (see Hayes et al., 2019). Causal questions require special care and are usually more demanding in terms of the resources needed to answer them. Here we focus on (marine) observational surveys, and in particular the design of surveys. Whilst the topics discussed in this section are relevant to investigating causal relationships, other considerations associated with the analysis of observational data would also be required to be addressed before undertaking causal research (and we do not deal with these issues here). For more information on the evidence hierarchy, and a more thorough description of the different design types for marine ecology, see Hayes et al. (2019).

A key concern in this scientific process is ensuring that survey data are trustworthy and fit-for-purpose (i.e. can answer the research question). To this end, it is important that surveys

and monitoring programs are designed and implemented in such a way that the resulting data are: (i) appropriate for the research question under consideration; (ii) representative of the population under investigation so that (for example) the sample mean is generalisable to the population mean; and (iii) information rich so that uncertainty around inferences is reduced as much as survey budgets will allow. We focus here on survey designs that will help ensure environmental monitoring programs deliver data with these characteristics.

Scope

This chapter will not follow the usual presentation for statistical design in ecology. Rather, we will focus on what we believe to be the most important aspects from a practical (and management) viewpoint. We do not intend it to be like a 'text-book' and explicitly do not include formulae or descriptions of tangential details. Readers will want to look elsewhere for such detail (Urquhart and Kincaid, 1999; Gitzen *et al.*, 2012; Thompson, 2012, are a good start, although there are many). We hope to only introduce the relevant concepts and stress that these are the things that should be thought about by all researchers involved with survey planning. In particular, we discuss: (i) setting the research objectives, (ii) randomisation, (iii) efficiency of design, (iv) uncertainty reduction, (v) sampling in space and time, and (vi) specifics for different gear types. This all leads to an illustrative example design, using the *MBHdesign* R-package. (available from CRAN, <https://cran.r-project.org/package=MBHdesign>). For those readers interested in acronyms: *MBHdesign* stands for *Marine Biodiversity Hub design*. The goals and techniques implemented in *MBHdesign* are outlined throughout this chapter.

Research Objective(s)

The first, and most important factor in developing an appropriate statistical design, is knowing the objective(s) you're asking (your scientific hypotheses) and hence the type of question(s) you are seeking answers to. The objectives will have direct, and sometimes obvious, implications for design. As an example, if the objective is to estimate the effect of implementing a management zone (e.g. a no-take reserve) - a casual type of research question - then at a minimum samples will have to be taken from outside the reserve (a control) as well as within it. We think it useful to consider the following list of probing questions before starting the design process.

- i) What is the primary research question? Is a comparison between areas of interest required (e.g. impacted/not-impacted)?
- ii) What is the appropriate metric to measure (and to subsequently analyse)? Often measurements will be taken on species (e.g. biomass, size, presence-absence, and/or abundance), but analysed as a different quantity (e.g. a diversity index).
- iii) What are the primary sources of potential difference, if any? This will depend on the research question (e.g. impacted/not-impacted areas) but may also include extraneous variables such as environmental conditions and human impact.
- iv) Could locations with different 'treatments' also differ in other important variables, so that there is potential for confounding? An example is whether the habitats within and outside no-take reserves are different.
- v) What resources are available for conducting the survey? Is there a particular type of sampling platform that should be used (see the remaining chapters)? What

previously-collected information is available to aid the current survey (e.g. bathymetry or back-scatter data)? How can we use the previously collected information? Does the previously collected information make one sampling platform or unit of measurement a better candidate than another? How many samples can be taken? This last question also directly affects the *power* of the survey to detect differences in contrasts of interest. See Section 2.5 for more discussion on power.

Random Sampling

In all areas of science (and where statistical methods are applied), representative samples are typically achieved by randomly selecting samples from a wider population (e.g. Thompson, 2012; Smith et al., 2017; Tillé and Wilhelm, 2017). Random sampling ensures that the information contained in the sample is generalisable to the population that it was obtained from (Fisher, 1925). Simply using some sort of random sampling ensures that the data are representative and thereby able to answer many types of research questions (see Table 2.1).

An alternative, which is unfortunately common in marine ecology, is to select sites based on other (non-random) properties. These properties could include their convenience to be sampled, or what a researcher expects to find. This is called 'ad-hoc', 'opportunistic', 'haphazard', 'judgemental', 'purposeful', or 'convenience' sampling. While at first glance this approach appears to be efficient, it in fact diminishes the ability to answer any questions about the population as a whole, which limits questions to those involving the specific sample only: descriptive and exploratory questions (unless non-testable assumptions are made). The reader is referred to Smith et al. (2017) and Dobson et al. (2020) for recent discussions on this topic in ecology.

The implication here is immediate and clear – **researchers should randomise** the sampling process if they expect the patterns observed in the sample to hold in the population. Researchers should not routinely perform haphazard sampling. Of course, there may be situations where a particular location appears so interesting that it could be *appended* to a randomised survey design, but its data can only be included into the analysis with additional (strong) assumptions and/or complexities in analysis approaches. The randomisation process is particularly important for monitoring programs where data from multiple surveys (through time and/or space) are combined.

An important side-effect of randomisation is that a researcher must specify what the statistical population under study is. Formally, for surveying geographic areas, the population is a collection of potential survey locations from which a random sample is taken, often called a *sample frame* in the literature. The formal specification of the sample frame is important as it gives the extent to which the results are legitimately generalisable. A sample frame may be delimited by some combination of: spatial extent, depth, habitat type, season and the type of sample that the selected gear can adequately collect. Generalisation beyond the sample frame requires assumptions, often quite strong assumptions, that the processes outside the sample frame are identical to those within it. It is best to try and avoid these assumptions by expanding the sample frame prior to undertaking the survey.

Efficient Designs

Simple randomisation – randomly scattering sampling locations through space – is not necessarily an efficient approach, and in many circumstances a large number of samples are necessary to obtain acceptably precise estimates of population parameters (e.g. Tillé and Wilhelm, 2017). This potential inefficiency is one of the reasons that haphazard sampling can initially although mistakenly appear quite attractive. There are, however, ways to address inefficiency, and to generate designs that require fewer samples and resources. Various researchers have proposed statistically valid restrictions on the randomisation process. In the environmental sciences this discussion has ultimately led to several forms of spatially balanced designs (Stevens and Olsen, 2004; Dobbie et al., 2008; Grafström et al., 2012; Grafström, 2012; Grafström and Tillé, 2013; Grafström, 2013; Robertson et al., 2013; Brown et al., 2015; Foster et al., 2017; Tillé and Wilhelm, 2017), with similar ideas known as ‘spatial coverage designs’ (Royle and Nychka, 1998; Brus et al., 1999, 2006; Minasny and McBratney, 2006; Walvoort et al., 2010) and ‘even sampling designs’ (Chen et al., 2012).

A spatially balanced design can be seen as an extreme form of stratification (Stevens and Olsen, 2004) that aims to reduce the frequency of placing samples close to each other (relative to simple randomisations). This process improves efficiency by reducing the amount of spatial auto-correlation between data implying that each sample is providing as much unique information as possible (Grafström and Tillé, 2013). Additionally, spatially balanced designs are more efficient than other types of randomised designs as they tend to increase balance on many environmental variables (also known as covariates), where the population’s covariate mean is equal to the sample’s covariate mean (Grafström, 2013). This is more than just stratifying for important environmental gradients, as that process does not ensure balance unless explicitly accounted for. Even if balance is sought in stratification, the simple randomisation process within strata lacks efficiency, can complicate analyses, and can be wasteful of ‘degrees of freedom’ in the analysis (reducing analytical power – where relevant). In summary, spatially balanced designs are used to enhance efficiency so that the greatest amount of information is obtained from any number of sample locations (compared to other forms of randomisation).

This type of efficiency is not the only consideration though - logistical considerations often impose practical constraints. Take for example baited remote underwater video (BRUV) surveys where there are often multiple BRUV cameras, each of which must undergo a ‘soak time’ before the data can be collected. In this case, it is inefficient to sit at a single station during the soak time of each and every BRUV deployment (not sampling and not travelling to other sample locations). Instead, it may be better to make multiple deployments in a spatial ‘cluster’ with all the available BRUVs sampling simultaneously. Similar arguments can be made for sampling gear that takes considerable time to deploy and retrieve (e.g. SCUBA), where multiple transects can be swum in a single dive. This type of design is known as a cluster design in the sampling literature (Thomson 2012) and has been successfully used for marine sampling (e.g. Lawrence et al., 2015 and Hill et al., 2018). The location of the clusters can still be spatially balanced (Lawrence et al., 2015), which gives the spatial balance over the survey area. We suggest trying to make the number of clusters as large as possible, especially if there is a trade-off between the number of clusters and the size of those clusters. In some situations, like the BRUV

example, the number within each cluster may be naturally specified by the number of BRUV cameras that are available or can be safely stored on the vessel. All design decisions have implications for analysis, and using nested designs is no exception. When analysing nested data, there should be some accounting for within-cluster correlation. This can be achieved using a cluster random-effect, or by using a geostatistical model (e.g. Diggle, P. & Ribeiro, 2007 and Banerjee, et al. 2004).

Some researchers will know spatially balanced designs as 'GRTS' (for generalized random tessellation stratified; Stevens and Olsen, 2004), but GRTS is just one type of spatially balanced design. It is a good design approach and it is the prime reason that spatially balanced designs are gaining popularity. However, it is not the most spatially balanced design, which implies that it is also not the most efficient (Grafström et al., 2012; Robertson et al., 2013; Foster et al., 2017). Between the various spatially-balanced design types, the differences in relative performance are minor. Computational methods for GRTS, via the *spsurvey* R-package (Kincaid and Olsen, 2016), in our experience can be cumbersome, time-consuming and in some ways inflexible. Experienced GRTS users can legitimately continue using it, as the efficiency cost is not large, and they have already overcome many of the more cumbersome aspects. However, we recommend that new users start with *MBHdesign*.

While we focus here on spatial balance, many (but not all) of the algorithms for producing spatial balance can be employed to sampling situations that involve more than just 2-dimensional space. In particular, the algorithms implemented in *MBHdesign* are equally applicable to space-time scenarios and even space-depth-time ones (where a 3-dimensional volume, such as a water mass, is sampled over time). In fact, the algorithms scale well with dimensions, and there is no limiting dimensionality, except what is practical in the application.

The efficiencies of spatially balanced designs can be further improved by increasing the probability of selecting sites (sampling locations) where the sampling variable is thought to have greater variance (e.g. Godambe and Joshi, 1965; Brewer et al., 1988; Chambers, 2011; Grafström and Tillé, 2013). Here, we use the term 'site' to mean the location where a single deployment is made. We note that others may have slightly different definitions. This is achieved by altering the so-called inclusion probabilities of each potential site. Inclusion probabilities specify the chance of each site being randomly chosen to be part of the survey and they can be chosen on the basis of data from a pilot study or from other sources (e.g. literature on similar species and/or regions). A very low inclusion probability (near zero) will imply that the site will almost never be sampled, whereas a site with very high inclusion probability will be chosen much more often. The inclusion probabilities are prescribed by the survey designer to indicate where the sampling effort should be placed (see Grafström and Tillé, 2013, for more information on how to perform this task).

In ecology, where univariate biological variables often have an increasing mean-variance relationship (e.g. through Taylor's power law; Taylor, 1961), this equates to increasing inclusion probabilities in locations where the variable being sampled is expected to have high abundance (noting again that this is often the motivation for judgemental samples but here we embed auxiliary information on abundance within a strictly probabilistic framework). If no prior knowledge exists about the variable under study, which may have been obtained from previous surveys or a pilot study, then the inclusion probabilities should be equal.

Special consideration is required in situations where there are multiple outcome variables to be measured, such as the sampling of multiple different community types or multiple species. In these cases, the inclusion probabilities should reflect 'combined usefulness'. For sampling multiple communities/species this means that each community/species should be effectively sampled and that the combined inclusion probabilities should reflect this. Whilst the 'combined usefulness' concept is vague, it should reflect the combined utility of each sampling site to each component of the multivariate observation. In situations where the multivariate observations are independent or even negatively correlated, such as communities occupying different habitats might be, then inclusion probabilities may be increased for each different habitat. The nett result of this process may be an inclusion probability surface that is quite even and so equal inclusion probabilities may be a good default..

Altering inclusion probabilities requires the identification of one or more measured covariates (available at time of design) that can be used to guide the variation in inclusion probabilities. It is beneficial only in situations where the inclusion probabilities are related to the sampling variable. When inclusion probabilities *do not* have this relationship, then this will cause a *loss* of efficiency (lower precision) than equal inclusion probabilities. We caution against using too many covariates in the design stage and point out that equal inclusion probabilities is a conservative and usually adequate approach. In fact, fewer covariates is better in many ways. The simple reason is that if they are used to define the design then they must also be used in the analysis (as the design is conditional on these covariates), see Gelman et al. (2013) and Foster et al. (2017) for discussion. This means that precious 'degrees of freedom' must then be used to estimate potentially non-helpful parameters, which has the effect of increasing analysis complexity and reducing the discrimination ability of the analysis. So, the survey designer must weigh up the anticipated reduction in variation due to incorporating the covariate against the necessity to use more terms in the model. When there are multiple sample variables of interest, altering the inclusion probabilities should be considered carefully as altering the probabilities to reduce the variation in one variable may be at the expense of others.

The concepts of stratification and altered inclusion probabilities are almost, but not quite, identical in situations where stratification is applicable. However, at the cost of being conceptually more sophisticated, the inclusion probability concept is more general and more flexible. The reasoning for the equivalence is that the inclusion probabilities can be designed to match the stratification, so that *on average* the specified number of survey sites is chosen within each strata, but this is not guaranteed for every randomised design. Contrastingly, all stratified designs will have the specified number of survey sites within each strata. To us, this is not a large difference and the ability to spatially balance the design is likely to lead to bigger benefits. We therefore recommend altering inclusion probabilities with spatial-balance in preference to formal stratification. However, stratification is not a bad option and is more efficient than simple randomisation (when the stratification is meaningful). We note that the *spdesign* software that implements GRTS allows for stratification *and* spatial balance by balancing within each spatially-contiguous strata.

When planning marine monitoring programs, the ability to incorporate any existing sites will often be advantageous, especially when those sites are part of a random sample. An example is when certain sites are mandatorily sampled to achieve regulatory compliance or where sites

must be sampled in the future to demonstrate compliance. In the NESP Marine Biodiversity Hub, methodology was developed to incorporate these *legacy* sites into a spatially balanced design. Legacy sites (or historical, reference or sentinel sites) are those sites that have been sampled in the past and the researcher wants to re-visit them as part of the upcoming survey for comparability, or sites that must be sampled in the future, for example to quantify the effects of decommissioning oil and gas platforms in-situ. Readers are referred to Foster et al. (2017) for details. Briefly however, spatial-balance is achieved by setting the inclusion probability of legacy sites to one and adjusting inclusion probabilities (within the proximity of legacy sites) downwards so that new samples are less likely to be placed very near them.

Software

There are many pieces of software that will generate spatially-balanced designs, most of which are based on different algorithms. For monitoring the marine environment, we developed a specific software – the R-package *MBHdesign*. It is intended to be easy to use and tailored to common situations in marine ecology. It also has the ability to make designs spatially balanced around existing legacy sites, see Foster et al. (2017), and also for designing surveys with sampling platforms that are transect-based, see Foster et al (2019). We will use *MBHdesign* in the example to follow.

Uncertainty, Precision, and Power

It is important to consider how to reduce the uncertainty (and increase precision) in statistical analyses of survey/monitoring data. Practically, there are two components to this: 1) increasing the information content in the dataset; and 2) reducing non-relevant variation (noise) during the collection process. Addressing the first component, increasing information content, begins by using an efficient survey/monitoring design, such as a spatially balanced design. More information implies that the signal in the data can be clarified with greater ease. The second component (noise reduction) refers to that part of the variation in the data that is induced by (for example) performing assays slightly differently each time. This includes the measurement tools used (e.g. CATAMI for image classification; Althaus et al. 2015). This type of noise can be reduced by adhering to well-defined, repeatable, measurement protocols and classification schemes so that two or more measurements on the same sample will generate identical or at least very similar observations. See the gear-specific chapters in this field manual package for detailed advice on reducing measurement noise. As an example, consider the scoring of an AUV mission. Scoring more images from the mission (less subsetting) and using a higher density of points within each image will reduce measurement variation. There is diminishing returns though, with more points becoming less and less better (Perkins et al.; 2016). What constitutes irrelevant noise depends on the objectives of the study, but additional sources of noise can be subtle and can encompass issues such as taxonomic inconsistency and inclusion of non-target species or individuals of a wide variety of life-stages. The latter may occur from sampling pelagic species in transit to the targeted habitat.

For some novel measurement platforms, measurement/scoring techniques are still being assessed and these updates should be incorporated where possible. Examples of this process

are Perkins et al. (2016) for scoring AUV images and Schobernd et al. (2014) for scoring BRUV deployments. We stress though, that whilst noise reduction is important, it is not the only consideration and that particular care should be taken to maintain protocols within already established monitoring programs, or calibrate new protocols with old. In addition to reducing 'noise', it will ensure that, for example, time-series do not get 'broken' and that data are directly comparable in time and space without unfortunate confounding due to a change in sampling methodology or other factors discussed above.

Most Chapters in this field manual package are variations on the noise-reduction theme as they provide a foundation for reducing variation between and within surveys. In particular, if adhered to, they will help minimise, or possibly even eliminate, inherent systematic variation (bias) between different surveys or within a monitoring program. This will have the effect of increasing the utility of combining data from different surveys (as there will be minimised bias between the two sets). We have unfortunately come across long-term studies that could not be used to estimate trends in the target species because of inconsistencies in sampling design and implementation (Hosack and Lawrence, 2013).

Any approach to reducing variance in the sample statistics should be welcomed whole-heartedly, so long as there is no introduction of confounding between it and any spatial/temporal signals or other important trends. This includes processes to eliminate obvious sources of measurement variation (e.g. non-uniform gear deployment, faulty measurement equipment, poor laboratory practices) and data entry errors. In well conducted studies, under most circumstances, measurement variation is likely to be relatively small compared to the variation in the ecological processes that are being sampled. Understanding this means that exorbitant amounts of time should not be placed in perfecting each measurement – especially not if the cost of perfection is a substantial reduction in the number of samples taken. Often a much richer sample is obtained (in terms of signal to noise) by taking more, slightly noisier, samples than fewer precise ones. Unfortunately, we are aware of no rules-of-thumb to guide researchers with this issue.

In certain situations, it may be pragmatic to alter the measurement process. For example, in the rare situations where the cost of performing the measurement assay is large compared to the cost of collecting the ecological material, then it *may* be useful to combine a number of samples prior to measurement. However, it is important to realise what is being lost in this case: the ability to understand the variability between samples within the same assay group. This is often a limitation if the combined samples originated in quite different environmental conditions (for example). It should also be noted that in analyses, the combined sample is the sampling unit, not the original ones that contribute to it. Another situation where measurement variation can be reduced is when the assay is both cheap and noisy. In this situation, it can be beneficial to use sub-samples (sometimes confusingly called replicates), which are measurements on the *same* biological material. When sub-samples are utilised, the analyst can partition variation into measurement and ecological components.

Some design experts advise that a power analysis be performed before any survey effort is undertaken. Recall that a power analysis calculates the probability that the survey will be able to detect a difference if there actually is one (a true positive). This is undoubtedly a good thing to

do when there is a clear hypothesis to be tested, a meaningful effect size can be stipulated and detected, and an (at least moderately) accurate estimate of variance is available prior to the study. However, this is not always the case. It has been observed that power analyses are often performed without great thought, leading to (perhaps) overly large stipulated sample sizes (e.g. Mapstone, 1995); probably larger than any reasonable budget will allow. The arguments outlined in Mapstone (1995) are, to us, quite compelling as they make a researcher undertaking a power analysis think critically about the relative environmental/economic/political costs of making a poor decision. Sometimes it will be more important to guard against making a false-negative (type II) errors than false-positive (type I). Such a situation could occur if the cost of falsely declaring significance is larger than that of falsely declaring *non*-significance (e.g. declaring impact may result in closure of a factory or imposing fishing quotas). This is quite contrary to many applications of hypothesis testing in other areas of science.

If a power analysis is undertaken, then there is some general advice that we offer to marine ecologists. First, don't blindly follow text-book recipes for power analyses. They make some strong assumptions that are unlikely to be met in ecology (e.g. normality of observations, independence of observations, and constancy of variance in space and/or time). Second, be prepared to do a lot of homework about the sizes of the components of variation that you are likely to observe: "How much overdispersion is there in your study region?" "Is there any spatial autocorrelation likely?" "What analysis methods are intended to be used?"

It is our opinion that a very useful, and often not too difficult, method for assessing power is to use simulation. There have recently been attempts to provide simplified R-based tools for this process (Green and MacLeod, 2016, for mixed models), and these show promise. The simulation approach consists of a small number of steps: 1) simulate some data under the alternative hypothesis (incorporating the effect that is being considered), 2) analyse the data and see if there is a significant effect, and 3) repeat steps 1) and 2) many times. The proportion of analyses (of simulated data) that produce a significant analysis will give one minus the power of the test. The simulation approach is not without detractors though, and many of these are shared with all power-analyses. Primarily, the simulation model describes a simplified version of reality, which is likely to be less 'noisy'. The reduction in noise stems from unaccounted for events, such as storms, unusual recruitment events, and so on. Irrespectively, power analyses are widely used and the simulation approach has been used in many places, including the marine realm (Foster et al., 2014, Perkins et al., 2017). Power is not the only piece of information that can come from the simulation though. In particular, it can be used to support the evaluation of how sample size and study design impacts more general monitoring objectives (e.g., the ability to estimate parameters in a model or predict future data).

Spatio-Temporal Sampling

Sampling in space is a task that requires plenty of thought, as demonstrated by the previous sections. Sampling in space *and* time (i.e. monitoring) requires even more thought as there are

even more options. Generally, if one wants to sample repeatedly then the focus will be (at least partly) on trends through time. It is commonly established in the survey literature, that the uncertainty around temporal signals is reduced by repeatedly visiting the same sites (e.g. Urquhart and Kincaid, 1999). This comes at a cost though – less sites are sampled and therefore the sample may not be as representative of the population as it could be. Extreme cases in marine sampling are when the sampling gear actively alters the population size (through extractive sampling) or its habitat (for example removal of epibenthic structure). In these cases, repeatedly sampling the same sites will not reflect the trends in the population.

Intuition tells us that, unless sampling is destructive, then you should revisit at *least some* of the sites. This is due to the reduction in variation in the temporal signal (the site-to-site variability is removed). The proportion of sites to be revisited, and the pattern of revisits (e.g. rotating panel, fixed panel, and so on – see McDonald, 2003), will depend upon the temporal (and spatial) variability of the biota under consideration (see Perkins et al., 2017, and references therein). Legacy sites, which come from a previous randomised survey, can, and should, be incorporated into a temporal monitoring program. Our advice is to try and make sure that some legacy sites get sampled during each revisit for new sites. This has the effect of ensuring 'a link back to the legacy site time-series' for each revisit. If the biota change randomly/unpredictably and rapidly, even at the same spatial location, then there is little point revisiting sites. This is especially so for monitoring programs with substantial time between revisits. In summary, think carefully about the relative importance of the temporal signal versus the generality. This will reflect the number of revisits to perform.

Gear-Specific Considerations

Some gear types need special consideration as they naturally force the survey designer into different modes of thinking and have limitations on the type of data collected. To our mind, the biggest distinction in sampling gears for marine biota, for design considerations at least, is whether the gear collects a single observation point from each deployment (e.g. a grab) or whether it collects many (e.g. an AUV). There is some grey area here: we usually class BRUVs as point collection methods and under certain circumstances (outlined below) we might also class trawls also as point source methods. BRUVs introduce uncertainty in the exact area sampled due to variation in bait plumes with different current conditions. Trawls integrate spatially contiguous locations along a transect by means of combining the catch in a cod-end (e.g. Foster et al., 2019). Trawls also introduce uncertainty in the exact area sampled due to the behaviour of the trawl ('digging in' or bouncing off the sea bed) and sometimes not knowing where the trawl has touched/left bottom or where the net is. These can be particular concerns in deeper water without a well-functioning positioning device (USBL).

When the spatial scale of the sample-frame is geographically large, in relation to the transect size (e.g. AUV or trawl) or field-of-view (e.g. BRUV), then all these methods can be treated as point collection and standard survey design principles apply. However, when the sample frame is geographically small in relation to the size of the area sampled by the sampling gear, then the position of the observation within the sampling unit becomes important as biota from two separate samples may be spatially close. The only design advice in the literature for the gear

types considered in this field manual package, that we are aware of, is to try and space samples well apart in space within the objectives of the study (Foster et al., 2014). However, recent Marine Biodiversity Hub research aims to provide greater utility around this (see Foster et al. 2019). Developed methods are implemented into the R-package *MBHdesign*.

There are more considerations when designing a transect-based survey when the transect footprint is large relative to the sample frame (e.g. AUV, ROV and towed video). Chiefly, one needs to consider how long the transects are and in what direction the transects should be performed. Our intuition tells us that, logistics aside, the length of the transect should be dependent on the spatial properties of the biota being surveyed. Biota with large spatial autocorrelation should be sampled with many short transects, whereas biota with short spatial autocorrelation could be sampled with longer transects. See Foster et al. (2014) for an example of identifying length and direction of spatial autocorrelation from image-based transect data and see Foster et al. (2019) for how to randomise, with spatial balance, transect samples. Of course, it may be cheaper to deploy the image-based sampling platform for longer and then simply sub-sample or account for the autocorrelation within an analysis model, but the reasoning will still provide advantages. In any situation, care needs to be taken in the analysis to account for this autocorrelation (see next paragraph for further elaboration). The direction of the transects might be gear dependent – for example it may be ‘safer’ to take transects down-slope or across-slope; or more efficient to tow into the prevailing current. However, irrespective of the restrictions on direction the design should aim to cover the study area as evenly as possible. Image-based transects have further considerations – how much effort to place in scoring each image versus how much effort to place in scoring more images. Perkins et al. (2016) suggests that this too depends on the spatial properties of the biota under consideration and suggests apportioning effort according to these properties.

When designing temporal surveys, it is important to consider if you can actually perform replicates with enough geographical accuracy to be useful. If the exact transect cannot be repeated then there is a confounding of temporal and spatial variation, and if the spatial patterns are quickly changing then the temporal uncertainty will also be inflated (Perkins et al., 2017). This is particularly concerning for gear types that are located only by the location of the deployment vessel. Even for accurately re-deployable gears the spatial repeatability is sometimes not sufficient to adequately determine within-site heterogeneity and detect a prescribed temporal change (Perkins et al., 2017).

Whilst this chapter is about statistical *design*, we feel it important to briefly mention statistical *analysis* of survey data, especially that resulting from transect-based sampling platforms. These produce data that are spatially close to each other, often very close. This naturally raises concerns about spatial autocorrelation and its impact on an analysis. Our advice for these platforms is to use geostatistical models (e.g. Diggle and Ribeiro, 2007; Banerjee et al., 2004). These naturally account for the spatial dependence between observations and adjust measures of uncertainty accordingly. This is not an easy approach and involves a steep learning curve for many practitioners. However, it does circumvent the unfortunate (and dangerous) consequence of falsely considering that there is less uncertainty in the data than there actually is, which is effectively what happens when one assumes that geographically close observations are independent. Subsetting the individual observations within a transect is likely to have some

beneficial effect on mitigating autocorrelation (e.g. Mitchell et al., 2017). However, doing so presupposes that the range of the autocorrelation is less than the distance between the subsetting observations.

Mapping as a Foundation

Seafloor data from multibeam sonar and sidescan sonar can cover an entire area (sampling frame) and is a real boon for efficient survey designs. In particular, multibeam data can facilitate the production of base maps covering tens or hundreds of square kilometres, with accurate geo-location. Multibeam data enables the survey design team to produce a design that picks out the major sources of variation in the ecosystem (typically depth and hard substrate), which can then be used to alter inclusion probabilities. To use these data one must consider how the multibeam data might be related to the variance in the target biota being sampled; -- under certain circumstances it is reasonable to spend greater survey effort on hard substrate to reduce variation in parameter estimates. For example, sponge abundance will have higher variance on hard bottom than on soft bottom and so a sponge survey should disproportionately target hard bottom. Once these areas have been (accurately!) identified, then the inclusion probabilities for those regions can be increased, which will increase the chance of sampling hard substrate but maintaining the ability to infer to the sampling frame. This is the intuition in the approach that was used in Lawrence et al. (2015).

Although our recommendation is to map the survey area using multibeam prior to designing biological surveys, it is not always possible. One alternative approach, which tries to leverage as much multibeam information as possible, is to stage the sampling: perform a limited amount of multibeam mapping and work within those limited areas. Done smartly, like in Lawrence et al. (2015) this approach can still offer good estimates of biota. However, it is not without difficulties (principally in the analysis stage) and these complications could be, in some cases, overly limiting. Another possible alternative is to predict the presence and location of important habitat features, such as reefs, using statistical models (!Ref) and use the resulting habitat map to guide inclusion probabilities. We would caution, however, that such an approach will only be advantageous with a well-validated model.

Case Study: Surveying a Marine Park in Tasmania

To illustrate some of the technical aspects of the design process, we plan a baseline survey design for the Governor Island Marine Reserve off Bicheno on the East Coast of Tasmania to establish the status of biota. There are no zones of interest in this example, as the survey is not designed to test a management action (or other source of variation). The marine reserve is geographically complex with boundaries governed by natural land formations. The depth profile of the reserve is decreasing away from the land-based boundary, and there are less 'shallow' regions in the reserve than 'deep' ones.

We will present three designs. The first is a plain (vanilla) spatial design where all sites within the reserve are equally likely to be sampled. The second design intentionally samples shallow

sites more often on the assumption that these sites are likely to be more heterogeneous and diverse than their deeper water counterparts. The third type of design is when there are legacy (reference) sites in the area that should be resampled as it is considered important to create a time-series for this reserve. The spatial balance should then account for the locations of these legacy sites when finding the new sites. For more details on how to perform this third type of design, please see the *MBHdesign* vignette (by loading the *MBHdesign* package into R and typing vignette 'MBHdesign'). Another good place to look is the paper describing the method: Foster et al. (2017). The inclusion of legacy sites in this example is somewhat artificial, as we have to first choose the legacy sites to incorporate. However, we hope that the process is illustrative, and stress that it mimics the consequences of real sampling with legacy sites – first the legacy sites are chosen randomly and then the new sites are chosen randomly around them (with spatial balance).

The example here is performed in R, an open source statistical platform. Importantly, there are other free and licensed software and programming languages that can also be used, depending on your proficiency and what is available to you. Some of the code may, at first glance, look a little daunting. Well, that's R for you. Most of the lines written here are for plotting purposes and for reading in data. Since this is a document, we have taken some care with how the plots appear. This produces pretty(er) pictures but it also produces longer and more detailed code. Users should feel free to use the code below as a template, but please don't blindly do so without thinking if the actions of the code are appropriate for your data. If you do re-use code, then please run checks to see if the code has done what you think that it ought to.

If you are new to R, then you could try to get an introduction by one of the many online tutorials (e.g. <https://cran.r-project.org/doc/manuals/R-intro.html>). That particular one is likely to be like R helpfiles (helpful but takes time) and it could be quite *dense*. Another option is the excellent book Venables and Ripley (2002), which introduces you to R *and* gives a good introduction to some types of data analysis. Other recommended introductions to R include: Crawley (2007); de Vries and Meys (2015). However, these are just suggestions, you should shop-around until you find a reference/tutorial that is at-your-level and in no time at all you will be reading in data, analysing it, plotting it, and summarising results.

Set Up R to Generate Design

To start we have to set up R for generating designs. This should not be onerous in this case. The most difficult thing is in setting up the data file in the first instance (usually through a GIS). Here we have used an asc file as this is relatively easy to read into R. This file is included in the field manual package, along with the R code to create the output below.

This document was created using the R-package *knitr* (Xie, 2014). It is a wonderful tool, but like any tool it requires interpretation. Most notable here is that the R-code is placed in a grey box, to enable readers to highlight the code versus the document text. Within the code sections, anything that comes after a '#' symbol is a comment that is not interpreted by R (most of these are a brown colour). Bold dark blue words are function names. Dark blue words are argument names. Green is for text and light blue for numbers.

```
#####
####   Read in Data from spatial data (.asc here) and Organise   ####
####   Foster et al. NESP Biodiversity Hub Field Manuals         ####
#####
```

```
##if you don't have MBHdesign installed, please do so using
# install.packages( "MBHdesign")
```

```
#Load required packages
library(MBHdesign) #For spatial sampling
library(fields)   #for lots of things, but for plotting in this example
library(sp)       #for reading the ascii file of cropped depths for the reserve
```

```
#Set a seed for reproducibility
set.seed(666)
```

```
#Read in depth as a asc file containing long, lat and depth
#This path/file only exists on the first author's system
# you will need to change it if running this code
#the projection will need to be changed for each region too
#bth.orig.grid <- read.asciigrid("./ExampleGovIsland/gov_bth.asc",
proj4string = CRS("+proj=utm +zone=55 +datum=WGS84"))
bth.orig.grid <- read.asciigrid("gov_bth.asc", proj4string = CRS("+proj=utm
+zone=55 +datum=WGS84"))
```

```
#convert to a data.frame for ease
DepthMat <- as.matrix( bth.orig.grid)
bth.orig.grid <- as.data.frame(
  cbind( coordinates( bth.orig.grid), as.numeric( DepthMat)))
colnames( bth.orig.grid) <- c("Easting", "Northing", "Depth")
bth.orig.grid <- bth.orig.grid[order( bth.orig.grid$Northing,
  bth.orig.grid$Easting),]
#Setting up plotting for now and later
uniqueEast <- unique( bth.orig.grid$Easting)
uniqueNorth <- unique( bth.orig.grid$Northing)
ELims <- range( na.exclude( bth.orig.grid)$Easting)
NLims <- range( na.exclude( bth.orig.grid)$Northing)
#Fix up ordering issue
DepthMat <- DepthMat[,rev(1:ncol(DepthMat))]
#plot it to see what we are dealing with.
image.plot( uniqueEast, uniqueNorth, DepthMat,
  xlab="Easting", ylab="Northing", main="Governor Island Reserve",
  legend.lab="Depth (m)", asp=1, ylim=NLims, xlim=ELims,
  col=rev(tim.colors()))
```

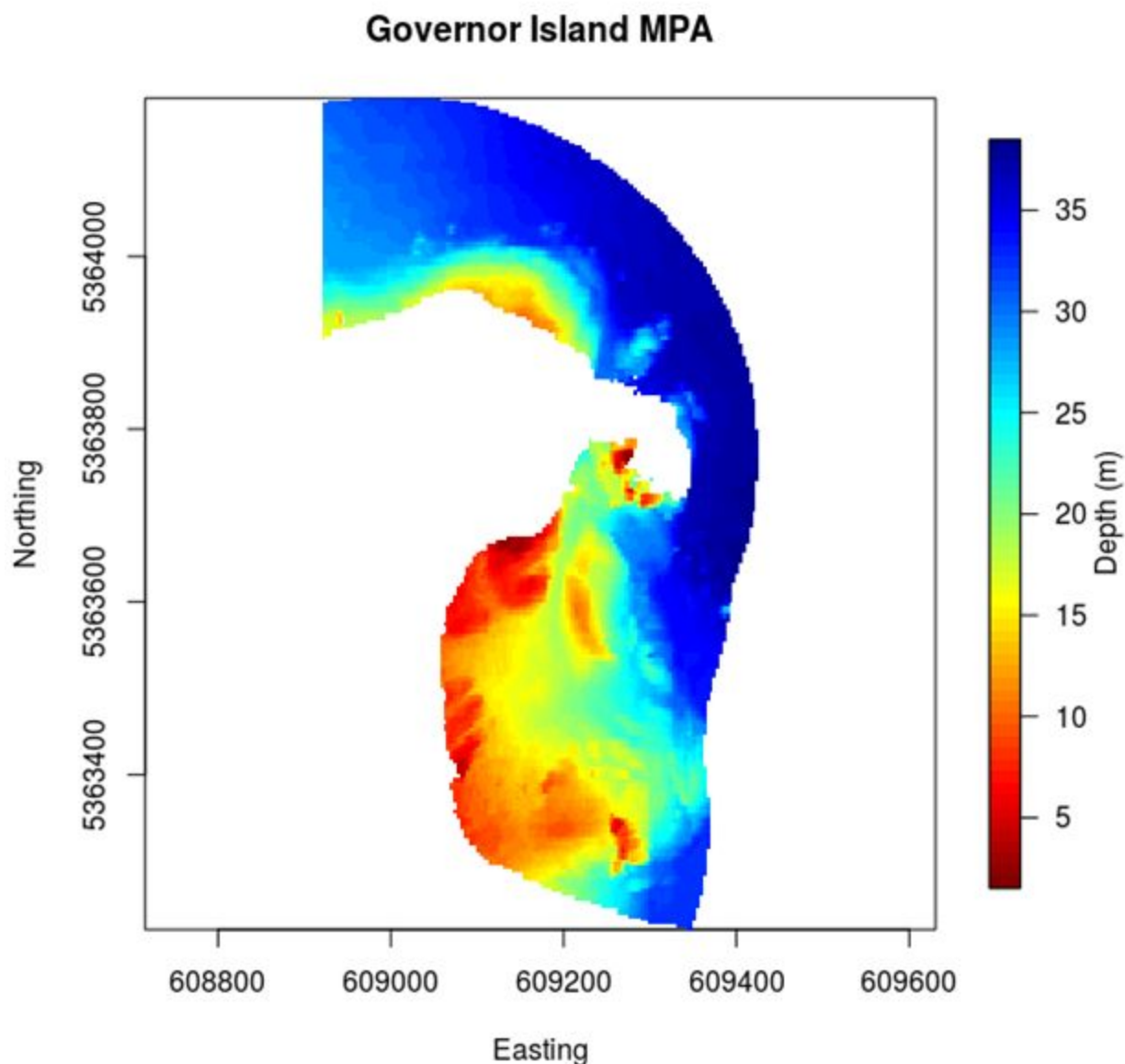


Figure 2.1: Map of Governor Island study region with depths. Note the non-regular shape and the non-uniformity of the regions depth profile.

Generate a spatially balanced design

Generating a spatially balanced design within the reserve is quite straight-forward using *MBHdesign*. Here we do it for 30 sampling sites spread throughout the reserve (Figure 2.1). Note that designs will vary from one realisation to the next, unless the random number generating seed is fixed (like we did in the previous subsection). Try it a few times, if you like, and see what happens between the realisations. Note that *on average* (over all realisations) the spatially balanced designs will have good spatial coverage.

```
#####  
####   Spatially balanced design -- uniform inclusion probs   ####  
####   Foster et al. NESP Biodiversity Hub Field Manuals      ####  
#####
```

```
#number of samples  
n <- 30  
#take the sample  
samp_spatialOnly <- quasiSamp( n=n, dimension=2,  
  potential.sites = bth.orig.grid[,c("Easting","Northing")],  
  inclusion.probs=!is.na( bth.orig.grid$Depth))  
with( bth.orig.grid, image.plot( uniqueEast, uniqueNorth, DepthMat,  
  xlab="Easting", ylab="Northing", main="Spatially Balanced Sample",  
  legend.lab="Depth (m)", asp=1, ylim=NLims, xlim=ELims,  
  col=rev(tim.colors())))  
points( samp_spatialOnly[,c("Easting","Northing")], pch=20, cex=2)  
write.csv(samp_spatialOnly, file="spatialOnly.csv", row.names=FALSE)
```

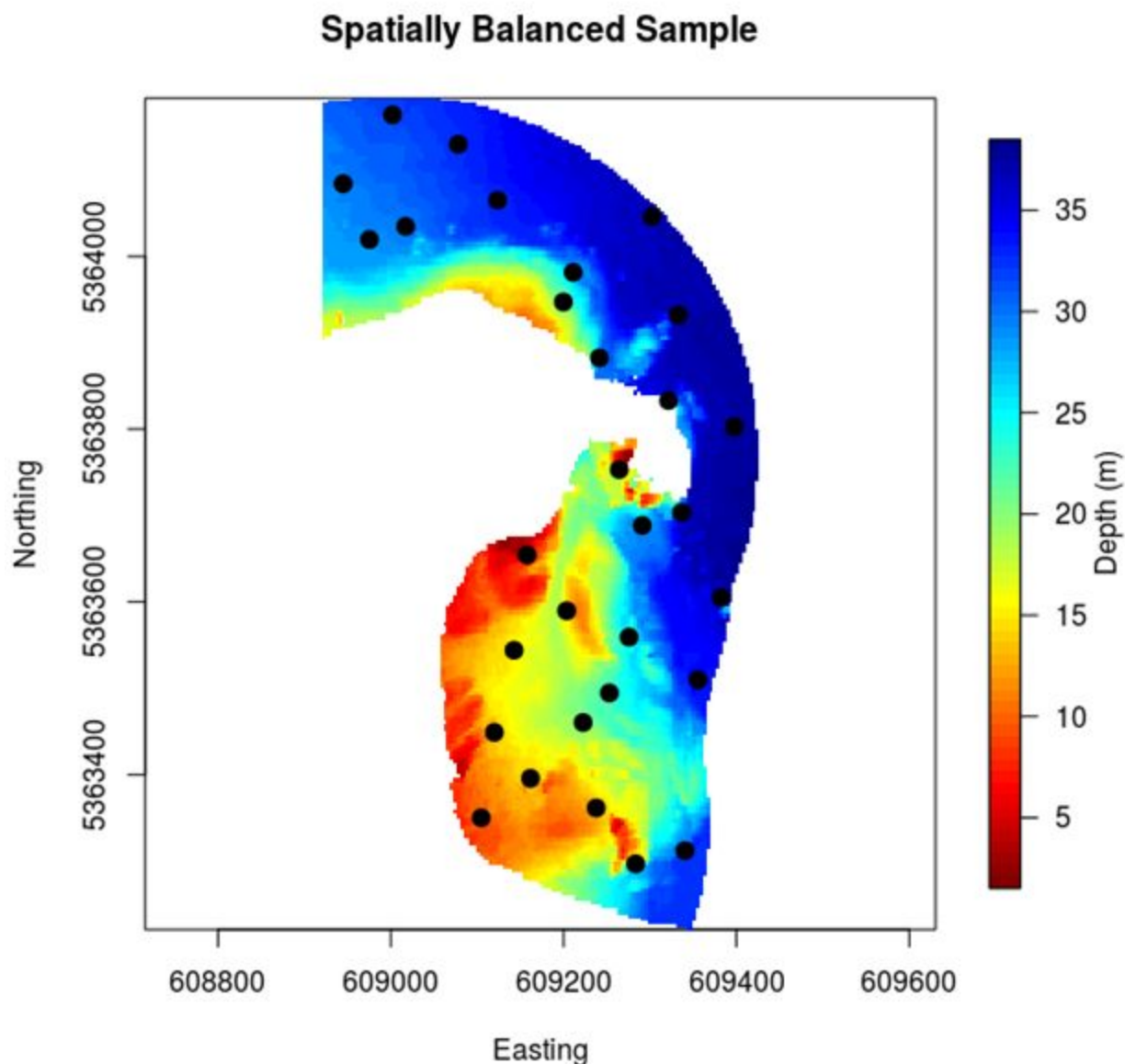



Figure 2.2: A uniform inclusion probability sample for Governor Island

Preference shallow environments

The equal inclusion probability design (Figure 2.2) assumes that all sites are equally advantageous to sample. Previously, we mentioned that this may not be an efficient approach to sampling. In particular, it can be advantageous to over-sample sites/regions that have greater variability. In the Governor Island reserve, this corresponds to the shallower depths as these typically are more heterogeneous and biodiverse on the east coast of Tasmania. We can design a survey with this in mind by increasing the probability that shallow sites will be sampled (i.e. by increasing their inclusion probabilities). This has the obvious effect of also decreasing the

probability that deeper sites will be sampled (Figure 2.3). The code below shows how this can be done. It is a little more involved, but most of the complexity comes from detail. The approach is simple though: 1) find the empirical distribution of depths in the reserve; 2) define the inclusion probabilities based on this empirical distribution; and 3) sample according to those inclusion probabilities. We will sample a few more sites ($n = 100$), just to make the effect of the depth adjustment clear.

```
#####
#### Spatially balanced design -- Depth biased inclusion probs ####
#### Foster et al. NESP Biodiversity Hub Field Manuals ####
#####

par( mfrow=c(1,3), mar=rep( 4, 4))
n <- 100
#The number of 'depth bins' to spread sampling effort over.
nbins <- 4
#force the breaks so R doesn't use 'pretty'
breaks <- seq( from=min( bth.orig.grid$Depth, na.rm=TRUE),
  to=max( bth.orig.grid$Depth, na.rm=TRUE), length=nbins+1)
#Find sensible depth bins using pre-packaged code
tmpHist <- hist( bth.orig.grid$Depth, breaks=breaks, plot=FALSE)
#Find the inclusion probability for each 'stratum'
tmpHist$inclProbs <- (n/(nbins)) / tmpHist$counts
#Matching up locations to probabilities
tmpHist$ID <- findInterval( bth.orig.grid$Depth, tmpHist$breaks)
#A container for the design
design <- data.frame( siteID=1:nrow( bth.orig.grid),
  Easting=bth.orig.grid$Easting, Northing=bth.orig.grid$Northing,
  Depth=bth.orig.grid$Depth, inclProb=tmpHist$inclProbs[tmpHist$ID])
#Plot the depths and the inclusion probabilities
with( design, plot( Depth, inclProb, main="Inclusion Probabilities",
  ylab="Inclusion Probabilities", xlab="Depth (m)", pch=20, cex=1.4))
#Plot the inclusion probabilities in space
with( design,
  image.plot( uniqueEast, uniqueNorth,
    matrix( inclProb, nrow=length( uniqueEast), byrow=FALSE),
    xlab="", ylab="", main="Inclusion Probability", asp=1,
    ylim=NLims, xlim=ELims))
#Take the Sample using the inclusion probabilities
samp <- quasiSamp( n=n, dimension=2,
  potential.sites = design[,c("Easting", "Northing")],
  inclusion.probs=design$inclProb, nSampsToConsider=100*n)
#Plot the design
with( design, image.plot( uniqueEast, uniqueNorth, DepthMat,
  xlab="", ylab="", main="Spatially-Balanced Sample", asp=1,
  ylim=NLims, xlim=ELims,
```

```
col=rev(tim.colors()))
points( samp[,c("Easting","Northing")], pch=20, cex=2)
write.csv( design, file="design.csv", row.names=FALSE)
```

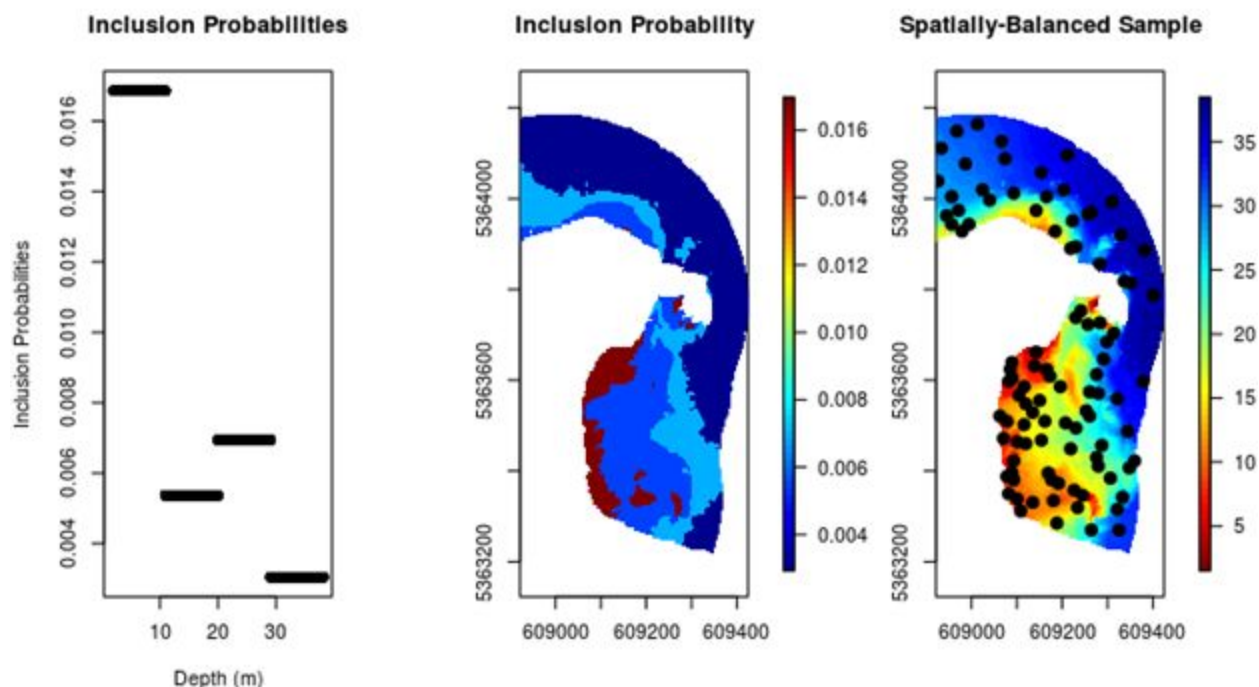


Figure 2.3: (Left panel) The empirical distribution of the 4 different depth bins. (Middle panel) The spatial distribution of the depth bins. (Right panel) A non-uniform spatially balanced sample, with inclusion probabilities based on the distribution of depths throughout the region. Shallow sites have been over-represented in the sample.

Incorporate legacy sites

Here, for edification purposes, we provide an illustration of how to design a spatially-balanced survey that accounts for the locations of legacy sites, which are those sites that we wish to include in the survey. The most likely reason for including legacy sites is that they have been sampled before as part of a previous randomisation process. Various names exist for legacy sites, including 'reference sites', and perhaps even 'sentinel sites' in some situations.

In our example, we first generate legacy sites and then generate more sites around them. To provide a little extra spice to the design we try to mimic the learning process: the $n = 6$ legacy sites are chosen with uniform probabilities (as we would do when there is no information about the area) and then the $n = 15$ new sites are chosen with a depth gradient altering the inclusion probabilities (Figure 2.4). This example therefore incorporates elements of the previous two examples.

```
#####
####   Spatially balanced design -- Legacy Sites (biased incl probs) ####
####   Foster et al. NESP Biodiversity Hub Field Manuals           #####
#####

#set up the plotting structure
par( mfrow=c(2,2), mar=c(3,3,3,3))
#number of samples
n_1 <- 6
##Take the sample for the legacy sites.
#Here they are a spatially balanced sample but in practice
# they would be supplied from a previous randomisation process
samp_legacy <- quasiSamp( n=n_1, dimension=2,
  potential.sites = bth.orig.grid[,c("Easting","Northing")],
  inclusion.probs=!is.na( bth.orig.grid$Depth))
#plot the legacy sites
with( bth.orig.grid, image.plot( uniqueEast, uniqueNorth, DepthMat,
  xlab="Easting", ylab="Northing", main="Legacy Sites",
  legend.lab="Depth (m)", asp=1, ylim=NLims, xlim=ELims,
  col=rev(tim.colors()), legend.mar=8.1))
points( samp_legacy[,c("Easting","Northing")], pch=17, cex=2)
#plot the depth-based inclusion probabilities
# scale first to sum to n=15
n <- 15
design$inclProb <- n * design$inclProb / sum( design$inclProb, na.rm=TRUE)
with( design,
  image.plot( uniqueEast, uniqueNorth,
    matrix( inclProb, nrow=length( uniqueEast)),
    xlab="", ylab="", main="Inclusion Probability", asp=1,
    ylim=NLims, xlim=ELims, legend.mar=8.1))
##Depth-based inclusion probabilities
#Alter the inclusion probabilities for the next sample
# inclusion probs taken from previous example
p2 <- alterInclProbs( legacy.sites=as.matrix(
  samp_legacy[,c("Easting","Northing")]),
  potential.sites=bth.orig.grid[,c("Easting","Northing")],
  inclusion.probs=design$inclProb)
#plot the altered inclusion probabilities
with( design,
  image.plot( uniqueEast, uniqueNorth,
    matrix( p2, nrow=length( uniqueEast)), ylim=NLims, xlim=ELims,
    xlab="", ylab="", main="Altered Inclusion Probability", asp=1, legend.mar=8.1))
##Take the new sample, spatially balanced around the legacy sites
samp <- quasiSamp( n=n, dimension=2,
  potential.sites = design[,c("Easting","Northing")],
  inclusion.probs=p2, nSampsToConsider=100*n)
```

```
#plot legacy sites and new sample sites.  
with( design, plot( Easting, Northing,  
  col=c('white',grey(0.9))[1+!is.na(inclProb)], ylim=NLims, xlim=ELims,  
  xlab="", ylab="", main="Combined Sample Locations", asp=1))  
points( samp_legacy[,c("Easting","Northing")], pch=17, cex=2, col='red')  
points( samp[,c("Easting","Northing")], pch=20, cex=2)  
legend( "bottomleft", c("Legacy Sites", "New Sites"), pch=c(17,20), pt.cex=2,  
  col=c('red','black'), bty='n')
```

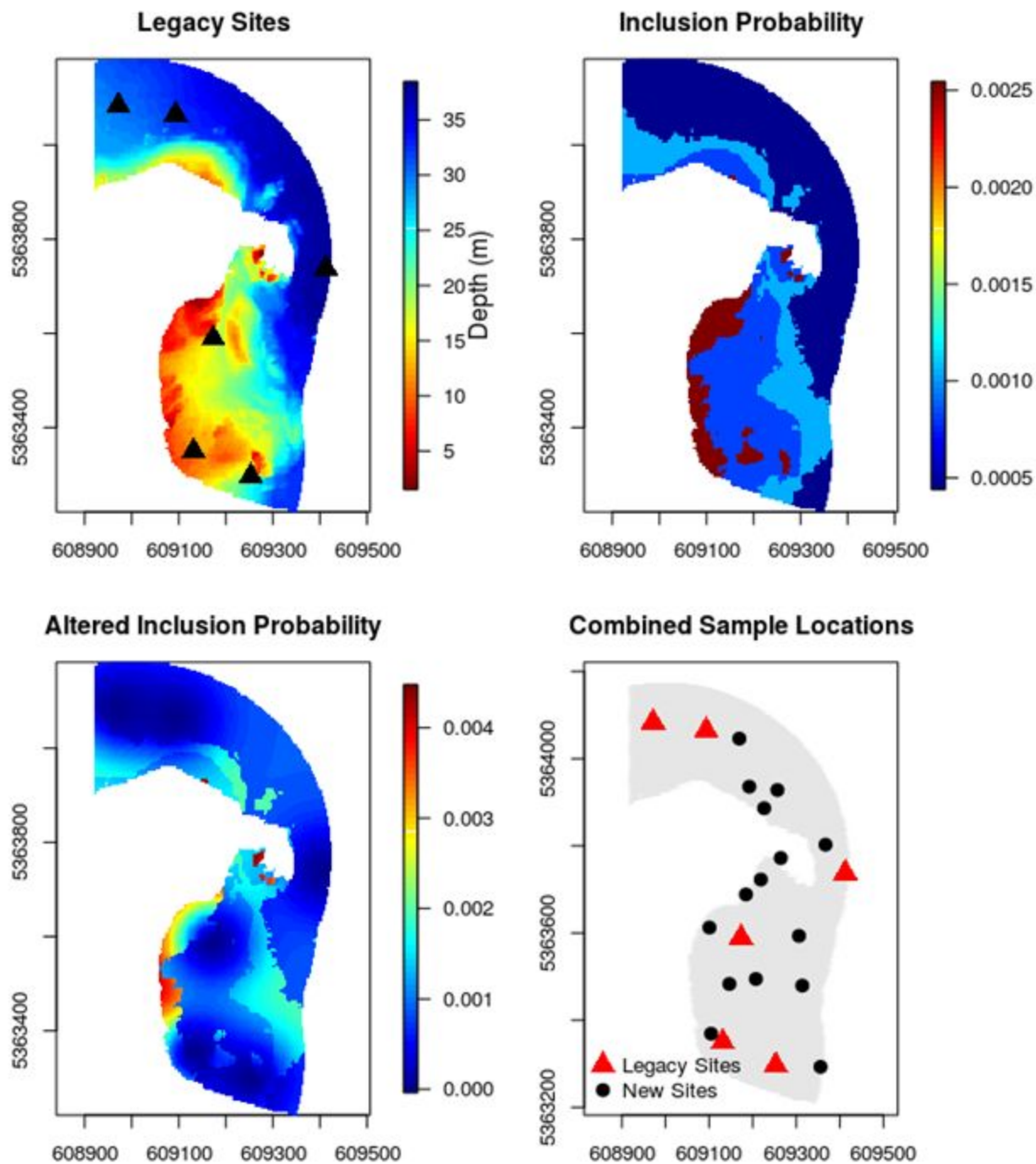



Figure 2.4: A spatially balanced design for Governor Island that incorporates legacy sites and has depth-varying inclusion probabilities (shallow sites are over-represented).

Case study summary

We have now seen how to generate three different kinds of designs: 1) a spatially balanced design with equal inclusion probabilities for when little is known about the sources of variation of the system; 2) a spatially balanced design with unequal inclusion probabilities for when we think we know where the locations with higher variance are likely to be; and 3) a spatially balanced design for when we have legacy sites that we want to take a repeat sample.

If future researchers wish to re-survey the area at some point in the future, then they have a choice to make: (i) Do they wish to revisit the same sites (to get a good temporal estimate)? (ii) Do they choose a new set of sites (to get a good spatial estimate)? Or (iii) Do they assume that the temporal change is not important and include the previous survey as part of the sample? This last scenario would be performed efficiently by using the original sample locations as legacy sites and spatially balance the new sample locations around those (as was done in the example). It will usually be sensible to combine these objectives by repeating some (not all) of the samples but choosing some new locations as well.

Field Manual Maintenance

In accordance with the universal field manual maintenance protocol described in Chapter 1 of the Field Manual package, this manual was updated in 2020 as Version 2. Updates reflect user feedback and new developments. There is currently no long-term plan or support for future updates. See Chapter 1 (Introduction to field manual package) for further details.

The version control for Chapter 2 (survey design) is below:

Version Number	Description	Date
0	Submitted for review (NESP Marine Hub, GA, external reviewers as listed Appendix A.	22 Dec 2017
1	Publicly released on www.nespmarine.edu	28 Feb 2018
2	Minor corrections, updates, added references and clarifications.	May 2020

Acknowledgements

The authors are grateful to Paul van Dam-Bates and Oliver Gansell (Department of Conservation, New Zealand) for reviewing Version 1 of this chapter.

References

Althaus, F.; Hill, N.; Ferrari, R.; Edwards, L.; Przeslawski, R.; Schönberg, C. H. L.; Stuart-Smith, R.; Barrett, N.; Edgar, G.; Colquhoun, J.; Tran, M.; Jordan, A.; Rees, T. & Gowlett-Holmes, K. A 2015 Standardised

- Vocabulary for Identifying Benthic Biota and Substrata from Underwater Imagery: The CATAMI Classification Scheme. *PLOS ONE, Public Library of Science*, 10, 1-18.
- Banerjee, S., B. Carlin, and A. Gelfand. 2004. Hierarchical Modeling and Analysis for Spatial Data. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, CRC Press
- Brewer, K., M. Hanif, and S. Tam. 1988. How Nearly Can Model-Based Prediction and Design-Based Estimation Be Reconciled? *Journal of the American Statistical Association* 83:128–132.
- Brown, J., B. Robertson, and T. McDonald. 2015. Spatially Balanced Sampling: Application to Environmental Surveys. *Procedia Environmental Sciences* 27:6 – 9.
- Brus, D., J. de Gruijter, and J. van Groenigen. 2006. Chapter 14 Designing Spatial Coverage Samples Using the k-means Clustering Algorithm. Pages 183 – 192 in A. M. P. Lagacherie and M. Voltz, editors. *Digital Soil Mapping An Introductory Perspective*, volume 31 of *Developments in Soil Science*. Elsevier.
- Brus, D., L. Spätjens, and J. de Gruijter. 1999. A sampling scheme for estimating the mean extractable phosphorus concentration of fields for environmental regulation. *Geoderma* 89:129 – 148.
- Chambers, R. 2011. Which Sample Survey Strategy? A Review of Three Different Approaches. *Pakistan Journal of Statistics* 27:337–357.
- Chen, B., Y. Pan, J. Wang, Z. Fu, and Y. Zhang, Y. Zhou. 2012. Even sampling designs generation by charges repulsion simulation. *Environmental Monitoring and Assessment* 184:3545–3556.
- Crawley, M. 2007. *The R Book*. Wiley.
- de Vries, A., and J. Meys. 2015. *R For Dummies*. Wiley.
- Diggle, P., and P. Ribeiro. 2007. *Model-based Geostatistics*. Springer Series in Statistics, Springer, New York.
- Dobbie, M., B. Henderson, and D. Stevens, Jr. 2008. Sparse sampling: Spatial design for monitoring stream networks. *Statist. Surv.* 2:113–153.
- Dobson, A.; Milner-Gulland, E.; Aebischer, N. J.; Beale, C. M.; Brozovic, R.; Coals, P.; Critchlow, R.; Dancer, A.; Greve, M.; Hinsley, A.; Ibbett, H.; Johnston, A.; Kuiper, T.; Le Comber, S.; Mahood, S. P.; Moore, J. F.; Nilsen, E. B.; Pocock, M. J.; Quinn, A.; Travers, H.; Wilfred, P.; Wright, J. & Keane, A. 2020 Making Messy Data Work for Conservation. *One Earth*, 2:455-465
- Fisher, R. 1925. *Statistical methods for research workers*. Edinburgh Oliver & Boyd.
- Foster, S., G. Hosack, N. Hill, N. Barrett, and V. Lucieer. 2014. Choosing between strategies for designing surveys: autonomous underwater vehicles. *Methods in Ecology and Evolution* 5:287–297.
- Foster, S., G. Hosack, E. Lawrence, R. Przeslawski, P. Hedge, M. Caley, N. Barrett, A. Williams, J. Li, T. Lynch, J. Dambacher, H. Sweatman, and K. Hayes. 2017. Spatially balanced designs that incorporate legacy sites. *Methods in Ecology and Evolution*.
- Foster, SD, Hosack, GR, Monk, J, Lawrence, E, Barrett, NS, Williams, A, Przewlaski, R. 2019. Spatially balanced designs for transect-based surveys. *Methods Ecol Evol.* 2019; 00: 1– 11.
- Gelman, A., J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin. 2013. *Bayesian Data Analysis*, Third Edition. Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis
- Gitzen, R. A., J. J. Millsaugh, A. B. Cooper, and D. S. Licht, editors. 2012. *Design and Analysis of Long-Term Ecological Monitoring Studies*. Cambridge University Press, Cambridge.
- Godambe, V. P., and V. M. Joshi. 1965. Admissibility and Bayes Estimation in Sampling Finite Populations. I. *Ann. Math. Statist.* 36:1707–1722.
- Grafström, A. 2012. Spatially correlated Poisson sampling. *Journal of Statistical Planning and Inference* 142:139–147.
- Grafström, A. 2013. Why Well Spread Probability Samples Are Balanced. *Open Journal of Statistics* 3:36–41.
- Grafström, A., N. L. P. Lundström, and L. Schelin. 2012. Spatially Balanced Sampling through the Pivotal Method. *Biometrics* 68:514–520.
- Grafström, A., and Y. Tillé. 2013. Doubly balanced spatial sampling with spreading and restitution of auxiliary totals. *Environmetrics* 24:120–131.
- Green, P. & C.J. MacLeod. 2016. SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7, 493-498.
- Hayes, K. R., G. R. Hosack, S. D. Foster, E. Lawrence, P. Hedge, N. S. Barrett, R. Przeslawski, and M. J. Caley. 2019. Designing monitoring programmes for Marine Protected Areas within an Evidence Based Decision Making paradigm. *Frontiers in Marine Science* 6, 746.
- Hill, N. A., N. Barrett, J. H. Ford, D. Peel, S. Foster, E. Lawrence, J. Monk, F. Althaus, K. R. Hayes. 2018. Developing indicators and a baseline for monitoring demersal fish in data-poor, offshore Marine Parks using probabilistic sampling. *Ecological Indicators*, 89, 610-621.
- Hosack, G.R., and E. Lawrence. 2013. Survey Design for Holothurians and *Tectus* at Ashmore Reef. A report prepared for the Australian Government Department of the Environment. 74 pp. CSIRO Wealth from Oceans Flagship, Hobart
- Kincaid, T. M. and Olsen, A. R. 2016. spsurvey: Spatial Survey Design and Analysis. R package version 3.3.

- Lawrence, E., K. Hayes, V. Lucieer, S. Nichol, J. Dambacher, N. Hill, N. Barrett, J. Kool, and J. Siwabessy. 2015. Mapping Habitats and Developing Baselines in Offshore Marine Reserves with Little Prior Knowledge: A Critical Evaluation of a New Approach. *PLoS ONE* 10:1–18.
- Leek, J. T., and R. D. Peng. 2015. What is the question? <http://www.sciencemag.org/content/early/2015/02/25/science.aaa6146.abstract>.
- Mapstone, B. D. 1995. Scalable Decision Rules for Environmental Impact Studies: Effect Size, Type I, and Type II Errors. *Ecological Applications* 5:401–410.
- McDonald, T. L. 2003. Review of Environmental Monitoring Methods: Survey Designs. *Environmental Monitoring and Assessment* 85:277–292.
- Minasny, B., and A. McBratney. 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers & Geosciences* 32:1378 – 1388.
- Mitchell, P. J., J. Monk, and L. Laurenson. 2017. Sensitivity of finescale species distribution models to locational uncertainty in occurrence data across multiple sample sizes. *Methods in Ecology and Evolution* 8:12–21.
- Perkins, N. R., S. D. Foster, N. A. Hill, and N. S. Barrett. 2016. Image subsampling and point scoring approaches for large-scale marine benthic monitoring programs. *Estuarine, Coastal and Shelf Science* 176:36 – 46.
- Perkins, N. R., S. D. Foster, N. A. Hill, M. P. Marzloff, and N. S. Barrett. 2017. Temporal and spatial variability in the cover of deep reef species: Implications for monitoring. *Ecological Indicators* 77:337 – 347.
- Robertson, B. L., J. A. Brown, T. McDonald, and P. Jaksons. 2013. BAS: Balanced Acceptance Sampling of Natural Resources. *Biometrics* 69:776–784.
- Royle, J., and D. Nychka. 1998. An algorithm for the construction of spatial coverage designs with implementation in SPLUS. *Computers & Geosciences* 24:479 – 488.
- Schobernd, Z. H., N. M. Bacheler, and P. B. Conn. 2014. Examining the utility of alternative video monitoring metrics for indexing reef fish abundance. *Canadian Journal of Fisheries and Aquatic Sciences* 71:464–471.
- Smith, A. N. H., M. J. Anderson, and M. D. M. Pawley. 2017. Could ecologists be more random? Straightforward alternatives to haphazard spatial sampling. *Ecography*
- Stevens, D., and A. Olsen. 2004. Spatially Balanced Sampling of Natural Resources. *Journal of the American Statistical Association* 99:262–278.
- Taylor, L. 1961. Aggregation, Variance and the Mean. *Nature* 189:732–735
- Thompson, S. 2012. *Sampling*. Third Edition. Wiley
- Tillé, Y., and M. Wilhelm. 2017. Probability Sampling Designs: Principles for Choice of Design and Balancing. *Statistical Science* 32:176–189
- Urquhart, N., and T. Kincaid. 1999. Designs for Detecting Trend from Repeated Surveys of Ecological Resources. *Journal of Agricultural, Biological, and Environmental Statistics* 4:404–414.
- Venables, W., and B. Ripley. 2002. *Modern Applied Statistics with S*. Fourth Edition. Springer.
- Walvoort, D., D. Brus, and J. J.J. de Gruijter. 2010. An R package for spatial coverage sampling and random sampling from compact geographical strata by k-means. *Computers & Geosciences* 36:1261 – 1267.
- Xie, Y. 2014. knitr: A Comprehensive Tool for Reproducible Research in R. in V. Stodden, F. Leisch, and R. D. Peng, editors. *Implementing Reproducible Computational Research*. Chapman and Hall/CRC.