Title
Effects of sample design and landscape features on a measure of environmental heterogeneity

Journal
Methods in Ecology and Evolution, 7(7)

ISSN
2041-210X

Authors
Christianson, Danielle S
Kaufman, Cari G

Publication Date
2016-07-01

DOI
10.1111/2041-210x.12539

Peer reviewed

Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2016, **7**, 770–782 doi: 10.1111/2041-210X.12539

Effects of sample design and landscape features on a measure of environmental heterogeneity

Danielle S. Christianson[1]* and Cari G. Kaufman[2]

[1]*Energy and Resources Group, University of California, 310 Barrows Hall, Berkeley, CA 94720-3050, USA; and*
[2] *Department of Statistics, University of California, 367 Evans Hall, Berkeley, CA, 94720-3860, USA*

Summary

1. Environmental heterogeneity, an important influence on organisms and ecological processes, can be quantified by the variance of an environmental characteristic over all locations within a study extent. However on landscapes with autocorrelation and gradient patterns, estimating this variance from a sample of locations may lead to errors that cannot be corrected with statistical techniques.

2. We analytically derived the relative expected sampling error of sample designs on landscapes with particular gradient pattern and autocorrelation features. We applied this closed-form approach to temperature observations from an existing study. The expected heterogeneity differed, both in magnitude and direction, amongst sample designs over the study site's likely range of autocorrelation and gradient features.

3. We conducted a simulation study to understand the effects of (i) landscape variability and (ii) design variability on an average sampling error. On 10 000 simulated landscapes with varying gradient and autocorrelation features, we compared estimates of variance from a variety of structured and random sample designs. While gradient patterns and autocorrelation cause large errors for some designs, others yield near-zero average sampling error. Sample location spacing is a key factor in sample design performance. Random designs have larger range of possible sampling errors than structured designs due to the potential for sample arrangements that over- and under-sample certain areas of the landscape.

4. When implementing a new sample design to quantify environmental heterogeneity via variance, we recommend using a simple structured design with appropriate sample spacing. For existing designs, we recommend calculating the relative expected sampling error via our analytical derivation.

Key-words: autocorrelation, gradient, monitoring, sampling, spatial, variability

Introduction

Environmental heterogeneity, a measure of variability in abiotic and biotic conditions, is important to many areas of ecological study, such as species diversity via niche theory (Holdridge 1947; Whittaker 1956), microbial biogeochemical processes (*e.g.* Fierer *et al.* 2006; Sierra *et al.* 2011) and population dynamics (*e.g.* García-Carreras & Reuman 2013). Increasingly, ecologists use environmental heterogeneity to understand how organisms and ecological processes may respond to a warmer future (*e.g.* Jentsch *et al.* 2011; Clark *et al.* 2013; Thornton *et al.* 2014). For example, heterogeneity in climate over small spatial scales results in climate refugia that may be of critical importance to the persistence of species with limited options for range expansion (Dobrowski 2011; Keppel & Wardell-Johnson 2012). Additionally, while many organisms optimize for average conditions, heterogeneity is important for understanding when and how likely thresholds (*e.g.* frost tolerances, heat stress) may be breached (Meehl *et al.* 2000; Jentsch & Beierkuhnlein 2008).

There are many definitions of heterogeneity. Some authors simply define a range of values, while others include aspects of scale, quantify spatial features such as clumping, or qualitatively describe composition (Kolasa & Rollo 1991; Wiens 2000). Quantitative metrics include coefficient of variance (standard deviation/mean), variance/mean ratio, variograms and others (see overview by Downing 1991). To be useful in many ecological applications, such a metric must describe the range of available environmental conditions, either in space or time. For example, consider a population of trees whose seeds disperse on the landscape. To maintain a viable population, there must be an adequate number of locations on the landscape through time that have environmental conditions suitable for the seeds to germinate and the resulting seedlings to survive. By quantifying the available set of environmental conditions, we can determine whether the environments for successful germination and establishment exist and thus predict whether early life-history stages will limit population viability.

Accurately estimating heterogeneity, however, is not straightforward when sampling autocorrelated quantities that may also have gradient patterns, a gradual directional change. To account for autocorrelation or gradients, ecologists employ

*Correspondence author. E-mail: dsvehla@berkeley.edu

a combination of sample designs (Dutilleul 2011) and well-developed statistical tools, including techniques such as analysis of variance (Dutilleul 1993; Legendre *et al*. 2004), spatial regression analysis (reviewed by Beale *et al*. 2010) and classical geostatistic metrics like Moran's I, correlograms and variograms (Legendre 1993). Researchers also employ methods from spatial statistics, modelling the distribution of an environmental characteristic with heterogeneity due to gradient patterns modelled in the mean and heterogeneity due to autocorrelation modelled in an autocorrelation matrix (Cressie 1993). However, these traditional methods that *account* for autocorrelation and gradient patterns are not sufficient to *quantify* the metric of our choice, which we will now describe.

Consider an environmental characteristic, Y, that varies across the landscape, for example air temperature at 2 metre height. Although Y may vary continuously across a broad region, we restrict our attention to a particular geographical area and spatial scale. We call the geographical area the *extent* and define the spatial scale by dividing the landscape into gridcells, whose resolution we refer to as the *grain*. The grain is ideally chosen such that heterogeneity at the chosen scale is appropriate to the organism or process under study, that is heterogeneity on a finer spatial scale is not practically important. In reality, however, the appropriate spatial scale may be unknown or variable, for example due to multiple scales of influence (Dutilleul 2011) and life-history change (Wiens 2000). For the purposes of this article, we assess spatial heterogeneity only, defining Y as a representative value (*e.g*. mean, minimum) over a period of time or at a single point in time. More broadly, the same exercise and issues can be applied to temporal heterogeneity.

To quantify the variability of available conditions on the landscape, we define heterogeneity as the variance of Y across all gridcells, which we call the complete variance, V_N. We can calculate V_N if we observe Y at all N gridcell locations. In addition, this calculation assumes that measurement error is negligible relative to variation across the landscape, so that V_N is essentially the 'true' quantification of the heterogeneity. This quantity might also be thought of as the variance of an 'empirical distribution' consisting of all values across the grid, as illustrated in Fig. 1(a).

Measuring environmental quantities at every gridcell location, however, is rarely possible, and we typically cannot calculate the complete variance exactly. Instead, we observe the environmental characteristic at a restricted set of locations in a sample design, estimating the complete variance, V_N, by the sample variance, $\hat{V}_s$. As shown in Fig. 1(b–c), different sample designs can over- or underestimate the complete variance on landscapes with autocorrelation and gradient patterns.

Why, then, are traditional ecological methods insufficient? Whereas many sample design recommendations and statistical tools are designed to account for the effects of autocorrelation and gradients separately, we want a description of heterogeneity that includes them. Additionally, statistical tools that account for autocorrelation and gradients, such as analysis of variance and others mentioned earlier, do not estimate the complete variance. In spatial statistics models, Y is considered a random variable drawn from an underlying stochastic process. However, none of the model parameters describing this process straightforwardly corresponds to the complete variance of the particular landscape that is observed.

Thus, we are interested in how the sample variance performs as an estimator of the complete variance. The average error (bias) can be calculated directly in closed form. We develop this analytical approach and illustrate its use with a forested study in Section 'Analytical derivation of expected sampling error'. To find the spread around the average error due to variability in landscape features, which cannot be calculated analytically, we use simulated landscapes to compare a variety of structured and random sampling designs in Section 'Simulation study'. Using the results from these two approaches, we make recommendations for (i) implementing a new sample design to best quantify heterogeneity and (ii) quantifying sampling error from existing sample designs.

Analytical derivation of expected sampling error

MODELLING LANDSCAPE CHARACTERISTICS

Suppose that the environmental characteristic Y on a gridded landscape with N gridcells can be described by a multivariate normal distribution:

$$\boldsymbol{Y_N} \sim MVN(\boldsymbol{\mu_N}, \boldsymbol{\Sigma}), \qquad \text{eqn 1}$$

where $\boldsymbol{Y_N}$ is the vector of Y values at gridcells 1 to N, and $\boldsymbol{\mu_N}$ is a vector of mean values at gridcells 1 to N (see Table 1 for a summary of notations). The $N \times N$ covariance matrix $\boldsymbol{\Sigma}$ (gridcells 1 to N across the columns and 1 to N down the rows) describes the pairwise relationship between gridcells. We used a multivariate normal distribution for conceptual ease; however, this derivation holds with any distribution.

For landscapes without a gradient, $\boldsymbol{\mu_N}$ has the same value, μ, for all N locations. For landscapes with a gradient, $\boldsymbol{\mu_N}$ has different values and can be modelled by multiplying the gridcell coordinates by a vector $\boldsymbol{\beta} = [\beta_o, \beta_x, \beta_y]$. We decomposed the covariance $\boldsymbol{\Sigma}$ into a variance term, σ^2, and an autocorrelation matrix, $\boldsymbol{K}$: $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{K}$. By doing this we assumed a constant variance across the landscape. We fixed $\sigma^2 = 1$ without loss of generality because σ^2 is a scaling factor.

We modelled autocorrelation with the exponential function in our example:

$$\boldsymbol{K} = e^{-\boldsymbol{d_N}/\rho}, \qquad \text{eqn 2}$$

where $\boldsymbol{d_N}$ is an $N \times N$ matrix of the pairwise distance between all gridcells, and $\rho > 0$ is a range parameter that determines the distance at which gridcells no longer have strong influence on one another. While we chose the commonly used exponential function (Zimmerman & Stein 2010), we derive the analytical solution for a general $\boldsymbol{K}$ in which other autocorrelation functions can be used. We report spatial autocorrelation in terms of ρ as measured in gridcells rather than distance units for interpretability given our unitless landscape.

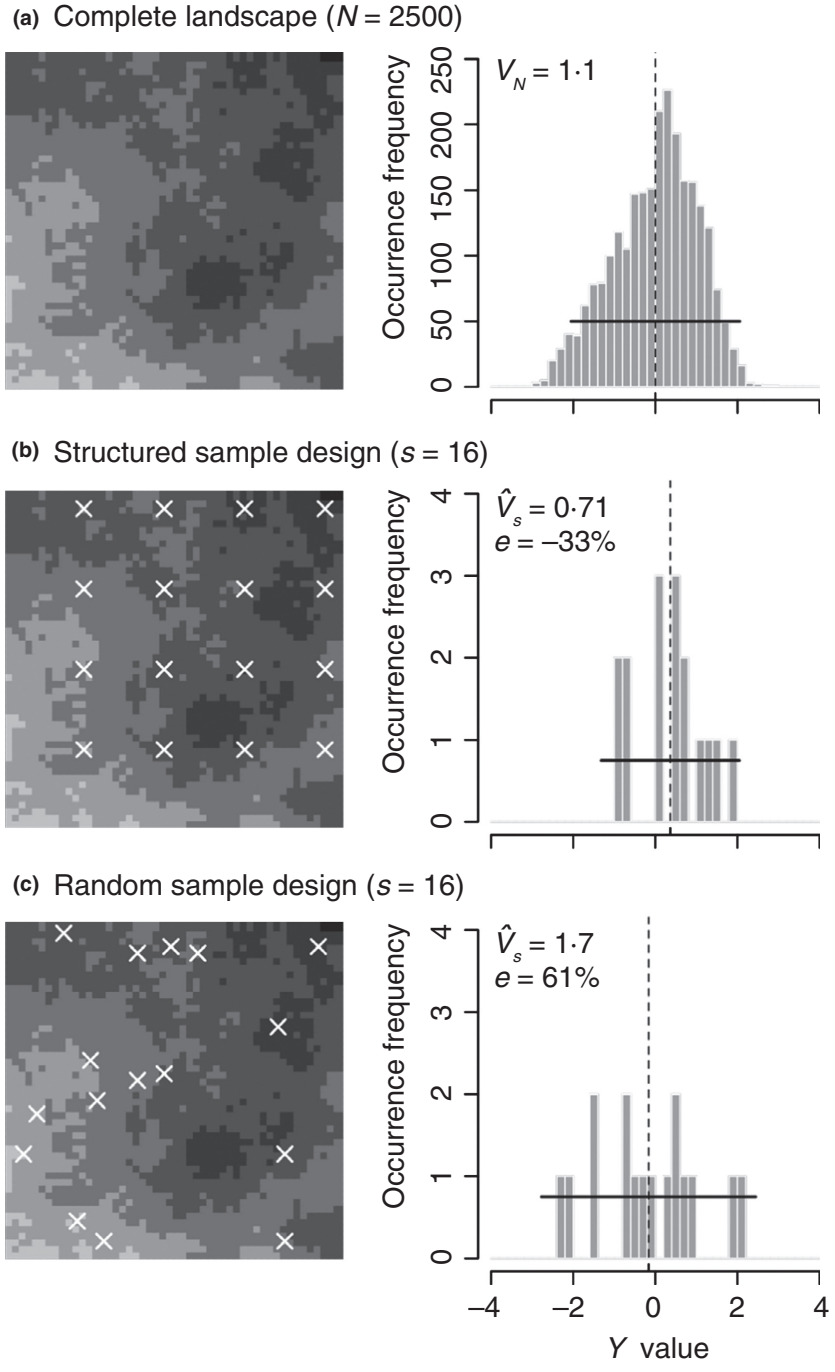

Fig. 1. Landscapes with gradient patterns and/or spatial autocorrelation, as in this example landscape that has both, can cause significant error in estimates of the complete variance V_N. In panel a, the empirical distribution for all gridcells within the extent is shown, with the mean as the dotted line and ± twice the standard deviation as the solid horizontal line. The sample variance, $\widehat{V}_s$, of the structured sample design in panel b underestimates the complete variance by 33%, while that of the random sample design in panel c overestimates the complete variance by 61%. For specific landscape parameterizations used in this and all subsequent figures, see Table S1 (Appendix S3).

EXPECTED VALUE OF COMPLETE AND SAMPLE VARIANCE

We define the complete variance according to the standard equation:

$$V_N = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y}_N)^2 \quad \text{with} \quad \bar{Y}_N = \frac{1}{N}\sum_{i=1}^{N} Y_i. \qquad \text{eqn 3}$$

In Appendix S1, we show that the expected value of the complete variance is

$$E[V_N] = \frac{1}{N}\left[\sigma^2 tr\left\{\left(\boldsymbol{I_N} - \frac{1}{N}\boldsymbol{J_N}\right)\boldsymbol{K}\right\} + \boldsymbol{\mu}_N^T\left(\boldsymbol{I_N} - \frac{1}{N}\boldsymbol{J_N}\right)\boldsymbol{\mu_N}\right], \qquad \text{eqn 4}$$

where $\boldsymbol{I_N}$ is an $N \times N$ identity matrix and $\boldsymbol{J_N}$ is an $N \times N$ matrix with all entries equal to one.

The observations under a given sample design are a subset of the $\boldsymbol{Y_N}$ vector, which we call $\boldsymbol{Y_s}$. We write $\boldsymbol{Y_s} = \boldsymbol{H}\boldsymbol{Y_N}$, where $\boldsymbol{H}$ is a $s \times N$ matrix consisting mostly of zeros and single entries in each row, equal to one, that correspond to the elements of $\boldsymbol{Y_N}$ to be selected (Appendix S1). Thus,

Table 1. Notation definitions

N, s	locations	N = number of total locations within extent s = number of sample locations
$Y, \boldsymbol{Y_N}, \boldsymbol{Y_s}$	environmental characteristic	Y = an environmental characteristic $\boldsymbol{Y_N}$ = vector of Y at N grid locations $\boldsymbol{Y_s}$ = vector of Y at s grid locations
$\mu, \boldsymbol{\mu_N}, \bar{Y}_N, \bar{Y}_s$	mean	μ = model mean (same for all N) $\boldsymbol{\mu_N}$ = vector of μ for all locations N $\bar{Y}_N$ = mean across all locations N $\bar{Y}_s$ = sample mean
$\sigma^2, V_N, \widehat{V}_s$	variance	σ^2 = model variance V_N = complete variance $\widehat{V}_s$ = sample variance
$SD_N, \widehat{SD_s}$	standard deviation	SD_N = complete standard deviation $\widehat{SD_s}$ = sample standard deviation
$\boldsymbol{\Sigma}, \boldsymbol{K}, \boldsymbol{d_N}$	covariance, autocorrelation, & distance matrices	$\boldsymbol{\Sigma}$ = model covariance matrix $\boldsymbol{K}$ = model autocorrelation matrix $\boldsymbol{d_N}$ = matrix of pairwise distances between all N locations
$\boldsymbol{\beta}, \rho$	landscape parameters	$\boldsymbol{\beta}$ = vector of gradient parameters $(\beta_o, \beta_x, \beta_y)$ = (intercept, x, y) ρ = autocorrelation range parameter
$\boldsymbol{I_N}, \boldsymbol{J_N}, \boldsymbol{H}$	analytical derivation	$\boldsymbol{I_N}$ = $N \times N$ identity matrix ($\boldsymbol{I_s}$ = $s \times s$ identity matrix) $\boldsymbol{J_N}$ = $N \times N$ matrix of all ones ($\boldsymbol{J_s}$ = $s \times s$ matrix of all ones) $\boldsymbol{H}$ = $s \times N$ matrix that samples s locations from N
$e_{jk}^{(i)}, \bar{e}_{jk}, R_s$	error	$e_{jk}^{(i)}$ = sampling error $\bar{e}_{jk}$ = average error R_s = relative expected sampling error
i, j, k	error indices	i = index for simulated landscapes j = index for a single design k = index for design configuration

$\boldsymbol{Y_s} = \boldsymbol{H}\boldsymbol{Y_N} \sim MVN(\boldsymbol{H}\boldsymbol{\mu_N}, \sigma^2\boldsymbol{H}\boldsymbol{K}\boldsymbol{H}^T)$. With sample variance defined as

$$\widehat{V}_s = \frac{1}{s-1}\sum_{i=1}^{s}(Y_i - \bar{Y}_s)^2 \quad \text{with} \quad \bar{Y}_s = \frac{1}{s}\sum_{i=1}^{s} Y_i, \qquad \text{eqn 5}$$

the expected value of the sample variance is

$$E[\widehat{V}_s] = \frac{1}{s-1}\left[\sigma^2 tr\left\{\left(\boldsymbol{I_s} - \frac{1}{s}\boldsymbol{J_s}\right)\boldsymbol{H}\boldsymbol{K}\boldsymbol{H}^T\right\} + \boldsymbol{\mu_N}^T\boldsymbol{H}^T\left(\boldsymbol{I_s} - \frac{1}{s}\boldsymbol{J_s}\right)\boldsymbol{H}\boldsymbol{\mu_N}\right], \qquad \text{eqn 6}$$

where $\boldsymbol{I_s}$ is an $s \times s$ identity matrix and $\boldsymbol{J_s}$ is an $s \times s$ matrix with all entries equal to one. For landscapes with no gradient, calculation of the expected variances is simplified as the term containing $\boldsymbol{\mu_N}$ is zero:

$$E[V_N] = \frac{\sigma^2}{N} tr\left\{\left(\boldsymbol{I_N} - \frac{1}{N}\boldsymbol{J_N}\right)\boldsymbol{K}\right\} \quad \text{and}$$
$$E[\widehat{V}_s] = \frac{\sigma^2}{s-1} tr\left\{\left(\boldsymbol{I_s} - \frac{1}{s}\boldsymbol{J_s}\right)\boldsymbol{H}\boldsymbol{K}\boldsymbol{H}^T\right\}. \qquad \text{eqns 7,8}$$

RELATIVE EXPECTED SAMPLING ERROR

We calculated a relative expected sampling error for each sample design as the per cent difference between the complete and sample expected variances:

$$\text{relative expected sampling error} = R_s = 100\% \times \frac{E[\widehat{V}_s] - E[V_N]}{E[V_N]}. \qquad \text{eqn 9}$$

We explored the general usefulness of this framework with sample designs in 1D space for landscapes with no gradient (Appendix S2). As expected, we found that amongst structured designs with the same sample size and fixed relative sampling locations, R_s was the same regardless of placement within the extent. However, R_s for random designs with the same sample size varied significantly as the sample locations changed from design to design. Additionally, we found that average sample spacing between sample locations was a key factor in determining the magnitude and direction of R_s. We explore this point more fully in our simulation study (Section 'Simulation study') and focus next on the practical application of R_s.

AN EXAMPLE CASE: ALPINE TREELINE WARMING EXPERIMENT (ATWE)

As an example, we applied the analytical derivation to temperature measurements from an existing sample design employed in the Alpine Treeline Warming Experiment (ATWE). ATWE is located along Niwot Ridge in the eastern Colorado Rockies, USA (40°3′ N, 105°36′ W). The Forest study site (elev. 3100 m) is located in a mature subalpine forest, primarily composed of conifer trees. Details of the study are described in Reinhardt *et al.* (2011). Five control plots each contained four soil temperature probes arranged in a 1 m square. Commensurate with this distance, we chose a grain of 1 m for our analysis. The five plots were chosen to span a range of landscape features including sun–shade patterns, topography and rockiness within the ~100 m × 50 m site extent (Fig. 2 inset). We compared the expected sampling error, R_s, between this clustered random design, as is, to a simple random version with only 1 measurement per plot. To carry out the analysis, we used eqn 6, calculating $\boldsymbol{H}$ by matching the geo-located positions of the temperature probes with a 100 × 50 gridded landscape.

With so few data, we were not able to assess spatial autocorrelation or gradient patterns via statistical methods to calculate $\boldsymbol{K}$ or $\boldsymbol{\mu_N}$. Rather, we identified a range of likely autocorrelation intensities, 5–20 m, based on the sun–shade patterns from the semi-open canopy. While other factors influence the soil

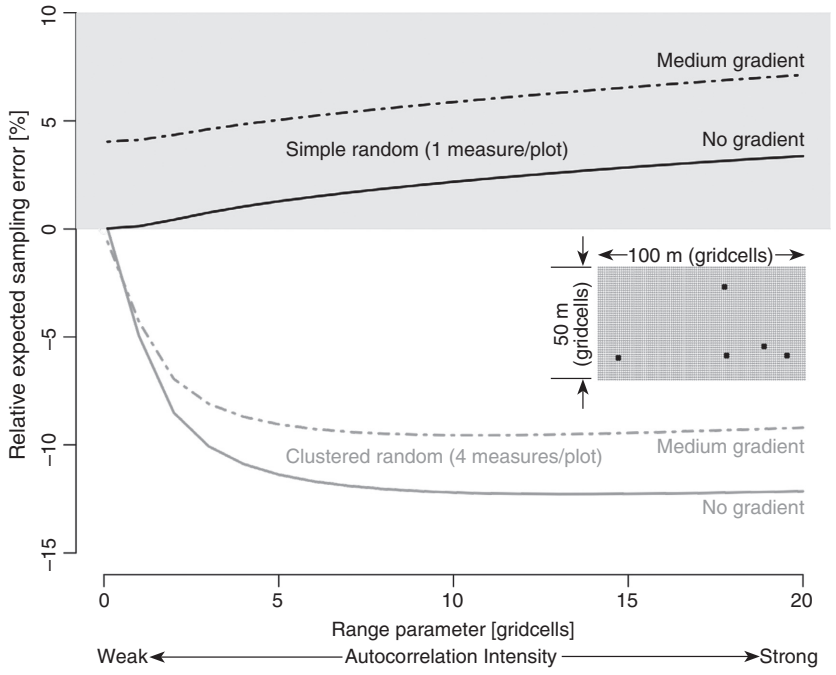

Fig. 2. Assessment of soil temperature heterogeneity in existing design at ATWE study site by calculating the relative expected sampling error, R_s, via our analytical derivation. The inset figure illustrates the study site, as a 100 × 50 gridcell landscape with 1 gridcell = 1 m and the five measurement plots in black.

temperature such as topography and soil properties, direct radiation is likely a primary driver. Potential drivers of a gradient pattern at this site are topographical: a gradual east–west elevation change of 15 m with minimal change in aspect. Thus, we suspect there to be essentially no gradient effect ($\boldsymbol{\beta}$ = [0,0,0]); however, we investigated a medium gradient to be conservative ($\boldsymbol{\beta}$ = [0,0,2]).

As shown in Fig. 2, the relative expected sampling error, R_s, differs in both magnitude and direction between the simple random design (1 measure/plot) and the clustered random design (4 measures/plot) over the range of autocorrelation intensities expected at the site. The simple design overestimates the complete variance while the clustered design underestimates it. On landscapes both with and without a gradient, the simple design, with smaller absolute R_s, yields a better estimate. When a gradient is included in the analysis, the magnitude of R_s for the simple design increases, while that for the clustered design decreases, however not enough to recommend it over the simple design. These results counter the common intuition that collecting more data is always better, which is only guaranteed to be true when the environmental characteristic at different sample locations can be thought of as independent of each other. This is not the case on landscapes with gradient patterns and spatial autocorrelation, which we explore further in the following simulation study.

Simulation study

While the analytical derivation is a useful tool to quickly assess the expected sampling error of a specific design, we completed a simulation study to systematically investigate the performance of a number of structured and random sample designs on landscapes with varying gradient patterns and spatial autocorrelation. In particular, this numerical approach allowed us to quantify the spread around R_s due to landscape variability. In addition, we assessed the importance of design variability in the case of random sample designs. We will demonstrate that sample designs estimate complete variance differently on different landscape types, sample spacing is a key factor in design performance, and design variability significantly affects random designs. We performed the analysis using R, version 3.0.2 (R Development Core Team 2013).

SIMULATION OF LANDSCAPE PATTERNS

We simulated 10 000 landscapes for each of 185 landscape types varying in gradient pattern and spatial autocorrelation. We used a square grid with 50 × 50 gridcells, yielding 2500 total locations, N. With edge length of 1 distance unit, the grain is 1 gridcell and the extent is 50 gridcells in each direction. By expressing the distance between two locations in units of gridcells, our results can be scaled to any real distance.

As in the analytical derivation, we simulated landscapes according to a multivariate normal distribution: $\boldsymbol{Y_N} \sim MVN(\boldsymbol{\mu_N}, \sigma^2\boldsymbol{K})$, which allowed us to vary gradient patterns via $\boldsymbol{\mu_N}$ and spatial autocorrelation via $\boldsymbol{K}$. We modelled landscapes with a 1D gradient in the y-direction by multiplying the y coordinates by a constant, taking values β_y = 0, 0·1, 2, or 10. The case β_y = 0 corresponds to no gradient. We again fixed σ^2 = 1 and used an exponential function with range parameter ρ to model spatial autocorrelation (eqn 2). We expected that the strength of autocorrelation, determined by ρ, would have a significant effect on efficacy of the sample designs. Thus, we varied ρ widely, from 0·1 to 100 gridcells (Fig. S4, Appendix S3). With ρ = 0·1, the elements of $\boldsymbol{Y_N}$ can be considered nearly independent and identically distributed, because neighbouring

gridcells have essentially zero correlation. At the opposite extreme, ρ = 100, each location is strongly influenced by all other locations.

SAMPLE DESIGNS

We tested a series of structured and random sample designs types, which we call:

- *Simple structured* – single samples located in a regularly spaced grid
- *Clustered structure*d – groups of samples located in a regularly spaced grid
- *Nested structured* – single samples located in two regularly spaced grids, one embedded within the other
- *Simple random* – single samples randomly located on the landscape
- *Clustered random* – groups of samples randomly located on the landscape.

In Fig. S6 (Appendix S3), we illustrate the sample designs and define the nomenclature used.

All sample designs have a *sample size* equal to the total number of sampled locations within the extent. The *spacing* is the distance in gridcells between sampled locations. For clustered structured designs, the spacing describes the distance between the clusters. For designs that have clusters or nested groups (nested structured), the *cluster size* is equal to the number of sample locations in each cluster or nest, and the *cluster spacing* describes the spacing between sample locations in each cluster or nest. While we present our results using a few representative configurations, we illustrate all designs configurations that we investigated in Fig. S7 (Appendix S3).

CALCULATION OF COMPLETE AND SAMPLE STANDARD DEVIATION

For a single simulated landscape of a given landscape type and for each sample design, we calculated:

$$SD_N = \sqrt{V_N} = \sqrt{\frac{1}{N}\sum_i^N (Y_i - \bar{Y}_N)^2} \quad \text{and}$$
$$\widehat{SD_s} = \sqrt{\widehat{V}_s} = \sqrt{\frac{1}{s-1}\sum_i^s (Y_i - \bar{Y}_s)^2}, \qquad \text{eqns 10,11}$$

whereas before N is the total number of locations, and s is the number of sample locations. We refer to SD_N as the complete standard deviation (what we cannot observe and wish to estimate) and to $\widehat{SD_s}$ as the sample standard deviation (our estimate based on s sample locations). We used standard deviation rather than variance as a metric describing heterogeneity because it has the same units as the environmental characteristic Y.

We report error between SD_N and $\widehat{SD_s}$ at two levels for each design configuration. First, we quantified a relative sampling error, e, defined as per cent difference between SD_N and $\widehat{SD_s}$:

$$\text{sampling error} = e_{jk}^{(i)} = 100\% \times \frac{\widehat{SD}_{s,jk}^{(i)} - SD_n^{(i)}}{SD_n^{(i)}}, \qquad \text{eqn 12}$$

where i represents a single simulated landscape (1–10 000) within a landscape type and j represents a specific design within the design configuration k. A negative sampling error means $\widehat{SD_s}$ underestimates SD_N, while a positive sampling error indicates that $\widehat{SD_s}$ is an overestimate. Then, we calculated an average error, $\bar{e}_{jk}$, for each design by averaging the sampling error over the 10 000 simulated landscapes:

$$\text{average error} = \bar{e}_{jk} = \frac{1}{10,000}\sum_{i=1}^{10,000} e_{jk}^{(i)}. \qquad \text{eqn 13}$$

Recall that this average error is the average of relative rather than absolute sampling errors. We examined the spread around average error from landscape variation by varying i in $e_{jk}^{(i)}$. Likewise, we varied j in $\bar{e}_{jk}$ within a design configuration k to examine the effects of design variation. For example, within the simple random configuration with sample size 16, we compared specific designs, each with a unique set of 16 sample locations.

To understand the impact of design variation, we randomly selected 100 specific designs for each random configuration and compared the average error for each of these 100 designs. To verify the results from our analytical derivation, we computed the average error of 20 randomly selected specific designs for each structured configuration. The differences in average error are indeed small (Fig. S8, Appendix S4). Thus, for structured configurations, we show results for a single representative structured design.

SIMULATION RESULTS

Our simulation results illustrate three main findings: (i) sampling error can be significant in the presence of autocorrelation and especially gradients; (ii) sample spacing indicates how the landscape is sampled and thus is a key factor in the sampling error observed; (iii) the range of possible sampling errors for random designs is large due to landscape and design variability.

RESULT 1: SAMPLING ERROR CAN BE SIGNIFICANT

Structured sample designs: autocorrelation can cause significant average error

The effect of autocorrelation is striking, as shown in Fig. 3-1. As autocorrelation intensity increases, the average error, $\bar{e}_{jk}$, can become quite large, varying from –38% to 12%. This increase in error occurs rapidly over range parameter values of 0–20 gridcells, approximately 25% of the maximum distance. within the extent Increasing the sample size without changing the sample spacing reduces $\bar{e}_{jk}$ (d-e in Fig. 3-1); however, $\bar{e}_{jk}$ may still be significant. Moreover, the configuration determines the sign of $\bar{e}_{jk}$. In the majority of design configurations, the

sample standard deviation, $\widehat{SD_s}$, underestimates the complete standard deviation, SD_N, that is, the error is negative (b-e in Fig. 3-1). However, $\widehat{SD_s}$ overestimates SD_N for a configuration in which the sample spacing is maximized across the extent (a).

Structured sample designs: gradients increase average error

As shown in Fig. 3-2, medium and strong gradient patterns degrade performance for all structured sample designs, the magnitude depending on both the strength of the gradient and the design configuration. In particular, design configurations that perform well without gradients (*e.g.* b) continued to do well. Design configurations that had poor performance without gradients (*e.g.* a and c) have even larger average error, $\bar{e}_{jk}$, on landscapes with gradients. For these configurations, medium strength gradients roughly double $\bar{e}_{jk}$, while strong gradients triple $\bar{e}_{jk}$. In Appendix S4·2 and S4·3, we examine clustered and nested designs. In summary, we find that these strategies generally lead to worse estimates of SD_N.

Random designs: effects of autocorrelation and gradients differ dependent on design

As shown in Fig. 4, random design configurations perform differently on landscapes with varying autocorrelation intensities and gradient patterns. By comparing the difference between simple Random 4 and simple Random 16 (b to a) and between simple Random 4 and clustered Random 16-4 (b to c), it is clear that increasing the sample size via a simple random configuration (*i.e.* not in clusters) leads to a greater decrease in the magnitude of error.

Additionally, the effects of autocorrelation and gradient patterns are complicated. For example, for design c-i on landscapes with no gradient in Fig. 4, $\bar{e}_{jk}$ worsens initially then improves as autocorrelation increases. Furthermore, significant change in $\bar{e}_{jk}$ can happen rapidly over weak autocorrelation intensities (0–20 gridcells). Gradients can also have beneficial (*e.g.* a-ii, b-ii and c-ii) or detrimental (*e.g.* all extreme designs) effects on $\bar{e}_{jk}$. Overall, increasing both autocorrelation intensity and gradient strengths increases the variability of $\bar{e}_{jk}$ across designs within a random configuration. It is this design

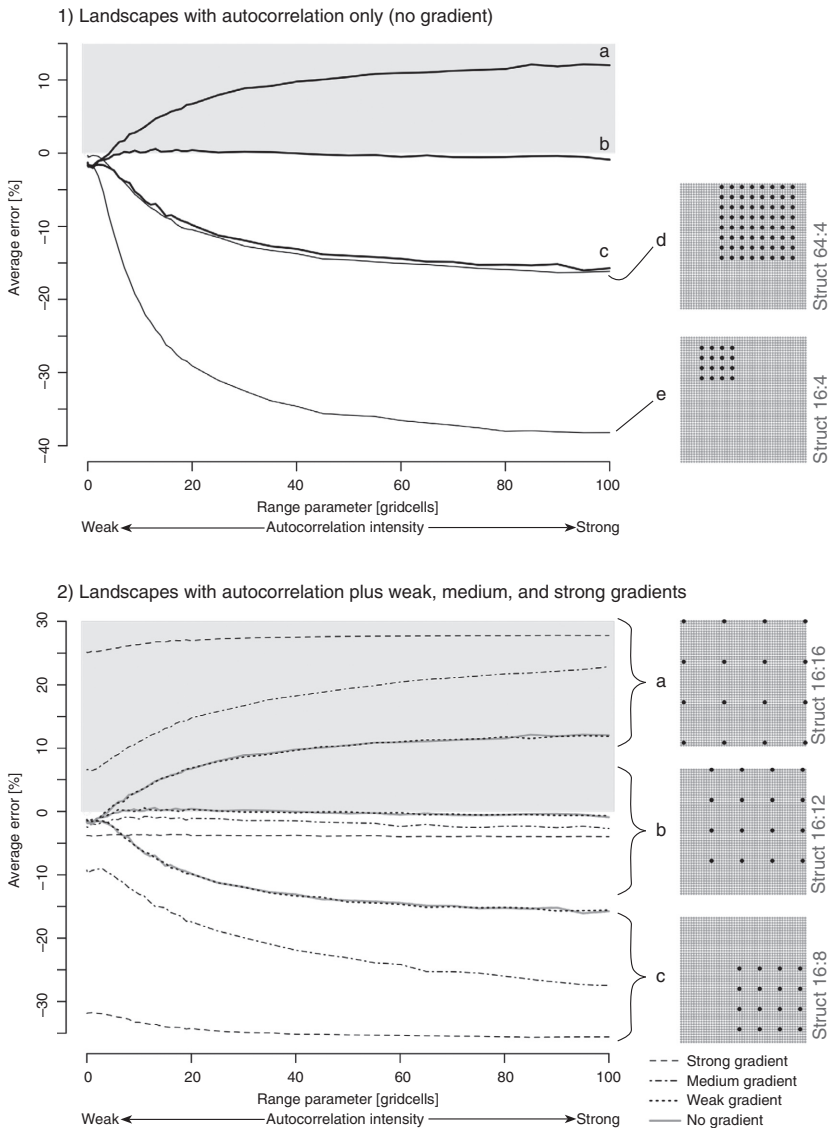

Fig. 3. Average error, $\bar{e}_{jk}$, for structured design configurations on landscapes with autocorrelation only, that is, no gradient (panel 1) and additionally with a series of weak to strong gradients (panel 2). The $\bar{e}_{jk}$ reported is for the representative design shown.

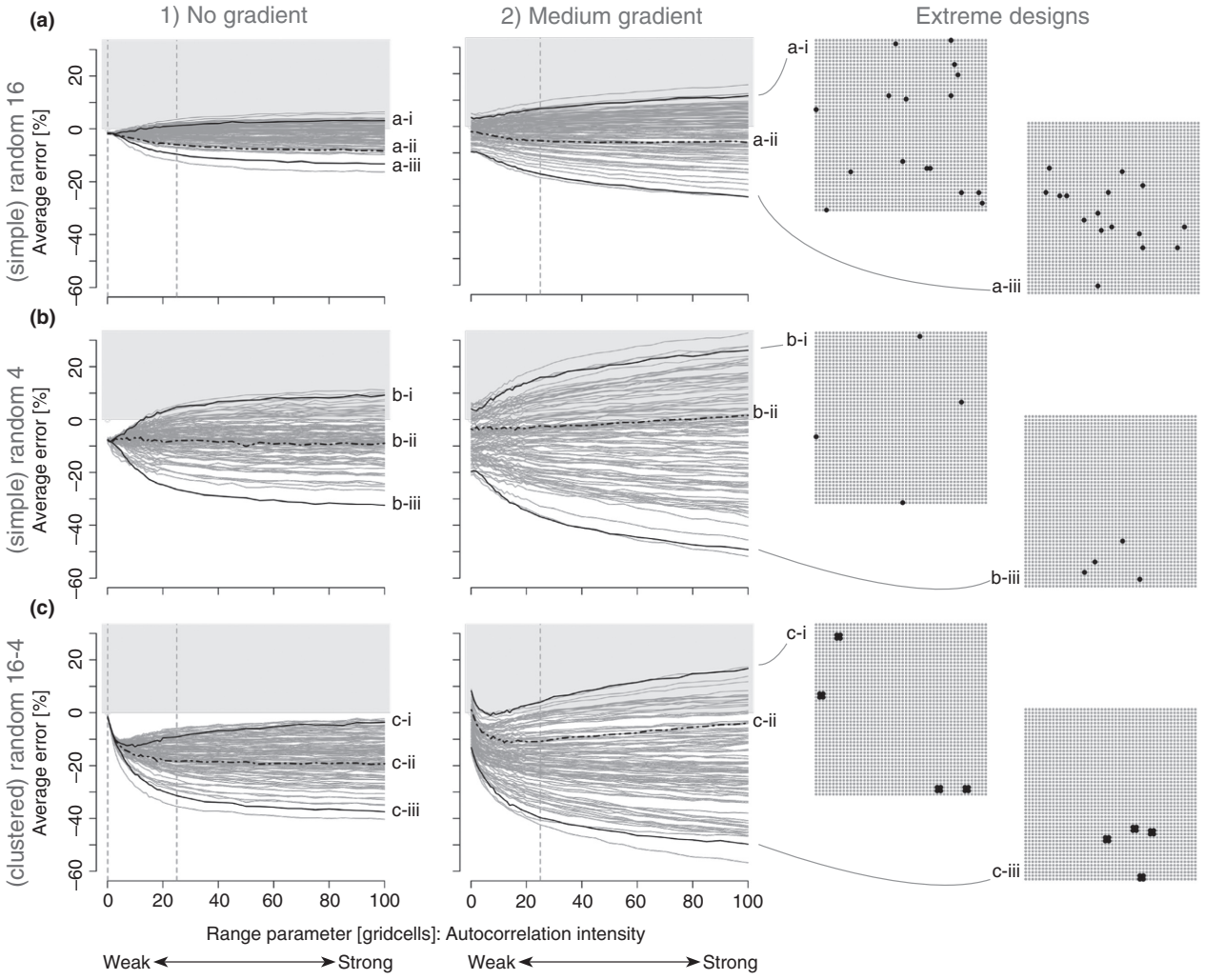

Fig. 4. Spread in average error, $\bar{e}_{jk}$, for random designs due to design variation. Each line is $\bar{e}_{jk}$ for a single design within each configuration: Random 16, 4 and 16-4 (rows a-c, respectively). For each design configuration, examples of extreme designs (i & iii) are highlighted in solid black. Examples of designs (ii) that have generally better estimates of the complete standard deviation on landscapes with gradients are shown in the dashed black line. Landscapes with no gradient (column 1) and with a medium gradient (column 2) are the same landscapes used in Fig. 3. The vertical lines are the representative landscapes used in subsequent Fig. 6.

variability that led us to investigate what design characteristics yield near-zero $\bar{e}_{jk}$ across autocorrelation intensities and gradient pattern strengths. As discussed in the next result, sample spacing is a key factor.

RESULT 2: SAMPLE SPACING IS A KEY FACTOR IN OBSERVED SAMPLING ERROR

For landscapes with autocorrelation, we observed a positive linear relationship between a design's average error, $\bar{e}_{jk}$, and its average distance between sampling locations. That is, for a given sample size, the average distance seems to determine the bias in $\bar{e}_{jk}$. As shown in Fig. 5, this relationship holds for autocorrelated landscapes with and without gradients and for both random and structured design configurations. For a sample size of 16, designs with average spacing of 25–30 gridcells consistently show $\bar{e}_{jk}$ near zero, where as designs with lower or higher average spacing have a non-zero $\bar{e}_{jk}$ that becomes more extreme as the autocorrelation intensity increases (not shown). Average spacing greater than 25–30 gridcells results in overestimates of SD_N (positive $\bar{e}_{jk}$), while average spacing less than 25–30 gridcells result in underestimates (negative $\bar{e}_{jk}$). Clustered designs almost always underestimate SD_N, with clusters reducing the average sample spacing (Appendix S4·2). We performed a similar analysis for the spread around $\bar{e}_{jk}$ due to landscape variation. We did not find strong relationships (see Fig. S11, Appendix S4). We discuss the importance of sample spacing further in the discussion section.

RESULT 3: RANDOM DESIGNS HAVE A LARGE RANGE OF SAMPLING ERROR DUE TO THE COMBINATION OF DESIGN AND LANDSCAPE VARIABILITY

When a random design configuration is chosen for a study's sample design, only one design is implemented. Therefore,

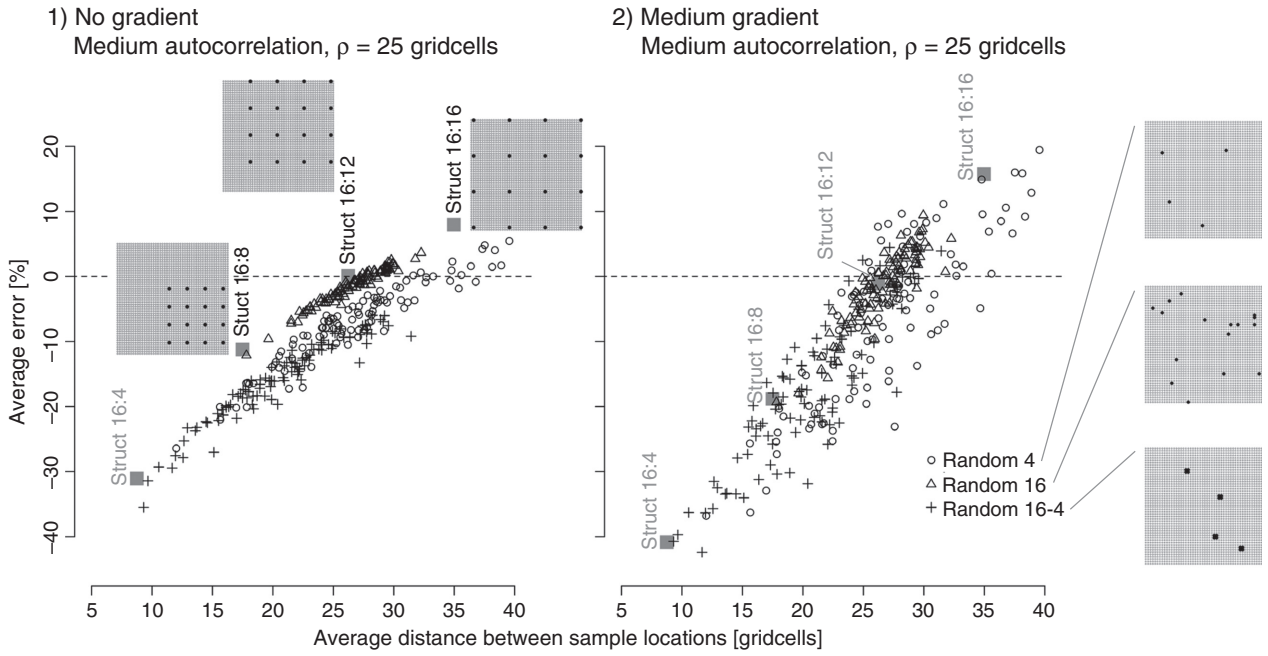

Fig. 5. Average error, $\bar{e}_{jk}$, versus average distance between sampling locations. Designs for the random design configurations Random 16, 4 and 16-4 are shown with circles, triangles and crosses, respectively. These are the same random designs for which results are reported in Fig. 4. Average error, $\bar{e}_{jk}$, for simple structured designs with increasing spacing are reported with grey squares for comparison.

the sampling error of the implemented design is a combination of both the design and landscape variability of that random design configuration. Thus, as we will show, the range of possible sampling error for a randomly selected random design is much larger than that of a structured design with the same sample size. This finding holds for environmental quantities, $\boldsymbol{Y_N}$, that do not repeat in a regular pattern on the landscape at multiples of the structured design frequency.

In Fig. 6, we illustrate the range of possible sampling error due to both design and landscape variation. For three representative landscape types, we show kernel estimates of the density of raw sampling error, $e_{jk}^{(i)}$, that underlie average errors for designs within (i) the best performing simple structured configuration Struct 16:12; (ii) the simple random configuration Random 16; and (iii) the clustered random configuration Random 16-4 (rows 1–3, respectively).

For landscapes with no autocorrelation or gradient (Fig. 6 column 1), little difference exists in the sampling errors between the three design configurations. These results are expected, as this landscape type is essentially a random surface. However, on landscapes with medium autocorrelation (column 2), as we saw in Fig. 4, design variability affects both the simple and clustered random configurations, that is the densities shift apart. Landscape variability also significantly affects the clustered configuration, that is the densities' widths increase. On landscapes with both autocorrelation and a medium gradient (column 3), the effect of design variability is exacerbated for the random designs, while the effect of landscape variability on the structured design is reduced as shown by a decrease in the density width.

When we consider the implications of the combined effects of design and landscape variability on landscapes with autocorrelation and gradients, the simple structured configuration clearly has a significantly smaller range of possible sampling errors than either of the random configurations. While any one simple random design may have a sampling error comparable to that of a simple structured design, the possible sampling error is considerably larger when choosing a design randomly.

Discussion

Our simulation study shows that good estimates of heterogeneity as measured by complete variance can be obtained with a well-chosen sample design over a range of gradient strengths and autocorrelation intensities. We have identified sample spacing as a key factor in determining designs with good estimation. The question then becomes what characteristics of sample designs are associated with optimal spacing. Additionally, how should a researcher practically apply an understanding of sample design to quantify heterogeneity when: (i) implementing a new sample design; (ii) using data from an existing sample design?

As shown in Fig. 5, for a given sample size, an ideal range of average sample spacing, which holds across gradients, leads to near-zero average sampling error. The same range holds across autocorrelation intensities (not shown). Because the proportion of the landscape covered and the evenness of that coverage determine the average sample spacing, the average sample spacing can be considered a proxy for adequate sampling of the landscape within a given sample size. In general, larger average spacing, up to

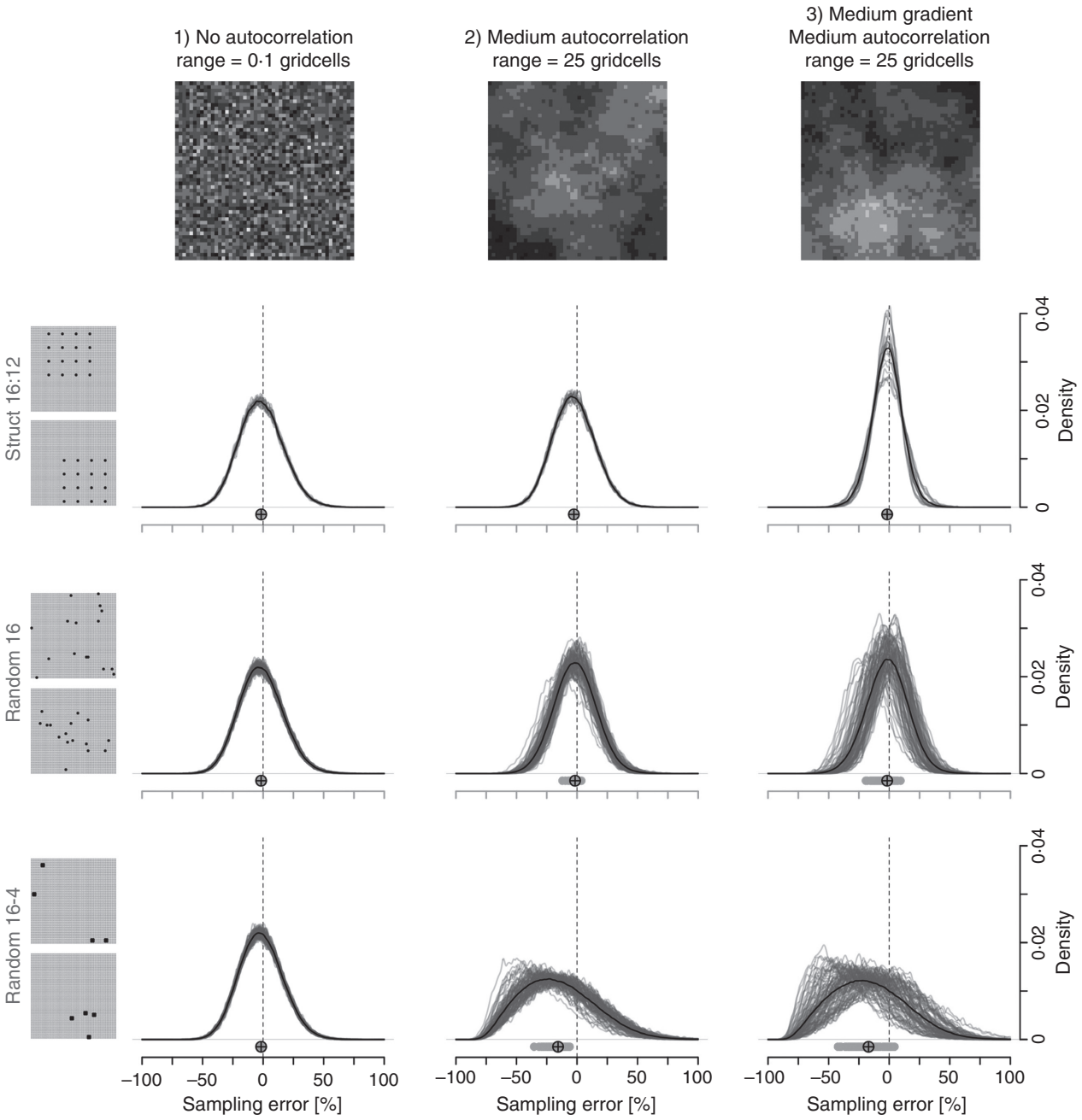

Fig. 6. The combined effects of landscape and design variation are shown by comparing the kernel estimates of the density of raw sampling error, $e_{jk}^{(i)}$, for structured and random designs calculated over the 10 000 simulated landscapes. Results are shown in columns for three landscapes types indicated by vertical dotted lines in Fig. 4. Each grey line is a single design within each of three configuration types shown in rows, each with sample size 16. The black line is the combined density distribution of raw sampling error for designs within each configuration type. The mean of the density, which is the average error, $\bar{e}_{jk}$, is represented by a grey dot below the densities. The black cross-haired dot is the mean of the average errors.

a distance constrained by the sample size, typically leads to greater landscape coverage. For structured designs, coverage increases evenly in both directions as average spacing increases; however, for random designs, the evenness may not be constant in both directions. Thus, the problem is that random designs, as well as clusters and nests, can have particular spatial arrangements that over-or under-sample certain regions of the extent, leading to larger sampling errors.

The combined effects of sample size and sample spacing determine a threshold sample size above which there are essentially no differences between simple structured and simple random designs' complete variance estimates, that is, the densities of raw sampling error $e_{jk}^{(i)}$, like those illustrated in Fig. 6, look similar. For random designs, the possibility for spatial sampling arrangements that over- or under-sample certain regions of the extent decreases as the sample size increases. Additionally, average sample spacing necessarily approaches the ideal

spacing in which a large portion of the landscape is sampled in a more-or-less evenly spaced fashion. Below the threshold sample size in a random design, too many possibilities exist for sampling arrangements that lead to poor complete variance estimates. An appropriately spaced simple structured design, with both adequate landscape coverage and even spacing within that covered portion, provides a better estimate.

Finally, as the strength of both gradient patterns and spatial autocorrelation increases, errors in estimating the complete variance can become quite large. Thus, when implementing a new sample design or evaluating a proposed design, it is helpful to have a sense of how strong a gradient or autocorrelation pattern may be. When sufficient data are not available, expert knowledge can be used to identify a range of autocorrelation strengths to consider, as employed in the ATWE example. In the case when existing data are used, some sample designs, such as structured or stratified random, may allow for estimation of gradient patterns (Dutilleul 1993, 2011). For data sets with sufficiently large sample sizes, geo- and spatial statistical techniques can be used to estimate the autocorrelation strength (Isaaks & Srivastava 1989; Zimmerman 2010).

In Fig. 7, we summarize our recommendations for practical application of our results. When implementing a new design to quantify heterogeneity via complete variance, we recommend a simple structured design when landscape features do not repeat at multiples of the appropriate sample spacing. We make this recommendation regardless of the sample size for 2 reasons: (i) an appropriately space simple structured design ensures a good estimate as discussed above, and (ii) the sample size threshold is complicated to determine as we discuss further below. Although motivated by a different purpose, authors have made a similar recommendation for using systematic sample designs to estimate spatial gradient and autocorrelation patterns (Zidek & Zimmerman 2010; Dutilleul 2011). In contrast, Zhu & Stein (2005) report that structured (regular) designs are outperformed by other non-random designs when estimating the autocorrelation parameters of a statistical model of the landscape.

We recommend the following rule of thumb to determine the appropriate spacing for simple structured designs on roughly square-shaped extents based on our results for multiple simple structured configurations (Appendices S3 and S4). Divide the length of a side by the number of sample locations in each row (or column) and then truncate the resulting number to get the sample spacing. For the example of our 50 × 50 gridded landscape and a sample size of 16 (4 × 4), a spacing of 12 (50/4 = 12·5 → 12) gives the best estimate of the complete standard deviation, that is, sampling error is closest to zero for all gradient and autocorrelation intensities.

As previously discussed, above a sample size threshold there is essentially no difference between an appropriately spaced simple structured design and a simple random design. A general threshold is difficult to determine because it depends on strength of the gradient and autocorrelation intensities, total number of locations (gridcells) within the extent and

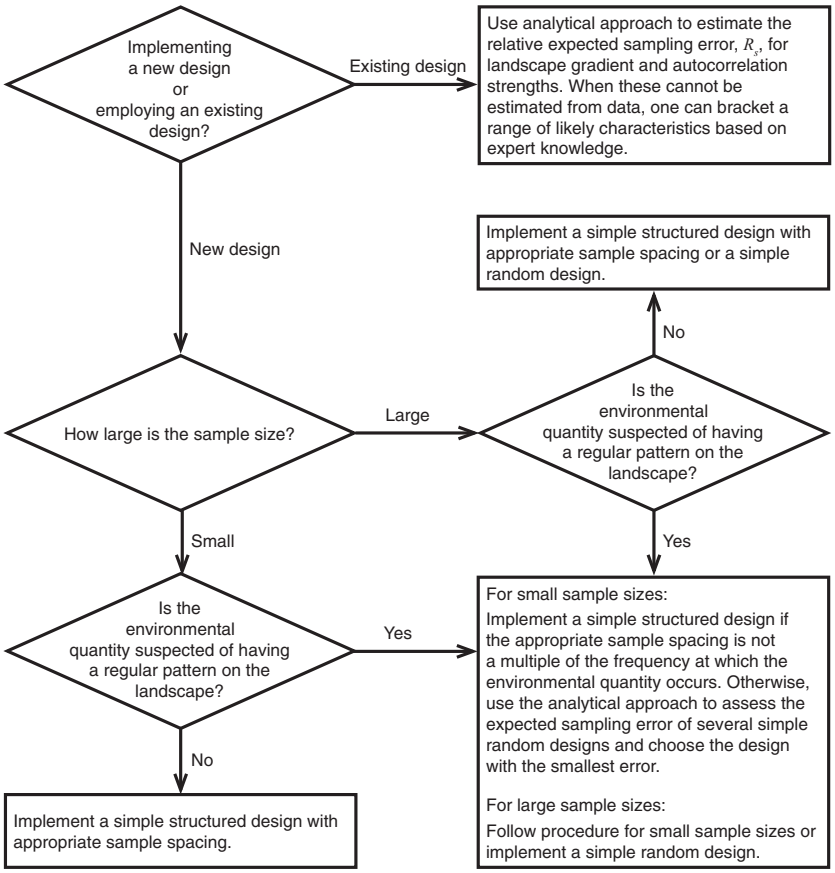

Fig. 7. Recommendations for quantifying environmental heterogeneity via complete variance when gradient patterns and spatial autocorrelation occur on the landscape.

potentially the extent shape. For our simulated 50 × 50 gridded landscape with no gradient, we estimate a threshold of ~1%, which is a sample size of 25. Sample sizes in ecological research are often smaller than this threshold, which we expect to increase with stronger gradients. Additionally, we expect extent shape to matter because we observed different thresholds between the 50 × 50 gridded landscape from the simulation study and the 1D gridded line from the analytical derivation.

Researchers, however, may have objectives other than heterogeneity when selecting a sample design. Random designs are often employed to avoid unintentional sampling bias and to satisfy criteria of inferential statistical methods (Zidek & Zimmerman 2010). As mentioned earlier, structured designs can be problematic for sampling landscape features that repeat at multiple of the sample spacing (Dutilleul 2011). Such patterns do occur in natural systems and care must be taken to implement an appropriate design. Fortunately, a number of particular simple random designs with near-zero sampling error across a range of autocorrelation intensities and gradient strengths do exist. Thus, if desired, one can select a particular design by iteratively evaluating it with our analytical derivation, as exemplified in the ATWE example. Additionally, while we did not include a stratified random sample design in our analysis, we expect that it would provide good estimates as the stratification criteria would likely limit over- and under-sampling.

To estimate the sampling error from a proposed design or existing data, we suggest using the analytical derivation to calculate the expected sampling error for the particular sample design at hand. The expectation can be calculated for a set of autocorrelation intensities and gradient strengths. We suggest bracketing these landscape features with a low and high value in cases where the value cannot be directly estimated, as illustrated in the ATWE example. This method allows a relatively quick assessment of the expected sampling error direction and magnitude. To quantify the range of sampling errors around the expected sampling error, the simulation methodology employed in this paper can be modified for specific study details. This error range due to landscape variability may be especially important for designs with low sample sizes.

We have begun a description of how different sample designs estimate heterogeneity via complete variance. In addition, we identified a tool to estimate the expected sampling error for landscapes with gradient patterns and spatial autocorrelation. Environmental heterogeneity is increasingly used in assessments of ecological pattern and function, in particular in climate change studies. To better represent this heterogeneity, future work should investigate more complex landscapes, for example those in which variance or autocorrelation is not constant (*i.e.* landscapes with non-stationary features). Metrics other than variance, such as skew and kurtosis are also important descriptors of environmental conditions. The influence of gradients and autocorrelation on estimates of such metrics should be examined. Finally, given the interest in heterogeneity across ecological disciplines, a summary of the different methods of defining and quantifying heterogeneity would serve as a guide to researchers, improving a common set of tools and enabling increased collaborative effort.

Acknowledgements

We thank Erin Conlisk, John Harte, Lara Kueppers, Lydia Smith and Cleo Woelfle-Erskine for helpful comments. A UC Berkeley Philomathia Graduate Student Fellowship in the Environmental Sciences supported DSC while conducting this work.

Data accessibility

R scripts: uploaded as online supporting information (Appendix S5).

References

Beale, C.M., Lennon, J.J., Yearsley, J.M., Brewer, M.J. & Elston, D.A. (2010) Regression analysis of spatial data. *Ecology Letters*, **13**, 246–264.

Clark, J.S., Melillo, J., Mohan, J. & Salk, C. (2013) The seasonal timing of warming that controls onset of the growing season. *Global Change Biology*, **20**, 1136–1145.

Cressie, N.A.C. (1993) *Statistics for spatial data, revised edn*. Wiley-Interscience, New York.

Dobrowski, S.Z. (2011) A climatic basis for microrefugia: the influence of terrain on climate. *Global Change Biology*, **17**, 1022–1035.

Downing, J.A. (1991) Biological Heterogeneity in Aquatic Ecosystems. *Ecological Heterogeneity* (eds J. Kolasa & S.T.A. Pickett), pp. 160–180. Springer, New York, NY.

Dutilleul, P. (1993) Spatial heterogeneity and the design of ecological field experiments. *Ecology*, **74**, 1646–1658.

Dutilleul, P. (2011) *Spatio-Temporal Heterogeneity*. Cambridge University Press, Cambridge.

Fierer, N., Colman, B.P., Schimel, J.P. & Jackson, R.B. (2006) Predicting the temperature dependence of microbial respiration in soil: a continental-scale analysis. *Global Biogeochemical Cycles*, **20**, GB3026.

García-Carreras, B. & Reuman, D.C. (2013) Are changes in the mean or variability of climate signals more important for long-term stochastic growth rate? (I. Gudelj, Ed.). *PLoS ONE*, **8**, e63974.

Holdridge, L.R. (1947) Determination of World plant formations from simple climatic data. *Science*, **105**, 367–368.

Isaaks, E.H. & Srivastava, R.M. (1989) *An Introduction to Applied Geostatistics*. Oxford University Press, New York.

Jentsch, A. & Beierkuhnlein, C. (2008) Research frontiers in climate change: effects of extreme meteorological events on ecosystems. *Comptes Rendus Geoscience*, **340**, 621–628.

Jentsch, A., Kreyling, J., Elmer, M., Gellesch, E., Glaser, B., Grant, K. *et al.* (2011) Climate extremes initiate ecosystem-regulating functions while maintaining productivity. *Journal of Ecology*, **99**, 689–702.

Keppel, G. & Wardell-Johnson, G.W. (2012) Refugia: keys to climate change management. *Global Change Biology*, **18**, 2389–2391.

Kolasa, J. & Rollo, C.D. (1991) Introduction: the Heterogeneity of Heterogeneity: a Glossary. *Ecological Heterogeneity* (eds J. Kolasa & S.T.A. Pickett), pp. 1–23. Springer, New York.

Legendre, P. (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology*, **74**, 1659.

Legendre, P., Dale, M., Fortin, M.J., Casgrain, P. & Gurevitch, J. (2004) Effects of spatial structures on the results of field experiments. *Ecology*, **85**, 3202–3214.

Meehl, G.A., Karl, T., Easterling, D.R., Changnon, S., Pielke, R., Changnon, D. *et al.* (2000) An introduction to trends in extreme weather and climate events: observations, socioeconomic impacts, terrestrial ecological impacts, and model projections. *Bulletin of the American Meteorological Society*, **81**, 413–416.

R Development Core Team (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. URL http://www.R-project.org [accessed 25 September 2013]

Reinhardt, K., Castanha, C., Germino, M.J. & Kueppers, L.M. (2011) Ecophysiological variation in two provenances of Pinus flexilis seedlings across an elevation gradient from forest to alpine. *Tree Physiology*, **31**, 615–625.

Sierra, C.A., Harmon, M.E., Thomann, E., Perakis, S.S. & Loescher, H.W. (2011) Amplification and dampening of soil respiration by changes in temperature variability. *Biogeosciences*, **8**, 951–961.

Thornton, P.K., Ericksen, P.J., Herrero, M. & Challinor, A.J. (2014) Climate variability and vulnerability to climate change: a review. *Global Change Biology*, **20**, 3313–3328.

Whittaker, R.H. (1956) Vegetation of the Great Smoky Mountains. *Ecological Monographs*, **26**, 1–69.

Wiens, J.A. (2000) Ecological heterogeneity: an otogeny of concepts and approaches. *The Ecological Consequences of Environmental Heterogeneity* (eds M.J. Hutchings, A.J. Libby & A.J.A. Stewart), pp. 9–31. Cambridge University Press, Cambridge.

Zhu, Z. & Stein, M.L. (2005) Spatial sampling design for parameter estimation of the covariance function. *Journal of Statistical Planning and Inference*, **134**, 583–603.

Zidek, J.V. & Zimmerman, D.L. (2010) Monitoring Network Design. *Handbook of Spatial Statistics* (eds A. Gelfand, P. Diggle, M. Fuentes & P. Guttorp), pp. 131–148. CRC Press, Boca Raton.

Zimmerman, D.L. (2010) Likelihood-Based Methods. *Handbook of Spatial Statistics* (eds A. Gelfand, P. Diggle, M. Fuentes & P. Guttorp), pp. 45–56. CRC Press, Boca Raton.

Zimmerman, D.L. & Stein, M.L. (2010) Classical Geostatistic Methods. *Handbook of Spatial Statistics* (eds A. Gelfand, P. Diggle, M. Fuentes & P. Guttorp), pp. 29–44. CRC Press, Boca Raton.

Received 7 December 2015; accepted 14 January 2016
Handling Editor: Robert Freckleton

Supporting Information

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

Appendix S1. Analytical derivation of the expected value of complete and sample variances.

Appendix S2. Analytical estimates of complete variance of a 1D landscape with varying autocorrelation (Figs S1–S3).

Appendix S3. Additional simulation methods (Table S1 and Figs S4–S7).

Appendix S4. Additional simulation results (Figs S8–S11).

Appendix S5. R scripts.