Journal of Applied Ecology
British Ecological Society

Maximum power for monitoring programmes: optimising sampling designs for multiple monitoring objectives

Journal:	*Journal of Applied Ecology*
Manuscript ID:	JAPPL-2009-00034
Manuscript Type:	Standard Paper
Date Submitted by the Author:	18-Jan-2009
Complete List of Authors:	Bijleveld, Allert; Royal Netherlands Institute for Sea Research (NIOZ), 1. Department of Marine Ecology van Gils, Jan; Royal Netherlands Institute for Sea Research (NIOZ), 1. Department of Marine Ecology van der Meer, Jaap; Royal Netherlands Institute for Sea Research (NIOZ), 1. Department of Marine Ecology Dekinga, Anne; Royal Netherlands Institute for Sea Research (NIOZ), 1. Department of Marine Ecology Kraan, Casper; Royal Netherlands Institute for Sea Research (NIOZ), 1. Department of Marine Ecology van der Veer, Henk; Royal Netherlands Institute for Sea Research (NIOZ), 1. Department of Marine Ecology Piersma, Theunis; Royal Netherlands Institute for Sea Research (NIOZ), 1. Department of Marine Ecology; University of Groningen, Centre for Ecological and Evolutionary Studies, Animal Ecology Group
Key-words:	macrobenthic invertebrates, intertidal, model-based inference, design-based inference, spatial autocorrelation, generalised least squares, power analysis, landscape ecology

Maximum power for monitoring programmes: optimising sampling designs for multiple monitoring objectives

Allert I. Bijleveld[1*], Jan A. van Gils[1], Jaap van der Meer[1], Anne Dekinga[1], Casper Kraan[1], Henk W. van der Veer[1] & Theunis Piersma[1,2]

1. Department of Marine Ecology, Royal Netherlands Institute for Sea Research (NIOZ), P.O. Box 59, 1790 AB Den Burg, Texel, The Netherlands
2. Animal Ecology Group, Centre for Ecological and Evolutionary Studies, University of Groningen, P.O. Box 14, 9750 AA Haren, The Netherlands

* Correspondence author. allert@nioz.nl, tel.: +31 (0) 222 369382 and fax: +31 (0) 222 319674

Running title: maximum power for monitoring programmes

Word count: 7,076

Summary

1. Spatial data on animal abundance underpin sound conservation and management advice. The expense of monitoring programs to determine species distributions and estimates of population sizes often limits sample size. For maximum effectiveness at minimal costs, optimisations of such monitoring efforts are critical. A monitoring programme can have multiple objectives with conflicting demands on the optimal sampling design. Here we develop an optimal sampling design for monitoring programmes with such conflicting objectives.

2. We distinguished three possible objectives: (1) estimation of temporal changes and spatial differences in abundance and (2) mapping, i.e. prediction of abundances at unsampled locations. Mapping abundances requires model-based analyses using autocorrelation models. Such analyses are as good as the model fits the data, therefore, an additional objective was (3) accurately estimating autocorrelation model parameters.

3. To compare sampling designs we used the following criteria: (1) the minimum detectable difference in mean between two time periods or two areas, (2) the mean prediction error, and (3) the estimation bias of autocorrelation parameters. Using Monte Carlo simulations we compared five common sampling designs with respect to these criteria at four levels of – naturally occurring – spatial autocorrelation.

4. The optimal sampling designs for objectives (1) and (2) was grid sampling and for objective (3) transect sampling with multiple samples per station and grid sampling with random replacements. The optimal sampling design that catered best for all three objectives combined was grid sampling with a number of random samples placed on

gridlines. This, at 0.5 km intervals between grid sampling-stations, is the optimal sampling design we recommend for the Wadden Sea intertidal flats.

5. *Syntheses and applications*. Grid sampling with additional random sampling is considered an accurate and powerful tool with the largest effectiveness/cost-ratio for monitoring programmes that allows for: (1) estimates of population sizes, (2) monitoring of population trends, (3) comparisons of populations/trends between years or areas, (4) modelling autocorrelation, (5) mapping of species distributions and (6) further understanding of species distribution processes.

Key-words: macrobenthic invertebrates, intertidal, model-based inference, design-based inference, spatial autocorrelation, generalised least squares, power analysis, landscape ecology

Introduction

Spatially explicit data on animal abundances comprise key data for ecologists and are essential for a sound underpinning of conservation and management plans (Underwood, 1997;Krebs, 2001). Often, spatial data are collected with monitoring programmes in which the abundances of one or several species are obtained according to specific sampling designs (Thompson, 1992). Monitoring programmes can have one or more objectives such as monitoring population trends, impact assessment and mapping of species distributions. Being expensive and labour intensive, monitoring programmes are practically constrained by the number of sampling units. With smaller sample sizes the accuracy of the estimates (e.g., population size), and thus the power to detect significant changes, is reduced (Quinn and Keough, 2005).

Sampling units from monitoring programmes and field surveys are separated in space, and such data typically exhibit a degree of spatial autocorrelation, e.g., sampling units closer together are more alike than sampling units further apart (Tobler, 1970;Sokal and Oden, 1978a;Legendre and Fortin, 1989;Legendre *et al.*, 2002). For the analysis of spatially autocorrelated data two statistical frameworks exist: design-based and model-based inference (Gregoire, 1998;Little, 2004). In design-based inference one considers the sampled population as fixed which makes this framework descriptive and useful for estimating quantities from the sample such as the population mean. A requirement for design-based inference is that the sampling units are obtained using a sampling design of probabilistic nature such as simple random sampling. Model-based inference is independent of the sampling design and – contrary to the design-based framework – the population sampled is not regarded as fixed, but as one of many possible realisations of

an underlying process. Using the model-based framework one tries to describe an underlying process which additionally allows for predictions at unsampled locations (Ripley, 1981;Cressie, 1993). The results generated with the two frameworks can differ, but – depending on the sampling design – both can be appropriate for analysing autocorrelated data (Brus and de Gruijter, 1997;Gregoire, 1998;Haining, 2003;Little, 2004). Here we adopted a model-based framework using a spatial autocorrelation model, because we are also interested in predicting species abundance at unsampled locations. Moreover, a model-based approach has the advantage that the focus is on the underlying process instead of on a single realisation of that process.

Spatial autocorrelation is generally modelled as a declining function of Euclidean distance between sampling units (Cliff and Ord, 1981;Upton and Fingleton, 1985). Such autocorrelation functions are fitted to field data and can be used to estimate covariance between sampling units. In ecology one most often observes positive spatial autocorrelation (Legendre and Fortin, 1989) and accounting for positive autocorrelation (i.e. positive covariances) increases variance. An increased variance reduces statistical power for comparisons in, for instance, mean abundance between two populations. On the other hand, autocorrelation is necessary for accurate interpolation of abundances at unsampled locations i.e. mapping (e.g., Koubbi *et al.*, 2006).

The amount of autocorrelation in the data is partly determined by the sampling design, because autocorrelation is a function of distance between sampling units. The optimal distance between sampling units is determined by the objective of the monitoring programme, e.g., small distance between sampling units for mapping species abundances or large for comparisons of abundances between two populations. Some monitoring

programmes have multiple and conflicting objectives regarding the distance between sampling units. In this case the distance between sampling units needs to be optimised between objectives.

The Royal Netherlands Institute for Sea Research (NIOZ) has a long term benthic monitoring program of which the objective is the detection of temporal and spatial changes in abundance from either natural or anthropogenic causes (Piersma *et al.*, 2001;Beukema and Dekker, 2006;van Gils *et al.*, 2006a;Dekker and Beukema, 2007;Kraan *et al.*, 2007;van Gils *et al.*, 2008). In addition, the mapping of macrobenthic invertebrates should allow predictions on the spatial distribution of their predators such as birds, fish and crustaceans (van Gils *et al.*, 2005;van Gils *et al.*, 2006b). In this study, building on the existing benthic monitoring efforts at the NIOZ, we aimed to determine an optimal sampling design for monitoring programmes that have multiple objectives with conflicting ideal sampling designs. Comparisons between years or areas depend on similar principles of analyses and can be combined into one objective. Therefore, we focus on the following objectives: (1) estimation of temporal change and spatial differences in abundance, e.g., the difference in abundance between year or area A and B, and (2) mapping of abundances. Model-based inference is as good as the model fits the data and therefore an additional objective was (3) accurately estimating autocorrelation parameters. Comparisons between sampling designs were based on: (1) the minimum detectable difference in mean between two time periods or areas, (2) the mean prediction error and (3) the estimation bias, i.e. the difference in simulated and estimated autocorrelation parameters. With respect to these criteria we compared five sampling designs which are regularly used.

Methods

GENERAL APPROACH

Using field data, realistic autocorrelation model parameters were estimated and four extreme autocorrelation models selected. These autocorrelation models were then used to simulate autocorrelated data according to different sampling designs and compared regarding the above criteria.

FIELD DATA

From 1996, building on a tradition of station-intensive and transect-based monitoring (Beukema, 1976;Beukema and Dekker, 2006;Dekker and Beukema, 2007), the Royal Netherlands Institute for Sea Research (NIOZ) has monitored population densities of macrobenthic invertebrates across 225 km^2 of intertidal mudflats in the western Dutch Wadden Sea (Piersma *et al.*, 2001). Between July and September each year, one sample was taken at between 1807 and 2762 stations in order to achieve large statistical power (van der Meer, 1997). The sample stations were arranged according to a grid sampling design with 0.25 km inter-sample distance. Sampling stations were located by handheld GPS. At each station a core with a surface area of 1/56 m^2 to a depth of 20-25 cm was collected, washed over a 1 mm mesh sieve and numbers of each species were counted. To allow for a comparison between two groups (objectives 1 and 2), the analyses were based on the difference in (numerical) densities between two successive years (2005 and 2006) and restricted to the five most abundant bivalve (*Cerastoderma edule*, *Macoma balthica*,

Mya arenaria, *Abra tenuis* and *Ensis americanus*) and worm species (*Scoloplos armiger*, *Heteromastus filiformis*, *Nereis diversicolor*, *Nephtys hombergii* and *Lanice conchilega*).

STATISTICAL FRAMEWORK

The Generalised Least Squares (GLS) method is a model-based analysis for spatially autocorrelated data as well as for spatial predictions necessary for the three objectives. GLS is widely used in spatial statistics (Cressie, 1993) and spatial ecology (see Dormann *et al.*, 2007). Spatial GLS assumes that autocorrelation (i.e. covariance) is a function of Euclidean distance between sampling units (Cliff and Ord, 1981;Upton and Fingleton, 1985) and fits such a spatial autocorrelation function (SAF) to field data in order to estimate covariance between sampling units.

Autocorrelation, expressed as the commonly used Moran's I, was calculated for discrete distance classes into a correlogram (Sokal and Oden, 1978a;Cliff and Ord, 1981;Legendre and Fortin, 1989). A SAF was fitted to the correlogram according to van der Meer & Leopold (1995):

$$AC(h) = \begin{cases} b_0 e^{b_1 h} & \text{if } h > 0 \\ 1 & \text{if } h = 0 \end{cases}$$

Autocorrelation AC was fitted as a continuous function of distance h with b_0 being the autocorrelation for distances close to zero (local autocorrelation) and b_1 denoting the decline in autocorrelation with distance (inversely related to the range of autocorrelation). Autocorrelation at distance zero is 1 by definition and therefore omitted for estimation of b_0 and b_1. The autocorrelation model was fitted to the distance matrix – which gives pair wise distances between all sampling units – and multiplied by the variance of the

response variable σ^2 to obtain an estimate of the variance-covariance matrix Σ (e.g., van der Meer and Leopold, 1995).

SAMPLING DESIGNS

Five designs were compared: (1) simple random sampling, (2) grid sampling, (3, 4) transect sampling (with one or with five sampling units per station respectively) and (5) grid sampling with random replacements. (1) Simple random sampling is the most common sampling method in ecology (Fig. 1a) and often combined with stratified sampling (e.g., Armonies and Reise, 2003). (2) For grid sampling, sampling stations are usually equally spaced in a lattice (e.g., Herman *et al.*, 2001) and, in this study, located in the centre of a grid cell (Fig. 1b). (3) The transect sampling design (Fig. 1c) consisted of transects with random starting locations and a random heading in which 9 additional stations were equally spaced (comparable to Beukema, 1976;Yates *et al.*, 1993). (4) Transect sampling with multiple sampling units is a design similar to transect sampling, but at each of 10 transect sampling stations an additional four sampling units were taken within 400 m^2 (comparable to Beukema, 1974). (5) Grid sampling with random replacements is based on the "lattice plus closed pair design" by Diggle & Lophaven (2006). Similar to grid sampling, sampling units are equally spaced on a grid, but 10% of these stations were replaced to a random position on both a vertical and horizontal gridline (Fig. 1d). *Replaced* instead of added to maintain equal sample sizes for between sampling design comparison, and replaced *onto gridlines*, because sampling stations are hereby more easily located in the field than is the case for completely random locations,

while maintaining some of the statistical advantages of random sampling (Diggle and Lophaven, 2006).

DATA SIMULATION

On a 10 x 10 km surface area, sampling stations were selected according to the different sampling designs. The distance between sampling stations (inter-sample distance) was 0.25, 0.5, 0.75 and 1 km, leading to sample sizes of 1681, 441, 196 and 121 respectively. This coincided with an expected averaged distance between sampling units of 0.12, 0.24, 0.36 and 0.45 km for simple random sampling (Clarke and Evans, 1954). At a given inter-sample distance, designs have different sample sizes. To compare power of sampling designs for each inter-sample distance, sampling designs were restrained to the sample size of grid sampling. For example, at an inter-sample distance of 1 km the sample size of grid sampling consisted of $11 \cdot 11 = 121$ sampling units. The sample size of transect sampling is a multiple of the length of one transect (i.e. 10 sampling units). To maintain equal sample sizes we truncated the last transect so the total sample size equalled that of grid sampling. Sample stations were simulated on the 100 km^2 surface area + $0.5 \cdot$ inter-sample distances, wherefore the grid sampling stations were located in the centre of a grid cell. Sample stations were restricted to this surface area, e.g., starting locations of transects were reassigned if any sample station would reach beyond this surface area. Therefore, diagonal transects are more likely to occur than transects parallel to the gridlines (Fig. 1c). This sampling bias will be large if the surface area is small relative to the inter-sample distance (Thompson, 1992). With an inter-sample distance of 1 km, for instance, the length of transects would measure the entire 10 km width or length

of the surface area. In the field this bias also occurs, and as we were interested in field implications of different sampling designs, it was accepted as realistic.

The variance-covariance matrix Σ was calculated with distance between sampling units using four extreme, but naturally occurring, levels of autocorrelation, i.e. spatial autocorrelation functions. Based on field data estimates of autocorrelation parameters, we modelled either weak or strong local autocorrelation (b_0) together with either a shallow or steep decline in autocorrelation with distance (b_1). Each of the four possible combinations of b_0 and b_1 were examined. Spatially autocorrelated response variables were simulated for each sampling design and inter-sample distance using Choleski decomposition (Cressie, 1993;Dormann *et al.*, 2007). A weight matrix W was derived from the variance-covariance matrix $\Sigma = W^T W$, and normally distributed, spatially autocorrelated response variables were then calculated by $\varepsilon = W^T \xi$ with ξ drawn from the standard normal distribution ($\mu = 0$ and $\sigma^2 = 1$).

COMPARISON CRITERIA OF SAMPLING DESIGNS

The minimum detectable difference (MDD) between two populations (objective 1) was calculated with the variance of the mean se: $MDD = \sqrt{se} \cdot (t_{\alpha,df} + t_{\gamma,df})$ and $\alpha = 0.05$ and $\gamma = 0.20$, i.e. the minimum detectable difference 80% of the time at a significance level of 0.05 (Quinn and Keough, 2005). The mean and variance of the mean were calculated with GLS following Cliff & Ord (1981). For detailed calculations see appendix (Appendix S1 in Supplementary Material). For comparison with design-based inference (where the existence of auto-correlation is basically irrelevant), we additionally calculated the mean and variance of the mean using ordinary least squares (OLS). This

corresponds to a GLS analyses with $b_0 = 0$ and $b_1 = 0$. Additionally, the relative number of independent data points in the autocorrelated sample (i.e. percentage effective sample size n*, Griffith, 2005) was estimated by dividing OLS- through GLS-variance.

A common method for spatial predictions at unsampled locations is kriging (see Ripley, 1981;Upton and Fingleton, 1985;Cressie, 1993;Haining, 2003). For objective (2) we calculated the mean prediction error using ordinary kriging with $Y = \mu + Z(h) + \varepsilon$ where Y is the interpolated response variable, μ is the overall mean, $Z(h)$ is a Gaussian stochastic process with mean zero and estimated variance-covariance Σ, and residual variance ε. In effect, the kriging interpolation Y is equal to the mean plus a value weighed by Σ. Details on the kriging calculations are available elsewhere (Ripley, 1981;Cressie, 1993;Fortin and Dale, 2005;Nychka, 2007).

For objective (3) we simulated autocorrelated data at the four autocorrelation levels and calculated the difference with the estimated autocorrelation parameter values after fitting the SAF i.e. estimation bias. The SAF was fitted over 2/3 of the maximum distance between pairs of sample units and the width of the distance classes was 1/3 of the inter-sample distance, hereby, the sample size per distance class was at least 10. Autocorrelation parameters were not estimable when the SAF could not be fitted or estimates of $b_0 > 2$, $b_1 > 0$ and $b_1 < -10$.

All analyses followed Monte Carlo simulations in which the above criteria were averaged over 1,000 runs. The estimation of the mean prediction error was calculated based on 200 rather than 1,000 runs, because of time consuming calculations and small Monte Carlo variance in the mean prediction error. For each run we calculated the mean

prediction error from 100 randomly chosen locations on the 100 km^2 simulated surface area.

SOFTWARE

All calculations and simulations were performed with R v2.6 (R-Development-Core-Team, 2008). Conversion of longitude and latitude to UTM coordinates were done with the package *PBSmapping* (Schnute *et al.*, 2008), the analysis of spatial data with *ncf* (Bjornstad, 2006) and *spatstat* (Baddeley and Turner, 2005) and kriging with *fields* (Nychka, 2007). See Appendix for the R calculations (Appendix S1).

Results

On the basis of 2,695 sampling stations covered both in 2005 and 2006, (numerical) density differences between years could be calculated. The data consisted of many zeros and were therefore not normally distributed. There are no transformation routines that adequately normalize the data, but sample sizes were large enough for the effect of non-normality to be small. Moreover, many zero counts do not change the pattern of the correlogram (Bergström *et al.*, 2002).

FIELD DATA

For each species, σ^2 was estimated and b_0 and b_1 were estimated from a correlogram (Fig. 2a). Parameter estimates for b_0 ranged from 0.03 to 0.66 and for b_1 from –3.12 to –0.34 (Table 1). The mean density differences between 2005 and 2006 for design-based (where the analysis is numerically equivalent to OLS) and model-based inference (GLS) were

similar, but as predicted, SE's were smaller for OLS than GLS (Table 1). Depending on the level of autocorrelation, the relative effective sample size (percentage of independent data points, n*) ranged from 3% to 28% (Table 1). MDD for OLS varied from 0.9 to 18.8 m^{-2} compared to 2.0 to 62.7 m^{-2} for GLS. Seven out of ten species showed a significant difference in densities between years for OLS compared to two out of ten for GLS (i.e. *N. hombergii* and *L. Conchilega,* Table 1).

SIMULATED DATA

Based on field estimates (Table 1), we used $b_0 = 0.1$ or $b_0 = 0.5$ and $b_1 = –0.5$ or $b_1 = –3$ (Fig. 2b) to simulate different levels of spatially autocorrelated data. The combinations of autocorrelation parameters approximated *C. edule* ($b_0 = 0.32$, $b_1 = –0.76$; strong local autocorrelation, long range of autocorrelation), *A. tenuis* ($b_0 = 0.66$, $b_1 = –3.12$; strong local autocorrelation, short range), *H. filiformis* ($b_0 = 0.13$, $b_1 = –0.58$; weak local autocorrelation, long range). None of the selected species showed the combination of weak local autocorrelation and a short range.

SIMULATED DATA: MDD

The level of autocorrelation decreased with increased inter-sample distance, because sampling units were increasingly outside each other's autocorrelation range. Nonetheless, the decrease in MDD (i.e. increased power) with inter-sample distance was outweighed by the increase in MDD caused by reduced sample sizes. Therefore, MDD increased for all sampling designs as inter-sample distance increased (Fig. 3). Grid sampling allowed the smallest MDD for most inter-sample distances. Simple random and grid sampling

with random replacements also provided relatively small MDD. Transect sampling and especially transect sampling with multiple sampling units consistently showed a larger MDD compared with the other sampling designs. Between autocorrelation levels, strong local autocorrelation (Fig. 3a-b) resulted in a large MDD compared to weak local autocorrelation (Fig. 3c-d). Additionally, a long range of autocorrelation (Fig. 3a and 3c) resulted in a large MDD compared to a short range (Fig. 3b and 3d). The differences in MDD between sampling designs were more pronounced for strong local autocorrelation over a short range (Fig. 3b).

SIMULATED DATA: KRIGING

Sample size and the level of autocorrelated data were reduced with an increase in inter-sample distance, and therefore, the prediction error increased with inter-sample distance (Fig. 4). With decreased autocorrelation, kriging interpolations became less accurate and the prediction error more or less approached the simulated variance of 1 (Fig. 4c-d). Grid sampling allowed smallest prediction errors for all inter-sample distances (Fig. 4a-d), followed by respectively grid sampling with random replacements, simple random sampling, transect sampling and transect sampling with multiple sampling units. Between autocorrelation levels, strong local autocorrelation (Fig. 4a-b) resulted in small prediction errors compared to weak local autocorrelation (Fig. 4c-d). Additionally, a long range of autocorrelation (Fig. 4a and 4c) resulted in small prediction errors compared to a short range of autocorrelation (Fig. 4b and 4d).

SIMULATED DATA: AUTOCORRELATION PARAMETER ESTIMATES

The smaller the level of autocorrelated data the less often the autocorrelation parameters were estimable (Fig. 5). Moreover, with an increase in inter-sample distance (i.e. reduced levels of autocorrelated data) the autocorrelation parameters were more often inestimable than with small inter-sample distances (Fig. 5).

In case the SAF was fitted, the estimate of local autocorrelation (b_0) was more accurate the smaller the sampling distance (Fig. 6). As inter-sample distance increased b_0 was overestimated using most sampling designs. Transect sampling with multiple sampling units was the most accurate for estimating b_0 (which was more pronounced for small b_0, Fig. 6 and Table 2), because multiple sampling units were taken within a small range. Transect sampling with one sample, random sampling and grid sampling with random replacements also showed small estimation bias (Fig 6 and Table 2). This was especially so for small inter-sample distances and large b_0. Grid sampling showed the largest bias, because the smallest distance over which b_0 can be estimated is the inter-sample distance (Fig. 6 and Table 2). Between autocorrelation levels, large b_0 (Fig. 6a-b and Table 2a-b) resulted in more accurate estimates compared to small b_0 (Fig. 6c-d and Table 2c-d). The steepness of the decline in autocorrelation with distance appeared to have little effect on the estimation of b_0 (Fig. 6 and Table 2). However, with a small range the SAF was less often fitted (Fig. 5).

With smaller inter-sample distances the estimation bias of the decline in autocorrelation with distance (b_1) was smaller (Fig. 7 and Table 2). Grid sampling with random replacements was the most accurate in estimating b_1 followed by random sampling and grid sampling (Fig. 7 and Table 2). Both transect sampling designs showed the largest bias. A large b_1 resulted in smaller estimation bias (Fig. 7 and Table 2).

Discussion

DESIGN- AND MODEL-BASED INFERENCE

The analysis of autocorrelated data without taking the autocorrelation into account is considered to be inappropriate (Legendre and Fortin, 1989;Legendre, 1993;Dale and Fortin, 2002;Legendre *et al.*, 2002;Liebhold and Gurevitch, 2002;Wagner and Fortin, 2005). Spatial autocorrelation causes spatial pseudoreplication which violates the assumption of independent error terms, because only a proportion of the sample consists of non-autocorrelated independent data points, i.e. the 'effective sample size' (Griffith, 2005). This violation, however, is a misconception (Brus and de Gruijter, 1997;Gregoire, 1998;Dorazio, 1999;Little, 2004). In design-based inference independence has a different meaning and is determined by the stochastic nature of the sampling design, whereas, in model-based inference the independence is determined by the postulated model (Brus and de Gruijter, 1997). Like model-based inference, design-based inference can be appropriate (i.e. if the assumption of stochastic sampling is met) for the analyses of autocorrelated data, but – as was shown here – the results between the two can differ.

In our study, the estimated mean of design-based (OLS) and model-based (GLS) inference were similar, but significance levels differed. Note that we have neglected the fact that the data was collected using a grid sampling design and not a formal random sampling design. Hence a regularity in the data could have existed which resembled the regularity in the grid and would bias OLS estimates. Though such resemblance cannot be ruled out, it is unlikely. OLS analysis revealed that seven species significantly changed in abundance between years. By contrast, GLS-analysis revealed that two species showed a

significant change in abundance between years. Both analyses are correct, but fundamentally differ in meaning. From OLS analysis we conclude that the observed changes in abundance between years are significant for seven species. With GLS analysis we can conclude that for two out of ten species the underlying process that generated changes in population numbers were significantly different. Because both frameworks differ in their results, it should be clear which framework has been chosen and which hypotheses were tested. Moreover, significance from design-based inference should not lead to conclusions on superpopulation level and *vice versa*.

The main advantage of the design-based framework for analysing spatially autocorrelated data is that no model assumptions of the underlying process are necessary. This contrasts with a model-based framework, for which the analyses are as good as the assumed models fit the data. We adopted a model-based framework, because it allows for more accurate predictions at unsampled locations (i.e. mapping) than design-based inference for which the best prediction is the population mean (Ripley, 1981;Cressie, 1993). Additional advantages are that model-based inference is independent of the sampling design (i.e. allows for sparse sampling) and that the (autocorrelation) model can provide additional biological information (Sokal and Oden, 1978b). Spatial autocorrelation can be caused by exogenous and endogenous processes or a combination of these (Fortin and Dale, 2005;Wagner and Fortin, 2005). Exogenous processes are independent of the variable of interest (e.g., environmental variables) and endogenous processes are caused by the biology of the variable of interest such as dispersal (Lagos *et al.*, 2007) and predation (Klaassen and Nolet, 2008). A model-based framework to analysing spatial data allows quantification of autocorrelation, the possibility of

distinguishing exogenous and endogenous processes (e.g., Kraan *et al.* under review-a) and understanding of the mechanisms behind the observed spatial distribution (e.g., Bergström *et al.*, 2002;Klaassen *et al.*, 2006;de Frutos *et al.*, 2007: Kraan *et al.* under review-b).

OPTIMAL SAMPLING DESIGN

Low levels of autocorrelation resulted in large power to detect changes between years or areas (objective 1). This suggests that largest power is obtained if the inter-sample distance exceeds the autocorrelation range to minimise autocorrelation in the data i.e. maximise the effective sample size. The opposite is true for predicting values at unsampled locations (objective 2) where low levels of autocorrelation resulted in increased prediction error. A trade-off between objectives exists. Nonetheless, the optimal sampling design for both objectives was grid sampling which revealed largest power for objective (1) and the smallest prediction error for objective (2). Grid sampling was the optimal sampling design for objective (1), because no samples were closer together than the inter-sample distance which reduced autocorrelation in the data. And optimal for objective (2), because it is *surface-covering* and therefore satisfies the uniformity condition necessary for accurate kriging (Pooler and Smith, 2005;Marchant and Lark, 2007). Additionally, other sampling designs showed 'holes' in the sampled surface area (Fig. 1). In these holes the prediction error was largest which increased mean prediction error even though these designs showed higher levels of autocorrelated data.

Grid sampling seemed the optimal sampling design for conflicting objectives (1) and (2). However, note that in our study we simulated autocorrelated data with known

autocorrelation parameters. In the analysis of field data these parameters need to be estimated from the data itself and how well they fit the data determines the validity of model-based inference (Gregoire, 1998;Haining, 2003;Little, 2004). Grid sampling provided the largest estimation bias for autocorrelation parameters, opposed to transect sampling with multiple samples and grid sampling with random replacements, which revealed smallest autocorrelation estimation bias. Opposite grid sampling the latter designs include small inter-sample distances which allow for accurate estimates of autocorrelation parameters (e.g., Diggle and Lophaven, 2006). Transect sampling with multiple samples was suboptimal for objectives (1) and (2), but grid sampling with random replacements performed well on all objectives: similar MDD (objective 1) and prediction error (objective 2) as grid sampling, but with more accurate estimates of autocorrelation parameters (objective 3). Therefore, grid sampling with random replacements is the optimal sampling design for monitoring programmes with similar objectives.

In this study, we *moved* 10% of grid sample stations to randomly selected sample positions on gridlines to maintain equal sample sizes for correct comparisons between sampling designs. Therefore, we lost homogenous surface coverage which increased the prediction error. The constraint of equal sample size does not apply in the field and, therefore, the optimal sampling design would be surface-covering grid sampling with a percentage (e.g., 10%) of sampling stations randomly placed on gridlines *additional* to the grid design. The grid sampling allows for large statistical power in comparisons between years or areas as well as small prediction errors at unsampled locations and the *additional* random sampling allows for accurate estimates of autocorrelation parameters.

The larger the percentage of random points the more accurate the estimates of autocorrelation parameters.

IMPLICATIONS FOR WADDEN SEA MONITORING PROGRAMMES

Currently, the NIOZ macrobenthic monitoring programmes follow either transect sampling (Beukema, 1976;Beukema and Dekker, 2006;Dekker and Beukema, 2007), or non-surface covering grid sampling with a inter-sample distance of 0.25 km (Piersma *et al.*, 2001;van Gils *et al.*, 2006a;van Gils *et al.*, 2006b;Kraan *et al.*, 2007;van Gils *et al.*, 2008). This study indicates that surface-covering grid sampling with additional random sampling is the optimal sampling design for detecting temporal and spatial changes in abundance as well as the mapping of macrobenthic invertebrates across the entire Dutch Wadden Sea. Given the surface area of the Dutch Wadden Sea, sampling at 0.25 km would inflate sample size beyond what is feasible within seasonal and logistical constraints. We, therefore, suggest the inter-sample distance should be increased to 0.50 km to allow surface-coverage of the entire western Dutch Wadden Sea according a grid sampling design with additional random samples.

Acknowledgements

We thank all volunteers who helped collecting the samples over the years and the crew of MS *Navicula* whom provided a helpful and welcoming atmosphere on board which made the benthic surveys so much more enjoyable. We thank Geert Aarts and Peter Herman for useful comments before submission. We additionally thank Hans Malschaert and Piet

Ruardij for the use of a *biocluster* supercomputer and the Nederlandse Aardolie Maatschappij (NAM) for financing AIB.

References

Armonies, W. & Reise, K. (2003) Empty habitat in coastal sediments for populations of macrozoobenthos. *Helgoland Marine Research,* **56,** 279–287.

Baddeley, A. & Turner, R. (2005) Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software,* **12,** 1–42.

Bergström, U., Englund, G. & Bonsdorff, E. (2002) Small-scale spatial structure of Baltic Sea zoobenthos - inferring processes from patterns. *Journal of Experimental Marine Biology and Ecology,* **281,** 123–136.

Beukema, J. J. (1974) Seasonal changes in the biomass of the macro-benthos of a tidal flat area in the Dutch Wadden Sea. *Netherlands Journal of Sea Research,* **8,** 94–107.

Beukema, J. J. (1976) Biomass and species richness of the macro-benthic animals living on the tidal flats of the Dutch Wadden Sea. *Netherlands Journal of Sea Research,* **10,** 236–261.

Beukema, J. J. & Dekker, R. (2006) Annual cockle *Cerastoderma edule* production in the Wadden Sea usually fails to sustain both wintering birds and a commercial fishery. *Marine Ecology-Progress Series,* **309,** 189–204.

Bjornstad, O. N. (2006) *Spatial nonparametric covariance functions.* http://onb.ent.psu.edu/onb1/R.

Brus, D. J. & de Gruijter, J. J. (1997) Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). *Geoderma,* **80,** 1–44.

Clarke, P. J. & Evans, F. C. (1954) Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology,* **35,** 445–453.

Cliff, A. D. & Ord, J. K. (1981) *Spatial Processes: Models and Applications*. Pion Ltd., London.

Cressie, N. A. C. (1993) *Spatial Statistics,* revised edn. John Wiley & Sons, Inc., New York.

Dale, M. R. T. & Fortin, M. J. (2002) Spatial autocorrelation and statistical tests in ecology. *Ecoscience,* **9,** 162–167.

de Frutos, A., Olea, P. P. & Vera, R. (2007) Analyzing and modelling spatial distribution of summering lesser kestrel: the role of spatial autocorrelation. *Ecological Modelling,* **200,** 33–44.

Dekker, R. & Beukema, J. J. (2007) Long-term and large-scale variability in productivity of the tellinid bivalve *Macoma balthica* on Wadden Sea tidal flats. *Marine Ecology-Progress Series,* **337,** 117–134.

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

Dorazio, R. M. (1999) Design-based and model-based inference in surveys of freshwater mollusks. *Journal of the North American Benthological Society,* **18,** 118–131.

Dormann, C. F., McPherson, J. M., Araujo, M. B., Bivand, R., Bolliger, J., Carl, G., Davies, R. G., Hirzel, A., Jetz, W., Kissling, W. D., Kuhn, I., Ohlemuller, R.,

Peres-Neto, P. R., Reineking, B., Schroder, B., Schurr, F. M. & Wilson, R. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography,* **30,** 609–628.

Fortin, M. J. & Dale, M. R. T. (2005) *Spatial Analysis: A Guide for Ecologists*. Cambridge University Press, Cambridge.

Gregoire, T. G. (1998) Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research,* **28,** 1429–1447.

Griffith, D. A. (2005) Effective geographic sample size in the presence of spatial autocorrelation. *Annals of the Association of American Geographers,* **95,** 740–760.

Haining, R. (2003) *Spatial Data Analysis: Theory and Practice*. Cambridge University Press, Cambridge.

Herman, P. M. J., Middelburg, J. J. & Heip, C. H. R. (2001) Benthic community structure and sediment processes on an intertidal flat: results from the ECOFLAT project. *Continental Shelf Research,* **21,** 2055–2071.

Klaassen, R. H. G. & Nolet, B. A. (2008) Persistence of spatial variance and spatial pattern in the abundance of a submerged plant. *Ecology,* **89,** 2973–2979.

Klaassen, R. H. G., Nolet, B. A., Van Gils, J. A. & Bauer, S. (2006) Optimal movement between patches under incomplete information about the spatial distribution of food items. *Theoretical Population Biology,* **70,** 452–463.

Koubbi, P., Loots, C., Cotonnec, G., Harlay, X., Grioche, A., Vaz, S., Martin, C., Walkey, M. & Carpentier, A. (2006) Spatial patterns and GIS habitat modelling of

Solea solea, *Pleuronectes flesus* and *Limanda limanda* fish larvae in the eastern English Channel during the spring. *Scientia Marina,* **70,** 147–157.

Kraan, C., Piersma, T., Dekinga, A., Koolhaas, A. & van der Meer, J. (2007) Dredging for edible cockles (*Cerastoderma edule*) on intertidal flats: short-term consequences of fisher patch-choice decisions for target and non-target benthic fauna. *ICES Journal of Marine Science,* **64,** 1735–1742.

Krebs, C. J. (2001) *Ecology*. Addison Wesley Longman, San Francisco.

Lagos, N. A., Tapia, F. J., Navarrete, S. A. & Castilla, J. C. (2007) Spatial synchrony in the recruitment of intertidal invertebrates along the coast of central Chile. *Marine Ecology-Progress Series,* **350,** 29–39.

Legendre, P. (1993) Spatial autocorrelation - trouble or new paradigm. *Ecology,* **74,** 1659–1673.

Legendre, P., Dale, M. R. T., Fortin, M. J., Gurevitch, J., Hohn, M. & Myers, D. (2002) The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography,* **25,** 601–615.

Legendre, P. & Fortin, M. J. (1989) Spatial pattern and ecological analysis. *Vegetatio,* **80,** 107–138.

Liebhold, A. M. & Gurevitch, J. (2002) Integrating the statistical analysis of spatial data in ecology. *Ecography,* **25,** 553–557.

Little, R. J. (2004) To model or not to model? Competing modes of inference for finite population sampling. *Journal of the American Statistical Association,* **99,** 546–556.

Marchant, B. P. & Lark, R. M. (2007) Optimized sample schemes for geostatistical surveys. *Mathematical Geology,* **39,** 113–134.

Nychka, D. (2007) *fields: tools for spatial data.* http://cran.r-project.org/web/packages/fields/index.html.

Piersma, T., Koolhaas, A., Dekinga, A., Beukema, J. J., Dekker, R. & Essink, K. (2001) Long-term indirect effects of mechanical cockle-dredging on intertidal bivalve stocks in the Wadden Sea. *Journal of Applied Ecology,* **38,** 976–990.

Pooler, P. S. & Smith, D. R. (2005) Optimal sampling design for estimating spatial distribution and abundance of a freshwater mussel population. *Journal of the North American Benthological Society,* **24,** 525–537.

Quinn, G. P. & Keough, M. J. (2005) *Experimental Design and Data Analysis for Biologists,* 4th edn. Cambridge University Press, Cambridge.

R-Development-Core-Team (2008) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Ripley, B. D. (1981) *Spatial Statistics*. John Wiley & Sons, Inc., New York.

Schnute, J., Boers, N. & Haigh , R. (2008) *PBSmapping version 2.55*. http://cran.r-project.org/web/packages/PBSmapping/index.html.

Sokal, R. R. & Oden, N. L. (1978a) Spatial autocorrelation in biology 1. Methodology. *Biological Journal of the Linnean Society,* **10,** 199–228.

Sokal, R. R. & Oden, N. L. (1978b) Spatial autocorrelation in biology 2. Some biological implications and four applications of evolutionairy and ecological interest. *Biological Journal of the Linnean Society,* **10,** 229–249.

Thompson, S. K. (1992) *Sampling*. John Wiley & Sons, Inc., New York.

Tobler, W. (1970) A computer movie simulating urban growth in the Detroit region. *Economic Geography,* **46,** 234–240.

Underwood, A. J. (1997) *Experiments in Ecology. Their Logical Design and Interpretation Using Analysis of Variance.* Cambridge University Press, Cambridge.

Upton, G. J. G. & Fingleton, B. (1985) *Spatial Data Analysis by Example.* John Wiley & Sons, Inc., New York.

van der Meer, J. (1997) Sampling design of monitoring programmes for marine benthos: a comparison between the use of fixed versus randomly selected stations. *Journal of Sea Research,* **37,** 167-179.

van der Meer, J. & Leopold, M. F. (1995) Assessing the population-size of the European storm-petrel (*Hydrobates pelagicus*) using spatial autocorrelation between counts from segments of crisscross ship transects. *ICES Journal of Marine Science,* **52,** 809–818.

van Gils, J. A., Dekinga, A., Spaans, B., Vahl, W. K. & Piersma, T. (2005) Digestive bottleneck affects foraging decisions in red knots *Calidris canutus*. II. Patch choice and length of working day. *Journal of Animal Ecology,* **74,** 120–130.

van Gils, J. A., Kraan, C., Dekinga, A., Koolhaas, A., Drent, J., de Goeij, P. & Piersma, T. (2008) Reversed optimality and predictive ecology: burrowing depth forecasts population change in a bivalve. *Biology Letters***,** doi: 10.1098/rsbl.2008.0452.

van Gils, J. A., Piersma, T., Dekinga, A., Spaans, B. & Kraan, C. (2006a) Shellfish dredging pushes a flexible avian top predator out of a marine protected area. *PLoS Biology,* **4,** 2399–2404.

van Gils, J. A., Spaans, B., Dekinga, A. & Piersma, T. (2006b) Foraging in a tidally structured environment by red knots (*Calidris canutus*): Ideal, but not free. *Ecology,* **87,** 1189–1202.

Wagner, H. H. & Fortin, M. J. (2005) Spatial analysis of landscapes: concepts and statistics. *Ecology,* **86,** 1975–1987.

Yates, M. G., Gosscustard, J. D., McGrorty, S., Lakhani, K. H., Durell, S., Clarke, R. T., Rispin, W. E., Moy, I., Yates, T., Plant, R. A. & Frost, A. J. (1993) Sediment characteristics, invertebrate densities and shorebird densities on the inner banks of the Wash. *Journal of Applied Ecology,* **30,** 599–614.

Supplementary material

The following supplementary material is available online from www.Blackwell-Synergy.com:

Appendix S1. R-code of calculations.

Table 1. Results from changes in macrobenthic invertebrate densities using design- and model-based analyses of field data. Estimates from design-based inference (OLS) and model-based inference (GLS) are presented for density changes between 2005 and 2006. For each species are given: local autocorrelation b_0, steepness of decline in autocorrelation with distance b_1, mean density change (m^{-2}), standard error of the mean (SE; m^{-2}), the minimum detectable density difference (MDD; m^{-2}), and the percentage effective sample size n*.

			OLS			GLS			
SPECIES	***b 0***	***b 1***	**mean (m^{-2})**	**SE**	**MDD**	**mean (m^{-2})**	**SE**	**MDD**	**n* (%)**
*Cerastoderma edule**	0.32	-0.76	-31.5	3.01	8.4	-21.8	13.85	38.8	5
*Macoma balthica**	0.05	-0.50	-4.2	1.39	3.9	-3.2	3.85	10.8	13
*Mya arenaria**	0.05	-0.34	-6.1	1.23	3.5	-5.3	4.30	12.1	8
*Abra tenuis**	0.66	-3.12	19.7	6.71	18.8	16.3	15.32	42.9	19
Ensis americanus	0.03	-0.42	0.4	0.31	0.9	0.1	0.72	2.0	18
*Scoloplos armiger**	0.21	-0.40	-27.0	3.90	10.9	-10.1	22.37	62.7	3
Heteromastus filiformis	0.13	-0.58	-6.0	5.47	15.3	-3.1	20.02	56.1	7
Nereis diversicolor	0.50	-2.11	8.0	3.78	10.6	6.0	10.26	28.8	14
*Nephtys hombergii***	0.38	-3.02	14.3	1.40	3.9	14.1	2.66	7.5	28
*Lanice conchilega***	0.23	-1.29	-24.5	3.58	10.0	-27.4	9.75	27.3	13

** significantly different from zero with OLS*

*** significantly different from zero with OLS and GLS*

Table 2. Estimation bias of autocorrelation parameters. The difference (in %) is given between the simulated and estimated local autocorrelation (Δb_0) and decline in autocorrelation with distance (Δb_1). The sampling designs are: transect sampling with either multiple (Transect M.) or a single sample per station (Transect), simple random sampling (Random), grid sampling with random replacements (Grid Rand.) and grid sampling (Grid). Tables A-D represent different levels of autocorrelation: (A) strong local autocorrelation and a long range of autocorrelation, (B) strong local autocorrelation and a short range, (C) weak local autocorrelation and a long range and (D) weak local autocorrelation and a short range.

		A		B		C		D	
	simulated value:	b_0=0.5	b_1=-0.5	b_0=0.5	b_1=-3	b_0=0.1	b_1=-0.5	b_0=0.1	b_1=-3
Sampling design	Sample distance (km)	Δb_0 (%)	Δb_1 (%)	Δb_0 (%)	Δb_1 (%)	Δb_0 (%)	Δb_1 (%)	Δb_0 (%)	Δb_1 (%)
Transect.M	0.25	-4	-128	1	-14	-6	-140	4	-19
	0.5	-8	-175	4	-24	-5	-202	10	-12
	0.75	-8	-209	2	-16	14	-217	26	9
	1	-6	-227	1	-1	22	-209	24	27
Transect	0.25	-3	-108	1	-7	-5	-125	11	-18
	0.5	-3	-131	8	-15	31	-227	70	-20
	0.75	-1	-151	22	-17	131	-294	191	-15
	1	9	-158	19	0	210	-321	244	2
Random	0.25	2	-86	2	-6	-1	-107	13	-18
	0.5	5	-90	14	-17	36	-168	85	-19
	0.75	9	-109	20	-11	162	-273	219	-11
	1	17	-117	17	6	233	-273	308	2
Grid Rand.	0.25	1	-88	1	-4	4	-111	15	-11
	0.5	4	-91	13	-16	55	-193	85	3
	0.75	11	-108	18	-3	162	-247	148	27
	1	19	-108	15	17	229	-208	221	35
Grid	0.25	11	-99	7	-8	7	-114	37	-19
	0.5	23	-118	27	-3	67	-165	160	8
	0.75	55	-148	-9	30	167	-167	229	29
	1	82	-136	-8	44	276	-149	757	16

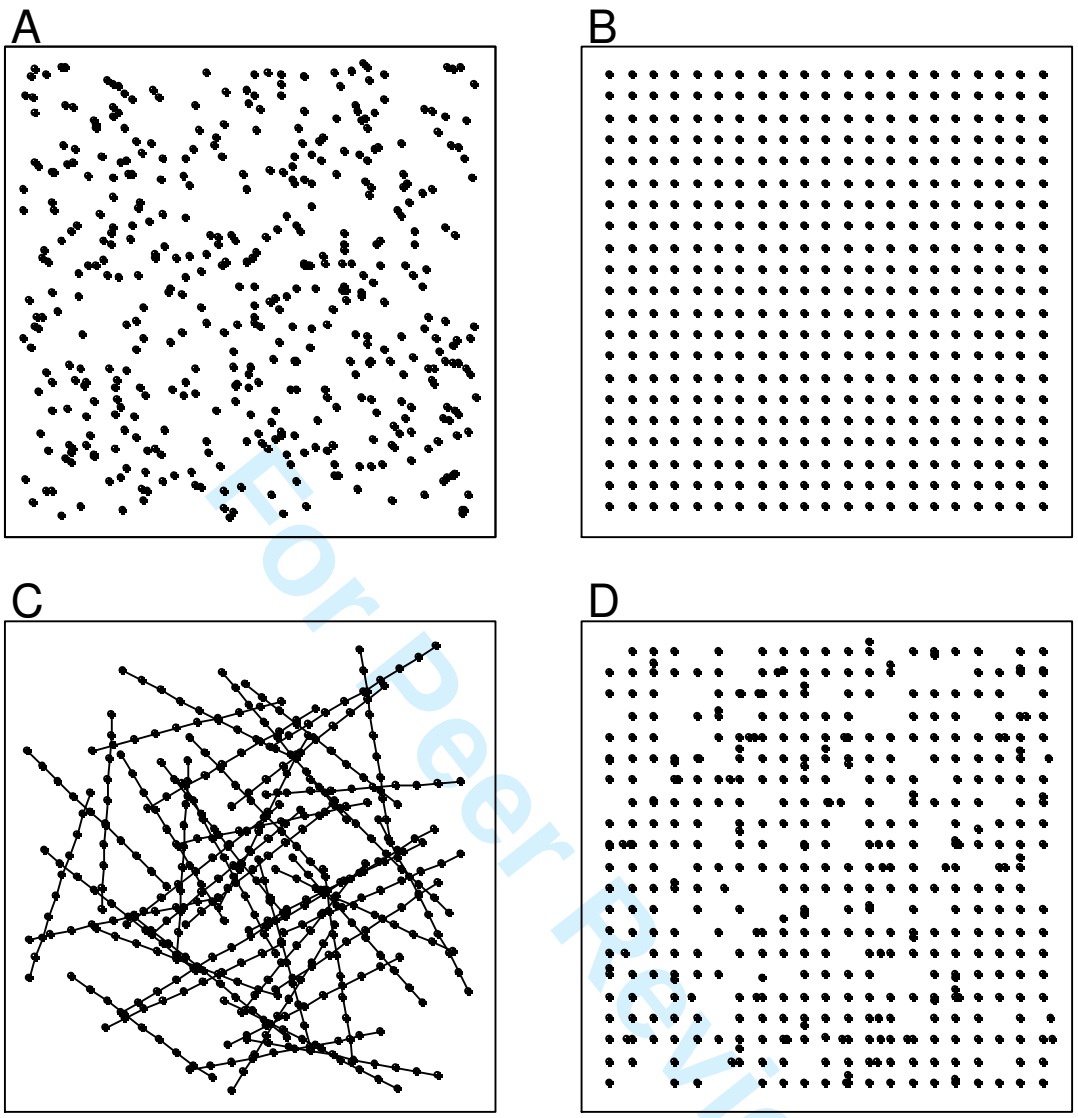

Fig. 1. The different sampling designs compared in this study. (A) Simple random sampling, (B) grid sampling, (C) transect sampling with either one or five sampling units per station and (D) grid sampling with random replacements.

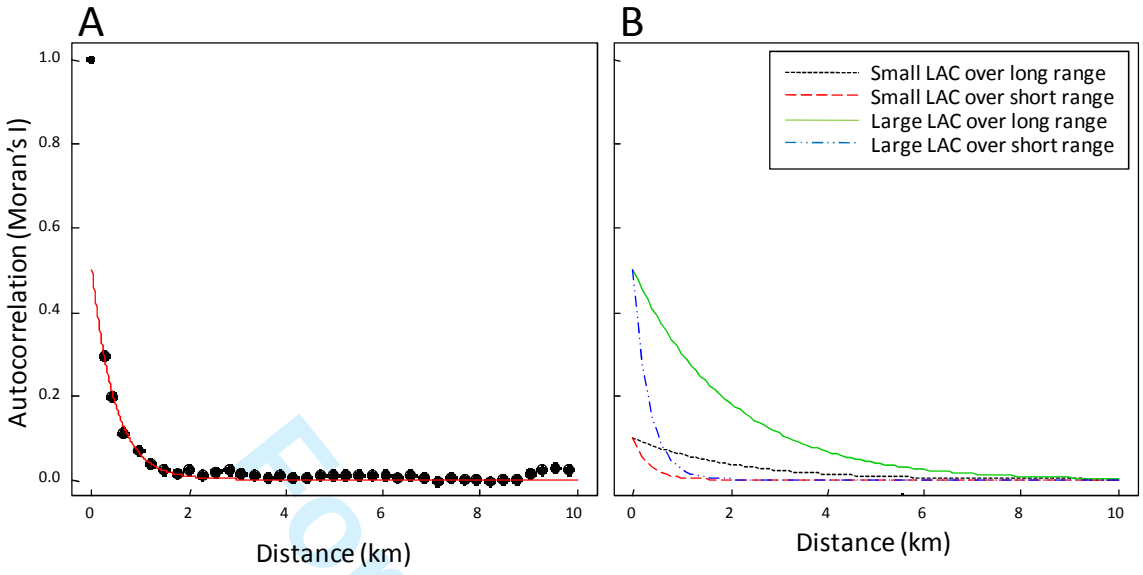

Fig. 2. Autocorrelation as function of distance for field and simulated data. (A) An example for fitting autocorrelation (*AC*) as function of distance (*h*) from field data for *Nereis diversicolor*, where $AC(h) = 0.50\ e^{-2.11h}$. Note that distance class zero is not included in the fit (see Methods). (B) Autocorrelation functions of four simulated levels of autocorrelation with weak or strong local autocorrelation (LAC) combined with a shallow or steep decline in autocorrelation with distance.

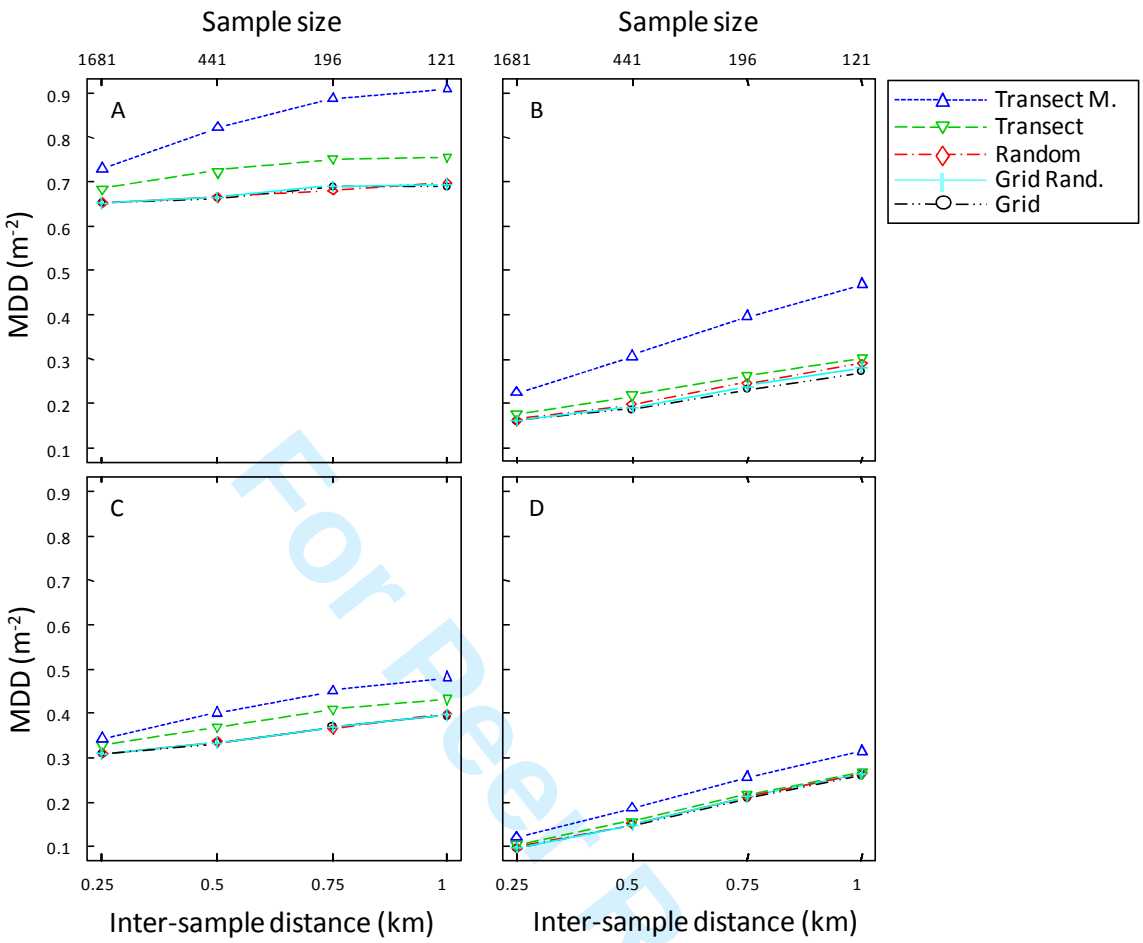

Fig. 3. Minimum detectable difference for sampling designs at different levels of autocorrelation. The minimum detectable difference (MDD) for: transect sampling with either multiple (Transect M.) or a single sample per station (Transect), simple random sampling (Random), grid sampling with random replacements (Grid Rand.) and grid sampling (Grid). The bottom axis gives the distance between sampling stations which is inversely related to sample size (top axis). Each panel represents different simulated levels of autocorrelation: (A) strong local autocorrelation and a long range of autocorrelation, (B) strong local autocorrelation and a short range, (C) weak local autocorrelation and a long range and (D) weak local autocorrelation and a short range.

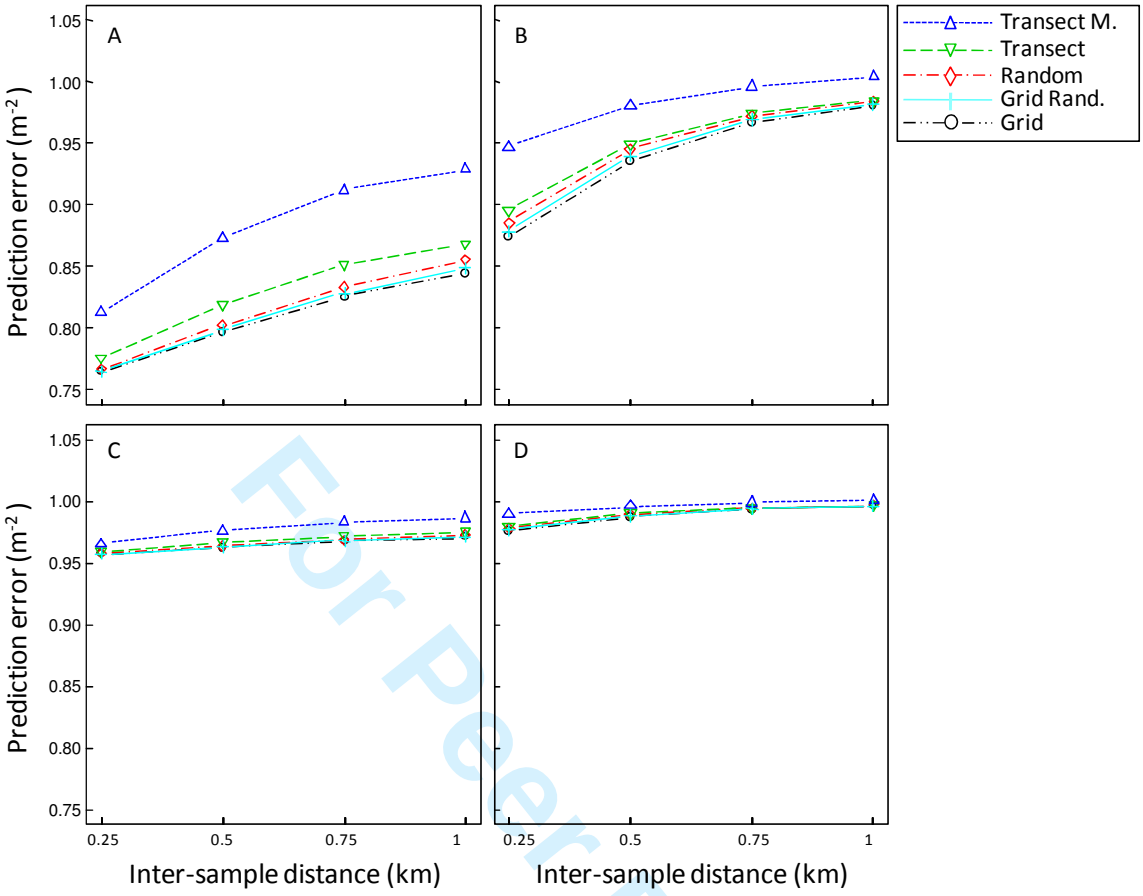

Fig. 4. The mean prediction error of kriging is given for sampling designs at different levels of autocorrelation. For an explanation on the x-axis, legend and panels A-D, see caption of Fig. 3.

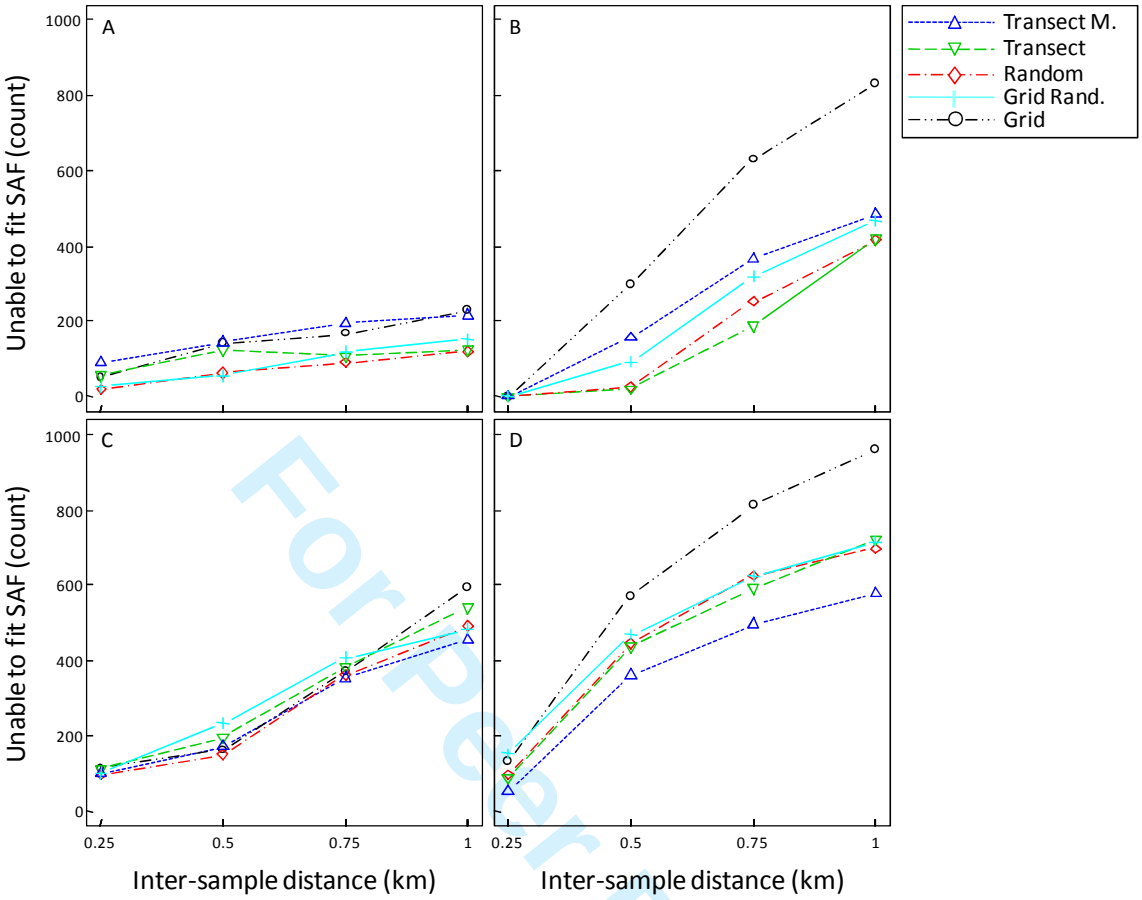

Fig. 5. Count of inestimable spatial autocorrelation function (SAF) from 1000 simulation runs for different sampling designs at different levels of autocorrelated data. For an explanation on the x-axis, legend and panels A-D, see caption of Fig. 3.

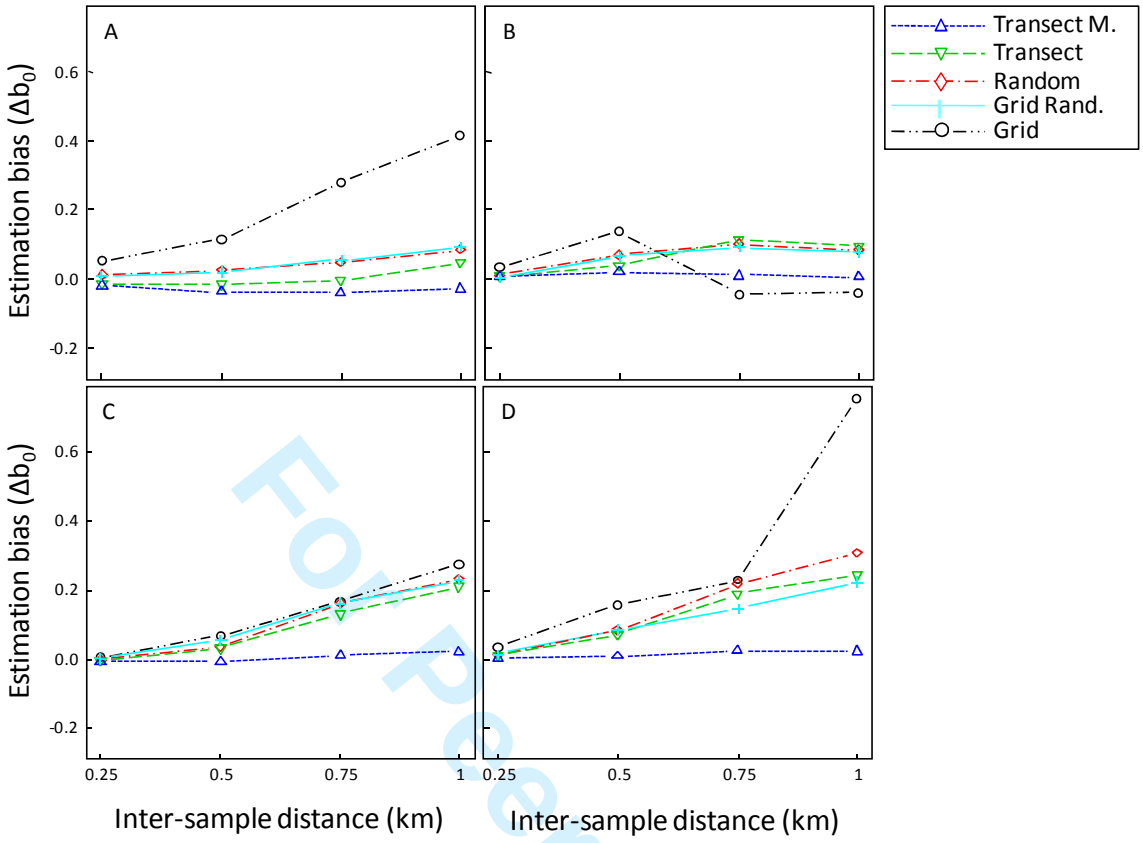

Fig. 6. Estimation bias of local autocorrelation for different sampling designs at different levels of autocorrelated data. The difference is given between the simulated and estimated local autocorrelation (Δb_0). For an explanation on the x-axis, legend and panels A-D, see caption of Fig. 3.

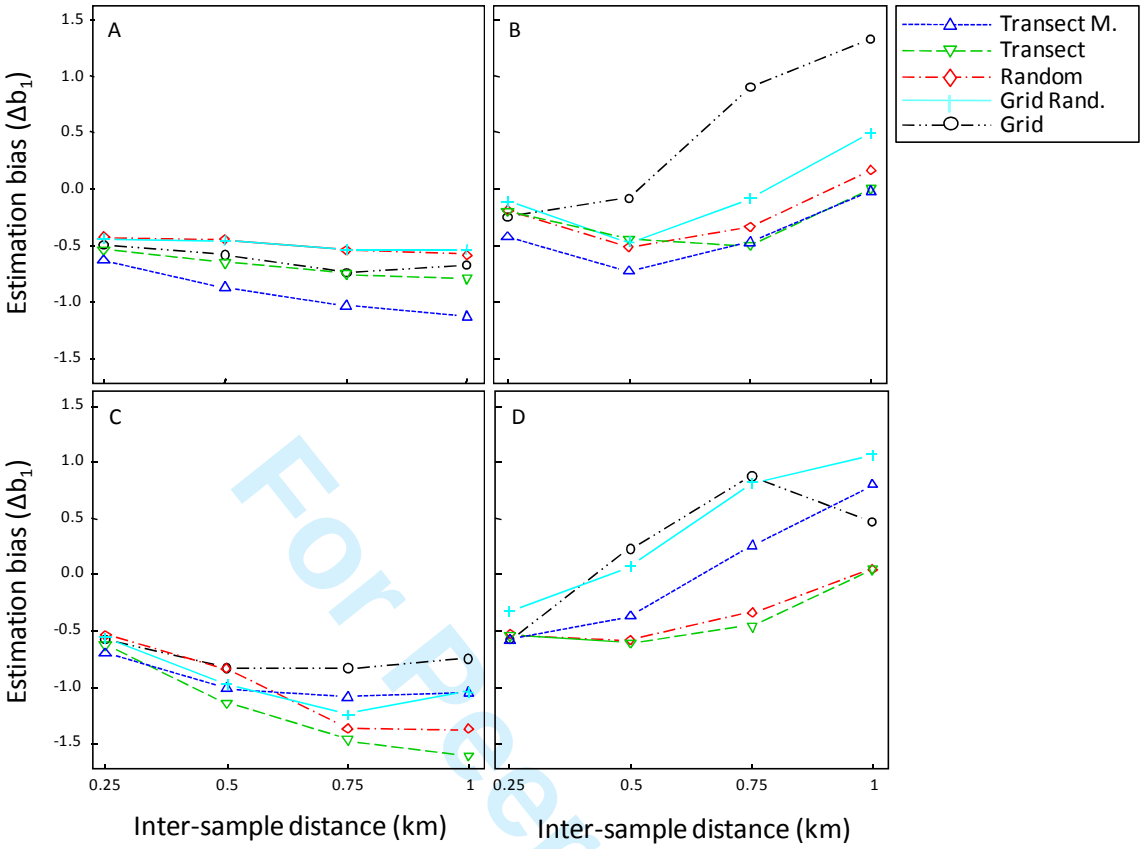

Fig. 7. Estimation bias of decline in autocorrelation for different sampling designs at different levels of autocorrelated data. The difference is given between the simulated and estimated decline of autocorrelation with distance (Δb_1). For an explanation on the x-axis, legend and panels A-D, see caption of Fig. 3.