

University of Groningen

## Untying the knot

Bijleveld, Allert Imre

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*  
2015

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Bijleveld, A. I. (2015). *Untying the knot: Mechanistically understanding the interactions between social foragers and their prey*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

## Chapter 2

# Designing a benthic monitoring programme with multiple conflicting objectives

Allert Bijleveld, Jan van Gils,  
Jaap van der Meer, Anne Dekinga,  
Casper Kraan, Henk van der Veer  
and Theunis Piersma

**SUMMARY** Sound conservation and management advice usually requires spatial data on animal and plant abundances. The expense of programmes to determine species distributions and estimates of population sizes often limits sample size. To maximize effectiveness at minimal costs, optimizations of such monitoring efforts are critical. A monitoring programme can have multiple objectives with demands on the optimal sampling design that are often in conflict. Here we develop an optimal sampling design for monitoring programmes with conflicting objectives, building on an existing intertidal benthic monitoring programme in the Dutch Wadden Sea and simulation models bounded in their parameter spaces by these data. We distinguish 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, the final objective was (3) accurately estimating model autocorrelation parameters. To compare sampling designs, we used the following criteria: (1) minimum detectable difference in mean between two time periods or two areas, (2) mean prediction error, and (3) estimation bias of autocorrelation parameters. Using Monte Carlo simulations we compared five sampling designs with respect to these criteria (i.e. simple random, grid, two types of transects, and grid with random replacements) at four levels of naturally occurring spatial autocorrelation. The ideal sampling design for objectives (1) and (2) was grid sampling and for objective (3) random sampling. The sampling design that catered best for all three objectives combined was grid sampling with a number of random samples placed on gridlines. Grid sampling with a number of random samples is considered an accurate and powerful tool with the highest effectiveness. This sampling design is widely applicable and allows for accurate estimates of population sizes, monitoring of population trends, comparisons of populations/trends between years or areas, modelling autocorrelation, mapping species distributions, and a mechanistic understanding of species distribution processes.

## 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). Collecting such data is expensive and labour intensive, and therefore monitoring programmes are practically constrained by the number of sampling units (Andrew and Mapstone 1987, Field et al. 2005). Smaller sample sizes reduce the accuracy of the estimates (e.g. total abundances), or the power to detect significant impacts (Quinn and Keough 2002). Hence, it pays selecting a sampling design that minimises the number of sampling units and maximises the accuracy of the estimates (Thompson 1992). Monitoring programmes can have multiple objectives such as describing spatial patterns and temporal trends in species abundance, or impact assessments and each objective, can have a different optimal sampling design. Optimising sampling designs between monitoring objectives that explicitly consider spatial autocorrelation has received little attention so far and is the objective of this paper.

Hitherto the ecological literature has paid much attention to designing sampling programmes aiming at detecting the impact of a specific treatment in an area (Green 1979, Underwood 1991, Stewart-Oaten et al. 1992, Underwood 1997, Stewart-Oaten and Bence 2001). So called 'beyond BACI designs' (Before-After Control-Impact) are now regarded as the most appropriate for spatial sampling for impact assessments (Underwood 1991, 1994, Schmitt and Osenberg 1996). Usually, multiple sites are sampled within an area, several locations per site, and several sampling units per location. The results are analyzed by nested ANOVA where the overall variance is allocated to different variance components according to the spatial scale of sampling. Such models are powerful for impact assessment, but they ignore spatial autocorrelation which can provide additional biological information (Sokal and Oden 1978b, Kraan et al. 2009a, Kraan et al. 2009b). As such, the monitoring of spatial autocorrelation warrants to become a monitoring objective itself.

In contrast to the nested ANOVA approach, the geostatistical literature (Diggle and Ribeiro 2007) and some of the ecological literature (Sokal and Oden 1978a, b, Legendre 1993, Keitt et al. 2002, Fortin and Dale 2005, Dormann et al. 2007), have emphasized explicitly modelling spatial autocorrelation. Usually spatial autocorrelation is modelled as a declining function of Euclidean distance between sampling units (Cliff and Ord 1981, Upton and Fingleton 1985). Hence, geostatistical approaches advocate model-based inference by estimating an underlying spatial autocorrelation model allowing for predictions at unsampled locations (i.e. mapping, Ripley 1981, Cressie 1993). Another advantage of explicitly modelling spatial autocorrelation is that this provides an understanding of the mechanisms (e.g. competition, landscape structure) underlying the observed spatial distributions (Bergström et al. 2002, Klaassen et al. 2006c, de Frutos et al. 2007, Lagos et al. 2007, Kraan et al. 2009b, van Gils 2010).

The NIOZ Royal Netherlands Institute for Sea Research maintains long-term benthic monitoring programmes for detecting 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. 2009). Additionally, mapping macrobenthic invertebrates enables predictions on the spatial distribution of their predators, such as birds and fish (van Gils et al. 2005b, van Gils et al. 2006b). Currently, the NIOZ monitoring programme is limited to the western Dutch Wadden Sea, but is to be extended to cover the entire Dutch Wadden Sea for monitoring effects of gas extraction. The aim of this study is twofold. First, building on the existing benthic monitoring efforts at NIOZ, we aim to determine an optimal sampling design for monitoring programmes that have multiple conflicting objectives. Second, we apply this sampling design to the Dutch Wadden Sea. We focus on the following objectives: (1) estimation of temporal change and spatial differences in abundance between two years or two areas. Because comparisons between years or areas depend on similar analytical principles, they can be combined into one objective. (2) Predicting species abundances at unsampled locations, i.e. mapping. Such predictions, using model-based inference, are only as good as the match between the estimated model parameters and the data, and therefore an additional objective was (3) accurately estimating autocorrelation model parameters. Comparisons between sampling designs were based on: (1) the minimum detectable difference between means of two time periods or areas, (2) the mean prediction error, and (3) the estimation bias, i.e. the number of times the autocorrelation parameters were inestimable and the difference in simulated and estimated autocorrelation parameters. With respect to these criteria, we compared one novel with four regularly applied sampling designs.

## **METHODS**

### **General approach**

Using field data, the most parsimonious autocorrelation structure was fitted, model parameters estimated and four extreme, but realistic levels of autocorrelation selected. These autocorrelation models were then used to simulate spatially autocorrelated data with a normal distribution and according to different sampling designs compared regarding the previously mentioned 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 NIOZ has monitored population densities of macrobenthic invertebrates across 225 km<sup>2</sup> of intertidal mudflats in the western Dutch Wadden Sea (van der Meer 1997, Piersma et al. 2001). Between July and September each year, between 1,807 and 2,762 stations were sampled. Sample stations were arranged according to a grid sampling design with 0.25 km inter-sample distance. Sampling stations were located by handheld Global Positioning System (GPS, Garmin 60, Olathe, Kansas, U.S.A.). At each station one core (1/56 m<sup>2</sup>) 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 comparisons between groups (objectives 1 and 2), the analyses

were based on the difference in 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 polychaete species (*Scoloplos armiger*, *Heteromastus filiformis*, *Nereis diversicolor*, *Nephtys hombergii* and *Lanice conchilega*).

### Statistic framework

Generalised Least Squares (GLS) methods are model-based analyses for spatially autocorrelated data, as well as for spatial predictions necessary for the three objectives. GLS are widely used in spatial statistics (Cressie 1993) and spatial ecology (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 a spatial autocorrelation function (SAF) to field data in order to estimate covariance between sampling units.

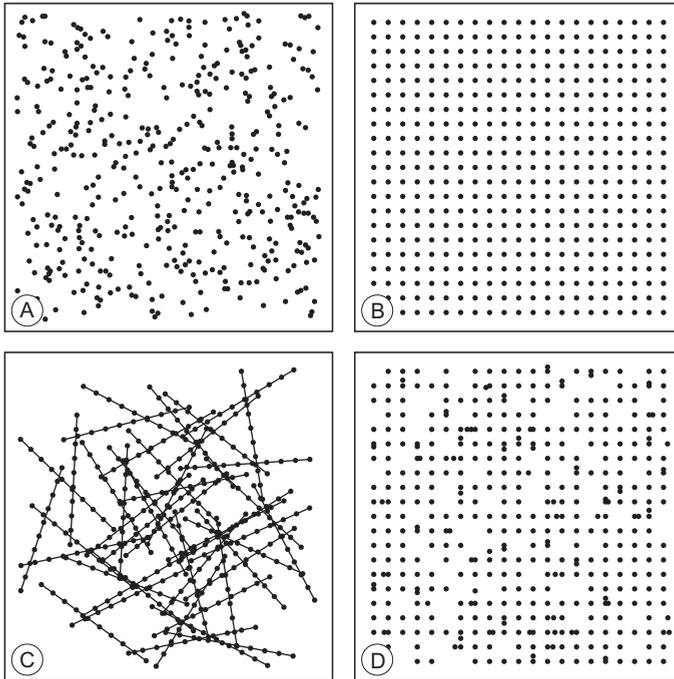
Autocorrelation, expressed as the commonly used Moran's I, was calculated for discrete distance classes into a correlogram (Sokal and Oden 1978a, Cliff and Ord 1981, Legendre and Fortin 1989). Several SAFs were fitted to the correlogram. The most parsimonious fit was provided by the exponential SAF (van der Meer and Leopold 1995):

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

Using nonlinear least-squares, 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. 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 sampling units, and multiplied by the variance of the response variable  $\sigma^2$  to obtain an estimate of the variance-covariance matrix  $\Sigma$  (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. 2.1A) and often combined with stratified sampling (Armonies and Reise 2003). (2) For grid sampling, sampling stations are usually spaced in a lattice (Herman et al. 2001) and, in this study, located in the centre of a grid cell (Fig. 2.1B). (3) Transect sampling (Fig. 2.1C) consisted of transects with random starting locations and a random heading in which 9 additional stations were equally spaced (Beukema 1976, Yates et al. 1993). (4) Transect sampling with multiple sampling units is similar to transect sampling, but at each of 10 transect sampling stations an additional four sampling units were taken within 400 m<sup>2</sup> (Beukema 1974). (5) Grid sampling with random replacements is based on the "lattice-plus-closed-pair-design" by Diggle & Lophaven (2006).



**Figure 2.1** 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.

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. 2.1D). *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. This reduces sampling costs while maintaining most statistical advantages of random sampling (Diggle and Lophaven 2006).

### Data simulation

On a 10 × 10 km surface area, sampling stations were selected according to the different sampling designs. The distance between sampling stations was 0.25, 0.5, 0.75 and 1 km (i.e. sample sizes of 1681, 441, 196 and 121 respectively). This coincided with an expected 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. 9 inter-sample distances). 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  $100 \text{ km}^2$  plus a margin of 0.5 times the inter-sample distance. Sample stations were restricted to this area and starting locations of transects were re-assigned if any sample station would reach beyond this area. Consequently, diagonal transects are more likely to occur than transects parallel to the gridlines (Fig. 2.1C). This sampling bias will be large if the 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 area. This bias also occurs in the field, and as we were interested in field implications of different sampling designs, it was accepted as realistic.

The variance-covariance matrix  $\Sigma$  was calculated using four extreme, but naturally occurring, levels of autocorrelation. Based on field data estimates of autocorrelation parameters, we modelled either weak or strong local autocorrelation ( $b_0$ ), with a shallow or steep decline in autocorrelation with distance ( $b_1$ ). Spatially autocorrelated response variables were simulated for each sampling design and inter-sample distance, using Cholesky decomposition (i.e. given a symmetric positive definite matrix, the Cholesky decomposition is an upper triangular matrix with strictly positive diagonal entries such that  $A = U^T U$ ) (Ripley 1981, 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 calculated by  $\varepsilon = W^T \xi$  with  $\xi$  drawn from the standard normal distribution ( $\mu = 0$  and  $\sigma^2 = 1$ ).

### Comparison criteria

The minimum detectable difference (MDD) between two populations (objective 1) was calculated with the standard error of the mean (se):  $MDD = se \cdot (t_{\alpha, n-1} + t_{\gamma, n-1})$  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 2002). The mean and se were calculated with GLS following Cliff & Ord (1981). We calculated the variance of the mean using ordinary least squares (OLS, corresponding to GLS analyses with  $b_0 = 0$  and  $b_1 = 0$ ). With OLS variance, the fraction of independent data points in the autocorrelated sample (i.e. effective sample size  $n^*$ , Griffith 2005) could be estimated by dividing OLS-variance 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 for which the calculations are available elsewhere (Ripley 1981, Cressie 1993, Fortin and Dale 2005, Furrer et al. 2013). To estimate the mean prediction error we randomly selected 100 locations on the  $100 \text{ km}^2$  simulated area. For each location we calculated the prediction error and the resulting 100 prediction errors were averaged.

For objective (3) we fitted a SAF to simulated autocorrelated data at four levels of autocorrelation. We recorded how often autocorrelation parameters were inestimable and calculated the difference between simulated and estimated autocorrelation parameters,

i.e. estimation bias. The SAF was fitted over 2/3 of the maximum distance between pairs of sample units and the width of 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.

All calculations and simulations were performed with R v2.6 (R Development Core Team 2008) using the following packages: *PBSmapping* (Schnute et al. 2008), *ncf* (Bjornstad 2006), *spatstat* (Baddeley and Turner 2005) and *fields* (Furrer et al. 2013).

## RESULTS

### Field data

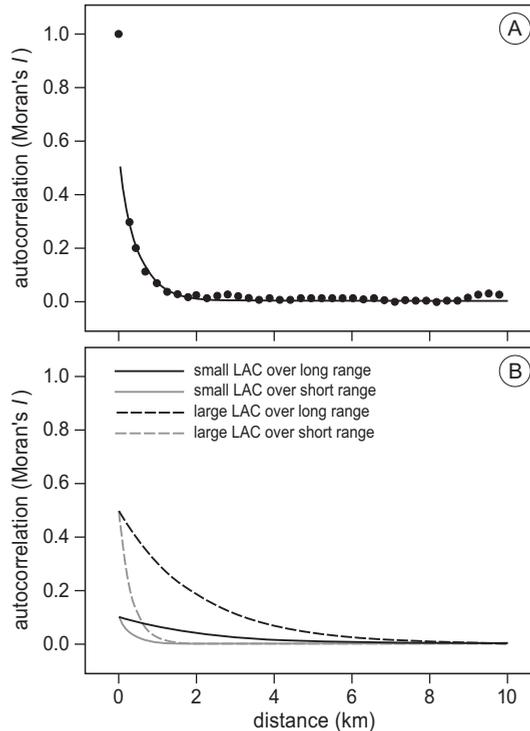
On the basis of 2,695 sampling stations covered both in 2005 and 2006, density differences between years could be calculated. These data, used to estimate a species correlogram, consisted of many zeros and were therefore not normally distributed. There are no transformation routines that could adequately 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). For each species,  $\sigma^2$  was estimated and  $b_0$  and  $b_1$  were estimated from a correlogram (Fig. 2.2A). The parameter estimates for  $b_0$  ranged from 0.03 to 0.66 and for  $b_1$  from -3.12 to -0.34 (Table 2.1). Depending on the level of autocorrelation, the effective sample size (percentage of independent data points,  $n^*$ ) ranged from 3% to 28% (Table 2.1).

### Simulated data

Based on field estimates (Table 2.1), we used  $b_0 = 0.1$  or  $b_0 = 0.5$  and  $b_1 = -0.5$  or  $b_1 = -3$  (Fig. 2.2B) to simulate different levels of spatially autocorrelated normally distributed 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.

### MDD - Objective (1)

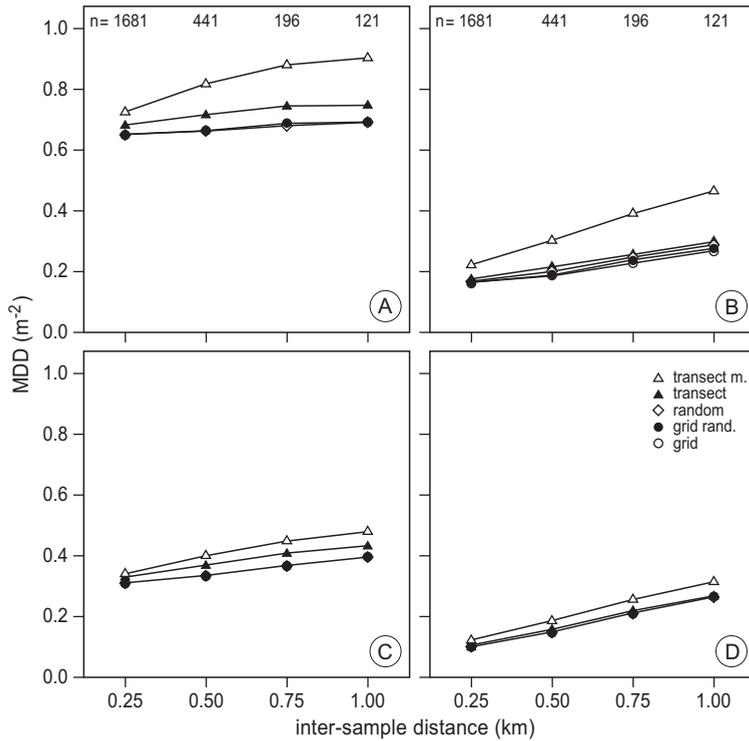
The level of autocorrelation decreased with increased inter-sample distance, because sampling units were increasingly outside each other's range of influence. Nonetheless, the decrease in MDD (i.e. increased power) with longer inter-sample distance was outweighed by the stronger increase in MDD caused by reduced sample sizes. Therefore, MDD increased for all sampling designs as inter-sample distance increased (Fig. 2.3). Grid sampling allowed for the smallest MDD for most inter-sample distances. Simple random



**Figure 2.2** Autocorrelation as function of distance for (A) field and (B) 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.

**Table 2.1** Estimates of spatial autocorrelation function parameters based on field data. For each species are given: local autocorrelation  $b_0$ , steepness of decline in autocorrelation with distance  $b_1$ , and percentage effective sample size  $n^*$  (see methods).

Species	$b_0$	$b_1$	$n^*$ (%)
<i>Cerastoderma edule</i>	0.32	-0.76	5
<i>Macoma balthica</i>	0.05	-0.50	13
<i>Mya arenaria</i>	0.05	-0.34	8
<i>Abra tenuis</i>	0.66	-3.12	19
<i>Ensis americanus</i>	0.03	-0.42	18
<i>Scoloplos armiger</i>	0.21	-0.40	3
<i>Heteromastus filiformis</i>	0.13	-0.58	7
<i>Nereis diversicolor</i>	0.50	-2.11	14
<i>Nephtys hombergii</i>	0.38	-3.02	28
<i>Lanice conchilega</i>	0.23	-1.29	13

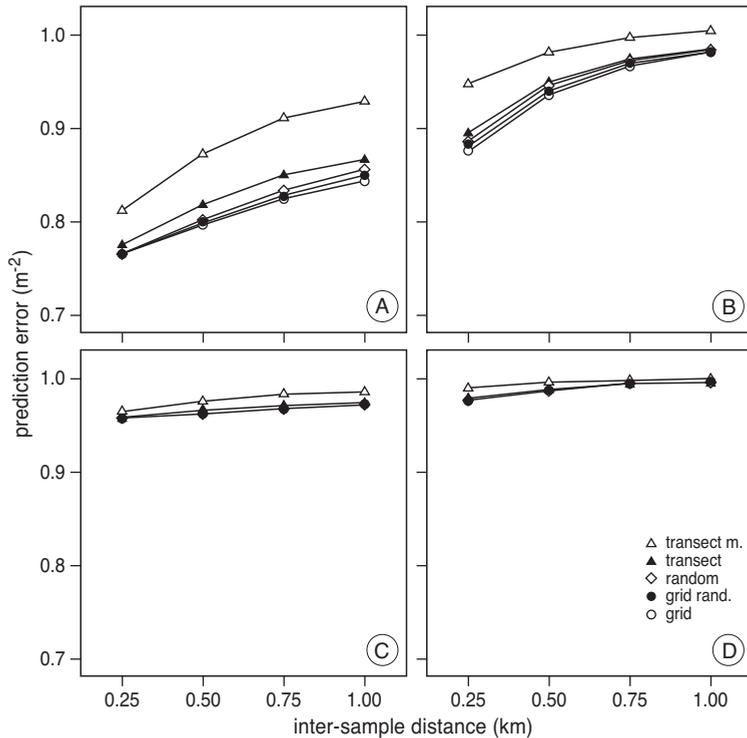


**Figure 2.3** Minimum detectable difference (MDD) at different levels of autocorrelation for: transect sampling with either multiple (transect m.) or a single sample per station (transect), simple random sampling (random), grid sampling with random replacements (grid rand.) and grid sampling (grid). The x-axis gives distance between sampling stations which is inversely related to sample size. 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.

and grid sampling with random replacements also provided small MDD. Both transect sampling designs consistently showed a larger MDD than the other sampling designs. Between autocorrelation levels, strong local autocorrelation (Fig. 2.3A-B) resulted in a larger MDD than weak local autocorrelation (Fig. 2.3C-D). Additionally, a long range of autocorrelation (Fig. 2.3A and 2.3C) resulted in a larger MDD than a short range (Fig. 2.3B and 2.3D). The differences in MDD between sampling designs were more pronounced for strong local autocorrelation over a short range (Fig. 2.3B).

### Prediction error - Objective (2)

Sample size and the level of autocorrelation were reduced with an increase in inter-sample distance, and therefore, the prediction error increased with inter-sample distance (Fig. 2.4). With decreased autocorrelation, kriging interpolations became less accurate and the prediction error more or less approached the simulated variance of 1 (Fig. 2.4C-D).



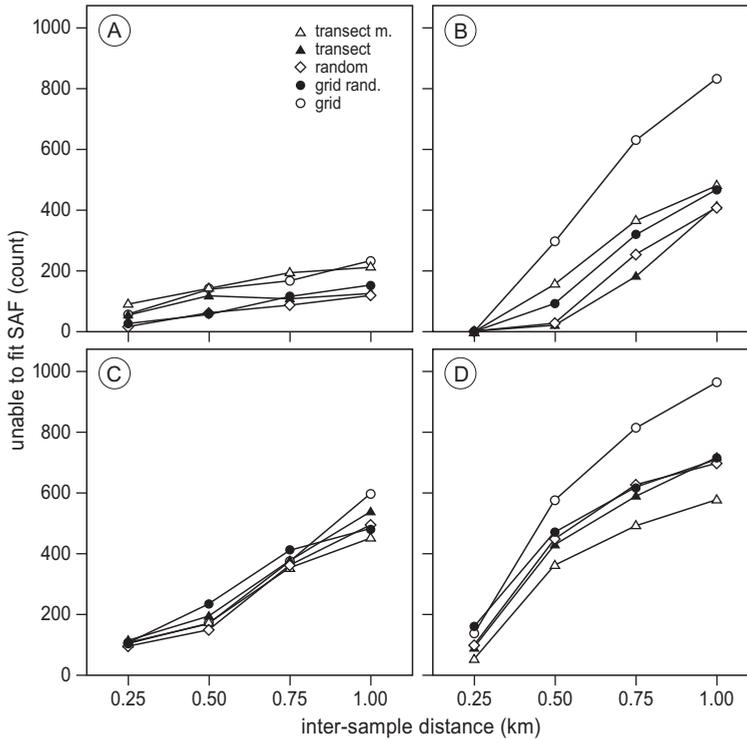
**Figure 2.4** Mean prediction error of kriging given for sampling designs at different levels of autocorrelation. For an explanation on the x-axis, legend and panels A-D, see caption of Fig. 2.3.

Grid sampling allowed for smallest prediction errors for all inter-sample distances (Fig. 2.4A-D), followed by grid sampling with random replacements, simple random sampling, transect sampling and transect sampling with multiple sampling units. Between autocorrelation levels, strong local autocorrelation (Fig. 2.4A-B) resulted in smaller prediction errors than weak local autocorrelation (Fig. 2.4C-D). Additionally, a long range of autocorrelation (Fig. 2.4A and 2.4C) resulted in smaller prediction errors than a short range of autocorrelation (Fig. 2.4B and 2.4D).

### Estimation bias of autocorrelation parameters - Objective (3)

The smaller the level of autocorrelation, the less often the autocorrelation parameters were estimable (Fig. 2.5). An increase in inter-sample distance, therefore, reduced the number of times the autocorrelation parameters were estimable (Fig. 2.5). Overall, random sampling allowed for estimating the SAF most often.

The smaller the sampling distance, the more accurate the estimate of local autocorrelation ( $b_0$ ) (Fig. 2.6). As inter-sample distance increased  $b_0$  was overestimated using most sampling designs. Because multiple sampling units were taken within a small range, transect sampling with multiple sampling units was most accurate for estimating  $b_0$  (Fig. 2.6),

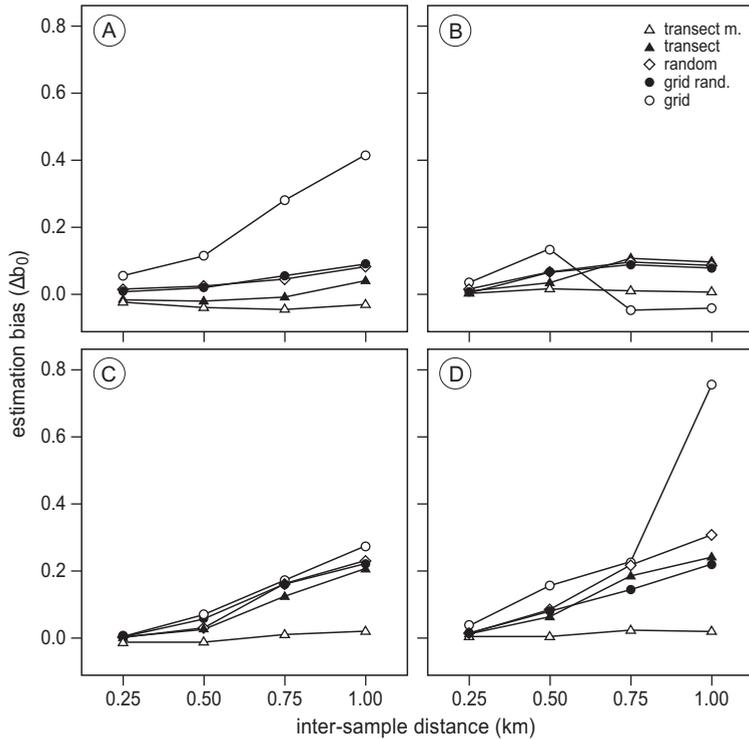


**Figure 2.5** Count of inestimable spatial autocorrelation function (SAF) from 1000 simulation runs for different sampling designs at different levels of autocorrelated data. For an explanation on the x-axis, legend and panels A-D, see caption of Fig. 2.3.

especially at low levels of autocorrelation (Fig. 2.6D). Grid sampling showed the largest estimation bias (Fig. 2.6).

The decline in autocorrelation with distance ( $b_1$ ) was often underestimated (Fig. 2.7). The estimation bias of  $b_1$  was larger with low levels of autocorrelation (Fig. 2.7D) than with high levels of autocorrelation (Fig. 2.7A) and increased with inter-sample distance (Fig. 2.7). Grid sampling with random replacements was the most accurate in estimating  $b_1$  followed by random sampling (Fig. 2.7). Both transect sampling designs showed the largest estimation bias.

Combining the three sub-criteria for estimating autocorrelation structure, the best performing sampling design was dependent on the level of autocorrelation. For low levels of autocorrelation, random sampling performed best, but for intermediate and high levels of autocorrelation, grid sampling with random replacements performed best. At our average overall level of autocorrelation random sampling performed best closely followed by grid sampling with random replacements.



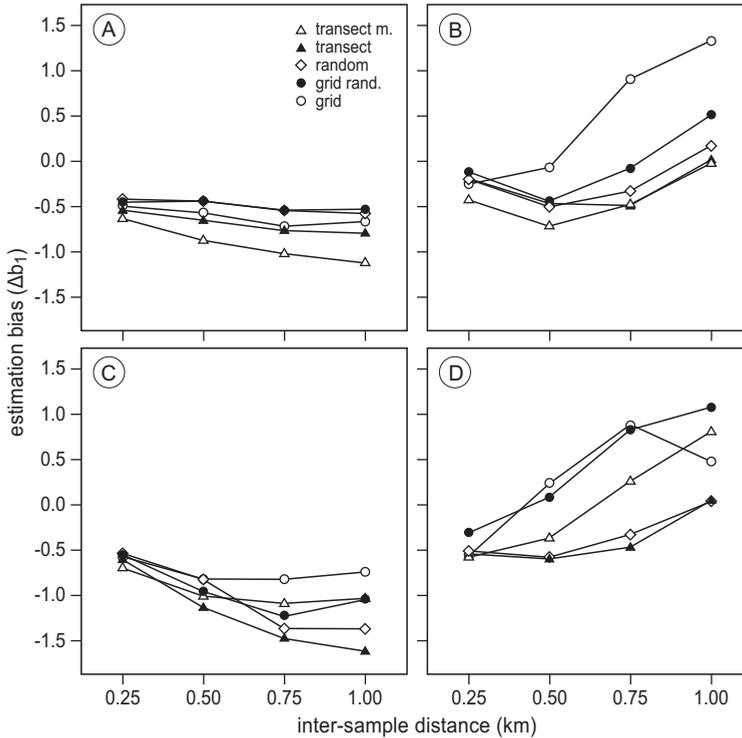
**Figure 2.6** Estimation bias of local autocorrelation for different sampling designs at different levels of autocorrelated data. The difference is given between the simulated and estimated local autocorrelation ( $\Delta b_0$ ). For an explanation on the x-axis, legend and panels A-D, see caption of Fig. 2.3.

## DISCUSSION

### The ideal sampling design per objective

#### COMPARISON BETWEEN YEARS OR AREAS

In ecology one often observes positive spatial autocorrelations (Legendre and Fortin 1989). Statistical power for comparisons between, for instance, the mean abundances of an organism in two areas, is thus reduced. This can be illustrated by the 'effective sample size' (Table 2.1), i.e. the proportion of sampling units that consists of non-autocorrelated independent data points (Griffith 2005). The higher the level of autocorrelation, the smaller the effective sample size and the smaller the power of model-based inference. Indeed, our results show that low levels of autocorrelation resulted in large power (i.e. small MDD) to detect changes between years or areas (objective 1). Between all levels of autocorrelation grid sampling revealed the largest power.



**Figure 2.7** 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 ( $\Delta b_1$ ). For an explanation on the x-axis, legend and panels A-D, see caption of Fig. 2.3.

#### MAPPING SPECIES ABUNDANCES

The stronger the spatial autocorrelation, the more accurate interpolations of abundances at unsampled locations as the interpolated values are weighed more strongly and by more surrounding sampling units (Cressie 1993, Diggle and Ribeiro 2007). Also, designs that satisfy the uniformity condition (e.g., surface-covering sampling designs) allow for more accurate kriging predictions (Pooler and Smith 2005, Marchant and Lark 2007). Our results are consistent with this understanding. The prediction error was smallest with the highest levels of autocorrelation and with grid sampling which covers the entire surface and conforms to the uniformity condition.

#### ESTIMATION OF AUTOCORRELATION PARAMETERS

Grid sampling was the best sampling design for objectives (1) and (2). However, note that in our study we simulated autocorrelated data with known autocorrelation parameters. In the analysis of field data autocorrelation parameters need to be estimated from the data itself. For estimating autocorrelation parameters (objective 3), grid sampling performed

worst, although the fit of these parameters to the data determine the validity of model-based inference (Gregoire 1998, Haining 2003, Little 2004). For accurate parameter estimations, spatial sampling designs should include small distances between sampling units (Diggle and Lophaven 2006). Our results showed that those designs that included small inter-sample distances allowed for the most times the SAF could be fitted and the most accurate estimates of autocorrelation parameters. Overall, random sampling performed best in estimating autocorrelation structure closely followed by grid sampling with random replacements.

### The ideal sampling design between objectives

For this study we were interested in a sampling design that allowed for the best results between three monitoring objectives: estimation of temporal changes and spatial differences in abundance, prediction of abundances at unsampled locations and accurately estimating autocorrelation model parameters. None of the sampling designs suited all objectives. Therefore, the objectives need to be compromised to find the best overall sampling design. A procedure ideally suited for finding a compromise between sampling designs is Pareto-optimization (Steuer 1986). Using Pareto-optimization we can identify superior sampling designs using the following criterion: no other sampling design produces improved results concerning a particular objective without at the same time producing worse results for another. A further selection from those sampling designs that fit the above criterion requires arbitrary weighing of sampling objectives. We ranked all sampling designs according to the different monitoring objectives (Table 2.2). For objective (3) we averaged the rankings for sub-objectives to obtain an overall ranking. For objectives (1) and (2), grid sampling was the best sampling design, closely followed by grid sampling with random replacements. For objective (3) random sampling performed best, closely followed by grid sampling with random replacements. The worst sampling

**Table 2.2** Ranking of sampling designs according to different monitoring objectives, i.e. minimum detectable difference (MDD), mean prediction error, the accuracy in fitting the spatial autocorrelation function (SAF). The different sampling designs are transect sampling with 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). Three Pareto-optimal solutions exist (indicated by \*): Random, Grid Rand. and Grid. Weighing all monitoring objectives equally, grid sampling with random replacements (indicated in bold) is the ideal compromise between objectives.

	MDD	prediction error	SAF
Transect M.	5	5	4
Transect	4	4	3
Random*	3	3	1
<b>Grid Rand.*</b>	<b>2</b>	<b>2</b>	<b>2</b>
Grid*	1	1	5

design for objective (1) and (2) was transect sampling with multiple sampling units and for objective (3) it was grid sampling. Following Pareto-optimization we identified three optimal solutions: grid sampling, random sampling and grid sampling with random replacements. For all objectives, grid sampling with random replacements was a close runner up and showed substantially improved performance compared to grid sampling on objective (3). Weighing all three monitoring objectives equally, grid sampling with random replacements is the best compromise between 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 ideal sampling design for similar objectives would be surface-covering grid sampling with a percentage of sampling stations randomly placed on gridlines *additional* to the grid design (see Online Supporting Information for R-code to create such a sampling design). The main effect of *adding* random samples instead of *replacing* is that the homogenous surface coverage is preserved which decreases the prediction error. 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 lower the level of autocorrelation, the higher the percentage of additional random sampling units needs to be for accurately estimating the autocorrelation function. The level of autocorrelation depends on the scale of the sampling effort, i.e. the ratio of inter-sampling distance to autocorrelation range. The higher the ratio, the lower the level of autocorrelation. Increasing the percentage of random sampling units will increase levels of autocorrelation and allow for more accurate estimates of autocorrelation parameters. On the other hand, the higher the level of autocorrelation in the data, the larger the MDD.

In practice sampling programmes are more complicated than we have simulated here. For instance, in order to increase power and reduce prediction error, one might want to use environmental variables as covariates or apply environmental stratification, where autocorrelation varies among strata. Optimising sampling designs in such cases will be slightly more difficult, but can be achieved along similar principles as we have opted here.

### **Issues of non-normality**

The field data used to estimate autocorrelation parameters, were not normally distributed. Nonetheless, we simulated normally distributed data and from this deduced the ideal sampling design according to the three criteria. Ideally, one would simulate data similar to the data observed. However, methods for simulating non-normally distributed data with a known autocorrelation structure are still in development (Jackson and Sellers 2008). Using simultaneously-specified models (Jackson and Sellers 2008), we explored the possibility of simulating Poisson data with a known autocorrelation structure. The general idea of this method is generating a normally distributed random vector  $\varepsilon$  with known covariance matrix, and add this to a vector of expected values (on the basis of environmental

data) to create a vector  $X$ . Then a Poisson variable is generated with the exponential of  $X$ . Following this method, we experienced that the simulated autocorrelation structure of the normally distributed variable  $X$  did not show up in the autocorrelation structure of the resulting Poisson data. We recommend more work on this topic to resolve this issue for non-normally distributed data. Regardless of this practical limitation, we have no reason to believe that our results are not robust to different underlying data distributions. Even though data with different distributions will probably alter the quantitative results (i.e. absolute values of the estimates), the qualitative results (i.e. ranking of sampling designs according to the objectives) are likely to remain similar. Nonetheless, when methods become available to simulate non-normally distributed data with a specific autocorrelation structure and computationally more efficient parameter estimation methods become available, we advocate further investigation into the effects of non-normality on selecting the most appropriate sampling design.

### **Implications for Wadden Sea monitoring programmes**

Currently, NIOZ macrobenthic monitoring programmes follow either transect sampling (Beukema 1976, Beukema and Dekker 2006, Dekker and Beukema 2007), or grid sampling with an 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. 2009). The NIOZ monitoring programme is to be extended to cover the entire Dutch Wadden Sea for monitoring the effects of gas exploitation. This study indicates that surface-covering grid sampling with additional random sampling is the ideal sampling design for detecting temporal and spatial changes in abundances as well as mapping macrobenthic invertebrates. Given the surface area of the Dutch Wadden Sea, extending the monitoring programme at the current inter-sample distance of 0.25 km would inflate sample size to 19,000 sampling units, beyond what is feasible within seasonal and logistical constraints. We, therefore, suggest the inter-sample distance should be increased to 0.50 km (corresponding to roughly 4,700 sampling units) to allow surface-coverage of the entire Dutch Wadden Sea.

### **Acknowledgements**

We thank all volunteers who helped collecting samples and the crew of MS *Navicula* whom provided a helpful and welcoming atmosphere on board. We thank Dick Visser for preparing the figures, Hans Malschaert and Piet Ruardij for use of the *biocluster* supercomputer, and the Nederlandse Aardolie Maatschappij (NAM) for financing AIB and JAvG. Finally, we thank the reviewers and editors for constructive comments.

