Environ Ecol Stat
DOI 10.1007/s10651-012-0223-2

Spatial variogram estimation from temporally aggregated seabird count data

B. Pérez-Lapeña · K. M. Wijnberg · A. Stein · S. J. M. H. Hulscher

Received: 4 March 2011 / Revised: 5 July 2012

Abstract Seabird abundance is an important indicator for assessing impact of human activities on the marine environment. However, data collection at sea is time consuming and surveys are carried out over several consecutive days for efficiency reasons. This study investigates the validity of aggregating those data over time to estimate a spatial variogram that is representative for spatial correlation in species abundance. For this purpose we simulate four-day surveys of seabird count data that contain spatial and temporal correlation arising from temporal changes in the spatial pattern of environmental conditions. Estimates of the aggregated spatial variogram are compared to a variogram that would arise when data were collected over a single day. The study reveals that, under changing environmental conditions over surveys days, aggregating data over a four-day survey increases both the non-spatial variation in the data and the scale of spatial correlation in seabird data. Next, the effect of using an aggregated variogram on the statistical power to test the significance of an impact is investigated. The impact concerns a case of establishing an offshore wind farm resulting in seabird displacement. The study shows that both overestimation and underestimation of statistical power occurs, with power estimates differing up to a factor of two. We conclude that the spatial variation in seabird abundance can be misrepresented by

Handling Editor: Pierre Dutilleul.

B. Pérez-Lapeña (✉) · K. M. Wijnberg · S. J. M. H. Hulscher
Department of Water Engineering and Management, Faculty of Engineering Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
e-mail: perez@itc.nl

A. Stein
Department of Earth Observation Science, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

Published online: 16 October 2012

using temporally aggregated data. In impact studies, such misrepresentation can lead to erroneous assessments of the ability to detect impact.

Keywords Ecological surveys · Impact assessment · Spatio-temporal correlation · Statistical power · Unconditional simulation

1 Introduction

The awareness that human activities may pose a threat to the environment has increased during the last decades (Crain et al. 2009; Gardner et al. 2009). In particular, the large number of wind farm developments at sea and their potential effects on marine fauna is of great concern (Gill 2005).

Several studies aim at assessing the displacement of seabirds due to the presence of an offshore wind farm, hereafter referred to as the impact. In such studies, seabird abundance may first define an undisturbed, or reference, situation. Collected count data on seabird abundance following the impact are then compared to the reference situation to investigate the effect of the impact (Petersen et al. 2006; Leopold et al. 2010; Pérez-Lapeña et al. 2010). Given the within-year variability in numbers of individual species, separate analyses are then carried out that can be repeated in different months. In such a way, data collected during the month before the construction of the wind farm can be used to build a statistical model that describes the reference situation. Recognizing the spatial variability in species abundance data, a geostatistical model is suited:

$$Y_{s,m} = \mu_{s,m} + e_{s,m}, \tag{1}$$

where Y is the bird density as the variable of interest at spatial location s and month m, μ is the deterministic component of the model, and e is the spatially autocorrelated stochastic component. The latter represents deviations from μ, and also is referred to as the residuals.

Knowledge on species behaviour derived from data or literature can be included in the deterministic component. For example, an estimate of average species abundance in a survey area in the particular month may be derived from data on the species population dynamics. To describe the spatial variation in the residuals a variogram can be employed in the stochastic component (Legendre et al. 2002; Perry et al. 2002; Pebesma et al. 2005).

Spatial dependence in the residuals can be caused by spatial dependence of the bird density on an environmental factor that affects species abundance at survey locations. It may serve as a proxy for data that for example at the survey dates are not available to be incorporated into the deterministic component (Diniz-Filho et al. 2003; Dormann et al. 2007). In such cases the spatial scale at which environmental factors vary, i.e. the scale of the spatial pattern, will be reflected in the scale of the spatial autocorrelation in the residuals. The latter is represented by the range parameter of the variogram model. For example, sea water temperature is an important factor determining the spatial distribution of several fish species (Fauchald et al. 2000). Studies have shown that there

is a spatial overlap between the scale of fish abundance patterns and that of several seabird species (Schneider and Piatt 1986; Fauchald et al. 2000; Skov et al. 2000). Sea water temperature may then be used to explain seabird patterns as a proxy for such direct factors, such as fish availability (Ainley et al. 2005). Varying spatial patterns in sea water temperature can result from movements of fresh water plumes that emerge from the mouth of a river located close to the study area (de Boer et al. 2009). For example, in the Netherlands an offshore wind farm is located at approximately 50 km from the mouth of the river Rhine. The images in Fig. 1 show that the spatial pattern of water temperature in and around the wind farm area is subject to changes during consecutive days due to discharges and movements of the river plume. As such data are often not available at the days of the survey, we will assume in this work that the scale and temporal variability of an environmental factor is reflected in the stochastic part of the geostatistical model.

Collecting species abundance data at sea is time consuming. Therefore, surveys are carried out over a time period of several consecutive days (e.g. four days). Data collected over a four-day survey are commonly combined to provide a snapshot representative of the species abundance in a particular month. This snapshot is often decomposed in terms of a mean value or a spatial trend surface (component μ in equation 1) and residuals e that may exhibit spatial dependence (Pebesma et al. 2005; Leopold et al. 2010; Pérez-Lapeña et al. 2010). The latter can be summarized in a spatial variogram of these snapshot residuals, which we refer to as a S-variogram. When estimating the S-variogram, pairs of residuals that are close in space may be as much as three days apart in time as a result of the data aggregation. This raises the question as to how this will affect the estimated variogram of the residuals. The

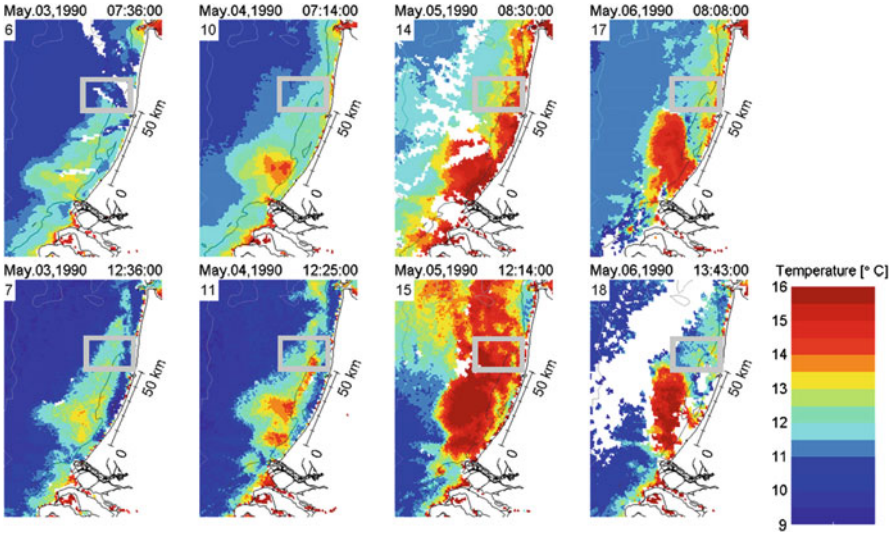

Fig. 1 Changes in water temperature along the Dutch coast over four days, during the morning and afternoon in May, 1990 (modified after de Boer et al. 2009). The *grey polygon* represents the location of the wind farm

S-variogram constructed with data collected over a four-day survey may not be representative of the true variogram, or T-variogram, as if data were collected on a single day. Whether or not the S-variogram is representative of a T-variogram is expected to be affected by the temporal correlations in the residuals. Changes in the spatial pattern of environmental conditions between consecutive days may produce changes in the temporal correlations in the residuals. Figure 2 shows four simulated spatial patterns of residuals (e.g. reflecting the spatial pattern of sea water temperature) for four days with a temporal correlation in the residuals of 0.5 between consecutive days. The dots represent survey locations on each single survey day. Note that the spatial pattern is changing but the scale of the pattern remains similar. Therefore, aggregating the data over a four-day survey to estimate the S-variogram of the residuals may considerably violate the assumption that data were collected on a single day, under the same environmental conditions.

The impact assessment method described in Pérez-Lapeña et al. (2010), used the S-variogram based on pre-construction surveys to define the reference situation. The environmental conditions during the days of the post-construction surveys will often differ from those during the days of the pre-construction surveys. In this scenario, the reference situation may not be well represented through the use of the pre-construction derived S-variogram. Observed changes in post-construction seabird abundance may be wrongly identified as being statistically different from the reference situation. In addition, the use of a S-variogram may also influence the ability to detect an impact when it actually occurred.

In this paper we investigate i) how temporal correlation in spatially correlated abundance data affects the approximation of the T-variogram by a S-variogram, and ii) to what extent the use of a S-variogram derived from pre-construction survey data affects

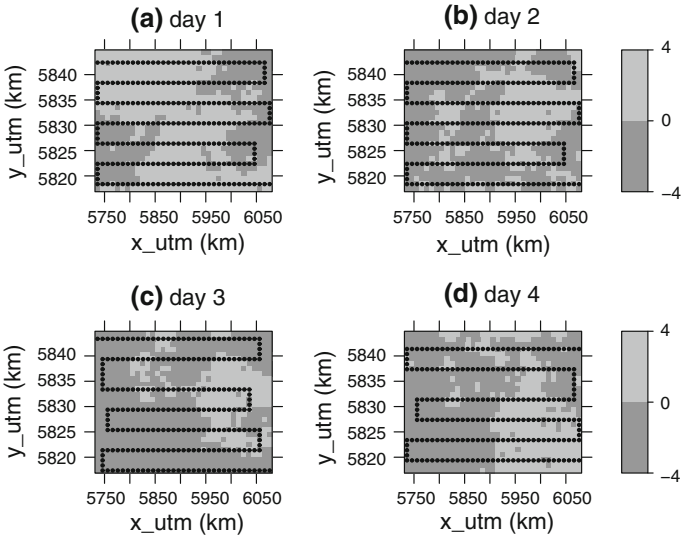

Fig. 2 Simulated spatial patterns in the residuals of a geostatistical model (e.g. reflecting the spatial pattern of sea water temperature) for four survey days. The *dots* represent survey locations in each day

the ability to detect impact as measured by statistical power. The content of the paper is organized as follows. First, the method to simulate spatially and temporally correlated abundance data is presented. Next, the analyses that experts carry out in estimating the S-variogram from aggregated data over a four-day survey are reproduced. Then the temporal correlation scenarios investigated are presented, and the validity of representing the T-variogram by a S-variogram is analysed. Finally, we show the effects of using a S-variogram on the ability to assess impact when the environmental conditions between pre- and post-construction periods differ.

2 Methodology

To investigate the above mentioned issues we need to know the true variogram to assess the effect of aggregating data on the estimate of the variogram. Therefore, we will generate artificial datasets that mimic seabird abundance data collection over four-day surveys under changing spatial patterns of environmental conditions. From these artificial datasets we will estimate the S-variograms of the simulated residuals.

2.1 Data simulation

We use artificial data consisting of simulated four-day surveys of species abundance. The survey design for the simulations, i.e. the configuration of survey locations, is adopted from an existing seabird survey that was carried out in June 2003 (prior to wind farm construction) off the coast of the Netherlands to investigate the impact of an offshore wind farm on seabirds (Leopold et al. 2004). The survey was carried out over four days t (t goes from 1 to 4). The study area covered the future wind farm area and a zone around it, or control area, that it is expected to be unaffected by the disturbance. Figure 3 shows the survey locations covering the study area of approximately 900 km^2 with different symbols indicating the spatial locations s that were visited on each day t. The total number of locations is $N = 461$.

At each spatial location s and day t, bird counts Y are treated as a Negative Binomial variable according to the model:

$$Y_{s,t} = \mu_{s,t} + e_{s,t}, \tag{2}$$

where $\mu_{s,t}$ is the expected number of a particular species at location s and day t, and e is the stochastic residual.

In this work, we model $\mu_{s,t}$ to be invariant in time and space to mimic that only an average abundance for the entire survey area and survey month could be derived, e.g. from data on population dynamics. Spatial and temporal correlation in the residuals may occur, for example, when spatially varying environmental factors affecting seabird abundance could not be included in the deterministic component, while the spatial pattern of these influential environmental factors changed between survey days.

The different spatial and temporal correlations in the residuals e are imposed in the Pearson residuals $p_{s,t}$ (hereafter referred to as residuals) which are calculated as follows:

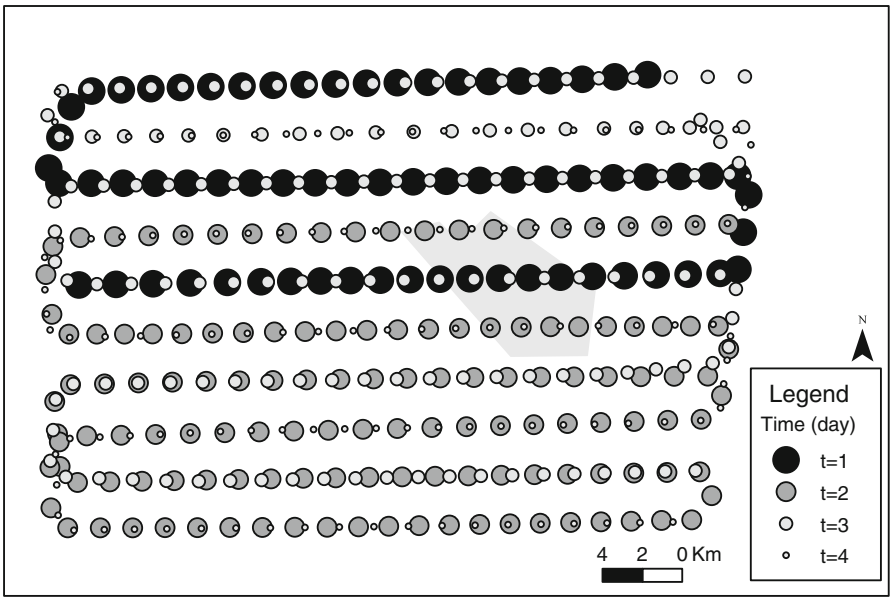

Fig. 3 Survey locations s where counts of a particular species were collected over four consecutive days t. The light *grey polygon* represents the wind farm area

$$p_{s,t} = \frac{Y_{s,t} - \mu}{\sqrt{\sigma^2}}, \tag{3}$$

where μ and σ^2 are the mean and the variance of the Negative Binomial distribution, respectively. In those situations where the distribution of the residuals are skewed and given that the simulation procedure generates correlated residuals from a standard normal distribution (refer to *Phase 1* later in the text), the simulated residuals are assumed to be normal score transformed Pearson residuals. Note that the analysis of spatial autocorrelation using real data would involve a transformation of the residuals so that these are normally distributed using the normal score transform, with the ties in the data broken at random (Goovaerts and Jacquez 2004).

We assume a separable space-time process (Genton 2007) with covariance:

$$C(s, s+h, t, t+u) = C_S(s, s+h)C_T(t, t+u), \tag{4}$$

where C_S is the spatial covariance, C_T is the temporal covariance, s is a spatial location, h is a distance in space, t is a temporal location, and u is a distance in time. In addition, we have assumed stationarity in space and, therefore, $C_S(s, s+h) = C_S(h)$. The temporal covariance does not only depend on the temporal lag but also on day t. For example, the temporal covariance between day one and day three does not need to be the same as the covariance between day two and day four. Therefore, the resulting space-time covariance becomes:

$$C(h, t, t+u) = C_S(h)C_T(t, t+u). \tag{5}$$

Furthermore, we assume homoscedasticity and a global variance equal to one, in which case C_S and C_T are the spatial correlation and the temporal correlation, respectively. In matrix notation, the space-time covariance matrix Σ is:

$$\Sigma = \mathbf{S} \otimes \mathbf{T}, \tag{6}$$

where $\mathbf{S}$ is the spatial correlation matrix which elements $s_{j,k} = C_S(s_j, s_k)$, $\mathbf{T}$ is the temporal correlation matrix which elements $t_{j,k} = C_T(t_j, t_k)$, and $\otimes$ represents the Kronecker product.

We illustrate the calculation of the space-time covariance matrix with a simple example. For two survey locations s_1 and s_2 we set the spatial correlation equal to 0.74, and for two survey dates t_1 and t_2 we set the temporal correlation equal to 0.3. The spatial correlation matrix $\mathbf{S}$ is then written as

$$\mathbf{S} = \begin{bmatrix} s_{1,1} & s_{1,2} \\ s_{2,1} & s_{2,2} \end{bmatrix} = \begin{bmatrix} 1.00 & 0.74 \\ 0.74 & 1.00 \end{bmatrix}$$

and the temporal correlation matrix $\mathbf{T}$ as

$$\mathbf{T} = \begin{bmatrix} t_{1,1} & t_{1,2} \\ t_{2,1} & t_{2,2} \end{bmatrix} = \begin{bmatrix} 1.0 & 0.3 \\ 0.3 & 1.0 \end{bmatrix}$$

Application of the Kronecker product between $\mathbf{S}$ and $\mathbf{T}$ results in the space-time covariance matrix:

$$\Sigma = \begin{bmatrix} s_{1,1}\mathbf{T} & s_{1,2}\mathbf{T} \\ s_{2,1}\mathbf{T} & s_{2,2}\mathbf{T} \end{bmatrix} = \begin{bmatrix} s_{1,1}t_{1,1} & s_{1,1}t_{1,2} & s_{1,2}t_{1,1} & s_{1,2}t_{1,2} \\ s_{1,1}t_{2,1} & s_{1,1}t_{2,2} & s_{1,2}t_{2,1} & s_{1,2}t_{2,2} \\ s_{2,1}t_{1,1} & s_{2,1}t_{1,2} & s_{2,2}t_{1,1} & s_{2,2}t_{1,2} \\ s_{2,1}t_{2,1} & s_{2,1}t_{2,2} & s_{2,2}t_{2,1} & s_{2,2}t_{2,2} \end{bmatrix}$$
$$= \begin{bmatrix} 1.000 & 0.300 & 0.740 & 0.222 \\ 0.300 & 1.000 & 0.222 & 0.740 \\ 0.740 & 0.222 & 1.000 & 0.300 \\ 0.222 & 0.740 & 0.300 & 1.000 \end{bmatrix}$$

Note that by applying the Kronecker product, the spatial correlation is maintained at each day t.

For the data analyzed in this study, the simulated residuals $p_{s,t}$ are spatially correlated on each day t according to an exponential variogram model and spatial covariance function, which are defined as:

$$\gamma(h) = c_0 + c_1 \left[1 - exp\left(-\frac{h}{a} \right) \right], \text{for } h > 0 \tag{7a}$$

$$C_S(h) = c_1 exp\left(-\frac{h}{a} \right), \text{for } h > 0 \tag{7b}$$

where h is the distance between spatial locations s, c_0 is the so-called nugget (discontinuity of the variogram at $h = 0$), c_1 is the partial sill, $c_0 + c_1$ is the sill (residual semivariance at distance h where spatial correlation is no longer present), and $3a$ is the effective range (distance h over which spatial autocorrelation is no longer present).

In the temporal dimension, the residuals $p_{s,t}$ are assumed to be temporally correlated from day to day. Consecutive days are correlated according to correlation $\rho_{(t,t+1)}$.

The generation of datasets that simulate seabird count data collected over a four-day survey is carried out in two phases.

Phase 1 Simulation of spatially and temporally correlated residuals

The simulation procedure for spatiotemporal correlated residuals consists of the following steps:

a. Construction of the spatial covariance matrix **S** and the temporal covariance matrix **T**.
b. The Kronecker product between **S** and **T** is calculated to obtain the space-time covariance matrix Σ.
c. The Cholesky decomposition **V** of Σ is calculated such that $\mathbf{V} = \Sigma\Sigma^T$.
d. The matrix **P** containing the spatially and temporally correlated residuals is obtained as follows:

$$\mathbf{P} = \mathbf{aV}, \tag{8}$$

where $\mathbf{P_{Nt\times 1}} = [p_{11}, p_{21}, \ldots, p_{N1}, p_{12}, \ldots, p_{Nt}]$,
$\mathbf{a_{Nt\times 1}} = [a_{11}, a_{21}, \ldots, a_{N1}, a_{12}, \ldots, a_{Nt}]$ with values a sampled from a standard normal distribution.

The result of phase 1 is a set of spatially and temporally correlated residuals at all N survey locations at all of the four days ($p_{s,t}$, for $s = 1, \ldots, N$ and $t = 1, \ldots, 4$).

Phase 2 Survey reproduction

Once the spatially and temporally correlated residuals are calculated in phase 1, a dataset of bird counts is obtained as follows:

a. Using the specified values for μ and σ^2, the residuals $p_{s,t}$ obtained in phase 1, and equation 3, the values of $Y_{s,t}$ are calculated and rounded to the nearest integer. In case that phase 1 involves normal score transformed residuals, the normally distributed residuals $p_{s,t}$ are first backtransformed so that these have their original distributional properties. Then, using equation 3, the values of $Y_{s,t}$ are calculated.
b. From each of the four realizations, only those counts $Y_{s,t}$ at locations s that were visited (in reality) on a particular day t are selected (see Fig. 3).

The obtained bird counts from phase 2 represent an example dataset that experts may collect during a four-day survey.

2.2 Variogram estimation procedure

From each simulated dataset a variogram will be estimated to investigate the effect of varying temporal correlation between environmental conditions in subsequent survey days.

The next steps are followed to estimate the variogram from aggregated data over four days, mimicking a procedure followed in reality with a single dataset:

a. Estimation of the mean bird count $\hat{\mu}$ and variance $\hat{\sigma}^2$ using all counts $Y_{s,t}$ from the four-day simulated survey.
b. Calculation of the Pearson residuals using the counts $Y_{s,t}$ and the estimated mean $\hat{\mu}$ and variance $\hat{\sigma}^2$.
c. Calculation of the empirical variogram and the fit of a variogram model (S-variogram).

2.3 Scenarios

We investigate several scenarios for temporal correlation between days, hence with varying parameter values for $\rho_{(t,t+1)}$. Parameter values for the μ, σ^2, and $\gamma(h)$ are fixed. The value for μ is set to 10, the value for σ^2 is set to twice the mean value, i.e. 20, and the values for the spatial variogram parameters of the residuals in equation 7a are $c_0 = 0, c_1 = 1$, and $3a = 6$ km.

The choice for an effective range (hereafter referred to as range) of 6 km is motivated by literature. In Schneider and Piatt (1986), the scale of the spatial pattern in Common Guillemots (*Uria aalge*) was estimated as ranging between 3 and 8.75 km. In addition, a significant spatial overlap was found between the scale of the pattern in the seabird species and the one of fish aggregations. In our work, it is assumed that fish aggregations are reflected in the scale of the spatial pattern of, for example, sea water temperature. Given that data on water temperature are not available, the scale of such pattern is modelled in the variogram of the residuals setting the range parameter equal to 6 km. A value of zero for the nugget parameter and of one for the partial sill parameter has been chosen for simplification.

The specified scenarios in terms of correlations between patterns in consecutive days are supposed to mimic the stability of spatial patterns in physical conditions in the area that influence bird abundance (such as sea water temperature). For instance, a spatial pattern in sea water temperature may exist due to river outflow in the neighbourhood of the surveyed area. Due to changes in river discharge or wind direction the location of this river plume may shift. In case of small shifts considerable correlation between days will exist (e.g. $\rho = 0.7$) whereas for larger shifts the correlations will become smaller (e.g. $\rho = 0.3$). In the extreme scenarios, the position of the river plume is constant over all days ($p\_1\_1\_1$) or extremely variable ($p\_0\_0\_0$).

Below, the scenarios and the types of analyses carried out for each scenario are explained in detail.

We impose a temporal correlation in the residuals $\rho_{(t,t+1)}$ which can vary from day to day. In Table 1, the correlations $\rho_{(t,t+1)}$ between days for each of the scenarios investigated are presented. The Cholesky decomposition requires a correlation matrix

Table 1 Scenarios investigated with corresponding temporal correlations in the residuals $\rho_{(t,t+1)}$

Scenario	Day 1–Day 2	Day 2–Day 3	Day 3–Day 4
p_1_1_1	0.999	0.999	0.999
p_07_1_1	0.7	0.999	0.999
p_07_07_1	0.7	0.7	0.999
p_07_07_07	0.7	0.7	0.7
p_05_1_1	0.5	0.999	0.999
p_05_05_1	0.5	0.5	0.999
p_05_05_05	0.5	0.5	0.5
p_03_1_1	0.3	0.999	0.999
p_03_03_1	0.3	0.3	0.999
p_03_03_03	0.3	0.3	0.3
p_0_1_1	0	0.999	0.999
p_0_0_1	0	0	0.999
p_0_0_0	0	0	0

that is positive definite. A correlation of exactly 1 between two time steps would result in a correlation matrix that is not positive definite. Therefore, for these cases a correlation of 0.999 has been imposed. In the remainder of this paper we refer to this correlation as a 'perfect' correlation.

For each scenario, 1000 four-day surveys (datasets) are simulated. For each dataset, an empirical variogram is calculated and the variogram model according to equation 7a is fitted. This results in 1000 estimates for the nugget, partial sill, and range. Hence, we obtain distributions of nugget, partial sill, and range values for each scenario. Two types of analysis are carried out on these distributions of variogram parameters.

The first analysis serves to investigate the validity to aggregate survey data over a four-day period as if all data were collected on a single day. For each scenario, estimates for the T-variogram and S-variograms are compared by plotting their 5th and 95th percentile variogram values. The 5% and 95% points of the variograms are obtained as follows. First, using the 1000 simulated variograms, the corresponding semivariance values for several distances are obtained. Then, for each distance the 5% and 95% percentiles over the semivariance values are calculated and an exponential variogram is fitted. The resulting 5% and 95% percentiles of both the T-variogram and S-variograms for each scenario are presented in Fig. 4. The T-variogram is estimated using the aggregated data over the four days but with a temporal correlation $\rho_{(t,t+1)}$ of 1 between each day. In other words, the T-variogram is estimated assuming that all data were collected in one day. Further, the overlap is calculated in variogram parameters (nugget, partial sill and range) between the T-variogram and S-variogram for all the scenarios (presented in Fig. 5). In this work, the term overlap refers to the area of overlap between two unit-area histograms of a given variogram parameter. For this particular analysis, the two histograms are those constructed with the fitted values of a parameter in the T-variogram and S-variogram. For each scenario, the multiplicative (temporal) correlation over the four days is calculated (x-axis) and plotted against the area of overlap (y-axis). The multiplicative correlation is calculated as the product of temporal correlations between consecutive days. For example, for

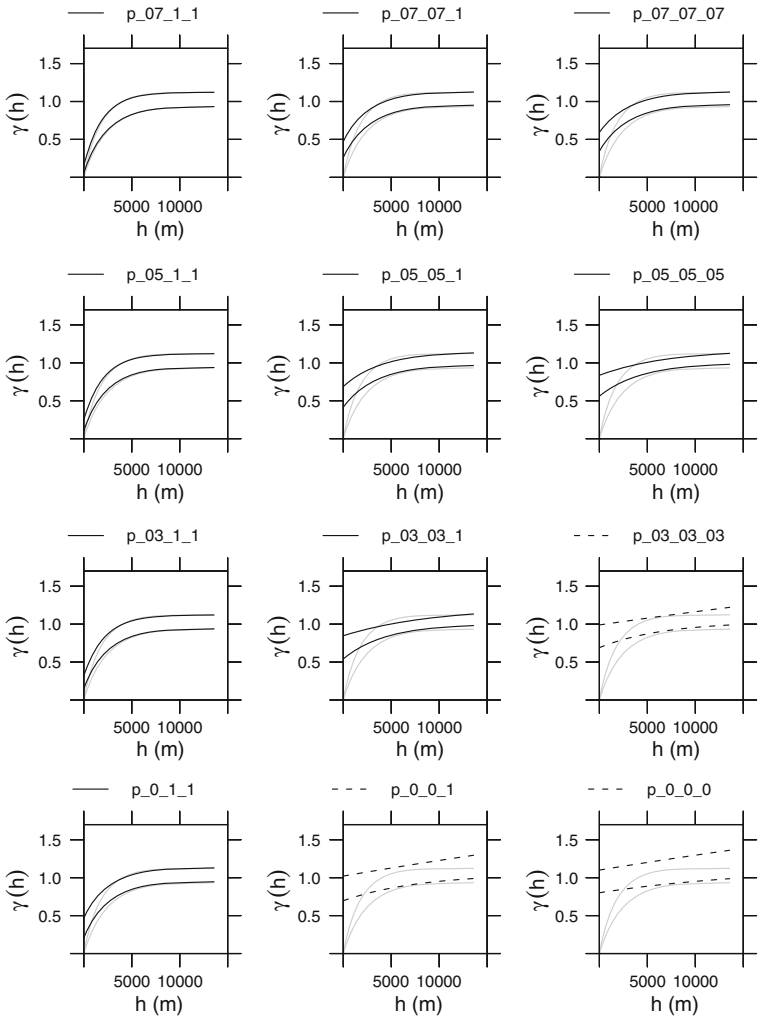

Fig. 4 5th and 95th percentiles of the fitted S-variogram, for each scenario, using all observations grouped over the four-day survey (in *black*) with temporal correlations $\rho_{(t,t+1)}$ between days (see Table 1). The 5th and 95th percentiles of the T-variogram are superimposed in light grey colour. For scenarios p_03_03_03, p_0_0_1, and p_0_0_0, the automatic fitting procedure of the exponential model resulted in an infinite number of possible combinations of sill and range. The corresponding variograms are shown with *dashed black lines*

scenario p_07_07_1 the multiplicative correlation equals 0.49. The full histograms for the parameters and scenarios are included in the Appendix.

In the second analysis, we investigate the overlap in the S-variogram parameters between pairs of scenarios within the framework of power analysis and impact assessment. The S-variogram from the first scenario of the pair is the one that would normally be used to construct the reference situation as it uses pre-construction collected data. The S-variogram from the second scenario of the pair is the one that, given the

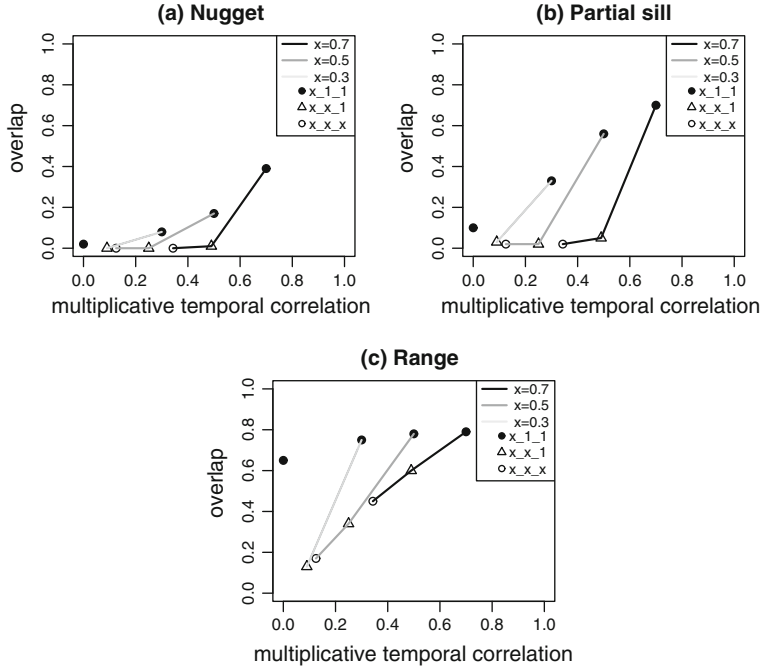

Fig. 5 Area of overlap in the empirical distribution of the variogram parameter values between the T-variogram and S-variogram. *Filled circles* represent the scenarios where the non-perfect temporal correlation spans over the first two days of the survey (x_1_1). *Triangles* represent the scenarios where the non-perfect temporal correlation spans over three days (x_x_1). *Open circles* represent the scenarios where the non-perfect temporal correlation spans over all days of the survey (x_x_x). The magnitude of the corresponding temporal correlations (x) is represented by the colour of the connecting lines

environmental conditions during the post-construction survey, should have been used in constructing the reference situation. The degree of overlap between S-variogram parameters for all combinations of investigated scenarios is visualized in a matrix containing the cartesian product of the scenarios as coordinates and the area of overlap in the empirical distributions of the variogram parameter value in each cell (presented in Fig. 6). Finally, the effect of using a S-variogram on the statistical power to test the significance of an impact is investigated. The power analysis is carried out in the context of assessing the displacement of seabirds from/to the wind farm area to/from the control area after wind farm construction. Statistical power is here defined as the probability that the test statistic used in hypothesis testing reports a statistically significant change in bird numbers using the wind farm area when a reduction $\Gamma (0 \leq \Gamma \leq 1)$ is imposed for the post-construction period (Maclean et al. 2007). This probability is the rejection rate of the null hypothesis H_0 at a chosen significance level α. The null hypothesis H_0 and alternative hypothesis H_a are stated as follows:

$$H_0 : R_{\text{ref}} - R_{\text{post}} = 0 \tag{9}$$

$$H_a : R_{\text{ref}} - R_{\text{post}} \neq 0, \tag{10}$$

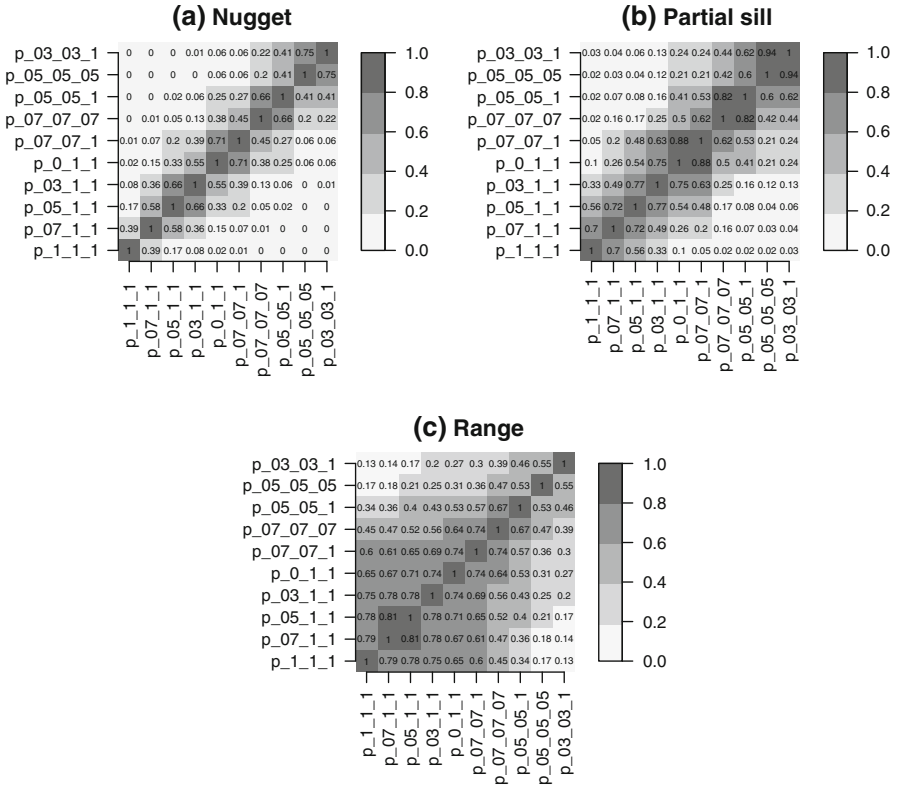

Fig. 6 Area of overlap in the empirical distribution of the variogram parameter values between S-variograms when comparing two scenarios

where R_{ref} and R_{post} are the ratio of the mean number of birds in the wind farm area as compared to the mean number of birds in the control area for the reference and the post-construction situation, respectively. In the power analysis, the number of birds in the reference situation is modelled using a particular input S-variogram. A more detailed description of the power analysis method can be found in Pérez-Lapeña et al. (2011).

All implementation of the analyses was done using the software R (R Development Core Team 2008) and related packages.

3 Results

3.1 T-variogram versus S-variogram

Percentiles

The fitted T-variogram and S-variogram for temporal correlations in the residuals $\rho_{(t,t+1)}$ are shown in Fig. 4. The panels in the first column show that when there is

non-perfect temporal correlation only from day 1 to day 2, the 5th and 95th percentiles in both the T-variogram and S-variogram remain similar as compared to the scenarios where there is non-perfect temporal correlations between the remaining days as well. This may be due to the configuration of the set of survey locations between day 1 and day 2, being more apart than sets of observations in other survey days (Fig. 3). Also, as the magnitude of change in environmental conditions over survey days increases, the non-spatial variation reflected in the nugget parameter increases.

For scenarios $p\_03\_03\_03$, $p\_0\_0\_1$, and $p\_0\_0\_0$ in Fig. 4 the experimental variograms were almost flat. The automatic fitting procedure of the exponential model that was used for these scenarios resulted in an infinite number of possible combinations of sill and range to fit to an almost flat line. Therefore, for these cases the 5th and 95th percentiles in the respective S-variograms are represented with a dashed line in Fig. 4.

Overlap in variogram parameters

Figure 5 shows the area of overlap between the two distributions of variogram parameter estimates that were derived from survey data simulations using the T-variogram and S-variogram, respectively. The area of overlap is an approximate measure to summarize how similar estimates of variogram parameters are when a S-variogram is used as an approximation of the T-variogram. The full histograms are shown in the Appendix. In Fig. 5, the area of overlap is related to the multiplicative (temporal) correlation. The multiplicative correlation for each scenario has been chosen to characterize the temporal correlation that the considered S-variograms represent.

Due to the variogram fitting limitations for scenarios with lower values of correlation (scenario $p\_03\_03\_03$, $p\_0\_0\_1$, and $p\_0\_0\_0$), the overlap for these multiplicative correlations over the four days is not shown in Fig. 5. The black point without connecting lines in each of the graphs corresponds to the scenario $p\_0\_1\_1$.

For those scenarios with a large overlap, the estimation error in variogram parameters of S-variograms and the T-variogram is such that one could use the S-variogram as being representative of the T-variogram. However, when the overlap is small also the separation between the mean parameter values should be considered. On the one hand, those scenarios for which that separation is large, the corresponding S-variograms will largely differ from the T-variogram. On the other hand, for scenarios where that separation is small, the corresponding S-variograms will also differ from the T-variogram but to a lesser extent.

From Fig. 5 it is observed that if the multiplicative correlation comes entirely from the correlation from day 1 to day 2 (filled circles), the overlap is much larger than in the cases where it is built up from the non-perfect correlations between the other days (open circles and triangles). The latter cases represent a more continuous change in the spatial pattern of environmental conditions over the survey period. Consequently, scenarios with similar multiplicative correlations, e.g., $p\_05\_1\_1$ and $p\_07\_07\_1$, can have a large difference in overlap. Therefore, the actual temporal variability over consecutive days remains important when interpreting the resulting overlap.

For the cases representing more continuous change (triangles and open circles) it can be observed that hardly any overlap exists in the estimates of nugget and partial sill

using the S-variogram versus the T-variogram, irrespective of the multiplicative temporal correlation (Fig. 5a, b). In these cases this implies a systematic overestimation of the nugget and underestimation of the partial sill by using a S-variogram instead of the T-variogram (also see Figs. 8, 9 in the Appendix). For the range parameter, to the contrary, a direct relationship seems to exist between the multiplicative temporal correlation and the amount of overlap in the estimates of the range parameter using the S-variogram versus T-variogram (Fig. 5c). The larger the multiplicative temporal correlation the more overlap exists. This implies that the extraction of the spatial scale of a pattern is somewhat less sensitive to temporal data aggregation than the estimation of the nugget and partial sill.

3.2 Statistical power of impact assessment

Figure 6 shows the overlap between the estimated variogram parameters from a pre-construction situation (y-axis) and the parameters that should have been estimated in a post-construction situation (x-axis). The nugget estimates show the least overlap for the scenarios investigated, while the range estimates show the largest overlap. Overall, there exist large differences in the overlap in variogram parameters. Such large differences motivate the analyses of the effects of using a S-variogram estimated in a pre-construction situation on statistical power when the environmental conditions in the post-construction period differ.

Figure 7a shows the power results for each scenario. Power for each scenario is calculated using the variogram fitted to the average semivariance values for several distances, also referred to as average S-variogram. In Fig. 7a it is shown that the effect of using the average S-variogram of each scenario on power is low. This is due to

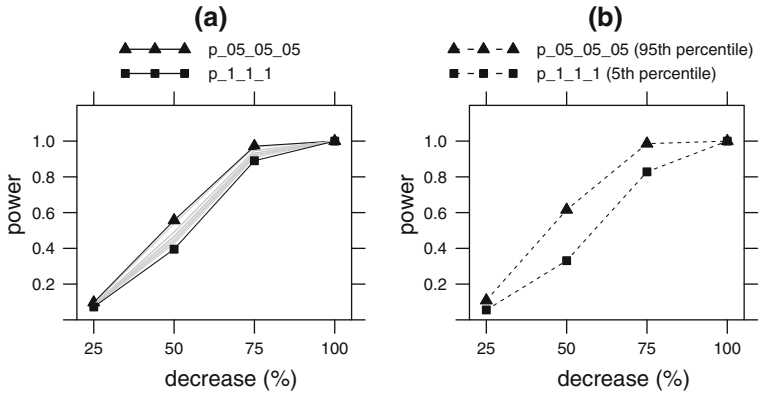

Fig. 7 Power analysis results for **a** all scenarios in Fig. 6 using the average S-variogram (power for scenarios *p*_05_05_05 and *p*_1_1_1 is shown in *black* and power for the remaining scenarios is shown in *grey*), and for **b** scenarios *p*_05_05_05 and *p*_1_1_1 using the fitted variogram parameters for the 95th percentile and the 5th percentile over the semivariance values, respectively. Power is calculated using a two-sided test and the probability of committing a type I error is set to $\alpha = 0.01$

the small number of observation pairs in the wind farm area. Scenarios having the maximum and the minimum power are $p\_05\_05\_05$ and $p\_1\_1\_1$, respectively.

In practice, one may find an estimate for the variogram that is not close to the average S-variogram. For example, Fig. 7b shows the power calculated for both scenarios $p\_05\_05\_05$ and $p\_1\_1\_1$ using the fitted variogram parameters for the 95th percentile and the 5th percentile over the semivariance values, respectively.

The power for scenario $p\_05\_05\_05$ is larger than for scenario $p\_1\_1\_1$ given that the variogram for the former scenario has a larger nugget parameter value. An increase in the nugget for scenario $p\_05\_05\_05$ has an effect on reducing the spatial correlation in the observations for distances lower than the range. As a result, observations are more independent and power increases.

Figure 7b illustrates that the estimate of the power to detect change in bird numbers may differ up to a factor of 2, depending on the collected sample of survey data in the pre-construction period and the environmental conditions assumed for the post-construction period. For example, let us consider detecting a 50% reduction in bird numbers in the wind farm area. If constant environmental conditions are assumed ($p\_1\_1\_1$) and the actually collected survey data resulted in the 5th percentile estimate of the variogram, then a power of 0.3 is estimated. If the conditions in the post-construction situation were such that the variogram should have been derived from $p\_05\_05\_05$ and the data resulted in the 95th percentile estimate of the variogram then a power of 0.6 would have been estimated, hence a factor of two difference.

4 Discussion

Our results reveal that, when data are aggregated over a four-day period, changes in environmental conditions over consecutive survey days are sources of variability in the estimated variogram of the residuals in a geostatistical model of species abundance.

This data aggregation procedure is usually employed in ecological studies in the North Sea when a representation of species abundance during a month relies on collected data in surveys that span several days (Pebesma et al. 2005; Leopold et al. 2004, 2010). With the survey design employed in this work and with non-perfect temporal correlations in the residuals only occurring from day 1 to day 2, we have shown that the estimated variogram using aggregated data is representative of the variogram estimated as if all data were collected on a single day. As the temporal correlation between days decreases, therefore the environmental conditions are more variable over the survey days, the validity of this assumption also decreases. In those situations where the goal of the study is to estimate the spatial dependence in the observations of a particular species, we have shown that aggregating data over a four-day survey increases the non-spatial variation in the data, as reflected in the nugget parameter. Therefore, the estimated spatial variation will not reflect the true spatial variation as if all data were collected on a single day. One approach to model the spatial variogram using observations that span over several days is presented in Pebesma et al. (2000). In their work, only pairs of observations that were collected on the same day are included

during variogram estimation. To follow this approach, however, one needs to ensure that enough data is collected on a single day.

In this work, spatial dependence in the residuals is attributed to the spatial pattern of environmental conditions for which data are not available and thus, cannot be incorporated in the deterministic part of the geostatistical model. We have used the example of sea water temperature as the factor responsible for spatially structuring the residuals in the geostatistical model (e in equation 2). Given that the deterministic mean μ has been kept constant in space and time ($\mu = 10$), the spatial pattern of water temperature reflects the preference of seabirds for certain locations in the study area but has no effect on the average number of birds in the study area. If one would like to incorporate in the analysis the effect of dynamic environmental conditions over days on changes in seabird mean abundance, a temporal correlation in the mean μ should be imposed in addition to the temporal correlation in the residuals.

Changes in the spatial pattern of sea water temperature over consecutive days have been incorporated as temporal correlations in the residuals of the geostatistical model (e in equation 2). In situations where large temporal changes occur, aggregating the data over a four-day survey may greatly differ from the working assumption that data were collected on a single day under the same pattern in sea water temperature. When high variability in the spatial pattern of influential environmental conditions between consecutive days is to be expected in the area, it may be advisable to carry out the bird count survey on a single day using a larger number of ships or to use another survey technique that allows surveying a larger area on a single day (e.g. aerial surveys). The same principle may apply to data collected during the post-construction surveys.

This study focused on temporal correlations in the residuals from day to day. Variability in the spatial pattern of, for example, sea water temperature within a single day can also occur. Such a situation is shown in Fig. 1. For example, for the fifth of May during the morning, the spatial pattern of sea water temperature was different than the one for the afternoon. If observations on seabird abundance were collected during the morning and afternoon, the variogram of the residuals may not be representative of the variogram that is estimated assuming that the pattern of water temperature has not changed over the day. The types of analyses we presented can be applied to investigate the effects of aggregating data over a single day whenever changes are occurring in the spatial pattern of influential environmental conditions at small temporal scales, e.g. within a day.

The overlap in variogram parameters is largely influenced by the configuration of survey locations. For example, a change in pattern of environmental conditions between day 1 and day 2 produces a higher overlap between the T-variogram and S-variogram because the set of observations in these two days are more distant as compared to sets of observations between other days.

We have seen that the estimated values for the nugget, partial sill and range parameters of the variogram model depend on the temporal correlation in environmental conditions amongst survey days. The estimated values have, in turn, an effect on power results. For example, when the differences in environmental conditions over days increase, the non-spatial variation in the residuals increases, as reflected by an

increase of the nugget value. As a result, pairs of observations at distances lower than the range are less correlated, leading to an increase in power.

Besides providing insight in the effect of temporal aggregation of datasets on the estimate for spatial correlation, the method presented could be used to perform a sensitivity analysis. This type of analysis would allow investigating the effects of temporal data aggregation on power in a particular impact study. If one would have qualitative knowledge on, for example, sea water temperature patterns in the study area one could estimate the variogram parameters for post-construction situations under different environmental conditions and determine whether overestimation, underestimation or the same power is likely to occur. Underestimation of power may affect the economical investment for a survey requiring unnecessary extra survey effort, and related costs, to increase the probability of detecting impact. On the other hand, overestimation of power results may be more serious as potential impact may continue more easily without being detected.

In the impact assessment method presented in Pérez-Lapeña et al. (2010), the variogram used for the reference situation is also estimated from an aggregated dataset assuming that data were collected on a single day. This variogram is assumed to be representative for the post-construction situation. If there was qualitative knowledge of water temperature patterns based on, for example, visual observations of turbidity at the moment of the pre- and post-construction surveys, the methodology we presented could serve to better interpret impact results.

5 Conclusions

A space-time analysis of residuals in a statistical model has been carried out. Under specific scenarios, we have looked at the effects of changing the spatial pattern of environmental conditions between survey days on the validity of using an aggregated dataset over a four-day period as a representation of a single day survey. The effects have been investigated in the context of impact assessment and power analysis. Our main conclusions are:

- In those situations where the goal of the study is to estimate the spatial dependence in the observations of a particular species to construct the reference situation, aggregating data over a four-day survey increases the non-spatial variation in the data. As a result, the estimated spatial variation will not reflect the true spatial variation as if all data were collected on a single day.
- The configuration of survey locations amongst the days is an important factor when assessing the validity to aggregate data over a four-day survey as if all data were collected on a single day.
- If the assumed changes in the spatial pattern of environmental conditions over days are different from the actual changes, one can underestimate or overestimate the ability to detect impacts that are possibly occurring. For the cases investigated, estimates of power could differ up to a factor of two.

Acknowledgments This work is part of the project PhD@Sea, which is substantially funded under the BSIK-programme of the Dutch Government and supported by the consortium WE@Sea. We acknowledge NoordzeeWind and IMARES for the baseline impact study dataset.

Appendix: Overlap in T-variogram and S-variogram parameters

Area of overlap between two unit-area histograms of estimated values of variogram parameters for the T-variogram and S-variogram (Figs. 8, 9, 10).

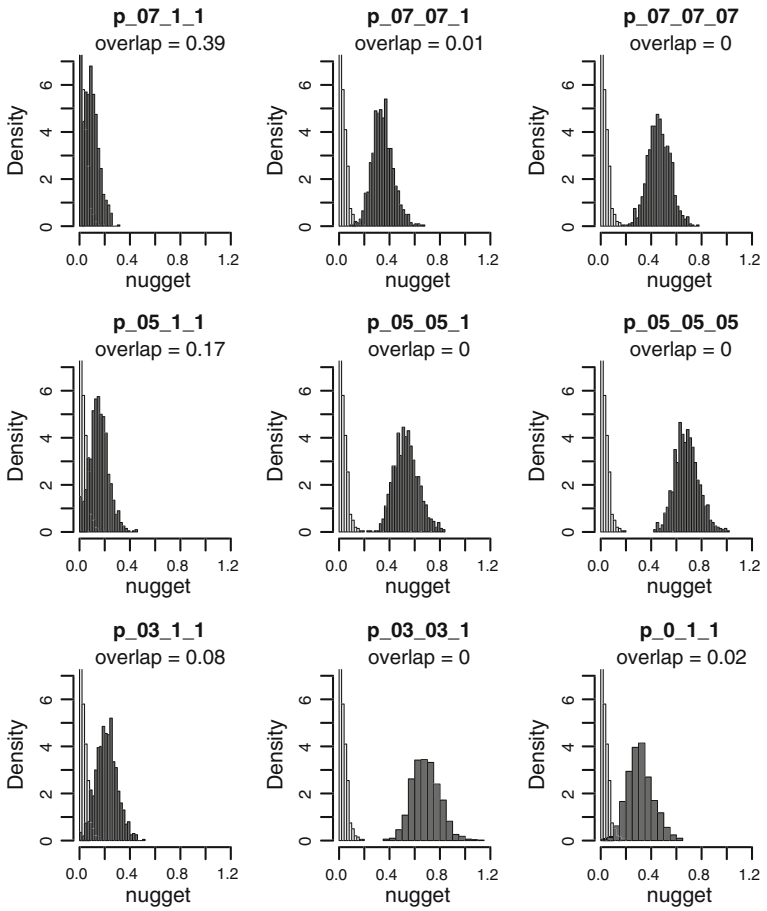

Fig. 8 Area of overlap between the unit-area histogram of estimated nugget values for the T-variogram (*light grey colour*) and S-variogram (*dark grey colour*)

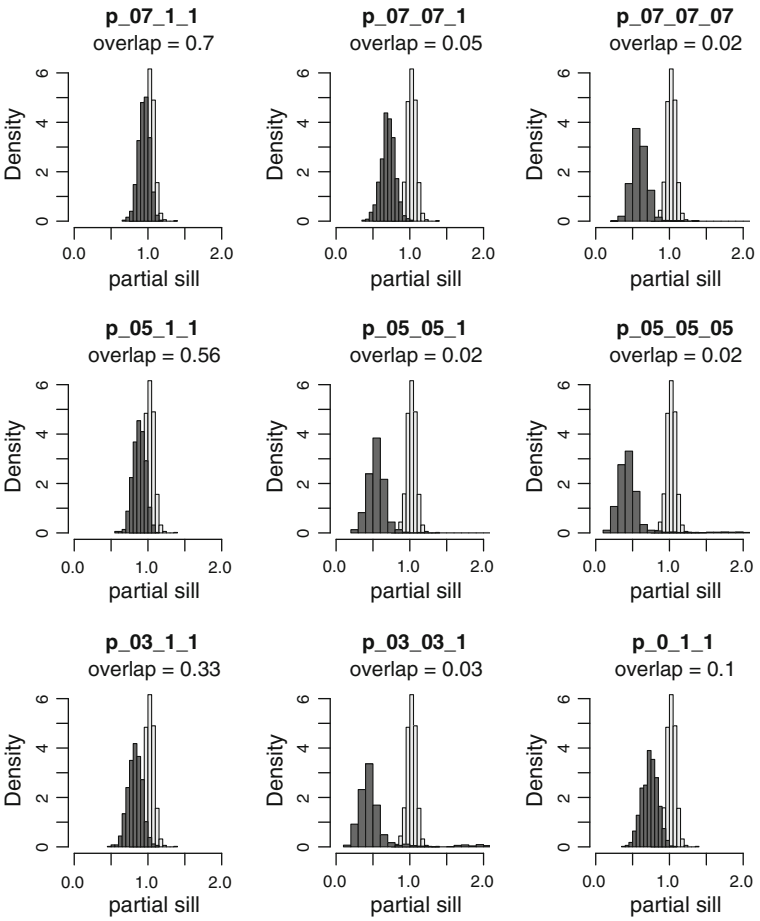

Fig. 9 Area of overlap between the unit-area histogram of estimated partial sill values for the T-variogram (*light grey colour*) and S-variogram (*dark grey colour*)

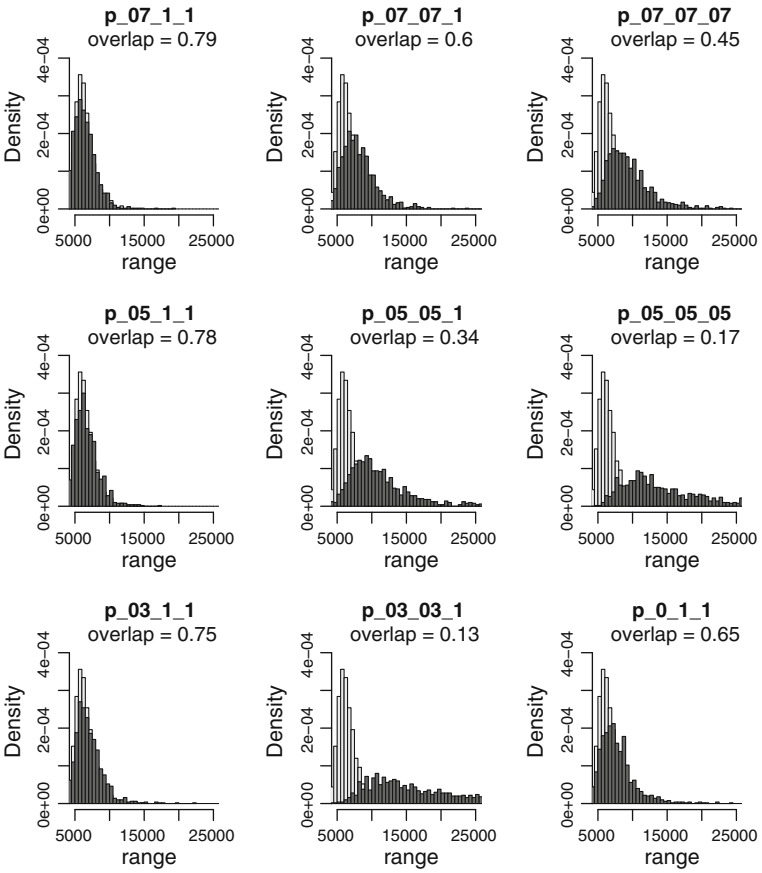

Fig. 10 Area of overlap between the unit-area histogram of estimated effective range values for the T-variogram (*light grey colour*) and S-variogram (*dark grey colour*)

References

Ainley DG, Spear LB, Tynan CT, Barth J, Pierce SD, Ford RG, Cowles TJ (2005) Physical and biological variables affecting seabird distributions during the upwelling season of the northern California current. Deep Sea Res Part II Top Stud Oceanogr 52(1–2): 123–143. doi:10.1016/j.dsr2.2004.08.016

Crain CM, Halpern BS, Beck MW, Kappel CV (2009) Understanding and managing human threats to the coastal marine environment. Ann N Y Acad Sci 1162: 39–62. doi:10.1111/j.1749-6632.2009.04496.x

de Boer GJ, Pietrzak JD, Winterwerp JC (2009) SST observations of upwelling induced by tidal straining in the Rhine ROFI. Cont Shelf Res 29(1):263–277. doi:10.1016/j.csr.2007.06.011

Diniz-Filho JAF, Bini LM, Hawkins BA (2003) Spatial autocorrelation and red herrings in geographical ecology. Global Ecol Biogeogr 12: 53–64. doi:10.1046/j.1466-822X.2003.00322.x

Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G, Davies RG, Hirzel A, Jetz W, Kissling WD, Kühn I, Ohlemüller R, Peres-Neto PR, Reineking B, Schröder B, Schurr FM, Wilson R (2007) Methods to acount for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30: 609–628. doi:10.1111/j.2007.0906-7590.05171.x

Fauchald P, Erikstad KE, Skarsfjord H (2000) Scale-dependent predator-prey interactions: the hierarchical spatial distribution of seabirds and prey. Ecology 81(3): 773–783. doi:10.1890/0012-9658(2000)081[0773:SDPPIT]2.0.CO;2

Gardner TA, Barlow J, Chazdon R, Ewers RM, Harvey CA, Peres CA, Sodhi NS (2009) Prospects for tropical forest biodiversity in a human-modified world. Ecol Lett 12(6): 561–582. doi:10.1111/j.1461-0248.2009.01294.x

Genton MG (2007) Separable approximations of space-time covariance matrices. Environmetrics 18: 681–695. doi:10.1002/env.854

Gill AB (2005) Offshore renewable energy: ecological implications of generating electricity in the coastal zone. J Appl Ecol 42(4): 605–615. doi:10.1111/j.1365-2664.2005.01060.x

Goovaerts P, Jacquez G (2004) Accounting for regional background and population size in the detection of spatial clusters and outliers using geostatistical filtering and spatial neutral models: the case of lung cancer in Long Island, New York. Int J Health Geogr 3: 1–23. doi:10.1186/1476-072X-3-14

Legendre P, Dale MRT, 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(5): 601–615. doi:10.1034/j.1600-0587.2002.250508.x

Leopold M, Camphuysen C, ter Braak C, Dijkman E, Kersting K, van Lieshout S (2004) Baseline studies North Sea Wind Farms: Lot 5 Marine Birds in and around the future sites Nearshore Windfarm (NSW) and Q7. Technical report, Alterra

Leopold M, Camphuysen C, Verdaat H, Dijkman E, Meesters H, Aarts G, Poot M, Fijn R (2010) Local birds in and around the Offshore Wind Park Egmond aan Zee (OWEZ) (T-0 & T-1). Technical report, IMARES Wageningen, UR

Maclean IM, Skov H, Rehfisch MM (2007) Further use of aerial surveys to detect bird displacements by offshore wind farms. Research Report 482, BTO, BTO, Thetford

Pebesma EJ, Duin R, Bio A (2000) Spatial interpolation of sea bird densities on the Dutch part of the North Sea. Technical report, Utrecht University

Pebesma EJ, Duin RNM, Burrough PA (2005) Mapping sea bird densities over the North Sea: spatially aggregated estimates and temporal changes. Environmetrics 16(6):573–587. doi:10.1002/env.723

Pérez-Lapeña B, Wijnberg KM, Hulscher SJMH, Stein A (2010) Environmental impact assessment of offshore wind farms: a simulation-based approach. J Appl Ecol 47(5): 1110–1118. doi:10.1111/j.1365-2664.2010.01850.x

Pérez-Lapeña B, Wijnberg KM, Stein A, Hulscher SJMH (2011) Spatial factors affecting statistical power in testing marine fauna displacement. Ecol Appl 21: 2756–2769. doi:10.1890/10-1887.1

Perry JN, Liebhold AM, Rosenberg MS, Dungan J, Miriti M, Jakomulska A, Citron-Pousty S (2002) Illustrations and guidelines for selecting statistical methods for quantifying spatial pattern in ecological data. Ecography 25(5): 578–600. doi:10.1034/j.1600-0587.2002.250507.x

Petersen IK, Christensen TK, Kahlert J, Desholm M, Fox AD (2006) Final results of bird studies at the offshore wind farms at Nysted and Horns Rev, Denmark. Technical report, National Environmental Research Institute (NERI)

R Development Core Team (2008) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0

Schneider DC, Piatt JF (1986) Scale-dependent correlation of seabirds with schooling fish in a coastal ecosystem. Marine Ecol Prog Ser 32:237–246

Skov H, Durinck J, Andell P (2000) Associations between wintering avian predators and schooling fish in the Skagerrak–Kattegat suggest reliance on predictable aggregations of herring *Clupeaharengus*. J Avian Biol 31(2): 135–143. doi:10.1034/j.1600-048X.2000.310205.x

Author Biographies

B. Pérez-Lapeña obtained her Ph.D. degree from the Water Engineering Department at the University of Twente (The Netherlands). Her research focused on the influence of seabird behavior and survey characteristics on the ability to detect the impact of seabird displacement caused by an offshore wind farm. She is currently working as a post-doctoral researcher at the Geography department of the Southern Illinois University Carbondale (USA) focusing on the impact of climate change on land use choices and their effects on water quality under a variety of technological, economic, and policy scenarios.

K. M. Wijnberg graduated in physical geography from Utrecht University, The Netherlands, in 1990. She received her Ph.D. degree in geographical sciences from that same university in 1995. Since then she has held a postdoctoral position at Oregon State University, College of Oceanic and Atmospheric Sciences (USA), and a postdoctoral position at Utrecht University, Department of Physical Geography (The Netherlands). In 1999 she was appointed as assistant professor at the latter department. In 2000 she chose to leave academia to pursue some personal interests. In 2003 she became a researcher/consultant at Delft Hydraulics in The Netherlands, in the marine and coastal section. In 2005 she resumed her academic career as an assistant professor at the University of Twente, Department of Water Engineering and Management (The Netherlands). Her research interests include, amongst others, the impact of human activities on the marine and coastal environment.

A. Stein graduated in mathematics and computer science from the University of Eindhoven, The Netherlands, in 1983. In 1985, he joined the Department of Soil Science and Geology as an assistant professor in statistics at Wageningen University. He received the Ph.D. degree in agricultural and environmental sciences from Wageningen University in 1991. In 1992 he was appointed as an associate professor and in 1995 as a visiting professor at the International Institute of Applied Earth Observation and Geo-Information (ITC). In 2000, he became a full professor of statistics at Wageningen University, and in 2002 he became a full professor of Mathematical and Statistical Methods for Geodata at Twente University. His research activities are focused on spatial statistics and spatial data quality, with special attention on earth observation and remote sensing.

S. J. M. H. Hulscher From 2002 Prof Dr. Suzanne J. M. H. Hulscher is head of the group Water Engineering and Management at the university of Twente. Originally she is educated as a theoretical physicist, MSc in 1991. She received her Ph.D. grade in 1996 at the faculty Physics and Astronomy, on the topic modeling of bed patterns in coastal seas. The research was conducted at Deltares (former WL|Delft Hydraulics) and IMAU, Utrecht University. Hereafter she held several scientific positions within the Cluster Civil Technology and Management at the University of Twente, from 2002 she holds the chair Physics of Watersystems. As guest scientist she stayed in Canada and Spain. In 2002 she won the Miverva-award of FOM (Fundamental Research on Matter). In 2003, she was awarded a VICI-grant of the NWO (Dutch organization of scientific research) to investigate roughness modeling for water management applications. Her research focuses on biogemorphology, evolution of morphological systems under human interference and sediment dynamics in rivers, coastal seas and lakes.