wadden sea
LONG-TERM ECOSYSTEM RESEARCH

Sound, innovative and connective monitoring
for the Wadden Sea area

SIMULATION APPROACH

A SIMULATON-BASED APPROACH TO EVALUATE DIFFERENT SAMPLING PLANS FOR LONG TERM ECOLOGICAL MONITORING IN THE WADDEN SEA

Emiel van Loon

Universiteit van Amsterdam

VERSION 20150831

AUG 2015

TABLE OF CONTENTS

1 INTRODUCTION.......... 5

2 THE CONCEPT OF SIMULATION-BASED EVALUATION.......... 7

3 SPATIAL SAMPLING OF A CONTINUOUS PROCESS USING ARTIFICIALLY GENERATED DATA.......... 10

3.1 Description.......... 10
3.2 Results.......... 11
3.3. Discussion.......... 14

4 SPATIAL SAMPLING OF A CONTINUOUS PROCESS USING FIELD DATA.......... 16

4.1 Description.......... 16
4.2 Results.......... 17
4.3 Discussion.......... 19

5 CONCLUSIONS.......... 21

6 REFERENCES.......... 22

7 APPENDIX 1 – CODE FOR CASE STUDIES.......... 23

ABSTRACT

This reports explains and demonstrates a simulation-based method to evaluate sampling plans in ecological monitoring. It is meant to provide an interdisciplinary-team of scientists with a means to specify, conduct and analyse similar experiments for relevant questions, field realities and systems of interest. In the method an ecological reality is simulated by a model, providing a 'truth', and predicted back by different approaches. When predicting the truth back, the (incorrect) prediction model and sampling properties are varied. Subsequently, the deviations from the truth for the different model and sampling properties are analysed – in this case via a linear model. After presenting the abstract workflow, it is applied in two case studies to further clarify its details: a) the evaluation of different interpolators and different observation densities to observe a continuous process in the sub-littoral part of the Wadden Sea and b) the evaluation of two existing monitoring programs to describe species occurrence in the sub-littoral part of the Wadden Sea. The results of the two case studies are briefly discussed and the codes used for the two case studies are provided as an example for future applications.

1 INTRODUCTION

The design and implementation of monitoring in the realm of long term ecological conservation and research is an area of active research where no general solutions exist that apply to a broad range of cases, but rather tailor-made solutions need to be created (e.g. Gitzen et al. 2012; Yoccoz et al. 2001). General guidelines may provide a direction to design an adequate monitoring plan, while more elaborate analysis is required to define the specifics and to evaluate the often many choices with (sometimes counter-intuitive) effects on monitoring outcomes. Usually this elaborate analysis involves the specification of monitoring aims and translation of practical constraints and knowledge (e.g. due to available equipment, staffing, dynamics of the system, prior knowledge about the system) in a coherent framework for statistical analysis. Once this has been achieved, a considerable literature exists on the statistical methodologies to design an adequate monitoring plan (e.g. de Guijter et al., 2006; Tompson, 2010) and by the building blocks from that literature relevant monitoring plans can be built. In this context, it is not the last step – finding an optimum plan once the complex questions and constraints involved in ecological monitoring have been cast in to a clean and consistent form – but the former step to structure the questions and constraints that form the major challenge. An important reason for this is that this step typically involves various experts with different backgrounds and different roles relating to the monitoring program. Hence, these different parties need to share information and develop a common understanding of the problem at hand. Another important reason is the fact that the problem is simply complex and not well-defined to start with: in the practice of (long term) ecological monitoring, monitoring questions can often not be defined very precisely, may be arbitrary to some degree, may change over time or may even lead to conflicting design-properties. The usual constraints on the resources available for monitoring (so that only a fraction of the possible system components could be monitored) often influence the monitoring questions implicitly. Next, considerable uncertainties with regard to observation accuracy of various field measurements (sometimes varying with environmental conditions) are common. And finally, there may be a general lack of knowledge about the important drivers and dynamics of the system under consideration.

This report is aimed at this first step in particular: to share and structure knowledge among different actors in the design of a monitoring program in the context of ecological monitoring. It presents a simulation-based approach to specify and evaluate monitoring questions and designs. Because it is conceptually simple to understand, the specification of different simulations as well as the interpretation of the results can be understood (and perhaps also conducted) by all the actors involved.

A monitoring plan (optimal or not) is not a goal in its own right, but forms part of a monitoring project where the formulation of policy, management and science goals and questions, the actual monitoring effort, storage, curation and analysis of data and decisions based on the results are all vital parts. According to Lindenmayer and Likens (2010) effective monitoring programs are characterized by: 1) good questions; 2) a conceptual model of an ecosystem or population; 3) strong partnerships between scientists, policy-makers and managers; and 4) frequent use of data collected. Hence it is probably not the technical design of a monitoring plan that contributes to the success of a monitoring program, but especially the way in which the monitoring plan is embedded in and providing support to these four characteristics. Simulation tools may help to achieve this.

This report first provides a general explanation and an abstract workflow for the simulation-based evaluation of a monitoring plan. Next, the abstract workflow is clarified by applying it in two case studies. The case studies investigate questions about the spatial layout of shellfish monitoring in the sub-littoral part of the Dutch Wadden Sea. The set-up of the simulation experiments well as the most interesting results are shown for each case study and the software (R scripts) used for each of the case studies is provided in two appendices.

2 THE CONCEPT OF SIMULATION-BASED EVALUATION

The general workflow of evaluating a specific procedure or measurement plan by simulation is conceptually straightforward and has been applied frequently in ecology (see e.g. Zurell et al. 2010). For brevity the ‘Simulation-based Evaluation of Monitoring’ approach will be called SEM in this document. Figure 1 sketches the SEM-workflow.

The procedure starts with a system concept to generate a data set (alternatively one could also draw sub-samples from a high-quality and high-resolution data set) (A). The data in A is treated as ‘truth’ and subsequently a sample is drawn according to some technique whereby the sampling procedure may add error to the truth (B). On the basis of the sample the truth is estimated, using knowledge about the data-generating system or by (purposively) assuming an erroneous system representation (C). The estimated truth is then evaluated by comparing it to the real truth trough a criterion function (D). Finally, the relation between one or more criteria and various data or sampling properties is analysed (E).

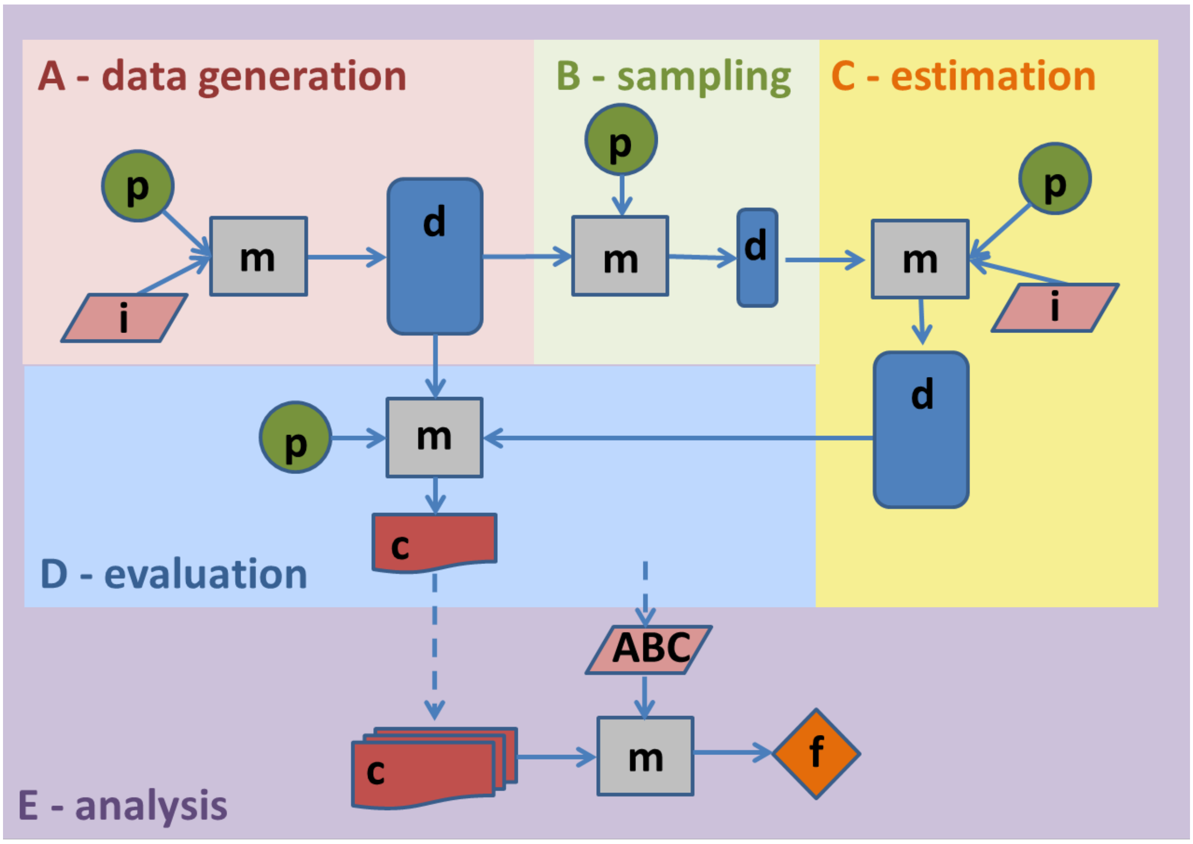

Figure 1. Sketch of the SEM-workflow. The small letters in this figure mean the following: c. criterion to measure the quality of a data set relative to a reference data set; d. output data from a model; f. final SEM-result; i. input data for a model; m. model; p. model parameters and settings. The SEM-procedure generates a ‘truth’ data set (A), samples from this set (B), estimates on the basis of this sample and an assumed model (C) and evaluates the result (D). Different combinations of truth, sampling and estimation can be evaluated and eventually analysed (E).

While Figure 1 presents the concept of a SEM-workflow to provide an overview, a more specific list of actions is required to build a workflow for a specific case at hand. Such a list is given below.

Definition:

1. Specify a model that describes the distribution of a species – to generate the truth (model mA). This model should contain as much ecological realism as possible (and/or necessary), e.g. dynamic and spatially distributed inputs or realistic spatio-temporal heterogeneity.
2. Specify one or more models that describe the way a population has been sampled (mB_1 to mB_S). These models should match the sampling practice as close as possible (e.g. realistic values for data-loss, and spatio-temporal support).
3. Specify one or more models that are used to estimate the truth (mC_1 to mC_P). One extreme would be a model that is identical to the data-generating model (mA), whereas the other would be a model that lacks any representation of the underlying system such as a heuristic interpolation procedure. The first represents the maximum attainable accuracy under a given sampling plan while the latter represents the minimum accuracy that may be expected from the sampling plan.
4. Specify one or more evaluation criteria to measure the aspects of the system that should be observed well in an eventual monitoring plan (mD_1 to mD_Q).
5. Specify an analysis procedure or model (model mE) to relate each of the Q evaluation criteria to choices in steps 1 to 4.

Calculation:

6. Conduct the computer experiment:
   a. apply model mA to generate *R* realisations (data sets dA_1 to dA_R);
   b. apply each model mB_s to sample from each realisation (data sets dB_1 to $dB_{R,S}$);
   c. apply each model mC_p to estimate data (data sets dC_1 to $dC_{R,S,P}$);
   d. calculate each evaluation criterion for the $dC_{R,S,P}$ data sets (evaluation criteria cD_1 to $cD_{R,S,P,Q}$).
7. Analyse the results from the computer experiment:
   a. relate each criterion to the different treatments, i.e. different (properties of) sampling plans mB and models mC;
   b. rank the results of the different criteria and decide about the preferred combination of sampling plan(s) and model(s) to be used.

In summary: the SEM-workflow comprises two main steps: definition and calculation. The definition-part of this workflow describes four models and an analysis procedure. The calculation-part sequentially applies each of the models to the relevant input-data and parameters and subsequently analyses the results through a meta-analysis. The analysis part (step 7) is typically least well defined a-priori and may involve some explorative analysis an interactive visualization to achieve a good understanding of the results and before a meta-model can be specified. By evaluating a number of different sampling plans for various system representations (and potentially also for different types of data), the procedure quickly leads to numerous combinations. However, if all combinations are simply evaluated, as in a balanced experimental design, the results can generally be analysed with standard linear models (step 7a) and a decision theoretic procedure (step 7b), which can handle large amounts of data and numerous treatment levels. In the two case studies that are being presented here, only one criterion is evaluated per SEM-experiment, and therefore step 7b is (however applicable in general) not considered in what follows.

3 SPATIAL SAMPLING OF A CONTINUOUS PROCESS USING ARTIFICIALLY GENERATED DATA

3.1 Description

In this case study the aim is to provide a relation between estimation-variance, sampling density and different interpolation methods for a spatially continuous process in the Dutch Wadden Sea. The process could represent a biophysical variable (such as salinity, sediment texture, phytoplankton biomass), which would be observed by spatial point sampling on a regular grid. The model used here is very similar to a correlation-model used in Bijleveld et al. (2012), resembling the autocorrelation of *Nereis diversicolor* density (Fig 2 in that study).
Similar to the approach in Bijleveld et al. (2012) here it is assumed that no (relevant) process knowledge about the system can be incorporated into the models, hence interpolators without any system-information are used for estimation. As an evaluation criterion, the RMSE is used, based on a sample of evaluation points that are placed between the 'observation points'. A quantitative summary of this experiment is given in Table 1.

Table 1. Summary of a SEM-experiment, evaluating the effect of spacing of the observation grid and different interpolation models on estimation accuracy.

A - Type of process:	
Continuous	Unconditional Gaussian simulation generating 20 realizations; point-values at a grid of 100 m in the intertidal area of the Dutch Wadden Sea, using the following exponential semi-variance function: sv = 0.5(1-exp(-lag/2000))+0.1
B - Sampling characteristics:	
Point sampling at regular grid	1) 500m 2) 1000m 3) 1500m 4) 2000m
C - Estimation methods:	
Interpolation and smoothing towards points	1) Linear 2) inverse dist. weighted 3) kriging with correct variogram 4) kriging with an incorrect variogram (sv = 0.6(1-exp(-(lag/2000)^2))) 5) loess smoother
D - Evaluation criteria:	
RMSE at evaluation points	At 700m, with distances of 1000m; omitting any points that would require extrapolating beyond the convex hull of observation points.
E - Analysis:	
Linear model	Relating the evaluation criterion to 4 levels of sampling and 5 different estimation methods

3.2 Results

The truth-data generated through unconditional Gaussian simulation generates a field with values ranging from approximately -4 to +4. Three examples of this field are shown in Figure 2.

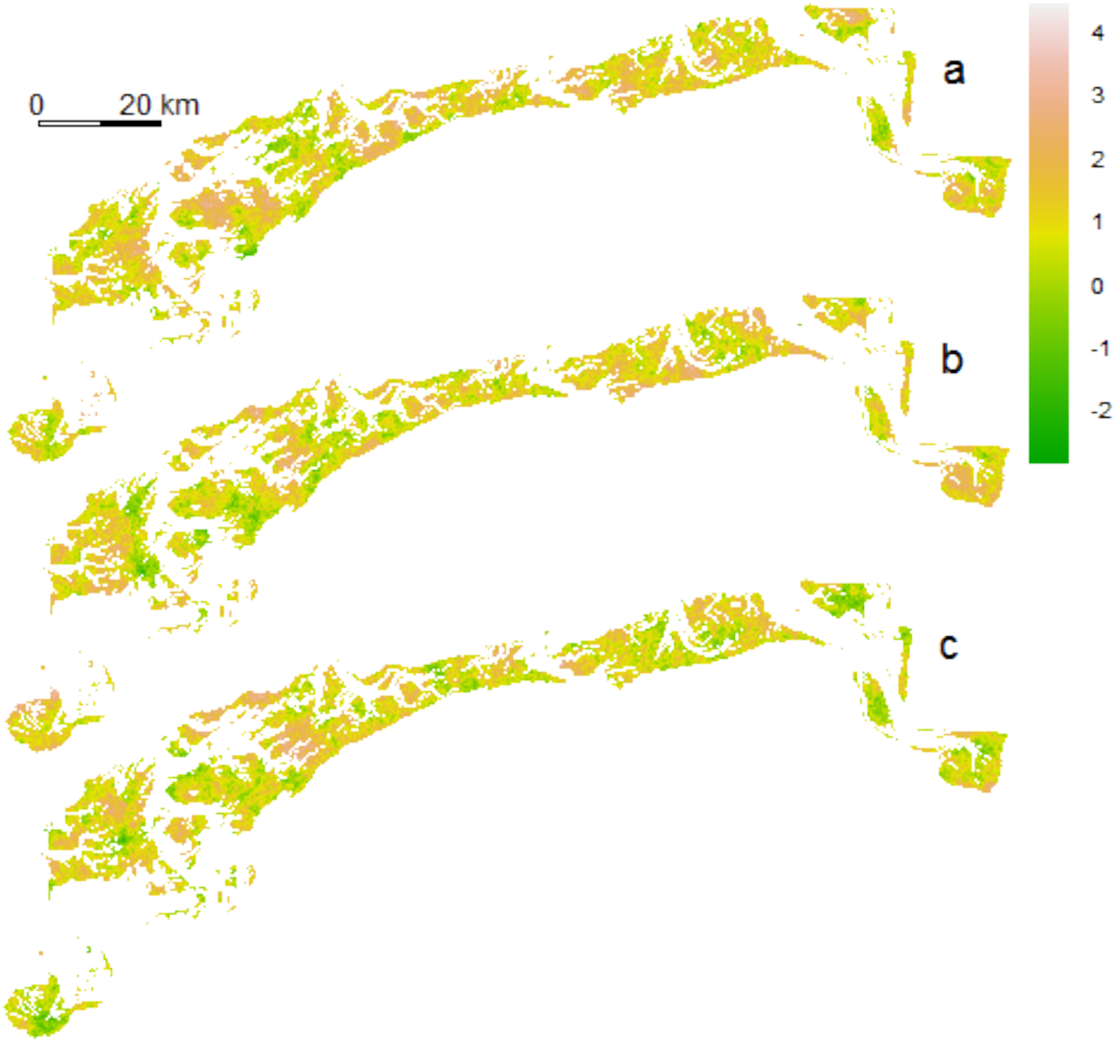

Figure 2. An impression of the fields generated in Case study 1 by Gaussian simulation (maps a to c represent the first three out of 20 fields)

When considering the different estimates, it becomes clear that in this case study the similarities are large (and declining slightly with increasing spacing of the observation grid). As an example, the results for linear interpolation and kriging (with correct variogram) are shown in Figure 3. The errors for the two interpolation methods are shown at the bottom (left and middle histogram) and the difference between the two estimates is shown in the histogram at the right. The errors for the different interpolators are strongly correlated (hence the narrower distribution in the rightmost histogram compared to the other histograms). The pattern shown in this figure, strong correspondence among interpolators, is prevalent throughout all the estimates in this case study, with one exception: the loess smoother gives a poor fit and does not correspond to the (other) methods.

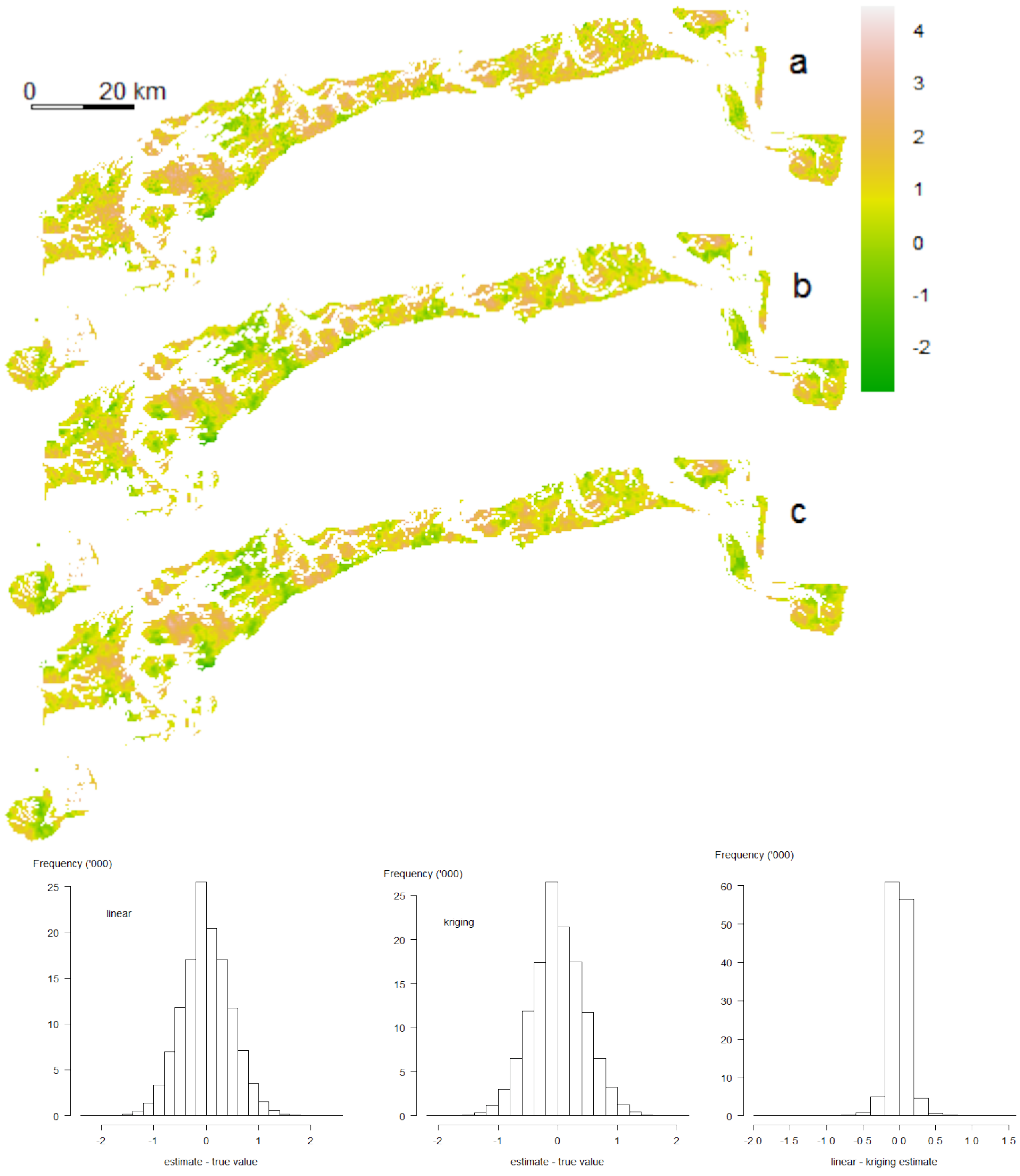

Figure 3. An example of the similarity between estimates based on a sample of the field on a 500 m grid via linear interpolation (b) and kriging using the correct system variogram (c) (the truth is shown in a). The bottom row shows the histograms of the residual difference for the different estimates as well as the difference between linear interpolation and kriging.

Figure 4a highlights the relatively poor performance of the loess smoother. As this smoother will not form an attractive estimation method in this case, and may obfuscate any interesting relations between interpolation method, observation density and RMSE, it is further omitted from the analysis.

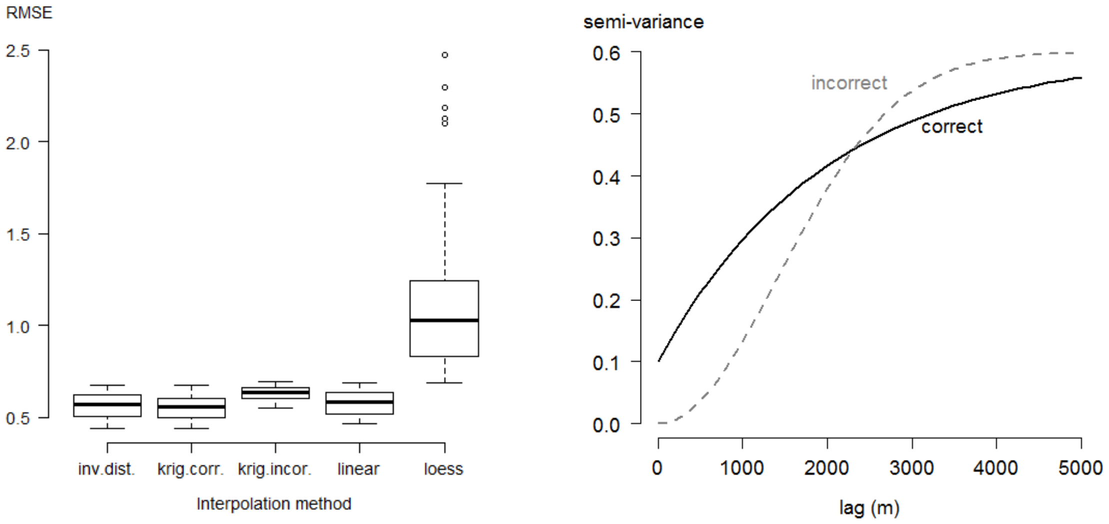

Figure 4. a) RMSE for estimating a continuous variable by different methods. The poor performance of the loess smoother stands out in this case. b) the difference between kriging with a correct and an incorrect semi-variogram are due to the differences between an exponential (shown in black solid) and Gaussian (grey dashed) semi-variogram. The actual semi-variograms that are used for estimation are fitted on the available sample-data so in practice the differences are smaller.

By considering both the effect of sampling density and interpolation method on RMSE, it turns out that a linear model which involves both terms with an interaction term, explains RMSE well (94% of the variance; residuals close to Gaussian). The observation density explains most of the variation in RMSE. While kriging with the correct semi-variogram leads on average to the lowest RMSE (especially at high observation densities), the differences between this method, inverse distance weighted interpolation, and linear interpolation are small and not relevant with respect to substantive questions (less than 5% of the RMSE values). Kriging with an incorrect variogram, however, leads to considerably higher RMSE-values (more than 20% of the RMSE values). The differences between the two kriging methods are completely due to the use of a different variance model (see Figure 4b). The results of this linear analysis are visualized in Figure 5 by considering each of the interpolators separately.

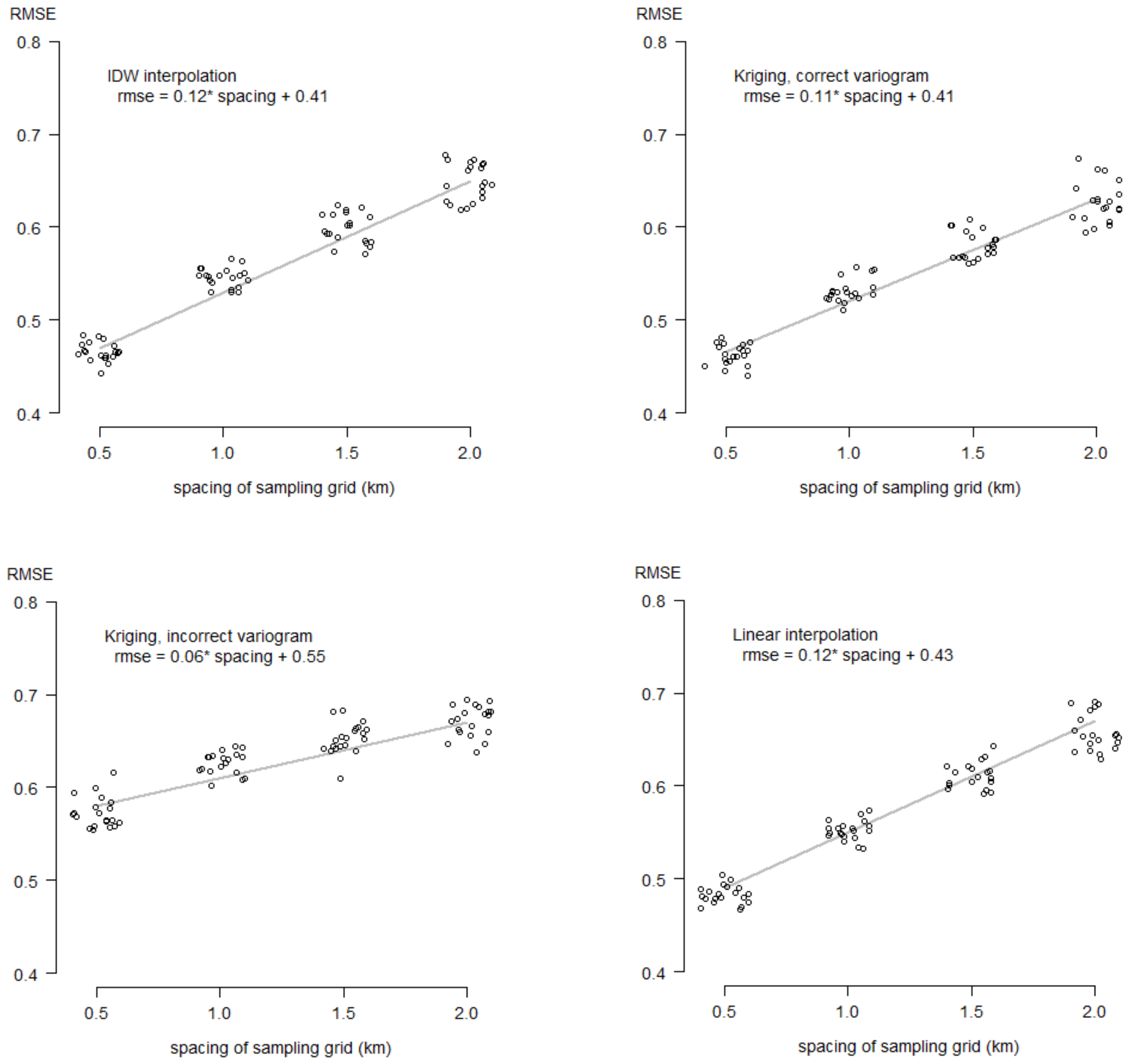

Figure 5. RMSE as a function of sampling density and different interpolation methods. For each method the results for the SEM-experiment are shown by the circles (the horizontal variation for the circles has been added for visualisation purposes) and a least squares estimate of the linear model per interpolation method is given by the equations and grey lines.

3.3. Discussion

The results from case study 1 show that the method to estimate a spatially varying process does matter (viz. loess versus the other methods) – but that one does not have to use the correct system representation (which will not be the case in a situation with real data anyway) to obtain an estimate that is close or equivalent to the optimum attainable. In this case, the relatively simple linear and inverse distance weighted interpolators gave good estimates of the property of interest.

This SEM-experiment is simple and incomplete for any realistic application or question in relation to field-research. One would probably like to evaluate a range of spatial distributions (not just applying a single data generating mechanism and/or a single underlying variance structure), and also evaluate robustness of various estimation methods in situations of missing data, different types of spatial-sampling (e.g. in clusters or stratified) and perhaps also consider situations with much sparser sampling.

However, the workflow to conduct more elaborate SEM-experiments under these adjustments remains unchanged, and the scripts (Appendix 1) can be used as a template for these more extended analyses.

4 SPATIAL SAMPLING OF A CONTINUOUS PROCESS USING FIELD DATA

4.1 Description

In this case study the aim is to provide a relation between a classification-performance criterion (the area under the receiving-operator curve - AUC), the spatial layout of two existing monitoring networks, observation densities (sub-sampling from the current monitoring networks), different levels of species commonness/rarity and two different interpolation methods. The two different monitoring networks that are considered are MosKok[1] and SIBES (Compton et al. 2013) respectively.

The two main questions of interest are: 1) whether there are systematic and large differences among the two monitoring networks (across the different levels of commonness), 2) whether there are systematic and large differences between the two interpolation procedures (across the different levels of commonness). The layout of this SEM-experiment is similar to that in case study 1, and detailed in Table 2.

The sampling characteristics in this experiment are a bit complex. The reason for this is that MosKok and SIBES surveys apply different spatial sampling plans. MosKok samples lay along north-south transects with irregular distances between the sampling points. There are larger gaps between transects (east-west direction) than within the transects. The SIBES sample points are on a 500 m regular grid with additional sampling points randomly placed in between the main sampling points. SIBES has approximately 4 times more sampling locations than the MosKok survey (4400 verusus 1200 respectively). Furthermore both surveys use sampling instruments that differ with respect to sampling support. SIBES samples at 0.0173 m^2 (by boat) or 0.0177 m^2 (on foot) whereas MosKok uses a 0.4 m^2 (by boat) and 0.1 m^2 (on foot). In this analysis only the difference in spatial sampling design of the two surveys is investigated, while a note regarding the expected effect of the differences in support is added in the discussion.

In order to compare both surveys in a meaningful way, similar data-densities are created by taking smaller samples from both data sets. SIBES data is sub-sampled from originally (around) 4400 to 2200, 1100, and 550 and 275 points. MosKok data is sub-sampled from originally around 1200 to 600 and 300 points. The down-sampling is done in a nested way (each smallest subset is also part of a larger sub-set), and points that are retained are selected randomly. In this way, the three smallest sub-sets from the SIBES data can be compared directly to the MosKok sets of corresponding size.

[1] Many IMARES/CVO publications that make use of the MosKok data (which forms part of a larger benthos monitoring project in the Wadden and North Seas). The following report describes the sampling properties of this survey: Craeymeersch JA, Baars D, Brummelhuis E, Bult TP (2004) Handboek bestandsopnames en routinematige bemonsteringen van schelpdieren. Stichting DLO, Centrum voor Visserijonderzoek (CVO), CVO Rapport: CVO 04.004, pp 74.

Table 2. Summary of a SEM-experiment to evaluate the spatial layout of two existing monitoring networks (MosKok and SIBES) for estimating distribution maps.

A - Type of process:	
Binary	Unconditional Gaussian simulation generating 20 realisations; point-values at a grid of 100 m in the intertidal area of the Dutch Wadden Sea, using the following semi-variance function: sv = 0.025(1-exp(-lag/1000)). An inverse logit-transform is applied to the resulting fields; subsequently four threshold values are used to determine presence or absence: 0.71 (very common, prevalence of 0.75), 0.74 (common, prevalence of 0.4), 0.77 (rare, prevalence of 0.1) and 0.80 (very rare, prevalence of 0.01).
B - Sampling characteristics:	
Existing sampling locations (MosKok and SIBES), downsampled to lower densities	1) 4400 points(SIBES) 2) 2200 points (SIBES) 3) 1200/1100 points (MosKok/SIBES) 4) 600/550 points (MosKok/SIBES) 5) 300/275 points (MosKok/SIBES)
C - Estimation methods:	
Interpolation towards points	1) inverse dist. weighted interpolation 2) kriging Hence, simple continuum-based interpolators rather than binomial models are used to model this binary data.
D - Evaluation criteria:	
AUC, on the basis of classification at evaluation points	At 700m, with distances of 1000m; omitting any points that would require extrapolating beyond the convex hull of observation points.
E - Analysis:	
Linear model	Relating the AUC to the 5 levels of sampling density, 4 levels of commonness and two different estimation methods

4.2 Results

An example of the fields generated in this case study is given in Figure 6, showing different levels of commonness. The four binary fields are created from a single continuous field by applying different threshold values. The huge range of commonness, covered by the four levels is clear from this example.

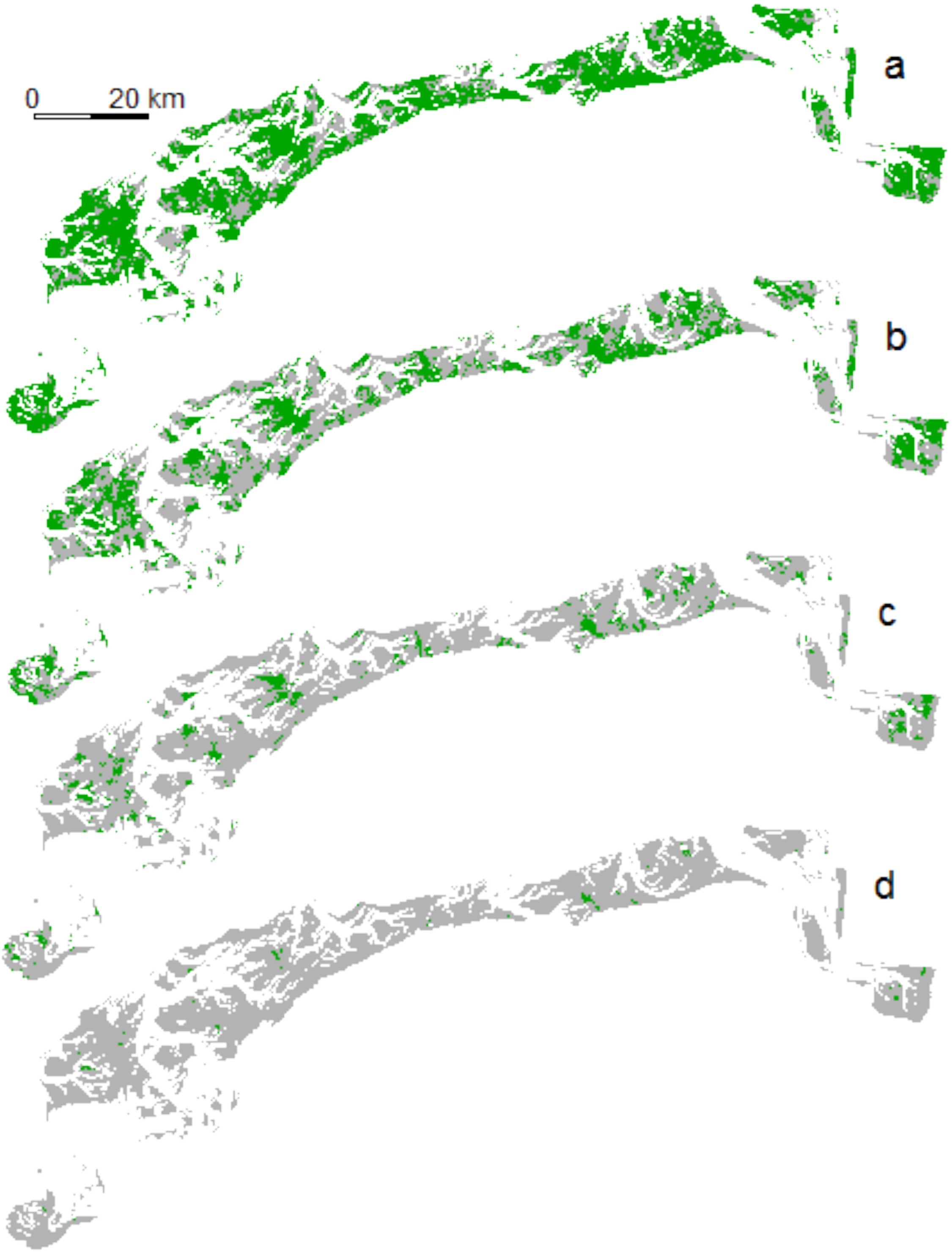

Figure 6. Example of a field with species occurrence (presence is indicated in green) at four levels of commonness. The upper map (a) shows a 'very common' species with a prevalence of 0.75, b) shows a 'common' species with a prevalence of 0.4, c) shows a 'rare' species with a prevalence of 0.1, and d) shows a 'very rare' species with a prevalence of approximately 0.01. It should be noted that prevalence is here defined as being present in a 100 by 100 m grid.

The main results of this case study are summarized in Figure 7, showing the AUC for all factors in this experiment. Along the y-axes of each plot the AUC-values (AUC varies from 0.5 up to 1, a value of 0.5 implies that a model has no better predictive ability than random classification). Along the x-axis of each plot the different sample sizes are shown (random perturbation in the x-direction is applied to enhance visibility). The two surveys are shown by different colors, the two methods by different symbols, and the four levels of commonness are shown in four different sub-plots.

The patterns that appear from this are:

1) The location of the SIBES observation points lead to a prediction that is systematically better than the MosKok locations; the difference varies with commonness (it is a bit larger for common species and almost disappears for rare species).
2) The AUC-value increments with sample size along a saturation-curve, does not vary much among common down to rare species (prevalence from 0.75 to 0.1) but drops considerably for very rare species (prevalence of 0.01).
3) The difference between interpolation methods is small and not systematic across all treatments. For larger sample sizes and common or very-common species, kriging outperforms inverse distance weighted interpolation. For small sample sizes and rare or very rare species, the two methods lead to the same AUC values.
4) From the rare to very rare class, it is not so much the decrease in AUC, but especially the increase in its variance along the entire range of sample sizes that is striking. Apparently the rarity is such that the distance among ‘presence-patches’ surpasses the correlation length for the data in this experiment.

If the experimental results for MosKok and SIBES are analysed separately, the model that describes the expected AUC best is comparable for both cases. It is a linear model with the number of observations, species commonness (4 classes), type of interpolator (kriging and inverse distance weighted interpolation) and an interaction term between species commonness and interpolator as predictors. The variance explained by this linear model for AUC is 0.70 for the MosKok data, while for the SIBES data (only analyzing the three smaller sample sizes – hence comparable to MosKok) it is 0.79. The details of this linear meta-analysis and its result are given in Appendix 2 (‘case2_analyse.r’).

4.3 Discussion

Similar to the case study for artificial data, it is clear that this experiment and its results cannot serve as a comprehensive evaluation of the MosKok and SIBES surveys. It does however provide a template to conduct such an evaluation. It also suggests that in order to obtain an overall impression of an optimal experimental plan, probably not the specific model form (two interpolators in this study) but other aspects like the rarity of the species/process under consideration and the specific error metric to be used are important factors to consider. In addition, the effect of variations in area and volume sampled and other field procedures deserve more in-depth investigation and could be incorporated in a SEM experiment.

Only the spatial placement of samples has been evaluated with regard to MosKok and SIBES, assuming a 100% observability of organisms at a 100 m grid resolution (i.e. if it is present in a 100 m grid cell, the observation procedure will find it). A 100% observability is a rather strong assumption, especially when considering the supports of 0.0173/0.0177 m^2 (SIBES) to 0.4/0.1 m^2 (MosKok). Knowing that patchiness in benthos is observed down to the meter-scale and that a considerable spatial randomness remains, it is likely that the 6 to 24 times larger supports in the MosKok survey has a considerable effect on the outcome (it may outweigh the differences in sample layout that have been studied here). To quantify these effects, SEM-experiments with fine-scale simulations of bivalve distributions would be required.

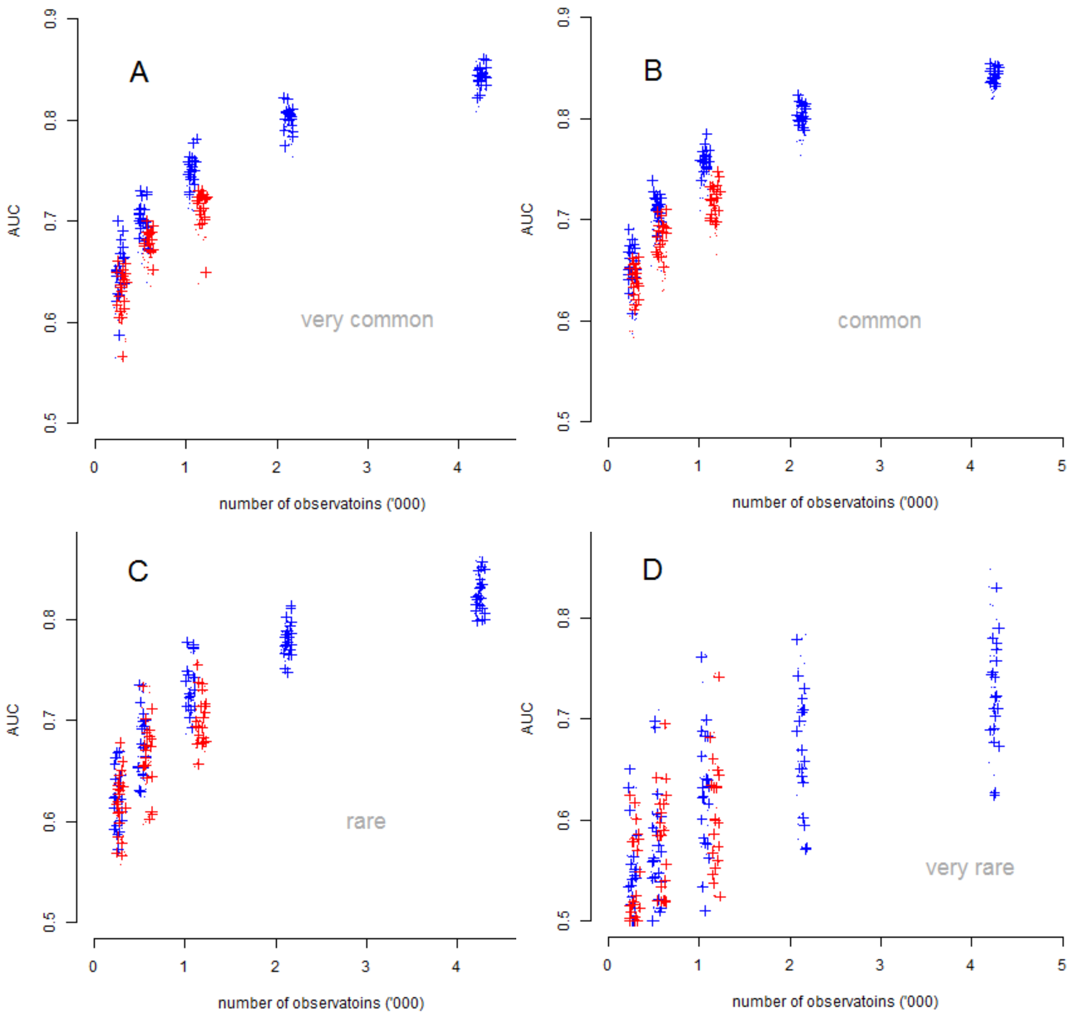

Figure 7. Relation between overall model performance (as measured by AUC), as a function of the number of observations. The data is broken down by survey type (red = MosKok, blue = SIBES survey) and type of interpolation technique inverse distance weighted interpolation (.) and kriging (+).

5 CONCLUSIONS

The development of suitable monitoring plans for (long term) ecological research, requires an oversight over policy, management or science questions (ranging from applied to theoretical), practical field methods and limitations, knowledge of bio-physical system functioning, statistical theory and state-of-the-art computational knowledge (e.g. Gitzen et al, 2012). It is and will always be difficult to bring this broad range of expertise together in an interdisciplinary project (leave alone a single person). And once the expertise has been brought together, it is difficult to translate all the relevant information between the various fields of expertise and generate a shared understanding. This report aimed to show that simulation-based experiments for monitoring (SEM) may be an appropriate communication platform among those different fields of expertise.

The description of the general SEM-workflow and two specific case-studies with R-code can hopefully help inter-disciplinary teams to specify, conduct and analyse similar experiments and thereby stimulate the exchange between (science, policy and management) questions, field reality and theoretical ideas. However, while aiming at keeping the code for this study easy to comprehend, the data-management structure has been kept simple. As a result, some of the code is not sufficiently abstract to be transferred directly to a different study (e.g. to analyse a monitoring plan for trend detection) and would need some reworking. Furthermore some extensions and changes may be required to make storage and organisation of in- and output more efficient in larger case-studies and make post-processing output more informative for the question at hand.

Getting back to the four characteristics of a good monitoring program by Lindenmayer and Likens (2010) (see introduction section), a SEM-experiment can provide a positive contribution to three of these: it does (by definition) provide a conceptual model of an ecosystem or population, and may in its function as communication platform also stimulate good questions and help to strengthen partnerships among scientists, policy-makers and managers.

Finally it should be stressed that there is a considerable literature on the topic of synthetic experiments in ecology (e.g. Zurell et al. 2010; Austin et al. 2006), and there are many ecological studies where simulation-based experiments are used. Hence there is nothing new about the SEM-experiments that are described in this report. Three examples of studies where SEM experiments were used in designing monitoring plans are Bijleveld et al. (2012), who compare five sampling designs for a benthic monitoring program to find a pareto-optimum to detect temporal change, generate accurate maps and estimate autocorrelation; Joseph et al. (2006), who compare two strategies for detecting trend, abundance, and presence–absence surveys; and Nuno et al. (2013), who evaluate a range of monitoring components to increase survey accuracy and precision for typical ungulates in a savannah ecosystem. Differently from these studies, this report has emphasized the general SEM-workflow and provided simple examples to capacitate a wider group of scientists in using SEM-experiments.

6 REFERENCES

Austin MP, Belbin L, Meyers JA, Doherty MD, Luoto M (2006) Evaluation of statistical models used for predicting plant species distributions: Role of artificial data and theory. *Ecological Modelling*, 199, 197–216.

Bijleveld AI, van Gils JA, van der Meer J, Dekinga A, Kraan C, van der Veer HW, Piersma T (2012) Designing a benthic monitoring programme with multiple conflicting objectives. *Methods in Ecology and Evolution*, 3, 526–536. doi:10.1111/j.2041-210X.2012.00192.x

T.J. Compton, S. Holthuijsen, A. Koolhaas, A. Dekinga, J. ten Horn, J. Smith, Y. Galama, M. Brugge, D. van der Wal, J. van der Meer, H.W. van der Veer, T. Piersma (2013) Distinctly variable mudscapes: Distribution gradients of intertidal macrofauna across the Dutch Wadden Sea. *Journal of Sea Research*, 82 (2013), pp. 103–116

Gitzen, R. A., Millspaugh, J. J., Cooper, A. B., & Licht, D. S. (Eds.). (2012). Design and Analysis of Long-term Ecological Monitoring Studies. Cambridge University Press.

de Gruijter JJ, Brus DJ, Bierkens MFP, Knotters M (2006) Sampling for Natural Resource Monitoring. Springer, New York.

Joseph, L. N., Field, S. A., Wilcox, C., & Possingham, H. P. (2006) Presence–Absence versus Abundance Data for Monitoring Threatened Species. *Conservation Biology*, 20, 1679–1687. doi:10.1111/j.1523-1739.2006.00529.x

Lindenmayer, David B., and Gene E. Likens (2010) The Science and Application of Ecological Monitoring. *Biological Conservation*, 143, 1317–1328. doi:10.1016/j.biocon.2010.02.013.

Nichols JD, Williams BK (2006) Monitoring for conservation. *Trends in Ecology & Evolution*, 21, 668–673. doi:10.1016/j.tree.2006.08.007

Nuno A, Bunnefeld N, Milner-Gulland EJ (2013) Matching observations and reality: using simulation models to improve monitoring under uncertainty in the Serengeti. *Journal of Applied Ecology*, 50,488-498. doi: 10.1111/1365-2664.12051

Thompson SK (2012) Sampling. 3rd edition. John Wiley & Sons, New York.

Yoccoz NG, Nichols JD, Boulinier T (2001) Monitoring of biological diversity in space and time. *Trends in Ecology & Evolution*, *16*, 446–453. doi:10.1016/S0169-5347(01)02205-4

Zurell D, Berger U, Cabral JS, Jeltsch F, Meynard CN, Munkemuller T, Nehrbass N, Pagel J, Reineking B, Schroder B, Grimm V (2010) The virtual ecologist approach: simulating data and observers. *Oikos*, 119, 622–635.

7 APPENDIX 1 – CODE FOR CASE STUDIES

The code for case studies 1 and 2 is stored at http://server3.walterwaddenmonitor.org/SEM_case_studies.zip

The archive contains R-code together with input and output data. It contains the code and data used to generate the results in this report and is meant to demonstrate how simulation-based evaluation of monitoring (SEM) can be conducted.

The code consists of a main script (separate for case 1 and 2: case1_main.r and case2_main.r) which calls the various sub-scripts that correspond to the different steps of the SEM workflow (see Figure 1). The main-scripts contain an explanation of the in- and outputs by each sub-script. The various parameter-choices in the analysis are coded explicitly in the scripts (and not put externally in a parameter file). In addition a file with helper-functions is included (case1_functions.r).