Centre for Geo-Information

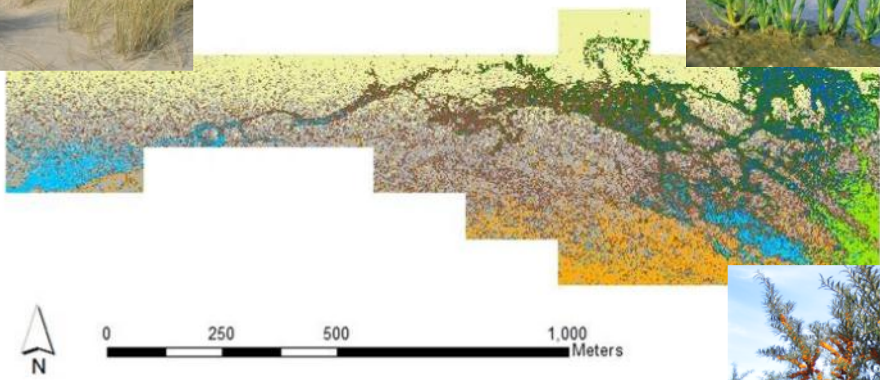
Thesis Report GIRS-2012-08

Uncertainty about mapped vegetation changes

A case study on the eastern part of Ameland

Anne-Ruth Sneep

May 2012



WAGENINGEN UNIVERSITY

WAGENINGEN UR

Uncertainty about mapped vegetation changes
A case study on the eastern part of Ameland

Anne-Ruth Sneep

Registration number 880420780071

Supervisors:

Gerard Heuvelink
Sytze de Bruin

A thesis submitted in partial fulfilment of the degree of Master of Science
at Wageningen University and Research Centre,
The Netherlands.

May 2012
Wageningen, The Netherlands

Thesis code number: GRS-80436
Thesis Report: GIRS-2012-08
Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing

Table of contents

List of figures

Photos on front page:

Salicornia europaea: http://www.britannica.com/bps/media-view/8426/1/0/0
Sea-buckthorn: http://www.lauwershof.com/page13.php
Marramdunes:http://nl.m.wikipedia.org/wiki/Bestand:Helmgras_kijkduin_februari_2005.JPG (last accessed 18-04-2012)

List of tables

Summary

Gas extraction below the eastern part of the island of Ameland causes lowering of the land surface, which affects the vegetation composition in two Natura 2000 areas. Vegetation changes are monitored by measuring vegetation types bi-annually at multiple locations in the area. Geostatistical interpolation is used to create vegetation maps from these point observations and spatially exhaustive covariates. However, the spatial interpolation technique that has so far been used does not quantify the uncertainty about the area occupied by the different vegetation types, which prohibits the analysis whether the observed areal changes for each vegetation type are true changes or can be attributed to spatial interpolation error. It is important to assess if the monitoring programme adequately estimates the areas of the vegetation types, as responsible parties base their decisions on the findings of the monitoring program.

The aim of the research presented in this thesis is to express and model the uncertainty about the area of different vegetation types at different times on the eastern part of the island of Ameland. This was achieved by developing a conditional sequential indicator simulation program with GSLIB in Fortran. The stochastic simulation program uses a regression indicator kriging model to define the joint probability for the presence of each vegetation type at every location in the grid. Next it draws multiple realizations from the joint distribution using a pseudo-random number generator. The implementation of the conditional simulation program implemented a technique to decrease computation time by using one random path for all realizations, calculate kriging weights only once and save the result.

For each realization the area per vegetation type was calculated to derive the expected value over all realizations. Box plots were used to express uncertainty about the area of each vegetation type at each of the moments in time. The uncertainty about the areas was low and most of the boxes in the plots did not overlap, so that most mapped vegetation changes can be considered significant. The coefficient of variation (CV) was used to compare the uncertainty between different vegetation types. Overall, the values were below 1%, which indicates a low uncertainty. The 'Gelderland' method was used for computing a trend in nature conservation value; it was found to be positive. The uncertainty about the nature conservation value is very low, since the average difference between the boundaries of the 90% confidence interval is 0.012% of the average nature conservation value.

From this research it is concluded that areas estimated for each of the vegetation types are considered adequate, as their uncertainty is very low. Gas extraction did not influence the vegetation in the study area in a negative way. Instead, considering nature conservation value gas extraction can be continued, since the nature conservation value increased significantly.

1 Introduction

The NAM[1] (Nederlandse Aardolie Maatschappij) has been extracting gas below the eastern part of the island of Ameland since 1985. This gas extraction causes lowering of the land surface and consequently higher water levels on the island (Eysink et al., 2000). In the Natura 2000 areas 'Duinen Ameland' and 'Noordzeekustzone' the vegetation composition may be affected by the changing water regime. The changes in vegetation are monitored by determining vegetation types at different places in the area, over a period of time. The monitoring is done by WL|Delft Hydraulics in cooperation with Alterra. To visualize the presence of different vegetation types in the whole area, vegetation maps are made using the spatial interpolation method regression kriging (e.g. Slim et al., 2011).

When comparing the vegetation maps and areas from different years, it appears that some vegetation types have become more abundant over the past ten years. However, no figures on the uncertainty about vegetation type abundance are currently available. Therefore it is not possible to determine whether observed changes are statistically significant or whether these changes may be attributed to spatial interpolation errors. It is important to assess if the monitoring programme adequately estimates the areas of the vegetation types, as responsible parties base their decisions on the findings of the monitoring program. This research aims at investigating the uncertainty about the changes in area, and providing data and an example for the application of an analysis to assess the significance of the changes.

There are two broadly different methods for this kind of research: design-based and model-based approaches. In the design-based approach, the pattern of values is regarded as fixed and the sampling locations are random, while in the model-based approach, the sampling locations are fixed and the pattern of values is regarded as random. Statistical inference with the design-based method is based upon the sampling design, while in model-based statistical inference is based on a stochastic model of the variation. Thus, a design-based method can use the weighted averages of the data, since the weights of the data are determined by the selection probabilities of the sampling locations. The standard deviation and confidence interval can be used for calculating the uncertainty. In the model-based approach the weights of the data are determined by the covariances between the observations, which are given by the model as a function of the coordinates of the sampling locations. The variable can take several possible values, each with a defined probability of occurring, thus forming a random variable. Utilizing a parameterized model of this random variable and semivariograms to incorporate spatial correlation, simulations can be performed to create other possible realities so as to assess the uncertainty and calculate, for example, the 95th percentile (Brus & De Gruijter, 1997; De Gruijter et al., 2006; Kott, 1991).

In this research a model-based approach is chosen. This choice is motivated by the desire to produce maps. These maps can visualise locations of different vegetation types, and more importantly, variations at these locations can be compared over time. In this way the change of location or abundance of different vegetation types can be assessed. Furthermore, the model-based approach was used in previous studies (Eysink et al., 2000; Slim et al., 2005;

[1] http://www.nam.nl/home/content/nam-en/general/

Slim et al., 2011) on which the current work is based. Also, the sample design used to acquire the data is not only probability sampling, which desired by the design-based method. Examples of other studies using the design-based approach are Brus (2000), Ter Braak et al. (2008) and Brus & De Gruijter (2011).

The research objective is to develop and apply a method to express and model the uncertainty about the area of different vegetation types (at different moments time) on the eastern part of the island of Ameland. The objective will be fulfilled with the use of stochastic simulation to investigate the uncertainty (Atkinson, 1999; De Bruin, 2000; Gómez-Hernández & Srivastava, 1990; Goovaerts, 1997; Goovaerts, 2001; Journel, 1996; Magnussen & De Bruin, 2003).

The research questions are:

1. How can spatial stochastic simulation be used for modelling the uncertainty about the area of each of the vegetation types in the study area for all measurement years?
2. How can spatial stochastic simulation be implemented, and which software environment is suitable?
3. What is the size and uncertainty about the area of the different vegetation types according to the model developed in this research?
4. What is the influence of uncertainty about the area of the different vegetation types on the nature conservation value in the study area?

The report is structured as follows. First, the ecological aspects of the research are discussed in Chapter 2. Here the motivation for vegetation monitoring is described, as well as the vegetation types that occur in the study area. Chapter 3 explains the statistical background: the sample design and methodology used in previous research. Furthermore, an introduction to spatial stochastic simulation is given and the methods used for analysis of the output of the simulation. The stochastic simulation program developed and applied in this research is explained in detail in Chapter 4. The analysis of the output of the program can be found in Chapter 5. First the output is verified, and trends are discussed. After this, the uncertainty is visualised and described. Finally, an example of significance testing is provided. In the last chapter, the methodology is discussed, the research questions are answered and recommendations are made for future research.

2 Study area and ecological background

This chapter describes the motivations for starting a vegetation monitoring project, which concerns the conservation of various vegetation types that are potentially influenced by gas extraction. The contours of the study area are shown in Chapter 2.1, and the vegetation types distinguished in the monitoring programme are described in Section 2.2.

2.1 Gas extraction and monitoring

By extracting hydrocarbons from gas-carrying strata, pressure in the pores of bedrocks is lowering, which causes a more compact structure of these strata. This structure manifests itself by a lowering of the surface (Ketelaar et al., 2011). The maximum expected subsidence is expected to be around 34 to 38 centimetres after finalization of gas extraction in 2025-2050. Figure 2.1 shows the subsidence up to 2009.

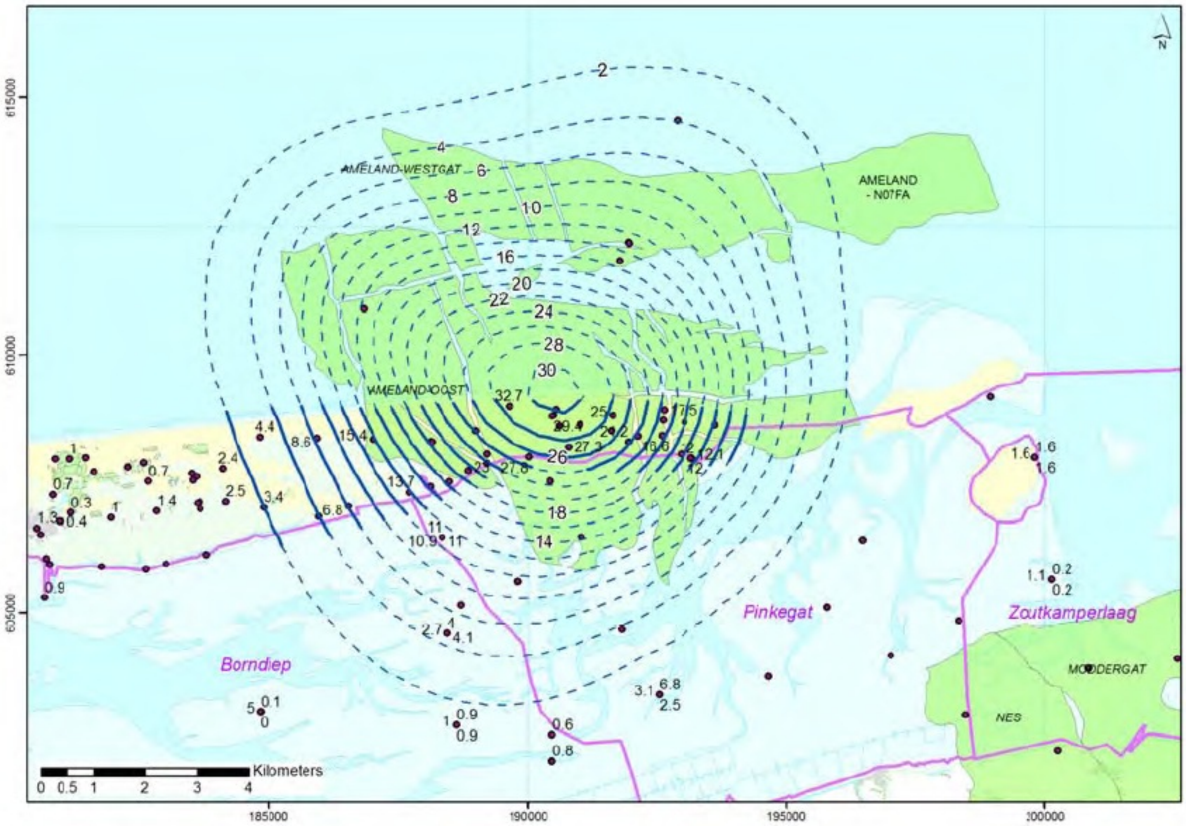


Figure 2.1: *Subsidence (in cm) in the period 1986-2009: figures represent the measured movements (benchmarks on the island and clusters in the Wadden Sea measured from 1986). Contour lines depict the subsidence prognosis. The locations of the gas fields are shaded in green (Ketelaar et al., 2011).*

In 1986, after gas extraction had started, several predictions were made about possible changes in vegetation. Examples are salt marshes returning to earlier succession states, wetter dune valleys, and because of more inundated flats: fewer birds. Negative scenarios like these were the driving force for nature conservation organization 'It Fryske Gea' to ask the NAM for a monitoring programme. A general first monitoring plan was designed to monitor the area from 1988 to 2000. In 1995 a first evaluation report was published (Eysink et al., 1995), followed by a second evaluation after 13 years of gas extraction in 2000 (Eysink et al., 2000).

In the first monitoring period, some unexpected changes occurred: the death of several areas with bushes Sea-buckthorn (*Hippophae rhamnoides*) (see Figure 2.2), Elderberry (*Sambucus nigra*) and Hawthorn (*Crataegus monogyna*). After investigation, it was derived that high groundwater levels (because of heavy rain) and inundation by the sea (because of high tide) were the main reasons for the dying shrubs. It was concluded that the dynamics of the natural environment had a bigger influence on the vegetation than the consequences of gas extraction in the area (Krol, 2004; Slim, 1997a; 1997b; Eysink et al., 2000). Later on, however, Krol suggested that not only rain and high tide caused the death of the shrubs. Long inundation of these bushes could also be induced by the subsidence. These uncertainties led to a more intensive monitoring of vegetation in dune valleys, from the year 2000 onwards (Krol, 2011).

Figure 2.2: *Dead Sea-buckthorns, in an inundated valley (Eysink et al., 2000).*

The area with dying Sea-buckthorn had an extent of approximately 25 hectares. The study area chosen for the monitoring is 70 hectares. This area contains the location with the dead Sea-buckthorn, but has a much bigger extent since it is important to also comprise parts where different or no ecological effects of subsidence are expected (Slim et al., 2011). The contours of the study area are presented in Figure 2.3.

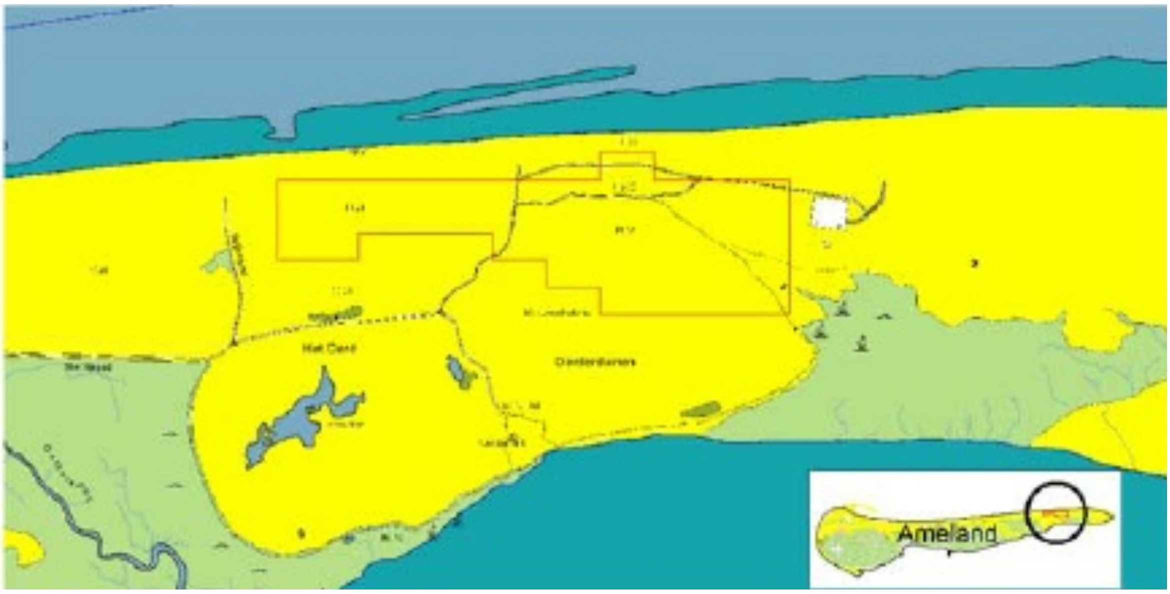


Figure 2.3: *Eastern part of Ameland. The study area is outlined in red (e.g.: Slim et al., 2011).*

2.2 Habitats and vegetation types

The study area is part of two Natura 2000 areas: 'Duinen Ameland' and 'Noordzeekustzone'. Dune valleys located in the western part of the study area belong to the first, and the eastern part of the study area belongs to the latter. Both areas are part of an ecological network of special protected areas all over the European Union, with the purpose to ensure biodiversity by conserving natural habitats and wild fauna and flora[2]. Several habitat types described in Annex I of the EG-habitat Directive (92/43/EEG) are present in the study area, see Table 2.1. In this table also a local typology is described, which corresponds well to European and national classification systems (Slim et al., 2011).

[2] http://europa.eu/legislation_summaries/environment/nature_and_biodiversity/l28076_en.htm (last accessed 12-04-2012)

Table 2.1: *Overview of the local vegetation classification system and the corresponding habitat types in the European classification system for Natura 2000 habitats (Slim et al., 2011).*

Type	Local typology	Habitat types Natura 2000
1	Seaside Marram dunes	H2120 Walking dunes on shoreline with *Ammophila arenaria* ('white dunes') H2170 Dunes with *Salix repens subsp. argentea* (Salicion arenariae)
2	Fixed grassy dunes	H2130 Fixed coastal dunes with herbaceous vegetation ('grey dunes')
3	Overgrown and thicketed dune slacks	H2190 Wet dune valleys
4	Sea-buckthorn thickets	H2160 Dunes with *Hippophae rhamnoides*
5	High salt marsh, more saline	H1330 Atlantic marshes (Glauco-Puccinellietalia maritimae)
6	High salt marsh, less saline	
7	Low salt marsh, less saline	
8	Low salt marsh, more saline	

The local typology was developed by Slim et al. (2011) using automatic classification of datasets of all years (2001, 2004, 2006, 2008 and 2010), which leads to a typology that is valid for all years.

The vegetation types can be roughly distinguished into two groups. The first four vegetation types are part of the dune landscape, while the latter are different types of salt marsh vegetation. In salt marsh vegetation, inundation with salt water plays an important role. The differences between the vegetation types are shortly discussed below.

- **Type 1: Seaside Marram dunes** are dunes on the North Sea side of the study area. This is an open dune landscape with some small pioneer species that can survive under the extreme circumstances of this habitat. Right behind the beach ridge, a small part of the area consists of habitat type H2170 (see Table 2.1) with Creeping Willow (*Salix repens)*
- **Type 2: Fixed grassy dunes** consist for the largest part of fixed dunes with grassy vegetation. Due to the activity of rabbits and gulls on a part of this habitat some other plants, like the Common Nettle and the Small Nettle (*Urtica dioica* and *U. urens*), also occur.
- **Type 3: Overgrown and thicketed dune slacks** consist of dune valleys, which are relatively wet due to the influence of inundation by the sea. The vegetation is bushy and contains some shrubs.
- **Type 4: Sea-buckthorn thickets** are situated higher and drier than type 3, and are not influenced by inundation. The main species is Sea-buckthorn, but also other species of bushes occur like Dewberry and Bramble (*Rubus caesius* and *R. fruticosus*)
- **Type 5: High salt marsh, more saline** contains high salt marsh, similar to type 6. On high salt marsh it is relatively dry, compared to the low salt marsh (type 7 and type 8). Type 5 has more halophyte vegetation compared to type 6, because of higher salt levels.
- **Type 6: High salt marsh, less saline** contains high salt marsh, but these locations have a lower salt level compared to type 5, and therefore less halophytes.
- **Type 7: Low salt marsh, less saline** is low salt marsh, where it is relatively more wet than on the high salt marsh, but less salt than in areas with type 8. This results in less

halophyte species, like Common Reed (*Phragmites australis*), that like wet circumstances.

- **Type 8: Low salt marsh, more saline** is low salt marsh, similar to type 7. Because of higher salt concentrations, more halophyte species, like Common Glasswort (*Salicornia europaea*) occur.

Within the 'Overgrown and thicketed dune slacks' (type 3) a very special plant species is present. In the Netherlands it is one of the three endangered species, mentioned by the EG-Habitat Directive in Annex II and Annex IV. The Fen orchid (*Liparis loeselii,* Figure 2.4) has a stable yearly population in the study area of a couple of hundred plants, slightly spreading to the east and west. In the study area, the very wet, young and calcareous dune valleys are habitat of this species. A limited amount of organic carbon and nutrients are available. The influence of the sea causes a buffered acidity, which also occurs in quaking bogs (also a habitat of the fen orchid). The fen orchid and its habitat are protected and may not be negatively influenced by humans, for example by gas extraction (Slim et al., 2011).

Figure 2.4: *Fen Orchid (Liparis loeselii) in the research area (Slim et al., 2011).*

3 Statistical methods

This chapter gives an overview of the statistical methods used in this research. First, the methods used in previous research that formed the basis of the current work are explained, starting with the sample design. In Paragraph 3.2 the basics of stochastic simulation are explained. Finally, the methods used to analyze the output of the stochastic simulation algorithm are explained.

3.1 Previous research

As explained in Chapter 2.1, this research is part of a monitoring programme with the objective to keep track of the influences of gas extraction. In 2001, the part of the monitoring programme that is used in this research started. This section describes the methods used in this monitoring programme.

3.1.1 Sample design

Measurement locations were selected using a statistical program. To derive a systematic unaligned sample, a grid was projected on the study area. The grid cell in the upper left corner was assigned a random X- and Y-coordinate. For every cell in the upper row, the same local X-coordinate of this cell was used and a new Y-coordinate was picked. The grid cell in the second row was assigned randomly a new X-coordinate, but kept the Y-coordinate. Just like the grid cells on the upper row, the cells below kept the Y-coordinate of the cell above, and got a new X- coordinate. This was repeated until all grid cells had coordinates. Each grid cell has an area of 1 hectare. With this methodology 70 measurement locations were picked in 2001. For 2004, 2006, 2008 and 2010 the same 70 locations were used, as well as 70 extra locations. The extra locations were picked using 35 strata, which consist of two adjacent cells of the projected grid. From these strata randomly 2 measurement locations were picked per stratum. With this methodology, variation over different distances could be explored, which is important when using geostatistical interpolation. The locations for 2006 were, by exception, the same as those of 2004. Figure 3.1 shows the measurement locations in 2010 and the vegetation types found in these locations (Slim et al., 2011).

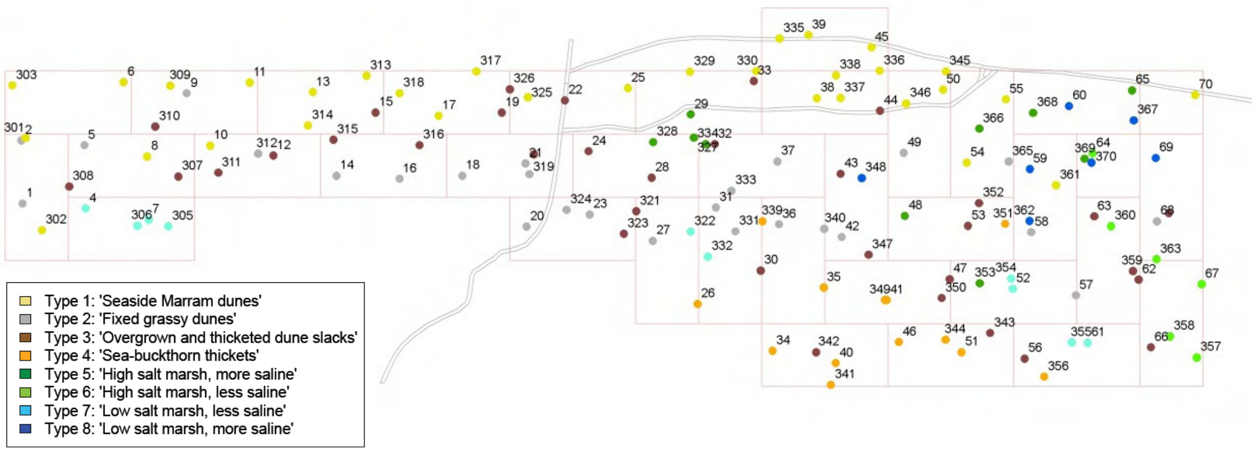


Figure 3.1: *Measurement locations and vegetation types on the monitoring site in 2010 (Slim et al., 2011).*

3.1.2 Regression kriging

Vreugdenhil (2011) described the geostatistical analysis of the vegetation data. Output of this analysis were regression maps and residual values, used as input for this research. The methodology for creating these data is described below.

The used analysis technique is called regression kriging, which is extensively discussed in Hengl et al. (2007). In regression kriging a regression model is used to predict the presence of a vegetation type. This predicted probability is corrected with residuals that are interpolated using kriging. In this study, the presence of every single vegetation type is a categorical variable with values 1 and 0, for presence or absence respectively. Hence, kriging for categorical variables was used, which is called indicator kriging (Bierkens & Burrough, 1993a; b). The model to express the presence of a vegetation type is:

$$Y_i(x) = \pi_i(x) + \varepsilon_i(x)$$

Here Y_i is an indicator that is 1 when a vegetation type is present and 0 when it is not. Pi is a local prediction of the probability of presence of vegetation type i, (i = 1,…,8) that was obtained by multinomial logistic regression, described below. The difference between the measured occurrence of vegetation types and their probabilities predicted by regression are $\varepsilon_i(x)$, the regression residuals. Since residuals are differences between the vegetation type predicted with regression and the type observed in the field, true residuals are only known on measurement locations. Elsewhere, residuals can be represented by a random variable (RV) in a geostatistical model. Below, regression kriging is explained below in two steps: regression and kriging.

Step 1: Regression

A multinomial logistic regression model (Kempen et al., 2009; Slim et al., 2011) was used to predict presence of vegetation types for every location in the study area. The formula used to calculate the probability of occurrence of a vegetation type for one grid cell is:

$$\log\left(\frac{\pi_i(x)}{1-\pi_i(x)}\right) = \beta_{0i} + \beta_{1i}h_1(x) + \beta_{2i}h_2(x) + \ldots + \beta_{ji}h_j(x)$$

In this formula, $h_j(x)$ (with j = 1,…6) are the explanatory variables on location x, β_{1i} … β_{6i} the regression coefficients, and β_{0i} the intercept. The regression coefficients and intercept can be found in Table 3.1. The explanatory variables used by Vreugdenhil (2011) are:

1. x coordinate
2. y coordinate
3. relative ground height in a circle with radius of 25 meters, to assess if a point is located in a valley or on the top of a dune
4. whether area is flooded with water levels of 1.9 meters above Amsterdam Ordnance Datum
5. slope
6. number of days per year the location is flooded

__Table 3.1:__ The coefficients used in the regression kriging by Vreugdenhil (2011).

Vegetation type	Regression coefficients						
	0	1	2	3	4	5	6
Type 1	6.09	-3.65	3.66	2.68	-2.37	4.00	-2.06
Type 2	6.48	-4.11	-0.48	2.84	-5.20	3.92	-2.13
Type 3	8.98	-3.88	-0.69	2.46	-0.98	3.44	-1.17
Type 4	3.13	-4.20	-3.15	2.36	-4.63	4.63	-1.86
Type 5	4.47	-1.75	0.41	0.80	0.37	3.63	-0.44
Type 6	-28.34	14.71	-2.06	-2.61	2.43	-4.81	0.23
Type 7	-0.73	-7.30	-4.63	2.62	0.45	-3.38	-0.15
Type 8	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Result of this regression is a series of maps showing predicted probabilities for every year and every vegetation type covering the whole study area.

Step 2: Kriging

In this step the geostatistical interpolation technique kriging is used to interpolate the residuals to correct the regression maps calculated in Step 1. Kriging predicts the residual on a certain location as a weighted linear combination of known values of the residual at observation points:

$$\hat{\varepsilon}_i(x_0) = \lambda_1 \varepsilon_i(x_1) + \lambda_2 \varepsilon_i(x_2) + \ldots + \lambda_n \varepsilon_i(x_n)$$

In this model $\hat{\varepsilon}_i$ is the predicted residual at location x_0, ε_i are the measured residuals at location x_1 to x_n, with n measurements involved in the interpolation. All measurement values have a kriging weight λ_i that determines the contribution of the residual at this measurement point to the interpolation. The weights are chosen in a way that the estimation error $\varepsilon_i(x_0) - \hat{\varepsilon}_i(x_0)$ has a mean of zero and a variance as small as possible. To calculate the kriging weights a semivariogram is used, which represents the spatial correlation structure of the residual ε_i.

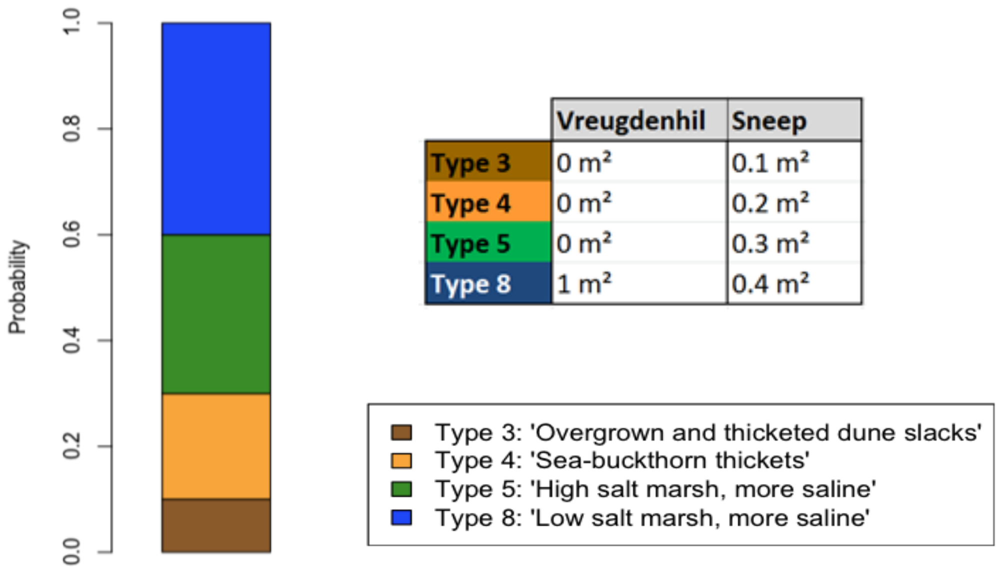
In this research the semivariograms of Vreugdenhil (2011) were used, which are presented in Table 3.2. The parameters are identical for all observation years. The residual of vegetation type 6 has a pure nugget semivariogram, which indicates no spatial correlation is present. For the other semivariograms the spatial correlation is present in the range of the model. For vegetation type 1 (Seaside Marram dunes), type 2 (Fixed grassy dunes) and type 7 (Low salt marsh, less saline) the ranges are quite small: 26.24759, 17.15274 and 24 respectively. Conversely, for type 4 (Sea-buckthorn thickets) the range is large: 514.981 (Vreugdenhil, 2011). A small range in semivariograms results in less influence of the measurement points on the distribution of the residual.

Table 3.2: *The semivariogram parameters used by Vreugdenhil (2011).*

Vegetation type	Semivariograms			
	Nugget	pSill	Sill	Range
Type 1	0.03698816	0.02456935	0.06155751	26.24759
Type 2	0.05462978	0.06207231	0.11670209	17.15274
Type 3	0.07879114	0.07783798	0.15662912	69.86289
Type 4	0.02827878	0.01864838	0.04692716	514.981
Type 5	0.02713797	0.02059868	0.04773665	91.32331
Type 6	0.008182	0	0	0
Type 7	0.014969311	0.006527021	0.021496332	24
Type 8	0.01993458	0.01650136	0.03643594	40.59852

3.1.3 Dominant vegetation maps

In previous research the method regression kriging was used to make dominant vegetation maps. This was done by picking the vegetation type with the highest probability for every location. These dominant vegetation maps were used by Vreugdenhil (2011) to calculate the size of the area of each vegetation type. The number of pixels of a vegetation type in the dominant vegetation map was summed and the area was derived by multiplying with the cell size. This methodology uses only the highest probabilities to calculate the area. A different method is to use the expected area, which is calculated by multiplying the cell size with the probability of a vegetation type on a certain location, and doing this for all locations and types. This results in a more accurate area calculation, which is used in this research. Figure 3.2 gives an example of the area calculation with the two methods.



	Vreugdenhil	Sneep
Type 3	0 m²	0.1 m²
Type 4	0 m²	0.2 m²
Type 5	0 m²	0.3 m²
Type 8	1 m²	0.4 m²

Figure 3.2: *Example showing the probabilities of the different vegetation types in one cell. The dominant vegetation type in this cell would be type 8. To calculate the area, with a cell size of 1m², Vreugdenhil (2011) would count 1 m² of Type 8. However, the expected value is 0.1m² of type 3, 0.2 m² of type 4, 0.3 m² of type 5 and 0.4 m² of type 8.*

3.2 Stochastic simulation

Stochastic simulation is used in this research to calculate the uncertainty of the area of the various vegetation types. This method creates multiple equiprobable realizations of the joint distribution of attribute values in space (Gómez-Hernández & Srivastava, 1990; Journel, 1996; Deutsch & Journel, 1998; Goovaerts, 1997; 1999). When measurement values are known on certain locations a simulation is conditional, since simulated values are conditioned to these values. A simulation is sequential when one grid cell after the other is simulated and previously simulated grid cells influence the value of the next simulated cell. This happens when a value at a location is simulated, this value is added to the conditioning data, hence more conditioning data is created as the simulation proceeds. In this way the final grid node to be simulated, will have the spatial structure desired and honour all initial data. With the use of sequential simulation spatial correlation is incorporated in the simulation algorithm (Gómez-Hernández & Srivastava, 1990; Oliver, 1996; Deutsch & Journel, 1998; Yao, 1998).

Unlike kriging, spatial stochastic simulation does not aim at minimizing a local error variance. Instead it focuses on the reproduction of statistics, such as the sample histogram or the semivariogram model, in addition to taking data values into account. The pattern of the simulated map looks more realistic than the map of statistically best estimates, because it reproduces the spatial variability modelled from the sample information (Goovaerts, 1999).

Simulation of multiple realizations of a categorical variable can be performed using stochastic indicator simulation. This procedure is used and reported by many studies (e.g. Gómez-Hernández & Srivastava, 1990; Goovaerts, 1996; 1997; 2001; Deutsch & Journel, 1998; Bierkens & Burrough, 1993a; b; Atkinson, 1999; Gotway & Rutherford, 1993; De Bruin, 2000).

The basic conditional sequential indicator simulation algorithm performs the following steps:

1. Define a random path through all grid nodes visiting each node only once;
2. Calculate the probability distribution of the target variable at the first location that is visited, using the conditioning data. In this case regression indicator kriging was used to calculate the probability of each vegetation type at this location, as described in Section 3.1.2. These values together are rescaled in the interval [0;1] to form a distribution representing the probabilities of each vegetation type on this location;
3. Draw a value from the distribution created in Step 2. First take a random number between 0 and 1. Then, pick from the probability distribution the vegetation type that corresponds to this random number. This vegetation type is chosen for this location;
4. To incorporate spatial correlation in a simulation, the value simulated in Step 3 is added to the conditional data. New simulated values are then conditioned both on the original survey data as well as previously simulated values;
5. Move to another node along the random path and repeat Steps 2, 3 and 4 until all nodes have been given a simulated value;
6. When all nodes have a simulated value, one realization is made.

Most of the times, more than one simulation is desired to be able to compare the realizations and assess uncertainty. When particular spatial features are present on the same location in (almost) all realizations, these patterns can be considered certain. On the other hand, features are unlikely when only seen in a few simulations (De Bruin, 2000).

3.3 Analysis of realizations

Averaged over many realizations, areas derived by conditional simulation should correspond to expected values from regression kriging. To verify this, dominant vegetation maps made with realizations of this research and dominant vegetation maps made with regression kriging by Vreugdenhil (2011) were compared. The method used to make the dominant vegetation maps is described in Section 3.1.3.

The spread and uncertainty in area of the different vegetation types obtained from the series of realizations were visualised using box plots. Input for the box plots were the realizations, which had all different areas for each vegetation type. A box plot shows interquartile distances: the bottom and top of the box represent the first quartile (Q1) and third quartile (Q3). The interquartile range (IQR) is the distance between the first and third quartile. The second quartile (Q2) is also called the median, which is represented by the line close to the middle of the box, 50% of all values are found within the box. In the graphs in Chapter 5, the whiskers of the box plot reach out to the lower and upper inner fence. The lower inner fence is defined by the maximum of the smallest data value and Q1 – 1.5(IQR). The upper inner fence is defined by the minimum of the largest data value and Q3 + 1.5(IQR). Any data value beyond the whiskers is considered an outlier, which are visualised by open circles. The smallest and largest data values that are not an outlier are represented by the whisker and are called the lower and upper adjacent value (e.g. Ott & Longnecker, 2010, p.100). The lower and upper adjacent values can be used to determine if the area of a vegetation type has changed more or less significantly over time. When the whiskers of two different observation years do not overlap, there is likely a significant change. This analysis method can only be used for the uncertainty about one vegetation type in time. The assessment of significance using adjacent values gives only an indication of significance. To be sure a change is significant a statistical test needs to be done.

To compare the variability in area between the different vegetation types the coefficient of variation (CV) was used. This coefficient measures the variability in area relative to the magnitude of the mean (e.g. Ott & Longnecker, 2010, p. 96). In this research the CV is expressed as a percentage:

$$CV = (\text{standard deviation} / \text{mean}) * 100$$

Apart from changes in areas of different vegetation types, also the consequences of these changes for the nature conservation value of the study area are assessed. The most used indicators for estimating conservation value are rarity and irreplaceability (Hertog & Rijken, 1996). Eysink et al. (2000) and Vreugdenhil (2011) used the ‘Gelderland’ method to assess the conservation value (NBW) of vegetation in the study area. This is a standardized method to assess the rarity, decline and international threat of the species occurring in a vegetation survey (e.g. Eysink et al., 2011). Appendix 3 in Sanders et al. (2004) describes the application of this method more extensively. Every vegetation type has a different score for NBW, which are given in Table 3.3 (Vreugdenhil, 2011).

Table 3.3: *The nature conservation value for all vegetation types in the study area (Vreugdenhil, 2011).*

	NBW
Type 1	12.649
Type 2	11.718
Type 3	11.038
Type 4	9.318
Type 5	14.538
Type 6	13.678
Type 7	12.571
Type 8	13.608

- Type 1: 'Seaside Marram dunes'
- Type 2: 'Fixed grassy dunes'
- Type 3: 'Overgrown and thicketed dune slacks'
- Type 4: 'Sea-buckthorn thickets'
- Type 5: 'High salt marsh, more saline'
- Type 6: 'High salt marsh, less saline'
- Type 7: 'Low salt marsh, less saline'
- Type 8: 'Low salt marsh, more saline'

The total NBW value for the study area was calculated by multiplying the area of a vegetation type by the NBW value for this vegetation type (Table 3.3) and divide by the total area.

To test significance of change between two observation years, areas of a chosen vegetation type were calculated for every realization of two chosen measurement years. The differences in area between all possible combinations of these two realizations were calculated (with 500 realizations per measurement year resulting in 250,000 differences). From the result a histogram was made that visualizes the spread. Using the statistical program R, percentiles of relevant differences were calculated. Also the probability of a change of a given amount of squared meters was calculated. An example of this test for significance is given in Section 5.5.

4 Implementation of the simulation program

This research uses the geostatistical package GSLIB to implement the conditional simulation program described in Chapter 3. In this chapter the choice for this software is motivated and the conditional simulation program developed during this research is described.

4.1 Software

In accordance with Goovaerts (2010), the choice for a geostatistical software package was based on three considerations, which are discussed in this section: the type of analysis, licences and availability, and the expertise of the user.

The analyses that needed to be performed with the software were kriging and conditional simulation. Not all packages provided these functionalities. In Goovaerts (2010) and Hengl et al. (2007) different software solutions are compared, and an overview of different aspects, like functionality, is given.

Taking into account restrictions like expensive licences and availability of software, GSTAT in R (Pebesma & Wessling 1998; Pebesma 1998) and GSLIB in FORTRAN (Deutsch & Journel, 1998) were both suitable software packages.

The choice between GSLIB and GSTAT was based on the last consideration: the expertise of the user. For both programs modifications to the basic algorithm were considered necessary:

- Combining regression kriging with indicator simulation required modifications to the source code in GSTAT. With regression kriging the regression values (the external drift of the kriging) need to be summed to the interpolated residual for every location in the grid, as described in Section 3.1.2. After, for the simulation, the corrected probabilities from the regression kriging, that form a distribution on this location, need to be rescaled to fit in the interval [0;1], before drawing from this distribution, and herewith realizing a vegetation type for this location. In GSTAT it is not possible to sum this external drift to the residual before a value is drawn from the distribution, which is required in this research.
- In GSLIB kriging weights were not saved, but calculated again for every grid cell, to derive the probabilities of all vegetation types on all cell locations. This calculation is time consuming, hence calculation time could be reduced when kriging weights were saved. Saving these weights results in the use of one random path (since in this case all neighbourhoods are the same) for all simulations.

GSLIB in Fortran was chosen, since a program in Fortran can be relatively straightforward modified and adapted (Personal communication: De Bruin, 2011). Modifying the necessary functions in GSTAT required advanced programming skills in the C language. Moreover, GSLIB has ready-to-use subroutines available that can be called from the main program.

4.2 Conditional sequential indicator simulation

Standard programs are available in GSLIB, which can be compiled and used directly. To use these programs, only the input files and parameters need to be customized by the user. The parameters used in a GSLIB program are stored in a .PAR file. In Appendix I the .PAR file used for the conditional simulation of the vegetation in 2001 is given. This file is read when the program is run. Among other parameters, the semivariograms of Vreugdenhil (2011) are stored in this file, as well as the filenames of the input data.

In this section the conditional simulation program is described, starting with program input. Next, the used algorithms and subroutines are described in the program description. Finally, the output of the program is discussed. The version of GSLIB that is modified in this research is owned by The Board of Trustees of the Leland Stanford Junior University, Copyright 1996. Everyone is granted permission to copy, modify and redistribute the programs in GSLIB, but only under the condition of notice of the copyright.

4.2.1 Input

There are two data flows into the model: regression maps and residuals. The first contain the output of the regression model described in Chapter 3.1.2, and have a resolution of 1 m^2. For every vegetation type, there is a regression map with the probability of occurrence on every grid cell in the study area. Since there are eight vegetation types classified, there are eight regression maps. For every observation year different parameters are input to the regression model. This results in different regression maps for all measurement years. The regression maps are .asc files.

The residuals are point data, since these files contain the residuals of each vegetation type at all measurement locations in one year (for example: Figure 3.1 contains the points for 2010). Each file contains three columns: X-coordinates, Y-coordinates and the residual value in the third column. These data values are the difference between the recorded vegetation type in the field and the vegetation type calculated with the regression model. The input regression maps and residuals were produced by Vreugdenhil (2011).

4.2.2 Structure of the program

The structure of the conditional simulation program is given in Figure 4.1. The boxes with calculation steps have different shadow colours: green and blue. The blue boxes indicate subroutines or pieces of code that are part of already existing GSLIB programs. The green boxes represent algorithms that were developed or modified during this research. The boxes without shadow are incorporated to clarify loop structures in the program. In this paragraph the structure of the program is explained with reference to the numbers next to the boxes in Figure 4.1.

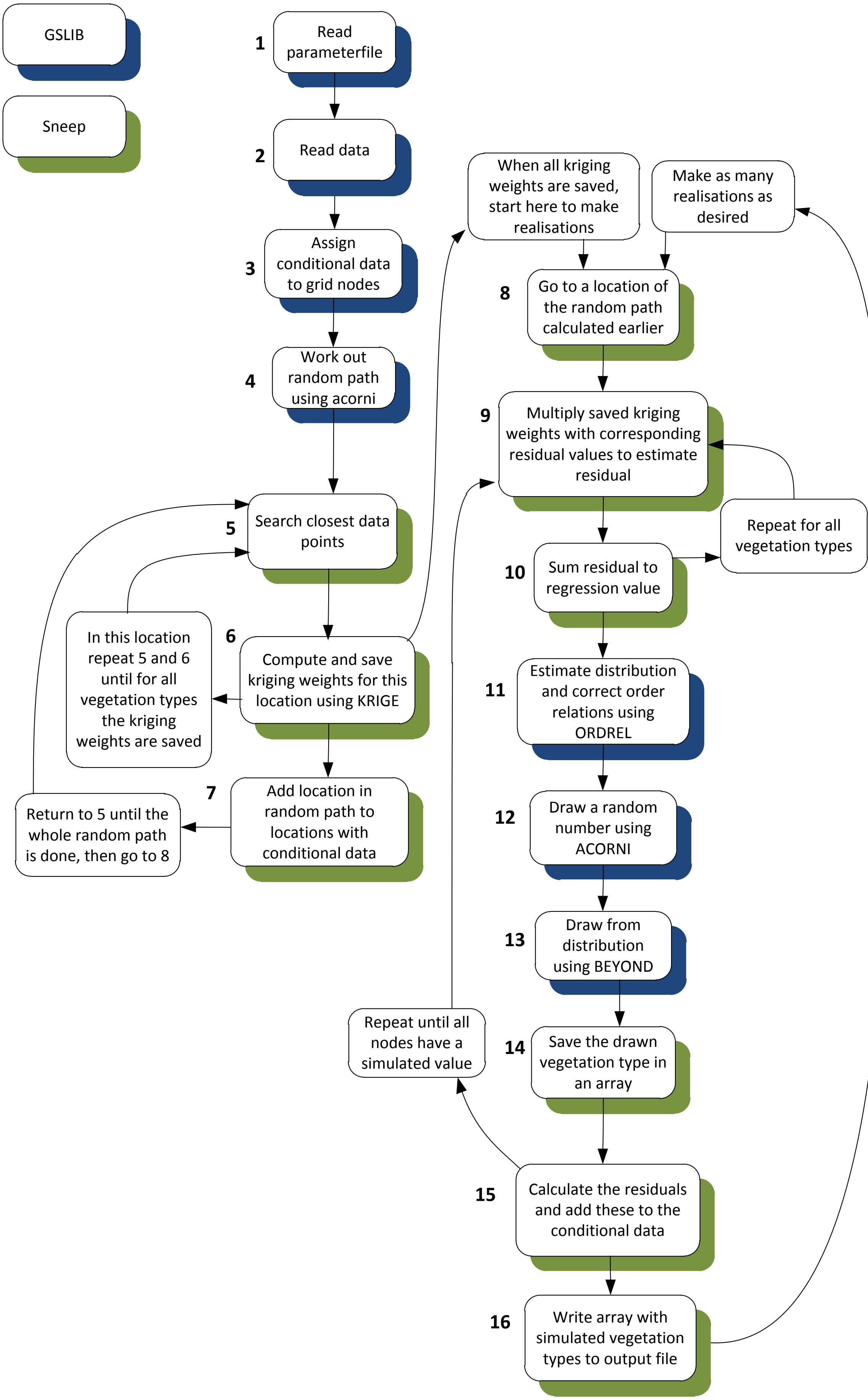


Figure 4.1: *Schematic overview of the conditional simulation program.*

1. **Read parameter file**
   The .PAR file is read, which contains the names of data files, the grid sizes, and semivariograms (see Appendix I for an example of a .PAR file of 2001). The use of a .PAR file is a generic part of GSLIB, and is used in almost every GSLIB program.

2. **Read data**
   The files specified in the .PAR file are opened and read. The .asc files with regression probabilities are read in rows (Y axis of the grid) and columns (X axis of the grid). The data values are stored in a matrix and flipped upside down. This is necessary because the .asc file is read starting from the top of the file, while in the grid definition of GSLIB the Y-axis starts on the bottom of the map (see Figure 4.2). Therefore the matrix is filled starting with the highest Y index. A file with residuals consists of three columns. For each column a matrix is used to store the information. The dimensions of the matrix depend on the number of vegetation types and the number of residual values.

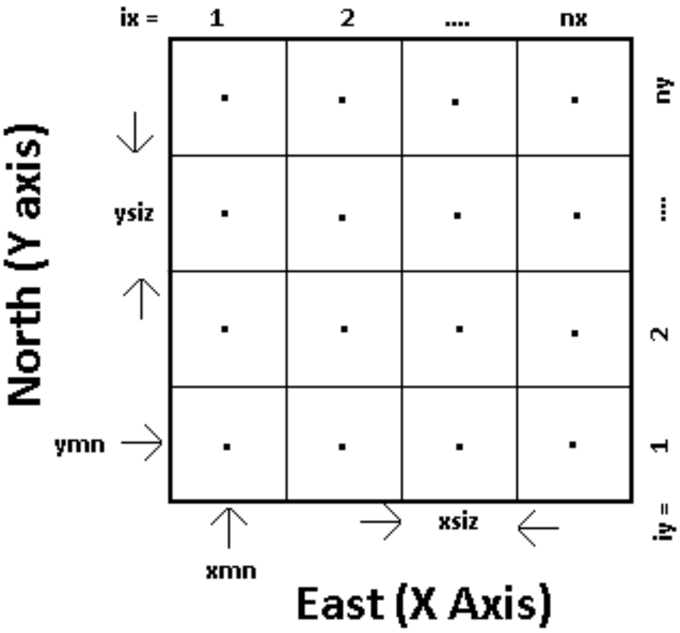


Figure 4.2: *Grid definition in GSLIB (Deutsch & Journel, 1998 (p. 22)).*

3. **Assign conditional data to grid nodes**
   To ensure that every residual read from a file is at the same location as a grid node with regression data, each data point with residuals is assigned to a grid node. According to the grid definition of GSLIB (Figure 4.2), the grid nodes are situated in the middle of the cells. As a result of the assignment of the data values to a grid node, the measurement points need to be slightly shifted to overlap with the regression data points. When all data values are situated on a grid node, every grid node is assigned a unique number that is a combination of the X-, Y- and Z-coordinates. The program written in this research works with a 2D grid, so no Z- coordinate is specified and consequently, $iz = 1$

in the formula used to calculate unique numbers for all grid nodes (Deutsch & Journel, 1998, p. 23):

$$loc = (iz - 1) * nx * \big(ny + (iy - 1)\big) * (nx + ix)$$

In the case of a two dimensional grid of 1900x600 the unique numbers run from 1 to 1,140,000 (1900*600). The assignment of a unique number secures easy overview and the possibility of having one-dimensional arrays corresponding to grid locations.

4. **Work out a random path using ACORNI**
The random number generator in GSLIB is part of the subroutine ACORNI. This subroutine is used to generate a random path. Since the grid nodes have a unique number these can be ordered in a random order. This is done by generating as many random numbers in the interval [0;1] as there are grid nodes. The grid nodes are then sorted in the same order as the random numbers, using the subroutine SORTEM.

5. **Search closest data points**
To calculate the residual on a certain location the KRIGE algorithm is used. To do the interpolation, first cells with a known value have to be found in the neighbourhood. All values outside the study area have been assigned the value -9999, all observed points and simulated nodes value 0, and all other locations value 100. Therefore, only cells with the value 100 need to be evaluated. Neighbours that can be used for the interpolation have the value 0. An array is used for the residual points, the locations with value 0 (which are visited yet, and will be modified in Step 15 when simulation starts), and the locations which are not visited yet that have value 100. The neighbours are found using a square search with a 'max radius' in cells, to be specified in the .PAR file. This search algorithm was modified from SRCHND in GLSIB by De Bruin. Only the cells in the border of the square are evaluated to find a location of a residual value (value 0). The locations of the grid nodes found are temporarily stored in an array to be used in the kriging algorithm.

6. **Compute and save kriging weights for this location using KRIGE**
In subroutine KRIGE the covariance between the points found in the previous step are calculated, as well as the covariance between these points and the location to be estimated. For this calculation the subroutine COVA3 is used. A minor adjustment was made in this subroutine, to avoid the use of a rotation matrix (which is of importance when dealing with anisotropy, but this is not the case in this research). To solve the equation used for kriging (Step 2 of Section 3.1.2) subroutine KSOL is called. The kriging weights, which are output of KSOL, are saved in a matrix. These weights will later on be used in the simulation program.

7. **Add location in random path to locations with conditional data**
After the neighbours of the first location in the random path are found, this location in the random path is added to the group of locations with known residuals, which have value 0. This makes it possible to find this node in the search algorithm when looking for neighbours for the next locations of the random path.

8. **Go to a location of the random path calculated earlier**
   When this step is reached, all nodes in the grid have been assigned kriging weights or are observation points. Consequently, simulations can be made. The random path that is used to loop over all grid nodes to save kriging weights is also used for all simulations. In this step a node in the random path is located to be used in the next steps. Each node will be visited sequentially. When a node has value 0 in the array discussed in Step 5, the next node of the random path will be located.

9. **Multiply saved kriging weights with corresponding residual values to estimate residual**
   For the every location of the random path kriging weights have been saved in Step 6. These weights are now used to calculate the residual of the located node in Step 8. This is done by multiplying the kriging weights with the residual values (found in Step 5, or calculated in Step 15), which are neighbours of this location in the random path.

10. **Sum residual to regression value**
    For the location in the random path found in Step 8, the corresponding value in the regression map is found. The regression value and residual (calculated in the previous step) are summed. By doing this the regression value is corrected using the interpolated residual, as was described in Section 3.1.2.

11. **Estimate distribution and correct order relations using ORDREL**
    The subroutine ORDREL is used to rescale the values in the distribution into the interval [0;1]. The probabilities of all the vegetation types on each location are summed and every probability is divided by the sum.

12. **Draw a random number using ACORNI**
    To draw from the distribution of probabilities of the different vegetation types on the current location of the random path, a random number is generated using ACORNI.

13. **Draw from distribution using BEYOND**
    The random number of Step 12 corresponds to a number in the distribution of probabilities of vegetation types. The probability interval that contains this number corresponds to a certain vegetation type. This type is assigned to this location using the subroutine BEYOND.

14. **Save the drawn vegetation type in an array**
    The vegetation type that is drawn from the distribution is saved in an array, using the unique location number as location in the array. After all locations of the random path are visited, this array will be filled with vegetation types for all locations.

15. **Calculate the residuals and add these to the conditional data**
    When a vegetation type is drawn from the distribution, this information needs to be added to the simulation. The new residual value is generated by calculating the difference between the realised probability of occurrence and the value in the regression maps on this location. The probability of occurrence is presented by the value 1 for the realised type and the value 0 for the other vegetation types. The calculated residual value of each vegetation type is inserted in the array that will be used in Step 9, where a new residual is estimated on the next location.

16. Write array with simulated vegetation types to output file

Once the random path is completed, the array with drawn vegetation types is written to an output file.

By repeating Steps 8 to 16 as many simulations as desired can be performed. The more simulations, the more precise the calculations will be that can be performed with the output.

4.2.3 Output

The maps that are output of the model have a data value from 1 to 8 (referring to the 8 vegetation types described in Chapter 2.2) for every location in the study area. For locations outside the study area, where no data is available, grid cells have value -9999. The name of the output file is Mapxxx.asc with xxx as consecutive number of the simulation.

5 Application of the simulation program to the study area

In this chapter the methodology described in Chapters 3 and 4 is applied to the study area. The chapter starts with the verification of the output that was generated with the conditional simulation program described in Chapter 4. Next, in Section 5.2, the area distribution of the different vegetation types over time is discussed, as well as temporal trends. The uncertainty of the calculated areas is also presented. In Section 5.3 these vegetation changes are linked to the nature conservation value in the study area. Uncertainty in the nature conservation value is also discussed. Finally, an example of significance testing is presented in Section 5.4.

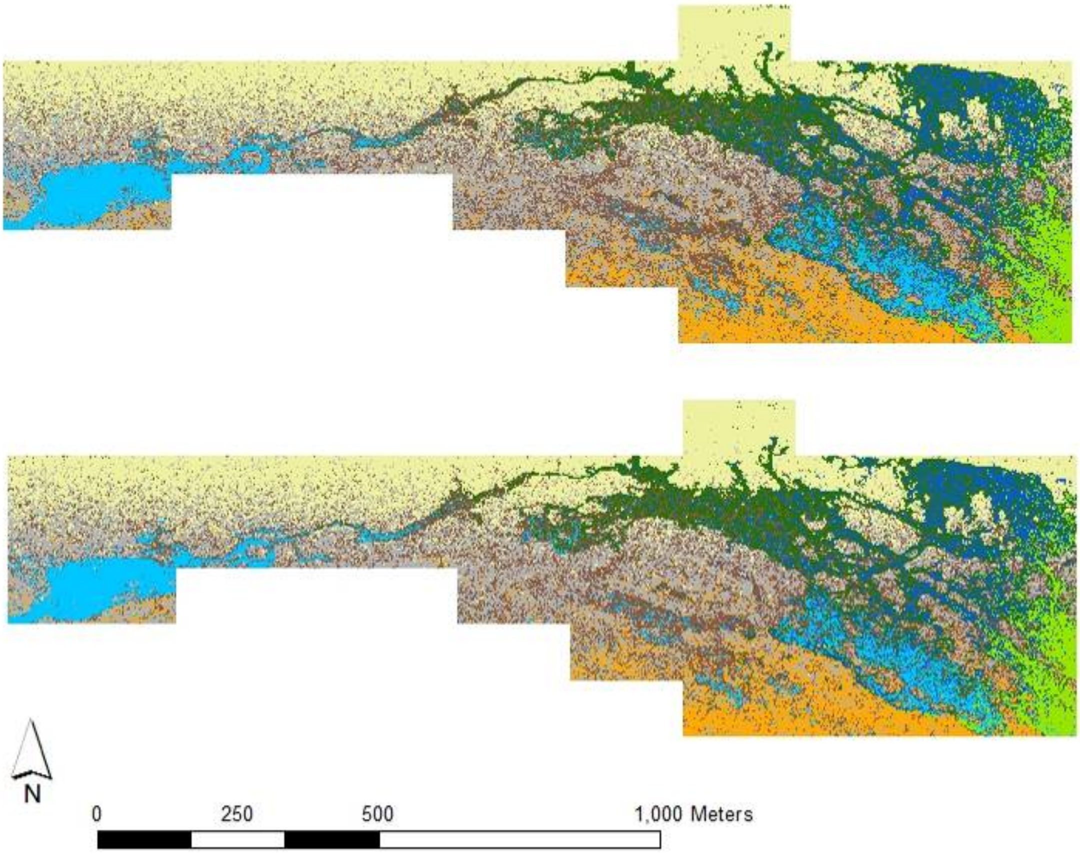


Figure 5.1: *Two different vegetation realizations of 2010, from the 500 realizations made.*

5.1 Verification of output

In Figure 5.1 two realisations of the year 2010 are visualized. These realizations look quite similar, which is an indication of a low uncertainty. There are however a lot of differences, since 28% of the grid cells do not have the same value in both realizations. In the realizations locations and vegetation types with a high spatial correlation contain less noise, since the probability distribution will be almost entirely occupied by one vegetation type. The noise on the locations with a low spatial correlation is caused by the probability distribution that contains similar probabilities for several vegetation types. The random number drawn to determine the vegetation type from the distribution may pick in this case for adjacent grid cells different vegetation types, which leads to noise in realizations.

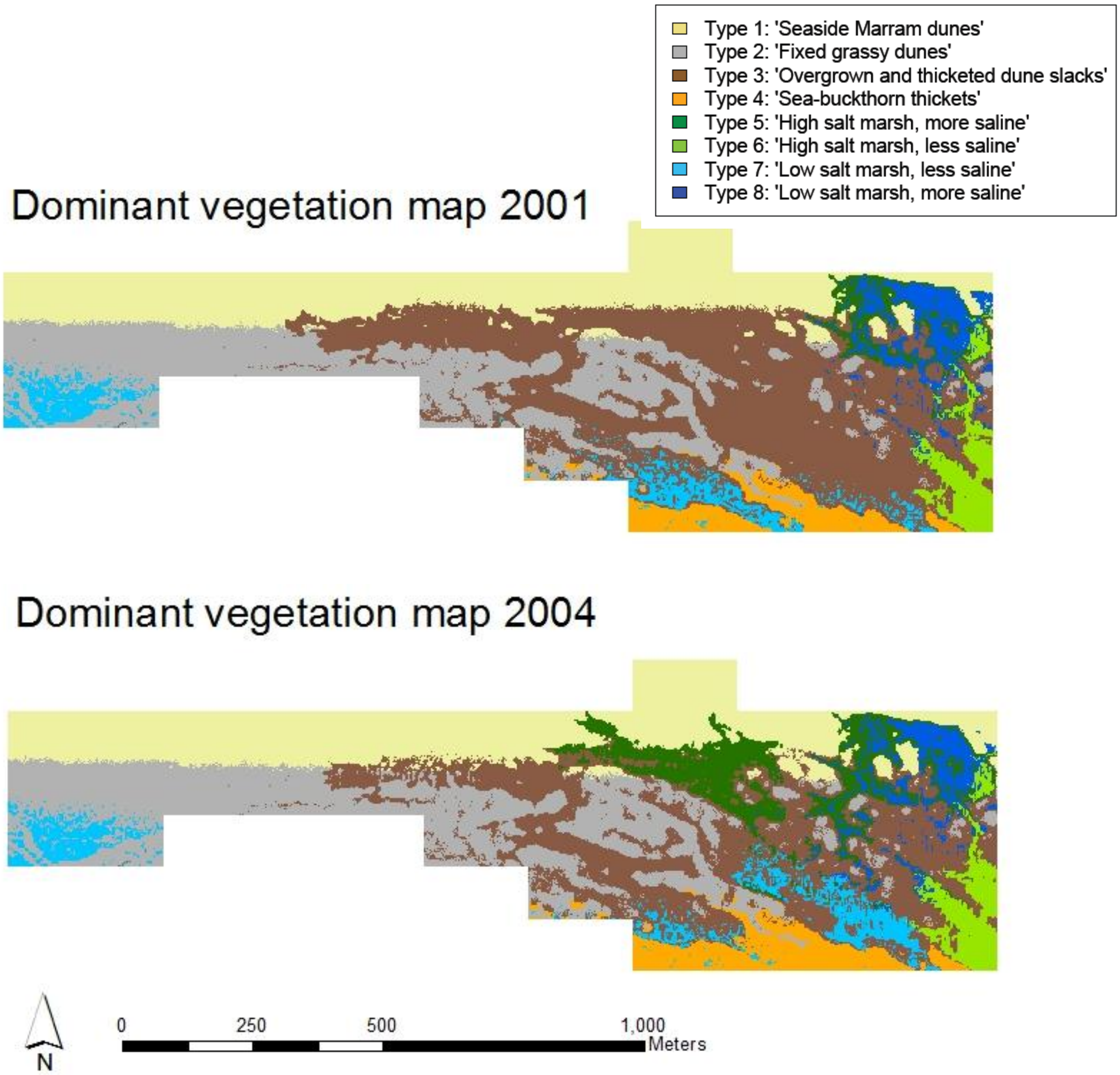


Figure 5.2: *Dominant vegetation maps of 2001 and 2004 made using the areas of the 500 realizations for these years and the input regression model of Vreugdenhil (2011).*

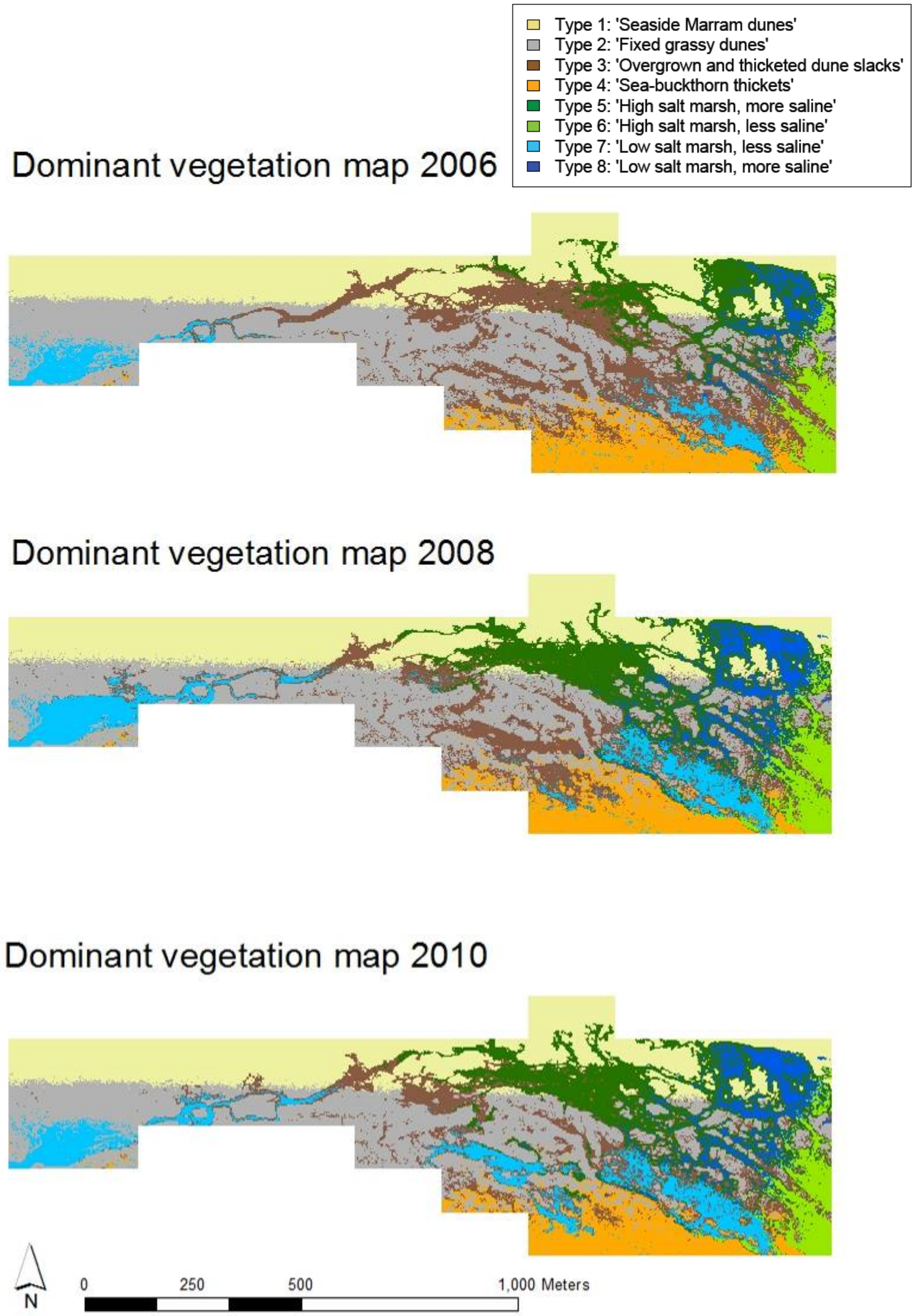


Figure 5.3: *Dominant vegetation maps of 2006, 2008 and 2010, made using the areas of the 500 realizations for these years and the input regression model of Vreugdenhil (2011).*

From the realizations areas of the different vegetation types were calculated. To compare the output of the conditional simulation program to the results of Vreugdenhil (2011), dominant maps were calculated for every observation year, as described in Section 3.1.3. In Figures 5.2 and 5.3 the dominant vegetation maps, made with the area calculations of the 500 realizations for all observation years, are shown.

The dominant vegetation maps have a less noisy character than the realizations, since these represent the vegetation types with the highest probability for every location. From 2004 onwards, a more detailed Digital Elevation Model (DEM) was used, which leads to more detailed maps. Changes in vegetation over time are evident from the maps, and are discussed in the next section.

The deviation (in percentage) of the areas of the dominant vegetation maps produced in this research from the results from Vreugdenhil (2011) is shown in Figure 5.4. The differences are positive when the realizations made in this research lead to a higher dominant area calculation than Vreugdenhil (2011) and negative percentages are the result of the calculation of a lower area.

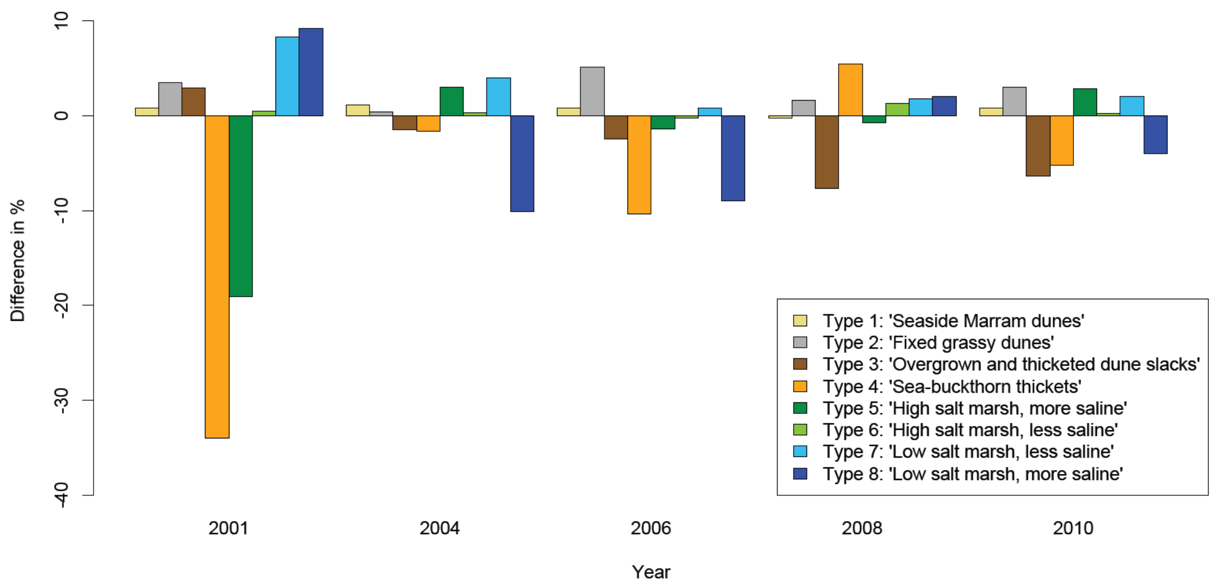


Figure 5.4: *Differences between the dominant area calculated by Vreugdenhil (2011) and this research.*

The differences in Figure 5.4 are mainly caused by the amount of realizations made. More realizations will lead to smaller deviations. The biggest differences are found in 2001. As discussed in Section 3.1.1 the amount of measurement points was 70 in 2001, while in the other observation years this was 140. Fewer observations cause higher uncertainty, hence a bigger magnitude of the differences. The biggest deviations concern vegetation types 4 and 5. The dominant areas of these vegetation types were estimated substantially lower with conditional simulation in comparison to regression kriging. Reason for this may be the small area occupied by these types in 2001 (see Figure 5.5). A smaller area easier leads to a higher percentage of difference. The average deviation was -0.74%, from 2004 onwards. This is a low deviation, from which is concluded that enough realizations are made.

5.2 Vegetation changes

The vegetation has clearly changed over time, as visible from the dominant vegetation maps in Figure 5.2 and 5.3. In Section 5.2.1, the changes in area of the different vegetation types are presented. In Section 5.2.2, the uncertainty about the changes is demonstrated using box plots and the coefficient of variation.

5.2.1 Trends

Figure 5.5 shows the relative distribution of the different vegetation types in the study area. The areas presented in this figure are averages of the realizations calculated using the expected value, as explained in Section 3.1.2. The area of vegetation type 3 (Overgrown and thicketed dune slacks) is clearly declining, while the areas with salt marsh vegetation are increasing (types 5, 6, 7 and 8).

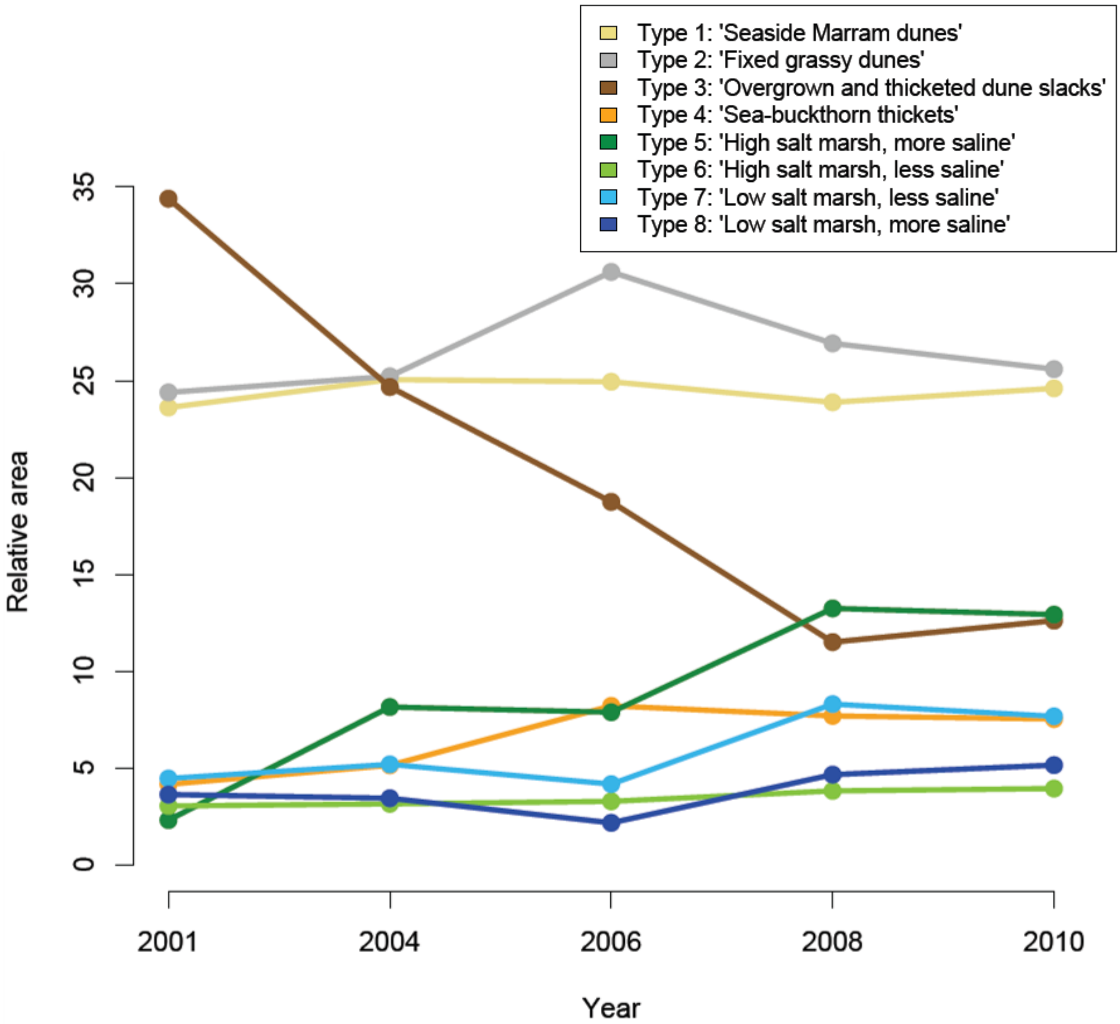


Figure 5.5: *The relative area of the different vegetation types. Calculations based on the regression model of Vreugdenhil (2011) and average areas of 500 realizations per measurement year.*

Krol (2011) and Slim et al. (2011) attributed the changes in vegetation composition mainly to soil moisture and salt levels in the dune valleys. These factors lead to an increase in salt marsh vegetation (Krol, 2011; Slim et al., 2011). Increased moisture in the study area is probably not caused by a single factor. Krol (2011) mentions four different aspects possible:

1. By chance the weather can be more wet;
2. The Wadden Sea is inundating the area more often. The valleys are inundated for several months each winter. The frequency of inundation has been evaluated and an increase from 3 to 5 times per year was found, with a subsidence level of 33 cm. The duration of inundation increased strongly since the beginning of the subsidence in 1986. According to Krol (2011), this is a combined result of flooding with seawater and increased rainfall. The relative level of the groundwater in the area has risen due to soil subsidence. As a result, especially after a flooding with seawater, it takes a long time before the ground surface is dry again. Krol (2011) reports this as the most important significant change, which also has impact on local nature;
3. The subsidence because of gas extraction can cause increased moisture;
4. Nature conservation organisation 'It Fryske Gea' performed a nature development project in the western part of the study area in 2005. During this project the ground was lowered artificially, and therefore moisture and salinity increased. This last aspect certainly accounts for a decrease in area for vegetation type 2 (Fixed grassy dunes) from 2006 onwards.

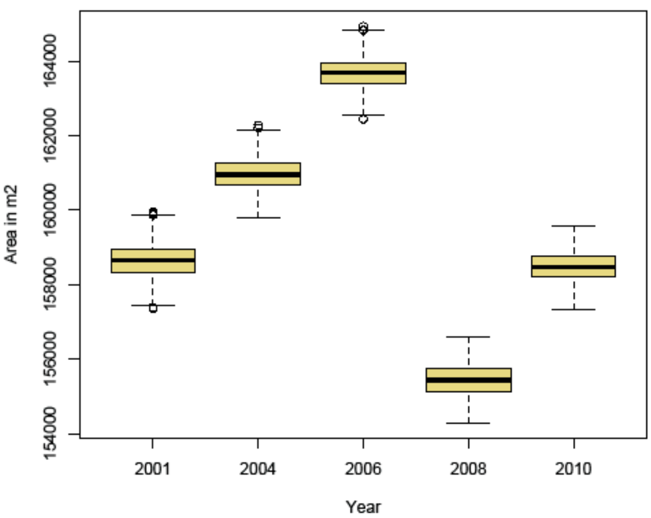
The monitoring report refers to an extreme event after 23 years of gas extraction (Krol, 2011; Slim et al., 2011), which occurred in spring 2007. In that year, late inundation (18 March) was followed by a long drought that caused a salt crust when the ground surface dried up at the end of April. The extremely high salinity resulted in a total mortality of the existing freshwater vegetation and a successive settlement of salt marsh plant species the same year. "Without this extreme event, the change in vegetation would have been gradually or would not have occurred" (Slim et al., 2011 p. 172).

It seems that the abrupt changes in vegetation composition were not caused by the significant monotonous subsidence effect, but rather by the nature development project in 2005 and the extreme event in spring 2007. The overall change was from the lower numbered vegetation types towards the higher numbers, which represent the more saline marsh vegetation (Slim et al., 2011). In Section 5.3, the implications of the changes in vegetation for the nature conservation value are discussed.

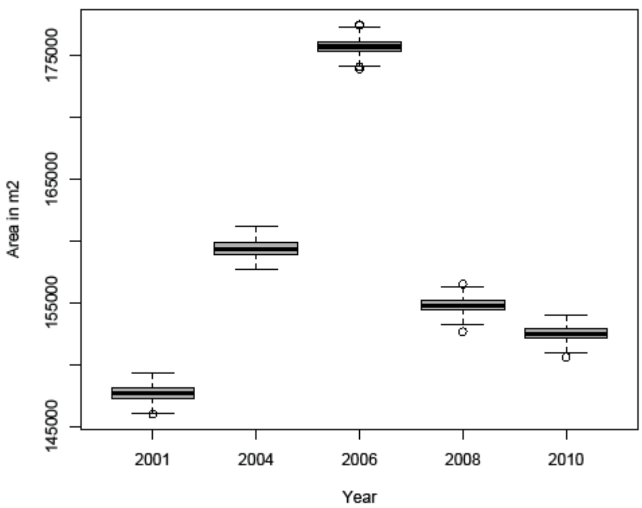
5.2.2 Uncertainty

In earlier research it was not possible to assess the influence of, for example, interpolation errors on area calculations. This section reveals the spread of the area predictions and the uncertainty about the area of the different vegetation types using box plots and the coefficient of variation.

Area of vegetation type 1: 'Seaside Marram dunes'



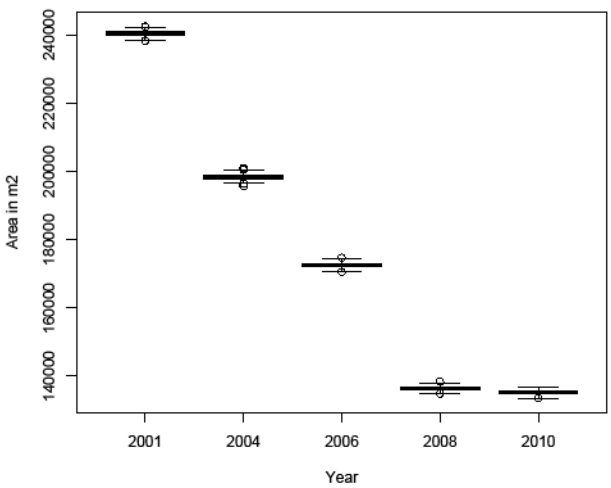
Area of vegetation type 2: 'Fixed grassy dunes'



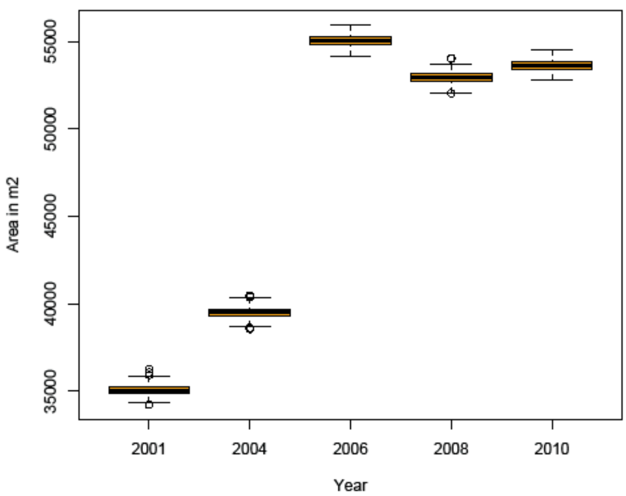
Type 1	2001	2004	2006	2008	2010
Lower adjacent value	157447	159802	162551	154290	157345
Q1	158323	160671	163391	155139	158198
Q2 (median)	158651	160954	163683	155437	158477
Q3	158941	161262	163964	155734	158780
Upper adjacent value	159862	162146	164820	156582	159590
Interquartile Range	618	591	574	595	582
Mean	158631	160961	163686	155437	158487
Standard deviation	483	438	438	421	417

Type 2	2001	2004	2006	2008	2010
Lower adjacent value	146035	157738	174147	153207	151002
Q1	147288	158926	175324	154420	152148
Q2 (median)	147715	159370	175737	154796	152510
Q3	148126	159842	176140	155249	152918
Upper adjacent value	149365	161175	177359	156341	154054
Interquartile Range	838	916	816	829	770
Mean	147718	159384	175750	154815	152515
Standard deviation	618	629	628	594	576

Area of vegetation type 3: 'Overgrown and thicketed dune slacks'



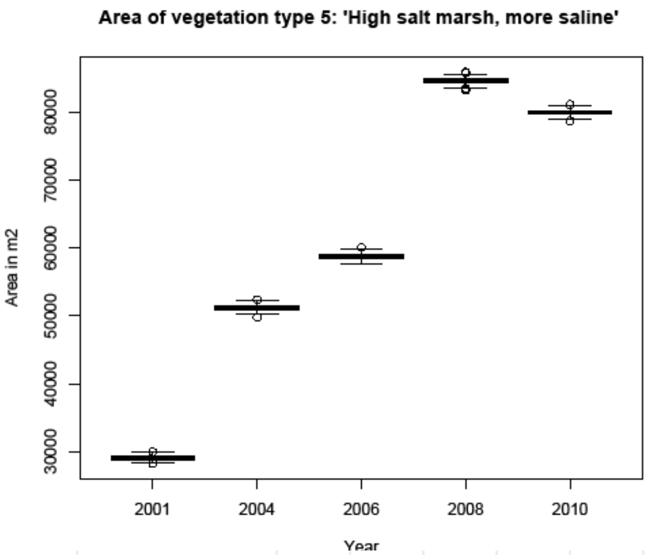
Area of vegetation type 4: 'Sea-buckthorn thickets'



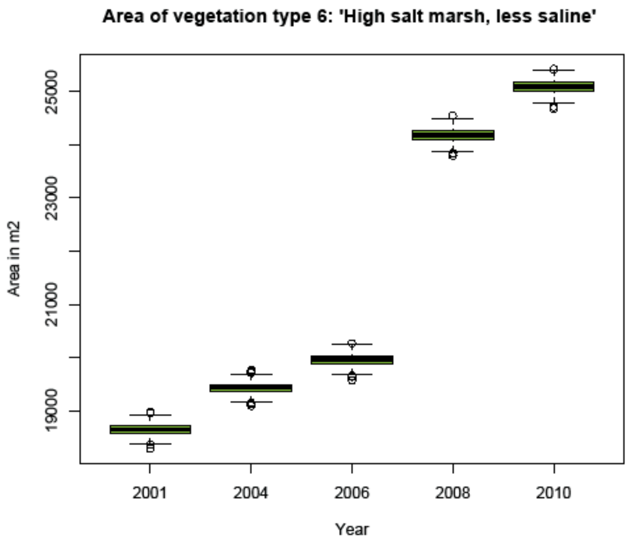
Type 3	2001	2004	2006	2008	2010
Lower adjacent value	238622	196723	170644	134964	133552
Q1	239998	197976	172060	136106	134782
Q2 (median)	240474	198440	172519	136474	135193
Q3	240976	198936	173005	136879	135606
Upper adjacent value	242060	200347	174326	138016	136731
Interquartile Range	978	961	945	773	824
Mean	240483	198456	172510	136485	135205
Standard deviation	733	736	698	588	575

Type 4	2001	2004	2006	2008	2010
Lower adjacent value	34317	38711	54123	52066	52771
Q1	34865	39275	54811	52707	53418
Q2 (median)	35054	39506	55041	52937	53643
Q3	35266	39692	55284	53147	53872
Upper adjacent value	35816	40315	55939	53731	54516
Interquartile Range	401	417	474	440	454
Mean	35066	39489	55044	52934	53644
Standard deviation	300	328	340	308	323

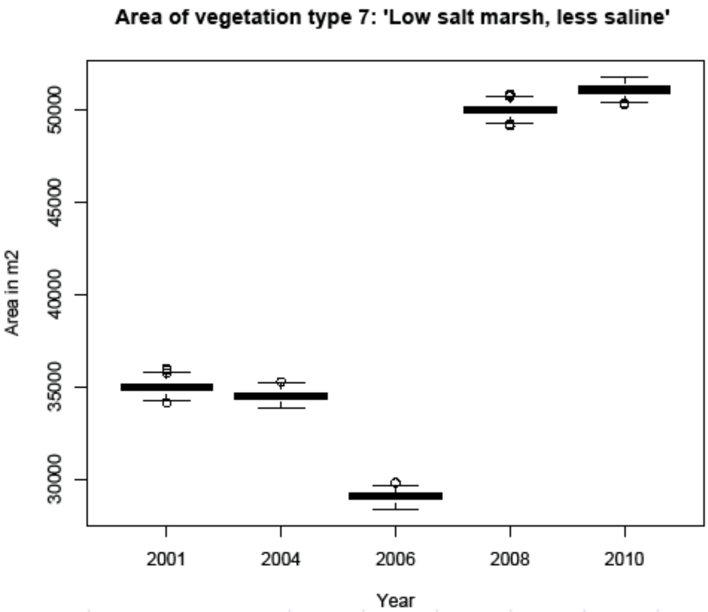
Figure 5.6: *Box plots with the area of vegetation type 1, type 2, type 3 and type 4. (Note the different scales on the Y-axis.)*



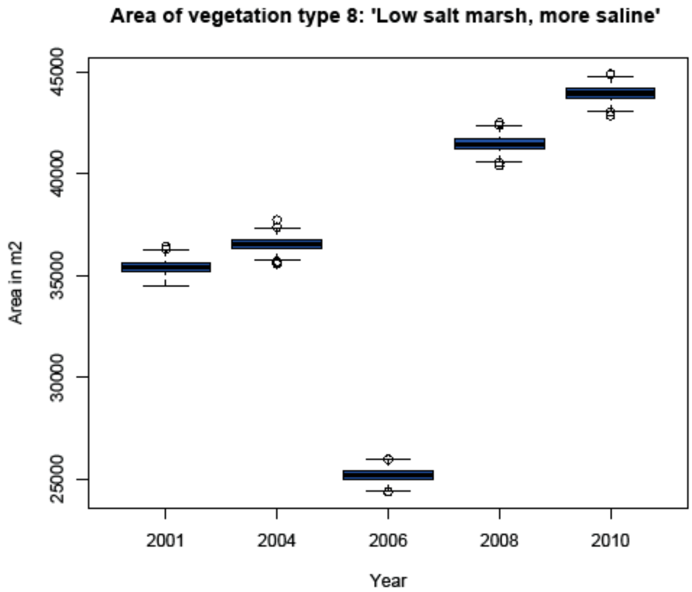
Type 5	2001	2004	2006	2008	2010
Lower adjacent value	28257	50192	57652	83609	78932
Q1	28850	50943	58505	84372	79723
Q2 (median)	29053	51229	58784	84653	79971
Q3	29262	51479	59079	84909	80278
Upper adjacent value	29852	52221	59910	85662	81015
Interquartile Range	412	537	574	537	555
Mean	29059	51225	58787	84649	79995
Standard deviation	301	382	406	428	412



Type 6	2001	2004	2006	2008	2010
Lower adjacent value	18380	19157	19686	23882	24779
Q1	18577	19360	19885	24099	25014
Q2 (median)	18640	19433	19964	24185	25096
Q3	18716	19497	20033	24262	25178
Upper adjacent value	18921	19692	20242	24481	25403
Interquartile Range	139	137	148	163	164
Mean	18642	19431	19959	24179	25094
Standard deviation	107	111	111	123	124



Type 7	2001	2004	2006	2008	2010
Lower adjacent value	34240	33840	28381	49330	50403
Q1	34800	34340	28890	49855	50931
Q2 (median)	34998	34506	29060	50053	51116
Q3	35185	34687	29239	50209	51289
Upper adjacent value	35763	35202	29673	50724	51793
Interquartile Range	386	347	349	354	358
Mean	34996	34512	29063	50036	51108
Standard deviation	293	261	246	276	246



Type 8	2001	2004	2006	2008	2010
Lower adjacent value	34511	35776	24451	40569	43082
Q1	35170	36330	25015	41241	43729
Q2 (median)	35402	36559	25202	41457	43954
Q3	35622	36746	25391	41690	44181
Upper adjacent value	36245	37359	25955	42324	44782
Interquartile Range	452	417	377	449	452
Mean	35404	36543	25200	41465	43951
Standard deviation	321	329	282	340	345

Figure 5.7: *Box plots with the area of vegetation type 5, type 6, type 7 and type 8. (Note the different scales on the Y-axis.)*

Figures 5.6 and 5.7 present box plots for the vegetation areas for all vegetation types and observation years, computed from the 500 realizations per measurement year. The tables under the graphs (in Figures 5.6 and 5.7) contain the lower adjacent value; the first quartile (Q1); second quartile (Q2, also called the median); third quartile (Q3), the upper adjacent value and the Interquartile Ranges. The latter express the width of the interval between the 25 and the 75 percentiles of the area of vegetation types (m^2) and can be used to compare uncertainty between vegetation types. The mean and standard deviation cannot be read from a box plot, but are given in the table to enhance interpretation of the data. Furthermore, the mean and standard deviation were used to calculate the coefficient of variation, which is presented later in this section.

The box plots in Figure 5.6 and 5.7 indicate a significant change for most consecutive measurement moments, while most whiskers do not overlap, as explained in Section 3.3. Eight changes between consecutive observation years do not have overlapping whiskers. For these cases the amount of overlap between the whiskers is described in Table 5.1. In the previous section the nature development project and extreme weather are mentioned as important factors influencing changes in vegetation. The vegetation changes that are consequences of these events appear as significant in the box plots. For example: between the measurements in 2006 and 2008 an extreme weather event occurred, which lead to an increase in salt marsh vegetation. In the box plots all salt marsh vegetation types (5, 6, 7 and 8) have increased significantly between 2006 and 2008.

***Table 5.1:** Amount of overlap of whiskers in box plots in Figure 5.6 and 5.7.*

	Observation years	**Area (m^2) between adjacent values**
Type 1	2001-2004	60
Type 2	2008-2010	847
Type 3	2008-2010	1767
Type 4	2008-2010	960
Type 6	2004-2006	6
Type 7	2001-2004	38
Type 7	2008-2010	321
Type 8	2001-2004	469

Type 1: 'Seaside Marram dunes'
Type 2: 'Fixed grassy dunes'
Type 3: 'Overgrown and thicketed dune slacks'
Type 4: 'Sea-buckthorn thickets'
Type 5: 'High salt marsh, more saline'
Type 6: 'High salt marsh, less saline'
Type 7: 'Low salt marsh, less saline'
Type 8: 'Low salt marsh, more saline'

Only for vegetation type 5 (High salt marsh, more saline) all changes are significant. Half of the cases in which the whiskers overlap are between 2008 and 2010, implying that between these measurement periods probably less change occurred. Also between 2001 and 2004 the area of Low salt marsh (type 7 and 8) did not change significantly and neither did the area of Sea-side Marram dunes (type 1). High salt marsh less saline (type 6) has an overlap of the adjacent values of 6 m^2. With such a small overlap, the probability of a significant change will be high.

Figure 5.8 visualizes the CV for all vegetation types over time. The CV indicates which vegetation type has a relatively large uncertainty compared to others.

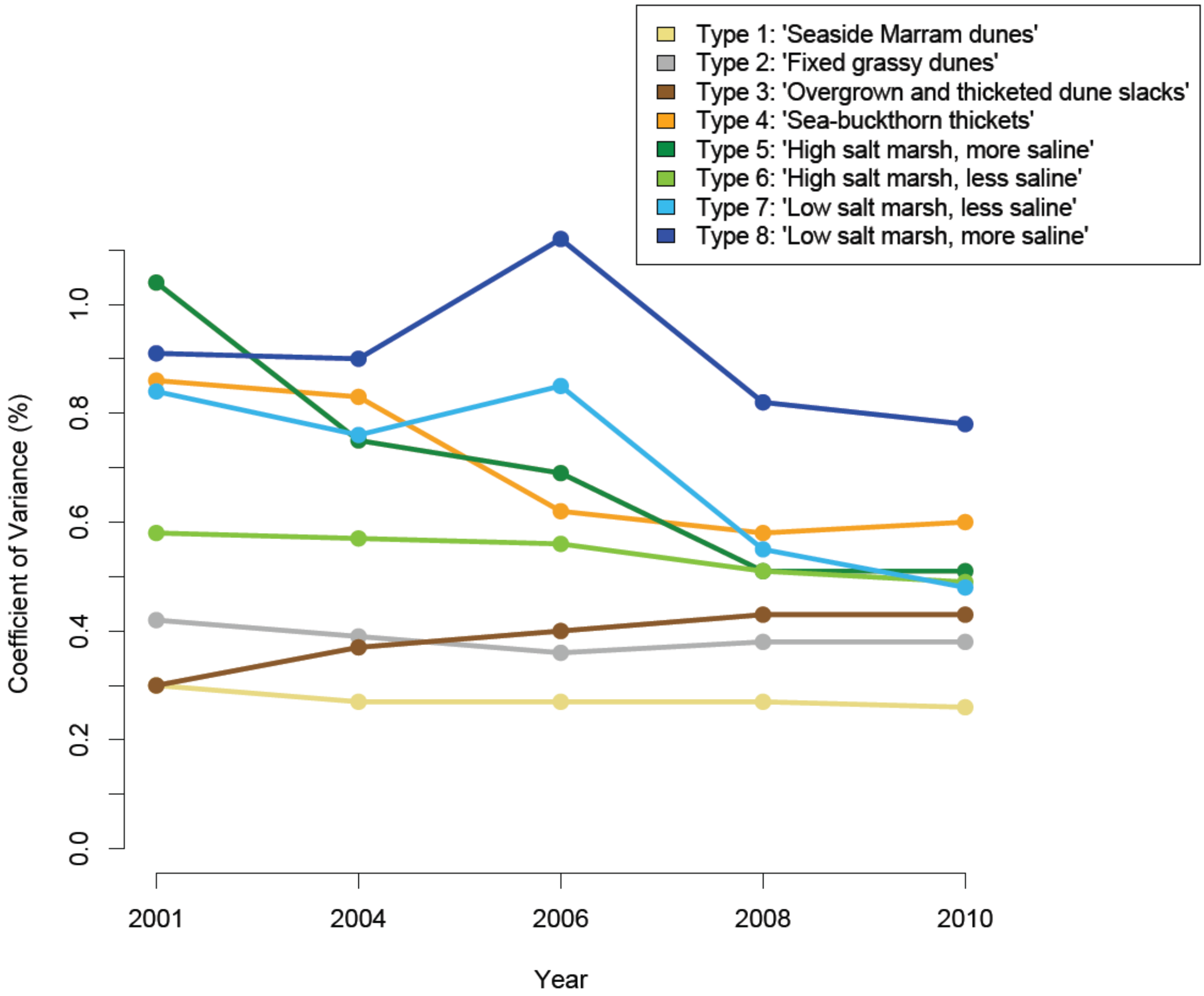


Figure 5.8: *The coefficient of variation for the different vegetation types for every observation year.*

When Figures 5.8 and 5.5 are compared, it is clear that a smaller area leads to a higher CV. This is explained by the smaller value in the denominator (see Section 3.3). There is relatively more spread in the area predictions when total area is smaller, but for almost all vegetation types the CV is smaller than 1%, i.e. the standard deviation is less than 1% of the mean of the data.

5.3 Nature conservation value

5.3.1 Trends

The nature conservation value is calculated for every observation year, according to the methodology described in Section 3.3. The result of this calculation is given in Figure 5.9. The trend in NBW in is positive. This trend is caused by the increase in area of salt marsh vegetation, which is considered valuable in the Gelderland method, as described in Table 3.3.

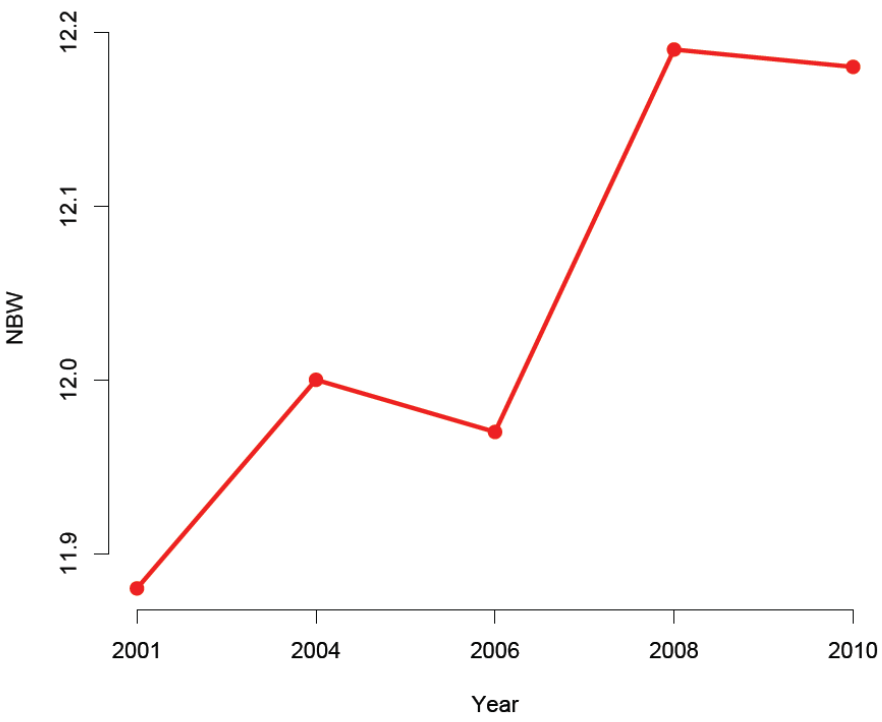


Figure 5.9: *The nature conservation value for every observation year.*

5.3.2 Uncertainty

Table 5.2 gives the lower and upper nature conservation value of the 90% confidence interval of the nature conservation value. There is very little uncertainty about the nature conservation value. The average difference between the boundaries of the confidence is 0.012% of the average NBW value.

Table 5.2: *The nature conservation value for the upper and lower bound of the 90% confidence interval in area.*

	5th percentile	95th percentile
2001	11.8811	11.8840
2004	12.0047	12.0063
2006	11.9747	11.9766
2008	12.1921	12.1926
2010	12.1868	12.1872

5.4 Significance of change

Figure 5.10 shows a histogram of the differences in area of vegetation type 2 (Fixed grassy dunes) between 2008 and 2010. This histogram illustrates the spread of differences in squared meters. The methodology used for this calculation is described in Section 3.3. Calculating the differences visualised in the histogram took a lot of calculation time, therefore the significance analysis was only performed for one vegetation type and one measurement interval.

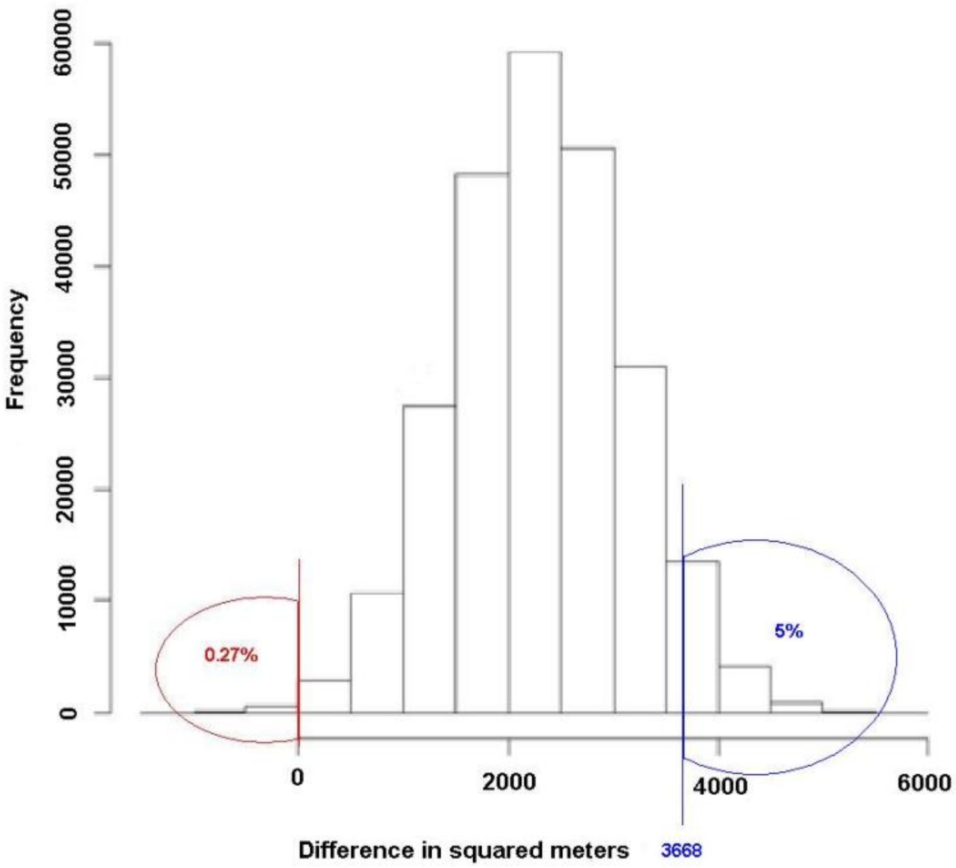


Figure 5.10*: Histogram of differences in squared meters of vegetation type 2 (Fixed, grassy dunes) between 2008 and 2010.*

The 95th percentile (the upper amount of change with a confidence level of 95%) was 3668 m², indicated on the right in Figure 5.10. From this was concluded with a probability of 95% that the area of this vegetation type has not increased more than 3668 m².

For an increase in area of 0 m² or less, the probability is 0.27%, indicated on the left in Figure 5.10. This probability is very low, hence the probability of a change in area bigger than 0 m² is very high with 99.73%. It is therefore highly probable (99.73%) that vegetation type 2 has not decreased in area between 2008 and 2010.

6 Discussion and conclusions

This chapter gives an overview of the work done in this research, as well as suggestions for possible improvements. First, the methodology used in this research is discussed, as well as the input. In Section 6.2, answers are provided to the research questions formulated in the Introduction and conclusions are drawn. Finally, future research is suggested and recommendations are given.

6.1 Discussion

In this section the methodology of the research is discussed, starting with the input of the program. Then, the implementation in terms of the conditional simulation program is critically reflected, as well as the analysis of the output of the program. Finally, the objectivity and repeatability of this research are discussed.

6.1.1 Input of the program

The sample design used random selection, which introduces the possibility that selected sites do not belong to the population. For example, in 2001, 1 of the 70 samples was in a barely vegetated area on the beach, which caused this measurement location to be skipped. Other possibilities of locations that have to be skipped when selected are for instance, roads or bike sheds. The fact that samples are skipped is not a problem when enough samples are taken. Otherwise, it will be considered as a weakness of the sampling scheme.

When point measurements are used to predict the vegetation cover in an area, there might be changes or groups of plants that are not measured, but which could have changed the outcome of the measurement. Additionally, these deviations can be temporally correlated: when in one year some change is missed, that change can persist in other years. These are errors for which no correction can be made.

The calculation of the explanatory variables is mostly based on a DEM. This DEM is not perfect: errors such as bushes that are measured as ground level occur, therefore the maps with explanatory variables may not be correct. Errors in a DEM influence the outcome of this research.

In this research the semivariograms of Vreugdenhil (2011) were used. These are different than the ones presented in Slim et al. (2011). Also, Vreugdenhil defined a negative pure nugget semivariogram for vegetation type 6 (High salt marsh, less saline), which is incorrect. This was adjusted to a nugget with the value of the variance of the residuals over all measurement moments.

6.1.2 Implementation

Only one random path was used for all realizations. In theory this does not result in a bias in the results, since for every location a new random number is determining the vegetation type on this location, and the data is conditioned to this drawn vegetation type. The sequence of the random path should therefore not influence the outcome.

On the other hand, the size of the neighbourhood can have an influence on the outcome. When neighbourhoods are smaller than the ranges of the semivariograms, the spatial correlation is not represented correctly, which can even lead to artefacts. In this research the size of the neighbourhood consisted of 12 nearest observations. This size, in combination with small ranges of most semivariograms, leads to the expectation that the influence of the neighbourhood is limited.

GSLIB has two predefined subroutines for searching, but these were not used since they depend on a covariance lookup table. This table was not incorporated in the program written in this research. The search algorithm used evaluates the border of a box, which increases in size until enough neighbours are found. The maximum radius of the box was set to 1000.

The reasons for using a neighbourhood instead of using the whole grid as nearby conditioning data are the limits of the CPU and memory of the computer used. Doubling the number of data leads to an eightfold increase in CPU time and a fourfold increase in the memory requirements (Deutsch & Journel, 1998, page 33).

For this research 500 simulations were performed. The more simulations, the more precise the estimation of the distribution with areas of vegetation types will be. However, the number of 500 simulations considered is large enough for adequately estimating the distribution, as follows from the comparison with Vreugdenhil (2011) in Section 5.1.

6.1.3 Analysis of the output

A longer time series could provide more accurate differentiation between influences on vegetation change: extreme weather events, the nature restoration project and the influence of subsidence.

The assessment of nature is always subjective and dependent on developments in knowledge and policy (Hertog & Rijken, 1996). Therefore, the nature conservation value determined in this research is also subjective. Different methods might discover different trends or values.

6.1.4 Objectivity and repeatability

Within scientific research objectivity and repeatability are cornerstones. But, since the model-based approach is chosen: “All uncertainty assessments are model-dependent, and are no better than the model on which they are built. In this regard no uncertainty assessment, no estimate, no statistical test can claim objectivity” (Deutsch & Journel, 1998: p. 19). The model on which this research is based is the regression kriging model with explanatory variables and semivariograms of Vreugdenhil (2011). This model influences the outcome of the stochastic simulation program.

Repeatability is achieved using the following measures:

- Using a fixed random seed number. This random number identifies each realization. Realizations are completely specified by the data, the algorithm and the random number seed, so the research can be repeated exactly when these parameters are the same (Deutsch & Journel, 1998: p.189). When another seed number is used, the results will be different, and so might be the areas calculated from the realizations, and the uncertainty. The random seed number used for all realizations is 69069.

- The program is flexible in the number of vegetation types. This enables application of the conditional simulation program to different case studies in the future.

6.2 Conclusions

The aim of the research was to develop and apply a method to express and model uncertainty about the area of different vegetation types (at different moments time) on the eastern part of the island of Ameland. Answering the research questions fulfilled this research objective. General conclusions are derived at the end of this section.

6.2.1 Research questions revisited

1. *How can spatial stochastic simulation be used for modelling uncertainty about the area of each of the vegetation types in the study area for all measurement years?*

With regression kriging, residuals (differences between the regression and the measured vegetation types in the field) were interpolated using kriging, to correct the regression model. The regression kriging method provided a probability for each vegetation type, starting on the first location of a random path. The probabilities of all vegetation types together on this location form a distribution, from which the conditional sequential indicator simulation algorithm randomly drew a vegetation type. The drawn vegetation type is added to the dataset used for regression kriging before the next location was visited. In this way, the simulated values were conditioned to the data and previously simulated vegetation types. The simulation visited every grid cell sequentially. With every simulation a realization was produced. Finally, for all realizations separately, the area per vegetation type was calculated. This calculation resulted in many different values for the area of a vegetation type. The distribution of the areas of one vegetation type was visualized using a box plot. From the box plots of different moments in time, the uncertainty of the changes was derived per vegetation type. To compare the uncertainty between the different vegetation types, the coefficient of variation (CV) was used.

2. *How can spatial stochastic simulation be implemented, and which software environment is suitable?*

In this research the regression kriging and conditional sequential indicator simulation were implemented using GSLIB. The desired functionality was not available in this program, so modifications had to be made to existing code. The choice for GSLIB software is motivated in Section 4.1. The implementation of the conditional simulation program is explained in Section 4.2. A technique was incorporated to increase speed: use one random path for all realizations and calculate kriging weights only once and save the result. When one random path is used, all cells will have the same neighbours (from which the data values are interpolated using kriging) for all simulations. When the weights and locations of these data values are stored, these can be used for all simulations, and need to be calculated only once.

3. *What is the size and uncertainty about the area of the different vegetation types according to the model developed in this research?*

Before calculating areas, the output of the simulation program was verified. In previous research the areas of vegetation types were calculated using dominant maps: the pixels in these maps have the value of the vegetation type with the biggest probability on this pixel. The dominant areas calculated with the simulation program were compared to the ones acquired in previous research (both studies used the same parameters). The verification showed acceptable results, since the magnitude of the deviations was on average 0.74% from 2001 onwards.

The sizes of the areas changed substantially over time towards more saline vegetation. These changes could for the greater part be explained by a nature development project and an extreme weather event.

The uncertainty about the areas is low. There are only 8 cases with overlapping whiskers between two consecutive measurement moments (see Chapter 5.2). To quantify the uncertainty an example of a significance calculation was done in Chapter 5.4, which lead to a probability of 99.73% on a change in area bigger than 0 m² for vegetation type 2 (Fixed, grassy dunes) between 2008 and 2010.

The CV, which was used to compare the uncertainty between the different vegetation types, was mostly below 1%. The highest values in the CV correspond with vegetation types with a small area, which results in a smaller denominator in the fraction, hence a greater CV.

4. *What is the influence of uncertainty about the area of the different vegetation types on the nature conservation value in the study area?*

The method used to determine the nature conservation value is the 'Gelderland' method (NBW, see Section 3.3). The trend in conservation value was positive, since the area of salt marsh vegetation, which has a high conservation value, increased over time. The uncertainty about the nature conservation value was very low: the 90% confidence interval of the conservation value is 0.012% of the average NBW value.

6.2.2 General conclusions

The estimates of the areas calculated by the monitoring project are considered as adequate, since their uncertainty is very low. This can be explained from the thorough sampling design, detailed maps of explanatory variables and a suitable interpolation method.

Also, it is concluded from this research that gas extraction does not influence the vegetation in the study area in a negative way. Because of changes towards more saline vegetation, the nature conservation value increased substantially.

Consequences of these conclusions are that gas extraction does not have to be discontinued at this point in time, since the nature conservation value did not decrease and the monitoring program is detecting changes adequately.

6.3 Recommendations and future research

It is highly recommended to repeat the simulations using the regression model of Slim et al. (2011), since this model is providing a much better prediction of the location of certain vegetation types. This recommendation is made, as the regression model is the foundation of this research. Slim et al. (2011) used more explanatory variables than Vreugdenhil (2011), which leads to a better regression model and smaller residuals. In this research the regression model of Vreugdenhil (2011) was used, since the model of Slim et al. (2011) was not available when the research started.

Other recommendations for further research are:

- When conclusions about significance for all changes are desired, it is recommended to test all deviations in Table 5.1, using the methodology in Chapter 5.4 or other methods (Besag & Clifford, 1989; Lin, 2005; Livezey & Chen, 1983; Hope, 1968).

- Currently, Alterra works on a design-based methodology for this case. It would be interesting to compare results between this design-based methodology and the model-based approach developed in this research.

- A topic for future research in stochastic simulation is to find the breakeven point in number of simulations from where it is less time consuming to save kriging weights, instead of calculating these again for every simulation.

- This research has used one method to assess the nature conservation value. It is interesting to compare results of different nature conservation value algorithms.

Hopefully, the program developed in this research will be beneficial to other projects that would like to asses uncertainty about changes in parameters for nature conservation purposes. Ideally, this research contributes to improved environmental awareness about the influences of gas extraction.

References

Atkinson, P.M. (1999) Geographical Information Science: Geostatistics and Uncertainty, *Progress in Physical Geography* 23; 134–142

Besag, J. & Clifford, P. (1989) Generalized Monte Carlo Significance Tests, *Biometrika* 76; 633-642

Bierkens, M.F.P. & Burrough, P.A. (1993a) The Indicator Approach to Categorical Soil Data: I. Theory, *Journal of Soil Science* 44; 361–368

Bierkens, M.F.P. & Burrough, P.A. (1993b) The Indicator Approach to Categorical Soil Data: II. Application to Mapping and Land Use Suitability Analysis, *Journal of Soil Science* 44; 369–381

Brus, D.J. (2000) Using Regression Models in Design-based Estimation of Spatial Means of Soil Properties, *European Journal of Soil Science* 51; 159-172

Brus, D.J. & De Gruijter, J.J. (1997) Random sampling or Geostatistical Modelling? Choosing Between Design-based and Model-based Sampling Strategies for Soil (with discussion), *Geoderma* 80; 1-44

Brus, D.J. & De Gruijter, J.J. (2011) Design-based Generalized Least Squares Estimation of Status and Trend of Soil Properties from Monitoring Data, *Geoderma* 164; 172-180

De Bruin, S. (2000) Predicting the Areal Extent of Land-Cover Types Using Classified Imagery and Geostatistics, *Remote Sensing of Environment* 74; 387-396

De Gruijter, J., Brus, D.J., Bierkens, M.P.F., Knotters, M. (2006) Sampling for Natural Resource Monitoring, *Springer-Verlag,* Berlin Heidelberg

Deutsch, C. V. & Journel, A. G. (1998) GSLIB Geostatistical Software Library and User's Guide, 2nd edition, *Oxford University Press*, New York

Eysink, W.D., Dankers, N., Dijkema, K.S., Van Dobben, H.F., Smit, C.J., De Vlas, J. (1995) Monitoring effekten van bodemdaling op Ameland-Oost, eerste evaluatie na 8 jaar gaswinning (interim report of the first monitoring program from 1988 – 2000*), Begeleidingscommissie Monitoring Bodemdaling Ameland*

Eysink, W.D., Dijkema, K.S., Van Dobben, H.F., Slim, P.A., Smit, C.J., De Vlas, J., Sanders, M.E., Wiertz, J., Schouwenberg, E.P.A.G. (2000) Monitoring Effecten van Bodemdaling op Ameland-Oost, Eerste Evaluatie na 13 jaar Gaswinning, *WL| Delft Hydraulics / Alterra*

Gómez-Hernández, J.J. & Srivastava, R.M. (1990) ISIM3D: An ANSI-C Three-Dimensional Multiple Indicator Conditional Simulation Program, *Computers & Geosciences* 16; 395-440

Goovaerts, P. (1997) Geostatistics for Natural Resources Evaluation, *Oxford University Press*, New York

Goovaerts, P. (1999) Geostatistics in Soil Science: State-of-the-art and Perspectives, *Geoderma* 89; 1–45

Goovaerts, P. (2001) Geostatistical Modelling of Uncertainty in Soil Science, *Geoderma* 103; 3–26

Goovaerts, P. (2010) Geostatistical software. In: Fischer, M.M. & Getis, A. (editors), Handbook of applied spatial analysis: Software tools methods and applications; 129–138, *Springer-Verlag*, Berlin

Gotway, C.A. & Rutherford, B.M. (1993) Stochastic Simulation for Imaging Spatial Uncertainty: Comparison and Evaluation of Available Algorithms, *Sandia National Laboratories*, Albuquerque, New Mexico, United States

Hengl, T., Heuvelink, G.B.M., Rossiter, D.G. (2007) About Regression-Kriging: From Equations to Case Studies, *Computers & Geosciences* 33; 1301–1315

Hertog, A.J. & Rijken, M. (1996) Geautomatiserde bepaling van natuurbehoudswaarde in vegetatie-opnamen, Appendix 5 in: Natuur, achtergronddocument bij de omgevingsplannen; 53-57, *Provincie Gelderland*, Arnhem, (http://www.natrix.demon.nl/gld/nbwmethode.html last accessed: 12-04-2012)

Hope, A.C.A. (1968) A Simplified Monte Carlo Significance Test Procedure, *Journal of the Royal Statistical Society, Series B* 30; 582-598

Journel, A.G. (1996) Modelling Uncertainty and Spatial Dependence: Stochastic imaging, *International Journal Geographical Information Systems* 10; 517–522

Kempen, B., Brus, D.J., Heuvelink, G.B.M., Stoorvogel, J.J. (2009) Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach, *Geoderma* 151; 311–326

Ketelaar, G., Van Der Veen, W., Doornhof, D. (2011) Monitoring effecten van bodemdaling op Ameland-Oost: 1. Bodemdaling, hoofdstuk 1 in: De Vlas, J. (2011) Monitoring effecten van bodemdaling op Ameland-Oost, Evaluatie na 23 jaar gaswinning, *Begeleidingscommissie Monitoring Bodemdaling Ameland*

Kott, P.S. (1991) A model-based look at linear regression with survey data, *The American Statistician* 45; 107-112

Krol, J. (2004) Monitoring van inundatie in duinvalleien op Oost-Ameland in 2001-2004, *Natuurcentrum Ameland*, Nes, Ameland

Krol, J. (2011) Monitoring van inundatie in duinvalleien op Oost-Ameland 2001-2011, hoofdstuk 4.1 in: De Vlas, J. (2011) Monitoring effecten van bodemdaling op Ameland-Oost, Evaluatie na 23 jaar gaswinning, *Begeleidingscommissie Monitoring Bodemdaling Ameland*

Lin, D. Y. (2005) An Efficient Monte Carlo Approach to Assessing Statistical Significance in Genomic Studies, *Bioinformatics* 21; 781-787

Livezey, R.E. & Chen, W.Y. (1983) Statistical Field Significance and its Determination by Monte Carlo Techniques, *Monthly Weather Review* 111; 46-59

Magnussen, S. & De Bruin, S. (2003) Updating Cover Type Maps Using Sequential Indicator Simulation, *Remote Sensing of Environment* 87; 161–170

Oliver, D.S. (1996) On Conditional Simulation to Inaccurate Data, *Mathematical Geology* 28; 811–17

Ott, R.L. & Longnecker, M. (2010) An Introduction to Statistical Methods and Data Analysis, 6th edition, *Brooks/Cole, Cengage Learning*

Pebesma, E.J. (1998) Gstat user's manual, *Utrecht University, Department of Physical Geography*, Utrecht

Pebesma, E.J. & Wesseling, C.G. (1998) GSTAT: A Program for Geostatistical Modelling, Prediction and Simulation, *Computers and Geosciences* 24; 17–31

Sanders, M.E., Slim, P.A., Van Dobben, H.F., Wegman, R.M.A., Schouwenberg, E.P.A.G. (2004) Effecten van eilandvarianten in de Noordzee op de ecologie van strand en duin, Alterra rapport 1092, *Alterra*, Wageningen

Slim, P.A. (1997a) Vooronderzoek duindoornsterfte duingebied Oost-Ameland, IBN-rapport 295*, Instituut voor bos en natuur onderzoek,* Wageningen

Slim, P.A. (1997b) Vooronderzoek meidoornsterfte duingebied Oost-Ameland, IBN-rapport 307, *Instituut voor bos en natuur onderzoek*, Wageningen

Slim, P.A., Heuvelink, G.B.M., Kuipers, H., Dirkse, G.M., Van Dobben, H.F. (2005) Vegetatiemonitoring en Geostatistische Vegetatiekartering Duinvalleien Ameland-Oost; 1-74, in: De Vlas, J. (2005) Monitoring effecten van bodemdaling op Ameland-Oost, Evaluatie na 18 jaar gaswinning, *Begeleidingscommissie Monitoring Bodemdaling Ameland*

Slim, P.A., Brus, D.J., Heuvelink, G.B.M., Heidema, A.H., Dirkse, G.M., Van Dobben, H.F. (2011) Vegetatiemonitoring en geostatistische vegetatiekartering duinvalleien Oost-Ameland 2001-2010, hoofdstuk 4.3 in: De Vlas, J. (2011) Monitoring effecten van bodemdaling op Ameland-Oost, Evaluatie na 23 jaar gaswinning, *Begeleidingscommissie Monitoring Bodemdaling Ameland*

Ter Braak, C. J. F., Brus, D. J., Pebesma, E. J. (2008) Comparing Sampling Patterns for Kriging the Spatial Mean Temporal Trend, *Journal of Agricultural Biological and Environmental Statistics* 13; 159-176

Vreugdenhil, C. (2011) Vegetatieverandering op Ameland-Oost, Bsc Thesis Bodem, Water en Atmosfeer, *Wageningen Universiteit*

Yao, T.T. (1998) Conditional Spectral Simulation with Phase Identification, *Mathematical Geology* 30; 285–308

Appendix I: Parameters for the conditional simulation program

This appendix gives parameters used for sequential conditional indicator simulations of vegetation types in 2001. For other years the same parameters are used except for the input file names in the first and second line.

```
          Parameters for SISIM
          ********************

START OF PARAMETERS:
res_1_2001.dat res_2_2001.dat res_3_2001.dat res_4_2001.dat res_5_2001.dat res_6_2001.dat
res_7_2001.dat res_8_2001.dat
Reg_type_1_2001.asc Reg_type_2_2001.asc Reg_type_3_2001.asc Reg_type_4_2001.asc
Reg_type_5_2001.asc Reg_type_6_2001.asc Reg_type_7_2001.asc Reg_type_8_2001.asc
1   2   0   3              -   columns for X,Y,Z, and variable
1900   188400    1.0            -nx,xmn,xsiz
600   608500    1.0            -ny,ymn,ysiz
1   1   1                      -nz, zmn, zsiz
69069                      -random number seed
12                         -maximum original data  for each kriging
12                         -maximum previous nodes for each kriging
1000                       - radius for search designed by Sytze
20.0  20.0  20.0           -maximum search radii (dummy variable for common block)
 0.0   0.0   0.0           -angles for search ellipsoid (dummy variable for common block)
1    0.03698816                -One   nst, nugget effect
1    0.06155751              -      it,cc
         26.24759               -      a_hmax
1    0.05462978                  -Two   nst, nugget effect
1    0.11670209                -      it,cc
         17.15274              -      a_hmax
1    0.07879114                  -Three nst, nugget effect
1    0.15662912                  -      it,cc
         69.86289                -      a_hmax
1    0.02827878                  -Four  nst, nugget effect
1    0.04692716               -      it,cc
         514.981                -      a_hmax
1    0.02713797                  -Five  nst, nugget effect
1    0.04773665                -      it,cc
         91.32331                -      a_hmax
1    0.008182                -    six, pure nugget (variance of residual values over all
years)
0    0.0                          - it,cc
        0.0                       -  a_hmax
1    0.014969311                 -seven   nst, nugget effect
1    0.00652702                  -      it,cc
         24                       -  a_hmax
1    0.01993458                   -eight   nst, nugget effect
1    0.021496332               -      it,cc
         40.59852                 -      a_hmax
```