Fine scale LiDAR DEM for modelling of plant distribution on a green beach, Schiermonnikoog Island

Ganna Novgorodova
March, 2011

Course Title:	Geo-Information Science and Earth Observation for Environmental Modelling and Management
Level:	Master of Science (MSc)
Course Duration:	September 2009 – March 2011
Consortium partners:	University of Southampton (UK) Lund University (Sweden) University of Warsaw (Poland) University of Twente, Faculty ITC (The Netherlands)

Fine scale LiDAR DEM for modelling of plant distribution on a green beach, Schiermonnikoog Island

by

Ganna Novgorodova

Thesis submitted to the University of Twente, faculty ITC, in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation for Environmental Modelling and Management

Thesis Assessment Board

Chair: Prof. Dr. A.K. Skidmore
External Examiner Dr. M. Roge-Wiśniewska
First supervisor: Drs. E.H. Kloosterman
Second supervisor: Dr. H.A.M.J. van Gils

UNIVERSITY OF TWENTE.
ITC
FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION

Disclaimer

This document describes work undertaken as part of a programme of study at the University of Twente, Faculty ITC. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the university.

Abstract

Green beach is a dynamic phenomenon of vegetation succession on a sandy beach. It contains a mosaic of vegetation from different habitats (dune, dune slack and salt marsh) that form a distribution gradient along changes in elevation. Elevation (a proxy for inundation frequency and duration), ground water quality (saline/fresh), soil clay content are the main abiotic factors influencing vegetation distribution.

The objectives of this research are: 1) modelling the relation between plant distribution on the green beach and the abiotic factors influencing it, 2) assessing the impact of the increase in raster resolution of the predictors on the performance of statistical modelling.

A set of indirect predictors are derived from a LiDAR DEM: elevation, cost-distance to the sea, distance to sea water inlet, distance to fresh water seepage and slope. The elevation values of the original DEM are used to interpolate DEMs of increased resolution. Three cell sizes are used: 5 m, 2 m and 20 cm. All the predictor variables are created using these resolutions.

The predictors of different resolution are used together with plants presence/absence data to estimate the empirical relation using logistic regression model. The model performance is assessed for each plant and cell size. The impact of the resolution change is assessed. Plant models that yield significant results are used to produce plant distribution maps.

The predictor variables show significant correlations with the plant distributions. Logistic regression yielded significant models for 7 out of 24 plants. The increase in resolution of the predictors shows an effect on all species modelled. However, a general pattern is not observed. The impact is different for different plants showing increase or decrease in model performance at some of the cell sizes.

The poor performance of the statistical model is mainly caused by lack of true absence data and limitations of the method.

The effects of the increased raster resolution of the predictors are thought to be connected to the ecology, scale of the plants distribution patterns, and to the effect on resolution increase on the predictors' accuracy.

Keywords: Green beach, Cell size impact, Logistic regression, Species distribution modelling, Presence/absence data

Acknowledgements

Дякую, Боже.
I thank my mum for being my mum, all my family for loving and caring for me no matter what happens. I thank my friends, those far at home for not forgetting me; and those nearby for sharing, caring, feeding me in times of hunger, and correcting my poor English in this thesis.

I am enormously grateful to my supervisors Henk Kloosterman and Hein van Gils for their guidance, good advice, support and always cheerful and sunny mood.
Also this work would not be the same without the help of Elske Koppenaal, who provided the 2006 validation data and some useful insight about the green beach.
I thank the Schiermonnikoog Natural Park staff for the assistance and shelter during the field work on the beautiful island.

I am also very grateful to the staff and professors of all universities of this programme (SU, LU, WU and ITC) for putting a lot of work and giving us useful knowledge, help and support throughout these 18 months. And of course, the Erasmus Mundus programme for making this a reality.

It's unbelievable, how time flies and how all of us, Gemmers, will go our own ways so soon. I hope we don't lose each other, no matter where we are! Thank you for being such a diverse and unforgettable group!
Є на світі добрії люди. ☺

Table of contents

List of figures

List of tables

1. Introduction

1.1. Research background and justification

1.1.1. Green beach

About 30% of the Dutch area is below sea level. Major cities such as Amsterdam, Rotterdam and the Hague, accommodating approximately 50% of the country's population, are located within this area. Mreover, important industrial and economic centres are also located within these areas providing for 50% of the Dutch GDP (CBS, 2011).

The majority of the coast of the Netherlands consista of coastal sand dunes or dikes built for protection from the sea. Big effort is put to manage the coastal areas (Schoeman, 2006).

Sea level rise and intensive management of the coastal areas do limit space for natural processes on the coastaline. Natural growth and development in these areas is a rarity. Therefore, the formation and development of a green beach along the Dutch coast represents an interesting subject for study.

Green beach is the phenomenon of vegetation succession on a sandy beach, where in the relative shelter of embryonic dunes a rare mosaic of dune, dune slack and salt-marsh vegetation develops. Species with radically different requirements are growing side by side forming a unique landscape (Edmondson et al., 2001). Another distinctive feature of the green beaches is their dynamic development. Without the impact of severe storms vegetation can develop very rapidly, leading to drastic vegetation changes in only a few years time. On the other hand one big storm can wash away or bury the green beach with sand.

Some studies refer to green beaches as a transitional phase of succession that passes into salt marsh or a dune slack community (Edmondson et al., 2001, Koppenaal, 2007).

The green beach community has not been discussed much in the literature, even though the phenomenon has been documented in the past (Allen, 1932). Several researches were dedicated to the issue in the recent past: green beach on the Sefton coast, UK (Edmondson et al., 2001, Smith, 2006), green beaches on the Frisian Islands, Netherlands (Koppenaal, 2007), green beaches of German East Frisian Islands (Petersen and Pott, 2005); however, more attention is needed for a better understanding of the dynamics and the driving forces of green beaches.

Studying this phenomenon may bring interesting findings. Even though the majority of species reported growing on green beaches may not be unique (note that several Red List (Tamis et al., 2004) species do occur) it is the dynamics and the patternt of vegetation growth that draws attention.

Jansen (2010, personal communication) reported nesting and breeding of migratory birds such as *Vanellus vanellus* on the beach, which was never observed before. Some birds have a preference for green beach situations for nesting (*Emberiza schoeniclus*, *Anthus pratensis*) and wintering (*Lymnocryptes minimus*, *Carduelis flavirostris*) and due to its uniqueness green beaches have a contribution to biodiversity in coastal environments.

Another aspect to consider is the possible impact of vegetation growth on sand accretion, important in the coastal area. The presence of vegetation may to some extent stabilise active erosion processes (Edmondson et al., 2001). However, taking into accout the dynamics of the green beach phenomemon it may have a very small impact.

A challenge for management arises when a balance between beach diversity protection and recreational activities needs to be met. Consequently, a better understanding of the dynamics and driving forces of a green beach is required.

1.1.2. Green beach formation

Some specific factors are required for a green beach formation to initiate. Important is the width of the shore providing more stable conditions, with some protection from embryonic dunes or beach ridges. Absence of disturbances such as driving vehicles (Birkdale beach, Edmondson et al., 2001), or no severe storms for a long enough period of time (Koppenaal, 2007) can trigger or accelerate the vegetation development. Initially, the microorganisms set in followed by pioneer species such as *Puccinellia maritima* (Edmondson et al., 2001, Koppenaal, 2007). This leads to enhancement of sand accretion and further vegetation succession.

A green beach can change drastically within a few years time.

Figure 1.1 below shows a simplified representation of green beach components.

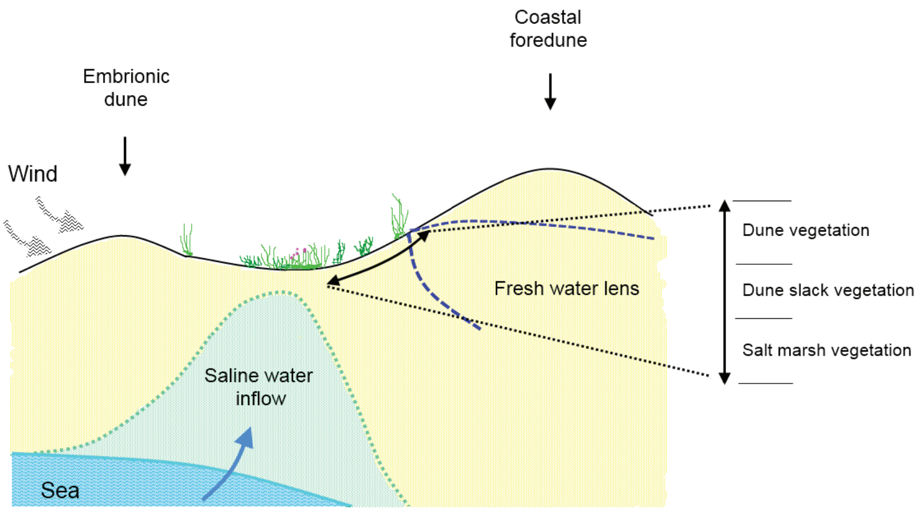


Figure 1.1 Schematic representation of a green beach and some abiotic factors influencing the vegetation composition

1.1.3. Previous green beach studies

Previous green beach studies have focused on the following: vegetation and animal classification (Smith, 2006, Smith, 2001), using remote sensing (time series of aerial photos) to detect changes and development of a green beach (Edmondson et al., 2001), finding relation between abiotic factors and vegetation types of a green beach (Koppenaal, 2007).

Smith (2006) and Edmondson et al (2001) are looking at the same study area – Birkdale green beach in UK. Both authors observe a fast exponential development and increase in species richness as well as changes in vegetation types. These developments occurred within 5-10 years. In these researches the relation between vegetation distribution and abiotic conditions were scarcely studied.

Koppenaal (2007) aimed at detecting the driving factors of vegetation succession as well as determining the vegetation types occurring. In her work she looked at elevation according to mean high tide (as a proxy for inundation frequency and duration), ground water electrical conductivity (a proxy for salinity), and presence of microorganisms as factors influencing vegetation distribution. The salinity and elevation have shown most relation to vegetation distribution. Koppenaal also found a relation between distribution of plants/vegetation types and fresh water seepage. However, fresh water seepage is only present where the beach borders sufficiently big dune areas, which contain fresh water lens underneath. Clay content in the soil and nutrient fixing microorganisms had some relation with the vegetation distribution. However, these factors were rarely found.

The pattern of species distribution along the elevation gradient is discussed in the study; the green beach is therefore divided into three types. First type consists of salt-marsh species hosted by areas with low elevation and high salinity. Second type contains some dune slack species that are distributed in areas with higher elevation and lower grown water salinity (occasionally, some fresh ground water). Finally, a transitional phase, hosting species of salt-marsh and dune slack vegetation is attributed to the third type.

Some correlations between conductivity measurements, elevation and vegetation distribution were detected, however the explanatory value of those was found weak.

Researchers suggest that depending on the conditions, a green beach can develop into salt marsh, dune slack, and dune or disappear again.

1.2. Research overview

Green beach is a very dynamic environment. It is difficult to predict how a green beach will develop in time. It can evolve into a dune slack, a salt marsh, or even a dune area, but it can also disappear in the course of one winter season with heavy storms.

According to a previous study (Koppenaal, 2007) briefly described above the following parameters are essential in vegetation succession:

- Inundation duration and frequency
- Water quality (saline/fresh) depending on seepage or position in the terrain
- Clay content

In this study elevation will be used as a proxy for the combined effect of inundation duration and frequency. To account for more abiotic influences mentioned (ground water salinity/fresh water seepage and other) some additional indirect factors will be derived from a digital elevation model (DEM) and tested. These are cost-distance to the sea, distance to the sea inlet, distance to fresh water seepage and slope in elevation (further referred to as DEM-derived variables). For more detailed explanation of these DEM-derived variables see chapter 3.3.

From the methodological part – the application of high resolution DEM derived from Light Detection And Ranging (LiDAR) data representing the topography of the study area will be used for obtaining all the variables in this study and for further statistical modelling. Can a detailed LiDAR DEM be used for modelling? Is its resolution fine enough for explaining the small scale variation in the green beach vegetation?

Some attention will be focused on the impact of raster cell resolution on the modelling results. It will be tested whether the increase of cell size improves the model performance or whether there is no significant effect.
Using DEM derived variables to describe variation in the vegetation is not new in predictive modelling, however, at this fine scale (centimetre elevation scale) this kind of modelling hasn’t been tried out.
And finally it will be tested whether retrospective modelling can be performed. The relations between dependant and explanatory variables derived for 2010 data will be applied for the data of 2006, the output will be assessed.

1.3. Research objectives

Overall objective:
To model the relation between elevation, other DEM-derived factors and plant distribution on a green beach.
Specific objectives:

- Application of LiDAR derived elevation data for extracting the predictor variables, they are elevation and several distance functions (see further chapters).
- Testing which variables derived have the prevailing influence on vegetation composition of a green beach and defining the empirical relation between plant species distribution and the DEM –derived factors.
- Testing the influence of the cell size on the outcome of modelling
- Testing the model using historical data of 2006

Research questions
1) Can LiDAR DEM-derived variables be used for modelling plant distribution on the green beach?
2) What relation is there between the DEM predictors and vegetation distribution, are these enough to explain variation in vegetation?
3) What is the influence of changing the cell size of the DEM-derived factor maps on the modelling output?
4) Does the relation remain the same over time; can the same empirical equations be used for both present and retrospective modelling?

1.4. Hypotheses

During the research the following hypotheses will be tested:
Hypothesis 1
H_0: There is no significant relationship between the plants distribution and the DEM-derived predictor variables.

H_1: There is significant relationship between the plants distribution and the DEM-derived predictor variables.

Hypothesis 2

H_0: The decrease in raster cell size of the predictors does not impact the accuracy of the output.

H_1: The decrease in raster cell size of the predictors has significant influence on the accuracy of the output.

Hypothesis 3

H_0: The empirical relation between topography and vegetation composition does not remain constant within the period from 2006 and 2010.

H_1: The empirical relation between topography and vegetation composition remains constant within the period from 2006 and 2010.

1.5. Species distribution models

1.5.1. Short classification

The number of models for spatial prediction of species distribution is increasing rapidly (Hegel et al., 2010). Most of these models have quantification of species-environment relationship as their underlying principle and are static and probabilistic (Guisan and Thuiller, 2005). Although the number of statistical tools increases it is quite important not to forget the underlying ecological knowledge while applying sophisticated statistical techniques (Austin, 2002).

Austin (2002) identifies three components in a statistical modelling framework: ecological model, data model and statistical model. They combine the ecological theory or knowledge as a base for the study; provide guidelines on which way and what data needs to be collected, which statistical method, error function and significance tests need to be used.

Nature is too complex and modelling it uses specific assumptions to simplify it. These assumptions can impact the output of spatial modelling and need to be mentioned here. Guisan et al. (2000) have reviewed and classified the predictive habitat models according to a few of these assumptions. Some considerations are given below.

1) Generality, reality and precision of the model. Theoretically a given model may only incorporate two of these factors leaving the third one out. According to this the models are divided into: a) analytical, designed to give precise and general prediction; b) process models, revealing causal relationships; and c) empirical models, that use statistical relations to make predictions rather than ecological theory or cause-effect relationships of the variables. In species distribution

modelling mostly empirical statistical models are used due to relative simplicity of use.

2) Models that use different predictor types or types of ecological gradient: a) resource gradient – relates directly to organism's consumption of energy or matter; b) direct gradients having direct physiological importance apart from consumption; and c) indirect gradients – being connected to the species indirectly (Austin, 2002). Resource or direct gradients are effecting on a large scale, whereas indirect on smaller scales.
3) Assumptions about environmental niche: fundamental versus realised niche. Describing the species occupying all theoretically suitable habitats, or else only part of it due to interactions with other species (Guisan and Thuiller, 2005). Statistical models are often simplified and only quantify realised niche based on field observations (Austin, 2002, Guisan and Zimmermann, 2000, Hegel et al., 2010).
4) Equilibrium/non-equilibrium assumption. Models are divided in two groups. Those that represent reality in a static pseudo-equilibrium way (assuming none or slow change in time in the system); and those that represent the dynamics of the system. However, the statistic models used presently do not incorporate dynamic elements, which is against the understanding of ecology.

Some considerations are also raised about whether to use individual species or communities for modelling. Individualistic approach is believed to be closer to reality as compared to arbitrary classifications (Guisan and Zimmermann, 2000).

Reality is always too complex to be modelled, hence some simplifications are made. For example, most species distribution models assume either equilibrium between species and predictors or represent the realised niche. These assumptions are not completely true to reality; rather, they are dictated by the statistical approach.

In this study field observations will be used to derive the relationship between the predictors and vegetation distribution. This implies that an empirical model is used to quantify the realised environmental niche.

The scale of the study is quite small, which implies that using indirect predictors is more appropriate. This is the case with DEM-derived predictors; they do not influence the vegetation distribution directly, but are used as proxies for other environmental factors. A static model is used, assuming no big changes in the relationship between the predictors and the plant distribution.

1.5.2. Statistical models

Statistical models are mostly variable-specific. The choice of a statistical approach depends on the probability distribution of the response variables (Guisan and

Zimmermann, 2000). This distribution needs to be known prior to selection of a particular model.

Nowadays there is a rich variety of techniques for statistical modelling. These may be regression models, classification trees, environmental envelopes, neutral networks and other. In this work a type of generalised linear model (GLM) will be used. GLMs have been used widely, they are simple to implement, and are good to use when the probability distribution of the response variable is from the exponential family, but not necessarily normal (Campbel, 1989). In our case the response variable data has binomial distribution, having two possible values – presence and absence.

1.5.3. Generalised linear models

GML consists of three components: 1) random component representing the response variable in the equation (*Y*), 2) systematic component representing the predictors (X_i) and 3) the link function linking the first two components together (Campbel, 1989, Guisan et al., 2002):

$$g(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ \dots \qquad (1)$$

Where $g(\mu)$ is the link function, β_i - coefficients to be estimated.

Binary response variables (eg. dead/alive, present/absent) are quite common. For this kind of distribution logistic regression model (Hosmer and Lemeshow, 2004) is commonly used where a logit link function is used $g(\mu) = \ \log\left[\frac{\mu}{1-\mu}\right]$.

The logistic model is:

$$\pi(x) = \ \frac{e^{\beta_0 + \beta_1 x_1 + \cdots}}{1 + e^{\beta_0 + \beta_1 x_1 + \cdots}} \qquad (2)$$

where $\pi(x)$ is the probability that the response variable equals to 1 for given predictors, β_i - coefficients to be estimated.

The adequacy of the logistic regression model is checked by estimating goodness-of-fit statistics by comparing predicted against observed values - Hosmer-Lemeshow statistics (Campbel, 1989, Hosmer and Lemeshow, 2004).

1.6. General approach

Figure 1.3 illustrates a simplified workflow of this study.

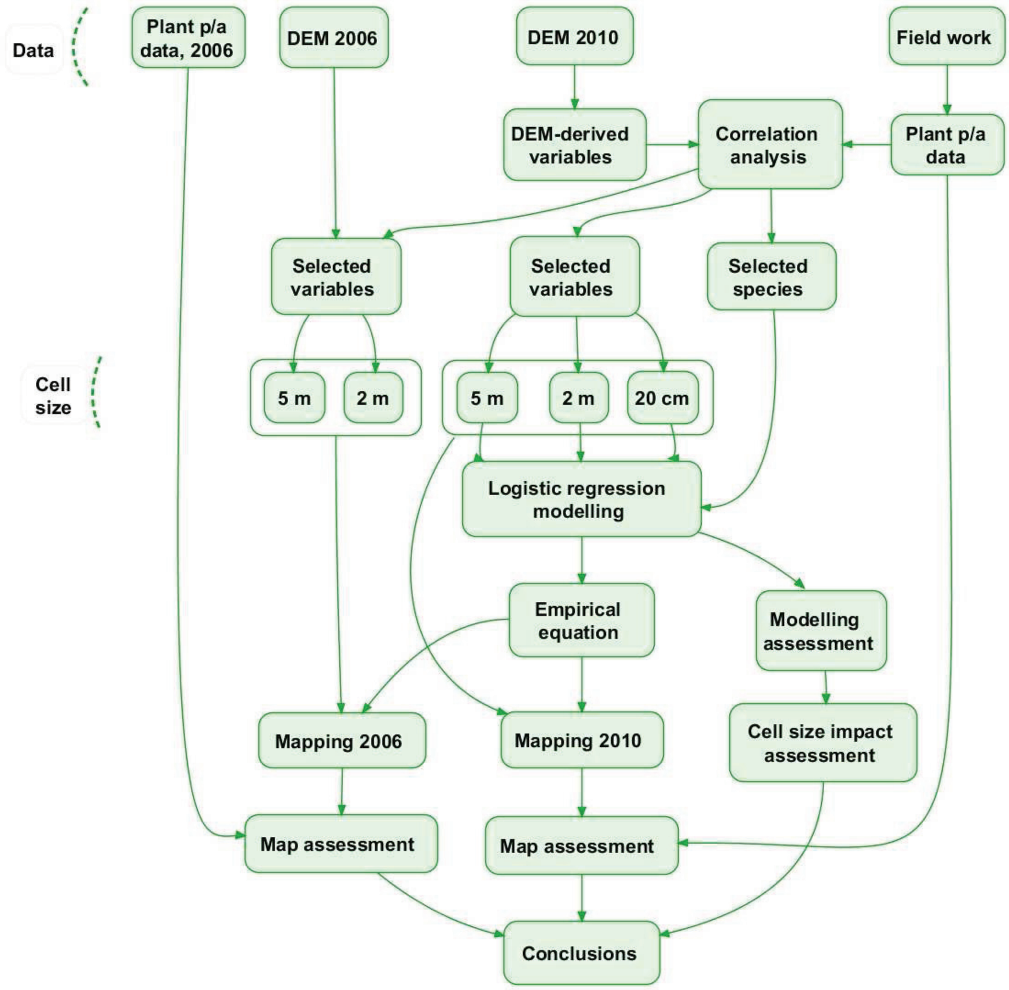


Figure 1.2 Simplified workflow of the study.

P/a stands for presence/absence

2. Study area

2.1. General information

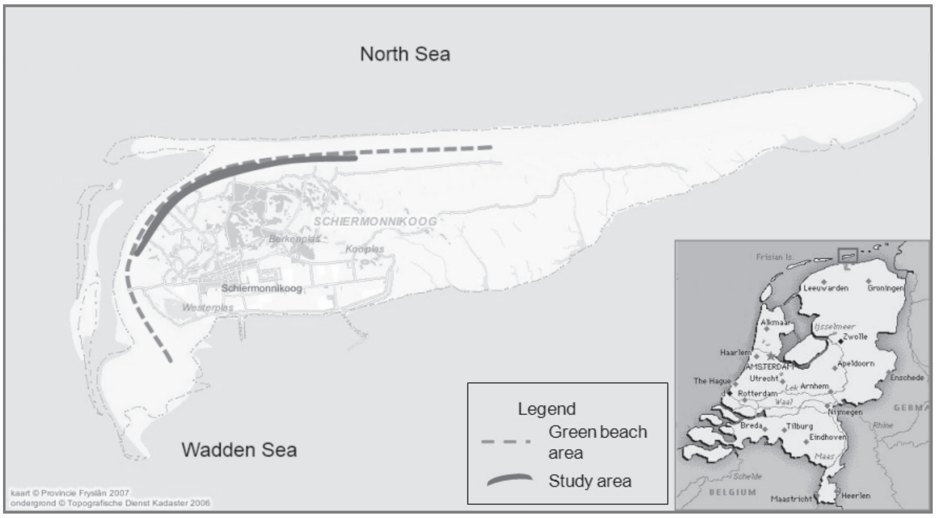


Figure 2.1 Schiermonnikoog Island; location of the green beach and the area of study are indicated (adapted from Frisian Province map, 2007).

Schiermonnikoog (Figure 2.1) is one of the barrier islands on the border between the Wadden Sea and the North Sea (figure 2). The area of the island is about 40 km^2, being about 16 km long and up to 4 km wide. The island hosts one village with a permanent population of 941 inhabitants (CBS, 2009).

Schiermonnikoog is the site of the Netherlands' first national park. Every year this place attracts up to 300 000 tourists, many of them staying for one day only (up to 4000 per day in July and August) (Wikipedia, 2010).

Tidal and wind interactions, as well as sea currents cause the island to slowly sift to the south-east. In 1250 it lay about 2 km to the north of its present position, and had a very different shape (National park Schiermonnikoog, 2011).

Although small in area, Schiermonnikoog has a variety of landscapes. Thanks to this, the island has an abundant population of animals and plants (National park Schiermonnikoog, 2011).

The island consists mainly of dune and salt-marsh areas, with elevation of up to 20 m and 1-2 m respectively (Beukeboom, 1976). The dune area is built-up with dunes oriented in south-east direction caused by prevailing westerly winds and sand supply from north-west. The southern and eastern area of the island is a salt-marsh – a flat area flooded regularly with tides and heavy storms. The southern part is a polder – land transformed from salt-marsh into agricultural area.

Figure 2.2 View of Schiermonnikoog from the air (Photo: Samuel Bekx)

In this research a part of the green beach area of Schiermonnikoog is studied. The study area is situated on the northern beach of the island, at the North Sea side (indicated in figure 2.1).

The coastline of the Netherlands has a reference system with beach poles placed at every 1 km (Hiller and Roelse, 1995). These beach poles were used by Rijkswaterstaat (Dutch government agency responsible for road and water infrastructure and protection against flooding) for coastal monitoring, and precise coordinates of the pole locations are known, although these poles are no longer in used and not maintained. The beach poles were used in this study as a reference for geodetic surveying. Figure 3.3 shows the locations of the beach poles used in this study.

2.2. Climate and Hydrology

The Wadden Islands have a humid, temperate, maritime climate. Mean annual precipitation is 500-1000 mm. There is no real dry period, but the months with most precipitation are September through December (see figure 2.3). The mean temperature varies from 2 °C in winter to 17 °C in summer. The peak in evaporation occurs in the summer months and the precipitation peak is in autumn and winter leading to some surplus in precipitation in these seasons.

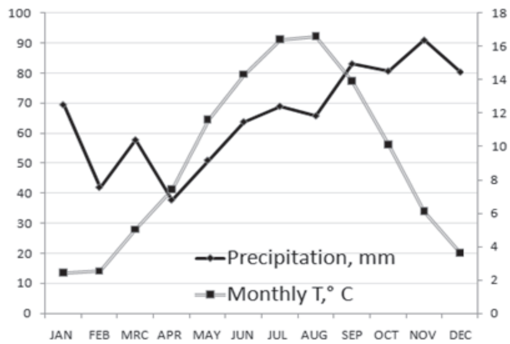


Figure 2.3 Mean monthly precipitation and temperature 1971-2000 for the Schiermonnikoog weather station (Adapted from KNMI, 2010).

Hydrology is an important feature on the island influencing the vegetation distribution including the vegetation in the green beach.

All Wadden Islands have similar hydrology. Generally a Wadden island consist of a dune area and sometimes a salt marsh or polder – land reclaimed from the salt marsh (Beukeboom, 1976). Salt water has higher density than fresh water. Precipitation surplus infiltrated in the soil brings fresh groundwater table up and creates pressure pushing salt water down. This way a fresh water lens is formed under the dune system (Figure 2.4).

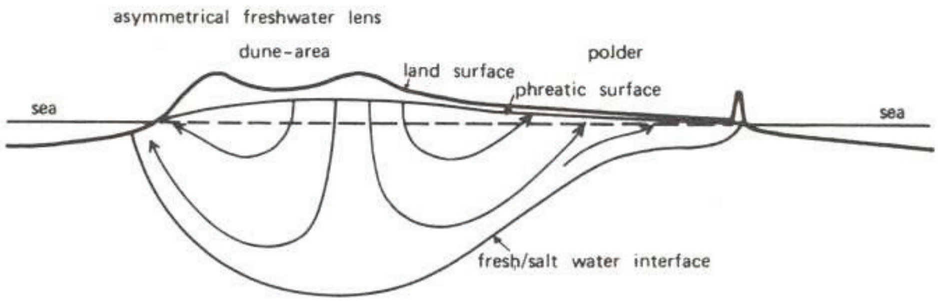


Figure 2.4 An asymmetrical freshwater lens, occurring under a dune area and polder. This is the case on Schiermonnikoog island (Adapted from Beukeboom, 1976).

At the break of slope at the edge of the dune system from the seashore side the freshwater table approaches the ground surface. This results in some fresh water seepage at the beach.

During autumn and winter time there is precipitation surplus and the ground water table rises. At these periods some areas at the beach and in between the dunes are below the water table level resulting in some standing fresh water throughout the season.

2.3. Shore erosion and sedimentation

Marine and Aeolian erosion and sedimentation are connected to the nature of the Wadden Islands making them slowly shift, as was briefly mentioned before. Without these processes they would not exist.

According to a case study of erosion on the Dutch Frisian Islands (Schoeman, 2006) the physical processes in the Wadden Sea area are influenced by the tide and waves. The alongshore sediment transport is induced by waves, whereas cross-shore transport happens through tidal inlets and is induced by tide. The island shores have erosion and accretion in different places: erosion of the western part of the island and sedimentation in the eastern area; this causes the islands shift eastwards.

Some causes of accelerated erosion and sedimentation are sea level rise, land subsidence due to gas mining, storms relocating sand, and sand waves - the sand volumes that are moved along the shore.

In this study the vegetation distribution is assumed to be indirectly related to topographic forms of the green beach area (elevation, distance to the sea, distance to fresh water seepage location and other). This relation is also assumed to be relatively constant throughout the period from 2006 to 2010. Consequently, change in vegetation cover, if any, would be caused by a change in the topography of the beach. Some initial insight is needed on the processes of erosion and accretion of the shore of Schiermonnikoog, more specifically the area of study.

Schoeman (2006) states that the erosion on Schiermonnikoog is not significant and the island is relatively stable, however some changes do occur.

To illustrate the change in the terrain a simple deduction of DEMs was made: DEM 2010 minus DEM 2006. The result shows the change in the elevation of the study area (Figure 2.5). Simply observing the digital elevation models from the two years it’s possible to notice changes in the shoreline: the sand bank on the west has moved slightly to the east, the beach has become less wide at one part and slightly increased on the eastern part. Overall, the embryonic dunes in the study area are more developed in 2010.

The area further away from the sea has gained some elevation, as a contrast to up to 2.5 m loss of elevation at the shore line. The area on the western part of the beach has suffered some loss of sand, but generally there has been an increase of elevation from about 25 to 50 cm along the dunes. The embryonic dunes in the centre of the study area have grown for 1-1.5 m. This increase presumably leads to the area behind these dunes being more protected from tidal inundation, which does not seem to occur there anymore, and the storms.

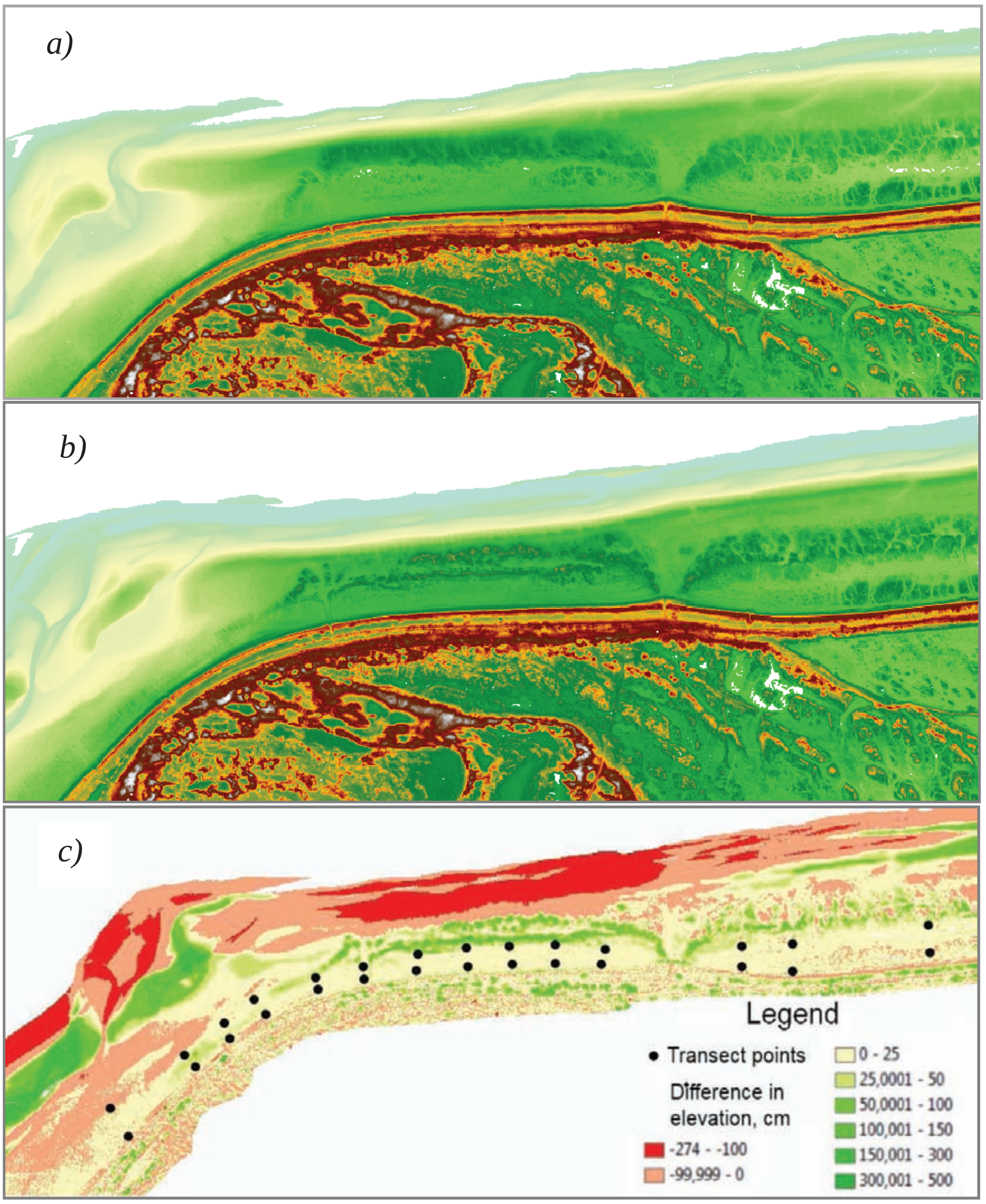


Figure 2.5 DEMs of the study area from a) 2006, b) 2010. Map c) shows the increase/decrease in elevation between 2006 and 2010. The red colour represents loss in elevation, yellow to green represent gain. The dots mark endpoints of sampling transects. Note big loss of elevation at the shore line and some increase of embryonic dunes and areas along the dune system, the sand being shifted inwards the island.

3. Materials and Methods

3.1. Field work

The field work aimed at sampling vegetation species occurring in the study area. The line-point intercept method was used with a systematic sampling design: the vegetation was sampled along transect lines with a constant interval. The field work was implemented in three stages: reconnaissance, surveying, vegetation sampling.

3.1.1. Reconnaissance

During reconnaissance the potential study area was observed. Some broad general patterns of vegetation growth were visually detected in the field. The widths of these patterns were approximately measured by steps and a conclusion was made about the possible interval between sampling transects. The extent of the study area was finally determined visually: the green beach between beach poles 3 and 8. The area eastwards of beach pole 8 was considered to be out of the scope of the study since it primarily hosted dune vegetation. As for area to the south-west of pole 3, the vegetation cover there was similar to the area between poles 3 and 4. For efficiency reasons and time constrains sampling the area south of beach pole 3 was excluded.

The area between poles 3 and 4 and around pole 5 is regularly inundated with high tide. The vegetation observed there is very similar to salt marsh vegetation with increase of brackish and fresh water species close to the foot of the foredunes. The area around beach pole 5 is higher in elevation hosting smaller number of halophytic plants (e.g. plants like *Salicornia europaea* are not found here).

Around pole number 6 the green beach area is situated on even higher plain protected by some high (about 1,5 m) embryonic dunes to the north, sheltering the green beach from direct influence of the sea; this area does not seem to get inundated by tides too often. Some plants requiring fresh water are regularly found (*Mentha aquatica*).

The area in between poles 7 and 8 hosts an interesting situation (see figure 3.1). There is a sea inlet between the embryonic dune ridge and the main dune ridge; this gets inundated with high tides, saline standing water was found there during the field work period. The vegetation gradient in this area runs perpendicular to the inlet (mainly with north-south and east-west orientation), with saline vegetation passing into brackish further away and into some rare fresh water plants growing near the dune ridge or on small dunes in the area.

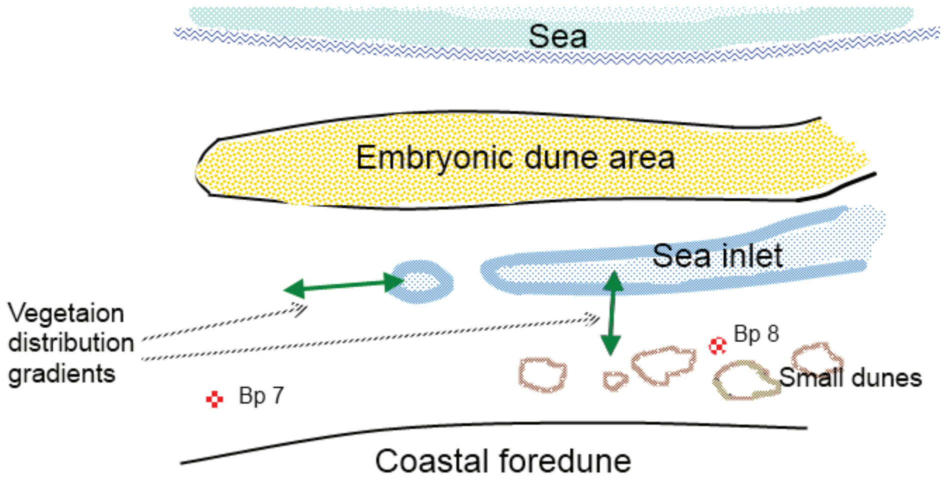


Figure 3.1 Sketch of the landscape features between poles 7 and 8. 'Bp' stands for beach pole.

The gradient occurring along the length of the island was observed in the field and estimated to be close to 250 m, the sampling transects were placed using this interval, although, the transects near areas of big human influence were taken out.

3.1.2. Geodetic surveying

The systematic sampling design was chosen partially because it enables to achieve higher positioning accuracy compared to random sampling. The positioning accuracy is quite important due to small scale of variation and small patch size of the vegetation cover in a green beach. The surveying of the positions of sampling transects took place after the reconnaissance. The transect end points were fixed temporarily in the field and GPS readings were taken to make sure the points are easy to be found in future during sampling. The transect lines were surveyed using an optical theodolite Wild Heerbrugg T5. Figure 3.2 shows an example surveying traverse. Each starting point of a transect was used as surveying point and the angles and distances were measured to parallel transects and the endpoint of each transect. As a coordinate reference the beach poles (marked *Hp* on figure) with known coordinates were included in the surveying network. The outcomes of the surveying – angle and distance measurements together with reference coordinates were used to calculate the coordinates of end points of each transect.

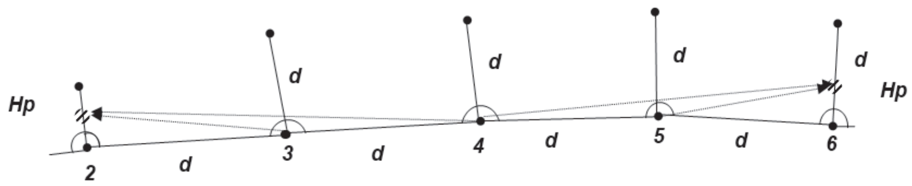


Figure 3.2 Example of part of a theodolite traverse. *Hp* – beach poles with known coordinates, the transect numbers are shown, *d* – measured distance, the angle measurements are shown with direction arrows.

3.1.3. Vegetation sampling

The transect lines were placed at a regular distance of 250 m from one another perpendicular to the ridge of coastal foredunes (see figure 3.3). The first transect was placed between beach poles 3 and 4 with the last transect being at beach pole 8. The total number of transects sampled is 14.

Line-point intercept method (Herrick et al., 2005), a variation of point intercept method, was used for presence/absence sampling of plant species. In this method the vegetation species are sampled along a transect with a regular interval. At each interval point a pin (about 50 cm long) was placed on the ground and all the leaves/plant species touching the pin were recorded.

The nomenclature of species followed Van der Meijden (2005).

In the field the measurements were taken according to the procedure described by Herrick et al (2005). A tape measure is placed along the transect. Starting from zero point of the tape the measurer goes from left to right to reach the end of the measuring tape. The starting measurement is taken at 20 cm mark and then every 20 cm for the whole length of the transect. The vegetation is sampled using a pin of about 1 m long that is placed vertically onto the ground at each interval. All the species touching the pin are recorded. Clearly a 'hit' is a true presence record and all 'no hits' were considered true absence data.

To decide on the scale of sampling a small interval of 10 cm was tested first. It became apparent that a 10 cm interval was too small (compared to the scale of the patterns in the field) and not time-efficient. Consequently, the working interval taken was 20 cm. Although this interval was still producing a big number of sampling points, it was considered appropriate. The sample was to be split in two sets: for model training and for model validation; that is why excess in the number of sampling points was needed.

3.2. Data

3.2.1. Point coordinates calculation

The outcome of the geodetic surveying was derivation of distance and angle measurements between transect points. These were used for coordinate calculations.

A series of geodetic techniques were used. The calculations were done using a geodetic calculator (Gribok, 2007).

After the transect endpoints' coordinates were derived (see section 3.1.2) the coordinates of the sample points were calculated using the distance from starting point and the direction angle of the transect.

This way the coordinates of the points were derived with higher precision than available GPS receiver. The estimates of the positioning accuracy for calculated coordinates varied from 50 cm to just over 1 m as compared to above 2 m GPS receiver accuracy.

3.2.2. Sampled data

The sample points with the species hits data were the main outcome of the field work. As in the field only the hits of plants were recorded, later the data was transformed into a matrix containing all species as header and point coordinates in each row; the present species were marked as 1, absent as 0.

The total number of sample points in 14 transects is 7387. The number of species recorded is 54; the list of species is given in appendix A. For modelling the plants having a low presence (less than 20) and the plants that were questionable (some species were difficult to distinguish without flowers and could have been recorded erroneously) were omitted.

The dataset was divided into a training and a validation set for further modelling of the species distribution. The division rate was 60/40 respectively, the data were split randomly.

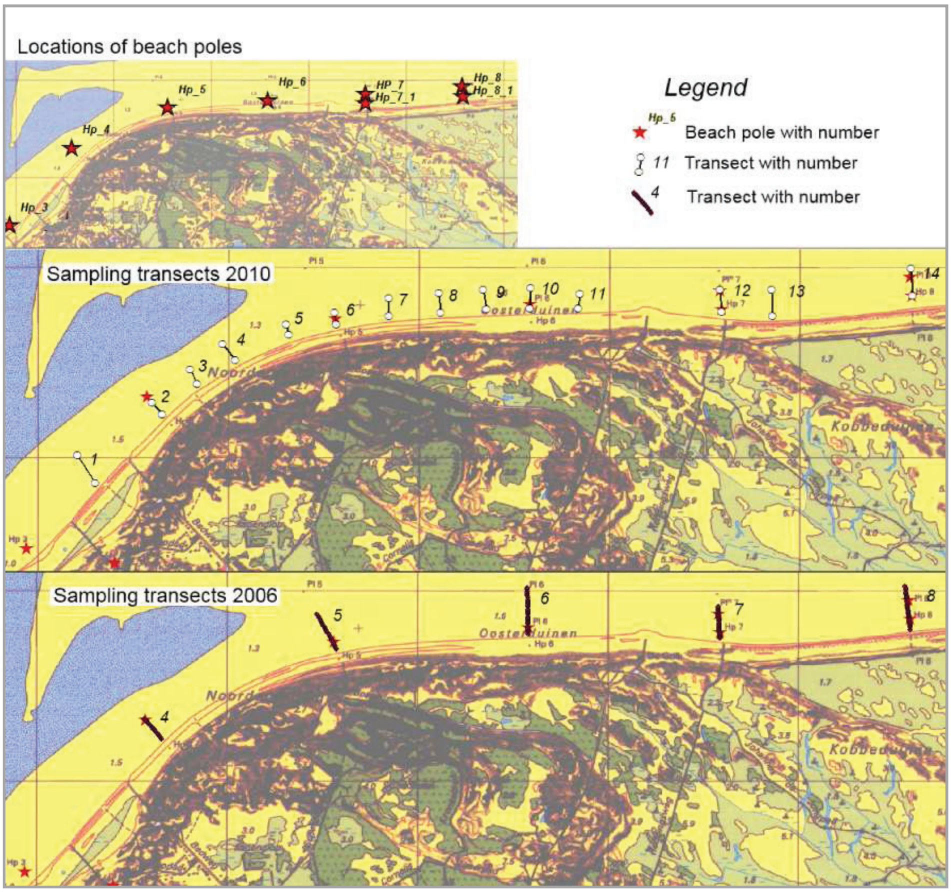


Figure 3.3 Locations of beach poles, sampling transects for 2010 and validation data for 2006.

3.2.3. LiDAR DEM

Laser altimetry (method used by LiDAR) became a well accepted approach for terrain data collection in the recent past (Flood, 2001). This system works similarly to radar: it uses laser scanning to derive a cloud of points with known elevations and known coordinates with relatively high accuracy. Among applications in various areas the terrain data collection and DEM generation is becoming most frequent (Liu, 2008). As the sensor measures the distance to the closest surfaces, the points generated not always represent ground surface (vegetation canopy, objects etc.), a DEM needs to be derived from the digital surface model. Filter algorithms are used to derive true elevation data. The height value of a pixel is calculated from the surrounding laser points of the filtered base file. This technique is called a weighted average interpolation (Rijkswaterstaat, 2010b).
In this study DEMs for the years 2006 and 2010 are used.

The original cell size of the DEMs used in this study is 5x5m. The value of a 5x5 meter grid cell is calculated from multiple laser points (the number depends on the point density of the base file). This reduces the influence of measurement noise and outliers, however, there is a slight degree of flattening (Rijkswaterstaat, 2010b).
The accuracy (standard deviation) of the height value is less than 5 cm, the horizontal accuracy is up to 50 cm (Rijkswaterstaat, 2010b).

3.2.4. Historical validation data

Validation dataset for 2006 was kindly provided by Elske Koppenaal who collected field data in 2006 on the green beach of Schiermonnikoog (Koppenaal, 2007). In the 2006 field work sampling was done using 2x2 m plots where the full floristic composition was recorded and the cover of each species visually estimated. The 2x2m plots were placed 10 m apart along the sampling transects that followed the beach pole locations. See figure 3.3 for locations of the sampling transects.
The sampling technique and scale of the 2006 and 2010 datasets do not match.
To solve this problem the vegetation cover data were transformed into presence/absence. The cell size was assumed to be 2x2 m and all the species recorded per plot were marked as 1 (present), the rest of the species – 0.

3.2.5. Raster cell size

As reported before (Guisan et al., 2007) the choice of cell size of the environmental layers that are used in modelling may impact the predicted output. Guisan et al. have used 10 times coarsening of the data to investigate the change in predictions. The result showed no severe changes, however, there was an unequal effect across regions and species types.
In this study it is proposed to use interpolation techniques for increasing the spatial resolution of the DEM.
As mentioned, the original cell size of the DEMs is 5x5m. As the scale of vegetation pattern on the green beach may be quite small, a lot of change might occur within 5 square meters. Thus the raster resolution was considered too coarse. To deal with this issue and test whether any improvement of the predictions occurs it was decided to increase the raster resolution. Three cell sizes were chosen to be tested for the two periods: 5 m – the original cell size, 2 m – the cell size correspondent to the validation data of 2006, and 20 cm – cell size to be used for modelling 2010 situation only. The 20 cm cell size was chosen to match the sampling distance used during the field work. These resolutions were used for extraction of the training data and for mapping the output.

3.3. Methods

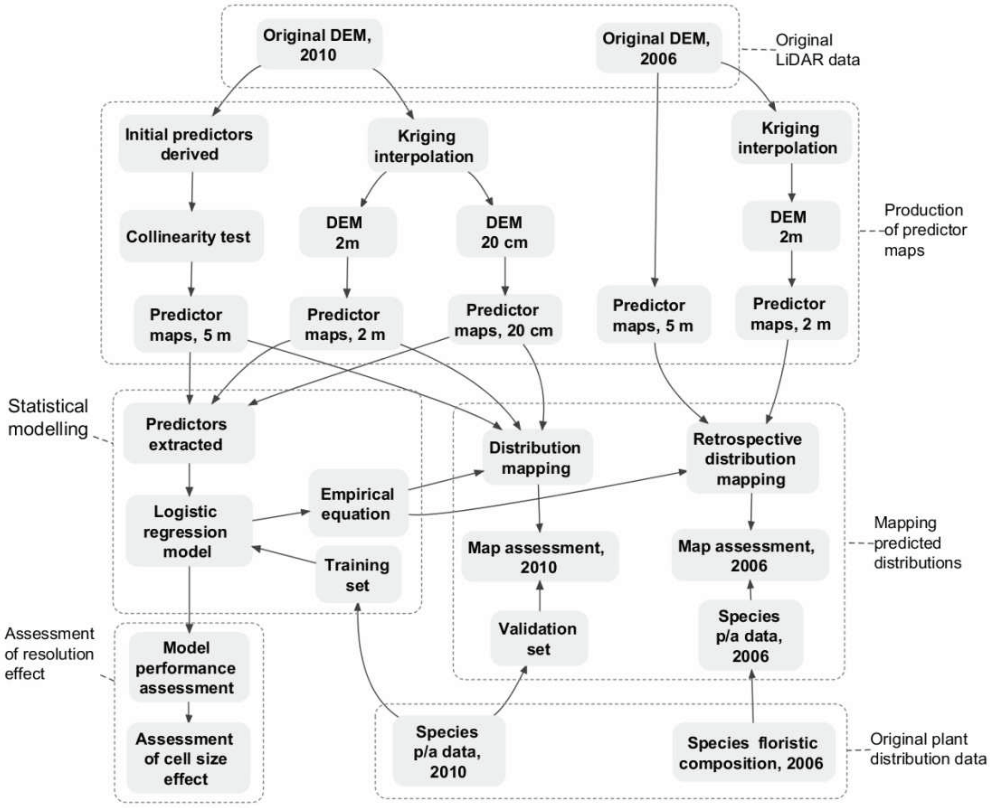


Figure 3.4 Methodology overview. (P/a – presence absence)

Figure 3.4 above shows overview of the methodology adopted in this study.

3.3.1. Explanatory variables

For modelling the plant distributions with logistic regression explanatory variables need to be chosen carefully. In case of this study where the scale of the phenomenon looked at is quite small, the general broader scale factors like light, temperature, precipitation, altitude remain constant across the whole area. Factors varying on a more local scale are defining the vegetation patterns, for example micro-topography of the area. One of the objectives of this study is the use of a detailed digital elevation model for deriving the factors influencing the vegetation pattern. Thus the variables chosen here were merely obtained from the DEM with some prior studies of the literature.

In a coastal environment one of the prevailing impacts is of course from the sea, coming in several forms: tidal inflow of saline water or presence of salt spray, saline

groundwater, destructive power of storms and exposure to strong winds and sand drifts. Presence of fresh groundwater is also very influential. The following factors were considered important in the study area: inundation frequency/duration, availability of fresh water (as described by Koppenaal, 2007) and exposure to the sea influences; these factors were also chosen since they can be modelled using a DEM.
Below each variable is described in more detail; the overview of the production of the variables can be seen in figure 3.6.

Elevation

Elevation in this study is used as a proxy variable for inundation frequency/duration, exposure of the vegetation to inflow of saline water. Elevation has been reported key in vegetation distribution on a green beach (Koppenaal, 2007).

Cost distance to the sea

This variable was used as a proxy for exposure to the sea, e.g. storms. Cost distance function is used to account for coastal relief (embryonic dunes, sand banks), that has some protective effect. The source location is taken to be the open sea area determined from the DEM.

Distance to sea water inlets

This variable is somewhat correlated to elevation, and perhaps to tidal inundation. The factors that it describes are connected to closeness of the sea water, salinity of ground water and salt spray.

This variable was derived by defining the elevation level that gets inundated with high tide and calculating a distance function to the areas below the elevation threshold. The threshold was set at 150 cm + NAP (Dutch ordnance level, Rijkswaterstaat) taken from tidal data available (Rijkswaterstaat, 2010a), corresponding to the high tide level on the 10th of September 2010.

Distance to fresh water seepage

To create this surface it was necessary to derive the location of the fresh water lens under the island and to define the break in the slope at the verge of the dune area and the beach, since that is where the seepage occurs (see figure 2.3).

The ground water table elevation was derived from the literature (Beukeboom, 1976). As can be seen in figure 3.5 the freshwater lens only extends under the dune system of the island and the polder, the salt marsh area has saline ground water.

The isolines of the water table elevation above the sea level were digitised and interpolated into a surface.

Slope was calculated for the study area and a line shape was derived at the break of the slope. It was selected where a rapid change in the slope occurred (approx. from 30° to 70°). The freshwater lens surface and the break in the slope were overlaid to define where the water seepage may occur (given the presence of fresh ground water at the location of the slope break). The potential seepage locations were derived and a distance function to these location was calculated.

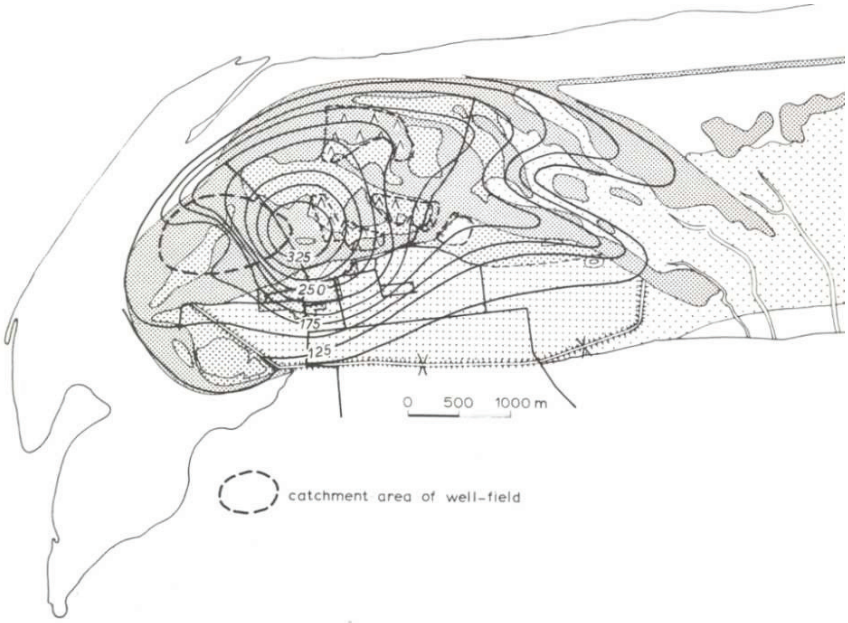


Figure 3.5 Contour lines of the capillary surface (Beukeboom, 1976). The figure shows the location of the fresh water lens in 1974 when a survey was conducted in the area. It is used in the present study because the features of the lens nowadays remain similar, except for the impact of water pumping in the well

Note that an attempt for more elaborated model of water seepage was made. The depth of fresh ground water was derived from the DEM and fresh water elevation, and the distance function was weighed by the closeness of the water to the surface. However, this variable yielded lower correlation with the presence/absence data as compared to the simpler model described above.

Slope

Slope was assumed as another proxy for topographic features of the study area. The slope itself does not influence the vegetation distribution; however, it indicates the areas of curvature (like small dunes or hummocks). Slope values would show small variations of the surface topography that result in different vegetation cover. After being tested for correlation with the presence/absence data, slope parameter yielded significant correlation, thus it was included in the further modelling.

Distance function values

It's important to mention one feature of the distance functions. The distance is calculated towards the target (e.g. fresh water seepage location), this means that as the target feature gets closer the values tend to be low or even zero, whereas moving away from the target increases the raster value.

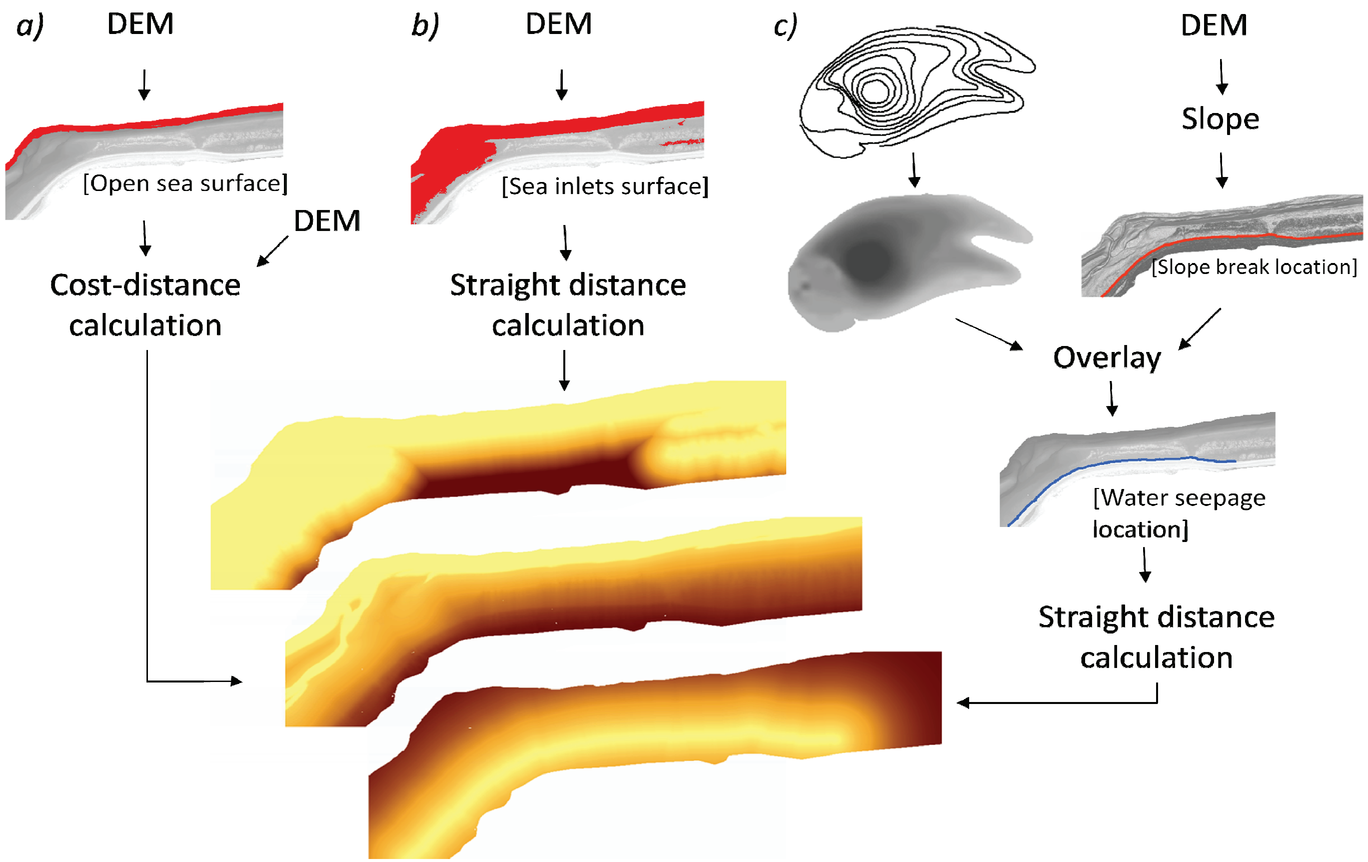


Figure 3.6 Steps of deriving the DEM variables. a) cost distance to the sea, b) distance to sea inlets, c) distance to fresh water seepage (contour lines of the fresh water lens derived from literature). For more detailed explanation refer to the text.

3.3.1.1. Multicolinearity

Multicolinearity of explanatory variables means that the variables are correlated. In logistic regression this may result in the equation coefficients being estimated with an error (Moore, 2006). Thus, firstly it is important to test the derived variables to see if they are correlated. The variable values were derived from 5 m DEM using sample point locations and a correlation table was calculated.

Table 3.1 Correlation table for predictors

	Elevation	Cost-distance to sea	Distance to sea inlet	Distance to fresh water	Slope
Elevation	1	,365	**,762**	-,343	,113
Cost-distance to sea	,365	1	,425	,338	,092
Distance to sea inlet	**,762**	,425	1	-,371	-,005
Distance to fresh water	-,343	,338	-,371	1	,000
Slope	,113	,092	-,005	,000	1

Here we can see that most variables are correlated. More attention needs to be paid to the 'Distance to sea inlet' variable because it has a high correlation with 'Elevation'. Normally correlations above 0,5 might mean multicolinearity in the data. A way to test for it is to calculate variance inflation factors (*VIF*). These are calculated by performing regression analyses for the variables one by one on the remaining variables. Next, R^2 values are obtained and *VIF* is calculated by the following formula:

$$VIF = \frac{1}{1-R^2} \tag{3}$$

For the explanatory variables here:

Table 3.2 VIF values for predictors

	Elevation	Cost-distance to sea	Distance to sea inlet	Distance to fresh water	Slope
VIF	2,54	1,94	3,03	1,84	1,05

Generally $VIF > 10$ is an indicator of multicolinearity which is equivalent of correlation of 95% (Chatterjee and Hadi, 2006). In our case the values are well below this threshold, so should imply that colinearity problem is not too serious.

3.3.2. Raster interpolation

The raster cell size was decreased from the original 5x5m to 2x2m and 20x20cm. First, the original raster was converted into a point feature class, with the centre point containing the cell value. This makes sense since the cell value of a raster is the value corresponding to cell centre (LU GIS-Centre, 2008). The extracted points were used for interpolating of elevation models of a finer resolution. The interpolation procedure used was kriging interpolation – a technique that is based on variogram and geostatistics. Kriging considers both the distance and the degree of variation between known data points when estimating values in unknown areas (Edward H. Isaaks, 1989).
The increase of cell resolution is believed to improve the accuracy of output. As the sampling had 20 cm interval, gives us 25 sampling points per 5 m cell. And if the scale of variation in the field is finer than 5 m, this variation may not be explained by predictors of this cell size. Figure 3.7 illustrates this problem. However, the exact scale of the variation in question is not known.

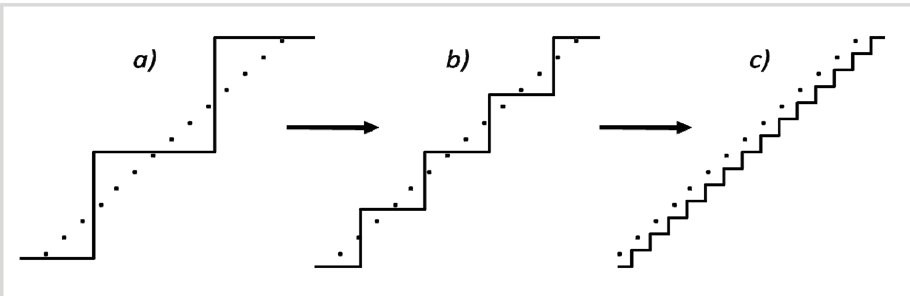


Figure 3.7 Schematic representation of profile views of DEMs of various resolution with each 'step' representing elevation of one cell: a) 5 m, b) 2 m, 3) 20 cm; the dots represent sampling points.

3.3.3. Analysis of predictor variables

The correlation between all plants presences and absences and environmental predictors was tested using two non-parametric tests: Kendall's τ (Kendall, 1938) and Spearman rank order test (Spearman, 1987). These approaches were used because they are suitable for testing correlations of binary data, such as the presence/absence data here. The correlations were investigated. The plants with correlations higher than 10% with the predictors and Red list plants were selected for further analysis.

3.3.4. Modelling plant distributions with logistic regression

The species that were selected after examining their correlations were then modelled using forward stepwise logistic regression. This approach was chosen in order to select the most suitable and significant explanatory variables for each plant, as it permits to compare the model performance as the predictors are added step by step. The coefficients for the equation 2 were derived (see section 1.5.3 for the equation).

The best model was selected from the stepwise procedure, the goodness of fit parameter of the model that compares predicted values against observed (Hosmer-Lemeshow test) and explained variance parameter (Cox & Snell R^2, Nagelkerke R^2). The statistical modelling was implemented in SPSS.

As the impact of the cell size was to be tested, modelling procedure was performed for each of selected species for 3 raster resolutions. The 3 sets of predictor rasters were used to extract the explanatory data. The training data was aggregated with increase in cell size.

3.3.5. Cell size impact assessment

After the modelling process was completed the change in the significance of goodness of fit parameter was assessed. This was done to assess effect of the changes of cell size for each selected plant.

To be able to compare changes in significance for the goodness of fit parameter between different plants, the parameters were standardised using equation 4 below:

$$g_i = \frac{GoF_i}{sum} \cdot 100\% - g_{5m}\,, \qquad (4)$$

Where *g* – is the standardised value, *i* – is the cell size, *GoF* – is the original goodness of fit significance value, *sum* – the sum of the parameters for 3 cell sizes.

The value of standardised parameter for 5 m resolution was subtracted from each standardised value to set the initial value to zero.

The values were then plotted and their behaviour assessed.

3.3.6. Mapping the predicted distributions.

The mapping of the probability distributions was implemented in ArcGIS Raster calculator using logistic regression equation (equation 2) and the coefficients derived from the statistical modelling.

As an outcome probability maps were derived. To transform these into presence/absence maps a threshold value was chosen. This cut value was used to determine at which probability it is assumed that the plant is occurring or is absent. Normally a cut value of 0,5 is used for modelling. However, logistic regression appears to be sensitive to a chosen threshold (Manel et al., 1999), especially in the case of this study, where some plants are rare and the probability of their occurrence in a given place is generally low.

For choosing suitable thresholds for each species a classification plot showing the actual and predicted values of the dichotomous dependent variable were used. An example of the classification plot is shown in figure 3.8. Here we can see how the actual probabilities are grouped by the model. If they are clustered separately on the plot, this implies that the model performance is good (SPSS, 2010).

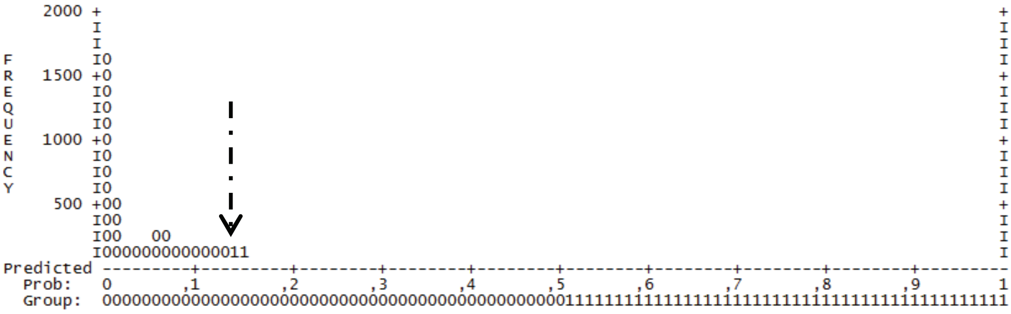


Figure 3.8 Example of a classification plot for a *Salix repens* model used for estimating probability cut-off; 0s represent cases of absence, and 1s represent cases of presence. Arrow indicates the possible probability threshold to be used.

The threshold of predicted probability is chosen at the point of change in the actual probabilities in the plot.

The overall procedure is repeated for 2010 (5m, 2m and 20 cm cell size) and 2006 (5m and 2m) using the same empirical equations and threshold values within years (not within changing resolutions).

3.3.7. Map assessment

The accuracy of the maps obtained was assessed using the validation dataset containing presences and absences recorded in the field. Error matrices were calculated and the accuracy values and Kappa statistics were derived (Cohen, 1960):

$$K = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \quad (5)$$

Where *K* – Kappa statistics, Pr(a) is observed probability and Pr(e) the probability expected by chance .

Error matrix (confusion matrix) is a table displaying counts of classified presences/absences and is used for assessing the mapping accuracy when predicted values are compared with ‘ground truth’ (Fielding and Bell, 1997). The mapped points are presented in columns and ground truth in rows. Diagonal values represent correctly classified value, other – misclassified. The false positive classification errors are referred to as type I and false positive type II (Morrison et al., 1992).

The Kappa value shows if the map quality in above or below (or equal to) random agreement. Kappa value of 1 this means perfect correspondence with the ground

truth, whereas -1 means that there is no agreement with ground truth and predicted data; Kappa equal to 0 means that the map is no better than random.
Overall accuracy was used for simple assessment of the prediction map; prevalence of positive cases statistic was also derived (Fielding and Bell, 1997).

$$Overall\ accuracy = \frac{Correctly\ classified\ cases}{All\ cases} \tag{6}$$

$$Prevalence = \frac{Posivively\ cassified\ cases}{All\ cases} \tag{7}$$

Overall accuracy may indicate good map predictions, even if in fact it is not true; for example, if prevalence is low a misclassified map would yield good overall accuracy. That is why prevalence value is used to be compared to overall accuracy.
For 2006 the same methodology was implemented using 2006 validation set.

4. Results

4.1. Explanatory variables

The variables showing to be significant for various species presence/absences are elevation, cost-distance to the sea, distance to sea inlet, distance to fresh water seepage and slope. The maps of the parameters can be seen in appendix B. All these variables were obtained from LiDAR DEM in three resolutions (5m, 2m and 20 cm) for 2010 and two resolutions (5m and 2m) for 2006.

4.2. Correlations of predictors and plant presence/absence data

The Kendall's τ and Spearman's ρ correlation tests showed significant correlations of some plant presence/absence data and the predictor variables. The correlation coefficients for all plant species tested are shown in appendix C. As explained above the distance functions increase when further away from the target. This has some specific effect on the sign of the correlation coefficients. For example, fresh water seepage has impact on fresh water plants, the correlation coefficient is negative, showing that the increase in the distance influences the presence values negatively (the closer the better). Similarly, positive sign of the coefficient means that increase of the distance function influences the presences of a species positively (the further the better).

The highest correlations appeared to be between *Puccinellia maritima* and 'Distance to sea inlet' variable (-40,3%) and with 'Elevation'(-37,9%); *Leontodon autumnalis* also shows high correlation to distance to sea inlet (36,6%).

In general highest correlations don't go higher than ±25-30%. Depending on species different predictors have prevalence. Slope appears to have the lowest correlation coefficients (not significant or low correlation coefficients with majority of the species).

Elevation is significantly correlated with the majority of species. In fact only four out of 43 tested plants do not appear to have significant correlation with the elevation variable (*Parnassia palustris*, *Sonchus maritimus*, *Samolus valeriandi*, *Centaurium pulchellum*). However, other predictors like Distance to sea inlet and Cost-distance to the sea show higher coefficients. Distance to fresh water seepage does not show significant correlations to a number of species, others show correlations of ±10-20% which is relatively high.

To reduce the number of modelled species based on the value of the correlation coefficients (above 10%), a number of plants were selected for further analysis (can be seen in table 4.1). The Red List species were also selected for further analysis regardless of the correlation they yielded.

Table 4.1 Table of correlations of species selected for further analysis and predictors. The values shown are correlation coefficients between predictors (in the header) and plants. The correlation tests use different ranking approaches and result in slightly different values. The Red List species are written in bold.

Plant name	Correlation test	Elevation	Cost-distance to sea	Distance to inlet	Distance to fresh water seepage	Slope
Agrostis stolonifera	Kendall's tau	,248**	,065**	,288**	-,243**	-,077**
	Spearman's rho	,302**	,080**	,344**	-,297**	-,093**
Aster tripolium	Kendall's tau	-,112**	-,006	-,130**	,035**	-,024
	Spearman's rho	-,136**	-,008	-,155**	,043**	-,029
Carex extensa	Kendall's tau	,096**	,153**	,182**	-,082**	-,010
	Spearman's rho	,117**	,187**	,217**	-,100**	-,013
Elymus pycnanthus	Kendall's tau	-,030*	-,096**	-,088**	-,134**	,001
	Spearman's rho	-,036*	-,117**	-,105**	-,163**	,001
Festuca rubra	Kendall's tau	,064**	,162**	,074**	-,086**	,099**
	Spearman's rho	,077**	,198**	,089**	-,104**	,121**
Glaux maritima	Kendall's tau	-,117**	-,090**	-,146**	,124**	,006
	Spearman's rho	-,142**	-,109**	-,175**	,151**	,007
Juncus alpinoarticulatis	Kendall's tau	,051**	,076**	,102**	-,006	,051**
	Spearman's rho	,062**	,093**	,122**	-,008	,062**
Juncus gerardii	Kendall's tau	,077**	,083**	,042**	,051**	-,050**
	Spearman's rho	,094**	,101**	,050**	,062**	-,061**
Juncus maritimus	Kendall's tau	,124**	,007	,127**	-,019	-,097**
	Spearman's rho	,150**	,008	,152**	-,023	-,119**
Leontondon autumnalis	Kendall's tau	,167**	,295**	,306**	-,045**	,070**
	Spearman's rho	,203**	,361**	,366**	-,055**	,086**
Linum catharticum	Kendall's tau	,120**	,110**	,139**	-,074**	,017
	Spearman's rho	,146**	,134**	,166**	-,091**	,020
Lythrum salicaria	Kendall's tau	,105**	,079**	,156**	-,125**	-,032**
	Spearman's rho	,128**	,096**	,186**	-,153**	-,039**
Mentha aquatica	Kendall's tau	,131**	,068**	,141**	-,137**	-,078**
	Spearman's rho	,159**	,084**	,169**	-,167**	-,095**
Odontitis vernus	Kendall's tau	,087**	,089**	,137**	-,055**	,007
	Spearman's rho	,106**	,108**	,164**	-,067**	,009
Parnassia palustris	Kendall's tau	,010	,040**	,028*	-,049**	,042**
	Spearman's rho	,012	,049**	,034*	-,060**	,051**

Plant name	Correlation test	Elevation	Cost-distance to sea	Distance to inlet	Distance to fresh water seepage	Slope
Plantago maritima	Kendall's tau	-,132**	,113**	-,128**	,177**	-,032**
	Spearman's rho	-,161**	,138**	-,153**	,215**	-,039**
Puccinellia maritima	Kendall's tau	-,312**	-,219**	-,337**	,158**	,016
	Spearman's rho	-,379**	-,267**	-,403**	,193**	,020
Phragmites australis	Kendall's tau	,186**	,005	,110**	,024	-,012
	Spearman's rho	,225**	,007	,132**	,029	-,014
Sagina nodosa	Kendall's tau	,056**	,073**	,033**	-,019	,104**
	Spearman's rho	,067**	,090**	,040**	-,023	,127**
Salicornia europaea	Kendall's tau	-,212**	-,047**	-,214**	,211**	-,024*
	Spearman's rho	-,257**	-,057**	-,256**	,257**	-,030*
Salix repens	Kendall's tau	,208**	,147**	,236**	-,098**	,002
	Spearman's rho	,252**	,180**	,282**	-,120**	,003
Schoenus nigricans	Kendall's tau	,162**	,076**	,136**	-,129**	-,034**
	Spearman's rho	,196**	,092**	,162**	-,157**	-,041**
Scirpus maritimus	Kendall's tau	-,159**	-,291**	-,229**	-,224**	-,046**
	Spearman's rho	-,193**	-,355**	-,274**	-,273**	-,056**
Spartina anglica	Kendall's tau	-,123**	-,114**	-,114**	,017	-,005
	Spearman's rho	-,149**	-,139**	-,137**	,021	-,006
Triglochin maritima	Kendall's tau	-,064**	,021	-,066**	,046**	-,008
	Spearman's rho	-,078**	,026	-,079**	,057**	-,010

**. Correlation is significant at the 0.01 level.

*. Correlation is significant at the 0.05 level.

4.3. Logistic regression modelling

For species that were selected after correlation analysis (total of 25 species, see table 4.1) a logistic regression model was run. The output of the modelling consists of equation parameters estimated (β_i in the equation 2, section 1.5.3) and characteristics of each model performance. The equation parameters obtained, the variables in each equation and the model performance summaries (R^2 and Hosmer-Lemeshow statistics) are given in appendix D.

Observations

Overall, including the intercept (constant) in the model was lowering the values of the R^2 (relative variation explained parameters). The intercept coefficient in the model gives the probability that the response variable is equal to 1 when all the predictor variables are equal to zero. By excluding the constant this probability becomes equal to 0,5. In equation 2 this would result in the following:

$$\pi(x) = \frac{e^0}{1+e^0} = 0{,}5$$

Excluding the constant might result in an erroneous model, if this assumption is not true. Therefore, this was not done.

In some cases, however, the constant in the model was not significant and was removed (for example for *Parnassia palustris*).

4.3.1. Modelling assessment

After modelling was finished for all the selected species the performance of the models was assessed and plant models with significant goodness of fit parameters were chosen for further mapping. Out of 25 species models only 7 presented significant goodness of fit. However, the R^2 parameters for these were not always high.

Table 4.2 Modelling performance assessment of selected plants. Here and further Elev – elevation variable, Sea dist – cost-distance to the sea, Inl dist – distance to sea inlet, Fr wat dist – distance to fresh water seepage, C&S R^2 – Cox & Snell R^2, Nagelk. R^2 - Nagelkerke R^2.

Sp. name	Const.	Elev	Sea dist	Inl dist	Frwat dist	Slope	C&S R^2	Nagelk. R^2	Hosm-Lem	df	Sign
Juncus alpino-articulatus	-8,41	0	2,30E-05	0,01	0	0,03	0,022	0,139	9,16	8	0,329
Linum catharticum	-15,42	0,04	4,80E-05	0	-0,01	0	0,045	0,202	6,09	8	0,637
Mentha aquatica	-11,38	0,04	0	0,007	-0,05	-0,04	0,057	0,335	6,5	8	0,591
Parnassia palustris	0	-0,03	2,88E-05	0	-0,03	0,03	0,724	0,965	5,89	8	0,66
Salix repens	-15,96	0,04	3,95E-05	0,01	0	0	0,103	0,274	10,89	8	0,208
Salicornia eurupaea	5,37	-0,03	-1,10E-04	0	0,02	0	0,124	0,468	5,29	8	0,726
Spartina anglica	6,37	-0,05	-2,33E-05	-0,03	0	0	0,033	0,295	8,75	8	0,364

4.3.2. Modelling performance and different species

It appears much easier to obtain a significant model for plant species with a narrow range of distribution, since these plants' variation can be explained by local factors. As opposed to generalist plants that had not resulted in good models. A comparison of model predictions may be seen in figures 4.1 – 4.2. The classification plots (see section 3.3.6 for more detailed explanation) show how the model groups the observed presences and absences. It is clear that the model for *Agrostis stolonifera* fails to correctly classify the presences and absences of the plant, as compared to the *Mentha aquatica* model, which groups the observed presences and absences well.

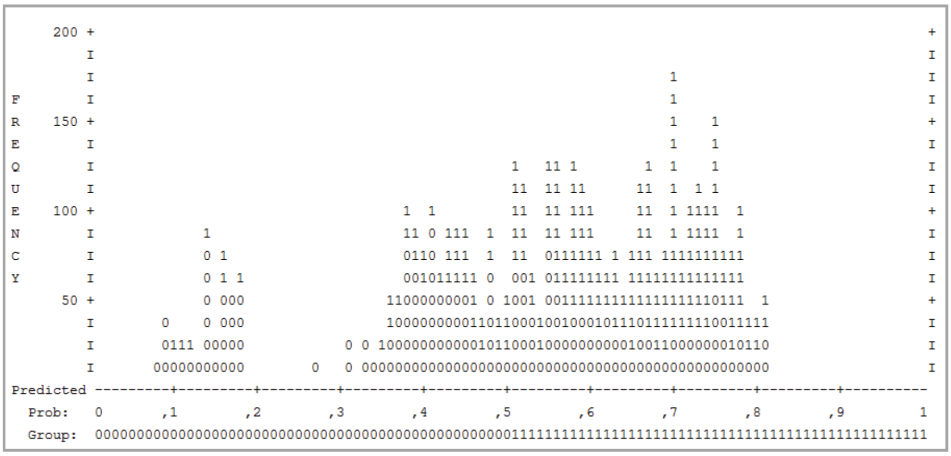


Figure 4.1 SPSS classification plot of predicted and observed probabilities of *Agrostis stolonifera*. 1 represents observed presence, 0 – observed absence. Note how the cases of presences and absences are not clustered separately, the observed presences are grouped together with absences. This indicates that the predictive performance of the model in not good enough for this species.

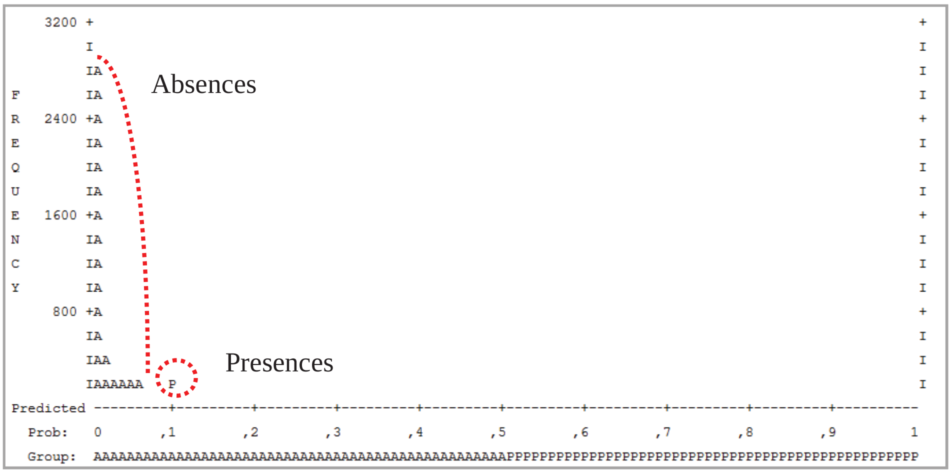


Figure 4.2 SPSS classification plot of predicted and observed probabilities of *Mentha aquatica*. Indicated on the graph the grouping of the observed presences and absences, the two groups are located separately, which means that the model differentiates them well.

Clearly, the species with wider range of distribution, like *Agrostis stolonifera*, require a different approach in modelling and possibly a different set of predictor variables.

4.4. Mapping species distributions

The maps produced showed the location of predicted species habitats at specified probability threshold. The example of prediction maps is shown in appendix C.

4.4.1. Mapping assessment

The maps were produced using the parameters obtained from modelling for both 2010 and 2006. The threshold used for both years was the same for same plants. The map assessment outputs can be seen in table 4.3.

Table 4.3 Mapping assessment of selected plants for 2001 and 2010.
Kappa values for 2006 for *Salicornia europaea* and *Linum catharticum* are equal to 0. This is because these plants were not detected in the sampling areas during the field work of that year.

		2010			2006		
Raster resolution	Probability threshold	Kappa value	Overall accuracy	Preva-lence	Kappa value	Overall accuracy	Preva-lence
Mentha aquatica							
5m	0,100	0,200	0,940	0,065	0,446	0,901	0,089
2m	0,100	0,191	0,941	0,060	0,480	0,912	0,078
0,2m	0,080	0,173	0,916	0,090			
Parnassia palustris							
5m	0,018	0,060	0,847	0,152	0,064	0,545	0,485
2m	0,011	0,012	0,548	0,015	0,108	0,735	0,275
0,2m	0,011	0,012	0,568	0,436			
Spartina anglica							
5m	0,075	0,196	0,947	0,055	0,070	0,861	0,119
2m	0,060	0,239	0,948	0,012	0,161	0,922	0,059
0,2m	0,074	0,272	0,958	0,047			
Salix repens							
5m	0,120	0,221	0,812	0,202	0,448	0,842	0,228
2m	0,040	0,395	0,866	0,171	0,446	0,901	0,422
0,2m	0,074	0,171	0,705	0,334			
Salicornia europaea							
5m	0,080	0,320	0,890	0,127	**0***	0,812	0,188
2m	0,094	0,332	0,904	0,121	**0***	0,657	0,343
0,2m	0,098	0,349	0,910	0,115			
Juncus alpinoarticulatus							
5m	0,02	0,0436	0,718	0,286	0,156	0,624	0,416
2m	0,02	0,038	0,696	0,309	0,201	0,667	0,373
0,2m	0,029	0,043	0,802	0,195			
Linum catharticum							
5m	0,04	0,143	0,77	0,211	**0***	0,812	0,188
2m	0,04	0,147	0,785	0,233	**0***	0,118	0,882
0,2m	0,05	0,151	0,808	0,206			

4.5. Cell size impact

4.6. Model performance

As modelling was performed for the selected plants using 3 raster resolutions (5m, 2m, 20cm) the model outputs of different resolutions for each plant varied. The resulting changes were as follows: a) changes in modelling performance (goodness of fit and approximate variance explained) and b) changes in predictors included in the model. The changes in the equation predictors also affect the modelling performance additionally to the effect of the cell size. The species with changes in the model variables were not included in the assessment of the cell size impact.

Table 4.4 contains lists of plants grouped according to whether or not the model variables were changed.

Table 4.4 Plant lists according to the change in model predictors. The left column contains plants that were used for further assessment of the cell size impact

No change in the model predictors	Change in the predictors included in the model
Agrostis stolonifera	*Elymus pycnantus*
Aster tripolium	*Juncus gerardii*
Carex extensa	*Leontodon autumnalis*
Festuca rubra	*Lythrum salicaria*
Glaux maritima	*Odontites vernus*
Juncus alpinoarticulatus	*Parnassia palustris*
Juncus maritimus	*Puccinellia maritima*
Linum catharticum	*Sagina nodosa*
Mentha aquatica	*Spartina anglica*
Phragmites australis	*Schoenus nigricans*
Plantago maritima	
Puccinellia maritima	
Salix repens	
Salicornia europaea	
Scirpus maritimus	
Triglochin maritima	

The assessment of the model performance with resolution change for selected plants is shown in table 4.5. The model assessment parameters for the plant models that have changes in their predictors can be seen in the appendix D.

Table 4.5 Model assessment with change in resolution. C&S R^2 – Cox & Snell R^2, Nagelk. R^2 - Nagelkerke R^2

Cell size	C&S R^2	Nagelk. R^2	Hosm-Lem	df	Sign
Agrostis stolonifera					
5 m	0,136	0,182	136,347	8	1,36E-25
2 m	0,137	0,183	123,432	8	6,48E-23
20 cm	0,136	0,182	118,096	8	8,19E-22
Aster tripolium					
5 m	0,706	0,941	44,392	8	4,80E-07
2 m	0,706	0,941	38,536	8	5,99E-06
20 cm	0,706	0,941	43,826	8	6,14E-07
Carex extensa					
5 m	0,064	0,1	64,074	8	7,36E-11
2 m	0,065	0,101	66,504	8	2,43E-11
20 cm	0,067	0,104	74,697	8	5,67E-13
Festuca rubra					
5 m	0,317	0,422	16,926	8	0,031
2 m	0,32	0,426	23,942	8	0,002
20 cm	0,321	0,428	16,623	8	0,034
Glaux maritima					
5 m	0,032	0,055	78,493	8	9,82E-14
2 m	0,032	0,054	72,81	8	1,35E-12
20 cm	0,032	0,054	71,807	8	2,14E-12
Juncus alpinoarticulatus					
5 m	0,022	0,139	9,16	8	0,329
2 m	0,021	0,134	13,524	8	0,095
20 cm	0,021	0,13	6,668	8	0,573
Juncus maritimus					
5 m	0,555	0,741	65,139	8	4,53E-11
2 m	0,552	0,736	66,743	8	2,18E-11
20 cm	0,552	0,736	47,706	8	1,13E-07
Linum catharticum					
5 m	0,045	0,202	6,094	8	0,637
2 m	0,047	0,211	12,351	8	0,136
20 cm	0,047	0,212	12,404	8	0,134

Cell size	C&S R^2	Nagelk. R^2	Hosm-Lem	df	Sign
Mentha aquatica					
5 m	0,715	0,953	7,78	8	0,455
2 m	0,715	0,953	8,103	8	0,423
20 cm	0,714	0,953	4,383	8	0,821
Phragmites australis					
5 m	0,056	0,129	47,739	8	1,11E-07
2 m	0,063	0,144	34,592	8	3,17E-05
20 cm	0,059	0,137	48,347	8	8,48E-08
Plantago maritima					
5 m	0,076	0,228	26,139	8	0,001
2 m	0,076	0,227	16,185	8	0,04
20 cm	0,075	0,224	21,792	8	0,005
Puccinellia maritima					
5 m	0,622	0,829	67,758	8	1,37E-11
2 m	0,623	0,83	91,705	8	2,10E-16
20 cm	0,623	0,831	113,887	8	6,04E-21
Salix repens					
5 m	0,103	0,274	10,889	8	0,208
2 m	0,107	0,284	12,649	8	0,125
20 cm	0,108	0,286	8,627	8	0,375
Salicornia europaea					
5 m	0,124	0,468	5,288	8	0,726
2 m	0,123	0,465	4,516	8	0,808
20 cm	0,123	0,465	4,413	8	0,818
Scirpus maritimus					
5 m	0,55	0,734	322,799	8	5,74E-65
2 m	0,55	0,734	132,042	8	3,09E-80
20 cm	0,55	0,733	384,699	8	3,51E-78
Triglochin maritima					
5 m	0,706	0,941	34,661	8	3,08E-05
2 m	0,706	0,941	17,889	8	0,022
20 cm	0,706	0,941	16,875	8	0,031

Changes in the significance of Hosmer-Lemeshow goodness of fit statistics were investigated for the selected species. Equation 4 in section 3.3.5 was used to standardise the values to make them comparable. The value can be seen in table 4.6.

Table 4.6 The percentage of change in the model assessment parameters with increase in predictor cell size for each plant

Cell size	*Agrostis stolonifera*	*Festuca rubra*	*Glaux maritima*	*Juncus alpino-articulatus*	*Plantago maritima*	*Salix repens*
5m	0	0	0	0	0	0
2m	7,31	-43,28	34,88	-23,47	84,78	-11,72
0,2m	92,64	4,48	56,92	24,47	8,70	23,59
	Linum catharticum	*Triglochin maritima*	*Mentha aquatica*	*Aster tripolium*	*Phragmites australis*	*Puccinellia maritima*
5m	0	0	0	0	0	0
2m	-55,24	41,43	-13,91	77,79	99,04	-100,00
0,2m	-55,46	58,40	12,01	1,89	-0,08	-100,00
	Salicornia europaea	*Juncus maritimus*	*Scirpus maritimus*	*Schoenus nigricans*	*Aster tripolium*	*Carex extensa*
5m	0	0	0	0	0	0
2m	3,49	-0,02	-100,00	0,45	77,79	-50,03
0,2m	3,91	99,90	-100,00	99,55	1,89	-74,15

Some patterns can be observed in the changes: decrease in goodness of fit with decrease in cell size; decrease at 2 m and increase at 20 cm; gradual increase of goodness of fit with decrease in cell size; a spike at 2 m resolution, slight increase at 2 m and a strong increase at 20 cm (see figures below). Figures 4.3 – 4.7 show the behaviour of the significance of the Hosmer-Lemeshow goodness of fit parameter with change in the raster resolution of the model predictors.

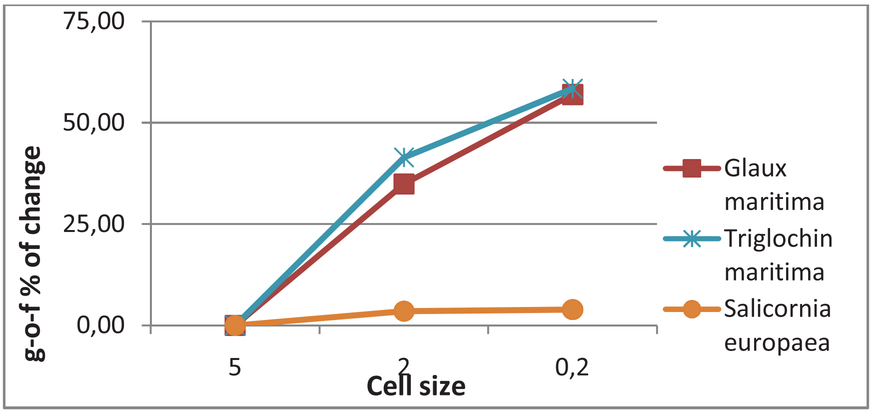


Figure 4.3 Resolution effect: gradual increase in goodness of fit

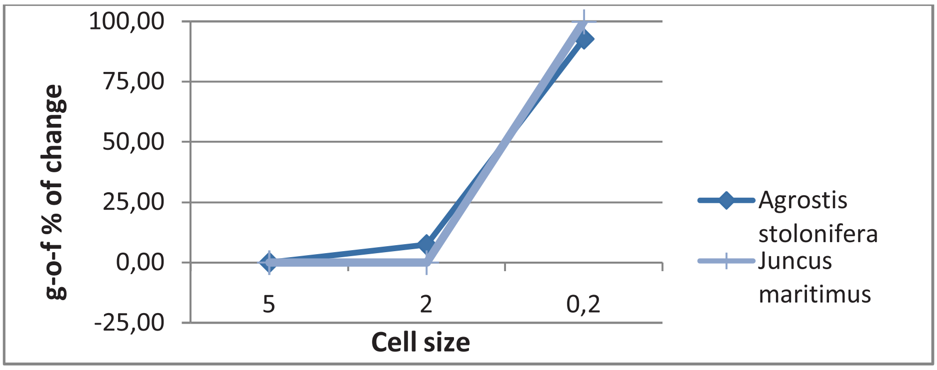


Figure 4.4 Resolution effect: strong increase in goodness of fit at 20 cm cell size

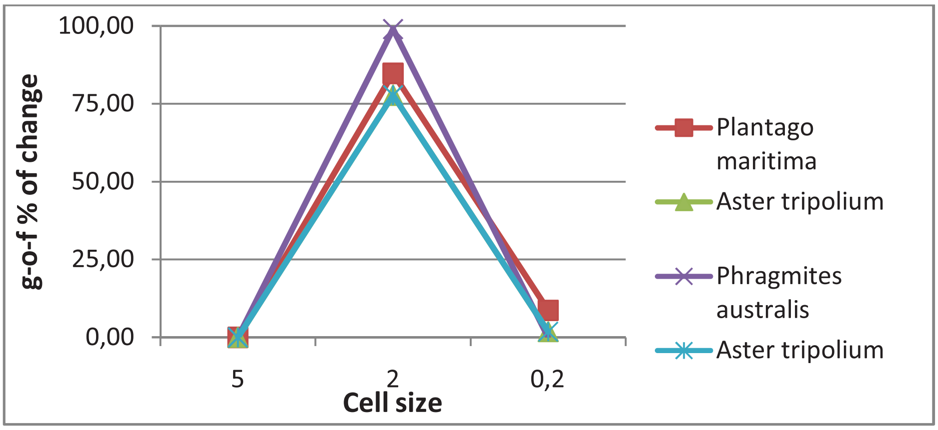


Figure 4.5 Resolution effect: peak in goodness of fit at 2 m cell size

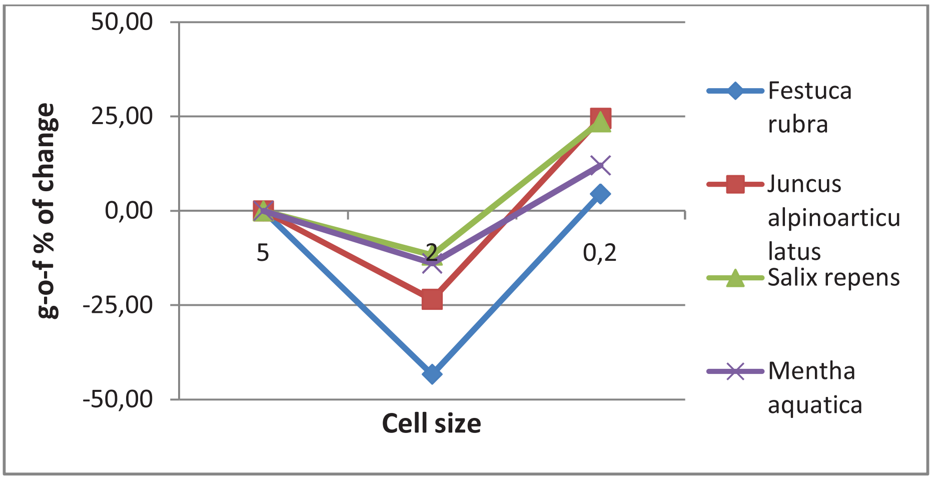


Figure 4.6 Resolution effect: decrease in goodness of fit at 2 m cell size

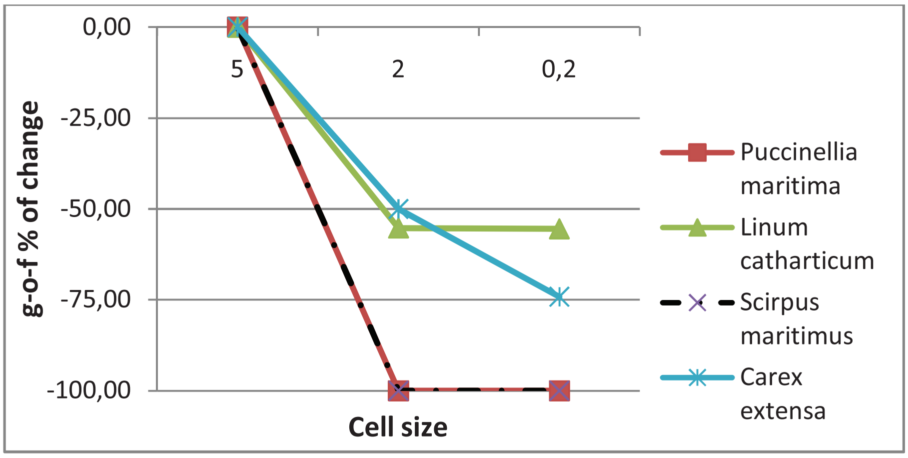


Figure 4.7 Resolution effect: general decrease in goodness of fit with increased resolution

4.6.1. Mapping accuracy

The increase in raster resolution also has some impact on the assessment of the accuracy of the mapped predictions. Some slight increase can be observed. However, these changes are quite small in general. Table 4.3 shows the values of the assessment parameters.

Below a graph of the change in Kappa statistics throughout the different resolutions is shown (figure 4.8).

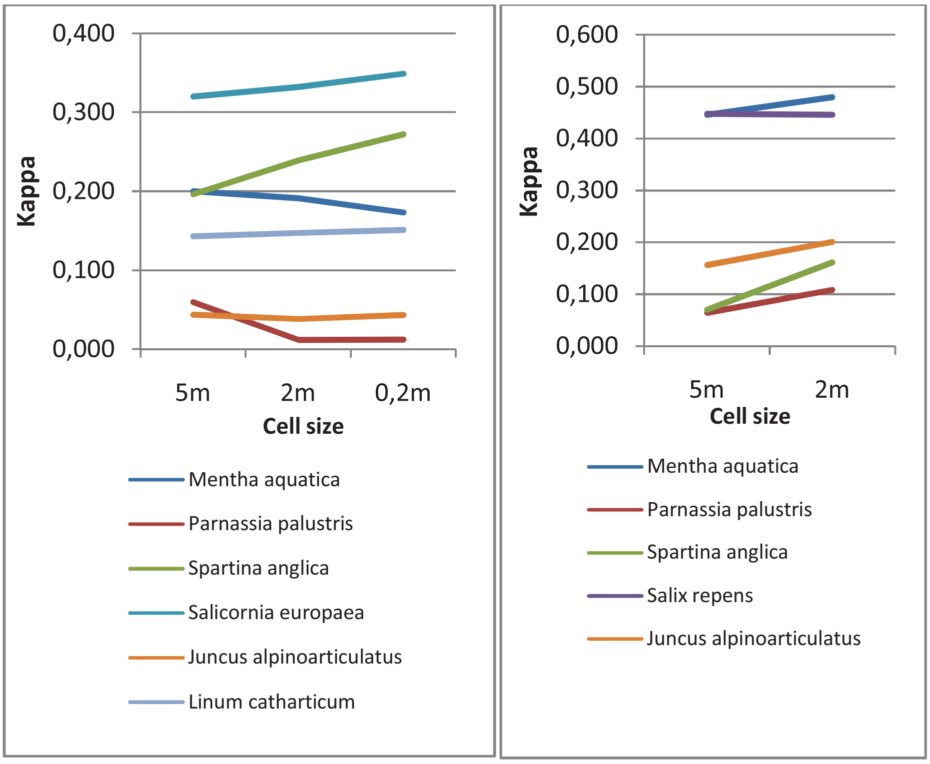


Figure 4.8 Kappa statistics for maps with increased resolutions (left 2010, b 2006 situation)

The graphs show some general slight increase in the accuracy, despite several plants having decrease in the accuracy.

5. Discussions

5.1. Positioning accuracy and DEM

As previously mentioned, quite high positioning accuracy is required for this kind of study to be successful since the variations in vegetation distribution depend on small scale variations in the topography and other factors. If this small scale variation is not reproduced by the elevation model or if the positioning is quite off, it may result in the wrong parameters being associated with presence/absence data. Consequently, an erroneous model may be a result, which does not represent the state of plant distribution in the study area.

In the case of this study the horizontal (x,y) accuracy of the original LiDAR DEM is approximately 50 cm. The accuracy of geodetic surveying which was used to calculate the x,y coordinates of the sample points ranges from 50 cm to 1 m. This means that the actual sample points may be more than one meter away from the points where the predictor variables were extracted from. Given the spatial scale of the vegetation patterns, that represents an obvious drawback.

The patterns observed in the field varied greatly in spatial extent. In the east-west direction the change in vegetation was quite gradual, with similar patterns expanding to up to hundreds of meters. On the other hand rapid changes in the plant distribution were also observed. For example, where the small hillocks in the terrain occurred the vegetation patterns were as small as 20-40 cm. This depended greatly on the degree of change in the elevation.

The spatial accuracy of the explanatory data is surely good enough for explaining plant variation of larger scale. However, this might not be the case with small rapidly changing patterns.

5.2. DEM interpolation

Next issue is the interpolation of the DEM to decrease the resolution. It is clear that the 5 m cell resolution of the DEM seems too coarse to be used for modelling a small scale phenomenon such as the one studied here. Hence an increase in the spatial resolution was needed.

The kriging algorithm seems to be a good choice, since it does not assume a linear relation between the neighbouring points but rather takes into account a specified radius and uses all the points giving them weight according to distance to the interpolation location. Thus the interpolated surface is continuous, smooth and close to the original data.

It is considered that the relation between the original resolution and the resulting resolutions influences the interpolation output. If the centre point of a cell falls on the border of the newly interpolated cells (e.g. if the cell is split evenly in 4 cells), the resulting raster may be erroneously interpolated, decreasing some values, see figure 5.1.

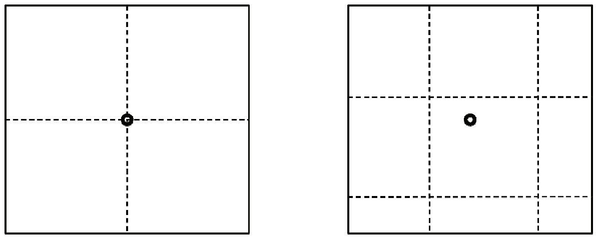


Figure 5.1 Different way of splitting the raster cell of original resolution with increase in cell size. On the left the boarder of newly interpolated cells hits the original centre point, on the left – the centre point is contained inside of an interpolated cell.

In this study the cell sizes were selected in such way that the original centre point falls inside the cell of the increased resolution.

One major drawback of this interpolation is lack of true data for validation. Here a positioning inaccuracy may be much accelerated by increasing the resolution. A coarser resolution may mitigate somewhat the positioning error as the sampling point falls in a cell with similar variable values as compared to its original predictors. Increased resolution has a greater impact putting the sample points in accordance to the values which are not necessarily correct.

It is difficult to assess the impact of increased resolution on the accuracy without true data to compare with. However, the study outcomes suggest that this influence was acceptable, since some increase in model performance is observed with increased resolution.

5.3. Species presence/absence data and logistic regression model

5.3.1. Data

Presence/absence

The presence data surely represent true presences. The absence data are not necessarily true absence, in the sense that the habitat conditions are suitable for a certain plant species, but it simply is not present. The abundance of some species is low by default, even within the optimum of their ecological amplitude. It is more realistic to say that the presence/absence data represents realised niche for the species, not fundamental (see section 1.5 for explanation). It is not only the abiotic factors that influence the growth of plants but also the interaction/competition between species that forms the true plant distribution. In this research only abiotic factors were incorporated in modelling.

Rare species with little presences as compared to a big number of absences would yield a low probability of presence even at locations where they are truly present (as the probability is calculated from the number of presences and absences). In logistic regression this results in very low threshold for mapping the predictions, as seen in this study. For instance, the probability threshold used for mapping the distribution of *Mentha aquatica* and *Slicornia europaea* was 0,1; for other plants it was even lower.

Possible ways to improve the problem with the absence data is to group the species that always occur together into species groups combining the presence data. This can be done by clustering the species with similar spatial behaviour. Although only visual interpretation of the data is required, this is quite time-consuming.

It's also possible to choose some dominant representative species for different vegetation types to be used for modelling. However, a strong justification has to be done for selecting dominant indicator species, some more extensive field work and great knowledge of the plant ecology is required

With these changes the study would yield habitat predictions for vegetation groups having broader borders than per plant species but also being more realistic.

One other thing for improving the modelling would be incorporating the interactions between the plants, possibly using presences of some competitive species as an explanatory variable. However, this requires knowledge in the species interactions and more time than given for present study.

5.3.2. Validation data for 2006

The data obtained from another source is always under question. How the data was collected and with which amount of accuracy is not known exactly. In this case some amount of consideration should be addressed towards the positioning accuracy. The data are said to be located within the same line as the beach poles on the shore. However, the 'true' coordinates were only retained for two transects (out of 5), thus leaving the other transects under question. This brings the question whether these data can still be used, taking the positional error into consideration. Surely, the 2006 data can be used to assess the general situation of that year. For example, it shows that *Saliconia europaea* and *Linum catharticum* were not found in the study area, to the contrary of the model predictions. Positioning accuracy in this case does not play a big role. Since the absence was reported in all transects, whereas modelled presence did occur in the range around two transects. See figure 5.2 below.

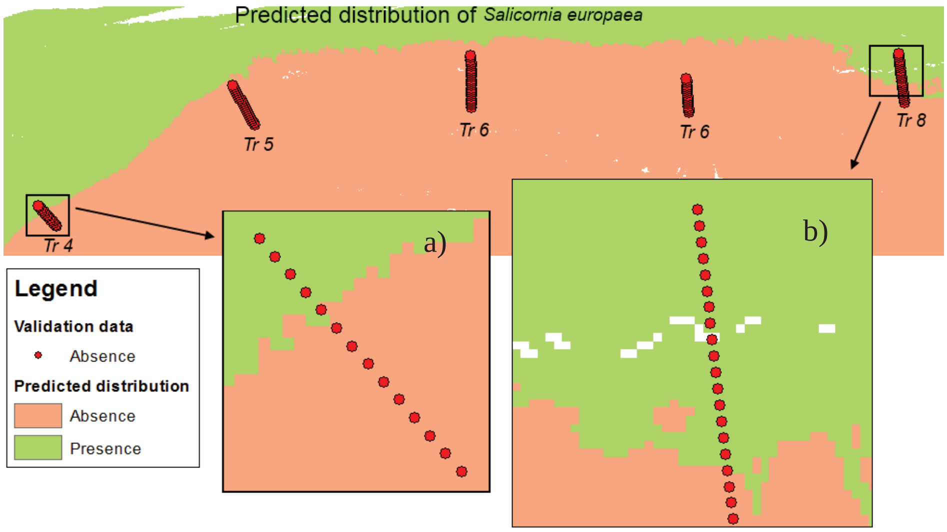


Figure 5.2 Predicted distribution of *Salicornia europaea* and 2006 field data showing the absence of the species. a) and b) show the areas where the model predicts presences, and the validation data states absences only.

This is not the case with other plants, since the validation data points would not necessarily correspond to the right locations on the prediction map resulting in an erroneous assessment of the prediction map.

Another issue is adjusting the data to fit the logistic regression model output. As mentioned earlier, the data was obtained as cover estimate for 2 m^2 plots. The cover was simply transformed into presence data. This means that across a 2x2 m cell each plant was detected as present, regardless of the percentage cover. Clearly, this kind of validation data does not match the validation data set derived for 2010, and may cause some inconsistencies in comparison of the map accuracies.

5.3.3. Logistic regression

Some other consideration should be given to the nature of the statistic model used in this study. Having its advantages like simplicity, ease of implementation and comprehensiveness, it also has apparent disadvantages, as seen here.

Logistic regression is threshold sensitive. It has been reported before by Manel et al. (1999) and also seen in this study. The accuracy of predicting presences strongly depended on choosing the more suitable threshold. However, lowering the threshold means introducing type II error when classifying the less suitable (or even unsuitable) areas for species distribution as suitable. If the probability threshold is set too high, the presences in their turn would be classified as absences (type I error). In this case a trade off between the two needs to be set.

Mantel et al. (1999) also state that logistic regression model predictions are affected by species prevalence (frequency of occurrence). As prevalence of species increases the effectiveness of the prediction of presences by the model also increases, and the opposite, with decrease of prevalence, the predictions of absences is more effective. (Manel et al., 2001). Since in our case in the majority of species sampled the absences prevail, this is the case where logistic regression model outputs may be affected.

The issues of threshold and prevalence are linked together. Species prevalence causes presences/absences to be overestimated and requires changing the probability threshold from standard 0,5 to a more suitable value.

Another drawback of using logistic regression model reported by Hegel et al. (Hegel et al., 2010) is the "true absence" of species. The use of logistic regression assumes that the p/a data are true. In practice this is often not the case due to errors/bias in sampling etc.

This brings us back to a previous discussion on the presence/absence data. The sampled absence data are not exactly true absences, which may also affect the logistic regression modelling. In this case grouping the species would have made an improvement, since the number of presences would increase, decreasing the number of absences.

Also, the habitat may be suitable for specific plant as far as abiotic factors are concerned, however, the biotic interactions may result is a species being absent in the sampling area. In this case abiotic explanatory variables may not result in good predictions.

Suggestions for alternative statistical approaches

Logistic regression uses a logit function for linear transformation of the relationship between the predictors and the probability of occurrence (Hegel et al., 2010). The assumption is somewhat restrictive, since a linear relation cannot be always expected. Other statistical approaches exist, that allow a more complex relationship. Moreover, approaches that would be more appropriate given the nature of the data at hand also exist. Hegel et al. (2010) suggest alternative statistical models.

Generalised additive models (GAMs), for instance, allow for more complex relationship to be modelled, they estimate a non-parametric smooth function for the predictors in a model. The smoothness of the relationship is to be decided by the analyst, and it's important not to overfit the data with a too smooth a model.

Some possible approaches to be used to account for the problem of 'true absences' are models using presence only data. These can be Ecological niche factor analysis (ENFA) or Entropy maximisation.
ENFA (Hirzel et al., 2002) compares the predictor variables where the species is present with these variable across the study area to estimate habitat suitability. Maximum entropy model (Phillips et al., 2006) is an algorithm based approach that identifies the probability distribution which is least biased according to the information contained in the random variable.
The models mentioned above would be suitable in case of this study, since they are used with binary or presence-only data, which can be derived from the field data at hand.

5.4. Predictors and correlations

The method greatly depends on choosing the suitable variables.
In this study only a small number of DEM derived variables were chosen. There may be more possible variables derived like, for example, slope curvature. Curvature indicates the concavity/convexity of the topography and can be used as another indirect variable.
Additionally, it makes sense that there are direct variables effecting plant distribution, which are not accounted for by indirect DEM-derived predictors. Some more possible factors influencing vegetation distribution on the green beach may connected to soil properties like clay content in soil, soil fertility etc. These were not considered in current study.
It is also possible that the variables chosen were not good proxies for the environmental factors (inundation frequency/duration, ground water salinity etc.), since the processes in nature are more complex than our attempt to represent them. For example, Bockelmann et al. (2002) have described that using elevation as a proxy for inundation frequency/duration may not yield good correlation with vegetation distribution. Tidal inundation depends on additional factors like winds, currents, sea bottom relief, and not only elevation.

5.4.1. Inter-correlations between variables

Looking at the correlation table of the predictors, a quite high correlation between the elevation and distance to the sea inlet variable can be observed. This may be due to the following. When calculating distance towards the sea inlet areas the function increases with distance from the target area. As the sea inlet is a lower area in the landscape, when moving away from it the elevation increases. In this way the distance function is correlated to elevation, they repeat a similar pattern. The

distance to inlet function increases linearly, whereas elevation is a more complex parameter.
Generally, since all the parameters are derived from the same digital elevation model they show correlations with elevation and each other as well. To assess the multicolinearity of the data the *VIF* parameter was estimated for all the predictors (see section 3.3.1.1). The *VIF* values (see table 3.2) did not indicate problems with multicolinearity, and it was concluded that the predictors are suitable for modelling.

5.4.2. Plant presence/absence data and predictors

The correlations between the dependant data and the independent variables are not always high. Some plants do not have correlations more than 3-5% with any of the predictors. This might indicate that the predictors are not suitable to explain variation in distribution of certain plants, and that other factors have higher influence on these distributions. Moreover, some plants may have occurred at a scale that was not captured by the raster resolution. Even though the resolution is increased, some small scale features present in the field, might not be captured by the interpolation.
This may also be connected to comparatively low correlations between elevation and plant distribution, as these correlations are never higher than correlation of the plant with other predictors. Clearly, it was expected that correlations with the elevation were higher, as elevation is key to inundation frequency and duration, one of the main ecological factors in the green beach (Hickey and Bruce, 2010).
In some way elevation being important factor in plant distribution is confirmed. Only four plants did not show significant correlation with elevation.
Distance to the sea inlet or cost-distance to the sea show higher correlations with plant distributions. This may mean that these variables explain the variation better and are better approximation (then elevation) of the natural factors like inundation.
Distance to fresh water seepage shows significant correlations to some plants requiring fresh water. However, it also shows strong positive correlations with saline plants. As mentioned before the distance functions have positive correlation with plants that are further away from the target areas.
Slope parameter does not seem to have high correlations to majority of species, even though the lower correlations are often flagged as significant. Here slope of the terrain indicates the areas with quick changes in elevation, such as small bumps and hillocks in the relatively flat area of the green beach.
The plants showing highest positive correlations with the slope parameter are *Hipophae rhamnoides*, *Plantago coronopus*, *Sagina nodosa* (see appendix C) which are species occurring in the dune environment. In fact, *Hipophae rhamnoides* shows highest correlation with the slope among all predictors. *Festuca rubra* also is highly

correlated with slope, in the field it appeared to be growing extensively on small bumps in the landscape as well.

One other thing to be addressed here is the lack of consistency between correlation coefficients of variables and modelling of the plant distributions. It seems that the plants having high correlations with the predictor variables would also result in a significant model, however, this is not the case.

The plant showing the highest correlation coefficients with the predictors is *Puccinellia maritima.* However, the model for this plant does not do yield a significant goodness of fit. And on the contrary, plants showing low correlations (e.g. *Parnassia palustris*) result in significant model performance.

It may be possible, that the false absence causes the inconsistencies between correlations and modelling performance to happen. These coupled with some limitation of the statistical tool may be considered as the reason. The limitations of the logistic regression model have already been discussed above.

5.5. Model performance

The modelling of 25 species has resulted in 7 significantly fitted models. This, of course, may be due to the drawbacks of the statistical approach discussed before. Additionally, the predictors may not be best for predicting the species variation, which was also discussed earlier.

It should be noted though, that the plants that did result in significantly fitted models had a very narrow distributions at the edge of the study area, being it *Spartina anglica* located along the seashore or *Mentha aquatica* along the coastal foredune. This indicates that the models of plants with narrower niches were more successful than the ones for more spread distributions. A narrow niche is said to facilitate more accurate models, since it may be more concise and more predictable (McPherson et al., 2006).

5.6. Mapping predictions

5.6.1. Mapping overview

The models mapped seemed to capture variation in the study area relatively well. However, outside the study area (where sampling was not performed, but the values were used for mapping) the model overestimated the probabilities of plant presence greatly. For example, areas of the seashore were shown to contain *Spartina anglica* and *Salicornia europaea*, or the embryonic dune areas (were there as not observed big vegetation cover) as well as sea shore area were shown to contain *Juncus alpinoarticulatus*.

Since the locations where the distribution was overestimated are outside if the study area, the sampling was not performed there, and plant absences were not sampled. This means that some species have no upper or lower limit in elevation, and the model assumes presence.

5.6.2. Assessment

Mapping the selected plants hasn't yielded accurate prediction maps, despite that the models were fit and with relatively high R^2. This depends to a large extent on the species prevalence. The assessment was done using an error matrix to derive Kappa statistic and overall accuracy. Generally overall accuracy values are high varying from 95% to about 50%. Kappa values, however, are quite low (the highest value is 0,39 for 2010). Kappa, of course, is threshold sensitive, but an optimal threshold was chosen for each plant species. Probably different thresholds would have yielded different values.

Another criticism of Kappa statistics is that it does not account for expected frequency of a given class (presence or absence) or distinguish between various sources of disagreement (Hegel et al., 2010). This way a map having a big number of absences would result in high accuracy even if created by chance. As Kappa is calculated the comparison between the probability of a random agreement and observed agreement does not result in significant difference, producing a low Kappa values.

Weighted Kappa (Cohen, 1968) is another option. Unlike Kappa, it does not treat all disagreements equally but introduces ratio-scale degrees for disagreements. This could be used as an option, however, first, a justifiable set of weights need to be developed for correct assessment.

5.6.3. Retrospective mapping

The retrospective mapping has not shown good results as the kappa statistics are very low. This can have various explanations.

The questionable positioning accuracy of the validation data is a problem.

Additionally, one cannot expect a better performance of the model on secondary 2006 data, than on the original 2010 training/validation data set. Better results compared to the original year would be misleading. The only thing that might work better for assessment of the map is the smaller quantity of the sample points (2 m^2 cells at 10 m distance). This produces fewer absences as compared to the 2010 data.

Finally, another conclusion that can be made is that the plant distribution does not have constant relation to selected environmental predictors. For instance *Salicornia europaea* and *Linum catharticum* were not present in the 2006 validation data; however, the 2010 model predicted some areas suitable for these plants.

Note that even though the plants' presences were misclassified, the overall accuracy of the maps is quite high (up to 81,2%). This represents a good example of a drawback of the overall accuracy in the statistic parameter.

Since the green beach is dynamic, the changes occurring in its environment may be quicker than the response of the plants. The difference in plant distributions may mean that some species of the green beach have not had enough time in 2006 to take/invade into their optimal niches.

The processes happening in the study area are dynamic and complex and involve other factors, than the simplified static model can account for.

5.7. Cell size impact

The influence of the cell size on species distribution modelling is an interesting topic. Changes in resolution of the distribution mapping has been studied before on various scales and for various reasons (Araujo et al., 2005, Johnson et al., 2002, McPherson et al., 2006, Thomas et al., 2002, Tobalske, 2002). One common investigation of this issue is connected to the need of deriving a finer resolution habitat models from existing coarse resolution data (Araujo et al., 2005, McPherson et al., 2006), so called downscaling, and assessing their accuracy. The data may be derived from, for instance, existing atlas data containing values of some 50x50 km resolution. In this case statistical approaches are used to increase the resolution of the data and some suitable ground data is used for assessment. In cases like this the downscaling still results at km wide cell resolutions.

More studies are looking at the effect of coarsening the predictor data on the output distribution models (Guisan et al., 2007, Tobalske, 2002). Here the original fine scale predictor data is being coarsened and used for detecting the effect on the model performance, predictions of various species.

In the case of this study, on the contrary, the data resolution is being increased; additionally the scale used is very fine – 5 m, 2 m and 20 cm resolutions. Looking at the change in cell size and its effects at this scale is somewhat new.

One other comment is that the training and validation data was of the same sampling scale as the fines resolution used for modelling, in the contrary to the studies that only had coarse resolution data to be downscaled (Araujo et al., 2005).

5.7.1. Changes in the model variables

During the modelling ten species have shown change in the model variables as the raster resolution was changed (see appendix C.2).

These changes may be due to an error of model selection, it is possible that an error occurred during the stepwise model selection.

Another option is that the change in cell size has resulted in the predictor values to be changed causing them to be removed or included in the model. The dominance of the environmental predictors controlling the distribution shifts with spatial scale (McPherson et al., 2006).

The first reason seems to be the case with *Elymus pycnantus*, *Puccinellia maritima, Odontites vernus*, *Parnassia pallustris, Lythrum salicaria,* since the changes do not imply improvement in the model in these cases, and also involve changes in the other parameter values.

Changes in the models for *Juncus gerardii* or *Sagina nodosa* (the distance to fresh water seepage parameter is removed in both cases) do slightly improve the assessment parameters. These changes also do not influence the values of the other predictor coefficients. This might indicate that at the level of 20 cm resolution this predictor does not influence the distribution as much as within the other resolutions.

Generally, it's hard to say if the changes of the model assessment parameters are caused by the change in the model variables directly or the cell size.

5.7.2. Changes in goodness of fit

An important outcome of this study is connected to this result.

The significance of goodness of fit of the modelling changed due to changes in the resolution of the predictors. These changes did not have the same pattern but rather depended on the species, since the location, the sampling, and the statistical approach remained the same. Guisan et al. (2007) having conducted a study of resolution effect on various factors conclude that three effects can be observed when coarsening cell size – improvement, no change and degradation. Generally they have not discovered a substantial effect of resolution change on the plant distribution models. In our case the outcome is similar, however, there has not been a no change effect observed. Possibly the higher impact of the cell size here is due to a finer resolution of the data.

Guisan et al. also mention that the effect on presence/absence data had shown a more significant result as compared to presence-only data. This may be the case here.

Looking specifically at the behaviour of different plants it can be inferred that increase in the model performance happens when the cell size matches the scale of the distribution pattern of the species. For instance, some dramatic change occurs in the model performance of *Agrostis stolnoifera* (the goodness of fit increases more than 90%) at 20 cm. This may be explained by the crisp boundary of the plant distribution in the field that occurs at slight increase in elevation at small dunes. This boundary may be not inferred from the data of the coarser resolution, but at 20 cm becomes much easier to be modelled causing change in model performance.

Some more plants indicate increase in the goodness of fit – a more gradual increase with increased resolution, as well as a peak of model performance at 2 m resolutions. This might indicate that the optimal cell size for given plants together with the predictors used for modelling.

One other important feature is degradation of model performance. This is not something that could be expected prior to the study. This could suggest that one or several key predictors used for these species do not represent the reality well at certain scales and increasing the cell size only causes bigger error in the modelling.

5.7.3. Changes in map accuracy

The mapping was performed using a) different empirical equations (with slightly different coefficients) derived for each resolution and b) different cell size. A slight increase in the output map accuracies may be resulted by both parameters. Figure 5.3 shows how maps of *Mentha aquatica* distribution in the same area look with increased resolution. It's clear that the 20 cm map gives a more smooth impression, although 2 m map also looks smooth enough. 5 m resolution map looks rather coarse for mapping a small scale distribution. However, this kind of conclusion should be made accordingly to the requirements for mapping.

As the resolution increases the map accuracies increase somewhat as well. This may be due to the ability of the smaller cell size have higher variation frequency to match the variation of the input or validation data, that has quite fine sampling scale.

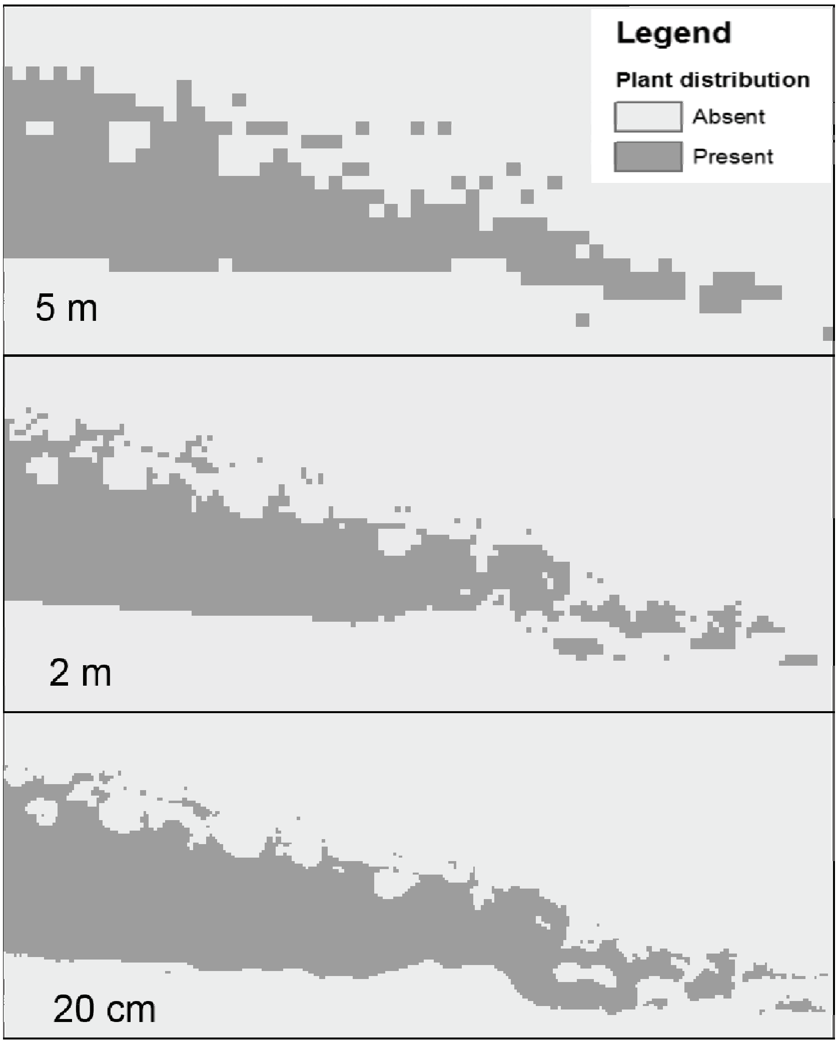


Figure 5.3 Maps of *Mentha aquatica* with different raster resolutions used

5.8. Short summary

This study has shown a few interesting features connected to plant ecology.

The model performance for narrow ranging plants was much better than for those with broader distribution. This is due to the narrow distributions being more predictable, since they require specific factors to be present. The plants with broader niche have more general requirements that are more difficult for modelling.

Another good insight of this study is the effect of the changes in predictor cell size on the model performance. It might be linked directly to the ecology of the species investigated here, as well as to the accuracies of the predictor variables that might change with resolution providing a more appropriate explanation for the plant variation or else greater inconsistencies putting the model performance down.

5.9. Possible improvements

To sum up some improvements proposed above.

Grouping the species would improve the presence/absence issue by increasing the number of presences, and also decreasing the prevalence of absences, that is greatly influencing the statistical model.

Grouping the species or selecting some representative ones to use for modelling would also yield some broader habitat borders.

Including biotic interactions within plants would help represent the realised environmental niche with higher accuracy.

Using a more appropriate statistical approach might yield a more accurate output. As for the DEM data, acknowledging the limitations is one of the few things to be done, apart from possibly finding some more suitable explanatory variables.

6. Conclusions and Recommendations

6.1. Conclusions

- The following indirect predictor variables were derived from LiDAR DEM: elevation, cost-distance to the sea, distance to sea water inlet, distance to fresh water seepage and slope. These variables were used as proxies for environmental factors influencing vegetation distribution of a green beach.
- The correlation between the predictors and the dependant data (plant presence/absence data) varied between species, according to the plants environmental preferences.
- Logistic regression modelling was used to derive the empirical relation between predictors and plant distribution. Only 7 out of 25 species models yielded significant results. This depended on the range of the original species distribution. Some characteristics of the data also had negatively influenced the modelling output.
- In general, given some improvements are made, modelling green beach species distribution with DEM-derived predictors could yield more acceptable results.
- Changing the predictor cell size has some effect on the modelling performance. The changes vary between species. Some species models show gradual or abrupt increase in the performance, some have decline in the performance, and some peak at specific cell size. These effects are thought to be connected to the ecology and scale of the plants distribution patterns, and to the effect on resolution increase on the predictors' accuracy.
- The assessment of the prediction maps for species distribution has produced poor results.
- The retrospective mapping of 2006 plant distribution performed with the predictors of that time and empirical relation derived from the present data has not shown high accuracy. This is to some extent due to errors in the positioning accuracy of the validation data. However, the assumption that the empirical relationship between vegetation distribution on the green beach and indirect predictors derived from elevation is questionable.

The following hypotheses were proved in this study:

- There is significant relationship between the plants distribution and the DEM-derived predictor variables.
- The decrease in raster cell size of the predictors has significant influence on the accuracy of the output.
- The empirical relation between topography and vegetation composition does not remain constant within the period from 2006 and 2010.

6.2. Recommendations

- It would be useful to find ways of improving or deriving new explanatory variables that would yield higher correlations with the plant distributions.
- To improve the general performance of the statistical model the problem with absences in the dataset needs to be accounted for. This can be done by grouping the plants occurring in similar conditions to increase the total number of true presences.
- Another option is to use a statistical model that requires presence-only data.
- Trying out the same approach with several statistical tools would also give some insight on the cell size impact on the species distribution model performance.
- The time limitation did not let more sampled species to be modelled. Modelling more plants' distribution and assessing the cell size impact would be good for examining the consistency of present results.
- Reassessing the models with the changes set of predictors would bring more light to why they are changed with the cell size decrease.

List of abbreviations

DEM – digial elevation model
GDP – gross national product
GPS – global positioning systems
GAM – generalised additive model
GLM – generalised linear model
e.g. – (*exempli gratia*, Lat.) for example
etc. – (*et cetera*, Lat.) and other things
ENFA - ecological niche factor analysis
LiDAR – light detection and ranging
NAP – Dutch ordnance level, Rijkswaterstaat
SDM – species distribution model
SPSS – *'Statistical Package for the Social Sciences'*, name of statistical software
p/a – presence/absence
VIF – variance inflation factor

References

Allen, M. J. 1932. Recent changes in the sea-beach flora at Ainsdale, Lancashire. *Northwestern Naturalist,* 24.

Araujo, M. B., Thuiller, W., Williams, P. H. & Reginster, I. 2005. Downscaling European species atlas distributions to a finer resolution: implications for conservation planning. *Global Ecology and Biogeography,* 14**,** 17-30.

Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling,* 157**,** 101-118.

Beukeboom, T. J. 1976. *Hydrology of the Frisian Islands.* Rodopi.

Bockelmann, A. C., Bakker, J. P., Neuhaus, R. & Lage, J. 2002. The relation between vegetation zonation, elevation and inundation frequency in a Wadden Sea salt marsh. *Aquatic Botany,* 73**,** 211-221.

Campbel, R. C. 1989. *Statistics for biologists,* Cambridge, Cambridge University Press.

CBS 2009. StatLine Database. Central Bureau for Statistics.

CBS. 2011. *Central Bureau for Statistics* [Online]. Available: http://www.cbs.nl [Accessed 17.02 2011].

Chatterjee, S. & Hadi, A. S. 2006. *Regression Analysis by Example*.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement,* 20**,** 37-46.

Cohen, J. 1968. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin,* 70**,** 213-220.

Edmondson, S. E., Traynor, H. & McKinnell, S. 2001. The development of a green beach on the Sefton coast, Merseyside, UK. *In:* Houston, J. A., Edmondson, S. E. & P.J., R. (eds.) *Coastal Dune Management. Shared experience of European Conservation Practice.* Liverpool: Liverpool University Press.

Edward H. Isaaks, R. M. S. 1989. *Applied Geostatistics,* New York, Oxford University Press.

Fielding, A. H. & Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation,* 24**,** 38-49.

Flood, M. 2001. Laser altimetry: From science to commercial lidar mapping. *Photogrammetric Engineering and Remote Sensing,* 67**,** 1209-+.

GIS-Centre 2008. *Geograpical Information Science - An Introduction,* Lund, Lund University.

Gribok, D. 2007. GCALC, GeodeticCalculator.

Guisan, A., Edwards, T. C. & Hastie, T. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling,* 157**,** 89-100.

Guisan, A., Graham, C. H., Elith, J., Huettmann, F. & Distri, N. S. 2007. Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions,* 13**,** 332-340.

Guisan, A. & Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters,* 8**,** 993-1009.

Guisan, A. & Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling,* 135**,** 147-186.

Hegel, T. M., Cushman, S. A., Evans, J. & Huettmann, F. 2010. Current State of the Art for Statistical Modelling of Species Distributions. *In:* Cushman, S. A. & Huettmann, F. (eds.) *Spatial Complexity, Informatics, and Wildlife Conservation.* Springer Japan.

Herrick, J., Van Zee, J., Havstad, K., Burkett, L. & Whitford, W. 2005 *Monitoring Manual for Grassland, Shrubland, and Savanna Ecosystems. Volume II: Design, Supplementary Methods, and Interpretation.,* Tucson, Arizona, University of Arizona Press.

Hickey, D. & Bruce, E. 2010. Examining Tidal Inundation and Salt Marsh Vegetation Distribution Patterns using Spatial Analysis (Botany Bay, Australia). *Journal of Coastal Research,* 26**,** 94-102.

Hiller, R. & Roelse, P. 1995. Dynamic preservation of the coastline in the Netherlands. *Journal of Coastal Conservation***,** 12.

Hirzel, A. H., Hausser, J., Chessel, D. & Perrin, N. 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology,* 83**,** 2027-2036.

Hosmer, D. W. & Lemeshow, S. 2004. *Applied logistic regression,* Interscience.

Jansen, E. 2010. *RE: Birdlife on Schiermonnikoog green beach.*

Johnson, C. M., Johnson, L. B., Richards, C. & Beasley, V. 2002. *Predicting the occurrence of amphibians: An assessment of multiple-scale models,* Washington, Island Press.

Kendall, M. G. 1938. A new measure of rank correlation. *Biometrika,* 30**,** 81-93.

Koppenaal, E. C. 2007. *Green Beaches; vegetation and abiotic conditions.* Master, University of Groningen.

Liu, X. 2008. Airborne LiDAR for DEM generation: some critical issues. *Progress in Physical Geography,* 32**,** 31-49.

Manel, S., Dias, J.-M. & Ormerod, S. J. 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling,* 120**,** 337-347.

Manel, S., Williams, H. C. & Ormerod, S. J. 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology,* 38**,** 921-931.

McPherson, J. M., Jetz, W. & Rogers, D. J. 2006. Using coarse-grained occurrence data to predict species distributions at finer spatial resolutions-possibilities and limitations. *Ecological Modelling,* 192**,** 499-522.

Meijden, R. V. D. 2005. *Heukels' flora van Nederland,* Groningen, Wolters-Noordhoff.

Moore, D. S. 2006. *Introduction to the practice of statistics,* New York, Freeman.

Morrison, M. L., Marcot, B. G. & Mannan, R. W. 1992. *Wildlife Habitat Relationships. Concepts and Applications.,* Madison, WI, USA, University Wisconsin Press.

Petersen, J. & Pott, R. 2005. *East Frisian islands, changing landscape and vegetation (in German).* Hannover, Schlütersche verlagsgesellschaft mbH & Co.KG

Phillips, S. J., Anderson, R. P. & Schapire, R. E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling,* 190**,** 231-259.

Rijkswaterstaat 2010a. Getijdentabel.

Rijkswaterstaat 2010b. Metadata information for DEM data.

Scherfose, V. 1987. Salz-Zeigerwerte von Gefässpflanzen der Salzmarschen, Tideröhrichte und Salzwasser Tümpel an der Deutchen Nord- und Ostseeküste. *Jahresberichte Forschungsstelle Küste*

Schiermonnikoog, N. 2011. *Nature and scenery of Schiermonnikoog* [Online]. Available: http://www.nationaalpark.nl/schiermonnikoog/ [Accessed 10.02 2011].

Schoeman, P. K. 2006. EUROSION. Case Study - Wadden Sea Islands (The Netherlands). *In:* MINISTRY OF TRANSPORT, P. W. A. W. M. (ed.). The Hague.

Smith, P. H. 2001. Vascular plants and vegetation of the Birkdale Green Beach. Sefton Coast Partnership.

Smith, P. H. 2006. The Ecology of Birkdale Green Beach. *British wildlife,* 19**,** 11-16.

Spearman, C. 1987. The Proof and Measurement of Association between Two Things. *The American Journal of Psychology,* 100**,** 441-471.

SPSS 2010. PASW Statistics user tutorial. IBM Corporation.

Tamis, W. L. M., R. van der Meijden, J. Runhaar, R.M. Bekker, W.A. Ozinga & B. Odé, I. H. 2004. Standaardlijst van de Nederlandse flora. *Gorteria***,** 101-195.

Thomas, K., Keeler-Wolf, T. & Franklin, J. 2002. *A comparison of fine- and coarse-resolution environmental variables toward predicting vegetation distribution in the Mojave Desert,* Washington, Island Press.

Tobalske, C. 2002. *Effects of spatial scale on the predictive ability of habitat models for the green woodpecker in Switzerland,* Washington, Island Press.

Wikipedia. 2010. *Schiermonnikoog* [Online]. Available: http://en.wikipedia.org/wiki/Schiermonnikoog [Accessed 10.02 2011].

Appendix A. List of species found during field work

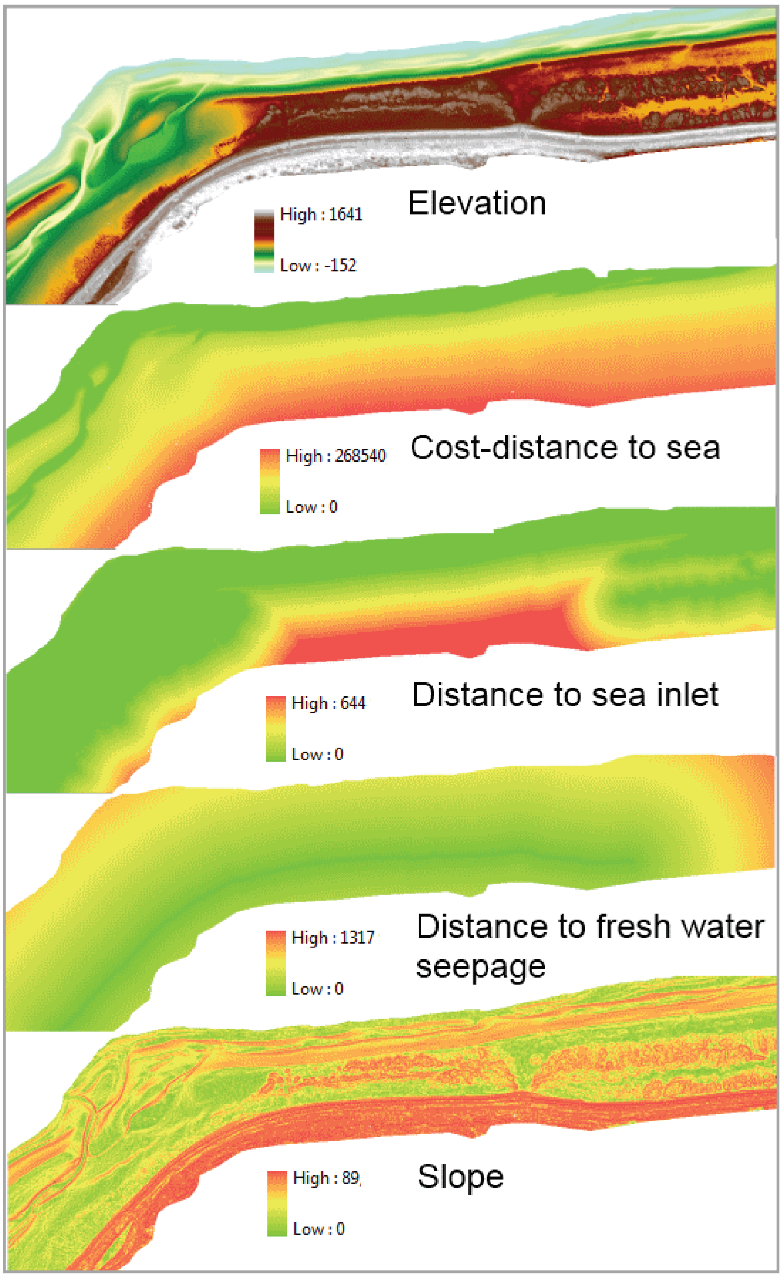
The list is given with the number of records and salinity values for plants (after Koppenaal, 2007, Scherfose, 1987) and indicating Red list species (Tamis et al., 2004).
Two names are stated, when there is possibility of error in species recognition.

No	Name	Number of records	Salt value	Red list
1	*Agrostis stolonifera*	3956	Brackish	
2	*Ammophila arenaria*	40	Fresh	
3	*Aster tripolium*	144	Saline	
4	*Artemisia maritima*	86	Saline	
5	*Atriplex littoralis/prostrata*	6	Saline	
6	*Calamagrostis epigejos*	23	Fresh	
7	*Carex arenaria*	21	Fresh	
8	*Carex distans*	238	Brackish	
9	*Carex extensa*	1569	Brackish	
10	*Carex nigra*	16	Fresh	
11	*Carex oederi*	98	Brackish	
12	*Carex panicea*	1	Fresh	
13	*Centaurium pulchellum*	123	Brackish	
14	*Eleocharis uniglumis*	493	Brackish	
15	*Epipactis palustris*	4	Fresh	x
16	*Elymus pycnanthus*	859	Brackish	
17	*Festuca rubra*	1660	Saline	
18	*Glaux maritima*	1191	Saline	
19	*Hippophae rhamnoides*	115	Fresh	
20	*Holcus lanatus*	29	Fresh	
21	*Hydrocotyle vulgaris*	23	Fresh	
22	*Juncus alpinoarticulatis*	132	Brackish	
23	*Juncus articulatus*	57	Brackish	
24	*Juncus cf. bulbosus*	26	Brackish	
25	*Juncus gerardii*	595	Saline	
26	*Juncus maritimus*	678	Saline	
27	*Leontondon autumnalis*	1036	Brackish	
28	*Linum catharticum*	225	Fresh	x
29	*Limonium vulgare*	38	Saline	
30	*Lythrum salicaria*	229	Fresh	
31	*Lotus corniculatus*	37	Brackish	
32	*Mentha aquatica*	126	Brackish	
33	*Odontites vernus*	367	Brackish	x
34	*Parnassia palustris*	89	Fresh	x
35	*Plantago coronopus*	281	Brackish	
36	*Plantago maritima*	359	Saline	x
37	*Potentilla anserina*	676	Brackish	
38	*Puccinellia maritima*	678	Saline	
39	*Phragmites australis*	606	Brackish	

No	Name	Number of records	Salt value	Red list
40	*Sagina nodosa*	198	Fresh	x
41	*Salicornia europaea*	250	Saline	
42	*Salix repens*	476	Fresh	
43	*Samolus valeriandi*	145	Brackish	
44	*Schoenus nigricans*	223	Fresh	
45	*Scirpus maritimus*	1210	Brackish	
46	*Scirpus lacustris* subsp. *tabernaemontani*	89	Brackish	
47	*Sedum acre*	17	Fresh	
48	*Sonchus maritimus*	74	Brackish	
49	*Spartina anglica*	84	Saline	
50	*Spergularia maritima*	50	Saline	
51	*Suaeda maritima*	2	Saline	
52	*Trifolium arvense*	3	Fresh	
53	*Trifolium repens/dubium*	52	Fresh	
54	*Triglochin maritima*	112	Saline	

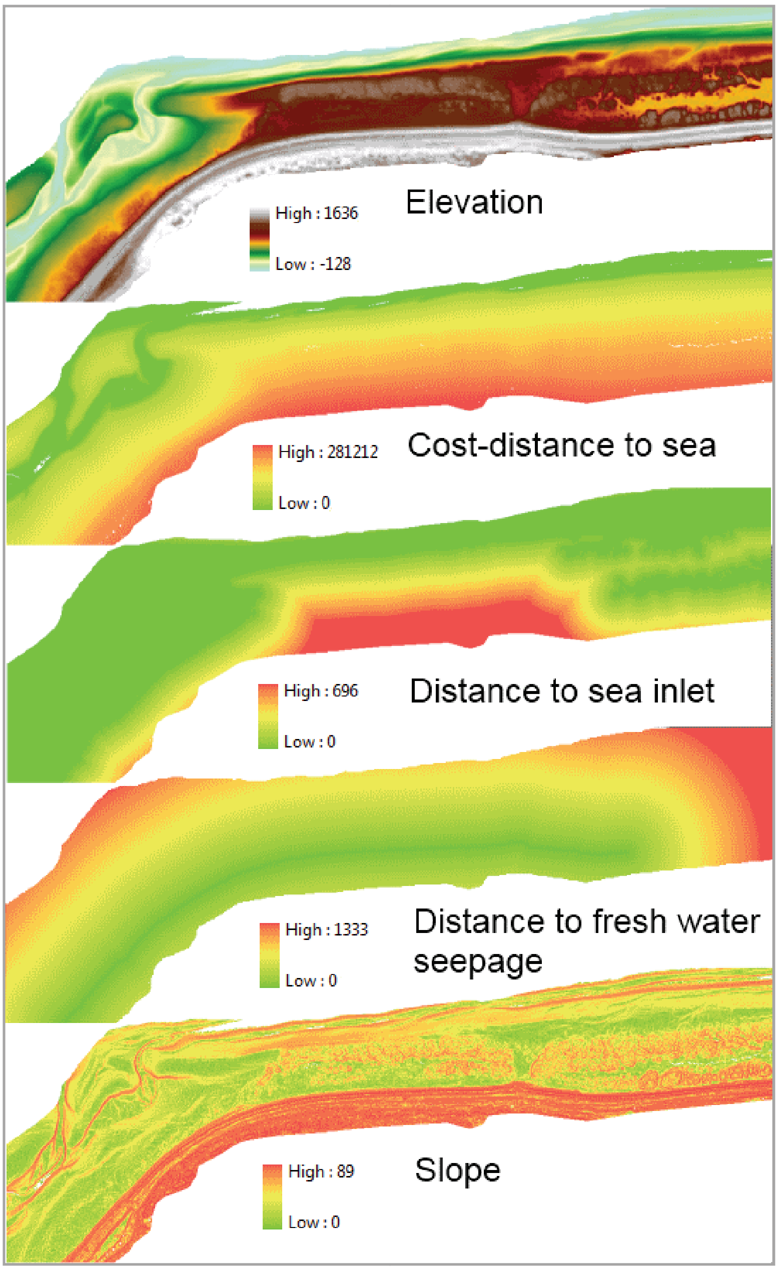
Appendix B.1. Explanatory variables, 2010

The units are Elevation, Dist. to sea inlet, Dist. to fresh water seepage – m, Slope – degrees.



Appendix B.2. Explanatory variables, 2006

The units are Elevation, Dist. to sea inlet, Dist. to fresh water seepage – m, Slope – degrees.



Appendix C. Correlation coefficients

Correlation coefficients between predictors and dependant variables
Red list plants are marked in bold.

Plant name	Correlation test	Elevation	Cost-distance to sea	Distance to inlet	Distance to fresh water seepage	Slope
Agrostis stolonifera	Kendall's tau	,248**	,065**	,288**	-,243**	-,077**
	Spearman's rho	,302**	,080**	,344**	-,297**	-,093**
Aster trifolium	Kendall's tau	-,112**	-,006	-,130**	,035**	-,024
	Spearman's rho	-,136**	-,008	-,155**	,043**	-,029
Atemisia maritima	Kendall's tau	-,088**	-,130**	-,099**	,062**	,017
	Spearman's rho	-,107**	-,159**	-,119**	,075**	,020
Carex distans	Kendall's tau	,090**	,075**	,093**	-,101**	-,013
	Spearman's rho	,110**	,092**	,111**	-,123**	-,016
Carex extensa	Kendall's tau	,096**	,153**	,182**	-,082**	-,010
	Spearman's rho	,117**	,187**	,217**	-,100**	-,013
Carex oederi	Kendall's tau	-,069**	,095**	-,080**	,115**	-,006
	Spearman's rho	-,084**	,116**	-,095**	,140**	-,007
Centaurium pulch/lit	Kendall's tau	,007	,074**	,017	,034**	,065**
	Spearman's rho	,008	,090**	,020	,042**	,079**
Eleocharis uniglumis	Kendall's tau	,091**	,066**	,144**	-,098**	-,045**
	Spearman's rho	,110**	,081**	,172**	-,120**	-,055**
Elymus pycnanthus	Kendall's tau	-,030*	-,096**	-,088**	-,134**	,001
	Spearman's rho	-,036*	-,117**	-,105**	-,163**	,001
Festuca rubra	Kendall's tau	,064**	,162**	,074**	-,086**	,099**
	Spearman's rho	,077**	,198**	,089**	-,104**	,121**
Glaux maritima	Kendall's tau	-,117**	-,090**	-,146**	,124**	,006
	Spearman's rho	-,142**	-,109**	-,175**	,151**	,007
Hipophae rhamnoides	Kendall's tau	,074**	,082**	,069**	-,006	,130**
	Spearman's rho	,090**	,100**	,082**	-,007	,158**
Juncus alpinoarticulatis	Kendall's tau	,051**	,076**	,102**	-,006	,051**
	Spearman's rho	,062**	,093**	,122**	-,008	,062**
Juncus articulatus	Kendall's tau	,039**	,004	,035**	-,042**	-,002
	Spearman's rho	,047**	,005	,042**	-,052**	-,002
Juncus cf. bulbosus	Kendall's tau	-,034**	-,045**	-,041**	-,056**	-,013
	Spearman's rho	-,041**	-,056**	-,049**	-,068**	-,015
Juncus gerardii	Kendall's tau	,077**	,083**	,042**	,051**	-,050**
	Spearman's rho	,094**	,101**	,050**	,062**	-,061**
Juncus maritimus	Kendall's tau	,124**	,007	,127**	-,019	-,097**
	Spearman's rho	,150**	,008	,152**	-,023	-,119**

Plant name	Correlation test	Elevatio n	Cost-distance to sea	Distance to inlet	Distance to fresh water seepage	Slope
Leontondon autumnalis	Kendall's tau	,167**	,295**	,306**	-,045**	,070**
	Spearman's rho	,203**	,361**	,366**	-,055**	,086**
Linum catharticum	Kendall's tau	,120**	,110**	,139**	-,074**	,017
	Spearman's rho	,146**	,134**	,166**	-,091**	,020
Limonium vulgare	Kendall's tau	-,044**	,041**	-,057**	,074**	-,018
	Spearman's rho	-,054**	,051**	-,068**	,090**	-,022
Lythrum salicaria	Kendall's tau	,105**	,079**	,156**	-,125**	-,032**
	Spearman's rho	,128**	,096**	,186**	-,153**	-,039**
Lotus corniculatus	Kendall's tau	,071**	,013	,046**	-,004	,015
	Spearman's rho	,087**	,016	,055**	-,005	,018
Mentha aquatica	Kendall's tau	,131**	,068**	,141**	-,137**	-,078**
	Spearman's rho	,159**	,084**	,169**	-,167**	-,095**
Odontitis vernus	Kendall's tau	,087**	,089**	,137**	-,055**	,007
	Spearman's rho	,106**	,108**	,164**	-,067**	,009
Parnassia palustris	Kendall's tau	,010	,040**	,028*	-,049**	,042**
	Spearman's rho	,012	,049**	,034*	-,060**	,051**
Plantago coronopus	Kendall's tau	,035**	,130**	,080**	,051**	,092**
	Spearman's rho	,043**	,159**	,096**	,062**	,113**
Plantago maritima	Kendall's tau	-,132**	,113**	-,128**	,177**	-,032**
	Spearman's rho	-,161**	,138**	-,153**	,215**	-,039**
Potentilla anserina	Kendall's tau	,160**	,061**	,197**	-,113**	-,077**
	Spearman's rho	,195**	,074**	,236**	-,138**	-,094**
Puccinellia maritima	Kendall's tau	-,312**	-,219**	-,337**	,158**	,016
	Spearman's rho	-,379**	-,267**	-,403**	,193**	,020
Phragmitis australis	Kendall's tau	,186**	,005	,110**	,024	-,012
	Spearman's rho	,225**	,007	,132**	,029	-,014
Sagina nodosa	Kendall's tau	,056**	,073**	,033**	-,019	,104**
	Spearman's rho	,067**	,090**	,040**	-,023	,127**
Salicornia europea	Kendall's tau	-,212**	-,047**	-,214**	,211**	-,024*
	Spearman's rho	-,257**	-,057**	-,256**	,257**	-,030*
Salix repens	Kendall's tau	,208**	,147**	,236**	-,098**	,002
	Spearman's rho	,252**	,180**	,282**	-,120**	,003
Samolus valeriandi	Kendall's tau	-,001	-,051**	,017	-,095**	-,066**
	Spearman's rho	-,001	-,062**	,020	-,116**	-,081**
Schoenus nigricans	Kendall's tau	,162**	,076**	,136**	-,129**	-,034**
	Spearman's rho	,196**	,092**	,162**	-,157**	-,041**
Scirpus maritimus	Kendall's tau	-,159**	-,291**	-,229**	-,224**	-,046**
	Spearman's rho	-,193**	-,355**	-,274**	-,273**	-,056**
Scirpus tabernaemo ntani	Kendall's tau	-,061**	-,088**	-,051**	-,026*	-,033**
	Spearman's rho	-,075**	-,107**	-,061**	-,032*	-,040**
Sonchus	Kendall's tau	,008	-,026*	-,009	-,054**	-,021

Plant name	Correlation test	Elevatio n	Cost-distance to sea	Distance to inlet	Distance to fresh water seepage	Slope
maritimus	Spearman's rho	,009	-,032*	-,011	-,065**	-,026
Spatina anglica	Kendall's tau	-,123**	-,114**	-,114**	,017	-,005
	Spearman's rho	-,149**	-,139**	-,137**	,021	-,006
Spergularia maritima	Kendall's tau	-,071**	-,057**	-,078**	,062**	,012
	Spearman's rho	-,086**	-,070**	-,093**	,076**	,015
Trifolium repens/dubiu m	Kendall's tau	,047**	,073**	,079**	-,054**	,027*
	Spearman's rho	,057**	,090**	,094**	-,065**	,033*
Triglochen maritima	Kendall's tau	-,064**	,021	-,066**	,046**	-,008
	Spearman's rho	-,078**	,026	-,079**	,057**	-,010

**. Correlation is significant at the 0.01 level.

*. Correlation is significant at the 0.05 level.

Appendix D.1. Parameters and assessment values for models with unchanged predictors

Elev – elevation variable, Sea dist – cost-distance to the sea, Inl dist – distance to sea inlet, Fr wat dist – distance to fresh water seepage, C&S R^2 – Cox & Snell R^2, Nagelk. R^2 - Nagelkerke R^2

Cell size	Const.	Elev	Sea dist	Inl dist	Fr wat dis	Slope	C&S R^2	Nagelk. R^2	Hosm-Lem	df	Sign
Agrostis stolonifera											
5 m	0	0,0016	0	0,0029	-0,0034	-0,0144	0,136	0,182	136,347	8	1,36E-25
2 m	0	0,0033	0	0,0030	-0,0031	-0,0143	0,137	0,183	123,432	8	6,48E-23
20 cm	0	0,0035	0	0,0030	-0,0032	-0,0141	0,136	0,182	118,096	8	8,19E-22
Aster tripolium											
5 m	0	-0,0195	0	-0,0112	0,0018	0	0,706	0,941	44,392	8	4,80E-07
2 m	0	-0,0195	0	-0,0110	0,0018	0	0,706	0,941	38,536	8	5,99E-06
20 cm	0	-0,0195	0	-0,0109	0,0018	0	0,706	0,941	43,826	8	6,14E-07
Carex extensa											
5 m	-2,5500	0	2,50E-05	0	-0,0027	-0,0066	0,064	0,1	64,074	8	7,36E-11
2 m	-2,5543	0	2,50E-05	0	-0,0027	-0,0078	0,065	0,101	66,504	8	2,43E-11
20 cm	-2,4660	0	2,53E-05	0	-0,0028	-0,0098	0,067	0,104	74,697	8	5,67E-13
Festuca rubra											
5 m	0	-0,0152	3,01E-05	-0,0022	-0,0036	0,0109	0,317	0,422	16,926	8	0,031
2 m	0	-0,0164	3,09E-05	-0,0021	-0,0036	0,0133	0,32	0,426	23,942	8	0,002
20 cm	0	-0,0168	3,08E-05	-0,0020	-0,0036	0,0142	0,321	0,428	16,623	8	0,034
Glaux maritima											
5 m	-0,6316	0	-1,96E-05	0	0,0024	0	0,032	0,055	78,493	8	9,82E-14
2 m	-0,6343	0	-1,96E-05	0	0,0024	0	0,032	0,054	72,81	8	1,35E-12
20 cm	-0,6391	0	-1,95E-05	0	0,0024	0	0,032	0,054	71,807	8	2,14E-12
Juncus alpinoarticulatus											
5 m	-8,4134	0	2,30E-05	0,0076	0	0,0268	0,022	0,139	9,16	8	0,329
2 m	-8,2277	0	2,23E-05	0,0074	0	0,0216	0,021	0,134	13,524	8	0,095
20 cm	-8,1895	0	2,26E-05	0,0072	0	0,0199	0,021	0,13	6,668	8	0,573
Juncus maritimus											
5 m	0	-0,0048	-2,34E-05	0,0074	0	-0,0301	0,555	0,741	65,139	8	4,53E-11
2 m	0	-0,0055	-2,36E-05	0,0073	0	-0,0187	0,552	0,736	66,743	8	2,18E-11
20 cm	0	-0,0051	-2,38E-05	0,0073	0	-0,0187	0,552	0,736	47,706	8	1,13E-07
Linum catharticum											
5 m	-15,4204	0,0366	4,80E-05	0,0025	-0,0062	0	0,045	0,202	6,094	8	0,637
2 m	-18,7292	0,0506	5,15E-05	0,0021	-0,0054	0	0,047	0,211	12,351	8	0,136
20 cm	-18,8181	0,0508	5,16E-05	0,0021	-0,0052	0	0,047	0,212	12,404	8	0,134
Mentha aquatica											
5 m	0	-0,0308	0	0,0121	-0,0576	-0,0478	0,715	0,953	7,78	8	0,455
2 m	0	-0,0140	0	0,0110	-0,0595	-0,0321	0,715	0,953	8,103	8	0,423
20 cm	0	-0,0146	0	0,0109	-0,0605	-0,0224	0,714	0,953	4,383	8	0,821
Phragmites australis											
5 m	-10,9074	0,0478	-1,77E-05	0	0,0012	-0,0114	0,056	0,129	47,739	8	1,11E-07
2 m	-12,4030	0,0570	-2,29E-05	0	0,0019	-0,0136	0,063	0,144	34,592	8	3,17E-05
20 cm	-11,9670	0,0545	-2,19E-05	0	0,0017	-0,0114	0,059	0,137	48,347	8	8,48E-08
Cell size	Const.	Elev	Sea dist	Inl dist	Fr wat dis	Slope	C&S R^2	Nagelk. R^2	Hosm-Lem	df	Sign

Cell size	Const.	Elev	Sea dist	Inl dist	Fr wat dis	Slope	C&S R^2	Nagelk. R^2	Hosm-Lem	df	Sign
Plantago maritima											
5 m	-4,0861	0	3,36E-05	-0,0118	0	-0,0122	0,076	0,228	26,139	8	0,001
2 m	-4,0066	0	3,31E-05	-0,0121	0	-0,0111	0,076	0,227	16,185	8	0,04
20 cm	-4,0522	0	3,27E-05	-0,0119	0	-0,0090	0,075	0,224	21,792	8	0,005
Puccinellia maritima											
5 m	0	0,0051	-5,61E-05	-0,0283	0,0055	0,0000	0,622	0,829	67,758	8	1,3715-11
2 m	0	0,0096	-6,37E-05	-0,0001	0,0061	-0,0132	0,623	0,83	91,705	8	2,10E-16
20 cm	0	0,0116	-6,74E-05	-0,0306	0,0064	-0,0180	0,623	0,831	113,887	8	6,04E-21
Salix repens											
5 m	-15,9612	0,0390	3,95E-05	0,0079	0	0	0,103	0,274	10,889	8	0,208
2 m	-18,8340	0,0512	4,30E-05	0,0075	0	0	0,107	0,284	12,649	8	0,125
20 cm	-19,0477	0,0521	4,33E-05	0,0075	0	0	0,108	0,286	8,627	8	0,375
Salicornia europaea											
5 m	5,3723	-0,0341	-1,10E-04	0	0,0158	0	0,124	0,468	5,288	8	0,726
2 m	5,0001	-0,0317	-1,11E-04	0	0,0157	0	0,123	0,465	4,516	8	0,808
20 cm	4,9485	-0,0315	-1,10E-04	0	0,0157	0	0,123	0,465	4,413	8	0,818
Scirpus maritimus											
5 m	0	0,0175	-3,74E-05	-0,0105	-0,0290	0	0,55	0,734	322,799	8	5,74E-65
2 m	0	0,0175	-3,72E-05	-0,0105	-0,0299	0	0,55	0,734	132,042	8	3,09E-80
20 cm	0	0,0175	-3,75E-05	-0,0105	-0,0298	0	0,55	0,733	384,699	8	3,51E-78
Triglochin maritima											
5 m	0	-0,0252	1,76E-05	-0,0047	0	0	0,706	0,941	34,661	8	3,08E-05
2 m	0	-0,0253	1,79E-05	-0,0047	0	0	0,706	0,941	17,889	8	0,022
20 cm	0	-0,0253	1,79E-05	-0,0047	0	0	0,706	0,941	16,875	8	0,031

Appendix D.2. Parameters and assessment values for models with changed predictors

The highlighted cells indicate change in the model parameters. Elev – elevation variable, Sea dist – cost-distance to the sea, Inl dist – distance to sea inlet, Fr wat dist – distance to fresh water seepage, C&S R^2 – Cox & Snell R^2, Nagelk. R^2 - Nagelkerke R^2

Cell size	Const.	Elev	Sea dist	Inl dist	Fr wat dist	Slope	C&S R2	Nagelk. R2	Hosm-Lem	df	Sign
Elymus pycnantus											
5 m	-8,44	4,07E-02	0	-1,29E-02	-4,35E-03	0	0,095	0,187	134,493	8	3,31E-25
2 m	-1,50	8,96E-06	0	-7,60E-03	-5,96E-03	5,44E-03	0,068	0,134	35,798	8	1,91E-05
20 cm	-9,10	4,41E-02	0	-1,35E-02	-4,24E-03	0	0,097	0,19	130,32	8	2,43E-24
Juncus gerardii											
5 m	-9,59	3,54E-02	0	0	3,69E-03	-2,10E-02	0,042	0,099	50,017	8	4,06E-08
2 m	-9,98	3,67E-02	0	0	3,75E-03	-1,48E-02	0,039	0,091	15,003	8	0,059
20 cm	-9,88	3,65E-02	0	0	0	-1,53E-02	0,039	0,092	15,933	8	0,043
Leontodon autumnalis											
5 m	-8,09	0	5,82E-05	6,46E-03	-2,51E-03	1,34E-02	0,204	0,374	107,796	8	1,08E-19
2 m	-10,04	1,04E-02	6,26E-05	5,50E-03	-2,17E-03	0	0,202	0,37	46,838	8	1,65E-07
20 cm	-9,91	9,87E-03	6,24E-05	5,50E-03	-2,19E-03	0	0,201	0,369	46,138	8	2,24E-07
Lythrum salicaria											
5 m	0	-2,12E-02	-1,06E-05	1,46E-02	-0,03	0	0,696	0,928	22,153	8	0,005
2 m	0	-2,24E-02	0	1,26E-02	-0,03	0	0,696	0,928	31,286	8	1,25E-04
20 cm	0	-2,23E-02	0	1,26E-02	-0,03	0	0,696	0,928	35,418	8	2,24E-05
Odontites vernus											
5 m	0	-2,29E-02	1,45E-05	5,68E-03	0	0	0,641	0,855	24,159	8	0,002
2 m	-4,75	0	1,88E-05	0	0	0	0,029	0,092	32,159	8	8,72E-05
20 cm	-4,75	0	1,88E-05	0	0	0	0,029	0,091	32,494	8	7,60E-05
Parnassia palustris											
5 m	0	-3,16E-02	2,88E-05	0	-2,61E-02	3,13E-02	0,724	0,965	5,889	8	0,66
2 m	0	-1,93E-02	0	0	-1,92E-02	0	0,721	0,962	11,984	8	0,151
20 cm	0	-1,92E-02	0	0	-1,93E-02	0	0,721	0,962	10,912	8	0,207
Puccinellia maritima											
5 m	0	5,10E-03	-5,61E-05	-2,83E-02	5,52E-03	0	0,622	0,829	67,758	8	1,3715-11
2 m	0	9,56E-03	-6,37E-05	-6,37E-05	6,10E-03	-1,32E-02	0,623	0,83	91,705	8	2,10E-16
20 cm	0	1,16E-02	-6,74E-05	-3,06E-02	6,35E-03	-1,80E-02	0,623	0,831	113,887	8	6,04E-21
Sagina nodosa											
5 m	-11,28	2,80E-02	2,18E-05	-5,10E-03	-1,57E-03	3,07E-02	0,029	0,127	28,556	8	3,79E-04
2 m	-13,83	4,21E-02	2,22E-05	-6,62E-03	-1,46E-03	2,09E-02	0,029	0,126	27,442	8	0,001
20 cm	-13,50	3,97E-02	1,84E-05	-5,39E-03	0	2,23E-02	0,027	0,117	17,872	8	0,022
Spartina anglica											
5 m	6,37	-5,16E-02	-2,33E-05	-2,71E-02	0	0	0,033	0,295	8,753	8	0,364
2 m	9,85	-6,63E-02	-3,04E-05	0	0	-2,50E-02	0,035	0,312	4,898	8	0,768
20 cm	9,83	-6,30E-02	-3,52E-05	0	0	-3,31E-02	0,036	0,32	2,163	8	0,976

Appendix E. Example of prediction maps for selected plants