Breeding White Storks (*Ciconia ciconia*)
in former East Prussia:
Comparing predicted relative occurrences across scales and time using a stochastic gradient boosting method (TreeNet), GIS and public data

University of Potsdam, Germany
Institute of Biochemistry and Biology

Diploma Thesis in Biology

by

Claudia Wickert

March 2007

Supervisor: Prof. Dr. Dieter Wallschläger, Ecoethology lab, Institute of Biochemistry and Biology, University of Potsdam

Co-Supervisor: PhD Falk Hüttmann, -EWHALE lab- Biology and Wildlife Dept., Institute of Arctic Biology, University of Alaska-Fairbanks

Table of Contents

List of Figures

List of Tables

List of Equations

Abstract

Different habitat models were created for the White Stork *(Ciconia ciconia)* in the region of the former German province of East Prussia (equals app. the current Russian oblast Kaliningrad and the Polish voivodship Warmia-Masuria). Different historical data sets describing the occurrence of the White Stork in the 1930s, as well as selected variables for the description of landscape and habitat, were employed. The processing and modeling of the applied data sets was done with a geographical information system (ArcGIS) and a statistical modeling approach that comes from the disciplines of machine-learning and data mining (TreeNet by Salford Systems Ltd.).

Applying historical habitat descriptors, as well as data on the occurrence of the White Stork, models on two different scales were created: (i) a point scale model applying a raster with a cell size of 1 km^2 and (ii) an administrative district scale model based on the organization of the former province of East Prussia.

The evaluation of the created models show that the occurrence of White Stork nesting grounds in the former East Prussia for most parts is defined by the variables 'forest', 'settlement area', 'pasture land' and 'proximity to coastline'. From this set of variables it can be assumed that a good food supply and nesting opportunities are provided to the White Stork in pasture and meadows as well as in the proximity to human settlements. These could be seen as crucial factors for the choice of nesting White Stork in East Prussia. Dense forest areas appear to be unsuited as nesting grounds of White Storks. The high influence of the variable 'coastline' is most likely explained by the specific landscape composition of East Prussia parallel to the coastline and is to be seen as a proximal factor for explaining the distribution of breeding White Storks.

In a second step, predictions for the period of 1981 to 1993 could be made applying both scales of the models created in this study. In doing so, a decline of potential nesting habitat was predicted on the point scale. In contrast, the predicted White Stork occurrence increases when applying the model of the administrative district scale. The difference between both predictions is to be seen in the application of different scales (density versus suitability as breeding ground) and partly dissimilar explanatory variables. More studies are needed to investigate this phenomenon.

The model predictions for the period 1981 to 1993 could be compared to the available inventories of that period. It shows that the figures predicted here were higher than the figures established by the census. This means that the models created here show rather a capacity of the habitat (potential niche). Other factors affecting the population size e.g. breeding success or mortality have to be investigated further.

A feasible approach on how to generate possible habitat models was shown employing the methods presented here and applying historical data as well as assessing the effects of changes in land use on the White Stork. The models present the first of their kind, and could be improved by means of further data regarding the structure of the habitat and more exact spatially explicit information on the location of the nesting sites of the White Stork. In a further step, a habitat model of the present times should be created. This would allow for a more precise comparison regarding the findings from the changes of land use and relevant conditions of the environment on the White Stork in the region of former East Prussia, e.g. in the light of coming landscape changes brought by the European Union (EU).

Zusammenfassung

In dieser Arbeit wurden verschiedene GIS-basierte Habitatmodelle für den Weißstorch *(Ciconia ciconia)* im Gebiet der ehemaligen deutschen Provinz Ostpreußen (ca. Gebiet der russischen Exklave Kaliningrad und der polnischen Woiwodschaft Ermland-Masuren) erstellt. Zur Charakterisierung der Beziehung zwischen dem Weißstorch und der Beschaffenheit seiner Umwelt wurden verschiedene historische Datensätze über den Bestand des Weißstorches in den 1930er Jahren sowie ausgewählte Variablen zur Habitat-Beschreibung genutzt. Die Aufbereitung und Modellierung der verwendeten Datensätze erfolgte mit Hilfe eines geographischen Informationssystems (ArcGIS) und einer statistisch-mathematischen Methode aus den Bereichen „Machine Learning“ und „Data-Mining“ (TreeNet, Salford Systems Ltd.).

Unter Verwendung der historischen Habitat-Parameter sowie der Daten zum Vorkommen des Weißstorches wurden quantitative Modelle auf zwei Maßstabs-Ebenen erstellt: (i) auf Punktskala unter Verwendung eines Rasters mit einer Zellgröße von 1 km und (ii) auf Verwaltungs-Kreisebene basierend auf der Gliederung der Provinz Ostpreußen in ihre Landkreise.

Die Auswertung der erstellten Modelle zeigt, dass das Vorkommen von Storchennestern im ehemaligen Ostpreußen, unter Berücksichtigung der hier verwendeten Variablen, maßgeblich durch die Variablen ‚forest', ‚settlement area', ‚pasture land' und ‚coastline' bestimmt wird. Folglich lässt sich davon ausgehen, dass eine gute Nahrungsverfügbarkeit, wie der Weißstorch sie auf Wiesen und Weiden findet, sowie die Nähe zu menschlichen Siedlungen ausschlaggebend für die Nistplatzwahl des Weißstorches in Ostpreußen sind. Geschlossene Waldgebiete zeigen sich in den Modellen als Standorte für Horste des Weißstorches ungeeignet. Der starke Einfluss der Variable ‚coastline' lässt sich höchstwahrscheinlich durch die starke naturräumliche Gliederung Ostpreußens parallel zur Küstenlinie erklären.

In einem zweiten Schritt konnte unter Verwendung der in dieser Arbeit erstellten Modelle auf beiden Skalen Vorhersagen für den Zeitraum 1981-1993 getroffen werden. Dabei wurde auf dem Punktmaßstab eine Abnahme an potentiellem Bruthabitat vorhergesagt. Im Gegensatz dazu steigt die vorhergesagte Weißstorchdichte unter Verwendung des Modells auf Verwaltungs-Kreisebene. Der Unterschied zwischen beiden Vorhersagen beruht vermutlich auf der Verwendung unterschiedlicher Skalen und von zum Teil voneinander verschiedenen erklärenden Variablen. Weiterführende Untersuchungen sind notwendig, um diesen Sachverhalt zu klären.

Des Weiteren konnten die Modellvorhersagen für den Zeitraum 1981-1993 mit den vorliegenden Bestandserfassungen aus dieser Zeit deskriptiv verglichen werden. Es zeigt sich hierbei, dass die hier vorhergesagten Bestandszahlen höher sind als die in den Zählungen ermittelten. Die hier erstellten Modelle beschreiben somit vielmehr die Kapazität des Habitats. Andere Faktoren, die die Größe der Weißstorch-Population bestimmen, wie z.B. Bruterfolg oder Mortalität sollten in zukünftige Untersuchungen mit einbezogen werden.

Es wurde ein möglicher Ansatz aufgezeigt, wie man mit den hier vorgestellten Methoden und unter Verwendung historischer Daten wertvolle Habitatmodelle erstellen sowie die Auswirkung von Landnutzungsänderungen auf den Weißstorch beurteilen kann. Die hier erstellten Modelle sind als erste Grundlage zu sehen und lassen sich mit Hilfe weitere Daten hinsichtlich Habitatstruktur und mit exakteren räumlich expliziten Angaben zu Neststandorten des Weißstorches weiter verfeinern. In einem weiteren Schritt sollte außerdem ein Habitatmodell für die heutige Zeit erstellt werden. Dadurch wäre ein besserer Vergleich möglich hinsichtlich erdenklicher Auswirkungen von Änderungen der Landnutzung und relevanten Umweltbedingungen auf den Weißstorch im Gebiet des ehemaligen Ostpreußens sowie in seinem gesamten Verbreitungsgebiet.

1 Introduction

1.1 Objectives

White Storks that nest in the former East Prussia are part of the Masurian-Baltic core population. This region shows a very high density in nesting (more than 10 nesting pairs per km^2 in the overall region and even 30 nesting pairs per 100 km^2 in the Polish section, Schimkat 2006). This finding contrasts the situation of the White Stork in large areas of Western Europe where the number of successfully nesting White Storks has been severely diminished due to human interference into nature by irrigation, monocultures and intensive agriculture (e.g. Dallinga & Schoenmakers 1987, Johst 2001). Thus, the question arises which environmental factors and conditions in East Prussia (in the past and in the present) have made this region such a successful nesting ground for the White Stork? And are there differences within the region or when exploring different scales?

Man has long shown a keen interest in the White Stork (Blotzheim 1987, Creutz 1988). From early on, the annual arrival times in the nesting grounds or the number of nests were recorded. The first large-scale inventory in parts of Poland was exercised as early as 1876 (Profus 2005). In the region of former East Prussia a first inventory was done in 1905 (Braun 1908, Schüz 1933). A first international census of the White Stork population of many European countries took place in 1934 and was repeated almost continuously at decadal intervals. Thus, comprehensive data material on the development of White Stork populations is available which has been extended by a variety of surveys, e.g. on the ecology of food supply (e.g. Böhning-Gaese 1992, Bairlein & Henneberg 2000) or on the migration behavior of the White Stork (e.g. van den Bossche et al. 2002, Berthold et al. 2004, Chernetsov et al. 2004).

An increasing digital availability of data sets (see Hüttmann 2005) on the biotic and abiotic characters of the landscapes, and in connection with GIS and statistical methods provide a significant modeling opportunity of demands species claim of their environment. Only in the last few years powerful methods in statistics were developed which show promising results in their applications, especially in regards to species-habitat-relationships (Elith et al. 2006). These methods originate, in part, from the disciplines of machine-learning and data mining and can even cope with faulty and opportunistic data. That makes them important tools in the interpretation of historic data, for which the method of assuring the data details often can no longer to be completely reconstructed. That is why faults, and a valid inference from such data, are so difficult to assess (Engler et al. 2004, Elith et al. 2006).

In this study, for the first time, a habitat model for the White Stork was created using GIS and advanced modeling techniques. Furthermore, the changes in land use and its effects on the White Stork in the former province of East Prussia were to be investigated. So, the following goals were defined:

- Creation of habitat models applying different historical opportunistic data sets
- Investigation of two different scales (point scale as well as administrative district scale)
- Evaluation of the performance of the applied historic data sets for the ecological modeling
- Inferring important factors on the habitat of the White Stork
- Predicting the distribution of the White Stork for the period 1981 to 1993 based on the models created
- Comparison of the historic models applying the population estimates and the results of available population inventories

Especially in regards to the eastern expansion of the European Union and the admission of the currently Polish part of East Prussia into the European Union these issues gain crucial importance for the protection and conservation of the White Stork populations in Eastern Europe. So it is that Schulz (1999) understands the introduction of modern agricultural methods and the accompanying negative ecological by-products as a threat to the White Stork (see also Tryjanowski et al. 2006). If the repercussions are to be recognized and possibly countervailed, it requires a comprehensive investigation and understanding of the quantitative relationship between the White Stork and the habitat.

1.2 The region of the former East Prussia

1.2.1 Location of the study area

The region of the former Prussian province of East Prussia is situated between the 53rd and the 56th northern latitude, and the 18th and 23rd eastern longitude. In the north it borders on the Baltic Sea and the River Neman ('Memel'), in the West on the Rivers Nogat and Vistula ('Weichsel') and in the East on Lithuania. In the South, the former province stretched to the southern tip of the Masurian Lakes Platteau ('Masurische Seenplatte'). From 1922 to 1939 East Prussia covered an area of 36,992 km^2 having approx. 2.5 million inhabitants in the year 1939 (Barran 1988). The capital of the province was Königsberg.

1.2.2 Brief geographical classification of the natural landscape

The landscape of East Prussia was mainly formed by the inland ice masses of the glacial age. The Baltic coast in the north of East Prussia is shaped by the Vistula Lagoon ('Frisches Haff') and the Courland Lagoon ('Kurisches Haff') with their off-shore spits, covered with dunes as well as the steep coast of the Sambian Peninsula (up to 61m high) positioned between both lagoons. Adjacent, an undulating stretch of land can be found, dispersed by isolated moraine mounds. Further south and almost parallel with the Baltic coastline is the Baltic land ridge with the Masurian Lakes Plateau. On average, it has an altitude of 100 to 150 m above sea level. Individual hill ranges such as the 'Seesker Berge' south of Goldap or the 'Kernsdorfer Höhen' near Osterode exceed 300 m above sea level. The highest elevation is the actual 'Kernsdorfer Höhe' (313 m above sea level). South of the ridge extensive sandy areas stretch out, e.g. 'Johannisburger Heide'. Fertile low lands characterize the country side along the underflows of the three main rivers Nogat, Pregel and Neman.

The climate in East Prussia is continental, having cold winters and dry hot summers. Only on the coast a narrow stretch with oceanic influence can be found. This is reflected in its slightly higher annual precipitation and milder temperatures. The average annual temperature reads 5.9 °C in Treuburg, located at the Masurian Lake Plateau in the southeast of East Prussia, and 7.2 °C in Königsberg located near the coast (Rohde 1957). According to Barran (1988), the average annual precipitation is 500 to 608 mm. On average, the period of vegetation lasts for 220 days per year (Schumacher 1977).

The former province of East Prussia is positioned in the vegetation zone of the green deciduous and mixed (deciduous and coniferous) forests. In 1939, there were approx. 700,000 hectares of forest, more than 540,000 of it was coniferous (Barran 1988). Large forest areas were found on the heights of the land ridge in the southern part of the province. Predominant types of tree species were the Pine tree (*Pinus*) and the Spruce (*Picea*). In the northern part of the country, especially in Sambia (‘Samland’) and along the Courland Lagoon, large forests could be found in which the Alder (*Alnus*) dominated in marshy areas (Rohde 1957).

In the south of the former province of East Prussia soils are found with a high sand content (Schüz 1933); they are occasionally used for some root crops or forestry. Best suited for agriculture and meadows are the soils which are located in the central and in the north of East Prussia (Rohde 1957).

1.2.3 Administration of the region and its changes

In the course of its eventful history, East Prussia experienced several major changes in its administrative boundaries and its nationality, among others, due to wars. A short overview of the region’s history high-lights the difficulties encountered in the definition of the region to be investigated, and in the acquisition of consistent and sound statistical data on the White Stork population.

In the years 1920 to 1939 the province of East Prussia consisted of the four administrative districts Königsberg, Gumbinnen, Allenstein and West Prussia (Figure 1.1). East Prussia belonged to the German Reich but as a consequence of World War I it was isolated from it by the ‘Polish Corridor’ since 1919. After the Second World War, East Prussia was divided into a Russian and a Polish sector. The border ran through the former counties of Heiligenbeil, Preußisch Ehlau, Bartenstein, Gerdauen, Darkehmen and Goldap. The Polish part makes up about 2/3 of the original extent of land area and was awarded to the newly founded voivodships Gdansk (Danzig), Olsztyn (Allenstein) and Suwalki. After the administrative reform of 1975 the Polish East Prussia was divided into the new voivodships Elblag, Olsztyn, Ciechanów and Suwalki. Since the repeated reform of the regions, fixed for the 1st January 1999, the area of the former province East Prussia encompasses nearly the entire voivodship Warmia-Masuria with its capital Olsztyn. The northern part of the former province of East Prussia today forms the Russian oblast Kaliningrad with its capital Kaliningrad. After the disintegration of the Soviet Union this oblast is now a Russian exclave and a free trade zone.

On the 1st of May 2004 Poland joined the European Union and now adheres to the land management of the EU and agricultural policies.

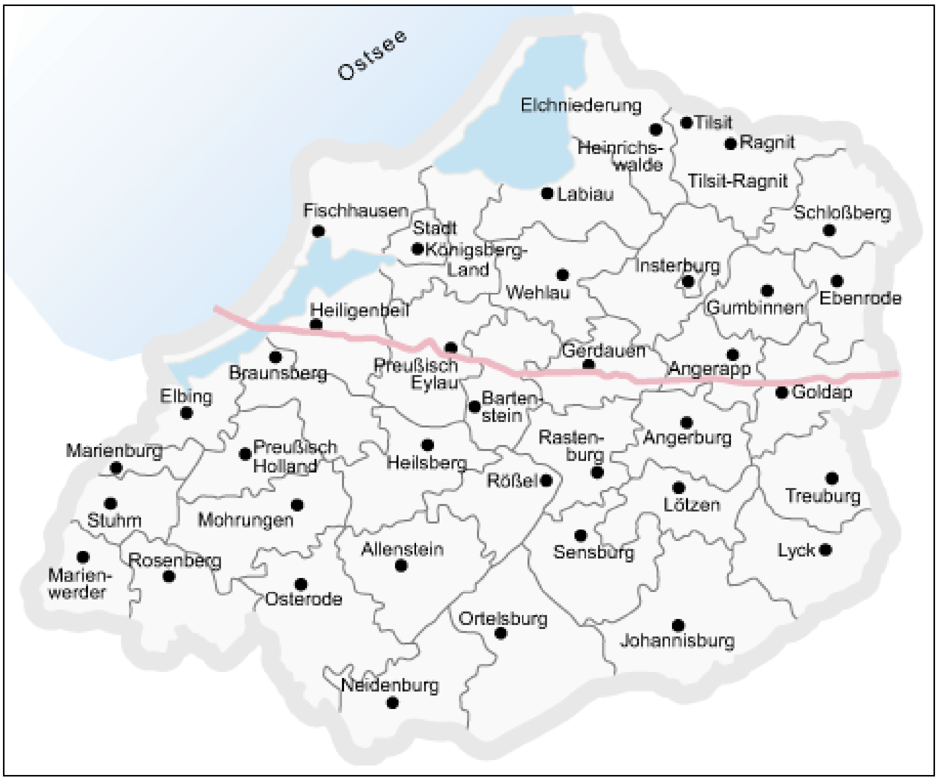

Figure 1.1: East Prussia and its administrative districts as in 1937, the thick line shows the division into a Russian part (north) and a Polish part (south) after World War II (http://www.mpichard.ca/ostpreussen.htm, downloaded March 2007)

1.2.4 Land use and population in the former East Prussia

With the development of the kingdom of Prussia under the reign of Emperor Frederick I at the beginning of the 18th century, East Prussia experienced a cultural and economic boost. Many colonists of German origin and immigrants from Polish and Lithuanian regions as well as religiously motivated refugees from other parts of Europe settled in this region earlier deserted by the plague and warfare, new settlements were founded and compulsory education was introduced (Schumacher 1977). The main sources of income and business were in agriculture and forestry sectors.

The East Prussian agriculture was generally organized in large manorial farms. According to the statistical year book for the German Reich from 1928 (Hertz-Eichenrode 1969) over 70 %

of the farms managed more than 20 hectares of ground, large manors with more than 100 hectares owned 39.2 % of the farmland of East Prussia. The average farm in the German Reich had about half that size with 20.2%. Cattle farming became another important source of income, since it was widely independent of unexpected climatic impairment which traditionally led to a collapse in the crop yield. Thus, more and more cropland was used for the growing of forage plants and as meadows and pastures. Industry concentrated mostly on the processing of agricultural produce, and was only more widely developed in the larger cities of Königsberg, Elbing, Tilsit and Allenstein. Also, the level of urbanization was very low. In 1925, 61% of the population (of East Prussia) lived in communities with less than 2,000 inhabitants compared to 36% in the whole German Reich (Hertz-Eichenrode 1969). Even with the beginning of industrialization little changed due to the lack of natural resources (coal deposits) and the dominance of the agricultural structure. In the year 1933 still more than half the population was occupied in agriculture. In addition to this came the geographical separation of the region from the German Reich after the First World War in 1919, which weakened the competitiveness of the already frail economy through additional costs for transport (Hertz-Eichenrode 1969). For several years after the First World War the crop was lower than in 1914. Agriculture was not further intensified. Fertilization was only rarely applied. In many places the contingent of pastures rose in comparison to the crop land, since the cultivation of grain was too risky due to the poor sales potential. In 1936, 47.2 % of the total area of East Prussia was used as crop land, 20 % as pastures and 19.3 % as forest acreage (Statistisches Handbuch für die Provinz Ostpreußen 1938). So overall, East Prussia remained rurally structured with less intensified agriculture than in the western parts of the German Reich. Highly structured ecosystems which sustained a high abundance of species were conserved (Peterson et al. 1999). After the Second World War East Prussia was divided into a Russian and a Polish part (Figure 1.1) and the original population was nearly completely driven out.

Since then, the Russian part has been named Kaliningrad region. It was nearly completely resettled by people from other Soviet republics. The number of settlements decreased by more than half from 2,500 settlements and approx. 1,000 individual farms in the year 1898 to 1,126 settlements in the year 1989 (Grishanov 1989/90). Most likely that means that the villages were abandoned, but the overall footprint remained in the landscape. Agriculture was led into the system of collective farming (kolkhoz and sovkhoz) and managed according to the Russian requirements of the command economy, whereby the integration into the supplies of the whole of the USSR was of prime importance and no consideration was given to regional demands or special features. From the 1980s onwards, the economy stagnated and attempts to revive it were in vain. With the comedown of the communism and consequential political change in 1989/90 the oblast of Kaliningrad became a Russian exclave. On the base of privatization and structural development the transformation to a free market economy took place (Galcov 1999).

The human population of the Polish part was also nearly completely exchanged after the Second World War. However, no agricultural reform was undertaken as the case in the Soviet Occupied Territory. The manorial farms were socialized, but the old land marks remained in the landscape. After the decline of socialism in Poland in 1990, however, the governmental financial support was completely stopped so that many of these farms faced ruin. Formerly farmed fields were abandoned or altered to pasture, as (Tryjanowski et al. 2005) reported for the Podhale region in Southern Poland for instance.

1.3 The White Stork (*Ciconia ciconia*)

1.3.1 Distribution and habitat of the White Stork

The breeding ground for the European subspecies of the White Stork *Ciconia ciconia ciconia* stretches from western North Africa across Europe right up to Western Asia. The limit of its expansion stretches from Denmark (up to 1938 also southern Sweden; now a region abandoned by White Storks) across the southern coast of the Baltic Sea up to the Finnish Gulf, then south to the Crimea along the Black Sea to Greece, Turkey, Iran and the Iraq. In the west the White Stork breeds on the Iberian Peninsula and in the North African countries of Morocco, Algeria, and in Tunisia south of the Sahara (Blotzheim 1987, Frenz 1995). Originally, this species occurred almost in the entire region described above, but middle Europe between France and West Germany is now only scarcely populated by breeding White Storks (Figure 1.2). Reasons for such declines are discussed e.g. in Creutz (1988).

A subspecies of the White Stork, the Asian White Stork *Ciconia ciconia asiatica*, finds its habitat in Central Asia between Lake Aral and the Province of Sinkiang in Western China (Frenz 1995).

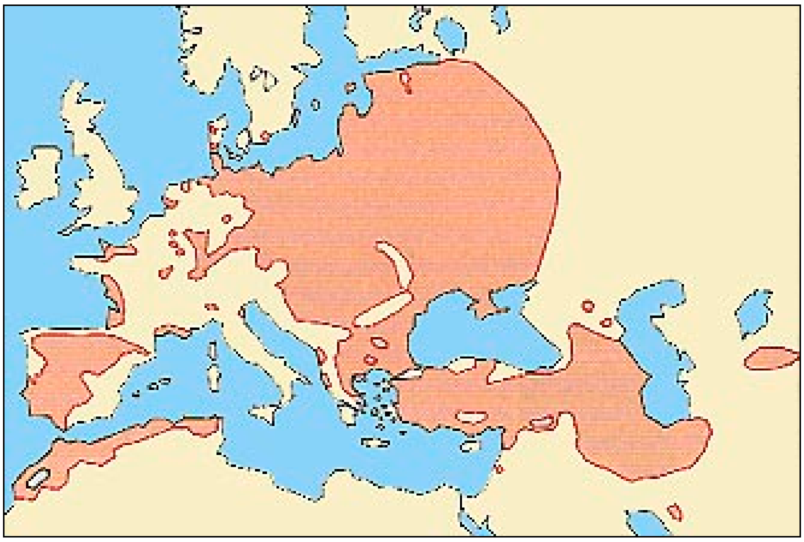

Figure 1.2: Distribution of the White Stork in its breeding ground (http://www.nabu.de/m05/m05_03/01450.html, downloaded July 2006)

In order to forage, the White Stork prefers open habitat with low-growing vegetation and sufficient food supply as to be found e.g. in plain tracts or irrigated terrain. Presently, it makes use of cultivated land such as short-tufted pastures and extensively exploited or fallow fields in its search for food (Blotzheim 1987). Due to its demands on the habitat it occurs mainly in the low lands. However, Creutz (1988) points out a number of breeding occurrences in higher altitudes. Especially in the Near East the White Stork has been traced at an altitude of 2,500 meters. The author comes to the conclusion that the influence of abundance in food is more important than that of the climate.

The White Stork mainly feeds on small mammals such as mice and moles, frogs, reptiles, insects, earthworms and fish. It also preys on carrion (Blotzheim 1987).

In areas densely populated by the White Stork, it often breeds in concentrations resembling colonies. Profus (2005) mentioned an accumulation of 47 pairs of breeding White Storks at the village of Lwowiec in Southern Poland in the year 2004. The nests were originally built on rocks or large solitary trees with strong branches, as it can still be seen in many places in North Africa and Southern Europe. However as a bird typically settling in cultivated areas these days, it usually places its nests, which are often used for several decades, on high-rising buildings, moreover on power pylons and other kinds of poles (Schulz 1993). Often, the White Stork is offered artificial nesting bases on buildings and on trees to serve as a substructure for its nest.

1.3.2 Population assessments in East Prussia for nesting White Storks

The first organized count of the White Stork breeding population in East Prussia took place as early as 1905. It was done by the fauna section of the physical-economical society (faunistische Sektion der physikalisch-ökonomischen Gesellschaft). It showed 13,565 nests occupied in the province of East Prussia in the historic borders of 1905 (Braun 1908). Another count was exercised in the year 1912, however, unfortunately no results were published due to the start of the First World War two years later (Schüz 1933). Another official assessment was carried out in 1931 by the local police force at the order of the Governor Supreme of the Province of East Prussia. In the year 1934 a first international inventory was exercised under the auspices of the ICBP (International Council for Bird Protection) enabling an overview of the population of the White Stork in large areas of Europe. In its entire breeding grounds approx. 46, 000 pairs were counted, 16,600 (=36%) of which in East Prussia alone (see Schüz 1936). In the county of Insterburg the population was surveyed by Hornberger in 1931, and then annually from 1933 to 1942 (Hornberger 1943). Profus (2005) later estimated approx. 8,700 breeding pairs for the Polish region in the year1934; however these numbers were not presented spatially explicit, e.g. with a map. In the 1950s and 1960s the White Stork population apparently declined severely, but revived again by about 1970. Counts repeated within the international White Stork census in the Polish region of former East Prussia took place in 1974 and 1984, and showed approx. 7,600 pairs and 7000 pairs counted. For the present Polish voivodship of Warmia-Masuria, which more or less comprise the Polish part of East Prussia, 8,200 to 8,600 pairs of White Stork were counted in a further international census in the years 1994/95 (Profus 2005). For the northern part of former East Prussia, today's Russian exclave of Kaliningrad, 8,000 pairs of White Stork were additionally assumed for 1934 (Hinkelmann 1995). Population assessments in the years 1974/75 showed a significant decline down to 855 to 1116 breeding pairs and 1255 breeding pairs in the years 1985 to 1989 (Grishanov 1989/90).

Due to repeated alterations in the administrative boundaries exact comparisons are impossible, however, unless nest densities are provided.

1.4 Habitat models

Habitat models help to characterize the relation of a species to its environment; they describe the ecological niche in quantified terms (Guisan & Zimmermann 2000). They can be expressed as spatially explicit static models which describe the condition in a limited area at a specific point in time (e.g. Manly et al. 2002, Schröder & Reineking 2004). The predictive modeling of species distribution is a common technique, applied in the disciplines of ecology, biogeography, evolution, conservation biology and climate change research for instance (Guisan & Thuiller 2005).

With regard to the statistic modeling of the distribution of animal species the term Resource Selection Functions (RSF) is frequently used. The concept of a RSF is defined as any mathematical function that is proportional to the probability of use of an available resource or area by an organism (Austin 2002, Manly et al. 2002). However, there is a larger overlap of this concept with species distribution models developed by plant ecologists (Guisan & Zimmermann 2000, Boyce et al. 2002).

For developing habitat models, presence/absence data is mostly applied as response variables. However, presence-only data (or rather presence/available data) as well as density or relative abundance can be modeled, too. As an output one receives maps of a relative index of occurrence, which depict the statistical probability of appearance or rather the incidence, i.e. the appearance or non-appearance, of the species in delimited homogenous research units (Schröder & Reineking 2004).

The number of habitat predictors to be consulted in the description of species-habitat relations can be very large and dependent on the species to be modeled. To describe a habitat, topographic (elevation, slope, aspect), climatic (precipitation, temperature), biotic (vegetation, predation, competition), anthropogenic factors (population density, land use, fragmentation), the structure of the landscape, etc. can be used. In advance, and based on theoretical and biological considerations, one must decide on these features, which is also influenced by the research design, availability, and sometimes even the convenience in the investigation of the data. Such decisions should be made in accordance to the hypothesis tested and the appropriate research design (Burnham & Anderson 1998, Manly et al. 2002). Depending on the position of the habitat factors in the chain of causation that would link them to the species under investigation, Austin (2002) distinguishes between proximal (close causal correlation) and distal (minor causal correlation) factors. Furthermore, the same author distinguishes resource (nutrients, water, light consumed by the organism), direct (environmental parameters with physiological relevance like temperature, pH) and indirect factors (variables without direct physiological relevance like slope, aspect, elevation). Indirect variables are often more readily measured and can replace combinations of resources and direct variables (Guisan & Zimmermann 2000). Since they often describe the existing patterns well, they are frequently used in modeling. Their disadvantage is that they describe well the relations between species and environment within the region for which the model was developed, in other regions, however, the indirect variables may contain different combinations of resource and direct gradients (Guisan & Zimmermann 2000). Models based on proximal resource and direct gradients are assumed to be among the most robust and widely applicable (Franklin 1995, Austin 2002).

One assumption in statistical habitat modeling is that the species to be modeled is in (pseudo-) equilibrium with the environment or rather that any change shows slow progress in comparison to the life span of the organism (Guisan & Zimmermann 2000, Austin 2002, Manly et al. 2002). Invasive species are an example for species which are usually not in equilibrium with their newly inhabited area, but are still in their dispersal stage (Guisan & Thuiller 2005). Since habitat models are estimated on the basis of distribution survey data of the species in a certain area, as a rule, they can describe merely a realized niche of that species in the specific area, i.e. biotic interactions and competitive exclusion are implicitly incorporated in the models (Guisan & Zimmermann 2000). Underlying processes e.g. competition, migration, predation and even succession for the most part can usually not be quantified, although they sometimes penetrate the model as predictors (see Austin 2002 for examples). This makes the transferability (generalizability) of models to other regions with possibly different biotic interactions for the species more difficult. Austin (2002) however also points out that parts of the distribution area of a species can also be sink-areas (source-sink model see Pulliam 1988), in which the species does not experience optimal habitat qualities and the population can only be sustained by immigration from source areas. In this case, the model would describe a mixture of realized niche and sink areas. Guisan & Thuiller (2005) point out that for the modeling of the distribution of plants it is not sufficient to exploit mere presence/absence data but in fact observations of the species as presence locations are necessary in which the species reproduces and so depicts the habitat as suitable for the preservation of the species.

"Nature is too complex and heterogeneous to be predicted accurately in every aspect of time and space from a single, although complex, model" (Guisan & Zimmermann 2000). Models for the description of species–habitat relations merely offer assumptions on the causal backgrounds to be verified by experimental research (Schröder 2000). Manly et al. (2002) points out that RSFs are correlational. They require detailed experiments to obtain the mechanistic link.

The model building process is usually divided into five steps by (Guisan & Zimmermann 2000): conceptual model formulation, statistical model formulation, model calibration, model prediction and model evaluation. Furthermore, an overview is given showing the multitude of statistical methods which could be exploited in the modeling (see also Segurado & Araujo 2004, Schröder & Reineking 2004, Elith et al. 2006). The mathematical algorithm applied to fit a RSF in this thesis is to be found in the region of data mining, and will be explained more closely in the following chapter. Here model evaluation is to be dealt with more explicitly, since it is an important factor which is often neglected (Fielding & Bell 1997).

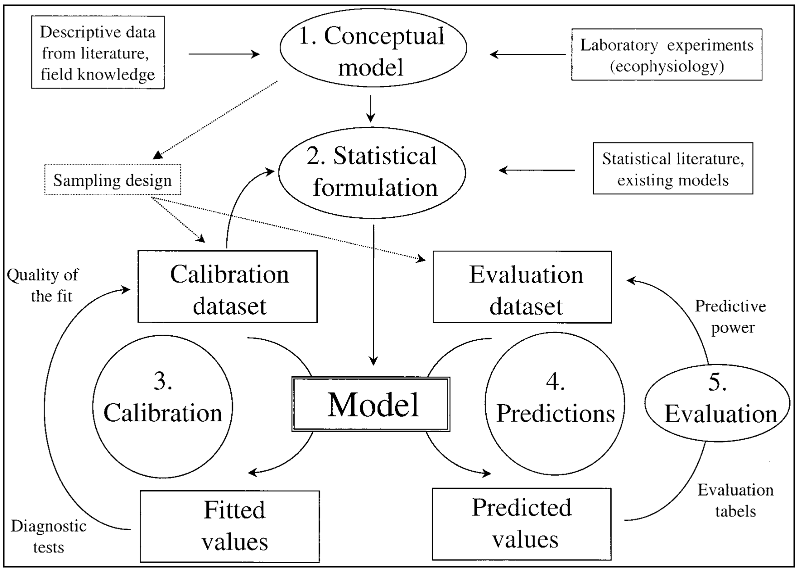

Figure 1.3: Successive steps of the model building process (published in Guisan & Zimmermann 2000)

The evaluation of a model offers an objective assessment of the prediction accuracy as well as a test of the area of validity, i.e. to what degree models can be applied to other areas and other periods of time. Beside the assessment of the adaptation of the model to the training data, the most important quality of the evaluation according to Boyce et al. (2002) is how accurately the model predicts the incidence of organisms. The determination of the accuracy of predictions is directly connected to the error rate (Guisan & Zimmermann 2000), calculated by the comparison of the predicted values with the observed values of a test data set. In a presence/absence model two different errors can occur (Fielding and Bell 1997): (i) prediction of an occurrence with observed non-occurrence (false positive (FP) or Type I error) and (ii) prediction of a non-occurrence with observed occurrence (false negative (FN) or Type II error). The causes might be limitations imposed by the classification algorithm, the data gathering process (algorithmic errors), the actual biological data processing, or the lack of specifying all of the ecologically-relevant processes (biotic errors) (Fielding & Bell 1997). Robust measures for the evaluation of a model apply an independent data set (Fielding & Bell 1997, Pearce & Ferrier 2000), gathered at a different time or in a different area. If only one data set is available for the development and testing of the model, it is usually partitioned into a training (also learning) and a testing (also validation) data set. Especially with small data sets a partitioning usually does not make sense, so that bootstrap or resampling methods as well as cross-validation get applied instead. Fielding & Bell (1997) as well as Guisan & Zimmermann (2000) give further a review of data partitioning methods. However, only with an independent data set can the scope of the model for other regions and periods of time truly be assessed. The evaluation of the model with the same data set than the one used for the development of the model depicts merely the discriminatory power of the model as it concerns occurrence and non-occurrence in the limited area (Guisan & Zimmermann 2000). For the assessment of the performance of a presence-absence model a confusion matrix, widely used in Remote Sensing, is frequently calculated that cross-tabulates the observed and predicted presence/absence patterns (Fielding & Bell 1997, Pearce & Ferrier 2000). If the model offers an occurrence probability as its output, this must be transformed again into presence and absence data applying a threshold value. Taken from the confusion matrix different measures can be calculated (explained, as well as their limitations discussed, in Fielding & Bell 1997). Measures taken from there and applied in the paper at hand are the correct classification rate (share of correctly classified cases in total), sensitivity (share of actual occurrence, correctly classified) and specificity (share of actual non-occurrence correctly classified). They were selected since they are often employed and so can provide a basis for a (limited) comparison of different models. A threshold-independent unbiased discrimination index can be derived from the definition of the Receiver Operating Characteristic (ROC)-Curve (Pearce & Ferrier 2000). For the generation of the curve, the sensitivity for every possible threshold is marked on the y-axis against the reading '1-specificity' on the x-axis. The area under the curve (AUC) serves as the metric of the performance of the model. Readings between 0.5 (prediction of a constant probability) and

1.0 (perfect distinction between occurrence and non-occurrence) can be achieved. Although one should be careful in the interpretation of thresholds in biological applications, Hosmer & Lemeshow (2000) specify the following values for assessing the classification quality of the AUC readings:

$0,7 \leq AUC \leq 0,8$ = acceptable
$0,8 \leq AUC \leq 0,9$ = excellent
$0,9 \leq AUC$ = outstanding.

However, if only presence-only data is available for the evaluation of a model, none of the methods applied by default for the estimation of the quality of the model can be utilized (Guisan et al. 2006a). Boyce et al. (2002) suggests applying a Spearman-Rank Correlation with the area adjusted frequency of the testing data set. But this approach does not offer a chance to compare the performance of different models, and merely shows how well model predictions are related to the probability of occurrence (Pearce & Boyce 2006).

An evaluation method worthwhile to consider is the one carried out in Engler et al. (2004), which is the calculation of the minimal predicted area (MPA). In this method, the predicted value above which 90 % of the occurrences were observed is used as a classification threshold to transform the predicted values for the whole habitat suitability map into presence and absence. The minimal predicted area equals the area of all cells containing the value 1 (presence locations). The MPA can be used for a comparison among different models. A model predicting a relatively small potential area of occurrence while still containing a maximum number of observed species presence locations is considered a good model (Engler et al. 2004).

1.4.1 TreeNet algorithm

For the predictive modeling a machine learning algorithm called TreeNet was used, the software is offered by Salford Systems Ltd. (http://www.salford-systems.com). TreeNet is a tree-based computational method within the realms of data mining, and presents one among the many modeling algorithms in the statistical modeling toolbox (see Elith et al. 2006 for overview). TreeNet reaches among the most powerful algorithms known to date, and it is well known and described in the machine-learning and statistical model community (Hastie et al. 2003). One of its strength is that it is non-parametric and that it can be used for regression as well as classification problems, with continuous and/or categorical predictors. Correlated variables do not present a problem for algorithms like TreeNet neither. The underlying mechanism used for building the model is called stochastic gradient boosting which was developed and is described by Friedman (1999). The behavior of these algorithms is well known, adding confidence to these methods. They prove very powerful and are likely playing a major role in future modeling applications worldwide (Elith et al. 2006, Hubley et al. in press).

A model constructed with TreeNet can consist of a large number (up to several thousand) of trees. The approximation to any function can be written as

$$F(X) = F_0 + \beta_1 T_1(X) + \beta_2 T_2(X) + \dots + \beta_M T_M(X)$$

where each T_i is a small tree (for more detail see Salford Systems Ltd. 2003). The approximation is built up stagewise and every single tree is constructed on the residuals of the previous tree. The individual trees of a model have the same fixed size (predetermined by the user). It depends on the individual dataset and its characteristics which amount of nodes performs best. Sizes usually used by TreeNet experts range from 2 to 12 nodes (Salford Systems Ltd. 2003). Models constructed with only 2-node trees, referred to as 'stumps', involve only one variable per tree and thus do not detect interactions between predictors (Hastie et al. 2003). However, they can be rather useful for finding generalizations in data.

Several mechanisms can help to prevent the learning algorithm to 'overfit' the model to the data which might in many cases not improve the predictive validity of the model, but deteriorate the generality by fitting less relevant esoteric aspects of the training data. This is achieved in the following way: First the dataset is divided into training and testing subsets. The model constructed only on the training data is applied to the test data. At the point where the error rate of the test data (percentage of incorrectly classified observations) re-increases the model process stops. Furthermore, never all the training data is used at any given time. A sub-sample fraction of the training dataset is randomly drawn for every single tree (but other options to sub-sample a training set from the overall data set exist in TreeNet as well). Likewise, usually a tiny learning rate is used (referred to as shrinkage in Friedman 1999, Hastie et al. 2003). This means that the model prediction changes by very small amounts in each training cycle.

TreeNet is often referred to as a 'black-box', and therefore it was widely rejected by biologist. However, there are several advantages of TreeNet compared with other techniques that are often used for building habitat models such as Generalized Linear Models (GLM) or Discriminant Function Analysis (DFA): (i) It automatically selects the important predictor variables thus no prior variable selection or data reduction is required, (ii) the results are invariant with regards to modifications of the data such as transformation or rescaling, (iii) the approach handles missing values automatically and in the best possible way, and (iv) it is immune to outliers in predictors or the target variable, i.e. if samples are coded incorrectly and the model prediction starts to diverge substantially from observed data, that data will not be used in further updates, and (v) it can be learned and applied very quickly by users. TreeNet constructs models convenient and without time-consuming pre-processing of the data. Furthermore, it is remarkably resistant to overfitting. Hastie et al. (2003) calls the approach of multiple additive regression trees, which TreeNet is based on, an effective off-the-shelf procedure for data mining. Further, and because TreeNet can be used in a huge variety of applications beyond data mining, modeling and multiple regressions Salford Systems Ltd. refers to TreeNet as *"the closest tool we have ever encountered to a fully automated statistician"* (Salford Systems Ltd. 2003).

Because of the complexity of the models and the potential non-linear relationship between predictors and response a shortcoming of machine learning algorithms is usually the lack of interpretability. But TreeNet provides a visualization of the influence of every single predictor with so-called partial dependence plots. Partial dependence functions represent the effect of a predictor after accounting for the effects of the other variables on the outcome (Hastie et al. 2003). TreeNet assists with 2D and 3D plots showing the effect of one single predictor and a pair of predictors respectively on the predicted response. In a resource modeling context, these plots might be described as Habitat ore Food Use functions, or when done with the appropriate research design taking resource availability into account, Resource Selection Function (RSF, Manly et al. 2002). Together with its speed and convenient data handling, this makes TreeNet an ideal tool for data exploration and for following up analysis in more detail.

Despite all these advantages, it is surprising to see that in the general wildlife and ecology literature no examples applying TreeNet to develop habitat models can be found up to now (but see Gräber 2006, Hubley et al. in press). Therefore, this thesis shows new methods and approaches for the scientific investigation of traditional problems.

2 Material and Methods

2.1 Habitat model at the point scale

2.1.1 Environmental data at the point scale

For the historical model, the characterization of the landscape structure was done using historical topographical maps that had a scale of 1:100,000. In order to describe the habitat in the years 1981-1992 the freely available Baltic Sea GIS online data was used (see section 2.1.1.2). Furthermore, a digital elevation model (DEM) derived from the GLOBE data set (see section 2.1.1.4) and a digital coastline data set extracted from the freely available online World Vector Shoreline data set (WVS, see section 2.1.1.3) was used in addition to the historical data set, as well as for the prediction made for the period 1981-1992.

2.1.1.1 Historical topographic maps (‘Großblätter’) for East Prussia

The historical topographic maps for East Prussia, so-called ‘Großblätter’, were issued by the former Reichsamt für Landesaufnahme (Office of the Reich for Land Surveying) which was located in Berlin. Today they can be obtained as reprints from the ‘Bundesamt für Kartographie und Geodäsie’ (Federal Office for Cartography and Geodesy). For the range of this survey 15 of these maps were used: Nos. 4, 5, 6, 14, 15, 16a, 16b, 27, 28, 29, 30a, 30b, 43, 44, and 45. However, map No. 42 of the south-west of East Prussia was unfortunately not available so that the districts of Rosenberg and Marienwerder had to be excluded from the investigation. The maps originated for the main part from the year 1939 (except Nos. 6 and 27 from 1941 and No. 43 from 1942). Also, they showed the division of the province of East Prussia in its districts as per 31st December 1937.

2.1.1.1.1 Digitizing and georeferencing of the topographic maps

The available maps were manually scanned as TIF files. In order to be able to open them in ArcMap they were converted to IMAGE files using ArcCatalog. For all GIS analysis ArcGIS 9.1 was applied. The georeferencing was done with the Georeferencing-Tool in ArcMap. For that task the marked longitudes and latitudes were used from the ‘Großblätter’. For every map the four corner points were chosen as the control points with known locations. Since the coordinates on the maps were originally provided in a degree, minute and second format, they were first converted into decimal degrees. The maximum total RMS (Root Mean Square) error reads 0.00303 m on the maps Nos. 5 and 14. The remaining maps showed a lower RMS error and thus a higher accuracy at the georeferencing.

2.1.1.1.2 Digitizing Habitat Features

The boundaries of the individual districts, as marked on the maps, were digitized. To achieve this task, a polygon shapefile was created using ArcCatalog in ArcGIS and named ‘administrative districts’. Every district was digitized as a polygon in ArcMap. The boundaries of the municipalities Elbing-Stadt, Insterburg-Stadt or Tilsit-Stadt were not clearly defined so that these three municipalities were combined with their distinctive rural districts and named in the subsequent steps Elbing, Insterburg or Tilsit-Ragnit, respectively.

In order to characterize the landscape structure of the survey area, the information on land use contained in the maps also had to be digitized. Since the maps were very detailed and a complete manual digitization would have been very labor intense and outside the scope of this project, only the land use classes relevant for White Storks (forest, lake, watercourse and settlement area) were selected to be used in this study. Land use classes like wetland, pasture/meadow or cropland, which are potentially important for the distribution of the White Stork as well, could not be digitized, since they were not clearly defined and distinguishable. For every land use class a new shapefile was opened in ArcCatalog according to the shape of the object to be found: a polygon shapefile for the classes forest and lake, a polyline shapefile for watercourse and a point shapefile for settlement area. In ArcMap the individual objects of the respective land use classes were digitized with the help of the Editor feature. For the layers ‘forest’ and ‘lake’ all forests and lakes marked on the maps showing an expanse of more than one kilometer in any direction (equivalent to one cm on the map) were digitized as individual objects. It was assumed that smaller objects have no relevant influence on the selection of breeding sites for the White Stork because foraging trips can reach a radius up to 5 km from the nest (Johst et al. 2001).

Any watercourse represented on the maps by a double line was digitized as a single line object. So all watercourses represented by a single line are not further considered in this study. This was done because the latter ones could not be distinguished without error from other objects like streets due to the black-and-white design of the maps. In several cases one watercourse feeds into another or rivers flow into lakes. When digitizing manually, inevitably some gaps remain between these lines or lines cross. To obtain a complete network of watercourses the lines were interrupted just before the next line and then lead on to the adjacent line using the Extend-Tool of the Editor feature.

In order to digitize the settlements according to their size the settlements were divided into six different categories. Due to the fact that some ‘Großblätter’ were partly copies of other maps the font sizes representing the names and the population size of the settlements were not identical throughout all the maps. However, the different size categories are well discernable within every single map. Unfortunately, no information could be obtained defining the exact population size of the settlements for the different size categories. The localities were each marked by a center point and then classified according to the individual size category. In some cases, localities consisted of two or more separate clusters of buildings. Where it seemed inappropriate to position the point in the centre of the settlement, since at the average location there were no buildings, each cluster of settlement was marked by a point and classified in the next lower size category. This was the case with some settlements of the size categories 4 and 5.

To minimize edge and fringe effects in the forthcoming modeling component, all objects that fitted one of the four land use classes were digitized up to about four kilometers beyond the boundaries of the survey area.

Finally, all layers were converted into Universal Transverse Mercator (UTM) Projection Zone 34 with the datum World Geodetic System 84 (WGS84). This projection was chosen because the underlying coordinate unit is in ‘meter’ and therefore makes it easier to model and interpret calculated distances. It was used for all other layers and data sets described in the following sections as well.

2.1.1.2 Baltic Sea GIS data from the period of 1981-1992

For the description of the habitat the Baltic Sea GIS data were used. It is available on the internet under http://www.grida.no/baltic/index.htm as an ArcView dataset. The dataset land cover was used, offering a division of the study area into six land use classes: forest, open land, open water, urban land, glacier and unknown land. It was available as a raster dataset with a grid cell size of 1 km * 1 km in the projection Lambert Azimuthal Equal Area. The primary data acquisition period spanned from 1981 to 1992.

2.1.1.3 Digital data for the coastline of the study area

For the coastline layer, the part relevant for the study could be extracted from the global data set of the World Vector Shoreline (WVS), 1990. Although this coastline is mapped at a higher resolution, such data are not available digitally, yet. This data set is made available free of charge online at the U.S. National Geographic Data Center (NGDC) website http://rimmer.ngdc.noaa.gov/coast/. The coastline was extracted for the study area 57° northern latitude, 15° eastern longitude (upper left corner) and 53° northern latitude, 23° eastern longitude (bottom right corner). The data set was downloaded as an ASCII flat file in the format Arc/Info Ungenerate and could be converted into a shapefile applying the script gen2shap1.ave obtained from the website above in ArcView 3.0.

Despite the temporal difference of about fifty years when compared to the year 1939, that coastline was chosen because it represents the best available data for this study as it can be assumed that any changes which may have occurred will have taken place on a scale which will have no influence on White Storks and the creation of the model.

2.1.1.4 Digital Elevation Model (DEM)

The DEM for the study area was taken from the data set of 1999 provided by the Global Land One-Kilometre Base Elevation (GLOBE), downloadable from the internet free of charge under www.ngdc.noaa.gov/mgg/topo/globe.html. The DEM was available as a raster dataset in geographic projection (latitude/longitude) with a horizontal grid spacing of 30 arc seconds. This meant that the size of a grid cell at the equator corresponded to less than 1 km * 1 km and continually declined towards the poles. At 50° northern latitude the extend of a grid cell represented 598 m (from east to west) * 927 m (from north to south) (see http://www.ngdc.noaa.gov/mgg/topo/report/s6/s6A.html). The elevation reading was given in meters. The DEM was extracted for the region 56° northern latitude, 18° eastern longitude (upper left corner) and 52° northern latitude, 24° eastern longitude (bottom right corner).

The difference of time between the year of the construction of the model (1939) and the production of the DEM amounts to a differential period of 60 years. However, it is assumed that the actual contour lines for the model construction remained constant and that the DEM can simply be applied without relevant biases for the historical setting. Major altitudinal changes for the study area and relevant for the White Storks are not known to the author.

2.1.1.5 Preparation of environmental data for modeling in TreeNet

2.1.1.5.1 Data derived from the historical topographic maps and coastline layer

For the complete study area, a raster with a cell size of 100 m * 100 m was created for each of the layers forest, lake, watercourse and coastline using the Distance-Tools 'Straight Line' of the Spatial Analyst in ArcMap. Each raster cell contained a value which showed in meters the distance to the next object of the layer or rather to the coastline. For the layer settlement area a classification according to different size categories was taken. This was done because it was assumed that localities of different size will have a varying impact on the distribution of White Stork nesting sites. According to the size categories different distance raster were created: 'size range 1 to 3', 'size range 1 to 4', 'size range 1 to 5' and 'size range 1 to 6'. For each of the created distance raster the value of the individual cells indicated the proximity to the next settlement of one of the included size categories. So it was possible to test which pool of size ranges showed the greatest effect on the distribution of breeding White Stork.

Further, several density raster with a cell size of 100 m * 100 m were generated for the layer 'settlement area' using the Density-Tool 'Kernel' of the Spatial Analyst with a radius setting of 5,000 m. Combined, and according to different size categories the following rasters were generated: 'size range 1 to 4', 'size range 1 to 5' and 'size range 1 to 6'. Each raster cell contained a value indicating how close the settlements were located to one another.

2.1.1.5.2 Baltic Sea GIS data from the period of 1981-1992

A raster representing a cell size of 1000 m * 1000 m for the land cover classes 'open water' and 'forest' of the grid land cover and derived from the Baltic Sea GIS each were created. For the land cover class 'urban land' from the same data set a point shapefile was created containing a point in the center of every group of adjacent raster cells belonging to this class. With the distance tool 'Straight Line' of the Spatial Analyst a raster for each of the three land use classes mentioned above and with a cell size of 100 m * 100 m depicting the distance to the next object of each individual class was created.

In Table 2.1 the used environmental variables for the model at point scale are shown.

Table 2.1: Environmental variables for the modeling at the point scale

<table>
<tr><th>Environmental variable</th><th>Unit</th><th>Cell size of Raster</th><th>Derivation</th><th>Origin of the data set</th></tr>
<tr><td>Distance to forest</td><td rowspan="3">Meter</td><td rowspan="5">100m * 100m</td><td>Polygon shapefile ‘forest’</td><td rowspan="5">Historical topographic maps, scale 1: 100,000 (edited mostly in 1939)</td></tr>
<tr><td>Distance to lake</td><td>Polygon shapefile ‘lake’</td></tr>
<tr><td>Distance to watercourse</td><td>Polyline shapefile ‘watercourse’</td></tr>
<tr><td>Distance to settlement area, different size ranges</td><td rowspan="2">No unit</td><td rowspan="2">Point shapefile ‘settlement area’</td></tr>
<tr><td>Density of settlement area, different size ranges</td></tr>
<tr><td>Elevation</td><td rowspan="2">Meter</td><td>30 arc seconds</td><td>DEM</td><td>GLOBE dataset (1999)</td></tr>
<tr><td>Distance to coastline</td><td>100m * 100m</td><td>Polyline shapefile ‘coastline’</td><td>WVS data set (1990)</td></tr>
</table>

Table 2.2 gives an overview of the environmental variables used to make a prediction for the period of 1981-1992.

Table 2.2: Environmental variables used for the prediction for the period of 1981-1992

<table>
<tr><th>Environmental variable</th><th>Unit</th><th>Cell size of Raster</th><th>Derivation</th><th>Origin of the data set</th></tr>
<tr><td>Distance to forest</td><td rowspan="5">Meter</td><td rowspan="3">100m * 100m</td><td rowspan="3">Land Cover Grid with a cell size of 1 km</td><td rowspan="3">Baltic Sea GIS (Acquisition period 1981-1992)</td></tr>
<tr><td>Distance to lake</td></tr>
<tr><td>Distance to settlement area</td></tr>
<tr><td>Elevation</td><td>30 arc seconds</td><td>DEM</td><td>GLOBE dataset (1999)</td></tr>
<tr><td>Distance to coastline</td><td>100m * 100m</td><td>Polyline shapefile ‘coastline’</td><td>WVS data set (1990)</td></tr>
</table>

2.1.2 White Stork data at the point scale

Two sets of Stork data were available for the development of a distribution model using the historical environmental data at point scale.

2.1.2.1 Data set 1: Banding locations according to the White Stork banding data set hosted and maintained by the German ornithological station 'Vogelwarte Radolfzell'

From the beginning of the 19th century up to the Second World War a great number of White Storks in East Prussia were 'ringed'/banded by members of the ornithological station 'Vogelwarte Rossitten'. When the ornithological station 'Vogelwarte Rossitten' gave up its work at the Courland Lagoon after the Second World War, the banding lists and a great number of documents on findings were basically lost. From former publications and notes these got partly reconstructed and, in the context of a diploma thesis (Sproll 2000), listed in the database of the presently responsible ornithological station 'Vogelwarte Radolfzell'. The relevant records for the area of survey were kindly made available as an Excel table by the 'Vogelwarte Radolfzell' provided by W. Fiedler. It contained the coordinates of the banding locations in geographic projection (latitude/longitude) as well as the year of the banding event amongst others. The entries used in this study show an accuracy of +/- 1 spatial minute (corresponds to approx. 1.2 km from east to west and 1.8 km from north to south). White Storks are usually banded as juveniles in their nest so that for the study at hand it can be assumed that there were nest sites close to the banding locations (Hornberger 1943, Sproll 2000). Since regular banding took place for many years and several juveniles were marked, many nest sites were repeatedly listed. However, in this study a nest site was used only once, and as a single record, in order to have unique locations preventing an over-presentation of individual nest sites in this study. In this process the entry was chosen that was closest to the year 1939, since the topographical maps exploited for the characterization of land use by the majority originated from that year. Thus, all the data exploited in the survey stem from the period 1908 to 1944.

For use in ArcGIS a shapefile was derived from the Excel table which contained the individual banding locations as points. In the whole area of survey there were 418 nest sites at which White Storks were banded; the survey effort is unequally and opportunistically distributed in the study area and not traceable anymore (Figure 2.1).

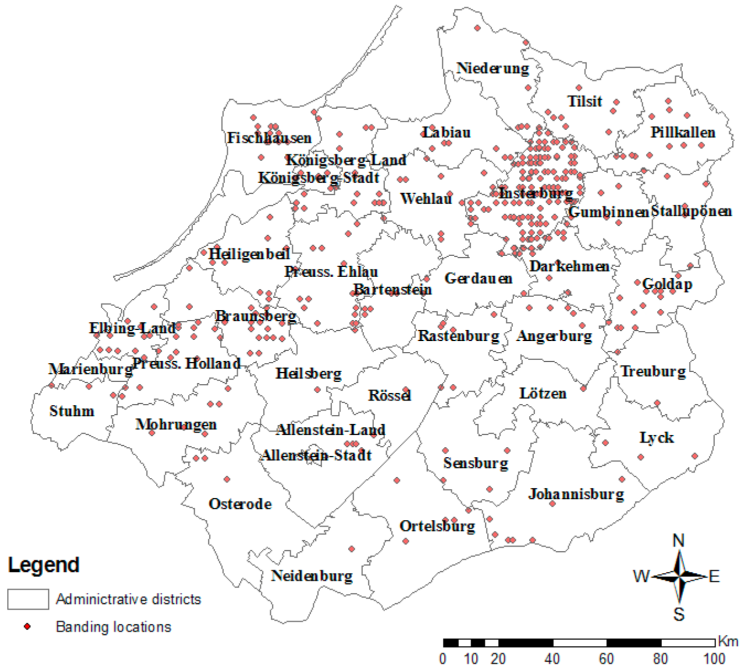

Figure 2.1: Study area showing district boundaries ('Landkreise') and banding locations from the banding data base of the German ornithological station 'Vogelwarte Radolfzell'

More than ¼ of the nest sites (148 of 418) were situated in the district of Insterburg. The reason for this unbalanced effort may be that in the years 1931 to 1942 (and possibly beyond that) intensive banding of White Storks was exercised by F. Hornberger in the district of Insterburg (Hornberger 1943). Thus, the possibility to track a White Stork banded in Insterburg and to list it in the database was somewhat more likely than to track a White Stork from any other district. In general, it can be said that the database does not reflect representatively the spatial distribution of nest sites of White Storks in East Prussia, but instead represents opportunistic banding activity, which is usually closely related to where people and roads are located (commonly referred to as access bias). Consequently, White Storks from regions were tracked more often in which banding was more extensively practiced.

When using all 418 presence points to build a model, the variable 'distance to coastline' emerged to be the only important predictor variable for the distribution of nest sites of White Storks. So in order to prevent an overestimation of habitat features of the district of Insterburg due to human activities, 19 points were chosen at random from the district of Insterburg. This number represented the maximum number of points where banding was exercised in one of the other districts (Königsberg-Land).

2.1.2.1.1 Preparation of Data set 1 for TreeNet

Data set 1 contained 289 presence points. In ArcMap, a new shapefile with 578 points (= double the number than presence points) was created using the tool 'Generate Random Points' with the extension 'Hawth's Tools' (Beyer 2004, available at http://www.spatialecology.com/htools). The points were distributed at random over the entire study area. These points did not represent a real absence; however, instead they represent pseudo-absences and are the points of the habitat available to nesting White Storks. Using pseudo-absences is a common method applied in such studies (see e.g. Manly et al. 2002, Engler et al. 2004).

One column each was added to the attribute table of both point shapefiles (absence and presence points) for the different environmental variables. The distance rasters described in section 2.1.1.5.1 as environmental variables was chosen for the layers 'forest', 'lake', 'watercourse', 'settlement area' and 'coastline'. For the environmental variable 'elevation' the elevation-reading of the DEM raster was used. With the tool 'Intersect Point-Layer' of the extension 'Hawth's Tools' for every point the value of the appropriate raster cell, in which the point was to be found, could be transferred to the attribute table. Both attribute tables were merged into one file in Excel, which was then exported as a TXT file. In order to distinguish between presence and absence points a column was added showing the label '1' for every presence point and the label '0' for all available points (for illustration see Table 2.3).

Table 2.3: Example of an Excel table containing the variable to be modeled (Pres_Avail) and different predictor variables

Pres_Avail	*Distance to watercourse*	*Distance to lake*	*Distance to forest*	*…*	*Distance to coastline*	*Elevation*
1	9848858032	2012461182	3622154053		8818645313	142
1	2236067963	1315294678	2662705322		8956003906	157
…				…		
0	1192057031	5315072754	500		1336450488	43
0	886397168	1976689063	6708203735		4576210156	33
…				…		

2.1.2.2 Data set 2: Map showing the breeding population of White Storks in former East Prussia in the 1931

In 1931 a comprehensive inventory of the White Stork breeding population was compiled by the local police force at the order of the Governor Supreme of the Province of East Prussia. The survey sheets received were evaluated by the ornithological station 'Vogelwarte Rossitten' and the results duly published (Schüz 1933). According to Gaupp (1936) the inventory of 1931 was somewhat faulty, since a large number of breeding sites were apparently overlooked. More exact inventory data was available from the count of 1934. Unfortunately, during the Second World War parts of the records of the ornithological station 'Vogelwarte Rossitten' were lost so that the survey at hand can only revert to the map of the inventory of 1931, published in Schüz (1933, Figure 2.2). The number of White Stork breeding sites, summarized by community, was shown in the community center. So the map does not show the exact location of the individual White Stork breeding ground, meaning no real presence/absence reading, but merely an outline on the level of communities (Schüz 1933). This presents currently the best available map for this subject.

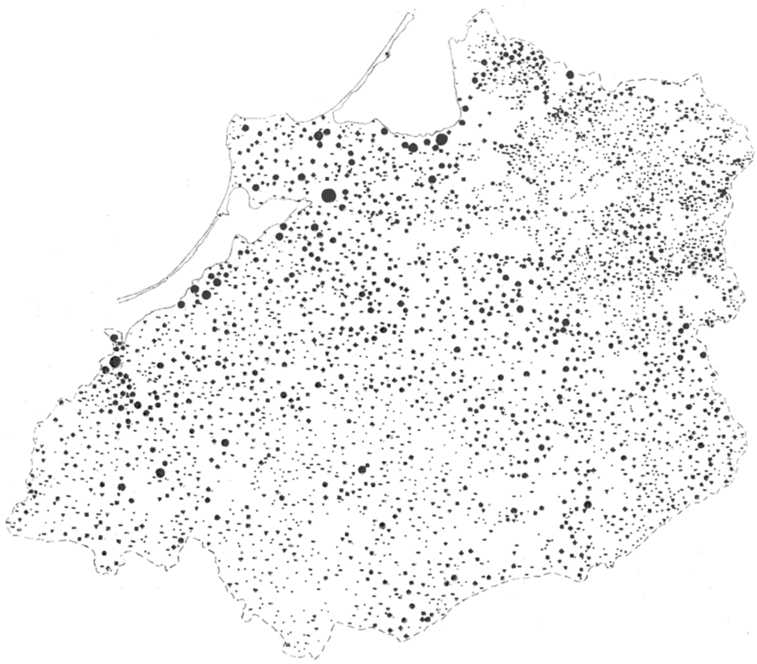

Figure 2.2: Map of White Stork breeding grounds in East Prussia 1931, filled circles: nests occupied, empty circles: nests unoccupied

For its use in ArcMap the map was scanned as a TIF file, converted to an IMAGE file in ArcCatalog and finally georeferenced. The georeferencing was obtained using the layer 'administrative districts' containing the boundaries of each administrative district. This data set was generated by digitalization from the historical maps as described in section 2.1.1.1.2. In the following seven control points along the border of the province of East Prussia were chosen which could be clearly identified in the map by Schüz (1933) as well as in the layer 'administrative districts'. The total RMS error was 0.02454.

2.1.2.2.1 Preparation of Data set 2 for TreeNet

Since the map was available only as a hardcopy in restricted quality the georeferenced scan showed some 'noise' in the form of small black pixels scattered throughout the entire map. Therefore a filter had to be used for the correction, if to be used as a raster. Using the tool 'Majority Filter' from Spatial Analyst in ArcMap every cell was allotted a new value of '0' or '1' according to the value reading of the four cells directly adjacent. Next, the raster was overlaid with a point shapefile in which points were regularly arranged at a distance of 500 m for the entire study area. All points on a raster cell valued '1' were coded as presence points (value '1'), all other points as absence points (value '0'). Since the representation of the White Stork count does not correspond with the exact and actual locations of the nests, but rather places them in the centers of the communities, only points at a distance of 2 km from the next presence point were used as absence points in the modeling. A selection of 5,000 locations from all the presence as well as absence points (a total of 10,000) was randomized and saved in two distinct data sets. As described in section 2.1.2.1.1, a column was added to the attribute tables of both data sets with the respective values applying the tool 'Intersect Point-Layer' for every environmental variable. Using Excel, a TXT file was created showing both attribute tables listed, as well as a column stating whether the reading showed a presence (= 1) or an absence (= 0) point.

2.1.3 Modeling in TreeNet (point scale)

Since the variable to be modeled was binomial (value '1' for nest locations and value '0' for available and absence points respectively), the algorithm for 'Binary Logistic Models' was selected in TreeNet. The output shows an index of relative importance between 0 and 1 as response. The settings shown in Table 2.4 were fixed for the modeling process in TreeNet.

Table 2.4: Fixed settings in TreeNet for the models at the point scale

Setting	*Reading*
Number of trees to use	10,000
Maximum number of trees	10,000
Minimum number of training observations in terminal nodes	10
Sub-sample fraction	0.5
Influence trimming factor	0.1
M-regression breakdown parameter	0.9
Regression loss criterion	M-Huber loss
Optimal logistic selection criterion	Cross Entropy (Likelihood)

For the different models only the parameters learn rate, maximum number of nodes per tree and the testing mode to validate the model during the modeling process of TreeNet were changed in order to optimize the result.

2.1.3.1 Model 1: Using presence/pseudo-absence data set derived from the Banding Database of the Ornithological Station Radolfzell, Germany (Data set 1)

For each model the variables ‘distance to forest’, ‘distance to lakes’, ‘distance to watercourse’, ‘distance to coastline’ and ‘elevation’ were used as predictor variables. Additionally, one of the variables derived from the layer ‘settlement area’ was used. These were ‘distance to village, size range 1 to 3’, ‘distance to village, size range 1 to 4’, ‘distance to village, size range 1 to 5’, ‘distance to village, size range 1 to 6’, ‘density of villages, size range 1 to 4’, ‘density of villages, size range 1 to 5’ and ‘density of villages, size range 1 to 6’. In order to find out which of the variables derived from the layer ‘settlement area’ showed the highest predictability for the distribution of White Stork breeding grounds, a model was created first of all using one of the different variables of the layer ‘settlement area’ (together with all other variables). For that the following settings in TreeNet were employed:

Learn rate:	0.001
Maximum nodes per tree:	6
Testing:	20 % of the cases are selected at random for testing.

Because of the unbalanced design of the two target classes (289 presence and 578 available locations) in TreeNet the class weights were set to 'balanced' for all models build in this step. Thus small classes were upweighted to the equal size of the largest target class by adjusting weights (Salford Systems Ltd. 2003).

After selecting a variable for the layer 'settlement area' to be used for the modeling individual settings in TreeNet were altered: The learn rate was set at 0.01 or 0.001. Maximum nodes per tree read at 2 (unable to detect interactions between predictor variables – 'main effects additive models') or 6 (well able to detect interactions). As a testing data set 20 % or 40 % of the data set were applied. The best model was established by the evaluation of an independent data set (see following section).

2.1.3.1.1 Evaluation of Model 1

For the evaluation the data set of the White Stork count from 1931 was employed, as mentioned in section 2.1.2.2. Using the tool 'Random Selection' of the extension 'Hawth's Tools' 1,000 presence and absence points each were randomized. On the base of these 2,000 points, a prediction was made using the program TreeNet. For every single model sensitivity, specificity and the area under the ROC-Curve (Area under the curve - AUC) were conveniently calculated for the test data set. For that a Delphi program written by B. Schröder was used (version January 2004), downloadable from the internet for free under http://brandenburg.geoecology.uni-potsdam.de/users/schroeder/download.html. With that program it was also possible to calculate bootstrapped confidence intervals for the AUC values with the percentile method referring to Buckland et al. (1997). The AUC values served for selecting the best model.

2.1.3.2 Model 2: Using a presence/absence data set derived from the White Stork count of 1931 (Data set 2)

As predictor variables 'distance to forest', 'distance to lakes', 'distance to watercourse', 'distance to coastline' and 'elevation' as well as one of the variables derived from the layer 'settlement area' 'distance to village, size range 1 to 3', 'distance to village, size range 1 to 4', 'distance to village, size range 1 to 5', 'distance to village, size range 1 to 6', 'density of villages, size range 1 to 4', 'density of villages, size range 1 to 5' or 'density of villages, size range 1 to 6' were used.

In order to select the variable derived from the layer 'settlement area', which showed the maximum predictability, a model was created using one of the these variables as well as the variables derived from the other layers. Therefore, in TreeNet a learning rate of 0.1, maximum 6 nodes per tree and 20 % of the data set applied for testing were used.

In order to optimize the model, the learn rate was changed from 0.1 and 0.05. Furthermore tests were run using maximum 2 or 6 nodes per tree. As a testing sub-sample 20 % or rather 40 % of the data set were used. The best model was established by the evaluation of an independent data set (see following section).

2.1.3.2.1 Evaluation of Model 2

Spearman-Rank Correlation

In order to evaluate the models created using TreeNet all 418 nest locations from the data set of 'Vogelwarte Radolfzell' (Data set 1) were used. Since this data set only dealt with presence points, no values for AUC, sensitivity or specificity, could be established as for the evaluation with presence-absence data. That is why a Spearman-rank Correlation was used as described in Boyce et al. (2002). As shown in section 2.1.4 a prediction was created on a regular point grid which covered the complete study area. The predicted values were allotted to ten bins of identical size as described in Boyce et al. (2002). In ArcMap the value of the original values of every point was converted to the number of the appropriate bins (1 to 10). Then, the area could be calculated that the individual bins covered (number of points per bin = area in square kilometer per bin). In the following step the number of the 418 nest locations per bin was calculated and the area-adjusted frequency established (number of points of the testing data sets per bin divided by the area of the respective bin). A Spearman-rank Correlation was calculated with the bin number as the categorical variable and the area-adjusted frequency as the continuous variable. The calculation was carried out with the program R (version 2.3.1) freely available in the Internet at http://cran.r-project.org/. To establish the deviation within the model the data set was divided randomly into five even-sized subsets for which the area-adjusted frequency each was calculated (Boyce et al. 2002).

Minimal predicted area (MPA)

Another measure to evaluate the performance of a model when only opportunistic data (presence-only) is available is the calculation of the minimal predicted area (MPA). The value above which 90 % of the locations of the test data set were observed is used as a threshold to transform the predicted values for the whole study area into occurrence and non-occurrence (as in Engler et al. 2004). The MPA is the area obtained by considering all raster cells of the study area showing an occurrence. Based on the rule of parsimony the smaller the minimal predicted area the better is the performance of the model (see also Guisan et al. 2006b).

2.1.4 Index of relative importance for the entire study area

To make predictions for the complete study area for both points in time, a point shapefile was established in ArcMap using the tool 'Generate Regular Points' of the extension 'Hawth's Tools' in which points were regularly placed throughout the study area at a distance of 1 km. Applying the tool 'Intersect Point Layer' of the extension 'Hawth's Tools' the value of the appropriate set of variables was allotted to each of these points. Applying a previously established model in TreeNet a prediction of an index of relative importance could be made stating whether a stork breeding ground could be found at that location. As output, TreeNet provided a TXT file in which a value between 0 and 1 was listed for every point. From this file a shapefile was generated for further use in ArcMap.

2.2 Habitat model at the administrative district scale (polygon)

2.2.1 Environmental data at the administrative district scale

2.2.1.1 Historical data and their preparation in ArcMap

Information on the character of the habitat was taken from the 'Statistical Handbook for the Province of East Prussia 1938' ('Statistisches Handbuch für die Provinz Ostpreußen 1938'). Table 69 contained the 'main types of land use according to the administrative districts of 1936'. From this table the columns 'arable land', 'pasture land' and 'forest' were taken showing in hectares the respective type of land use for the individual administrative districts. An additional column showed the total area in hectares for every administrative unit. It then was employed for the calculation of the area of the three land use classes (given in percent).

The digitized layers ‘settlement area’ and ‘lake’ as well as the information on the coastline and the DEM (see section 2.1.1) were also used. The lake layer was converted to a raster covering the whole study area with a cell size of 100 m * 100 m using the tool ‘Convert Feature to Raster’ of the Spatial Analyst in ArcMap. Subsequently, the percentage of area which was lake (‘water’) could be calculated for every administrative district applying the tool ‘Zonal Statistics’ of the Spatial Analyst. For the layer ‘settlement area’, the number of settlements in the districts were counted using the tool ‘Count Points in Polygon’ of the extension ‘Hawth’s Tools’. Here, a sub-division was taken, creating three different variables: ‘number of settlement areas, size range 1 to 4’, ‘number of settlement areas, size range 1 to 5’, and ‘number of settlement areas, size range 1 to 6’ (compare section 2.1.1.5.1). The average reading of all the raster cells of the DEM in the administrative district was calculated as the variable ‘elevation’ using the tool ‘Zonal Statistics’ of the Spatial Analyst. For defining the variable ‘distance to coastline’ the average of all raster cells in the created raster showing the distance to coastline (see section 2.1.1.5.1) was calculated.

2.2.1.2 Data for the period of 1981-1993 and their preparation in ArcMap

From the Baltic Sea GIS the Land Cover Data set was applied. For the generated raster data sets for the land use classes ‘open water’ and ‘forest’ (see section 2.1.1.5.2) the number of points per district were determined using the tool ‘Zonal Statistics’ of the Spatial Analyst. The percentage of the total area the land use classes had in the appropriate districts could be taken from these readings.

The Baltic Sea GIS furthermore contained a data set about the arable land and the pasture land (acquisition period 1987 to 1993). It had a grid spacing of 10 km * 10 km. Every raster cell contained a value about the percentage of arable land and pasture/meadow respectively. Using the above-mentioned tool ‘Zonal Statistics’ for every administrative district the mean percentage of the two variables could be determined. In the same way a mean value for the variables ‘elevation’ and ‘distance to coastline’ was calculated.

Table 2.5 and Table 2.6 give an overview of the environmental variables used to create Model 3 and to estimate the densities of breeding White Storks for the period of 1981-1993.

Table 2.5: Environmental variables for the modeling at the administrative district scale

Environmental variable	Units per district	Derivation	Origin of the data set
Pasture	Percentage	Table 69	‚Statistisches Handbuch für die Provinz Ostpreußen 1938', acquired in 1936
Arable land			
Forest			
Lake		Polygon shapefile 'lake'	Historical topographic maps, Scale 1: 100,000, edited mostly in 1939
Number of settlement areas, different size ranges	No unit	Point shapefile 'settlement area'	
Elevation	Meter	DEM (cell size of 30 arc seconds)	GLOBE dataset (1999)
Distance to coastline		Polyline shapefile 'coastline'	WVS data set (1990)

Table 2.6: Environmental variables used for the prediction for the period of 1981-1993

Environmental variable	Units per district	Derivation	Origin of the data set
Pasture	Percentage	Pasture Lands Grid (cell size of 10 km)	Baltic Sea GIS (Acquisition period 1981-1992)
Arable land		Arable Lands Grid (cell size of 10 km)	
Forest		Land Cover Grid (cell size of 1 km)	Baltic Sea GIS (Acquisition period 1987-1993)
Lake			
Number of Settlement areas per km^2, size range 1 to 4	No unit		
Elevation	Meter	DEM (cell size of 30 arc seconds)	GLOBE dataset (1999)
Distance to coastline		Polyline shapefile 'coastline'	WVS data set (1990)

2.2.2 White Stork data at the administrative district scale

In the year 1934 an international population assessment was carried out. Figures of the census for the individual administrative districts of the province of East Prussia were published in Tischler (1941). The counted number of White Stork breeding pairs per 100 km^2 was quoted for every administrative district.

2.2.3 Modeling in TreeNet (Model 3)

The variable 'number of storks per km^2' served as target variable. Since it concerns a continuous variable, in TreeNet the algorithm for Logistic Regression Models was chosen. The variables 'arable land', 'pasture land', 'forest', 'water', 'distance to coastline' and 'elevation' served as predictor variables as well as one of the three variables 'number of settlement area, size range 1 to 4', 'number of settlement area, size range 1 to 5' and 'number of settlement area, size range 1 to 6'.

For Model 3, the following settings were chosen in TreeNet: The number of terminal nodes was set at 1 due to the small number of samples (37). The maximum number of nodes per tree read 2 or 6, the learn rate 0.01 or 0.005. Because no independent data set was available for evaluating the models, the testing was done using a 10-fold cross-validation. This is recommended for small data sets when one cannot afford to reserve some data for testing (Fielding & Bell 1997). The data set is partitioned into 10 bins. Then a model is calculated for nine bins while the 10th bin serves as test data set. This is repeated 10 times until every bin was once used as test data. After all 10 folds are completed, the results from each fold are averaged to get a fair test estimate of the all-data model performance (Salford Systems Ltd. 2003).

2.2.3.1 Evaluation of Model 3

For every model the mean absolute error (MAE) between observed and predicted values was calculated. The MAE is often used as a similar measure than to determine the goodness-of-fit of models (Legates & McCabe 1999, Hall 2001):

$$MAE = N^{-1} \sum_{i=1}^{N} | O_i - P_i | \qquad \text{(Equation 2.1)}$$

The model with the lowest MAE was selected as best model. In addition the Coefficient of Efficiency (E) was calculated for better interpreting the results:

$$E = 1 - \frac{\sum_{i=1}^{N} (O_i - P_i)^2}{\sum_{i=1}^{N} (O_i - \overline{O})^2} \qquad \text{(Equation 2.2)}$$

The Coefficient of Efficiency can have values between minus infinite and 1 (perfect model).

2.2.4 Estimation of White Stork densities for the period of 1981-1993

In order to make a prediction for the period 1981-1993 the best created model using the variables ‘arable land’, ‘pasture land’, ‘forest’, ‘water’, ‘distance to coastline’, ‘elevation’ and ‘number of settlement area, size range 1 to 4’ was selected. Based on this model and the environmental variables for the period 1981-1993 as described in section 2.2.1.2, a predicted response value for every administrative district was computed.

3 Results

3.1 Model 1: Presence-available model based on banding locations (point scale)

Model 1 was generated applying banding locations taken from the Banding Database of the Ornithological Station Radolfzell (Data set 1) as the target variable to be modeled. As predictor variables it contained 'distance to lake', 'distance to forest', 'distance to watercourse', 'distance to coastline', 'elevation' and 'density of settlement areas, size range 1 to 5'. It was achieved using the following settings in TreeNet:

learning rate:	0.01
maximum nodes per tree:	3
percentage of the dataset applied for testing:	40 %.

For each predictor variable TreeNet offers a relative importance score (Table 3.1). *„The relative importance score provides a relative measure of each variable's contribution to the model's predictive power. The raw importance scores are rescaled so that the most important variable always gets a score of 100. The raw variable importance score is computed as the cumulative sum of improvements of all splits associated with the given variable across all trees up to a specific model size."* (Salford Systems Ltd. 2003). So TreeNet allows a ranking of the used predictor variables according to their importance in the model. However, one should keep in mind that such a ranking is different from using p-values or AICs, because it is derived from a different method than log-likelihood.

Table 3.1: Importance of the predictor variables of Model 1

Predictor variable	*Relative Importance Score*
Distance to coastline	100.00
Density of villages, size range 1 to 5	92.26
Distance to forest	76.82
Elevation	57.65
Distance to lake	47.64
Distance to watercourse	44.48

One option for the interpretation of the model is offered by the partial dependence plots, provided by TreeNet (Figure 3.1 a-f). They show the effect of the respective predictor on the response including the interdependency with the other predictors applied.

The partial dependence plot for the most important predictor ‘distance to coastline’ (Figure 3.1 a), shows a positive influence of the predictor on the occurrence of nest locations up to a distance of about 40 km from the coastline with a reading of 0.16. With increasing distance to the coastline the graph drops continually and the partial dependence reads negative at a distance of about 60 km from the coastline. From about 80 km the reading remains negative at a constant value of about -0.16.

Regarding the influence of the predictor ’density of settlement areas, size range 1 to 5’ on the response (Figure 3.1 b) it can be established that with a low density of settlement (up to 0.1) a negative partial dependence of -0.17 is on hand. Up to a density of 0.25 the partial dependence increases up to a reading of about 0.15. A further increase of density, however, does not increase the positive result any further.

The correlation of the predictor ‘distance to forest’ and the predicted response (Figure 3.1 c) result in marginally negative readings of -0.1 at a short distance of about 2 km. The highest readings for the partial dependence of about 0.07 appear at a distance of approx. 4 km from the nest locations to forest. Thereafter, the partial dependence shows a slight decline but remains somewhat positive.

The partial dependence plots of the following last three predictors which had lower relative importance scores (between 44.5 and 57.7) generally show a very slight influence on the variable. The predictor ’elevation’ (Figure 3.1 d) shows positive values around 0.05 for the region of 50 to 100 m above mean sea level. There is a sharp decline in the partial dependence between 100 and 120 m above mean sea level. In the following it remains constant at a low negative reading of about -0.04.

The partial dependence plot for the predictor ‘distance to lake’ (Figure 3.1 e) has a maximum reading of 0.04 at a distance of 5 km to the nearest lake. If the distance is shorter the influence becomes negative (-0.03 to -0.04). At a distance of more than 5 km to the nearest lake the partial dependence declines and from a distance of 15 km on remains stable at a reading of about 0.01.

Regarding the distance of locations to the nearest watercourse (Figure 3.1 f), the influence on shorter distances with a reading of 0.05 is slightly positive, declines constantly, however, up to a negative reading of -0.05 for the partial dependence.

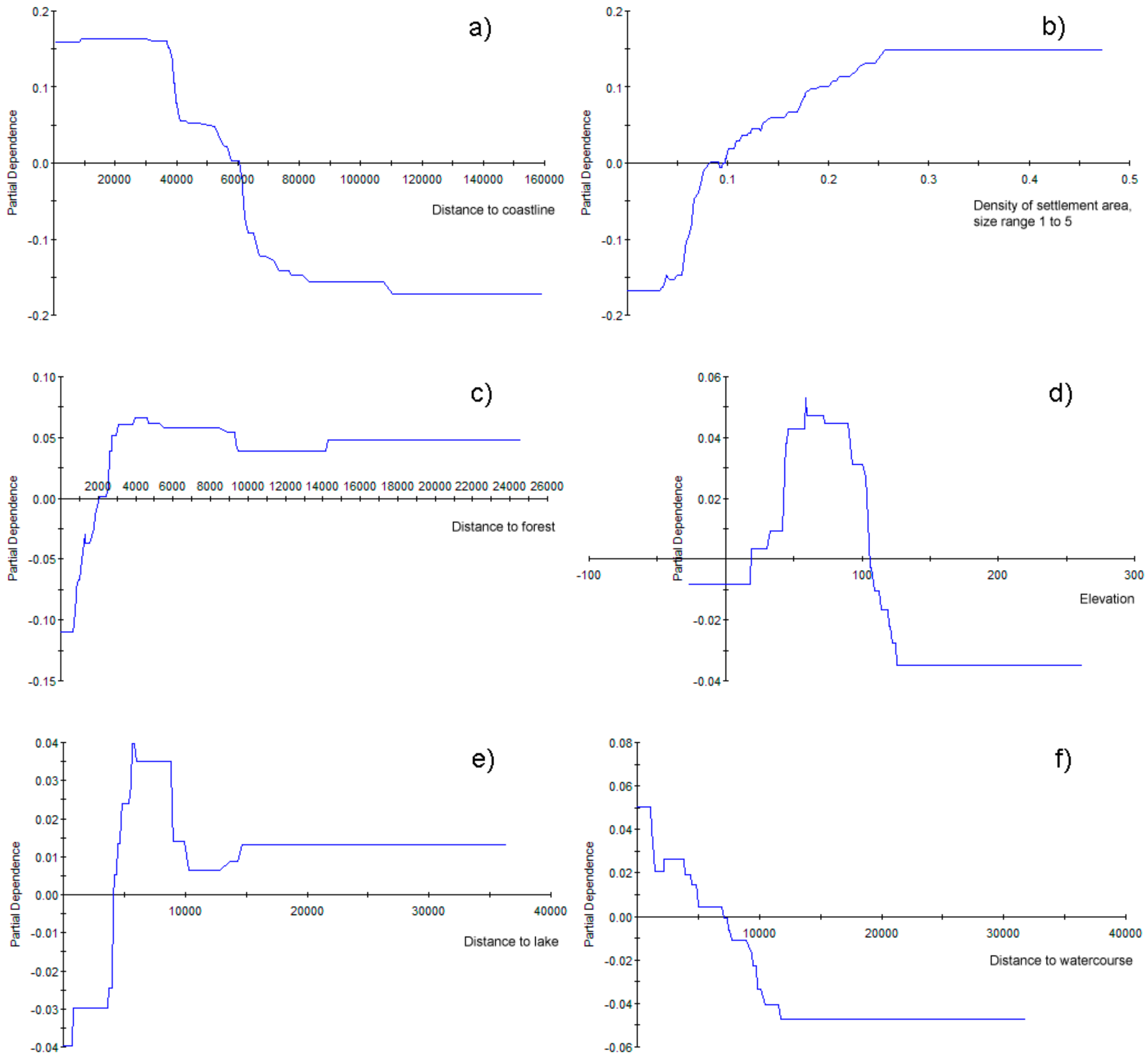

Figure 3.1: Partial dependence plots for predictor variables employed in Model 1; a) 'distance to coastline', b) 'density of settlement areas, size range 1 to 5', c) 'distance to forest', d) 'elevation', e) 'distance to lake' and f) 'distance to watercourse'

3.1.1 Evaluation of Model 1

The calculated AUC value for the test data set (per 1,000 randomly chosen presence and absence locations of data set 2) represent 0.790 with a confidence interval of 0.771 - 0.809.

An AUC of 0.790 means that when arbitrarily selecting a breeding pair from a presence and an absence location the predicted reading for the presence location is higher than the predicted value for the absence location by 79%. According to Hosmer & Lemeshow (2000) a model with an AUC of between 0.7 and 0.8 is to be considered as 'acceptable'.

In Figure 3.2 the Receiver Operating Characteristic (ROC) – Curve for Model 1 is shown.

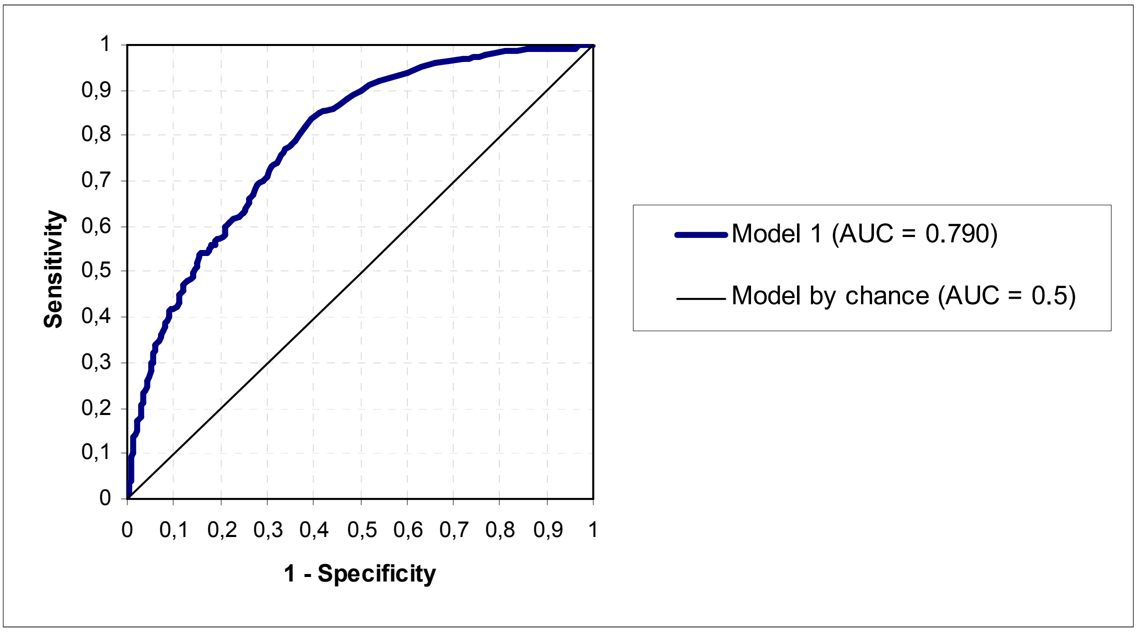

Figure 3.2: Receiver Operating Characteristic (ROC) – Curve for Model 1

In Table 3.2 correct classification rate (CCR), sensitivity and specificity are calculated for Model 1 with a threshold of p= 0.5 as well as with the optimal threshold (p_opt = 0.36). P_opt maximised the correct classification rate as described in Zweig & Campell (1993).

Table 3.2: Confusion Matrix for a threshold of a) p= 0.5 and b) p_opt = 0.36, calculated applying Schröder's program for ROC_AUC (version 2004)

a)

P = 0.5	**Prediction**	
Observed data	**0**	**1**
0	870	130
1	518	482
CCR	0.676	
Sensitivity	0.482	
Specificity	0.870	

b)

P_opt = 0.36	**Prediction**	
Observed data	**0**	**1**
0	592	408
1	153	847
CCR	0.720	
Sensitivity	0.847	
Specificity	0.592	

The predicted values of the test data set were divided into classes with an interval of 0.05 separated into presence and absence locations. The two histograms in Figure 3.3 clarify the distinction between the predicted readings of the absence and presence locations. For more than half of the absence locations a value of less than 0.4 was predicted. In the categories with readings of more than 0.4 the number of absence locations shows a continued decline.

The maximum of the predicted values for the presence locations read at 0.4 and 0.45 for about 180 of 1000 locations (=18 %). With probabilities between 9 % and 14 % of presence locations, readings up to the category of 0.7 to 0.75 in the higher ranking bins were predicted.

However, there is an overlap of the predicted values for presence and absence locations between the readings 0.35 and 0.45. In this region there are about 300 presence and 520 absence locations.

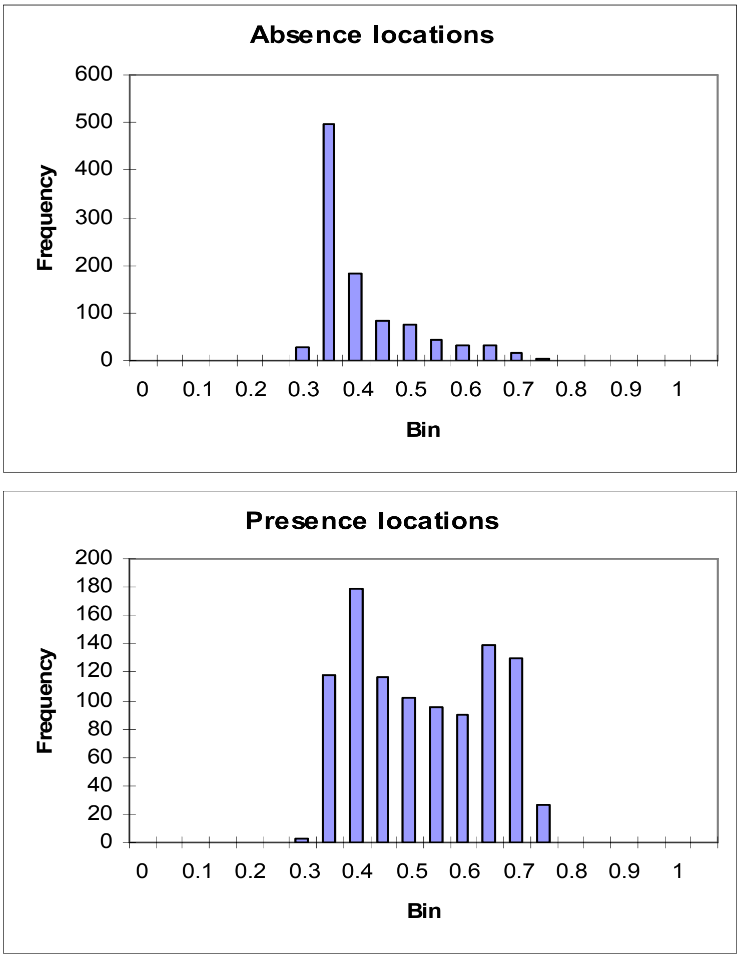

Figure 3.3: Histogram for predicted values of the test data set divided by absence (above) and presence locations (below)

3.1.2 Index of relative importance for the year 1939 applying Model 1

Applying the regular point grid generated in section 2.1.4 (distance between points = 1,000 m) a map showing an index of relative importance for the complete range of the survey was created, based on Model 1 (Figure 3.4). Because the environmental data used to characterize the locations originated mainly from the year 1939 (see Table 2.1) the prediction is related to this year. The predicted relative occurrences are located between 0.29 and 0.74.

Figure 3.4 explains once again the evidence of the partial dependence plots and the relative importance score of the predictor variables: the variable 'distance to coastline' is the most important predictor and is most influential in the distribution of nest locations. Thus, in a strip of about 60 km along the coastline high occurrence indices of nest locations of the White Stork are predicted. In contrast for the southern section of the survey area, situated at a greater distance from the coastline, low occurrence indices were predicted. Throughout the complete area that received surveys locations on lakes and in forests or rather in their close proximity, showed as being virtually unsuitable for breeding. In these regions the concentration of the settlement area is minimal, too. The environmental variable 'elevation' also shows a differentiated gradient in the survey area from north to south since the height above mean sea level along the coastline (in the north) offers lower readings, increases, however, towards the heartland (towards the south east). Higher occurrence indices were predicted for the lower lying areas of the north.

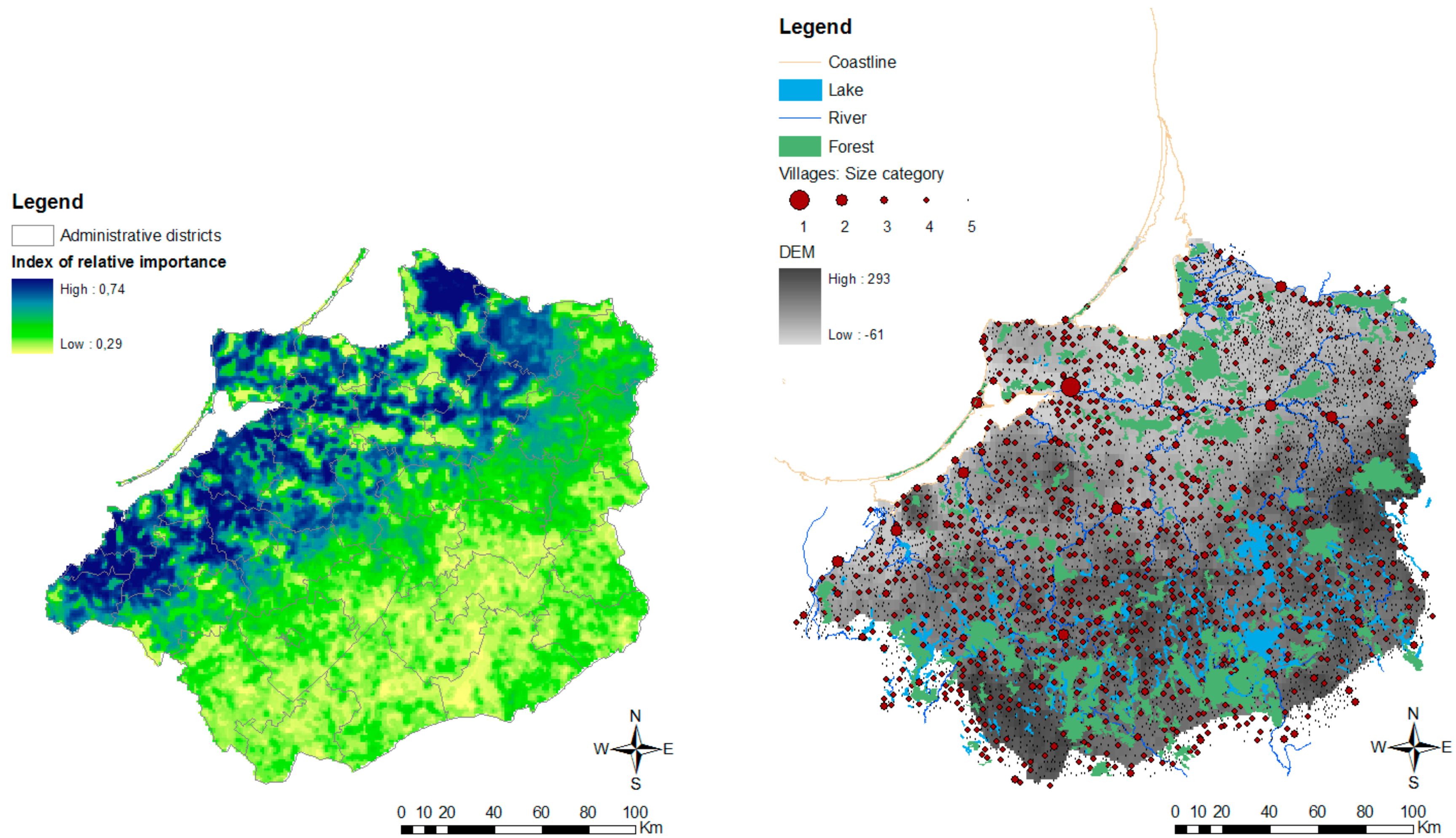

Figure 3.4: Predicted index of relative importance for 1939 applying Model 1 (left) and the study area with the applied GIS layers for comparison (right)

In order to relate the predicted relative occurrences to one of the two classes absence or presence, a convenient threshold must be selected. Using the program ROC_AUC the threshold could be established with which the correct classification rate was maximized; it has a value of 0.36. In Figure 3.5 the prediction for the complete survey area is depicted applying this threshold. According to this scheme, potentially in the north of East Prussia, the White Stork appears all over the country, apart from woodland areas. In the south, rich in lakes and forests, there is a larger area in which an absence of breeding sites of the White Stork was predicted.

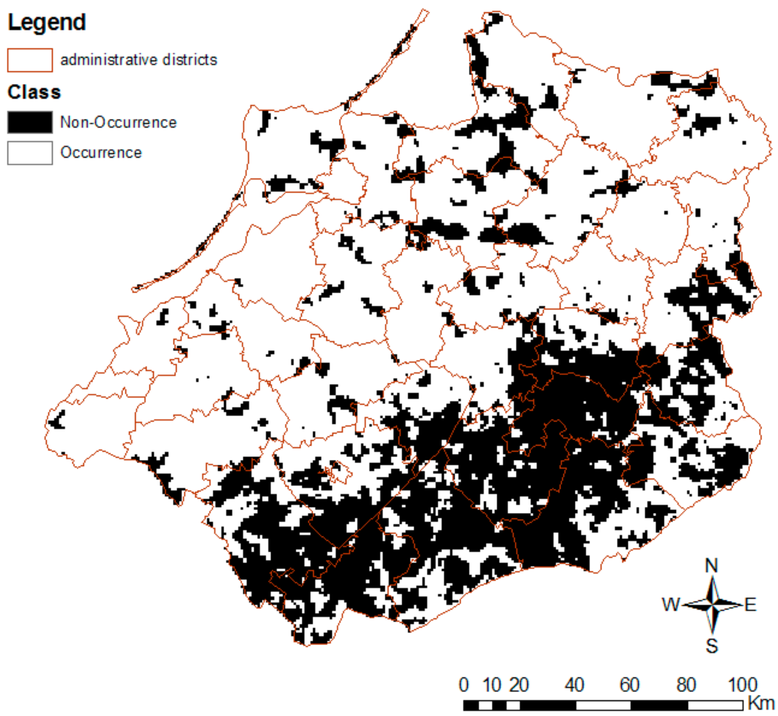

Figure 3.5: Classification of the predicted values for the year 1939 for occurrence and non-occurrence applying a threshold of 0.36

3.2 Model 2: Presence-absence model based on a White Stork count (point scale)

For the construction of the model the data set derived from the map for the White Stork census of 1931 (Data set 2) was chosen as the target variable. Model 2 contained the predictor variables ’distance to forest’, ’distance to settlement area, size range 1 to 5’, ‘distance to coastline’, ‘distance to lake’, ‘distance to watercourse’ and ‘elevation’ and was constructed in TreeNet with the following settings:

learning rate:	0.1
maximum nodes per tree:	2
percentage of the dataset applied for testing:	40 %.

Table 3.3 first shows the variable importance of the applied predictors, calculated by TreeNet The predictor variables ‘distance to forest’ and ‘distance to settlement areas, size range 1 to5’ show the greatest influence on the predicted response.

Table 3.3: Importance of the predictor variables of the Model 2

Predictor variable	*Relative Importance Score*
Distance to forest	100.00
Distance to settlement area, size range 1 to 5	97.44
Distance to coastline	78.69
Distance to lake	58.25
Distance to watercourse	56.65
Elevation	50.19

For an easier interpretation of the model a partial dependence plot was created for each predictor variable in TreeNet (Figure 3.6 a-f).

The predictor ’distance to forest’ is the most important one (Figure 3.6 a). Locations in close proximity to or rather in woodland are considered somewhat unsuitable as breeding sites. They show a partial dependence with readings of less than -0.5. As from a distance of about 1 km to the nearest forest the partial dependence shows positive readings of more than 0.3. However, up to10 km it shows a slow decline and reaches nearly 0. High positive values of 0.6 are reached once again at a distance of 13 km from the next forest.

For the predictor ’distance to settlement area, size range 1 to 5’ (Figure 3.6 b) a positive effect can be seen for locations at a distance of up to about 1.5 km from the next settlement of the size range 1 to 5. There is a maximum for the reading of 0.5 at a distance of 1 km to the nearest settlement area. If the distance increases, the partial dependence becomes negative and reaches a high negative effect with a partial dependence of -1 at distances of more than 3 km to the nearest settlement.

The predictor ‘distance to coastline’ (Figure 3.6 c) shows a high negative effect with readings under -0.7 in close proximity to the coast, it then rises considerably, however, and becomes positive. At a distance of about 2 km from the coast the partial dependence reaches a high positive reading of about 0.6. Up to a distance of about 20 km to the coastline the partial dependence is positive, however, it then shows a steady decline. With growing distance to the coastline the influence of the predictor is almost 0 to slightly negative. Only as from a distance of about 150 km does the partial dependence reach positive readings again with a second maximum of about 0.6.

The predictor ‘distance to lake’ (Figure 3.6 d) with readings of up to 0.4 has a negative effect on the predicted response of locations in close proximity of lakes. As from a distance of about 7.5 km to the next lake the partial dependence becomes slightly positive with readings between 0.1 and 0.2. It drops towards zero at a distance of 30 km, but then shows high positive readings of more than 0.4 at a distance of more than 35 km to the nearest lake.

The graph of partial dependence for the predictor ‘distance to watercourse’ (Figure 3.6 e) shows slightly positive readings around 0.1 for distances of up to 3 km to the next watercourse. At a distance of 3 km to 20 km the readings vary around the zero point. At distances over 20 km the partial dependence achieves high negative values of -0.6, dropping to -0.7 across a distance of 30 km.

The partial dependence of the predictor ‘elevation’ (Figure 3.6 f) with a reading of -0.2 is negative at first. At altitudes of 10m and 80m above mean sea level the readings vary between 0.03 and 0.12. Between 80m and 120m the partial dependence then shows slightly negative readings of up to -0.05 rising again up to nearly zero. As from a height of about 170 m the partial dependence shows a sharp decline and achieves a negative minimum with a reading of about -0.3.

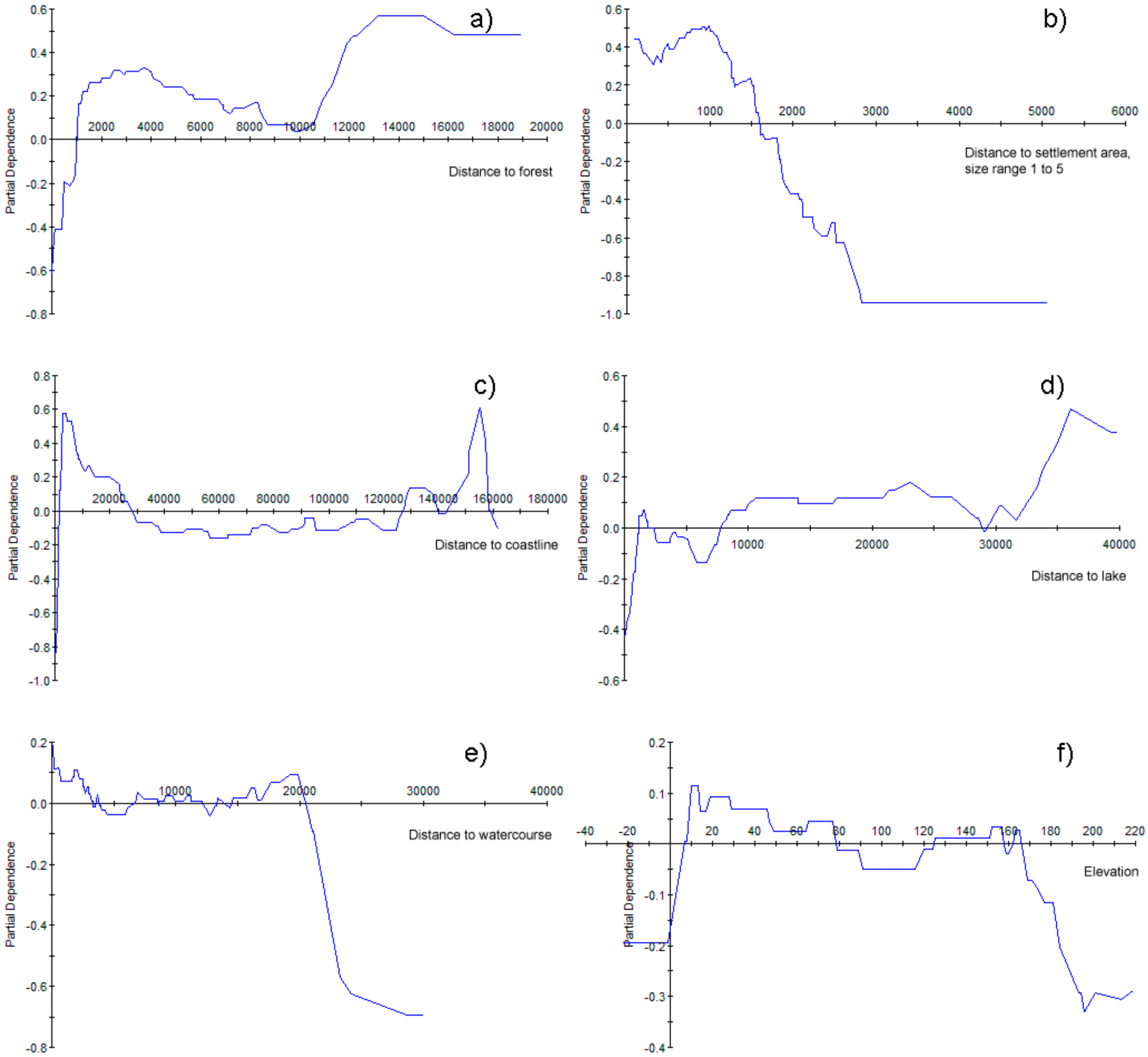

Figure 3.6: Partial dependence plots for predictor variables employed in Model 2; a) 'distance to forest', b) 'distance to settlement area, size range 1 to 5', c) 'distance to coastline', d) 'distance to lake', e) 'distance to watercourse' and f) 'elevation'

3.2.1 Evaluation of Model 2

Spearmans-Rank Correlation

After the classification of the predicted values of the breeding occurrence in 10 equal-interval bins the number of points of the test data set per bin (418 points from data set 1) were established and these adjusted to the area of the total survey covered by this bin (area adjusted frequency, as in Boyce et al. 2002). For Model 2 a value of 0.976 was calculated for the Spearman-Rank Correlation. For the determination of the variance within the test data set it was divided into 5 subsets and the medium area adjusted frequency as well as the standard deviation were calculated (Figure 3.7). As the high reading for the correlation states, the area adjusted frequency increases with the number of bins. 71% of the locations are to be found in bins 6 to 10. Most of the presence locations per area fell in bins 8 to 10.

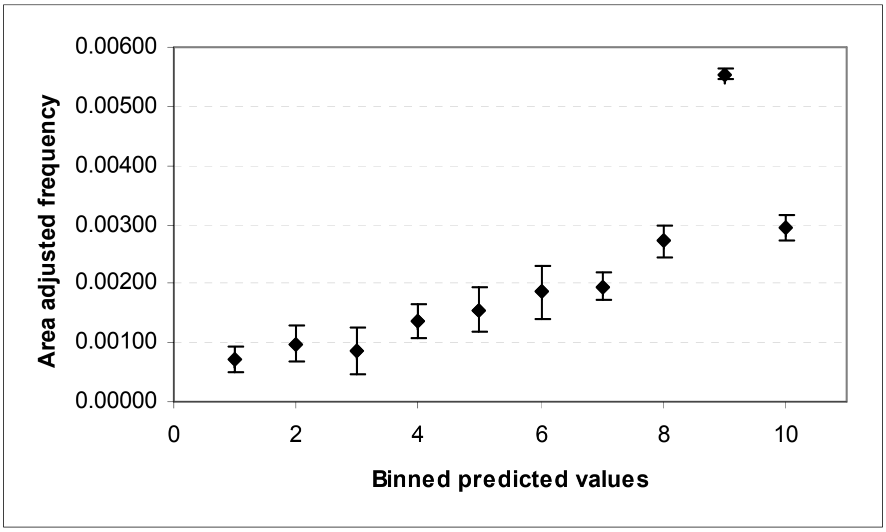

Figure 3.7: Average area adjusted frequency with standard deviation per bin for the predicted values of the test data set (418 presence locations in an area of 35,628 km^2, divided into five sub-samples)

Minimal predicted area (MPA)

Another measure to evaluate the performance of a model constructed with a presence-only data set is the calculation of the MPA (see section 2.1.3.2.1). For this one the reading of the appropriate grid cell of the relative occurrence index was established for the 418 nest locations of the test data set. 90 % of the nest locations showed readings larger than 0.34. If this reading is applied as the threshold for the transformation of the predicted values of the map generated applying Model 2 in presence/absence (as in Engler et al. 2004) 75 % of the total survey area lay within the occurrence area (minimal predicted area). This shows that the White Stork is predicted to be a widespread species within the study area (Figure 3.8).

In Table 3.4 the three models with the highest MPA are specified for comparison. They all contain the same predictor variables as Model 2 and only differ in their settings used for creating the model in TreeNet. The MPA of Model 2 was 2.6 % higher than the lowest obtained MPA.

Table 3.4: MPA and Spearman Rank Correlation for the three models with the highest MPA and for Model 2. The description of the models comprises the settings applied in TreeNet (lr-learning rate, nd – number of maximum nodes per tree, ts – percentage of data set applied for testing)

Description of the model	*Threshold*	*MPA [km2]*	*MPA [%]*	*Spearman-Rank Correlation*
Lr005_nd6_ts40	0.29	25,676	72.1	0.903
Lr01_nd6_ts20	0.23	26,089	73.2	0.939
Lr01_nd2_ts20	0.35	26,288	73.8	0.903
Lr01_nd2_ts40 (Model 2)	0.34	26,602	74.7	0.976

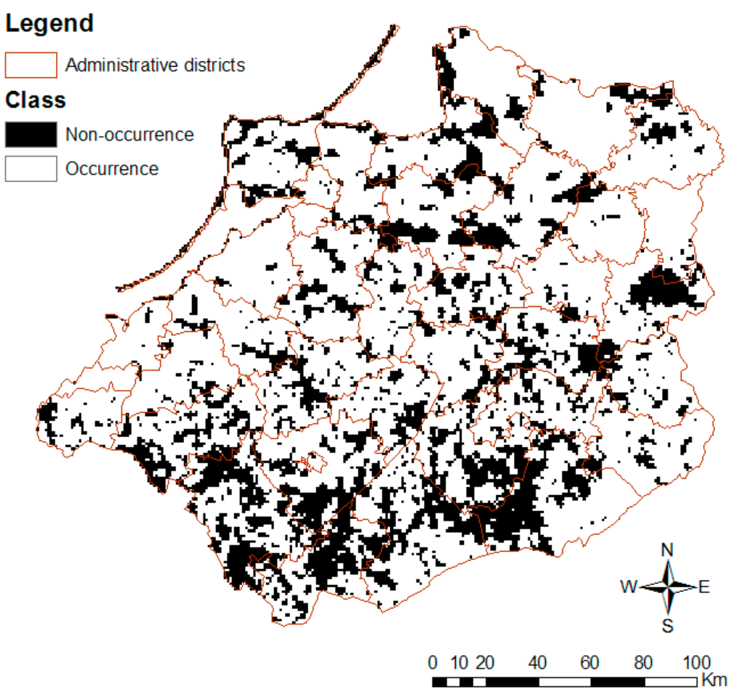

Figure 3.8: MPA for Model 2 applying a classification threshold of 0.34

The difference between the MPA for the created models was considered to be relatively small. Therefore and because the Spearman-Rank Correlation is an often used measure in the literature for evaluating presence-only models (Boyce et al. 2002), the model with the highest correlation value was selected as the best model (which was Model 2) although it did not show the lowest MPA.

3.2.2 Index of relative importance for the year 1939 applying Model 2

Applying the regular point grid (see section 2.1.4) and the subsequent conversion into a raster data set, a map showing an index of relative importance was generated for the complete study area for Model 2 (Figure 3.9). The predicted values are between 0.001 and 0.972. High values were predicted in a strip of about 20 km along the coastline as well as at a distance of about 150 km from the coastline at the southern edge of the study area. A further region with high indices of relative importance is found in the northwest of the survey area. However, it can be established that closed forests and lakes are classified as unsuitable breeding habitats since very low indices were predicted for these areas. Extensive forests and water bodies are situated especially in the south of the study area and correspond with the areas of a low predicted index of the presence of nest sites of White Storks. Locations in extensive regions of forest and water bodies as a rule also show a large distance to the nearest settlement, which also leads to a low predicted occurrence of breeding sites.

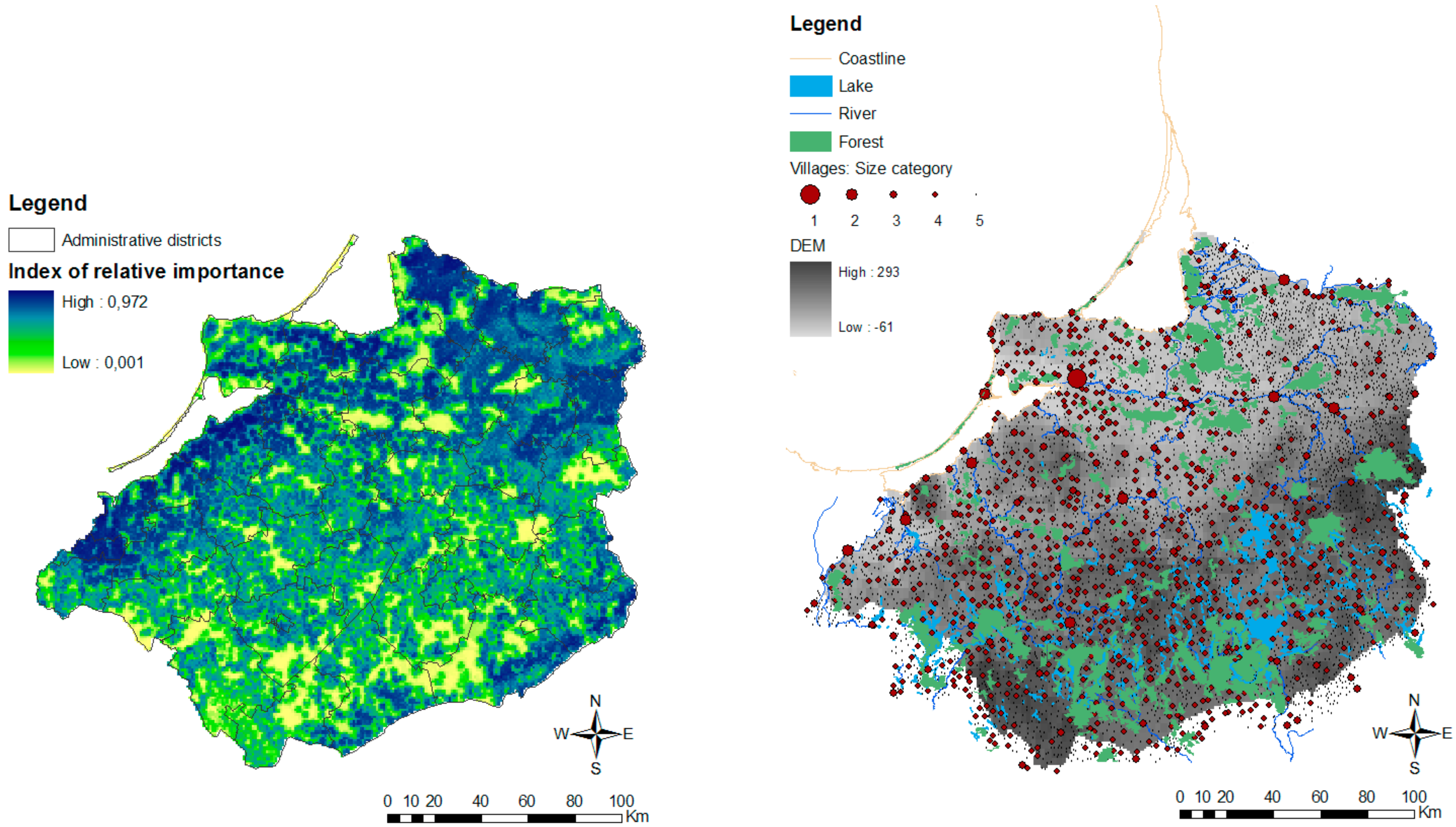

Figure 3.9: Predicted index of relative importance for 1939 applying Model 2 (left) and the study area with the applied GIS layers in comparison (right)

3.2.3 Index of relative importance for the period 1981-1992 applying Model 2_mod

Model 2, which is based on the results of the White Stork count from 1931 (Data set 2), was employed as the basis for a prediction of the occurrence of the White Stork for the period 1981-1992. This model was selected because unlike Model 1 (presence-available model) it is based on a presence-absence data set. Presence-absence models show a higher performance to distinguish between presence and absence of organisms (Boyce et al. 2002).

For the modified Model 2 (Model 2_mod in the following) the variable 'distance to settlement area, size range 1 to 4' was employed as the predictor variable instead of the variable 'distance to settlement area, size range 1 to 5'. Furthermore the predictor variable 'distance to watercourse' could not be used because unfortunately there was no digital data available about the course of the rivers in the years 1981-1992 at the required scale. All other predictor variables and the settings in TreeNet were applied as in Model 2.

Using Model 2_mod a map showing an index of relative importance was created for the year 1939 as well as for the period 1981-1992. The mean difference of the predicted values of the individual locations between both maps reads at -0.059. This means that on average marginally lower readings were predicted for the period 1981-1992. In Figure 3.10 the frequency distribution of the calculated difference between the predicted values for each location within the study area for the period 1981-1992 and the year 1939 is depicted.

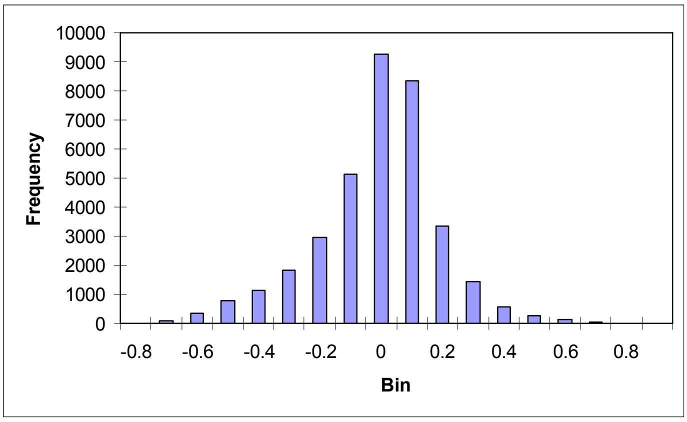

Figure 3.10: Difference between the predicted values for the period 1981-1992 and for the year 1939

In Figure 3.11 maps showing an index of relative importance for the year 1939 and the period 1981-1992 are shown. For the year 1939 the values between 0.002 and 0.994 were predicted. For the period 1981-92 these values are between 0.001 and 0.992.

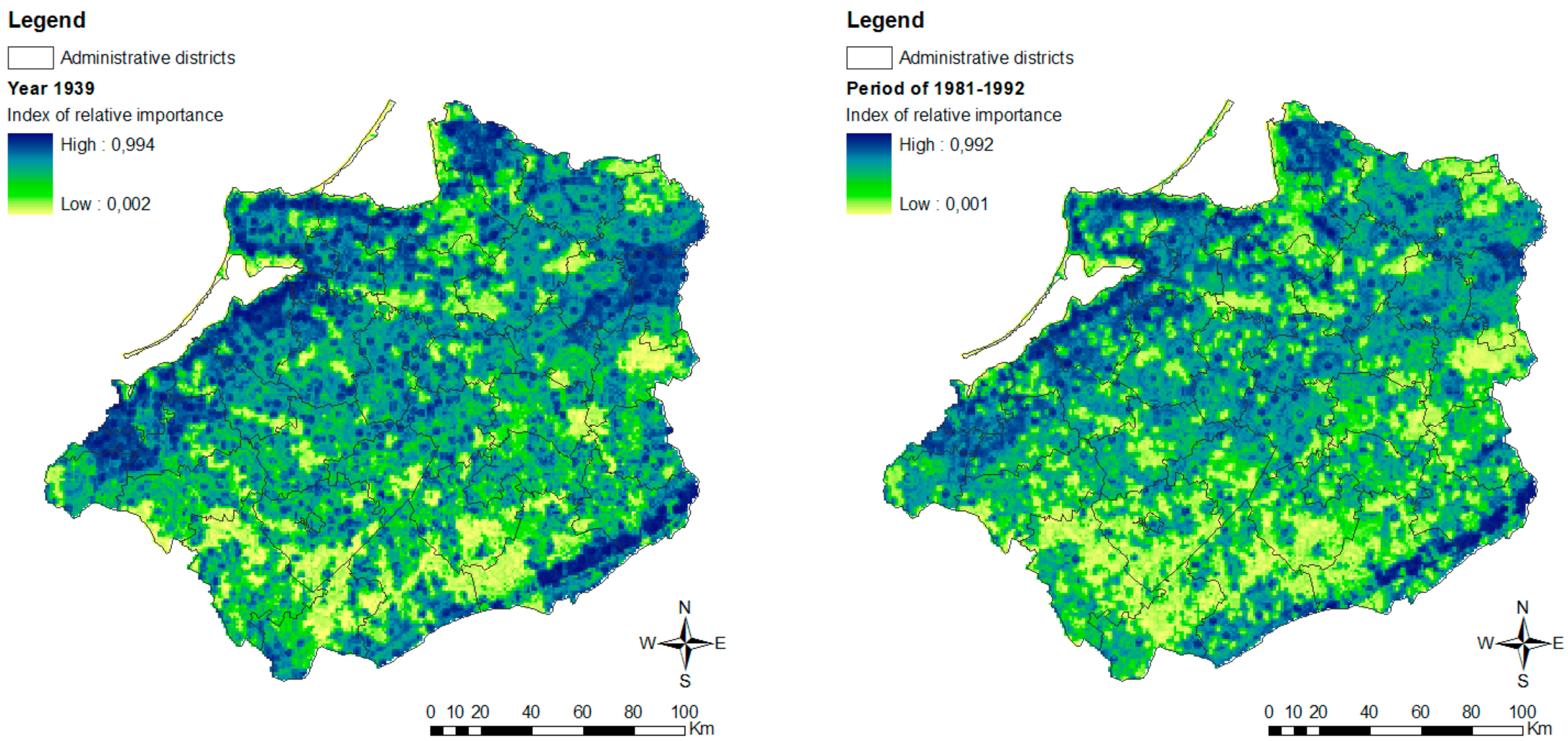

Figure 3.11: Index of relative importance for 1939 (left) as for the period of 1981-1992 applying Model 2_mod (right)

For a better comparison of the predictions for the year 1939 and the period of 1981-1992 a classification of the predicted values in each case with a threshold of 0.5 was carried out. According to Liu et al. (2005) this threshold was chosen because the data set showed a balanced sampling design consisting of 5,000 absence and 5,000 presence locations (prevalence = 0.5). Thus, all predicted values over 0.5 correspond with an occurrence, readings less or equal to 0.5 mean a non-occurrence. Subsequently, for every grid cell of the study area the difference between the values predicted for the period of 1981-1992 and for the year 1939 was calculated (Figure 3.12).

Grid cells with a value -1 corresponded to areas for which an occurrence of nest sites was predicted in 1939, for the period of 1981-1992, however, a non-occurrence. Grid boxes with a value of 1 correspond to the opposite. The grid boxes with the value 0 account for no change as to the prediction of occurrence and absence (non-occurrence) of breeding sites. In total there are 5601 grid boxes with the value -1 (= 5601 km^2) and 2771 with the value 1 (= 2771 km^2). Thus, in comparison of both outputs, a reduction of predicted potential breeding sites can be noticed for the period 1981-1992 compared with the year 1939 (15.7 % of the area of East Prussia). In 76.6 % of the area of East Prussia the suitability as breeding ground for the White Stork remained unchanged.

Comparing the underlying environmental variables for the year 1939 and the period of 1981-1992, a marked difference for these years can only be found in the variables ’forest’ and ’settlement area’: The forest area increased from 4890 km^2 in 1939 to 6383 km^2 in the period 1981-1992. The number of settlements dropped by about a third from 788 settlements (size categories 1 to 4) in 1939 to 243 settlements in the period 1981-1992. The surface area of the lakes with 1049 km^2 in 1939 and 1016 km^2 in the period 1981-1992 remained almost the same. For the coastline and the DEM the same data sets for both years were applied.

As seen in Figure 3.12 the changes in the predicted values are area-wide and spatially rather limited. In addition, the layer ‘forest’ for the two different points in time, which was the base of the most important predictor ‘distance to forest’ in Model 2_mod, is shown in detail in the same figure for the districts of Labiau und Insterburg and parts of the related districts. It is remarkable that in most of the cases where forest extended from 1939 to the years 1981-1992 the prediction for the relevant grid cells changed from occurrence to non-occurrence and vice versa.

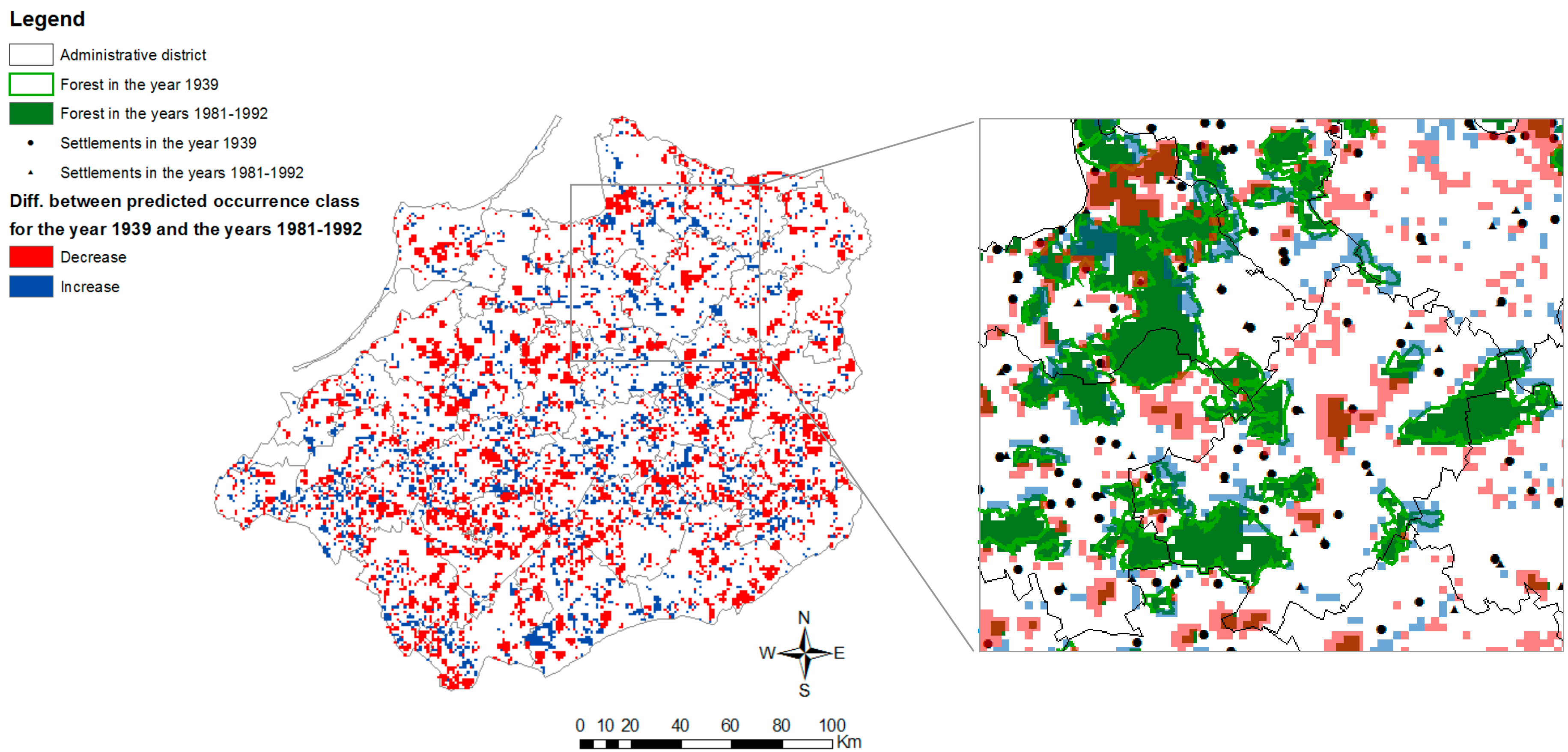

Figure 3.12: Difference between the predicted occurrence class for the year 1939 and the period 1981-1992 (left) and in detail the districts of Labiau and Insterburg with related districts and the layers ‘forest’ and ‘settlement area’ for both years in addition (right)

3.3 Model 3: Based on observed White Stork densities (administrative district scale)

Model 3 contained the predictor variables ‘percentage of pasture’, ‘percentage of forest’, ‘percentage of water’, ‘percentage of arable land’, ‘elevation’, ‘distance to coastline’ and ‘number of settlement areas per km^2, size range 1 to 5’ and was generated in TreeNet using the following settings:

learning rate:	0.05
number of maximum nodes per tree:	6
testing:	10-fold cross-validation.

Table 3.5 shows the variable importance of the applied predictor variables calculated by TreeNet. Both predictor variables ‘number of settlement areas per km^2, size range 1 to 5’ and ‘percentage of pasture’ had considerable bearing on the modeling. Notably less important were the two variables ‘percentage of arable land’ and ‘distance to coastline’.

Table 3.5: Importance of the predictor variables of Model 3

Predictor variable	*Relative Importance Score*
Number of settlement areas per km^2, size range 1 to 5	100.00
Percentage of pasture	90.67
Percentage of forest	54.40
Percentage of water bodies	53.97
Elevation	40.16
Percentage of arable land	27.29
Distance to Coastline	23.77

For the graphic visualization of the effect which the individual predictor has on the predicted response the partial dependence plots are shown in Figure 4.2.1 e-g.

The largest influence on the predicted response was contributed by the predictor ‘number of settlement areas per km^2, size range 1 to 5’ (Figure 3.13 a). Up to a number of 0.11 settlements per square km a negative partial dependence of -4 was existent. With a growing number of settlements the partial dependence increased continuously reaching zero point at about 0.13 settlements per square kilometer and had a high positive reading of 9 from about 0.19 on.

The predictor ’percentage of pasture’ (Figure 3.13 b) afforded a high negative partial dependence of -0.8 for districts that showed a reading lower than 15 %. From there the partial dependence rose considerably, reaching zero point at about 17 % and showing a positive partial dependence of 3 as from a percentage of 22 % pasture in the district.

The predictor variable ’percentage of forest’ (Figure 3.13 c) showed a high partial dependence for districts with a percentage of up to 15 % forest. The partial dependence declined with increasing percentage, reaching zero point in districts with roughly 17 % forest and showed a negative reading of nearly -3 as from 25 % forest in the district.

For the predictor ’percentage of water bodies’ (Figure 3.13 d) first a positive partial dependence of up to 1.5 for districts with a percentage of less than 0.8 % was apparent. Between 0.8 and 2 % it presented values at -0.7. In districts with a larger proportion of water bodies the partial dependence declined to values of around -1.5.

For the predictor ’elevation’ (Figure 3.13 e) there was a positive partial dependence of approx. 0.6 for districts with a medial altitude of up to 90 m. With increasing altitude the partial dependence dropped considerably and scored nearly zero at an average altitude of about 145 m. For districts with an altitude of more than 145 m the predictor showed a negative influence with a partial dependence of -1.3.

The predictor ’percentage of arable land’ (Figure 3.13 f) showed a positive partial dependence with a maximum value of 0.7 for districts with a percentage of between 42 % and 50 %. For districts with a minor or major percentage of arable land the predictor showed a lesser negative effect with a value of less than -0.15.

The predictor ’distance to coastline’ (Figure 3.13 g) showed the least influence on the model. For distances of less than 63 km to the coastline a negative partial dependence of maximally -0.45 exists. At greater distances the partial dependence becomes positive and achieves a reading of 0.35 at a distance of 80 km onwards from the coastline.

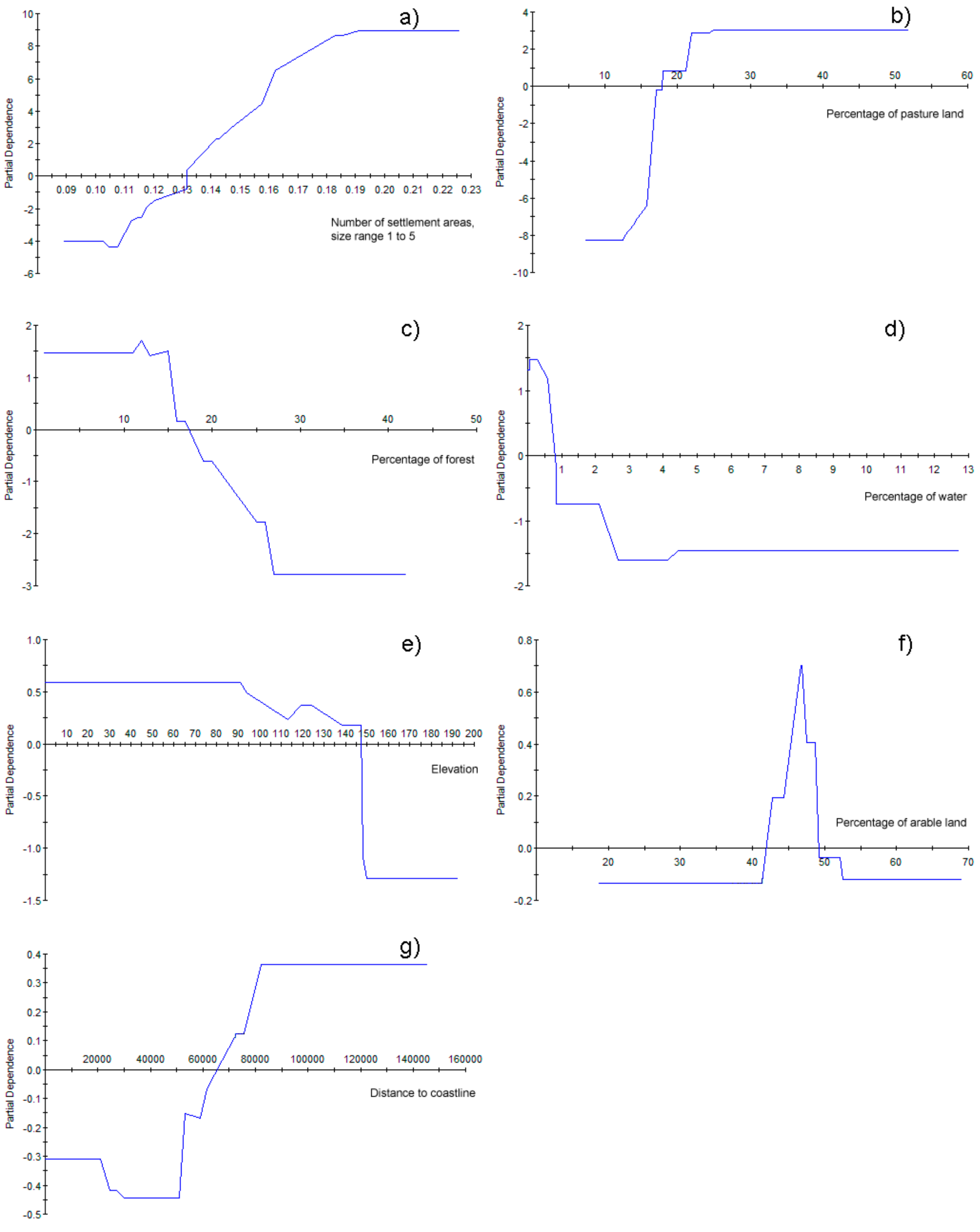

Figure 3.13: Partial dependence plots for the predictor variables employed in Model 3; a) 'number of settlement areas, size range 1 to 5', b) 'percentage of pasture land', c) 'percentage of forest', d) 'percentage of water', e) 'elevation', f) 'percentage of arable land', and g) 'distance to coastline'

3.3.1 Evaluation of Model 3

Since no independent test data set for the evaluation of the model was available the model could only be evaluated internally. To achieve this, a 10-fold cross-validation procedure was chosen as the test mode in TreeNet. Following, a prediction for the complete data set (n=37) was generated and various readings calculated for the definition of the goodness-of-fit of the model. Model 3 had a mean absolute error (MAE) of 5.22. The Coefficient of Efficiency had a value of 0.80. The MAE in per cent of the predicted value from the observed value amounts to 15.5 %.

3.3.2 Predicted White Stork densities for the year 1939

For a graphical display of the model a prediction was made for every administrative district for the year 1939 applying Model 3 and compared to the results of the White Stork census of 1934. This is shown in Figure 3.14 (for exact numbers see Appendix 8.5).

For the southern part of the survey area a low number of breeding pairs was predicted for the year 1939. High predicted densities of breeding pairs are to be found in the north west of the survey area. This tendency corresponds with the results of the count of 1934. In the comparison of the count with the values predicted in the model, however, it becomes obvious that the range of the values of the count is higher than the values of the model. The readings from observation are between 15.4 and 75.9 pairs per district, whereas the predicted values only range between 29.2 and 63.2 pairs per district.

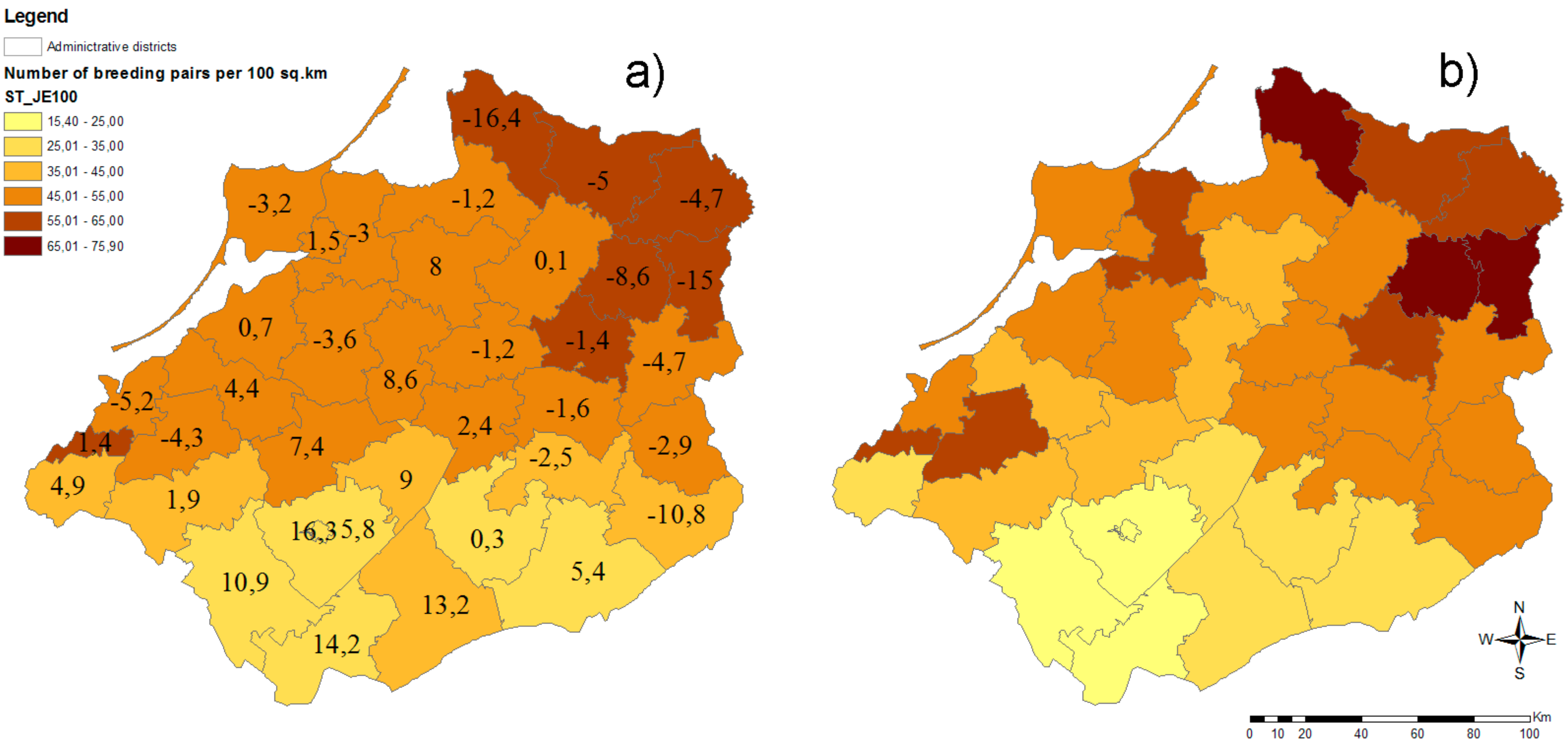

Figure 3.14: a) Predicted number of breeding pairs per administrative district for the year 1939 based on Model 3 and the difference to the count of 1934 in numerical value, and b) Number of breeding pairs as counted in the census of 1934

3.3.3 Predicted White Stork densities for the period 1981-1993

Due to the scale of the Baltic Sea GIS data Model 3 had to be modified for the estimation of White Stork densities for the period 1981-1993 with regard to the size categories of the settlement area applied. Instead of the variable ‘settlement areas per km^2, size range 1 to 5’ in Model 3, the variable ‘settlement areas per km^2, size range 1 to 4’ in Model 3_mod (MAE = 5.72) was applied. All other settings of both models correspond.

In Figure 3.15 the estimated densities of White Stork breeding pairs for the period 1981-1993 as well as the difference to the prediction for the year 1939 are shown. Especially in the central and southern section of the study area higher readings per district were predicted for the period 1981-1993. A reduction in the number of breeding pairs is predicted for the north and south as also the districts in the west of the survey area except for the district Stuhm.

If the total number of breeding pairs in the study area is calculated, the result is 16,395 in the year 1939 and 17,346 in the period of 1981-1993 applying Model 3_mod. The number determined in the count of breeding pairs in 1934 resulted in 16,092 (as in Tischler 1941). So, for the year 1939 already 303 breeding pairs more were predicted than actually counted in 1934. For the period 1981-1993 an increase in the number of breeding pairs in the total area of survey by 1255 is predicted in comparison to the count in the year 1934 and 951 breeding pairs in comparison to the year 1939.

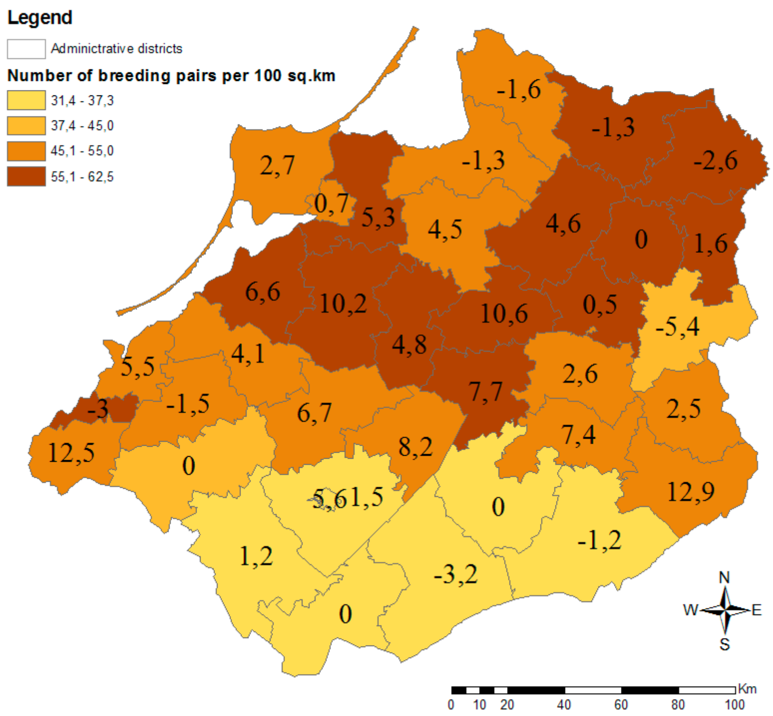

Figure 3.15: Estimated number of White Stork breeding pairs per 100 km^2 for the period of 1981-1993 applying Model 3_mod; labeled figures: Difference between estimated densities for 1939 and for the years 1981-1993

In Table 3.6 the environmental variables applied for both points in time are shown except the data set for the course of the coastline and the DEM, which were assumed to be unchanged.

Table 3.6: Mean ratio of the area of forest, water, pasture land and arable land as well as the mean number of settlements per area of administrative district for 1939 and for the period of 1981-1993

Variable	*Area in %*				*Mean number of Villages, size range 1 to 4*
	Forest	*Water*	*Pasture land*	*Arable land*	
1939	18.3	2.7	21.6	47.1	0.024
1981-1993	15.9	2.5	21.3	35.7	0.008

The largest change in surface area can be seen in the variable 'arable land’. Thus, the area of arable land has increased on average 11 % per district. Furthermore the number of settlements has decreased by nearly 1/3 to 243. The share of forest area has decreased by 2.4 % per district, whilst the area of lake land and pasture land has remained nearly unchanged.

4 Discussion

4.1 The White Storks' choice of nesting sites in former East Prussia

In Model 1 and Model 2 the three environmental variables 'forest', 'coastline' and 'settlement area' are most influential in the White Stork's choice of nesting place in East Prussia (Table 3.1 and Table 3.3).

In Model 3, which models the density of White Stork breeding pairs in the different administrative districts, the variables 'settlement area' and 'forest' also show a great effect on the modeling with a relative importance score of 100.00 or rather 54.40 (Table 3.5). The variable 'coastline' plays only a subordinate role. The variable 'pasture', which depicts how much area per administrative district is exploited agriculturally as pasture or meadow, is the second most important predictor in Model 3 having a relative importance score of 90.67. For the Model 1 and Model 2 no data on agricultural usage was available and so this aspect has to go unconsidered for now.

Surveys on the feeding ecology of the White Stork in agriculturally effected areas have shown that to a great extent the White Stork chooses pasture and meadow with a low height of vegetation for foraging (Struwe & Thomsen 1991, Böhning-Gaese 1992, Johst et al. 2001). Agricultural crop land plays an inferior role and shows only a greater availability of food supply at the time of working the land and after the harvest, i.e. at the end of the nestling phase (Bairlein & Henneberg 2000, Bäßler et al. 2000).

The 'forest' predictor is also of major importance in the modeling. Thus, the vicinity to forests has a negative effect on the probability of the White Stork occurrence. According to the created models dense woodland must be regarded as an inadequate habitat for the White Stork, since the optically orientated stalking predator has apparently difficulties in finding food supply there (Schüz 1933, Latus et al. 2000).

Beside the availability of food supply the choice of nesting ground is also determined by the vicinity to human settlements, since nests are often built on emerging buildings (Schulz 1993). Thus, in the year 1934 in former East Prussia (except for the districts of Rosenberg and Marienwerder) 92 % of the storks nested on roof tops (Tischler 1941). This explains the great influence the variable 'settlement area' has in the three models presented here.

The influence of the distance to coastline in the White Stork's choice of nesting ground (in Model 1 and Model 2) has not been examined yet. In particular in Model 1 this variable influences the distribution of the White Stork considerably and leads to the prediction of a high occurrence near the coastline. A reason for that could be the increasing continentality with growing distance to the coastline. According to Schüz (1933) this influences the point on time for arrival, breeding, and departure, but not the actual distribution of the White Stork population. However, considering the result of Model 2, there is a 'band of high abundances' with a high predicted occurrence 'probability' for a distance of 160 km from the coastline. This somewhat contradicts the influence of continentality on the distribution of the White Stork; perhaps this is caused by the spatial arrangement of the wetlands in the study area rather than continentality as such. Wet marshland, which can often be found near the coast, is a habitat preferred by the White Stork. Unfortunately the exact spatial arrangement of wetlands could not be considered in this study. This topic deserves still more research.

Moreover, the great influence of the coastline could be based on a combination of several factors. For instance, in East Prussia a specific landscape composition parallel to the coastline can be found. The height above sea level increases with a growing distance to the coastline. About 150 km from the coastline, the Baltic land ridge with an elevation of maximum 313 m above sea level parallels the coast. South of the Baltic land ridge extensive sandy territory expands, whereas in the north of East Prussia loamy soils can be found (Schüz 1933). This also influences the availability of food supply for the White Stork, since there is more agriculture on the sandy soil of the south, whereas in the loamy north there are more pastures and meadows for cattle breeding (Profus 1989 cited in Hinkelmann 1995). Furthermore, there are extended regions of lake-land and forest in the south of East Prussia and so at a greater distance to the coastline which makes them probably unsuitable as nesting grounds for the White Stork. Therefore, the distance to the coastline could be exploited as a proximal factor for the interaction of the different variables that influence the White Stork's choice of nesting ground. However, this may only be applicable to East Prussia with its characteristic landscape described here and should not simply be applied to other regions without checking the importance of this variable.

The specific structure of the data is possibly a further reason for the extremely high influence of the 'coastline' predictor on the occurrence of the White Stork in Model 1. Half of all the items of Data set 1 are within a distance of 41.5 km to the coast (the maximum distance to the coast is 164 km, Figure 2.1). This could be due to a more intensive banding activity in northern East Prussia and might lead to an over-estimation of the quality of the nesting grounds near the coast. However, the modeling approach chosen in this study should consider the Ecological Niche and help to overcome such problems of survey effort in space (see Kadmon et al. 2004).

4.2 Population Estimates

4.2.1 For the period of 1981-1992 at point scale

The predicted decrease in the occurrence of the White Stork for the period 1981-1992 results from the increased forest area and the reduced number of settlements in comparison to those of 1939 (section 3.2.3). Over time, it appears, that both variables have changed to the disadvantage of the White Stork (see partial dependence plots in Figure 8.7). This also corresponds with the information in the respective literature: the White Stork does not occur in dense forest areas and also it looks for the vicinity of human settlement areas for breeding grounds (Schüz 1933, Creutz 1988). White Storks appear to follow rural human settlements, and prefer the open habitats by which they are mostly surrounded.

However, it must be pointed out that both layers 'settlement area' from the year 1939 and the period 1981-1992, the best available layer and information at hand, can only be compared with care. The settlement areas for the year 1939 were digitized manually from topographical maps. The appropriate layer refers to settlements which exceed a certain number of inhabitants ('settlement area, size range 1 to 4', compare section 2.1.1.1.2) disregarding the size of expansion. In contrast, the variable 'settlement area' for the period 1981-1992 was derived from a raster grid with a cell size of 1 km. The settlement areas in this layer therefore must have shown a certain expanse. Most likely, this must be seen as the main reason for the lesser number of settlements in the period of 1981-1992 (243 in comparison to 788 settlements in the year 1939).

4.2.2 For the period of 1981-1993 at administrative district scale

Applying Model 3_mod an increase in the number of breeding pairs of about 5% was estimated for the period of 1981-1993. Comparing the applied environmental variables only the variables 'settlement area' and 'arable land' have changed considerably. The number of settlements (see above) and also the area of arable land have decreased. Considering the partial dependence plots for Model 3 _mod (see Figure 8.8) both findings have a positive effect on the occurrence of White Storks in the individual administrative districts.

The actual effect of the variable ’settlement area’ on the occurrence of White Storks in Model 3_mod is the opposite of its effect in Model 3 (and also Model 1 and Model 2), since in the latter a decrease in the number of settlement areas per administrative district shows a negative effect on the occurrence of White Storks. Despite it only shows a relative importance score of 43.3 % in Model 3_mod in comparison with 100 % in Model 3. This could indicate an unstable model. However, it is most likely attributed to the fact that due to the reasons stated in section 3.3.3 only the variable ’settlement area, size ranges 1 to 4’ was adopted as an explanatory variable into Model 3_mod. Smaller settlements and villages of the size range 5 and 6, in the vicinity of which the White Stork prefers to breed, were excluded. That is why this variable only plays a minor role in Model 3_mod in contrast to Model 3.

4.2.3 Comparison between point scale and administrative district scale

Both models predict a different development of the population of the White Stork: according to Model 2_mod the area of the predicted occurrences decreases, whilst Model 3_mod predicts an increase in the densities of White Storks in the administrative districts.

Both models, however, are only comparable in a limited way, since, for one, they were developed using different scales. Model 2_mod is based on the exact location of the breeding ground (point scale). So the close proximity of the breeding ground can be described exactly, applying the available environmental variables. In Model 3_mod the density of the White Stork per administrative district was modeled. The environmental variables could only be applied as average readings for the different districts. Thus Model 3_mod gives a real population estimate (as density) while Model 2_mod shows the suitability of different locations for the White Stork as breeding ground. The differing development of the predicted White Stork population can be caused by a specific spatial arrangement of the described habitat features which will not be detected when modeling at administrative district scale. Therefore the aim of such models as presented here should be to work at the smallest possible resolution.

Furthermore for the creation of Model 2_mod on the point scale the variable 'pasture' was unavailable since the location of pasture land could not be clearly defined on the used topographical maps (see section 2.1.1.1.2). Grassland, however, is the most important feeding habitat of the White Stork (Johst et al. 2001, Latus & Kujawa 2005). Beside the settlements this variable subsequently proved to be the most important predictor for the model on the district level (Model 3).

Another limitation in the comparison of both models is brought by the application of variables which were surveyed differently and taken from different years (see section 2.1.1 and section 2.2.1) to build the 'historic' models.

4.2.4 Comparison with inventories done in the 1980s and 1990s

A review of the prediction of Model 2_mod was not possible since there were no inventories available for the 1980s or 1990s from which the exact location of the stork breeding ground could be taken.

A direct possibility for a comparison of a prediction with an actually assessed inventory of White Storks is only possible for Model 3_mod, which predicted the densities of White Stork breeding pairs per administrative district. Thus, from the results of the international census of White Storks Profus (2005) estimated an inventory of about 7,000 breeding pairs for the year 1984 and between 8,200 and 8,600 breeding pairs for the years 1994/95 for the part which today is the Polish part of East Prussia (more or less the equivalent of today's Polish province (voivodship) of Warmia-Masuria. In an inventory of the years 1991/92 Grishanov (cited in Hinkelmann 1995) determined 1,270 breeding pairs for the Russian section of the former province of East Prussia (the current Kaliningrad region). Due to repeated alterations of the boundaries in the region and the exclusion of the districts Rosenberg and Marienwerder from the study area no exact inventory figures can be established. It can be seen, however, that the total inventory in the 1990s with a maximum total of 9,870 (= 8,600 + 1,270) breeding pairs in the study area is distinctly lower than in the year 1934 with its 16,092 nesting pairs. This is mainly due to the fact that the determined inventory in the Russian exclave was very low. For the year 1934 Profus (cited in Hinkelmann 1995) estimated approximately 8,000 breeding pairs for Kaliningrad. As reasons for the decline Hinkelmann (1995) mentions the dwindling numbers of farms and the mechanization of agriculture in combination with the use of chemicals like pesticides and fertilizer after World War II in the Kaliningrad region. However since the breakdown of the Soviet Union the population of the White Stork is increasing again due to the abandonment of the intensive agriculture (Hinkelmann 1995). It can be assumed that for this reason better food conditions are available for the White Storks.

The prediction made for the period 1981-1993 applying Model 3_mod totaled 17,346 White Stork breeding pairs for the former province of East Prussia, excluding the districts of Rosenberg and Marienwerder. This surpasses by far the maximum inventory count for this region by the factor ‘1.8’.

It is believed that the models created here are more likely to describe the capacity of the habitat taking the applied environmental variables into consideration. According to this logic, large surfaces of the former East Prussia are a potential habitat for the White Stork. Merely in the southern part there are larger areas with extensive forests and lake area, which are unsuited for the White Stork (details above). The reason for the high number of White Storks even in today’s inventory, especially in the Polish section (accounting for 30 breeding pairs per 100 km^2 in the years 1994/95, according to Schimkat (2006)) is to be seen in the extensive agriculture with an open landscape and the highly structured habitat allowing for an abundance of species, which offers the White Stork enough food supply (Peterson et al. 1999).

The population of the White Stork varies considerably with an annual fluctuation, which cannot solely be assigned to the character of the habitat. And so Peterson et al. (1999) show that the comparison of only two successive inventories can be misleading and only the observation of the development of the population within a larger period of time allows for more exact conclusions.

Successful breeding as well as the development of the population is influenced by the specific character of the habitat in the breeding area. This also includes the predominant climatic situations in the breeding area which are not explicitly included in the models presented here. However, climatic factors such as temperature and precipitation are significantly influenced by continentality or rather marine characteristics and are indirectly considered in the models by the predictor variable ’distance to coastline’.

Unfavorable weather conditions such as periods of rain or drought as well as fluctuations in the population of prey still remain unconsidered in the models. They show their influence in the successful breeding mainly by the commencement of breeding, the mortality of nestlings as well as the infestation by parasites (Schüz 1933, Creutz 1988) and can thereby lead to a limited number of breeding pairs in the following years (so called 'interruption years').

Beside the conditions in the breeding area the population is also influenced by the conditions in the wintering areas and during migration. However the wintering area of the White Storks migrating east (White Storks from East Prussian) stretches over several climate zones from Uganda to South Africa so that they might always find suitable conditions somewhere in their wintering area (Schulz 1999, Tryjanowksi 2006). Dangers during migration are manifold and can originate from anthropogenic causes (e.g. shooting, electrocution), bad thermic conditions, thunderstorms and lack of food supply (Sproll 2000). More studies on these effects are needed. Here the first baseline information for such investigations was provided focusing on the conditions at the breeding ground alone.

4.3 Restrictions and constraints of the models

4.3.1 Presence/Pseudo-Absence approach

Both data sets applied for the modeling on a point scale (Data set 1 and 2) contained no real absence locations to show at which locations the White Stork is actually not nesting (=confirmed absences). Data set 1 only showed presence locations which registered as retraced locations by banded White Storks and expanded into the banding data set of the ornithological station 'Vogelwarte Radolfzell'. For the modeling 578 locations were randomly scattered over the study area (pseudo-absence locations). The pseudo-absence locations may also arbitrarily contain locations suited as breeding sites for the White Stork (Boyce 2006).

Although data set 2 is based on a nearly complete mapping of all the White Stork breeding grounds in East Prussia in 1931 (section 2.1.2.2) a representation was chosen, however, in which the number of nesting grounds were both combined and presented in one community centre (Figure 2.2). This is why the exact locations (in coordinates) can only be deduced with some deviation. Locations were chosen at random that were at a distance of at least 2 km from the registered nesting site. For reasons just mentioned, confirmed absence locations may be at sites where a pair of White Storks has been breeding. This can lead to a decrease in the performance of the model, if absence locations turn out in fact to be suitable nesting grounds (Fielding & Bell 1997, Pearce & Boyce 2006).

4.3.2 Selected variables

It must be assumed that not all biologically relevant variables are included in the modeling (e.g. see Gottschalk 2002). This may happen because important variables are not recognized as such or also because they cannot be surveyed (complexity, time and effort). This is a common feature in current GIS models (e.g. see Guisan and Zimmermann 2000) but supposed to improve further. Here, modeling methods were started and a culture within the White Stork community emphasizing the importance of such available GIS data was set up. Especially in historic modeling only a limited choice of variables is commonly possible to apply, since only data can be used that is already available. A survey of data which have proven to be relevant for the species to be modeled is not always possible with hindsight. Instead, historical data sets originate from museums or archives (see Graham et al. 2004 for overview and applications). Often, it is not known who did the surveys and how the data was surveyed (Engler et al. 2004, Elith et al. 2006). This includes a multitude of sources of error concerning the quality of data applied (Elith et al. 2006, Hüttmann in press), and therefore limits a direct inference. The predictors are thus primarily showing correlations (Manly et al. 2002), and more hypothesis and on the ground work is required to assess their validity further.

4.4 Perspective for conservation and further research

Since the beginning of population assessments at the end of the nineteenth century a high number of nesting White Storks has been documented for the region of former East Prussia. In comparison, the occurrence of the White Stork in large parts of Central Europe has steadily declined. According to surveys on the population dynamics of east migrating White Storks by Schimkat (2006) the East German populations depend on the immigration of individuals by dismigration from South East and Eastern Europe to conserve their population. Thus, clear sources and sink populations can be delineated. It is the sources that keep the overall population alive. The protection of the Eastern European populations of White Storks therefore would have a positive effect on the development of the entire populations in Central Europe.

Currently, in the course of the expansion of the European Union to the east, the Polish White Stork population is also threatened by the intensification of agriculture and the application of increased amounts of pesticides and fertilizers (Schulz 1999, Tryjanowski et. al 2006). Fragmentation, e.g. caused by extensive road networks, is another habitat feature to consider.

The consequences of anthropogenic influences on the behavior of White Storks can already be seen in the choice of nesting ground. Thus, the White Stork is now building its nest between electric poles rather than on the roofs of houses as it still did some decades ago (e.g. Daniluk et al. 2006, Rubacha & Jerzak 2006). Effects of the electrification and further urbanization of the landscapes need to be taken very seriously. Effects such as global change, pollution, toxicology, hunting and the entire flyway and wintering ground effects are not addressed in this model, yet, but should contribute further to a cumulative impact assessment of White Storks for this important population.

Data are a key for modeling and for an informed decision-making (Hüttmann 2005). Any data, e.g. surveys, habitats, population status and health, and literature, relating to the management of White Storks need to be freely available online. Centralized databases are to be developed that handle this component efficiently and which do not constrain any progress.

The construction of a further model offers itself as the continuation of the study presented here. In such a model data on the current occurrence of the White Stork should be linked to the presently predominant environmental parameters. Relevant explanatory GIS layers, which should additionally be included in this model, are e.g. power supply networks, use of pesticides and pollution, the structure of roofs on houses, nitrogen input as well as the fragmentation of the landscape. A comparison of 'historic', 'current', and 'future' habitat models may help to realize differences in behavior and assess them in a pro-active fashion before they occur (see Onyeahialam et al. 2005, Hüttmann et al. 2005). This knowledge may be used to define adequate strategies for the preservation of nature in order to preserve the occurrence of White Storks. Considering the environmental situation, such efforts are crucial to pursue if the White Storks should be safeguarded in the future.

To carry out a spatially explicit model of the relation between the nesting ground and the state of the environment it is of great importance that the locations of White Stork nesting grounds are represented as accurately as possible (as geographical coordinates) and come from a controlled research design. In the future it must be avoided that the results of surveys for larger areas (e.g. municipalities, counties etc) are pooled and merged, especially since such administrative units can change in the course of history and thus impede a correct evaluation. Also, a documentation of nesting grounds which show an absence of the White Stork or an abandonment of nesting grounds is important for a significant presence- absence modeling to be used for a sustainable future, including viable White Stork populations.

5 Conclusion

This study shows that opportunistic, historical data sets can be used successfully with GIS and with a robust machine-learning model method (TreeNet) to derive robust species habitat relation models and new biological knowledge.

Using the example of the White Stork a model for the region of former East Prussia was created across political borders, cultures and management regimes.

The models generated here demonstrate that the vicinity of human settlements as well as the availability of sufficient feeding habitats (grass land with low vegetation, no close forest regions) have a strong influence on the distribution of the White Stork in this region.

Without further testing, the models generated are only valid for the region of former East Prussia so far. However, in order to transfer the relations found here between the distribution of the White Stork and the state of the habitat to other regions as well the models would need to be checked using data sets surveyed in other regions. Model assessments are crucial. Only then can a statement on the general validity of the model be made. The applicability and value of this approach would be rather large because a global and quantitative White Stork nesting model could be achieved, which would likely improve much of its sustainable management.

The models can be improved using additional variables. Thus, the quality of the habitat can be more closely described from the perspective of the White Stork with data on prey availability, the level of precipitation in June (when the young White Storks are especially susceptible to wet conditions) or the temperature for example.

According to this aim, modeling should be carried out on different scales, since important influencing factors might be ignored or rather the influence of factors can vary with the scale (Gottschalk 2002, Hüttmann & Diamond 2006).

To create species distribution models with the highest possible accuracy it is important that all relevant stakeholders (e.g. public authorities, archives, nature conservation agencies, NGOs and scientists) co-operate and mutually exchange data: only then can it be achieved that already existing data is exploited more effectively than currently done. The ideal state would be a freely accessible online data bank in which relevant data could be queried in digital form. This would further contribute to an improved and sustainable management of White Storks, their habitats world-wide and natural resources as a whole.

6 Acknowledgement

Here I would like to thank everyone who has offered professional and personal support in making this survey a success.

First and foremost my gratitude is extended to Prof. Dieter Wallschläger (University of Potsdam, Germany), who initiated this study and offered me kind support.

A great thank you to Falk Hüttmann, who made a work place available to me at the EWHALE lab of the University of Alaska and with his knowledge and his ideas greatly contributed to the outcome of the survey. My thanks also goes out to all the people at the EWHALE lab who accepted me so willingly into their team.

In the months of toilsome data retrieval the uncomplicated assistance of two people in particular restored my courage and enforced me to continue my way. For that I thank Wolfgang Fiedler, managing director of Vogelwarte Radolfzell, who quickly and non-selfishly made the data from the banding data bank available to me. Furthermore I thank Kai-Michael Thomsen of the Michael-Otto-Institut at NABU - Naturschutzzentrum Bergenhusen, (Germany) for the supply of literature and his endeavours to make the results of an inventory in the East Prussian administrative district of Gerdauen available to me. In addition Christoph Hinkelmann, managing director of the Ostpreußisches Landesmuseum (East Prussian County Museum) in Lüneburg (Germany), as well as Gerd Dahms provided literature and information on East Prussia and on the White Stork.

I thank the team of Salford Systems Ltd. for the provision of the software TreeNet and their flexible support.

Also my thanks go out to Dr. Boris Schröder (University of Potsdam, Germany) for the gratuitous provision of his personally created program ROC_AUC.

A great helper in writing this diploma thesis in English was Michael Schmitz. Also I am very grateful to Fabian Schmitz and my family for their caring support.

7 References

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

Bairlein, F. & Henneberg, H. R. (2000): *Der Weißstorch (Ciconia ciconia) im Oldenburger Land*. – Isensee, Oldenburg.

Barran, F. R. (1988): *Städte-Atlas Ostpreußen*. – Rautenberg-Verlag, Leer.

Bäßler, R., Schimkat, J. & Ulbricht, J. (2000): *Artenschutzprogramm Weißstorch in Sachsen*. – Sächsisches Landesamt für Umwelt und Geologie, Dresden.

Berthold, P., Kaatz, M. & Querner, U. (2004): Long-term satellite tracking of white stork *(Ciconia ciconia)* migration: constancy versus variability. – *Journal of Ornithology*, 145:356-359.

Beyer, H. L. (2004): *Hawth's Analysis Tools for ArcGIS.* – Available at http://www.spatialecology.com/htools

Blotzheim, G. v. (1987): *Handbuch der Vögel Mitteleuropas*. – Aula-Verlag, Wiesbaden.

Böhning-Gaese, K. (1992): Zur Nahrungsökologie des Weißstorches *(Ciconia ciconia)* in Oberschwaben: Beobachtungen an zwei Paaren. – *Journal of Ornithology*, 133:61-71.

Boyce, M. S., Vernier, P. R., Nielsen, S. E., & Schmiegelow, F. K. A. (2002): Evaluating resource selection functions. – *Ecological Modelling*, 157:281-300.

Braun, M. (1908): Die Nistweise des Storches. – *Schriften Physikalisch-Ökonomische Gesellschaft Königsberg*, 49:280-290.

Buckland, S. T., Burnham, K. P. & Augustin, N. H. (1997): Model Selection: An Integral Part of Inference. – *Biometrics*, 53:603-618.

Burnham, K. P. & Anderson, D. R. (1998): *Model selection and inference: a practical information-theoretic approach.* – Springer, New York.

Chernetsov, N., Berthold, P. & Querner, U. (2004): Migratory orientation of first-year White Storks (Ciconia ciconia): inherited information and social interaction – *Journal of Experimental Biology*, 207:937-943.

Creutz, G. (1988): *Der Weißstorch*. – Westarp Wissenschaften, Ziemsen.

Dallinga, J. H. & Schoenmakers, S. (1987): Regional decrease in the number of White Storks *(Ciconia ciconia)* in relation to food resources. – *Colonial Waterbirds*, 10:167-177.

Daniluk, J., Korbal-Daniluk, A. & Mitrus, C. (2006): Changes in population size, breeding success and nest location of local White Stork *Ciconia ciconia* population in Eastern Poland. – In: Tryjanowski P., Sparks T. H. & Jerzak L. (eds). *The White Stork in Poland: studies in biology, ecology and conservation.* Bogucki Wydawnictwo Naukowe, Poznan 2006.

Elith, J., Ferrier, S., Hüttmann, F. & Leathwick, J. (2005): The evaluation strip: A new and robust method for plotting predicted responses from species distribution models. – *Ecological Modelling*, 186:280-289.

Elith, J., Graham, H., Anderson, P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, J., Hüttmann, F., Leathwick, R., Lehmann, A., Li, J., Lohmann, G., Loiselle, A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, C. M., Townsend Peterson, A., Phillips, J., Richardson, K., Scachetti-Pereira, R., Schapire, E., Soberon, J., Williams, S., Wisz, S. & Zimmermann, E. (2006): Novel methods improve prediction of species' distributions from occurrence data. – *Ecography*, 29:129-151.

Engler, R., Guisan, A. & Rechsteiner, L. (2004): An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. – *Journal of Applied Ecology*, 41:263-274.

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.

Franklin, J. (1995): Predictive vegetation mapping: Geographic modelling of biospatial patterns in relation to environmental gradients. – *Progress in Physical Geography*, 19:474-499.

Frenz, L. (1995): *Störche in Deutschland*. – Scheibel, Berlin.

Friedman, J. H. (1999): Stochastic Gradient Boosting. – In: *Technical Discussion of TreeNet*. http://www.salford-systems.com/treenet.html

Galcov, V. (1999): Besonderheiten des Kaliningrader/Königsberger Gebiets. – In: *Annaberger Annalen Nr.7*, Hermann, A. und Lepa, A. im Auftrag des Baltischen Christlichen Studentenbundes, Bonn.

Gaupp, W. (1936): Der Bestand des Weißen Storches in Ostpreußen 1934. – *Beiträge zur Fortpflanzungsbiologie der Vögel mit Berücksichtigung der Oologie* 12:41-43.

GLOBE Task Team and others (Hastings, David A., Paula K. Dunbar, Gerald M. Elphingstone, Mark Bootz, Hiroshi Murakami, Hiroshi Maruyama, Hiroshi Masaharu, Peter Holland, John Payne, Nevin A. Bryant, Thomas L. Logan, J.-P. Muller, Gunter Schreier, and John S. MacDonald), eds. 1999. *The Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, Version 1.0*. National Oceanic and Atmospheric Administration, National Geophysical Data Center, 325 Broadway, Boulder, Colorado 80303, U.S.A. Digital data base on the World Wide Web (URL: http://www.ngdc.noaa.gov/mgg/topo/globe.html) and CD-ROMs.

Gottschalk, T. K. (2002): *A remote sensing and GIS-based model of avian species habitat and its potential as a part of an environmental monitoring programme*. – Materialien Umweltwissenschaften Vechta 14, Vechta.

Gräber, R. (2006): Towards a Biodiversity Assessment of the Pacific Rim: Large-scale GIS-Modeling of Brown Bear Distribution (Canada, Alaska, Russian Far East and Japan) in Estuaries using compiled Coastal Data. – *Diploma Thesis*, University of Hannover, Germany.

Graham, C. H., Ferrier, S., Hüttmann, F. Moritz, C. & Peterson, A. T. (2004): New developments in museum-based informatics and applications in biodiversity analysis. – *Trends in Ecology and Evolution* 19: 497-503.

Grishanov, G. B. (1989/90): Verbreitung und Bestandsentwicklung des Weißstorches in der Kaliningrad-Region. – In: Störche – Vorkommen, Ökologie, Schutz – *Referate der I. (Tallin, 1989) und II. (Minsk, 1990) Unions-Konferenz der Storchenarbeitsgruppe der Ornithologischen Gesellschaft der Sowjetunion.* Deutsche Übersetzung im Auftrag des NABU Institut für Wiesen und Feuchtgebiete – Naturschutzzentrum Bergenhusen.

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J. M. C., Aspinall, R., & Hastie, T. (2006a): Making better biogeographical predictions of species' distributions. – *Journal of Applied Ecology*, 43:386-392.

Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N. G., Lehmann, A. & Zimmermann, N. E. (2006b): Using Niche-Based Models to Improve the Sampling of Rare Species. – *Conservation Biology*, 20:501-511.

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.

Hall, M. J. (2001): How well does your model fit the data? – *Journal of Hydroinformatics*, 3:49-55.

Hastie, T., Tibshirani, R., & Friedman, J. (2003): *The elements of statistical learning: data mining, inference, and prediction*. – Springer, New York.

Hertz-Eichenrode, D. (1969): *Politik und Landwirtschaft in Ostpreußen 1919-1930*. – Westdeutscher Verlag, Köln und Opladen.

Hinkelmann, C. (1995): Der Weißstorch *(Ciconia ciconia)* im ehemaligen Ostpreußen. – *Blätter Naumann-Museum*, 15:24-52.

Hornberger, F. (1943): Einige Ergebnisse zehnjähriger Planarbeit im „Storchforschungskreis Insterburg" der Vogelwarte Rossitten. – *Journal of Ornithology*, 91:341-355.

Hosmer, D. W. & Lemeshow, S. (2000): *Applied logistic regression*. – John Wiley & Sons, New York.

Hubley, J., Neubauer, D., Paciulli, L. & Hüttmann, F. (in press): Using TreeNet for identifying management thresholds of Mantled Howling Monkeys' habitat preferences on Ometepe Island, Nicaragua, on a tree and home range scale – *Science Journal International*.

Hüttmann, F. (in press): Constraints, Suggested Solutions and an Outlook towards a New Digital Culture for the Oceans and beyond: Experiences from 5 predictive GIS Models that contribute to Global Management, Conservation and Study of Marine Wildlife and Habitat. – In: *Vanden Berghe, E. et al. (Eds). 2005. Proceedings 'Ocean Biodiversity Informatics' – International Conference on Marine Biodiversity Data Management*, Hamburg, Germany, 29 November-1 December 2004 IOC Workshop Report BSH/ VLIZ Special Publication 20.

Hüttmann, F. (2005): Research and management viewpoint: Databases and science-based management in the context of wildlife and habitat: Toward a certified ISO standard for objective decision-making for the global community by using the internet. – *Journal of Wildlife Management*, 69:466-472.

Hüttmann, F., Franklin, S. E. & Stenhouse, G. B. (2005): Predictive spatial modeling of landscape change in the Foothills Model Forest. – *The Forestry Chronicle* 81:1-13.

Hüttmann, F. & Diamond, A. W. (2006): Large-scale effects on the spatial distribution of seabirds in the Northwest Atlantic. – *Landscape Ecology*, 21:1089-1108.

Johst, K., Brandl, R., & Pfeifer, R. (2001): Foraging in a patchy and dynamic landscape: Human land use and the White Stork. – *Ecological Applications*, 11:60-69.

Kadmon, R., Farber, O. & Danin, A. (2004): Effect of roadside bias on the accuracy of predictive maps produced by Bioclimatic Models – *Ecological Applications*, 14:401-413.

Latus, C. & Kujawa, K. (2005): The effect of land cover and fragmentation of agricultural landscape on the density of White Stork (*Ciconia Ciconia L.*) in Brandenburg, Germany. – *Polish Journal of Ecology*, 53:535-543.

Latus, C., Kujawa, K. & Glemnitz, M. (2000): The influence of landscape structure on White Stork's *Ciconia ciconia* nest distribution. – *Acta Ornithologica, 35:97-102.*

Legates, D. R. & McCabe, J. W. Jr. (1999): Evaluating the use of 'goodness-of-fit' measures in hydrologic and hydroclimatic model validation. – *Water Resources Research*, 35:233-241.

Manly, B. F. J., McDonald, L. L., Thomas, D. L., McDonald, T. L., & Erickson, W. P. (2002): *Resource Selection by Animals: statistical analysis and design for field studies*. – Kluwer, Boston.

Onyeahialam, A., Hüttmann, F. & Bertazzon, S. (2005). Modeling sage grouse: Progressive computational methods for linking a complex set of local biodiversity and habitat data towards global conservation statements and decision support systems. *Lecture Notes in Computer Science (LNCS) 3482, International Conference on Computational Science and its Applications (ICCSA) Proceedings Part* III:152-161.

Pearce, J. & Boyce, M. S. (2006): Modelling distribution and abundance with presence-only data. – *Journal of Applied Ecology*, 43:405-412.

Pearce, J. & Ferrier, S. (2000): Evaluating the predictive performance of habitat models developed using logistic regression. – *Ecological Modelling*, 133:225-245.

Peterson, U., Jakubiec, Z., Okulewicz, J., Profus, P., & Haecks, J. (1999): Der Weißstorchbestand im Kreis Ketrzyn (Rastenburg), Masuren/Polen. – In: *Weißstorch im Aufwind? - White storks on the up?* - Proceedings, Internat. Symp. on the White Stork, Hamburg 1996, edited by H.Schulz, pp. 395-412. NABU (Naturschutzbund Deutschland e.V.), Bonn.

Profus, P. (2005): Bestandsveränderungen des Weißstorchs *Ciconia ciconia* in Polen. – *Charadrius*, 41:12-20.

Pulliam, H. R. (1988): Sources, sinks, and population regulation. – *American Naturalist*, 132:652-661.

Rohde, G. (1957): *Die Ostgebiete des Deutschen Reiches*. – Holzner-Verlag, Würzburg.

Rubacha, S. & Jerzak, L. (2006): Changes in the White Stork *Ciconia ciconia* population number, density and breeding places in Zielona Gora region 1926-2004. – In: Tryjanowski P., Sparks T. H. & Jerzak L. (eds). *The White Stork in Poland: studies in biology, ecology and conservation.* Bogucki Wydawnictwo Naukowe, Poznan 2006.

Salford Systems Ltd. Ltd. (2003). *TreeNet 2.0 User Manual - Software TreeNet*. San Diego, CA. http://www.salfordsystems.com/

Schimkat, J. (2006): Untersuchungen der Populationsdynamik von Regionalbeständen ostziehender Weißstörche *(Ciconia ciconia)* mittels eines Simulationsmodells – *Dissertation*, Universität Potsdam, Germany.

Schröder, B. (2000): Zwischen Naturschutz und Theoretischer Ökologie: Modelle zur Habitateignung und räumlichen Populationsdynamik für Heuschrecken im Niedermoor. – *Dissertation*, Technische Universität Braunschweig, Germany.

Schröder, B. & Reineking, B. (2004): Modellierung der Art-Habitat-Beziehung - ein Überblick über die Verfahren der Habitatmodellierung. – In: Dormann, C. F. et al. (eds): *Habitatmodelle - Methodik, Anwendung, Nutzen*, pp. 5-26. UFZ-Berichte, 2004

Schulz, H. (1993): *Der Weißstorch - Lebensweise und Schutz*. – Naturbuch-Verlag, Augsburg.

Schulz, H. (1999): Der Weltbestand des Weißstorchs (Ciconia ciconia): Ergebnisse des 5. Internationalen Weißstorchzensus 1994/95. – In: *Schulz, H. (Hrsg.) (1999): Weißstorch im Aufwind? - White Storks on the up?* - Proceedings, Internat. Symp. on the White Stork, Hamburg 1996. NABU (Naturschutzbund Deutschland e.V.), Bonn: 335-350.

Schumacher, B. (1977): *Geschichte Ost- und Westpreußens*. – Holzner-Verlag, Würzburg.

Schüz, E. (1933): Der Bestand des Weißen Storches *(Ciconia c. ciconia)* in Ostpreußen 1931. – *Verh.Orn.Ges.Bay.*, XX:191-225.

Schüz, E. (1936): Internationale Bestands-Aufnahme am Weißen Storch 1934. – *Ornithologische Monatsberichte*, 44:33-41.

Segurado, P. & Araujo, M. B. (2004): An evaluation of methods for modelling species distributions. – *Journal of Biogeography*, 31:1555-1568.

Sproll, A. (2000): Zugverhalten und Mortalität des Weißstorches (*Ciconia Ciconia*) nach Ringfunden der Vogelwarte Rositten (Ostpreussen). – *Diploma Thesis*, Fachhochschule für Technik, Wirtschaft und Sozialwesen Zittau/Görlitz.

Statistisches Handbuch für die Provinz Ostpreußen 1938 (1938): – Grenzlandverlag G. Boettcher, Schloßberg & Leipzig.

Struwe, B. & Thomsen, K.-M. (1991): Untersuchungen zur Nahrungsökologie des Weißstorches (*Ciconia ciconia*, L. 1758) in Bergenhusen 1989. – *Corax* 14:210-238.

Tischler, F. (1941): *Die Vögel Ostpreußens*. – Ost-Europa-Verlag, Königsberg & Berlin.

Tryjanowski, P., Sparks, T. H. & Jerzak, L. (2006): Introduction. White Stork *Ciconia ciconia* research in Poland: where we are and where we are going? – In: *Tryjanowski P., Sparks T. H. & Jerzak L. (eds). The White Stork in Poland: studies in biology, ecology and conservation.* Bogucki Wydawnictwo Naukowe, Poznan 2006.

Tryjanowski, P., Sparks, T. H., & Profus, P. (2005): Uphill shifts in the distribution of the white stork *Ciconia ciconia* in southern Poland: the importance of nest quality. – *Diversity and Distributions*, 11:219-223.

van den Bossche W., Berthold P., Kaatz M., Nowak E. & Querner U., 2002. *Eastern european White Stork populations: migration studies and elaboration of conservation measures*. – Bonn, BfN Scripten 66.

Zweig, M. H. & Campbell, G. (1993): Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine [published erratum appears in Clin Chem 1993 Aug;39(8):1589]. – *Clinical Chemistry*, 39:561-577.

8 Appendix

8.1 Habitat data: GIS-Layer

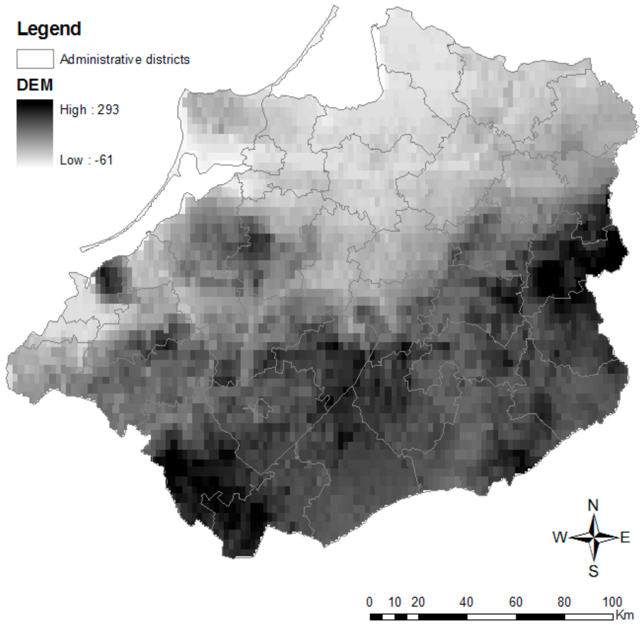

Figure 8.1: DEM-Layer

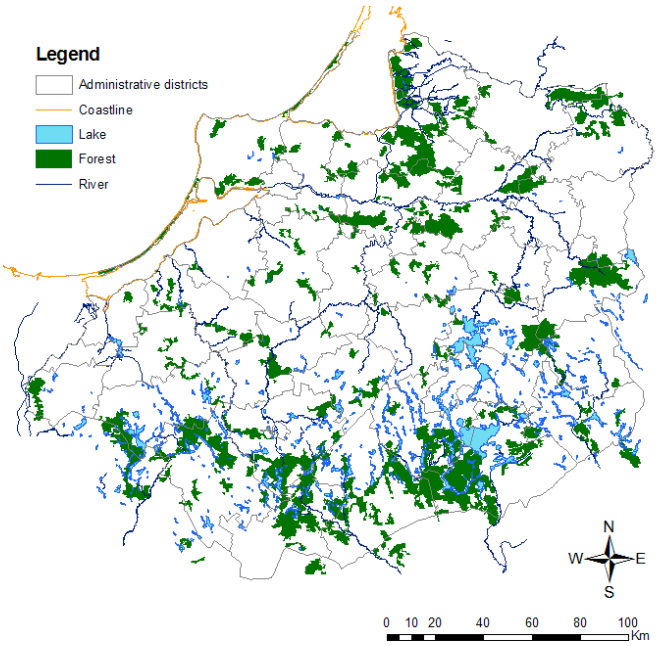

Figure 8.2: Lake-, Forest- and River-Layer as digitized from the historical topographic maps (‘Großblätter’) and Coastline-Layer

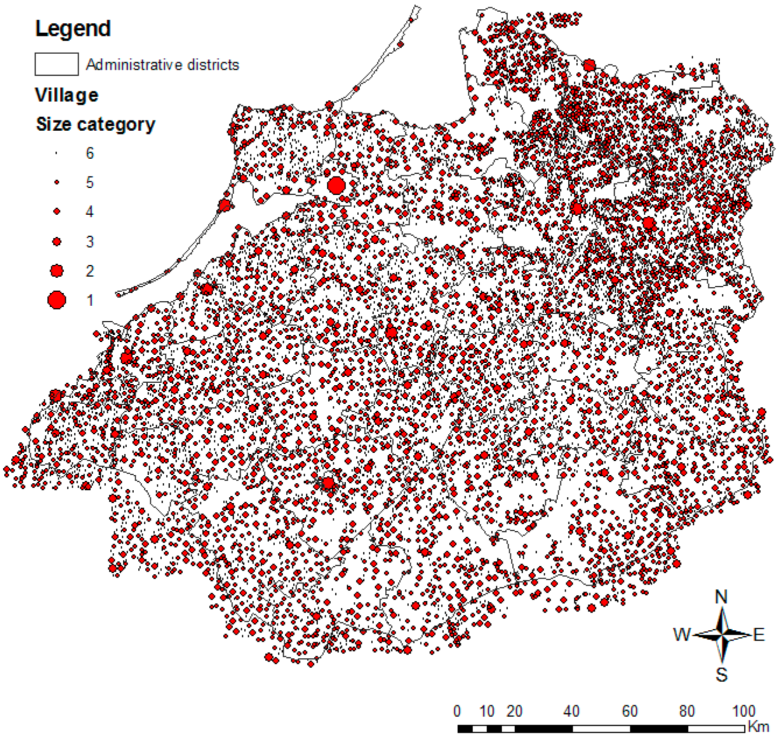

Figure 8.3: Village-Layer showing size categories '1' to '6' as digitized from the historical topographic maps ('Großblätter)

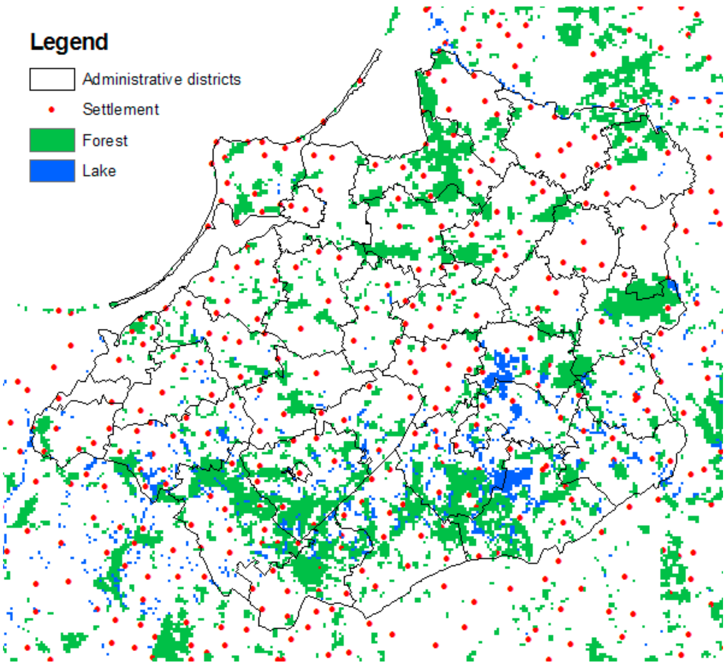

Figure 8.4: Lake-, Forest and Settlement-Layer as derived from the Baltic Sea GIS (acquisition period 1981-1992)

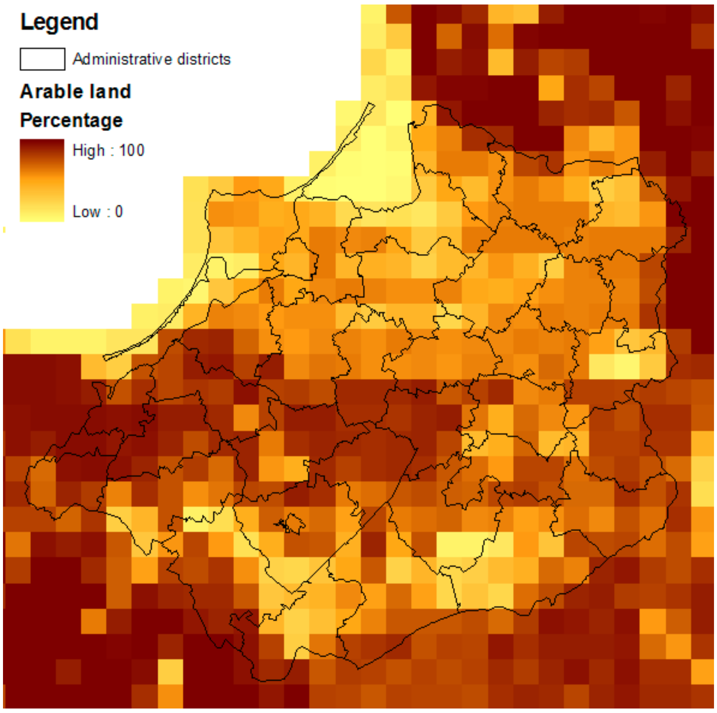

Figure 8.5: Arable land-Layer as derived from the Baltic Sea GIS (Acquisition period 1987-1993)

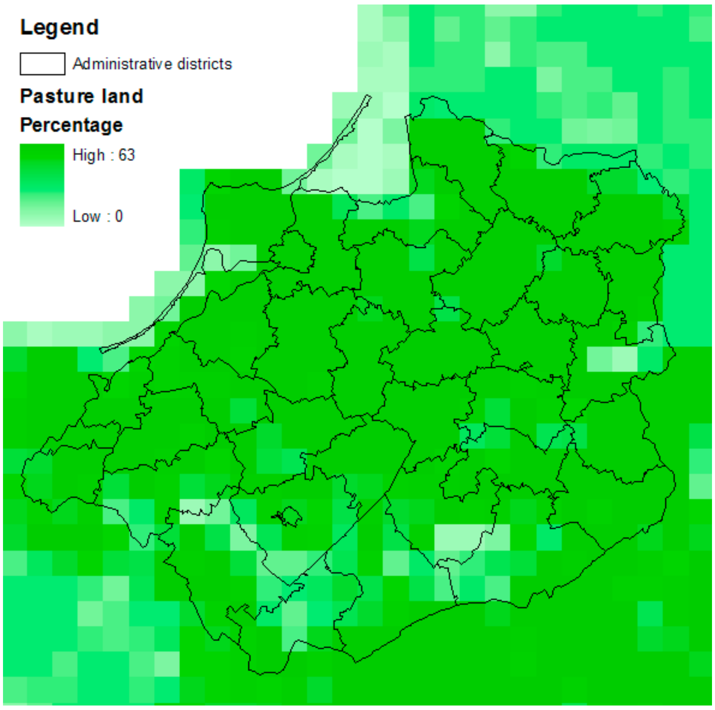

Figure 8.6: Pasture land-Layer as derived from the Baltic Sea GIS (acquisition period 1987-1993)

8.2 Habitat data and White Stork data used for Model 3

Table 8.1: Data used to build Model 3 at administrative district scale

Administrative district	*Stork [je 100 km²]*	*Area [km²]*	*Arable land [%]*	*Pasture and meadow [%]*	*Forest [%]*	*v1t4*	*v1t5*	*v1t6*	*DEM [m]*	*Mean Distance to coastline [m]*	*River [%]*	*Lake [%]*
Allenstein-Land	24,7	1304	46,7	12,5	27,0	42	139	278	149,9	86510	1,08	4,45
Allenstein-Stadt	15,4	52	18,9	7,4	42,0	1	5	8	138,5	82894	2,89	10,66
Angerburg	48,4	929	42,8	18,2	16,0	20	87	250	147,1	92308	0,84	12,71
Bartenstein	44,8	866	54,2	23,7	11,0	18	89	261	62,5	50921	1,16	0,12
Braunsberg	44,7	946	49,3	23,9	17,0	20	99	261	73,7	24482	1,51	0,26
Darkehmen	61,7	760	53,8	24,4	12,0	4	141	390	122,1	75630	1,27	0,11
Elbing	54,4	518	37,2	25,0	26,0	14	76	157	66,9	7550	0,92	1,74
Fischhausen	53,8	1048	45,6	22,6	21,0	32	104	320	25,2	5312	0,00	0,23
Gerdauen	51,4	844	52,7	23,4	15,0	24	97	298	65,5	61534	0,08	0,83
Goldap	54,3	993	43,3	18,0	26,0	6	161	338	192,6	102793	0,70	2,11
Gumbinnen	68,2	729	56,2	24,2	9,0	7	152	340	76,2	72595	1,81	0,14
Heiligenbeil	52,0	908	52,9	25,5	12,0	22	107	323	88,9	11773	0,67	0,13
Heilsberg	39,2	1096	52,7	23,7	13,0	33	116	227	119,7	59080	1,23	1,08
Insterburg	52,3	1205	69,1	51,9	35,0	15	190	550	44,8	46483	1,22	0,00
Johannisburg	27,8	1684	27,9	15,7	35,0	26	170	264	147,9	145687	0,40	9,85
Königsberg-Land	56,7	970	48,7	31,4	10,0	39	99	229	26,4	12892	0,53	0,27
Königsberg-Stadt	52,0	98	18,7	23,8	2,0	9	14	26	0,0	0	0,00	0,00
Labiau	50,2	1066	32,2	23,8	32,0	21	120	284	17,9	12357	1,53	0,00
Lötzen	46,1	897	50,2	17,4	10,0	23	95	247	146,6	111884	0,00	11,67
Lyck	49,7	1115	50,8	14,0	15,0	23	158	272	150,2	143623	0,41	6,65
Marienburg	58,1	226	44,0	38,5	1,0	4	44	98	24,8	21100	2,27	0,05
Mohrungen	40,0	1265	47,5	17,9	20,0	34	113	322	124,1	47507	0,33	6,65
Neidenburg	17,2	1146	44,4	10,8	32,0	20	108	162	172,5	117187	0,35	2,83
Niederung	72,3	1004	40,1	29,0	19,0	24	192	382	11,9	17386	3,38	0,06
Ortelsburg	26,9	1703	35,6	21,0	31,0	40	165	291	153,6	126884	0,43	2,67

Administrative district	*Stork [je 100 km^2]*	*Area [km^2]*	*Arable land [%]*	*Pasture and meadow [%]*	*Forest [%]*	*v1t4*	*v1t5*	*v1t6*	*DEM [m]*	*Mean Distance to coastline [m]*	*River [%]*	*Lake [%]*
Osterode	19,9	1537	50,4	10,8	25,0	37	178	305	171,5	83490	0,31	4,13
Pillkallen	64,5	1061	53,6	21,1	15,0	11	194	584	57,5	83497	1,05	0,09
Preussisch Ehlau	52,7	1242	46,9	23,6	15,0	33	126	325	91,0	31810	0,44	0,03
Preussisch Holland	57,4	858	51,8	22,4	15,0	26	113	242	94,1	27097	0,80	0,58
Rastenburg	46,2	873	57,1	21,9	11,0	24	101	303	113,3	82305	1,16	0,90
Rössel	33,8	851	52,1	20,9	15,0	26	90	276	156,9	82931	0,29	3,28
Sensburg	31,7	1232	41,3	12,7	26,0	34	121	289	156,6	115720	0,45	11,46
Stallupönen	75,9	704	60,9	21,9	7,0	12	147	322	119,2	94249	0,16	0,06
Stuhm	34,7	623	63,9	14,2	11,0	14	82	220	68,5	37466	0,60	0,85
Tilsit-Ragnit	64,4	1159	52,5	23,4	8,0	19	262	550	35,0	53232	1,30	0,00
Treuburg	52,0	856	55,5	17,1	12,0	17	103	244	167,3	125085	0,74	3,62
Wehlau	40,6	1067	39,6	23,3	29,0	14	115	267	27,1	30200	1,19	0,04

8.3 Model 2_mod: Variable Importance and Partial dependence plots

Table 8.2: Importance of the predictor variables of Model 2_mod

Predictor variable	*Relative Importance Score*
Distance to forest	100
Distance to coastline	95.3
Elevation	87.0
Distance to settlement area, size range 1 to 4	85.7
Distance to lake	84.1

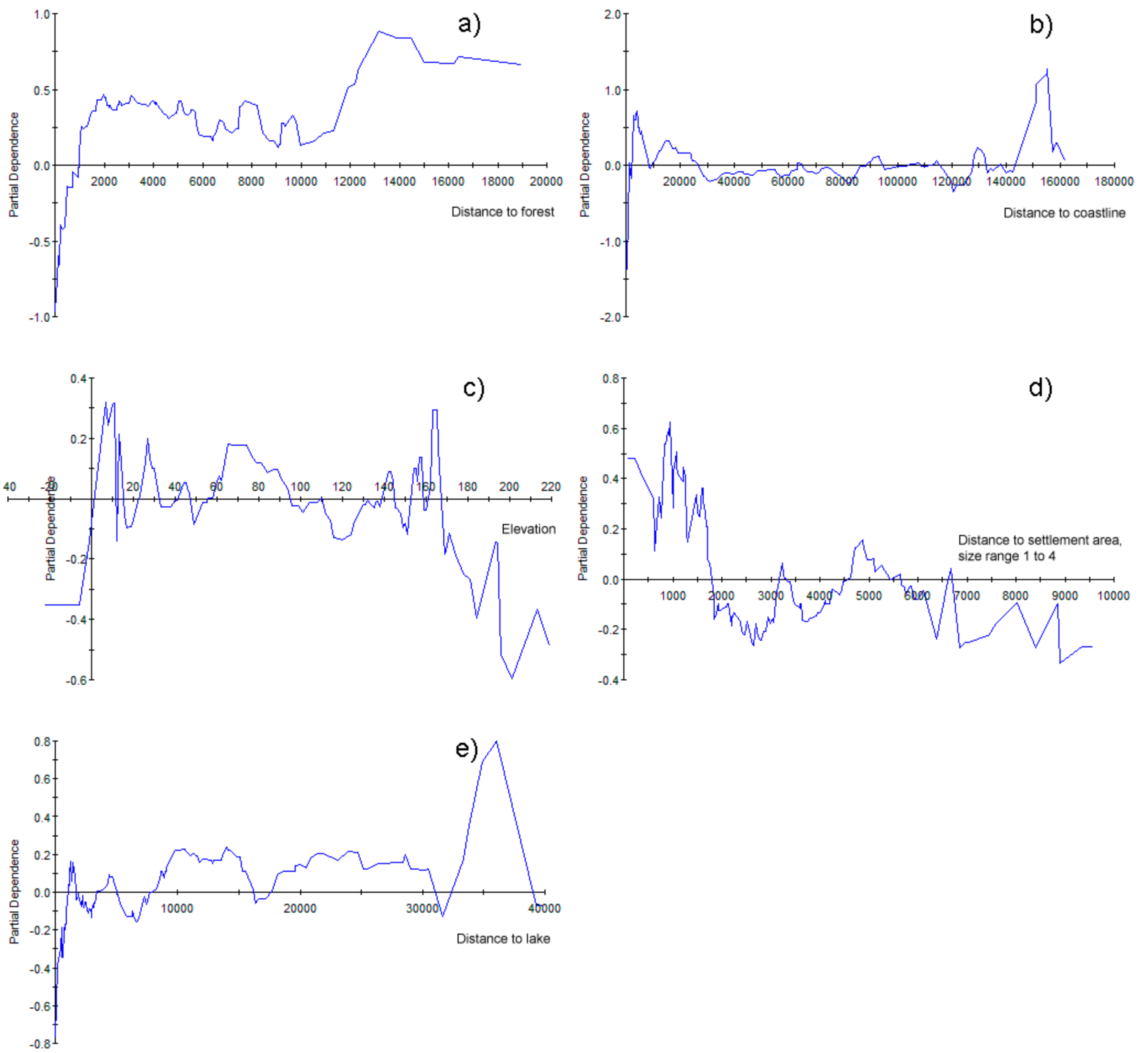

Figure 8.7: Partial dependence plots for predictor variables employed in Model 2_mod

8.4 Model 3: Observed and predicted number of White Stork breeding sites and MAE shown for every administrative district

Table 8.3: Observed and predicted densities of White Stork breeding pairs, Mean absolute error (MAE) and difference shown for every administrative district for Model 3

Administrative district	*Observed number of breeding sites per 100 km^2 (1934)*	*Predicted number of breeding sites*	*MAE*	*Difference in %*
Allenstein-Land	24.70	29.49	4.79	-19.39
Allenstein-Stadt	15.40	31.64	16.24	-105.47
Angerburg	48.40	44.07	4.33	8.94
Bartenstein	44.80	50.11	5.31	-11.86
Braunsberg	44.70	48.60	3.90	-8.73
Darkehmen	61.70	62.43	0.73	-1.19
Elbing-Land	54.40	50.60	3.80	6.98
Fischhausen	53.80	49.36	4.44	8.25
Gerdauen	51.40	49.98	1.42	2.76
Goldap	54.30	52.88	1.42	2.62
Gumbinnen	68.20	62.59	5.61	8.23
Heiligenbeil	52.00	52.00	0.00	0.01
Heilsberg	39.20	46.57	7.37	-18.79
Insterburg	52.30	53.07	0.77	-1.47
Johannisburg	27.80	31.58	3.78	-13.61
Koenigsberg-Land	56.70	51.51	5.19	9.15
Koenigsberg-Stadt	52.00	54.48	2.48	-4.77
Labiau	50.20	46.54	3.66	7.30
Loetzen	46.10	43.61	2.49	5.41
Lyck	49.70	42.80	6.90	13.88
Marienburg	58.10	60.91	2.81	-4.84
Mohrungen	40.00	43.20	3.20	-8.01
Neidenburg	17.20	29.16	11.96	-69.54
Niederung	72.30	59.39	12.91	17.86
Ortelsburg	26.90	37.68	10.78	-40.08
Osterode	19.90	32.27	12.37	-62.16
Pillkallen	64.50	60.21	4.29	6.65
Preussisch Ehlau	57.40	51.62	5.78	10.07
Preussisch Holland	52.70	52.09	0.61	1.16
Rastenburg	46.20	48.96	2.76	-5.98
Roessel	33.80	41.78	7.98	-23.60
Sensburg	31.70	30.57	1.13	3.58
Stallupoenen	75.90	63.18	12.72	16.76
Stuhm	34.70	40.39	5.69	-16.41
Tilsit	64.40	61.48	2.92	4.53
Treuburg	52.00	45.70	6.30	12.11
Wehlau	40.60	45.04	4.44	-10.93
Mean	46.92	47.50	5.22	-7.58

8.5 Model 3_mod: Observed and predicted number of White Stork breeding sites

Table 8.4: Observed and predicted densities of White Stork breeding pairs shown for every administrative district for Model 3_mod

Administrative district	*Number of breeding sites per 100 km²*			*Difference between 1981-1993 and 1936*
	Count 1934	*Prediction for the year 1934*	*Prediction for the period 1981-1993*	
Allenstein-Land	24.70	30.5	32.0	1.5
Allenstein-Stadt	15.40	31.7	37.3	5.6
Angerburg	48.40	46.8	49.4	2.5
Bartenstein	44.80	53.4	58.2	4.8
Braunsberg	44.70	49.1	53.2	4.1
Darkehmen	61.70	60.3	60.8	0.5
Elbing-Land	54.40	49.2	54.7	5.5
Fischhausen	53.80	50.6	53.3	2.7
Gerdauen	51.40	50.2	60.8	10.6
Goldap	54.30	49.6	44.2	-5.4
Gumbinnen	68.20	59.6	59.6	0.0
Heiligenbeil	52.00	52.7	59.3	6.7
Heilsberg	39.20	46.6	53.3	6.7
Insterburg	52.30	52.4	57.0	4.6
Johannisburg	27.80	33.2	32.0	-1.2
Koenigsberg-Land	56.70	53.7	59.0	5.3
Koenigsberg-Stadt	52.00	53.5	54.2	0.7
Labiau	50.20	49.0	47.7	-1.3
Loetzen	46.10	43.6	51.0	7.4
Lyck	49.70	38.9	51.8	12.9
Marienburg	58.10	59.5	56.5	-3.0
Mohrungen	40.00	41.9	41.9	0.0
Neidenburg	17.20	31.4	31.4	0.0
Niederung	72.30	55.9	54.3	-1.6
Ortelsburg	26.90	40.1	36.9	-3.3
Osterode	19.90	30.8	32.0	1.2
Pillkallen	64.50	59.8	57.2	-2.5
Preussisch Ehlau	57.40	53.1	51.6	-1.5
Preuss. Holland	52.70	49.1	59.3	10.2
Rastenburg	46.20	48.6	56.3	7.7
Roessel	33.80	42.8	51.0	8.3
Sensburg	31.70	32.0	32.0	0.1
Stallupoenen	75.90	60.9	62.5	1.6
Stuhm	34.70	39.6	52.1	12.5
Tilsit	64.40	59.4	58.1	-1.3
Treuburg	52.00	49.1	51.6	2.6
Wehlau	40.60	48.6	53.1	4.5
Mean	46.92	47.5	50.4	3.0

8.6 Model 3_mod: Variable Importance and Partial dependence plots

Table 8.5: Importance of the predictor variables of Model 3_mod

Predictor variable	*Relative Importance Score*
Percentage of pasture land	100
Percentage of water	75.2
Percentage of forest	64.0
Elevation	45.2
Number of settlement areas, size range 1 to 4	43.3
Distance to coastline	32.1
Percentage of arable land	26.0

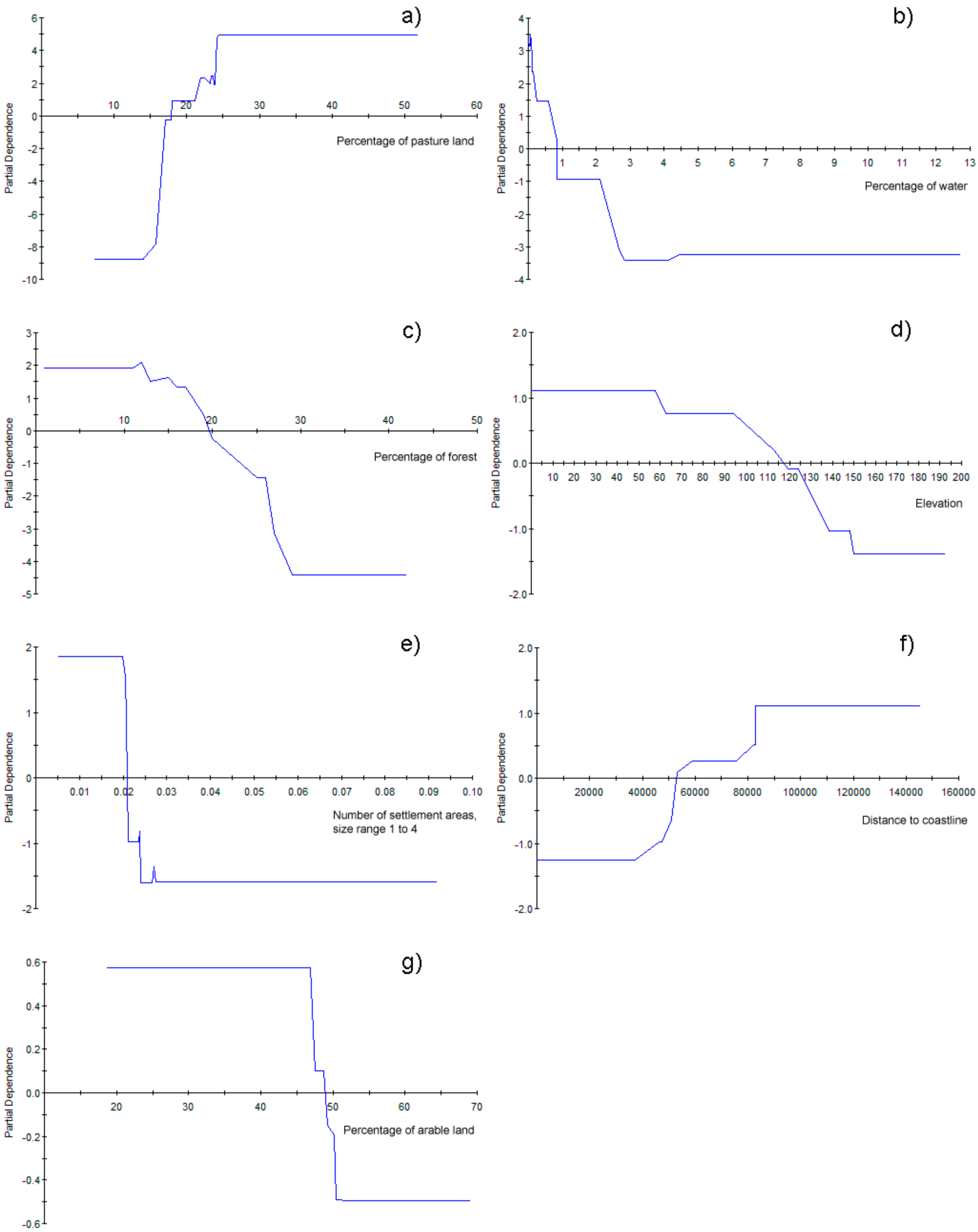

Figure 8.8: Partial dependence plots for the predictor variables employed in Model 3_mod: a) 'percentage of pasture land', b) 'percentage of water', c) 'percentage of forest', d) 'elevation', e) 'number of settlement areas, size range 1 to 4', f) 'distance to coastline', and g) 'percentage of arable land'

8.7 CD containing the used data

The CD contains the data used to create the different models. Furthermore the derived files for the modeling in TreeNet and the resulting models are given.

If you have only the PDF version of the diploma thesis at hand but are interested in the used data or if you have any question regarding the used data or the project carried out please contact:

Claudia Wickert, Ecoethology lab, Institute of Biochemistry and Biology, University of Potsdam (Email: cla_wi@hotmail.com),

PhD Falk Hüttmann, EWHALE lab - Biology and Wildlife Dept., Institute of Arctic Biology, University of Alaska-Fairbanks (Email: fffh@uaf.edu),

Professor Dieter Wallschläger, Ecoethology lab, Institute of Biochemistry and Biology, University of Potsdam (Email: wallsch@uni-potsdam.de).