



## **Analysis of kNN dataset according to its use in wind throw risk simulation.**



**Bartosz Standio**

**Supervisors:** Kristina Blennow, Mikael Andersson

---

M.Sc. Final Thesis no. 70  
Southern Swedish Forest Research Center  
Alnarp January 2006

---

# Index

- ABSTRACT..... 2**
- 1. INTRODUCTION..... 1**
  - 1.1. GENERAL INFORMATION..... 1
  - 1.2. RISK AS DECISION MAKING FACTOR - RISK MANAGEMENT ..... 2
  - 1.3. WIND DAMAGE MODELLING ..... 2
    - 1.3.1. *WINDA* ..... 3
    - 1.3.2. *The k Nearest Neighbour (kNN) and Segmentation Methods*..... 4
- 2. MATERIAL & METHODS ..... 5**
  - 2.1. THE AREA OF RESEARCH..... 5
  - 2.2. DATA DESCRIPTION ..... 5
  - 2.3. DATA PREPARATION & CONVERSION ..... 6
    - 2.3.1. *Asa Research Park Inventory Data* ..... 6
    - 2.3.2. *Asa Research Park the kNN Data* ..... 7
  - 2.4. DATA PROCESSING AND ANALYSIS..... 7
    - 2.4.1. *Defining Exposed Stand Edges*..... 7
    - 2.4.2. *Length of Exposed Edges* ..... 8
    - 2.4.3. *Mean Height Test*..... 9
- 3. RESULTS..... 10**
  - 3.1. STANDS STRUCTURE..... 10
  - 3.2. THE NUMBER OF EXPOSED POINTS ..... 10
  - 3.3. EXPOSED LINES ..... 11
  - 3.4. MEAN HEIGHT TEST ..... 13
- 4. THE DISCUSSION ..... 14**
- 5. CONCLUSIONS..... 17**
- ACKNOWLEDGMENTS ..... 17**
- LITERATURE ..... 18**
- APPENDIX..... 20**

## Abstract

Forestry is a very specific branch of industry, dependant on biotic and abiotic hazards of production where wind can be listed as one of the most important one. Each year damage caused by this factor to the production can be valued up to \$ 150 million. A way to achieve a reduction of loses and costs can be application of adequate risk management methods.

Risk assessment is a main element of risk management, unfortunately high cost makes it extremely difficult to apply in most of forest estates. A solution to this problem can be a simulation of wind damage probability provided i.e.: in WINDA simulator with the use of the  $k$ NN data. The aim of this paper is to test usefulness of the  $k$ NN Dataset and its precision in WINDA simulation. The results show significant similarities in comparison with traditional dataset. The highest discrepancies between dataset were visible in such elements as: number of points exposed to the wind where traditional dataset was 4,5 times bigger than the  $k$ NN; and the total length of exposed stands edges created on base of exposed points was already only twice times bigger. Despite visible differences tested number and length of exposed stands edges in relative comparison draw a similar trend. Statistical test on mean height of stands confirms that there are no significant differences between traditional and satellite acquired data.

Obtained in the research results showing similarities and usefulness of the  $k$ NN data promotes this method in risk assessment to future development and deeper studies.

*Keywords:* risk management, risk assessment, forest, damage, wind, wind throw, simulation,  $k$ NN, WINDA,

# 1. Introduction

## 1.1. General Information

Forest estate management is a demanding task, the proper performance of which requires a vast knowledge and orientation in fields of biology, ecology and, what is very important, in economy. Indeed, forest manager is expected to predict the growth of the forest in the first place, but also to skilfully manage it. In other words, the forester's knowledge has to be supported with the manager's ability to plan and predict future development of forest on both small estate and landscape level with respect to the potential shifts in the wood market. Therefore, the interaction between what is generally referred to as the environment and the forest must be instrumental in making strategic decisions. What can not be overlooked at this point is an element of risk, inherent in any long term investment. Forest management is as well subject to the rule. A study of risk is an important factor in the decision making process.

Due to its specific character, forestry and forest management, unlike any other industrial branch, differs from industrial safety management, and are affected by exogenous hazards that cannot be controlled. Abiotic risks such as fire, drought, wind damage, snow, ice, frost, flooding, and land-slippage (Gardiner and Quine 2000) constitute the major part of all risk factors. Windstorms are the main disturbance agent in European forests, with both ecological and economical impact upon the forest (Olofsson and Blennow 2005). In Sweden approximately 4 million cubic meters are damaged annually, what roughly corresponds to the worth of \$ 150 million (Valinger and Fridman 1999). The importance of this abiotic hazard can be confirmed by the particular examples of damage caused to forest. In 1990 about 100 million cubic meters of forest were blown down in one night by windstorm which swept over Europe (Peltola, Gardiner et al. 2000). A more up to date example is a storm which hit Southern Sweden at the beginning of January 2005. The new estimation, based on the aerial inventory, shows that 69.7 million cubic metres of forest has been damaged by this storm (Skogsstyrelsen - National Board of Forestry 2005).

A practical approach which can be employed in forestry planning models is to estimate age-dependent cumulative survival rates for a given set of hazard factors (von Gadow 2000). Tools based on a variety of models are developed to improve level of forest planning. Models used in forestry are applied for many purposes yet the risk analyse are surprisingly rare. Such a small number of such an application is triggered by problems with compilation data sufficient to develop a reliable model. Huge costs are inevitable in such a situation, for data compilation appears to be a very expensive process.

High costs and the need for extended data are a limiting factor in case of risk assessment applications similar to WINDA (WINDA described by Blennow and Sallnäs (2004)). The WINDA is a system of models for assessing the probability of wind damage. One of the ways to increase the employment of the application for a larger and bigger number of forest estate is to improve the process of data acquisition. For this purpose - satellite information could reduce costs and enhance accessibility of information for more and bigger areas. One of many methods of satellite data analysis is segmented  $k$  nearest neighbour ( $k$ NN) method by which complete area coverage of forest information may be modelled.

The interest in risk management in Swedish forest sector increased after storm in January 2005 and can boost the demand for risk assessment. The potential use of the  $k$ NN data in WINDA simulator as easily accessible data could provides means of extending the practicability of this application. The aim of this work was to prepare segmented  $k$ NN data to use in WINDA and to analyse its usefulness for the risk simulator in comparison with data gathered in a traditional way based on a field inventory.

## **1.2. Risk as Decision Making Factor - Risk Management**

Risk in forestry has been an issue of high interest. An example of that are countries of Central Europe. The extent of hazards and their influence on forest production as well as the economy has led to research efforts trying to facilitate risk management in practical forestry. In a short, risk management includes strategies and actions for reduce the risk (Hollenstein 1997). Risk management is a complex process that can be divided into four stages:

- the identification of risk agents/hazards
- the assessment of risk - the probability and predictable consequences
- the assessment of alternative responses
- the implementation of chosen course of action

Active risk management could result in substantial reduction of economic losses and in other types of benefits. In this work I have focused on the second step listed above, namely on the assessment of risk. In order to provide a valid assessment of hazard, the subject of the scrutiny must be well known and relatively easy to describe. However, a reliable representation of a natural phenomenon such as wind and windstorm is highly problematic. It is difficult to assess the probability of wind damage with purely statistical approaches because they do not define the casual links between tree parameters and susceptibility to wind damage, which can be described in a mechanistic approaches (Gardiner, Peltola et al. 2000). A method to do this is to use mechanistic models which are able to simulate reaction of single tree or stand for strong wind. Received in this way a vulnerability have to be extended with estimation of the probability of exceeding the wind speed threshold for damage.

## **1.3. Wind Damage Modelling**

Wind damage models are risk management tools for forestry. Models are designed to assess the risk and evaluate forest management with respect to the risk. A fundamental element of this process is a realistic representation of wind. Two independently developed models: GALES and WINDA can serve as illustrative examples.

The GALES model was developed to deal with the wind damage in the interior of unthinned or lightly thinned British commercial stands (Gardiner and Quine 2000). To calculate the wind forces acting on the tree certain steps have to be followed: to present forces acting on a tree this model use relationship between the drag of the air on a surface and the aerodynamic roughness of the surface; resistance to breakage is based on a assumption that the wind induces stress in the outer fibres of the tree stem is constant at all points between the base of the canopy and the butt swell at the stem base (Morgan and Cannell 1994); resistance to overturning is based on tree pulling experiments, provided on a variety of conifer tree species and on a range of typical soil types (Gardiner, Peltola et al. 2000); calculation of probabilities of exceeding wind speeds for damage uses a classification of the land according to wind risk classes. The GALES model does not include pure wind model

The WINDA model used in this work was developed at Southern Swedish Forest Research Centre at Swedish University of Agriculture Sciences. WINDA is a system of models for assessing the probability of wind damage (Blennow and Sallnäs 2004) and provides as a result wind damage risk probability for every stand. The program calculates the wind loading and the stability of trees at stand edges. These calculation include geographical computations using a GIS. The outcome of the model, expressed in terms of annual probability of damaged

by wind forest stand edges, is determined for six different wind directions and is calculated from assessment of the annual probability of wind damage at points along the exposed edges of the forest stand (Olofsson and Blennow 2005). A big influence on the structure of the simulator has an assumption that wind damage is initiated at forest stand edge.

### 1.3.1. WINDA

During simulation WINDA uses a number of sub-models which can be grouped due to function in a component groups:

- The first component identifies exposed stand edges at least 10 m high. These calculations are done by the sub-models : Roughness, Prefetch, Postfetch. There are points determined every 50 m along selected stand edges. In case of small and/or irregularly shaped stands where two consecutive nodes in the digitised polygon representing the forest stand are closer than 50 m from each other, a point defined midway between two nodes is used. The wind is divided into six direction sectors to which every point (and further stand edge) is assigned assuming exposition to wind from directions within  $\pm 30^{\circ}$  from the direction perpendicular to the edge. Wind direction sectors are divided as follow: I -  $0^{\circ}$ ; II -  $60^{\circ}$ ; III -  $120^{\circ}$ ; IV -  $180^{\circ}$ ; V -  $240^{\circ}$ ; VI -  $300^{\circ}$ .
- Dose/response component model HWIND is designed to calculate the critical wind-speed at exposed stand edges which was described by (Peltola, Kellomaki et al. 1999). The HWIND model was developed to quantify the vulnerability of Finnish forests to wind at stand edges following the creation of new edges and after thinning (Gardiner, Peltola et al. 2000). It simulates forces acting upon a tree and divides them into horizontal force due to wind and the vertical force due to gravity. The model is based on the assumption that a tree deflects to a point of no return when exposed to the wind of constant mean velocity and direction. Calculated mean wind load on a tree uses predicted wind profile at the stand edge and vertical distribution of stem and crown weights combined with gravity-base forces at each height in the canopy.
- Free-stream wind used in WINDA calculations is provided by model component WASP - The Wind Atlas and Application Program (Mortensen, Landberg et al. 1998). This sub-model is used to calculate the free-stream wind which is cleaned from effects of obstacles, roughness changes (i.e.: roughness of stand canopy) and orography.
- By combination of all elements listed above and further calculation results in probability of wind damage is calculated.

All geographical computations are carried out with the employment of the ArcInfo application using polygon coverage (Anon. 2001) file format. Forest description data for every stand within the area under investigation is contained in a data table of ArcInfo coverage. Forest inventory data includes the information on tree species, species-wise average tree height, the diameter at breast height, and the number of stems per hectare (Blennow and Sallnäs 2004). Other input data indispensable for WINDA simulation processes are also: a digital elevation model (DEM) for study landscape and its surroundings; forest map of surrounding and an obstacle map for a circle centred on the meteorological observing station.

### 1.3.2. The $k$ Nearest Neighbour ( $k$ NN) and Segmentation Methods

The  $k$  nearest neighbour estimation method is under an intensive research among forest inventory groups (Tomppo, Czaplewski et al. 2002), The  $k$  nearest neighbour ( $k$ NN) estimation method has been used in the Finnish multisource national forest inventory (NFI) since 1990 (Tomppo 1990), and integrates field and satellite data (Holmström, Nilsson et al. 2001). Areas only known due to their spectral signatures in the satellite image are assigned field data values as weighted mean values of the  $k$  nearest field plots (Reese, Nilsson et al. 2002); nearness is taken in a feature space defined by the different spectral wave-length bands of the satellite image (Holmström 2001). The  $k$ NN methods simultaneously can provide estimates for all parameters available at the reference plots. Another advantage of the method is the simplicity with which new sources of information can be used to strengthen the association between reference plots and areas to assigned forest data (Holmström, Nilsson et al. 2001).

The  $k$ NN data had been segmented into compartment method. The segmentation method used in this project is so called t-ratio segmentation method, which refers to a type of region growing algorithm originally developed by the SLU Remote Sensing Lab (Hagner 1990). The basic idea underlying the method is that spatially adjacent regions should be merged into larger regions if they can not be separated with a given certainty.

A criterion for merging regions should be described by the probability for two adjacent regions to represent for example the same tree height. It should result in the conclusion that the spectral intensities of two adjacent regions are in fact observations of the same height group. Hence, the significance of an absolute distance in feature space between regions is tested by relating it to the population variance and the number of observations – pixels (Mats Nilsson - personal communication). This merging process provides to a certain generalization of image which renders it more readable.

## 2. Material & Methods

### 2.1. The Area of Research

The subject of investigation presented in this paper is Asa Research Forest (57° 10'N/14° 47'E) situated in southern Sweden in Lammhult Municipality (Fig. 1). The forest estate covers 728 ha (Appendix 1 - map of estate) and is run by the Swedish University of Agricultural Sciences. On this time Asa Research Park consists 382 forest stands (Table 1) with significant domination of pure or mixed stands with Norway spruce (*Picea abies* L.), Scots pine (*Pinus sylvestris* L.) and birch (*Betula* spp.). The surrounding of the study land is covered by forest, except in the east, where Lake Asasjön is located.

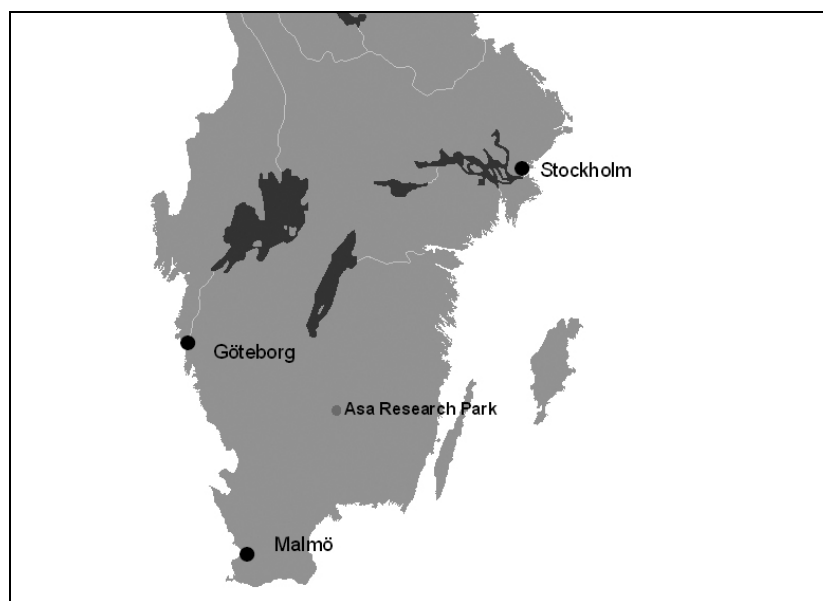


Figure 1. A map of Southern Sweden with localisation of Asa Research Park.

### 2.2. Data Description

Two different datasets were employed to compare two different types of datasets. Forest inventory data for Asa Research Park was used as a basic dataset.. This dataset is drawn on stand information obtained in 1997 for the purpose of forest management plan. As the second dataset (the *k*NN data) dates from 2000, all the changes which have taken place between 1997 and year 2000 were updated. The data was available only in shape file format (ArcView format with \*.shp extension) and as it is required of a traditional inventory data it contains a full description of stands. The height of forest stands, identifies as the main tested feature corresponds to basal-area-weighted mean tree height which makes it comparable to the *k*NN Data.

The second dataset (in comparison) was the segmented *k*NN data for southern Sweden (Reese, Nilsson et al. 2002) with the local coordinates: left top – 1150000, 6901000; right bottom – 1931000, 6901000. The precision of height information in the data is one meter and grid resolution of the map is 25x25 meters.



## 2.3. Data Preparation & Conversion

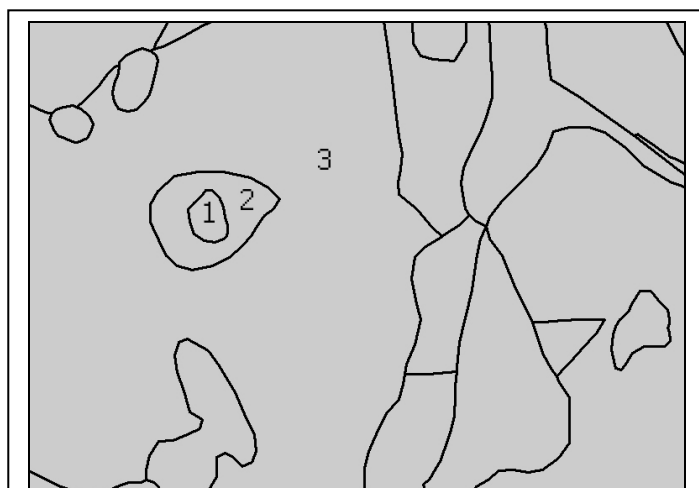
To be able to use the data as input to WINDA both datasets had to be converted and organized in a certain way into ArcInfo coverage format (Anon. 2001). The information table accepted by WINDA included in coverage format must contain specific columns arranged in a proper way. All the parameters of input files are described in manual available at Swedish University of Agriculture Sciences in Alnarp. The transformation of files was completed in ArcInfo 9 program (Anon. 2002). Moreover, the visualisation of simulated results and the preparation for tests such as “mean height test” were executed in ArcMap 9 program (Anon. 2002).

To achieve the final result WINDA requires information about each stand stored in the input file. WINDA works with information on stand height, diameter and number of stems per hectare. In this study only the *k*NN dataset containing height information was used. Due to limited information in used the *k*NN dataset, in both datasets (the *k*NN dataset and The Inventory dataset) fixed values for rest variables were set for all records. A data manipulation by setting untrue values in input files does not have any influence on the result of simulation at this stage of simulation process. The number and the distribution of the defined exposed points is not affected by these manipulations, yet absence of data in input files rendered the successive stages of simulation impossible.

### 2.3.1. Asa Research Park Inventory Data

Some difficulties were encountered during the map transformation process. The first one was related to “inside polygons”, polygons located in a form of an island in another one (a bigger polygon) - (Fig. 2). Two different polygons are in this case recognized in the data table as one. This kind of data arrangement is not accepted by ArcInfo coverage format (the only format accepted by WINDA), which necessitated the division of them into single polygons. This step has contributed a marked increase in stand number from 342 before reorganisation to 382 in the final coverage file.

Surprisingly with no data manipulation and with an altered stand number also the area of estate has increased from 656,8 ha to 728,3 ha. The underlying reason for this phenomenon is the employment of different methods of area calculation between shape file format and coverage format. The area in shape file must be calculated manually after every single data change. The area in coverage format is a variable calculated automatically.



**Figure 2. Fragment of stands map with visible “inside polygon”. Stands 1 and 3 were represented as one in the shape file data table.**

### 2.3.2. Asa Research Park the *k*NN Data

In order to prepare the *k*NN dataset a series of transformations had to be carried out. A rectangle containing map of Asa Research Park was cut out from main source file. It was accomplished with the use of the SELECTBOX command in the ArcInfo program. The selected area was transformed to coverage with the use of GRIDTOPOLY command and with no specified (default) weed parameter (default value was 5). The next step was to change the angular shape of polygons (Figure 3) to obtain the shape relatively analogous to real stands in Asa Research Park. To achieve this, a SPLIN command was employed with changed environment parameters: “vertex distance” (base value before transformation was 1.579), “arc span” (before 7,986) and “node span” (before 7,986). To provide the desirable shapes of the polygons’ edges the following values were used: in the first splin operation “vertex” distance was set on 60 and “arc” on 30, in the second splin operation “vertex” and “arc” were set on 30. Those values used during shape transformation were set after repeated attempts of combining different values. The relatively smooth curves but without losing angular elements were regarded as the satisfying shape (Appendix 2). The file contains polygons which later in this paper are also referred to as stands, but not in the literal sense of the word (*k*NN and segmentation description). Required data table was arranged with names, order and the number of columns requested by WINDA.

## 2.4. Data Processing and Analysis

### 2.4.1. Defining Exposed Stand Edges

The simulation’s results analysed in this research were the outcome of not all modules implemented in WINDA simulator. The range of analysis does not cover the whole simulation process but only first group of component modules. The following modules were employed in the project: Roughness, RChange, Displace, Prefetch, WASP and Postfetch. They are designed to define the exposed stand edges and generate the exposed points. Postfetch module is used to assign the exposed points into six wind-exposition sectors. The output of the simulation on this level is a text file with the coordinates of exposed points, attributed to the exposition sectors. The text file was imported to ArcInfo coverage file as a result of which further study was followed (Fig. 3).

To define exposed edges and exposed for stands located at the edge of estate WINDA simulator is considering also forest cover for neighbouring with them stands which are not part of the estate. After simulation made on the Inventory Dataset there was no exposed stand edges defined at the edge of estate. This phenomenon was observed even in such places



Figure 3. View on particular stand represented by raw *k*NN Data (background), stand borders after data preparation (white line). Black dots shows Edge Points simulated in “Postfetch” module presented here as another layer.

where forest was adjacent to meadows or lake, and difference in height was more than 10 meters. After later investigation a conclusion can be posed that there was an error in the input data describing surrounding of the estate. The source of the error are different classification of stands and their belonging to the estate. This phenomenon was not a problem in case of the *k*NN Dataset where simulation was provided on a bigger area than the estate and afterwards estate property was separated. To make both datasets comparable exposed points on the edge of the estate from the *k*NN Data were removed.

The first step of data comparison is exposed points where a given calculation checks the number of exposed points and its distribution between six different sectors.

#### **2.4.2. Length of Exposed Edges**

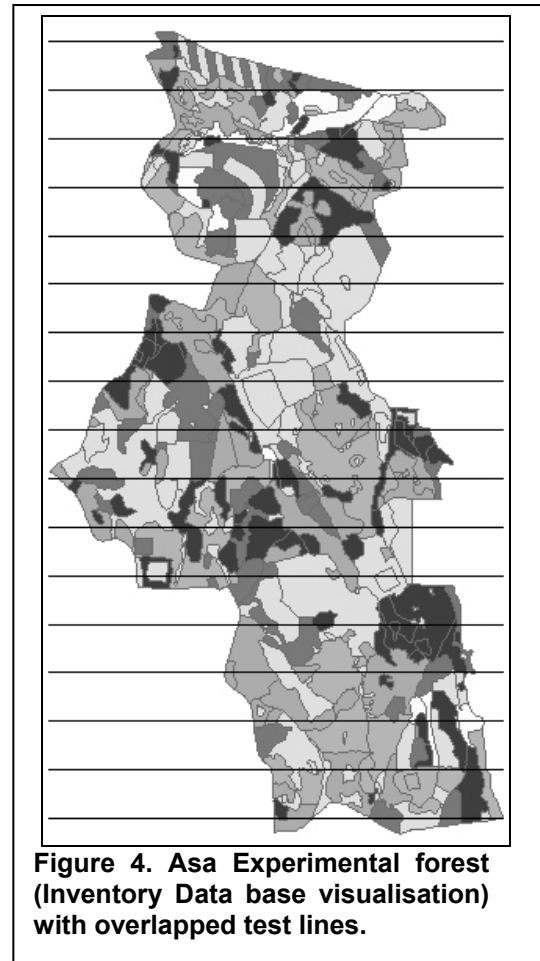
Another feature of analysed data was the length of the exposed stand edges. While exposed points do not create lines and can not be used to create stand lines, which would result in stand edge line incompatible with the original ones. One way to achieve this feature was to select fragments of exposed stand edges overlapping with exposed points.

The selection with the use of overlapping points can select only the whole lines of polygons without the possibility to select only a part of line (for example a part of the edge which is exposed). For that reason it was necessary to perform additional operation. Coverage file representing map of estate was edited in ArcEdit tool and every complex line (arc) which included pseudo-nods was broken into single segments. The created segments inherited coordinates of consecutive nodes and pseudo-nods. Next, segments overlapping with exposed points were selected. During this operation it was assumed that the total number of selected segments corresponds with the number of defined exposed points. This operation was performed for each sector separately. A certain toleration ratio had to be used during selection process, because of different points placement in relation to the lines. The toleration ratio was changed to get the number of selected segments corresponding to the number of exposed points in every sector. Selected segments were assumed as the exposed stand edges and merged into (longer elements) – complex lines, where RENODE command was used to facilitate the reading of the map. From this moment selected lines represented exposed edges defined by WINDA.

### 2.4.3. Mean Height Test

It is extremely interesting thing how the two forest maps differ. It can be proved by making a comparison such properties as for example number of stands, size structure. However checking if there was any similarity in stand height appeared to be an adequate idea. The examined datasets contain the information about stands height where, in both cases, the measuring method is analogous to a method based on basal-area-weighted mean tree height. The correspondence of the height information renders such comparison reasonable.

The area of estate was crossed with horizontal orientated (west – east) lines – transects (Fig. 4). 17 lines were drawn every 300 meters over the area. All the lines were gathered as one layer and used for both datasets. A set of stands which were crossed by these transects was selected. For the selected stands along each line a mean height was counted. All the preparatory operations were done in ArcMap 9 program. The following step was to statistically calculate where a T-Welsh test was appropriate for the gathered data. Statistical analysis was made in a Spanish program “R” (R Development Core Team 2004).



**Figure 4. Asa Experimental forest (Inventory Data base visualisation) with overlapped test lines.**

### 3. Results

#### 3.1. Stands Structure

The adoption of different methods for creating datasets resulted in their different composition. Inventory Data serving as the dataset, which represents the real forest structure and state, was characterised by over three times as big number of polygon - stands as the *k*NN Data (Table 1). The total discrepancy of estate area amounts to about 41 ha and the larger area was represented in the *k*NN Data. The average size of a stand (Table 1) in this comparison for the *k*NN Data is 6,9 ha whereas for the Inventory Data 1,9 ha. A higher precision of the Inventory Data is represented by the smallest stand – 0,03 ha while in the *k*NN Data smallest stand was almost 0,3 ha.

**Table 1. Stands description for both datasets, the area provided in hectares.**

	<b>kNN Data</b>	<b>Inventory Data</b>
Number of stands	112	382
Total area	769	728
Average area	6,9	1,9
Min area	0,29	0,03
Max area	39,9	21,6
StdDev [ha]	7,7	2,8

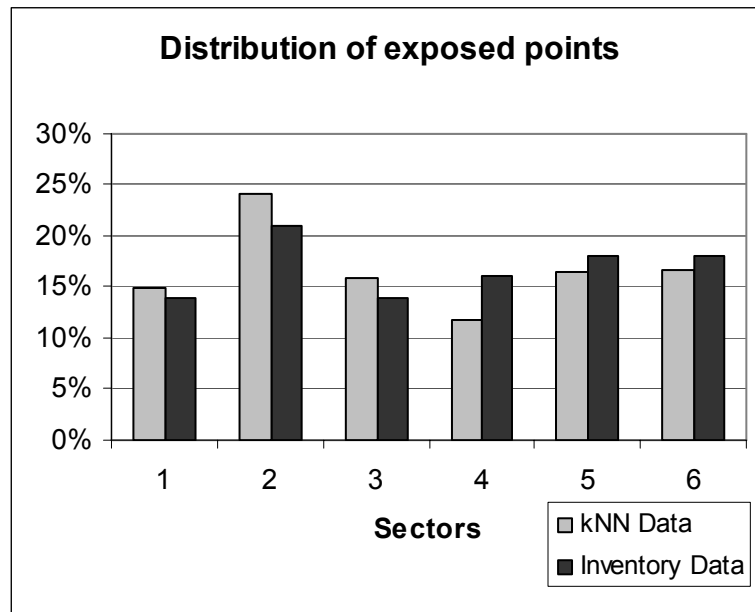
#### 3.2. The Number of Exposed Points

Exposed points represent places on the stand edge where the difference in tree height between two neighbouring stands exceeds 10 m. The number of simulated points in the two compared datasets varies considerably. It can be attributed to a higher precision of Inventory Data. The total number of points for the *k*NN dataset was 658 whereas for Inventory dataset - 4068 (Tab. 2).

**Table 2. The number of simulated points, their percentage share in the total number and the distribution into six sectors.**

SECTOR	<b>kNN Data</b>		<b>Inventory Data</b>	
	amount	% in total	amount	% in total
1	98	15%	554	14%
2	159	24%	842	21%
3	104	16%	565	14%
4	78	12%	642	16%
5	109	17%	752	18%
6	110	17%	713	18%
Total	658	100%	4068	100%

Table 2 represents the distribution of points between six sectors as well as their percentage share. The proportion between sectors and datasets shows similar values. The general pattern is that the *k*NN Data include relatively more points and has a proportionally greater percentage share in sector 1, 2 and 3 while the Inventory Data presents higher values in sectors 4, 5 and 6 (Fig. 5).



**Fig. 5. The relative distribution of edge points defined in Postfetch module, and divided into 6 sectors.**

### 3.3. Exposed Lines

It must be pointed out that as a result of the conducted extraction of exposed edges some edge fragments were assigned to two or more sectors. Such a situation occurred owing to the small distance between exposed point and surrounding lines, as well as flexible tolerance ratio in the SELECT command, different for all sectors. It allowed to assign two different lines with the use of one exposed point. A high number of exposed points and a small length of stand's edge fragments in the Inventory Data account for such occurrence in this dataset. However, the phenomenon of "double assignment" has a very modest share in the total number of exposed points and appears to have no influence on the final result.

The selection process resulted in 741 exposed lines (658 exposed points) in the case of *kNN* Data and 3963 lines (4068 exposed points) in the case of the Inventory Data, which is shown in (Table 3, 4).

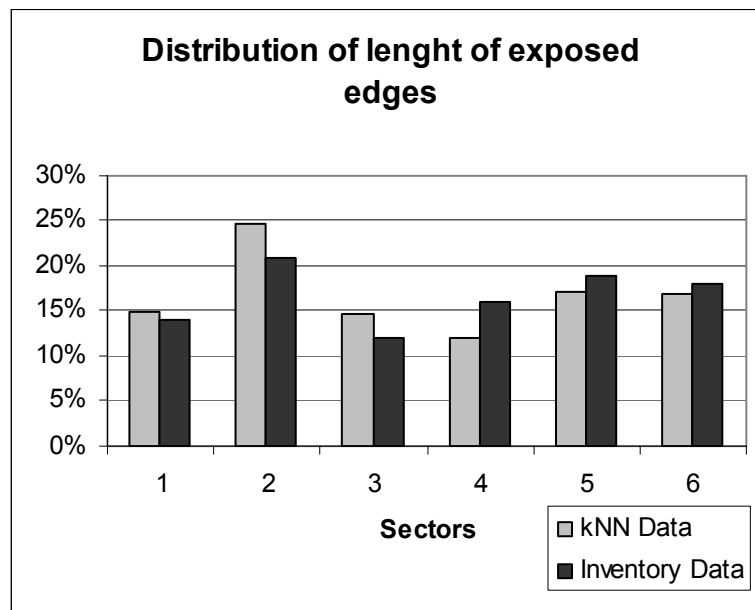
**Table 3. Extracted stand edges – result of line selection from *kNN* Data with the number of lines and their length in meters.**

<b>kNN Data</b>										
Sector	Count		Suma		Average	Max	Min	StDev [m]	Variance	
	amount	% in total	amount	% in total						
1	98	13%	2980	15%	32,7	79,1	13,3	8,9	79,3	
2	201	27%	4955	25%	32,4	70,7	20,1	7,3	53,4	
3	108	15%	2914	15%	32,6	70,7	18,0	7,5	56,8	
4	97	13%	2392	12%	32,2	79,1	23,6	7,3	54,0	
5	120	16%	3412	17%	33,3	79,1	14,5	9,8	96,7	
6	117	16%	3362	17%	32,5	79,1	21,9	8,0	63,5	
<b>Total</b>	<b>741</b>	<b>100%</b>	<b>20018</b>	<b>100%</b>	<b>32,6</b>	<b>79,1</b>	<b>13,3</b>	<b>8,1</b>	<b>65,4</b>	

**Table 4. Extracted stand edges – result of line selection from Inventory Data with the number of lines and their length in meters.**

<b>Inventory Data</b>										
Sector	Count		Suma		Average	Max	Min	StDev [m]	Variance	
	amount	% in total	amount	% in total						
1	550	14%	5630	13%	10,2	100,3	1,4	7,0	49,5	
2	842	21%	9626	22%	11,4	101,5	1,4	7,1	50,8	
3	485	12%	5442	12%	11,2	121,3	1,4	8,2	67,7	
4	635	16%	6783	15%	10,7	110,1	2,0	9,1	83,1	
5	745	19%	8764	20%	11,8	225,3	1,4	13,2	174,4	
6	706	18%	7610	17%	10,8	99,6	1,4	7,1	50,7	
<b>Total</b>	<b>3963</b>	<b>100%</b>	<b>43855</b>	<b>100%</b>	<b>11,1</b>	<b>225,3</b>	<b>1,4</b>	<b>9,0</b>	<b>81,2</b>	

The Proportion between the number of lines in both datasets is very similar to the number of exposed points. The percentage share is comparable in both cases too. The sum of lines reflects the total length of exposed forest edges and the length for every single sector separately. In this case the disproportion between datasets is smaller (Table 3, 4), for the *k*NN Data the total length was 20 018 m and for the Inventory Data it was 43 855 m. A general pattern emerging in this comparison is corresponding to the distribution of exposed points, yet with much smaller disproportion in the total length of exposed stand edges (Fig. 6).

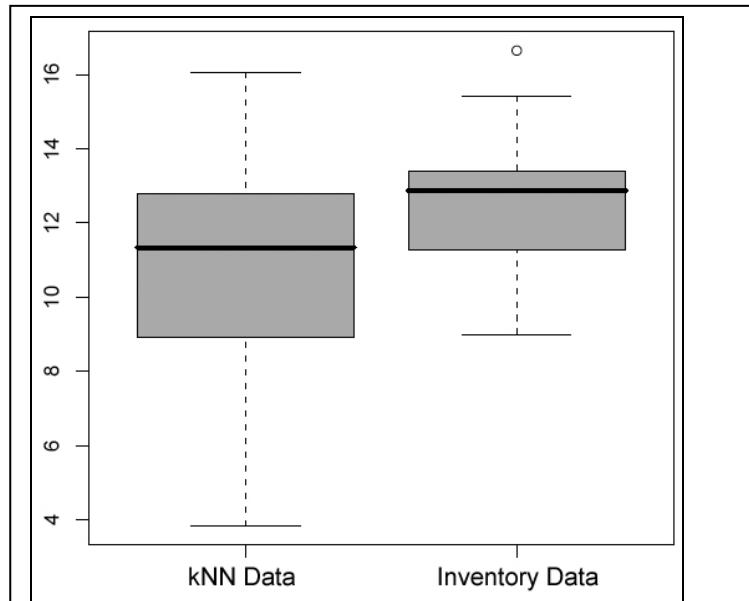


**Fig. 6. Relative distribution of length of exposed edges divided into 6 sectors.**

Figure 6 shows that the values in sectors 1, 2 and 3 are higher than in the remaining sectors which corresponds to the distribution pattern of exposed points. This situation can be caused by influence of the lake and its surrounding. Even when exposed points located on the edges of estate were removed before comparison some influence of the Asa Lake could remain.

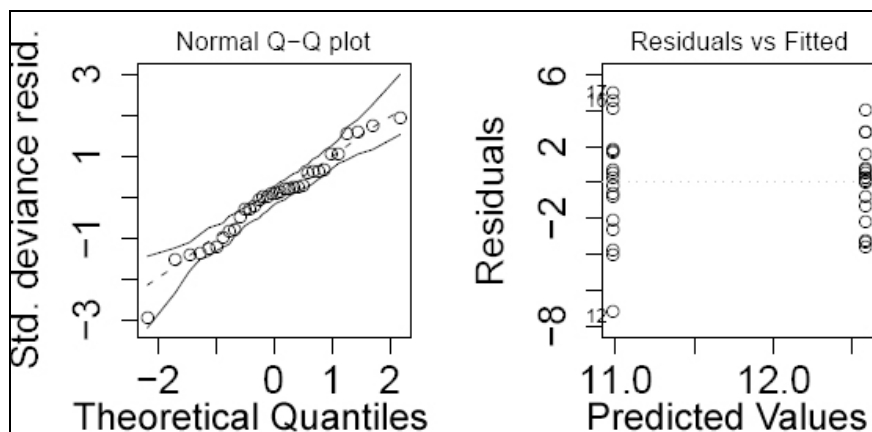
### 3.4. Mean Height Test

Height is a feature which provides a basis for all the tests performed in this research. To compare both the datasets a test of mean heights was conducted. Its aim was to examine if the maps are showing a similar mean heights. The stands height along 17 transects were analysed. The data gathered from all the transects and from both the datasets was statistically tested. Boxplot graph (Fig. 7) shows a nearly normal distribution but the inhomogeneity in variances (Fig. 7) necessitates a T-Welch test to be provided. The residual analysis gave good results. The points in Normal Q-Q plot are located inside the envelope and oscillate around 0-line (Fig. 8). The predicted values of residual plot are evenly distributed (Fig. 8).



**Figure 7. Boxplot graph shows that data are almost in normal distribution. Group A -represents kNN Data and B - Inventory Data.**

With the confidence level of 5% the hypothesis that maps are equal in means cannot be rejected. Consequently, the data can be assumed to be equal. The mean height in the all transects for the *k*NN Data was 10,98 m and for the Inventory Data - 12,59 m.



**Figure 8. Residuals analyses, values in Normal Q-Q plot are located inside an envelope, scattered and oscillate around zero value. Predicted values are distributed equally.**

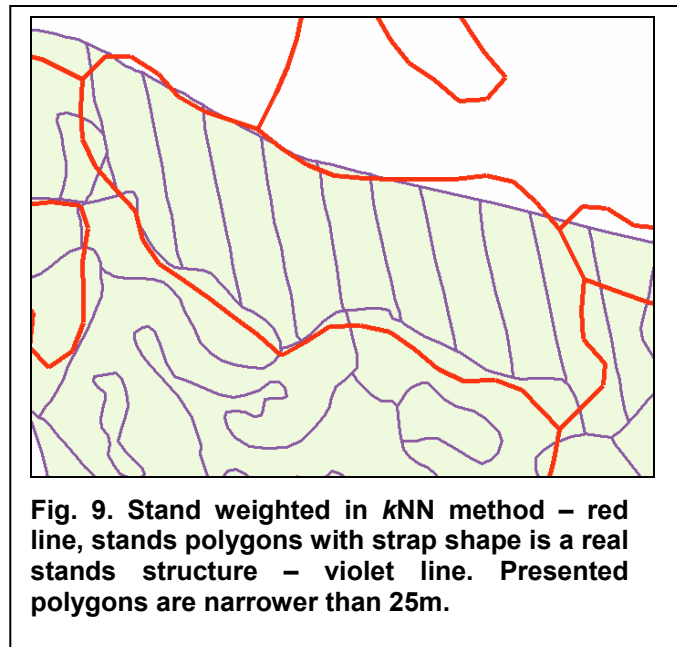


## 4. The Discussion

The aim of my paper was to devise a method to prepare the segmented  $k$ NN Data for further use in WINDA and to investigate if it can yield comparable results with reference to the traditional dataset. After a few transformations on datasets (mostly on the  $k$ NN Dataset because of its specific characterisation) allowed to carry out spatial part of WINDA simulation process. The correspondence between the datasets appeared to be good in the stages which spatial factors were used for computations. The relative distribution of stands' edges exposed to the damaging effect of the wind and mean heights reveal close similarities in results.

The classical method applied in the Inventory Data, is regarded as the most reliable one. However, it has certain undeniable drawback, it is costly and time – consuming. To evaluate the usefulness of the  $k$ NN Data in WINDA risk simulator the following question must be answered: is there any difference in the representation of forest stands? If so, what kind of differences are they and how do they influence the result of simulation? The conclusion is drawn upon the following steps:

- The general characterisation of stand structure;
- The comparison of exposed points – their number and the distribution between wind direction sectors;
- The comparison of exposed stand edges - their total length and the distribution between wind direction sectors
- Mean height test – the comparison of mean stands' heights for both datasets



The general description of forest estate is presented in Table 1 and shows certain discrepancies. The  $k$ NN Data has a significantly smaller number of stands and a larger area of estate where the average size of stand is bigger.

Such a difference is attributed to the precision of satellite images and the  $k$ NN method. The resolution of the sensor influences directly the size of the pixel in the recorded image. The data acquired by Landsat TM satellite with the size of the pixel: 25x25 m is regarded as a high resolution picture (Tomppo, Czaplewski et al. 2002). Satellite images considered as high resolution appeared to be very simplified and generalised in comparison to the Inventory Data. The level of precision is the reason why small stands or those with a very narrow shape can be skipped or weighted as one during the calculation (Fig. 9). It seems to be one of the most plausible explanation, as the number of polygons is two times lower in the  $k$ NN Data whereas the average size of stand is almost three times bigger. Of course average size of stand in the  $k$ NN Data depends also on the segmentation process where in reality the two or more neighbouring stands of similar height can create one polygon.

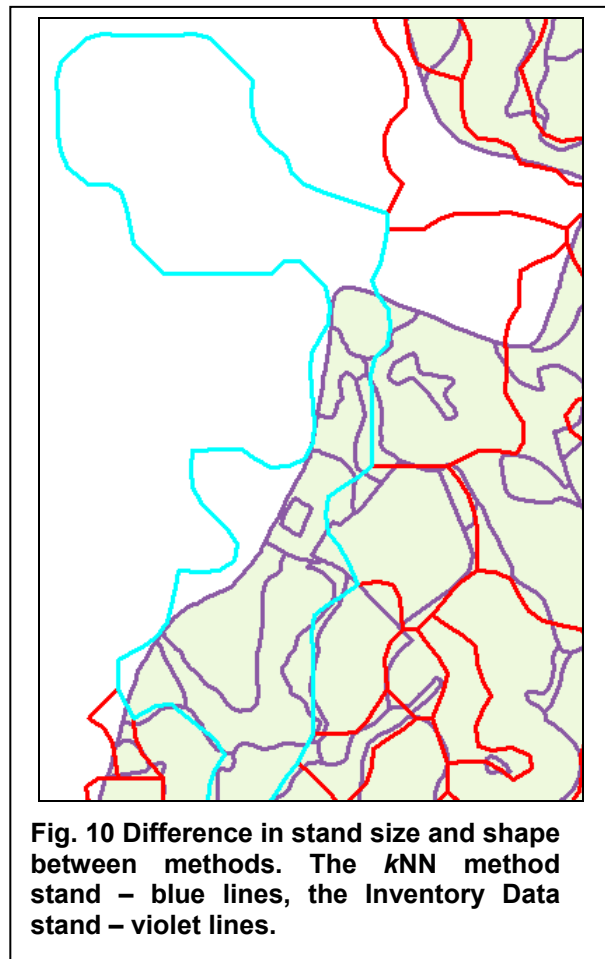
The discrepancy in the total area of estate, almost 41 ha more in the case of  $k$ NN, is caused by the specific shape of polygons in the segmented data, which is not always compatible with the real stand borders. It was problematic to select polygons which would contain only Asa

Research Park property, for some of them included an area larger than a real property of estate (Fig. 10). However, such a problem is to be encountered when two datasets with different levels of generalisation are compared.

Edge points defined by WINDA represent exposed stand borders, the parts of a stand exposed to the wind damage. The Inventory Data is characterised by 4,5 times as big number of exposed points (Tab. 2) as in the case of the *k*NN Data. Data precision accounts for that. The *k*NN Data contains a much lower number of stand-border elements. Consequently smaller number of nodes in the digitised polygons caused such a difference in the number of defined exposed points where WINDA locates exposed points between two consecutive nodes in the process of simulation. Despite the difference in the methods for stands' representation and in absolute values there is a similar pattern in distribution of simulated points into six sectors (Fig. 5, 6). It also shows clearly that the distribution of the length of exposed edges does not differ to such an extent as in the case of the total number of points.

The number of exposed points gives only a very synthetic overview of the characteristics of exposed edges. The length of exposed edges is a far more readable feature, it presents a realistic picture of the stand. The total length of exposed edges and the disproportion between datasets is much smaller than it was in case of number of exposed points only. The total length of exposed stand edges in the Inventory Data is 2 times bigger than in the *k*NN Data (Table 3, 4). It means that the smaller number of exposed edges in the *k*NN Data is compensated by greater length of edge element. Such an extent of disproportion is also caused by much more complicated stand borders in the Inventory Data than in the generalised *k*NN.

Figure 5 and 6 show that sector 1, 2 and 3 have bigger values for the *k*NN Data than for the Inventory Data. But the more important result is that in sectors 4, 5 and 6 the Inventory Data has shown higher values. Exposed edges in these sectors are facing in direction between  $150^{\circ}$  -  $330^{\circ}$  what makes them turned to the dominating wind direction in Sweden. The Inventory Data shows that there is slightly more edges exposed to the western winds than it is visible in case of the *k*NN Data. It raises a question about the reason for such pattern but with only one repetition of this experiment it is rather impossible to give any precise explanation. The differences are on a level of few percent only so it can be said that it is not significant. This statement is confirmed by the Mean Height Test made on data. However, this pattern where exposed edges to the west are underestimated in the *k*NN Data can be influenced slightly by a local relief with its localisation of Asa lake in the closest neighbourhood of the estate. The Asa lake is located on the eastern side of the forest estate, which is the direction faced by exposed points and edges, assigned to sector 2 and 3 and also partly 1 and 4.



Concluding this two features like the number of exposed points and the length of exposed stand edges, appear to yield similar results. The test on mean height of stands may serve as a confirmation of the statement. The mean height test shows that there are no significant differences in the case of stand height, perceived as the most important feature in this research. With the confidence level of 5% the hypothesis that maps are comparable could not be rejected. In terms of statistics no major discrepancies could be found. The very same method of tree height calculation verifies the reliability of the test. However, it must not be overlooked that assessment of height in the  $k$ NN method vary in accordance with the size of analysed area. There is no ground to claim the data is similar on a pixel level yet with the area aggregation it can be improved by 17% RMSE for 19 ha (Reese, Nilsson et al. 2002). If we assume that the reduction of RMSE will correspond to a larger size of the analysed forest, the use of the  $k$ NN on an estate level seems to be a viable solution in data collection.

All the results obtained in this thesis indicate that there are certain differences between the two sources of information about the forest, which can be used in risk simulators like WINDA. The question arises whether it is reasonable to employ the  $k$ NN method in such an application as WINDA. Indeed, the level of precision which is not adequate to provide sufficient information to represent a single stand (Fig. 9), and to calculate wind damage probability for such a given stand. The result of segmentation does not give us precise stand structure but close to the real forest height structure. It must be stressed at this point that in spatial computations provided in this research full stands structure description is not necessary and determining the exposed to wind damage places along forest estate is possible with the use of both types of data. One may conclude that on the level of forest estate both the methods show similarities and the  $k$ NN method appears to be a very promising solution. The analyses of big estates, regions or bigger units on the level of landscape have been and will be limited to the shortage of necessary forest data. It may be attributed to, for example, various standards and systems of digitalized forest data used by different forest owners and forest associations. The  $k$ NN method serves as a relatively cheap, reliable and easily accessible source of required information.

In my research I have tested only segmented data containing the information concerning the height of ground vegetation only, which does not suffice to provide the whole simulation process in WINDA. Therefore, the simultaneous use of another  $k$ NN satellite datasets together with the rest of forest information seems to be an effective solution to the problem. The preparation of this kind of data requires much computer and analytical work as well as time, yet, as far as I am concerned it is the cheapest way to carry out such a simulation on a bigger area-level. The combination of datasets with different variables can also significantly increase the precision of the prepared datasets. A precise description of forest area is obtained as a result, which is in my opinion worth the effort of further investigation.

## 5. Conclusions

The segmented  $k$ NN Data was successfully machined and used in WINDA - wind damage probability simulator. The simulation process was conducted without any problems as it was stated at the beginning of the research. Substantial discrepancies in the number of defined exposed points come from the precision level of both datasets. Broad generalisation in the  $k$ NN Data does not impinge on the relative values within wind-direction sectors. The similar distribution of exposed points and edges shows that high precision the Inventory Data does not differ much from the precision of segmented data. The most noticeable difference in the higher number of exposed points and edges in the three east-directed sectors is not significant and can be caused by the local influence of relief. The correspondence of datasets is strengthened by positive results of mean height test. The results of the simulation can be regarded as acceptable for the level of estate or bigger area.

Further studies with use of  $k$ NN method in risk assessment are promising and the accuracy of assessments can be improved by the use of more  $k$ NN datasets with important forest information.

## Acknowledgments

I thank Kristina Blenow for giving me a chance to work on this thesis, the inspiration, co-operation and supervision in great atmosphere. Thanks to Olla Sallnäs for checking this paper and examination. Big thanks to Mike Anderson for his patience and enormous long hours combating the mysterious of Arc-problems that occurred unfailingly during data preparation. I am grateful to Erika Olofsson for help with WINDA simulation process. I also thank for the access to computer laboratory to Southern Swedish Forest Research Centre, Swedish University of Agricultural Sciences. Big thank you for friendliness in the department.

## Literature

- Anon. (2002). ArcInfo 8.1. Redlands, CA, USA, ESRI Inc.
- Anon. (2002). ArcMap 8.1. Redlands, CA, USA, ESRI Inc.
- Blennow, K. and O. Sallnäs (2004). "WINDA - a system of for assessing the probability of wind damage to forest stands within a landscape level." Ecological modelling **175**: 87 - 99.
- Gardiner, B., H. Peltola, et al. (2000). "Comparison of two models for predicting the critical wind speeds required to damage coniferous trees." Ecological Modelling - **129**(- 1): - 23.
- Gardiner, B. A. and C. P. Quine (2000). "Management of forests to reduce the risk of abiotic damage - a review with particular reference to the effects of strong winds." Forest Ecology And Management **135**(1-3): 261-277.
- Hollenstein, K. (1997). "Analyse, Bewertung und Management von Naturrisiken." Hochschulverlag AG der ETZ Zurich.: 191.
- Holmström, H. (2001). Data Acquisition for Forestry Planning by Remote Sensing Based Sample Plot Imputation. Department of Forest Resource Management and Geometrics. Umea, Swedish University of Agricultural Sciences. **Doctoral Thesis**.
- Holmström, H., M. Nilsson, et al. (2001). "Simultaneous estimations of forest parameters using aerial photograph interpreted data and the k nearest neighbour method." Scandinavian Journal Of Forest Research **16**(1): 67-78.
- Morgan, J. and M. G. R. Cannell (1994). "Shape of tree stems: a re-examination of the uniform stress hypothesis." Tree Physiology **5**: 63 - 74.
- Mortensen, N. G., L. Landberg, et al. (1998). Wind Atlas and Application Program (WASP). Roskilde, Denmark, Riso National Laboratory.
- Olofsson, E. and K. Blennow (2005). "Decision support for identifying spruce forest stand edges with high probability of wind damage." Forest Ecology And Management **207**(1-2): 87-98.
- Peltola, H., B. A. Gardiner, et al. (2000). "Wind and other abiotic risks to forest." Forest Ecology And Management **135**: 1 - 2.
- Peltola, H., S. Kellomaki, et al. (1999). "A mechanistic model for assessing the risk of wind and snow damage to single trees and stands of Scots pine, Norway spruce, and birch." J. For. Res. **29**: 647 - 661.
- R Development Core Team (2004). A Language and Environment for Statistical Computing (Version 2.0.1., 2004-11-15).
- Reese, H., M. Nilsson, et al. (2002). "Applications using estimates of forest parameters derived from satellite and forest inventory data." Computers and Electronics in Agriculture **37**: 37 - 55.
- Skogsstyrelsen - National Board of Forestry (2005). Inventory by plane confirms earlier estimation of forest damages, <http://www.svo.se/minskog/templates/Page.asp?id=15432>.
- Tomppo, E. (1990). Designing a satellite image-aided national forest survey in Finland. The Usability of Remote Sensing for Forest Inventory and Planning. SNS/IUFRO

workshop, Department of Forest Resource Management and Geomatics. Umea, Swedish University of Agricultural Sciences.

Tomppo, E., R. Czaplewski, et al. (2002). The role of remote sensing in global forest assessment. Forest Resources Assessment Programme. Forestry Department Food and Agriculture Organization of the United Nations. Rome, FAO.

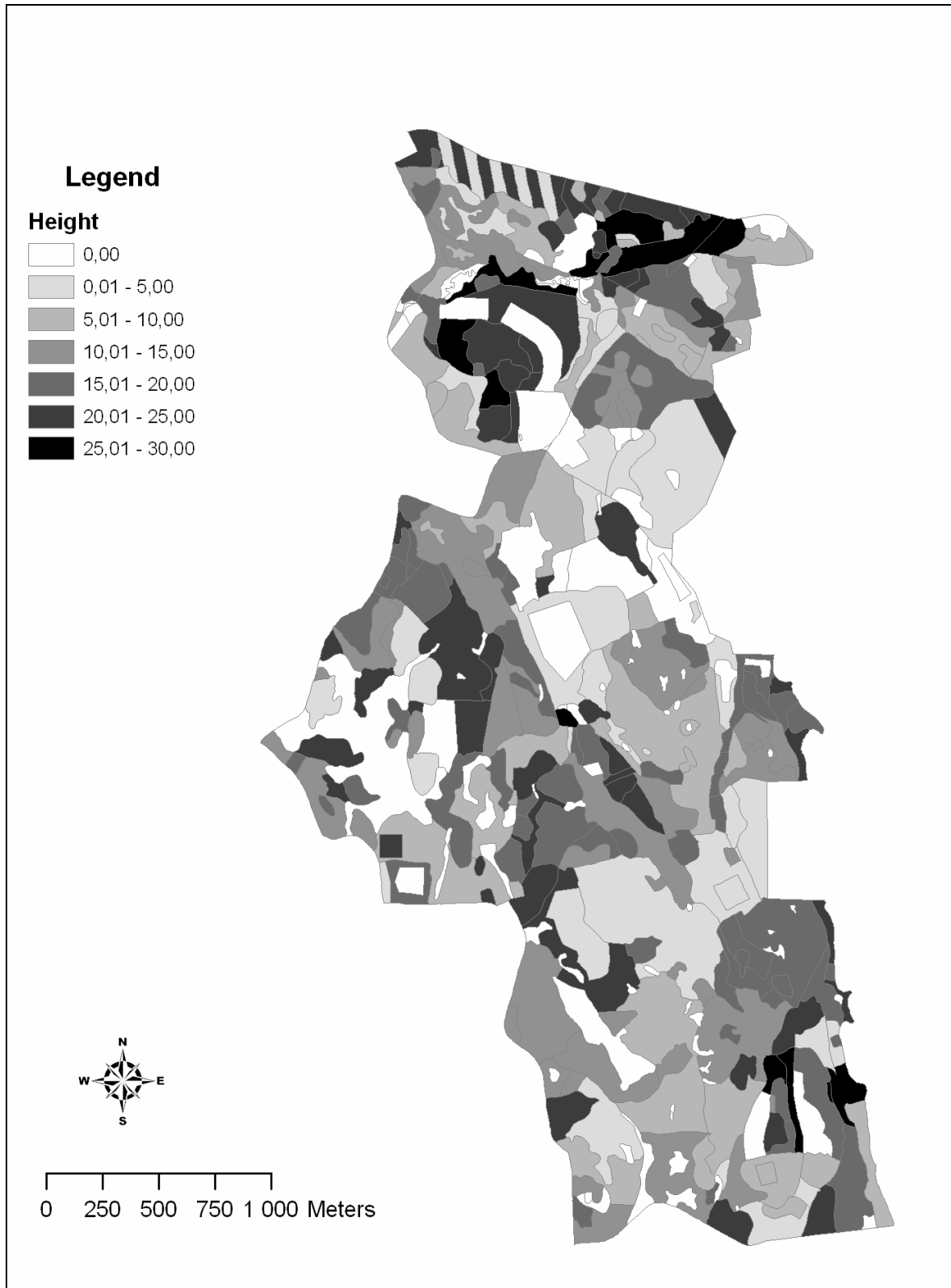
Valinger, E. and J. Fridman (1999). "Models to Assess the Risk of Snow and Wind Damage in Pine, Spruce, and Birch Forests in Sweden." Environmental Management **24**(2): 209 - 217.

von Gadow, K. (2000). "Evaluating risk in forest planning models." Silva Fennica **34**(2): 181-191.

# Appendix

## Appendix 1.

Asa Research Park height map based on forest management plan, values divided into 5 meters divisions.



Appendix 2.

Asa Research Park height map based on the segmented *k*NN Data, values divided into 5 meters divisions.

