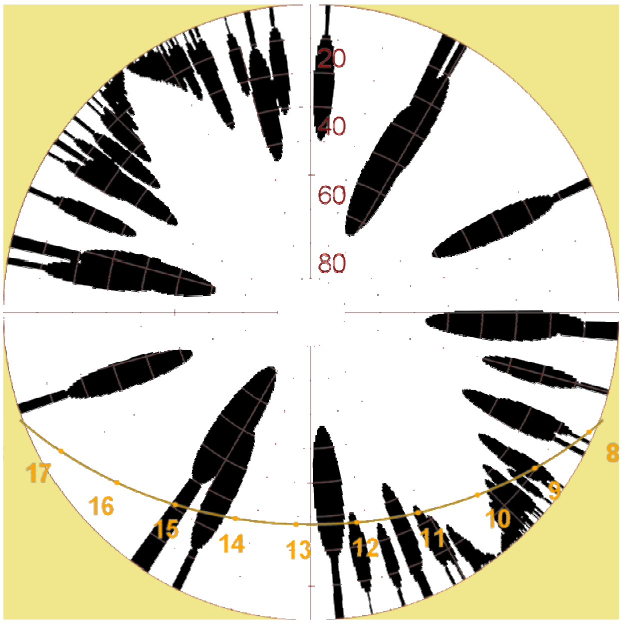
Centre for Geo-Information

Thesis Report GIRS-2011-02

Modeling the influence of road surroundings on the Sky View Factor using GIS

L.M.V. van der Toorn



April 12th, 2011

WAGENINGEN UNIVERSITY
WAGENINGEN UR

MODELING THE INFLUENCE OF ROAD SURROUNDINGS ON THE SKY VIEW FACTOR USING GIS

L.M.V. van der Toorn

Registration number 831111-839090

Supervisors:

John Stuiver
Ron van Lammeren

In collaboration with:

Meteo Consult

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

April 12th, 2011

Wageningen, the Netherlands

Thesis code number: GRS-80436
Thesis Report: GIRS-2011-02

Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing

ACKNOWLEDGEMENTS

First and foremost, a deep bow to my supervisor, John Stuiver, for convincing me to continue my studies in the first place. You've always seen potential in me and made me believe in what I was doing. Without your profound trust and patience this thesis would not exist.

Ron, thank you for seeing endless possibilities and always reaching beyond the sky. Although your workload did not allow you for deep involvement in my topic, your presence in the building (especially during the calm and peaceful hours) was enough to motivate me.

A deep bow to Daniël van Dijke, Ingeborg Smeding-Zuurendonk and Marcel Wokke of Meteo Consult. Thank you so much for your feedback, knowledge and support during the research.

I was very grateful to receive vegetation data from both the province of Gelderland and the municipality of Ede. Prof. Matzarakis was very kind to provide both software and feedback on how to use it.

On a personal note, I want to thank my parents for patiently informing about my progress from a distance, never pushing. I know how much you'd like to be here, to celebrate the completion of my thesis. Stooor klem!!!!

A large chunk of gratitude goes out to my dearest Arjuna, for opening my eyes to the universe outside the thesis constellation. Thank you for both keeping me from work and for urging me on, and for always staying true to yourself. Mwah, mwah, MWAH!!!

Teşekkür Gülsüm! My lovely lady, thank you for being there, ALWAYS generous, always passionate, in one way or the other ;) Çok çok öpücük!!!

A huuuge hug to my fellow thesis-strugglers, nerds, and party animals: Ban-An, Dzjen, Da PartyMeister, Judith, Mark, Tomas ("THANK YOU TOM!!!!" ;)), and Sarah. Thanks a million for your wonderful company!

During the last stretch of my thesis two small creatures decided to enter this world before their time and to show everyone around them the true meaning of hope, struggle and unconditional love. Whether on this earth or not, you will always be loved and always be remembered.

EXECUTIVE SUMMARY

During nights when temperatures fluctuate around freezing, road temperatures are often underestimated which leads to excessive salting. In order to optimize salting procedures, an accurate forecast of road conditions is essential. Meteo Consult, a commercial weather forecast company in the Netherlands, has made significant progress in developing a model to predict road conditions. This model generates a forecast of the road surface temperature based on the weather forecast, the road construction and the road surroundings.

The influence of the road surroundings is expressed by two factors: the sun view and sky view factor. The sun view is a measure of how much radiation is received by the sun. The Sky View Factor (SVF) represents what proportion of the sky is visible from a given point on the road. Several studies have shown that the SVF has a significant impact on the road surface temperature when the nights are clear and calm.

Currently, Meteo Consult derives these two factors from fish-eye photographs, a method with many drawbacks. In order to find a viable GIS-based replacement for this method, Meteo Consult has co-supervised two studies done at Wageningen University. The last study, done by Joseph Steenbergen in 2009, showed that the use of GIS has a very high potential of replacing the photographic method if the modeling of trees is improved.

Therefore, this research focused specifically on modeling the influence of road surroundings along roads where trees are the main cause for sky view obstruction. In addition, the quality of the photographic method was analyzed in order to provide guidelines for the quality of the GIS model.

The GIS model developed during this study simulates the road surroundings using ground elevation data and a solid 3D tree model to represent leafless trees. A proof-of concept showed that this approach produces sky view factors that come very close to the photographic method, but that is cannot yet serve as a replacement.

The most important reason for this is that SkyHelios, the program that visualizes the surroundings in 3D, cannot supply the sun view factor. It is expected that this can easily be implemented.

The research showed that the model requires accurate tree height data in order to produce accurate results. Therefore, a method needs to be developed to automatically derive tree heights (and possibly other tree characteristics) of individual trees from the AHN vegetation dataset; the most accurate source of tree height data.

Furthermore, it is recommended to take more road surroundings into account in order to provide a better simulation of the road surroundings. These objects can be derived from the Top10NL and provided with their heights using the AHN.

Last but not least, the conversion method used to create 3D trees should be improved. The model creates 3D trees based solely on the tree height. The dimensions of the 3D tree are calculated by multiplying the tree height by specific parameters. It is to be expected that each tree species requires a different set of parameters in order to simulate their shape. In addition, the creation of 3D trees representing leafless trees would probably be easier if the tree model would be extended with a transparency element.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS V

EXECUTIVE SUMMARY VII

1. INTRODUCTION 1

1.1 Context 1

1.2 Problem Definition 2

Determining the sky view factor 2

Improving GIS modeling 2

Input and modeling 3

Validation 4

1.3 Research Objectives 4

1.4 General Methodology 5

1.5 Content structure 5

2. QUALITY OF THE REFERENCE DATA 7

2.1 Introduction 7

2.2 Methodology 7

Acquisition and processing of the reference data 7

Comparability 9

Implementation 11

2.3 Results and discussion 15

Step a. Selecting comparable SVF data 15

Step b. Enabling point comparison by distance from a common point 16

Step c. Analyzing the relationship between road surroundings and positioning error 18

Step d. Selecting comparable point sets 21

Step e. Analyzing the differences in SVF using four comparisons 22

2.4 Conclusions and recommendations 47

Location of compared points 47

Types of surroundings 47

Different lane or year 48

Processing error 48

Quantifying quality 48

3. DESIGN AND DEVELOPMENT OF THE GIS MODEL 49

3.1 Introduction 49

Creating a new GIS model 49

Development process 50

3.2 Requirements 50

3.3 Design 52
Software 52
Data 57
Conceptual design of the GIS model 58
3.4 Building a basic model 59
Step 1. Pre-processing 60
Step 2. Calculation of tree elements 60
Step 3. SVF Calculation 62
3.5 Discussion and conclusions 63
Modeling method 63
Road surroundings 64
3.6 Recommendations 64
SkyHelios 64
Tree model 65
Road surroundings 66
4. APPLICATION OF THE GIS MODEL 67
4.1 Introduction 67
Chapter content and structure 67
Adoptions to the model 67
4.2 Calibration 68
Location 68
Approach 70
Results 73
4.3 Model sensitivity 75
Tree elements parameters 75
Height data accuracy 77
Additional surroundings 82
SVF calculation methods 84
4.4 Validation 84
Locations 85
Approach 87
Results and discussion 88
4.5 Conclusions and recommendations 91
Applicability 92
Sensitivity 92
5. DISCUSSION 93
Quality of the reference data 93
Road segments 93
Sun view factor 94
Model input 94
Model performance and applicability 94

6. CONCLUSIONS 95
Question 1: What is the quality of the reference data? 95
Question 2: What is the definition of a road segment and how can it be created using geo-data? 96
Question 3: Can the sky view factors of road segments be calculated using a GIS model? 96
Question 4: What is the quality of the GIS model? 97
Main conclusion 97

7. RECOMMENDATIONS 99
7.1 Quality analysis of the reference data 99
Location of compared points 99
Comparing points using SVF plots 99
7.2 GIS model 99
SkyHelios 99
Increasing the model's accuracy 100

8. REFERENCES 103

APPENDICES 105

I. STEP C: CORRELATION BETWEEN GPS POSITIONING ERROR AND SVF 106

II. STEP D: DIFFERENCES IN RESULTS DEPENDING ON THE INPUT DATASET 109

III. GIS MODEL PROCESSING IN DETAIL 112

IV. TRANSECTS: TREE HEIGHTS SAMPLED FROM THE AHN VEGETATION DATASET 113
First validation location: N224 30.0 – 31.4km 113
Second validation location: N224 32.0 – 33.0km 115

V. CHANGES MADE TO THE TREE HEIGHTS BASED ON THE AHN VEGETATION DATSET 117
First validation location: N224 30.0 – 31.4km 117
Second validation location: N224 32.0 – 33.0km 119

1. INTRODUCTION

1.1 Context

Winter road management in the Netherlands is a challenging task. Every time the cold weather seems to pose a threat to road safety, the necessity of salting is weighed carefully in order to reduce both costs and environmental impact. The question is not whether or not to treat the roads, but also when, where and to what extent exactly (Wassenaar, 2009; Wisse, Zuurendonk, & Wokke, 2008). In order to optimize salting procedures, an accurate forecast of road conditions is essential (Sass, 1997; Wokke & Wisse, 2007).

The most common methods to predict slippery roads in the Netherlands are based on road sensor measurements and the weather forecast alone (Meteo Consult et al., 2009). Each road management area has a few road sensors positioned at the coldest spots of the road network. Based on the sensor measurements a forecast can be made for each sensor position. During nights when temperatures fluctuate around freezing, the road temperatures for the rest of the network are often underestimated, inevitably leading to excessive salting (Wisse, et al., 2008).

In order to make winter road management more effective, meteorologists and road management parties joined forces in the RGI[1] project 'Prevention of slippery roads based on local meteorological, thermal mapping and GPS data'[2]. During this cooperation Meteo Consult, a commercial weather forecast company in the Netherlands, made significant progress in developing a model to predict road conditions. This so called 'network model' generates an hourly forecast of the road surface temperature (RST) for every segment of the road network, for a specific road management area (Meteo Consult, et al., 2009).

The network model has three main sources of input: the weather forecast, the road construction and the road surroundings (Meteo Consult, et al., 2009; Wokke & Wisse, 2007). The influence of the road surroundings is expressed by two factors: the sun view and sky view factor. The sun view is a measure of how much radiation is received by the sun (Wokke & Wisse, 2007). The Sky View Factor (SVF) represents what proportion of the sky is visible from a given point on the road (Fry, Slade, Taylor, & Davy, 2007). Several studies have shown that the SVF has a significant impact on the road surface temperature, when the nights are clear and calm (Barring, Mattsson, & Lindqvist, 1985; Chapman, Thornes, & Bradley, 2001a; Eliasson, 1996; Wokke & Wisse, 2007).

The SVF is currently being measured using fish-eye photographs. Due to the drawbacks of this method (see chapter 1.2: 'Problem Definition') Meteo Consult is keen to know if GIS can be used to complement it. Therefore, this research will focus on modeling the SVF of road segments using GIS.

---

[1] RGI stands for 'Ruimte voor Geo-Informatie', Dutch for 'Space for Geo-Information'. RGI funded innovative projects in the field of geo-information up to September 2009, stimulating efficient use and exchange of knowledge in order to empower the industry. Source: www.rgi.nl

[2] Original title: 'Preventieve Gladheidsbestrijding op basis van meteorologische, thermal-mapping en GPS data ter plekke'

1.2 Problem Definition

Determining the sky view factor

As mentioned earlier, the SVF indicates what proportion of the sky is visible from a given point on the road. This is expressed in a dimensionless value between 0, when the sky is completely obscured, and 1, in a completely open area where the whole sky can be seen (Thornes, Cavan, & Chapman, 2005). During clear and calm nights, a road location with a high SVF cools more quickly, because most of its long wave radiation is lost to the cold sky instead of being trapped and returned by surrounding objects (Chapman, Thornes, & Bradley, 2001b).

There are many methods to determine the SVF, of which fish-eye photography is the most accurate at the moment (Chapman, Thornes, & Bradley, 2002): photographs capture the entire surroundings and the SVF can be derived automatically by counting the amount of 'sky' pixels in the photograph (Chapman, et al., 2001b). Another surveying method uses GPS satellite visibility to calculate the SVF. This method has proven accurate in urban environments, but not in suburban and rural environments as trees cause noisy data (Chapman, et al., 2002). Recently, GIS modeling has also been used to calculate the SVF, predominantly in urban climate studies. Using building data and Digital Elevation Models (DEMs) as the main source of input, the SVF can be calculated quite accurately (Lindberg, 2007; Souza, Rodrigues, & Mendes, 2003; Unger, 2009). Vegetation data is not included as they are not part of any (available) GIS datasets of the research areas in question (Fry, Slade, Taylor, & Davy, 2007; Lindberg, 2007) and vegetation height is difficult to derive from available national DEMs. DEMs are namely produced to portray the ground based height, not the vegetation height, and are therefore surveyed in the winter, when vegetation is often leafless. Even if accurate tree height data is used, characteristics like crown shape and seasonal variety in foliage make trees difficult objects to model[3].

Improving GIS modeling

Meteo Consult uses fish-eye photography to calculate the SVF. However, this developed method has substantial drawbacks[4]:

- The (labor) costs of surveying and processing are high;
- The photos are sensitive to specific weather conditions. Homogeneous cloudy weather is required in order to be able to process the image correctly. This restricts the surveying window;
- Traffic is often a source of error: by obstructing the sky view it causes an underestimation of the SVF;
- The fish-eye camera is mounted on the top of a car, approximately 1.5 m higher than the road surface, which may lead to a different SVF than when the sky is seen from the road surface.

[3] Steenbergen, J. J. M. (2009). *Computing Sky View Factors from geo-data using a GIS* (Unpublished Master thesis). Laboratory of Geo-Information Science and Remote Sensing, Wageningen University and Research Centre, Wageningen.

[4] Personal communication with D. van Dijke, Meteo Consult, September 11, 2009

For these reasons, Meteo Consult is interested in finding out if GIS modeling can be used to replace fish-eye photography. Although it might be too much of a challenge to replace the photographic method altogether, GIS modeling could be used to lower the surveying frequency (by taking local changes in road surroundings into account) and to aid in error detection of the fish-eye photographs.

In collaboration with the Laboratory of Geo-Information Science and Remote Sensing of the Wageningen University, Meteo Consult has co-supervised two studies on this subject. A team of MSc students conducted a preliminary study on how GIS could be used to model environmental influences on the RST[5]. MSc student Joseph Steenbergen continued where the previous team left off, focusing his Master thesis on improving the accuracy of SVF calculations by including trees[6]. Since a significant part of Meteo Consult's RST predictions are done on roads in a rural environment, it was imperative that the model would include trees.

Although Steenbergen's results were promising, recommendations were made to improve tree modeling. This research will build on three of these recommendations:

- Check the quality of the reference data;
- Use new and improved input datasets;
- Improve the model calculations by taking more variables into account (e.g. the species of trees).

Input and modeling

Steenbergen's GIS model calculated the SVF by modeling individual trees based on their height and the transparency of the crown. The trees were added to a raster DEM using 2.5D extrusion. The transparency was calculated based on the tree height only (thus unrelated to e.g. shape or species). Unfortunately, tree management data (provided by the municipality of Ede) were found inadequate to act as input, because tree height is stored in just 3 categories. Therefore, he used heights based on AHN height data accompanied by estimations done in the field instead.

To avoid limitations concerning data inadequacy, the focus will shift from modeling the SVF for points along a route, to modeling the SVF of road segments. This shift in methodology is based on the assumption that the SVF varies within a certain range where the road surroundings are the same.

---

[5] Giffen, H. v., Huis, L., Narieswari, L., Roetman, J. M., & Spekken, M. (2008). *Assessment of sky-view and other environmental properties based on digital maps* (Unpublished Academic Master Cluster report). Laboratory of Geo-Information Science and Remote Sensing, Wageningen University and Research Centre, Wageningen.

[6] Steenbergen, J. J. M. (2009). *Computing Sky View Factors from geo-data using a GIS* (Unpublished Master thesis). Laboratory of Geo-Information Science and Remote Sensing, Wageningen University and Research Centre, Wageningen.

The road segments will be calculated based on the homogeneity of the road surroundings: if a stretch of road has the same surroundings for 200 meters for example, it is expected that the SVF will be the same throughout that segment. (This can be validated using Meteo Consult's reference data.) Given this new approach, the precision of the tree management data is assumed to be sufficient. The sensitivity of the model will be tested in order to quantify the effect of the inexact input on the output of the model.

As a replacement for the laborious photographic method, the execution of the GIS model should require less man-hours and minimal fieldwork of any kind. Therefore, the new GIS model will be based on existing geo-data only.

Validation

The validation possibilities of Steenbergen's results were limited in two ways.

Firstly, his research area consisted of a trajectory chosen to include "complex 3D landscape geometry"[7]. Unintentionally, this limited the ability to validate the results of the model to similar road segments in the area. In order to prevent this, several roads with comparable surroundings will be selected to see if the calculated sky view factors correspond.

Secondly, the reference data consisted of a single series of fish-eye photographs provided by Meteo Consult. Without an impression of the accuracy and reliability of the reference data, it was not possible to judge the quality of the GIS model. As Meteo Consult also has a high interest in improving the validation possibilities of their own model, a second series of fish-eye photographs are made as a part of this research.

1.3 Research Objectives

The main objective of this research is to improve GIS modeling of the SVF along roads, by investigating the quality of the reference data provided by Meteo Consult and developing a model that assigns sky view factors to road segments. The SVF is calculated based on specific characteristics of the surroundings that can be derived from geo-data.

The main research question is:

Can a GIS model produce sky view factors for road segments of a quality equal to that of the photographic method?

In order to answer the main research question four research questions have to be answered first:

1. What is the quality of the reference data; is there a significant difference in the sky view factor when comparing fish-eye photos taken in the same area, at different moments in time (in the same season of the year)?
2. What is the definition of a road segment and how can it be created using geo-data?

[7] Steenbergen, J. J. M. (2009). *Computing Sky View Factors from geo-data using a GIS* (Unpublished Master thesis). Laboratory of Geo-Information Science and Remote Sensing, Wageningen University and Research Centre, Wageningen, p.10.

3. Can the sky view factors of road segments be calculated using a GIS model?
    3.a. Which modeling method is best suited?
    3.b. Which road surroundings are (not) taken into account?
4. What is the quality of the GIS model?
    4.a. How sensitive is the model?
    4.b. How does the model perform on roads with strong variations in surroundings?

1.4 General Methodology

The research questions can be arranged into three subjects:

- Analysis of the quality of the reference data (RQ 1);
- Design and development of the GIS model (RQ 2 and 3);
- Analysis of the application of the GIS model (RQ 4).

These points will be investigated in the order they are mentioned; each point is handled in a separate chapter.

The analysis of the reference data is done first to find out how accurate the current method is. New photographs are taken (in both driving directions) along roads in rural areas, where trees are the main cause of sky view obstruction. The SVF values along several road stretches are then analyzed in further detail, by comparing SVF values of one lane to values in the opposite lane, or by comparing them to values in the same lane (if the lane has been surveyed before). The information on the accuracy of the current method can be used in next steps, where a new method is developed and tested.

The second part involves designing and developing a GIS model. The purpose of the new model is to calculate the SVF along rural roads that are mainly surrounded by trees. When developing the model, the focus is therefore on the conversion of tree data into 3D trees that are defined by certain tree elements. In this chapter only the foundation of the model is laid; a workflow is developed that defines which data should be processed in which way in order to produce sky view factors in an area of interest.

In the next chapter the model will be calibrated in order to find which values should be used to define certain tree elements, so that the surroundings are mimicked as realistically as possible. By comparing the outcome to the reference data, the model is calibrated on one location and validated on two other locations. At the end of the chapter the model's accuracy is quantified and its reliability is discussed.

1.5 Content structure

The three subjects mentioned above are addressed in chapters 2, 3, and 4. Each of these chapters is written like a separate report, where the research question is answered as if it was the main research question.

Chapter 2 starts with an ‘Introduction’, followed by ‘Methodology’, which explains how the reference data was acquired and which steps were taken in order to compare them. The results of these analyses are presented per step in ‘Results and discussion’. Due to the numerous findings, the results are discussed as they are presented. The final paragraph, ‘Conclusions and recommendations’ delineates what affects the quality of the reference data to which extent and provides recommendations on how these findings can be applied during the development of the GIS model.

In chapter 3 the focus of the basic model is outlined in the ‘Introduction’ first. Thereafter, the chapter is structured according to the development process, described in paragraphs ‘Requirements’, ‘Design’, and ‘Building a basic model’. Afterwards, ‘Discussion and conclusions’ answers the third research question and discusses the pros and cons of the developed model. The chapter ends with ‘Recommendations’, where advice is given on how the basic model can be improved.

Chapter 4 first introduces the aim of the chapter and its structure. ‘Calibration’ explains how the model was fine-tuned to match the reference data on a small location. The next paragraph, ‘Model sensitivity’ displays how changes to the model’s parameters affect the output. Subsequently, the model is applied on two long road stretches in paragraph ‘Validation’ to see if the parameters found during calibration also provide good results under similar conditions on other locations. Finally, in ‘Conclusions and recommendations’, the forth research question is answered and recommendations are given as to how the model’s accuracy can be improved.

In chapter 5, ‘Discussion’, all points of discussion of chapter 2 through 4 are brought together.

Chapter 6 (‘Conclusions’) answers the main research question and summarizes the findings in chapters 2 through 4.

The following and last chapter, ‘Recommendations’, summarizes the recommendations given in chapters 2 through 4 to provide advice on how the model should be improved to become a viable alternative to the photographic method.

2. QUALITY OF THE REFERENCE DATA

2.1 Introduction

A good estimation of the quality of the SVF is needed to know which level of accuracy can be reached using the photographic method. Reference datasets (fish-eye photos accompanied by their derived values, like SVF and sun view) have been made in 2007 and 2010. New reference data were needed for two reasons:

1. to get photos in both driving directions to see if it is sufficient to measure the SVF of just one lane;
2. to get photos in the same driving direction to see if (and why) there is a difference in SVF in the same traffic lane after a few years.

The reference datasets will be compared to get an estimate of the accuracy of the SVF in relation to the road surroundings. The next paragraph describes the steps taken in order to compare the data. The results of the implementation of these steps are showed afterwards, accompanied by a discussion. The chapter ends by drawing conclusions on the findings and answering the first research question, followed by recommendations on how the comparison could be done better.

2.2 Methodology

Acquisition and processing of the reference data

As mentioned, the first reference dataset was made in January 2007 and the new reference dataset was made in February 2010. The fish-eye photographs were taken approximately in the middle of the winter to be able to use the SVF values as average values for the whole winter season. (No photos were made at other moments in the winter season to account for seasonal change in SVF. Seasonal change in sun view is accounted for during processing.)

The new route has a partial overlap with the old route along the provincial road N224 and the highway A12 (see Figure 1). Extra roads outside the old route were selected to get more reference data for tree rows and forests.

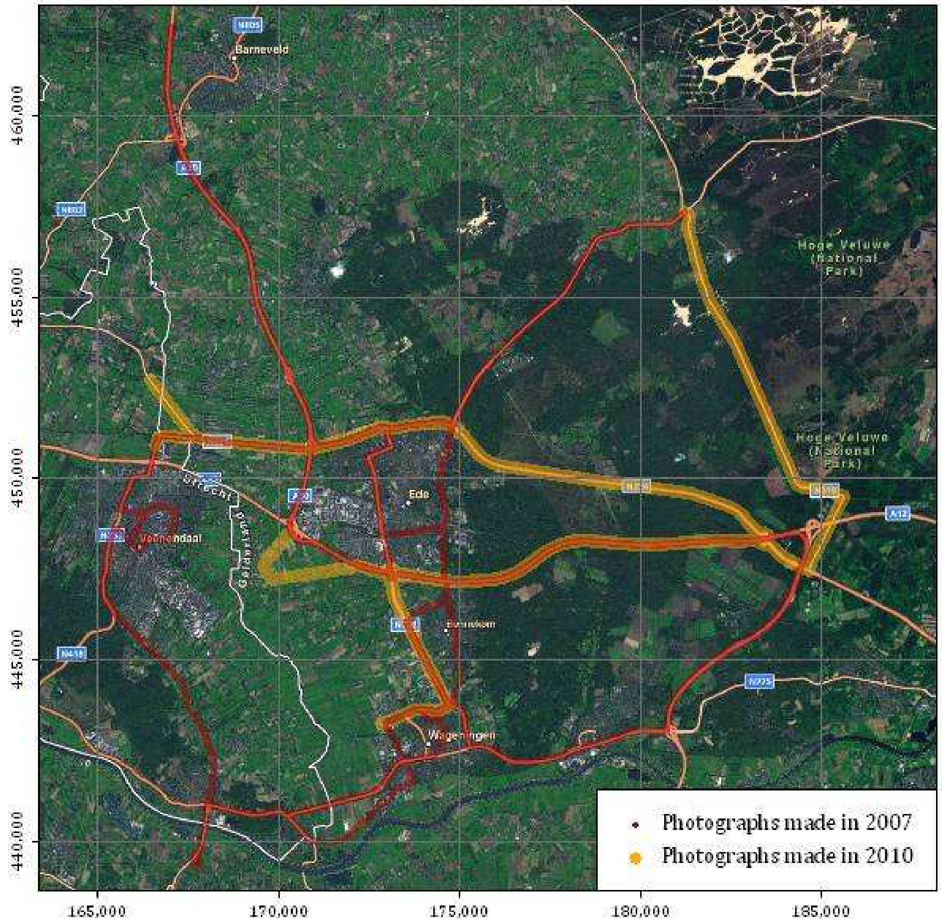


Figure 1: Locations of the fish-eye photographs taken in 2007 and 2010

Meteo Consult provided the equipment, software and expertise necessary for the acquisition and processing of the new reference data.

The process of calculating the SVF value of a fish-eye photograph is seen in Figure 2.

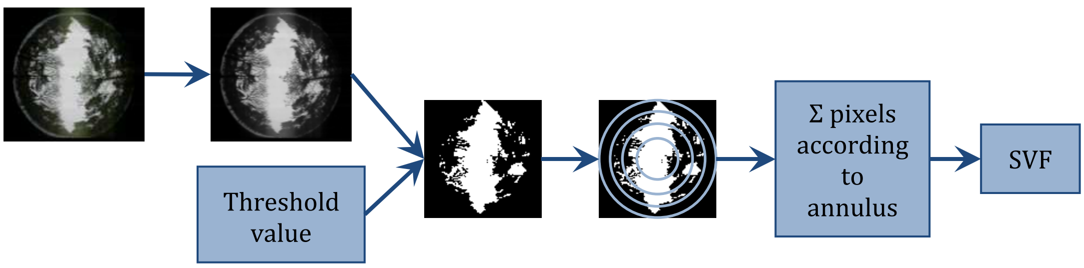


Figure 2: From fish-eye photograph to SVF value: The fish-eye photograph is converted from color image to binary image via a grayscale image. A threshold value is given to convert a grayscale image into a binary image. The SVF is calculated by summing the amount of sky binary pixels according to the formula of 'Chapman' (Chapman & Thornes, 2004).

Ordinarily, Meteo Consult processes small sections of a survey at a time, paying attention to setting the right threshold value and deleting images that contain dynamic objects. This is done to provide a high quality RST prediction for the customer. The new reference dataset is only used for this research, allowing two concessions to reduce processing time:

1. The new photographs were processed using one threshold value for all photographs. A test performed with 900 photographs with a shift of 4% in threshold value to convert grayscale images to binary images. This produced a difference in SVF with a mean of -0.009, a standard deviation of 0.005 and a maximum deviation of -0.026. This provides an insignificant contribution to SVF fluctuations.
2. Ideally the photographs only depict road surroundings that are always there and no dynamic objects like large vehicles passing by. These objects obscure the sky more than normally would be the case, causing an underestimation of the SVF.

   For the production of the 2007 reference dataset this error was avoided by checking each photograph manually and deleting the unsuitable photographs from the dataset. Due to the laboriousness of this method, this was not done for the 2010 dataset. The effect of skipping this step may become apparent if a reference point shows a very high difference in SVF in comparison to its neighbors, but if the difference is modest it may well be overlooked. This all depends on which deviations are common to certain road surroundings.

Comparability

To enable a good comparison of the SVF values of different datasets, it would be ideal to compare several photographs of the exact same spot on the road (or at least in the same lane) shot within days or weeks of each other. This would provide a good opportunity to calculate the systematic error of the photographic method. Unfortunately, this kind of comparison is not possible with the current data. The photos available are either taken:

- on the same day but in an opposite driving direction, or
- in the same driving direction but with 3 years time difference.

Since the camera is set to take pictures at a temporal interval, it means that the distance between subsequent photo locations varies according to the speed of the vehicle. This makes comparison between two reference datasets more complicated (see Figure 3).

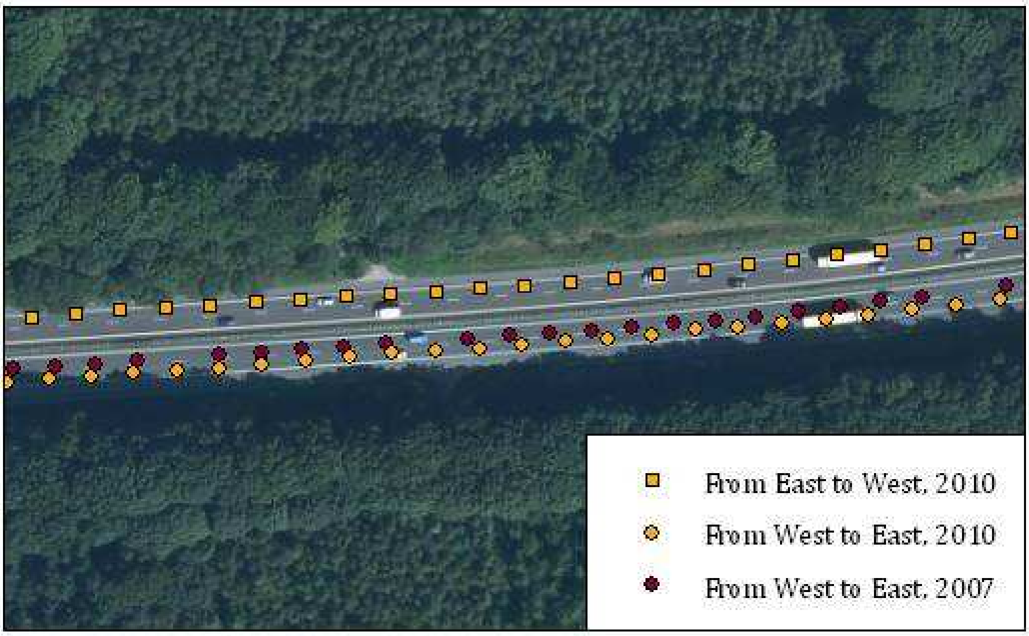


Figure 3: Example of photo locations along highway A12, in different directions and different years

The positioning accuracy also plays a role here. The position of the reference points is established by GPS measurements during the drive, meaning that its position is not acquired with great precision. This may lead to a situation where two points that seem close together on a map, might lie farther apart in reality, making them less suitable for comparison. This effect may be larger when the sky is obscured more, as less sky visibility might affect the GPS reception (Chapman, 2002). In other words: it is expected that the 'distance' between two points does not represent the actual distance, especially not along roads with dense surroundings. This assumed effect will be tested for every comparison made, by challenging the assumption that the closer the compared points are positioned, the smaller the difference in SVF must be. This assumption should be true if both the positioning error and the distance between the points are very low.

Despite of the problems concerning comparability, there are positive notes to make. Firstly, assuming that roads with two adjacent lanes (one for each driving direction) can be salted in one go, the sky view factors of opposite lanes can be averaged into one value. This is the case for most provincial roads, so measurements done in opposite lanes will be compared to see how much the SVF differs. Meteo Consult has also used this assumption for highways, where the lanes lie much further apart. A comparison of the data of opposite lanes of both road types will show if it is sufficient to measure the SVF of just one lane or not (comparison nr. 1 and 2 in Table 1).

Secondly, the SVF should be comparable if road surroundings have not changed significantly in the past three years. A significant change may be a new overpass or the removal of several adjacent trees in a tree row, while the growth of trees in a dense forest will probably go unnoticed. Photographs taken in the same lane will be compared to see if this is the case (comparison nr. 3 and 4 in Table 1).

Based on the availability of two comparable measurement sets ('same time, different direction' and 'same direction, different time') and two different kinds of roads (provincial roads and highways), four different comparisons can be made (see Table 1).

Table 1: Four comparisons of the SVF measurements

Comparison Nr.	Road type	Driving direction	Year
1	Highway	Opposite	2010
2	Provincial road	Opposite	2010
3	Highway	Equal	2007 & 2010
4	Provincial road	Equal	2007 & 2010

For every comparison two hypotheses will be either confirmed or disproved: the general hypothesis and one for the driving direction.

General hypothesis:

The closer the reference points are positioned, the smaller the difference in SVF.

Hypotheses for the driving direction:

- Hypothesis for comparisons 1 and 2 ('same time, different direction'):

  The SVF is the same if the sky is obscured equally on both sides of the road.

- Hypothesis for comparisons 3 and 4 ('same direction, different time'):

  The SVF is the same if the road surroundings have not changed significantly in the last three years.

Questions to assist in testing the hypotheses:

- What role does the GPS positioning error play?
- When comparing two SVF measurements, which deviation in SVF is normal and how large are the outliers?
- How can the differences in SVF be explained? (Change in surroundings? Measurement error? Processing error?)

Implementation

This paragraph describes the steps of the process executed to get the necessary results for answering the first research question. Figure 4 shows the steps in a flow chart.

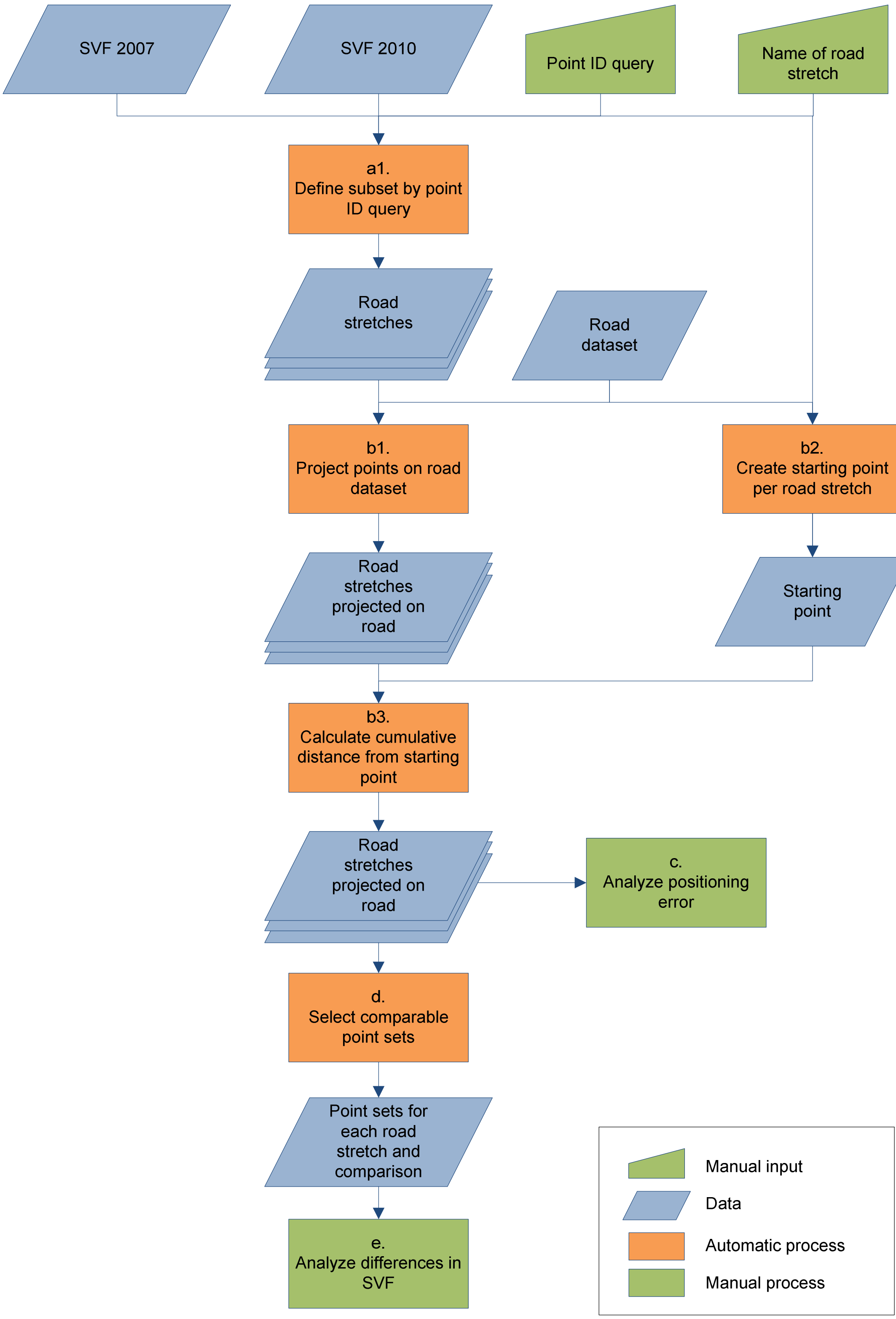


Figure 4: Flow chart of the implementation process

The SVF measurements are compared by:

a. Selecting a subset of reference points to represent the road stretches;

   Due to the large amount of data available a subset of the reference points is selected for the quality analysis. The points are defined manually by browsing the map, looking for provincial roads and highways that are surrounded by tree rows or forest.

   A road stretch is defined as a route along a road: it has a starting point and an end point. For each road stretch datasets are made depending on how this route has been surveyed: a subset is created for each driving direction and year. The maximum amount of subsets per road stretch is 4: a road can only have two driving directions and can only have been surveyed twice (in 2007 and 2010).

b. Enabling point comparison by distance by projecting SVF measurements on the centerlines of the roads and calculating the distance along the road from a common point;

   During the exploration phase of the thesis, the SVF values were plotted based on the X or Y coordinates and accompanied by a map using the same scale. Comparison between the points was not possible using this plotting method, because even though points seemed close together on the plot, the real distance between them was different (for example in road bends). So the idea arose to plot the points according to their distance along a common line, the centerline of the road, enabling a better interpretation of trends and deviations of the SVF by eye.

   Due to the costs involved of acquiring and processing fish-eye photographs, the SVF plots produced will not be displayed in full length in the report. If necessary, SVF of a part of a road stretch is provided for explanation.

   The reference points are positioned on the Top10NL centerline by creating a new point on the part of the centerline that is closest to the measurement[8] (see Figure 5).

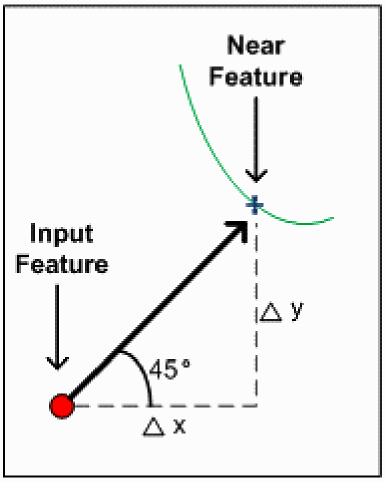


Figure 5: Euclidean distance to nearest point on the centerline (Source: ESRI.com)

[8] This calculation is performed using the 'Near (Analysis)' tool in ArcGIS 9.3.

Afterwards, a common point is established per road stretch: a point on the centerline at a small distance from the most western reference point. The distance to this point is then calculated for every point in each road stretch dataset.

c. Analyzing the relationship between road surroundings and positioning error;

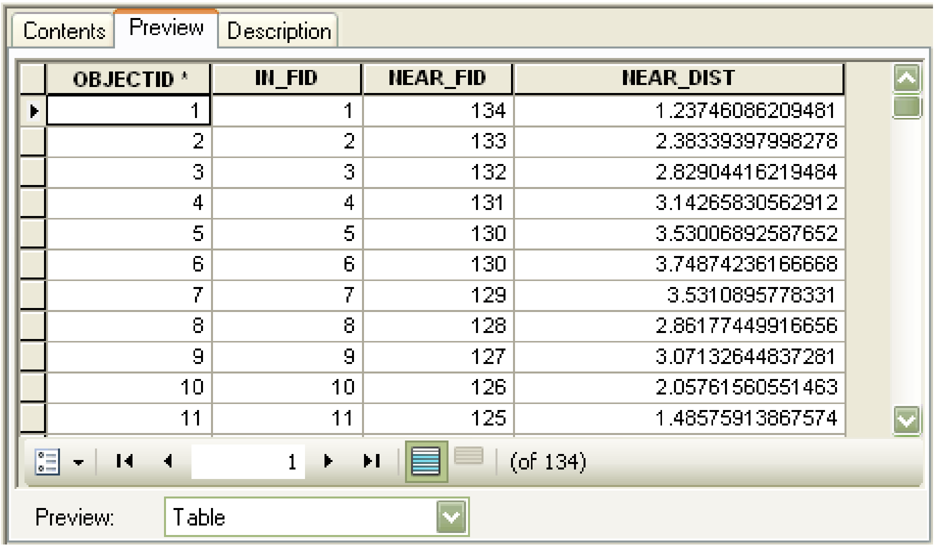
To analyze the relationship between the GPS positioning error and the SVF the correlation coefficient between them is calculated for every road stretch dataset. Plots are also made to be able to compare the results more easily.

This analysis is only performed on the survey done in 2010.

d. Selecting comparable point sets;

Every comparison needs two datasets per road stretch (data of the same lane surveyed in different years or data from two lanes surveyed in the same year).

The amount of resulting point sets depends on the amount of points in the input dataset: for every point in one dataset the closest point in the other dataset is found[9]. The result is stored in a table with three attributes: the FID of the feature in the input dataset ('IN_FID'), the FID of the nearest point in the other dataset ('NEAR_FID') and the Euclidean distance between these points ('NEAR_DIST').



OBJECTID *	IN_FID	NEAR_FID	NEAR_DIST
1	1	134	1.23746086209481
2	2	133	2.38339397998278
3	3	132	2.82904416219484
4	4	131	3.14265830562912
5	5	130	3.53006892587652
6	6	130	3.74874236166668
7	7	129	3.5310895778331
8	8	128	2.86177449916656
9	9	127	3.07132644837281
10	10	126	2.05761560551463
11	11	125	1.48575913867574

Figure 6: Part of the resulting table for road stretch A

This is a faster method than finding the closest point along the centerline using the distances already calculated in step b. The difference in the calculated point distance between these methods is minimal (namely centimeters).

e. Analyzing the differences in SVF using four comparisons.

This step compares the data of the point sets created in the previous step.

In order to prove or disprove the general hypothesis, the relationship between the distance between two points in a point set and their SVF difference is quantified by a correlation coefficient (R^2). A regression diagram is made to visualize the results.

[9] This calculation is performed using the 'Generate Near Table (Analysis)' tool in ArcGIS 9.3.

For the second and third hypothesis the correlation coefficient of the SVF of the two compared datasets is calculated and accompanied by a regression diagram. (The expected trends in the regression diagrams will be explained in detail in the 'Results and discussion' section.)

Fish-eye photographs, aerial photographs, and GIS topography are used to explain trends and outliers in the SVF plots.

2.3 Results and discussion

The results will be presented according to the implementation steps. Trends and outliers are discussed immediately afterwards.

Step a. Selecting comparable SVF data

The road stretches that have been selected for the quality analysis are highways and provincial roads surrounded by either tree rows or forest (see Figure 7).

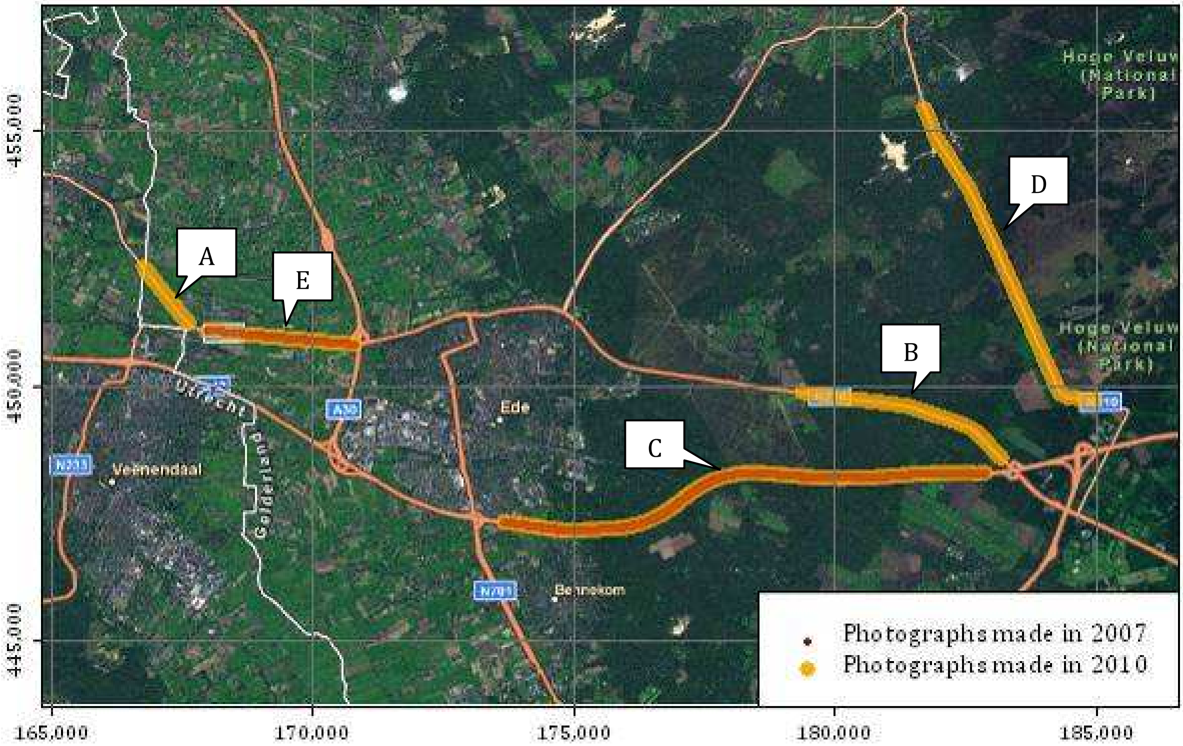


Figure 7: Location of the road stretches suitable for comparison

Road stretches A and E are provincial roads surrounded by tree rows. Road stretches B and D are provincial roads mostly surrounded by dense forest. Road stretch C is a highway partly surrounded by forest and open areas.

Road stretches C and E are surveyed twice in one lane: in 2007 and in 2010.

Step b. Enabling point comparison by distance from a common point

This method provides the possibility to plot the reference points according to their distance from a common point along the road centerline. Figure 8 shows an example of the result for a small part of highway.

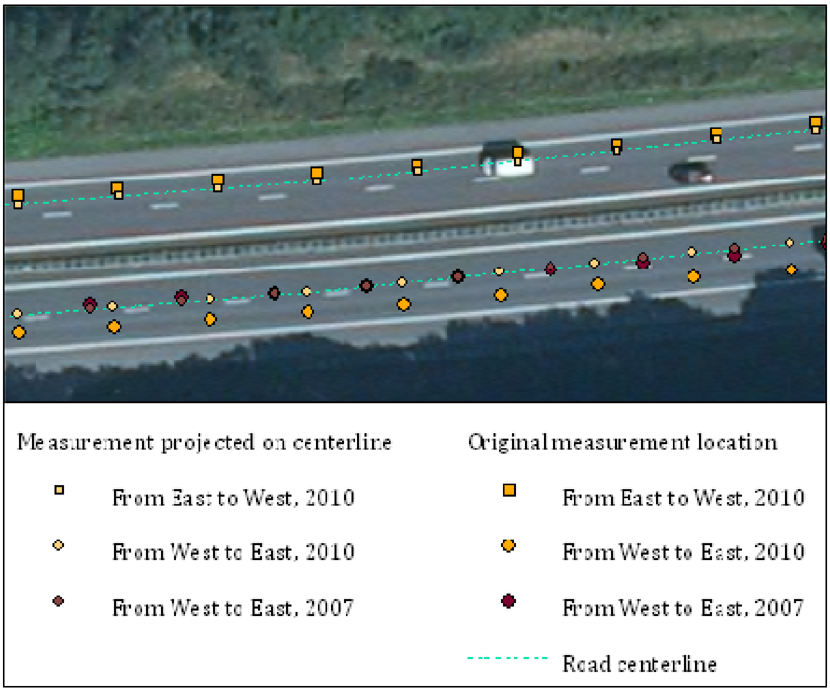


Figure 8: Example of reference points projected on the centerline of the road (highway A12)

When the projected points are plotted in a graph, as in Figure 9, it is easy to see trends in the SVF deviations. For example, for road stretch A you can see that the SVF in driving direction East-West (indicated as 'A EW 2010' with orange crosses) is systematically higher than in the opposite direction. This may indicate a slight systematic processing error or different surroundings on the sides of the road.

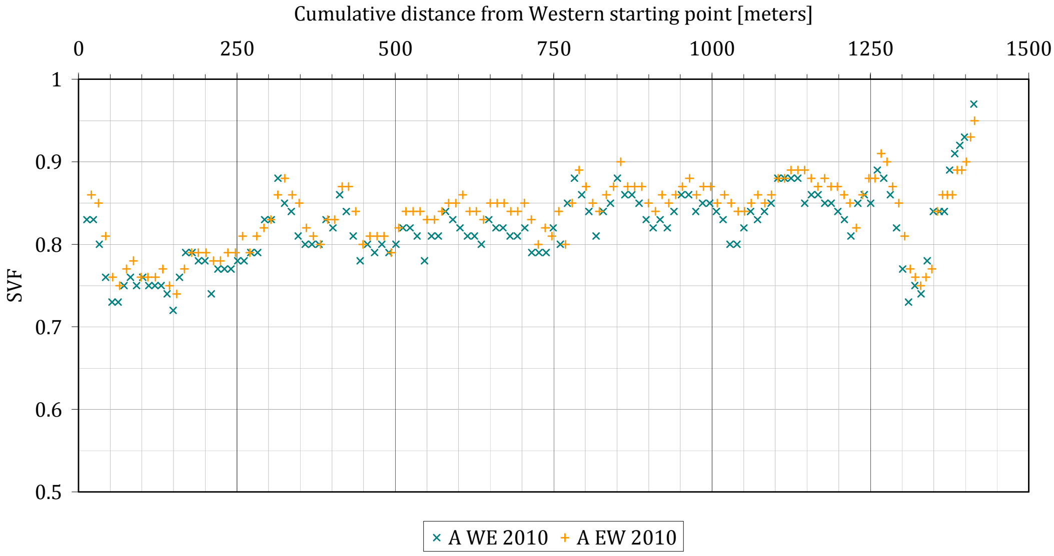


Figure 9: SVF values of all points of road stretch A plotted according to their distance from a common starting point

The method used for calculating the length of road between points was not suited for the situation at the traffic lights at the most eastern part of road stretch E. One of the errors occurring is illustrated in Figure 10: a clip box is created between two adjacent points and the length of the road inside this clip box is added to the cumulative distance. If the road does not lie (entirely) within this box, the distance between two points is incorrect. As seen in the figure the distance between the two points at the crossing is calculated at about 1 meter while the distance along the centerline is approximately 10 meters. These miscalculations lead to errors in the cumulative distances of all the points located there (approx. 25% of the total amount of points of this road stretch). Since the error affects so many points so severely, it was decided to leave them out of the analyses in step e. It does not affect the results of step c.

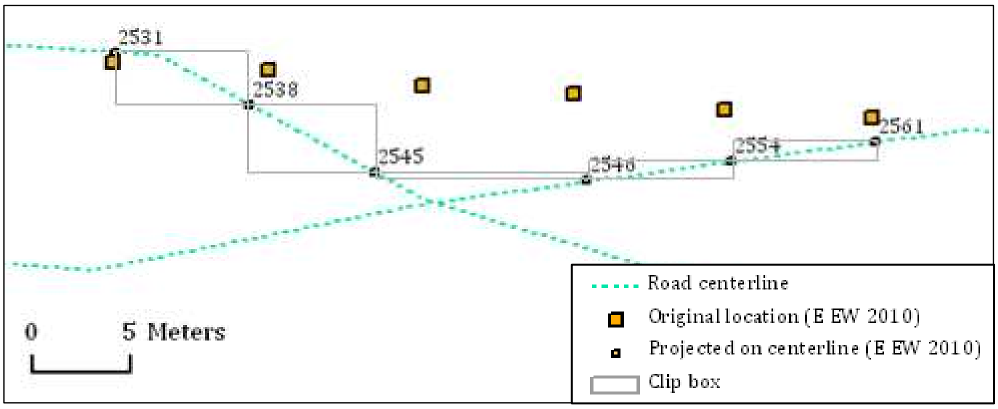


Figure 10: Example of an error in the process of calculating the cumulative distance. The points are labeled with the cumulative distance.

Step c. Analyzing the relationship between road surroundings and positioning error

The low correlation coefficients in Table 2 indicate that there is no significant relationship between the GPS positioning error and the SVF of the reference points. Only road stretches D and E seem to show a slight correlation. (The regression diagrams can be found in Appendix I.)

The positioning error was expected to be higher along the road stretches with dense forests, thus on road stretches B and D. Based on the higher standard deviation and slightly higher mean (compared to the other road stretches) this is indeed the case.

Table 2: Results of the analyses per driving direction

Surroundings	**Road stretch**	**Driving direction**	**R^2 (x = SVF, y = GPS pos. error)**	**Mean GPS pos. error [meters]**	**S.d. GPS pos. error [meters]**
Mixed	C	EW	0.080	2.0	0.061
		WE	0.175	2.4	0.116
Tree rows	A	EW	0.008	2.8	0.189
		WE	0.037	2.8	0.090
	E	EW	0.520	2.4	0.265
		WE	0.315	2.4	0.182
Forest	B	EW	0.032	2.9	0.303
		WE	0.068	2.9	0.234
	D	EW	0.421	2.9	0.514
		WE	0.458	3.3	0.502

Two extremes in Table 2, road stretches A and D, will be looked more closely to find out what causes the large difference in the correlation coefficients and standard deviations.

Road stretch A is surrounded by tree rows along the whole stretch except for a short part in the east (see Figure 11). A small part of road stretch D was selected to see how the GPS positioning error behaves in a situation where the surroundings vary a lot over short distances (see Figure 12). Starting from the West the road is open first, then it is surrounded by dense vegetation, then by tree rows, and in the last part it varies between dense and open.

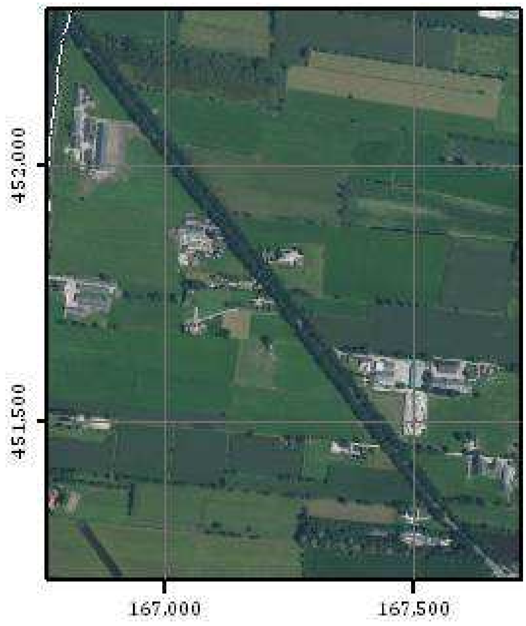


Figure 11: Surroundings of road stretch A

Figure 12: Surroundings of a part of road stretch D with mixed surroundings

The tables below show the relationship between the SVF and the GPS accuracy. The SVF (crosses) is plotted above the GPS accuracy (triangles). Mark that the values of the second axis, indicating the GPS positioning error, is reversed. This makes it easier to see if a high SVF is linked to a high positioning accuracy.

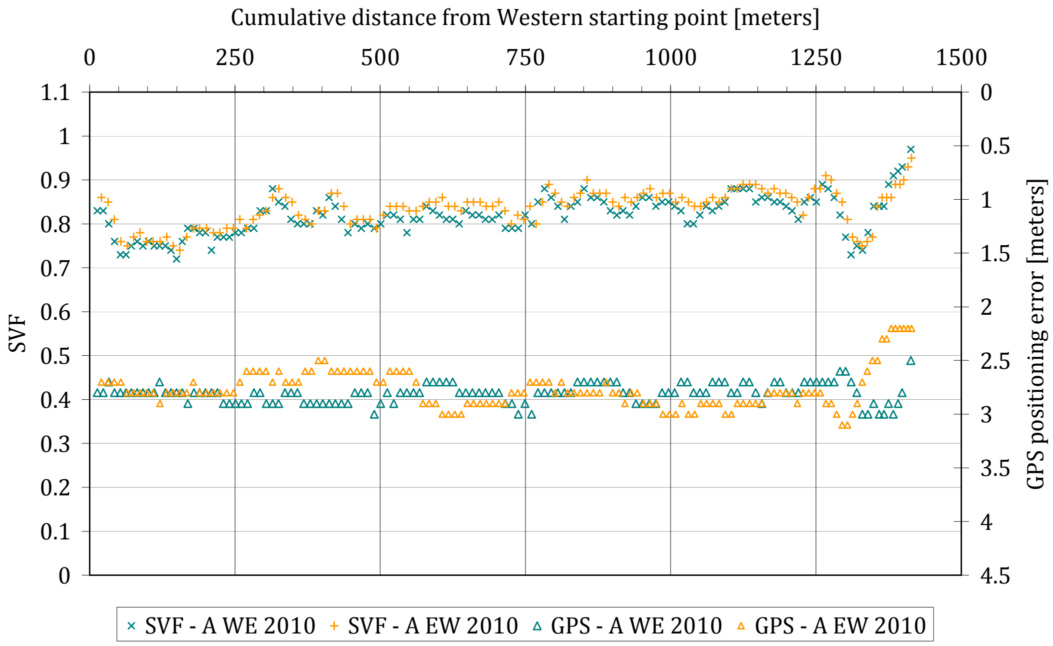


Figure 13: SVF and GPS positioning error of road stretch A plotted according to the cumulative distance

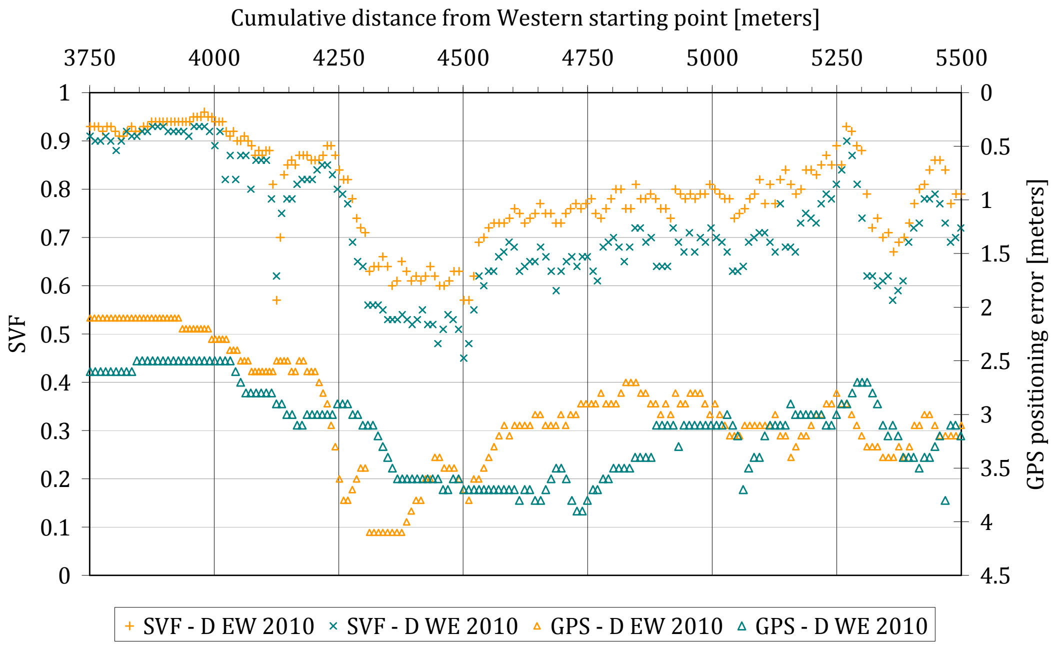


Figure 14: SVF and GPS positioning error of a portion of road stretch D plotted according to the cumulative distance

The change in surroundings at the most eastern, open part of road stretch A is reflected in the SVF and the GPS accuracy of driving direction East-West, but not so much in the GPS accuracy of the other driving direction (see Figure 13). Both data along this road stretch do not provide enough fluctuation to establish a relationship between them.

The data of road stretch D has far more fluctuation in both data types. Figure 14 shows the SVF values and GPS positioning error for this portion of the road stretch. The trend of both data is quite similar when it comes to large changes, but the positioning error can also fluctuate a lot when the SVF does not and vice versa. Furthermore, it is expected that since the SVF in driving direction East-West is higher than the SVF in the other driving direction along the whole stretch, that the GPS accuracy should also be higher; this is not the case.

So, as the correlation coefficient already suggests, the positioning error is related to the SVF in some degree, but not entirely.

Step d. Selecting comparable point sets

In this step point sets are created to enable a comparison between the data of these two points in the next step. When a point set is created, the direction of comparison defines the amount of resulting point sets: for each point in the 'input dataset' the closest point in the 'near dataset' is selected. The amount of points in these two datasets may vary quite a lot for some road stretches (see Table 3).

The input datasets have been chosen by their alphabetical order, except for the inputs for road stretch C, where the common dataset was explicitly chosen to be able to compare the results better. The impact that this 'random' choice of input datasets may have on the results is discussed in Appendix II.

Table 3: Amount of point sets per road stretch per comparison (n)

Comparison nr.	Road stretch	Input dataset	n (Input dataset)	n (Near dataset)
1	C	WE 2010	849	820
2	A	EW 2010	134	134
	B	EW 2010	387	372
	D	EW 2010	830	647
	E	EW 2010	223	221
3	C	WE 2010	849	748
4	E	WE 2007	237	221

Step e. Analyzing the differences in SVF using four comparisons

Overview

This paragraph provides an overview of the results of the comparisons as well as a short explanation on the interpretation of the numbers.

As mentioned before in the 'Comparability' section of the Methodology, the first two comparisons compare the datasets of two different driving directions surveyed in the same year, while the third and forth comparison compare the surveys of different years done in the same lane. Comparisons 1 and 3 focus on the highway, while the other two comparisons are specific for provincial roads.

For every comparison two hypotheses were formulated: a general hypothesis and one for the driving direction. For the general hypothesis we look at the dependability of the SVF difference on the distance to the closest point in a point set (as calculated in step d). For the hypotheses for the driving direction the SVF values of each point set are compared. Correlation coefficients (R^2 values) are used to prove or disprove both these hypotheses: if the correlation coefficient is high the hypotheses can be considered proven.

Table 4: Results of all comparisons

Comparison Nr.	**Driving direction**	**Year**	**Road type**	**Surroundings**	**Road stretch**	**R^2 (distance vs. SVF diff.)**	**R^2 (different direction/year)**	**Mean (SVF diff.)**	**S.d. (SVF diff.)**
1	Opposite	2010	Highway	Mixed	C	0.007	0.539	0.02	0.042
2	Opposite	2010	Provincial	Tree rows	A	0.005	0.780	0.02	0.013
					E	0.008	0.536	0.03	0.020
				Forest	B	0.007	0.182	0.07	0.057
					D	0.003	0.739	0.05	0.046
3	Equal (W-E)	2007 & 2010	Highway	Mixed	C	0.002	0.667	0.01	0.036
4	Equal (W-E)	2007 & 2010	Provincial	Tree rows	E	0.011	0.265	0.03	0.020

In the first R^2 column in Table 4 the correlation coefficients represent the results of the general hypothesis. Here we see that all values are very low, the highest being 0.011 for road stretch E in comparison 4.

In the second R^2 column in Table 4 the results of the hypotheses for the driving direction can be found. All but two road stretches show a high correlation; road stretches B (comparison 2) and E (comparison 4) have very low values (0.182 and 0.265 respectively) while the rest range between 0.539 and 0.780. This phenomenon is looked at more closely in the comparison paragraphs to follow.

The mean and standard deviation (s.d.) of the SVF differences are added to quantify the accuracy of the SVF along the road stretches. The accuracies are needed to know how accurate the GIS model needs to be. The accuracy is expected to vary for different kinds of surroundings. We can see this in Table 4 where the accuracy is lower along roads surrounded by forest and very high along the highway (comparison 3) for example.

Each comparison will be dealt with in detail in the following paragraphs where each road stretch will be looked at separately as well as in relation to other road stretches. The aim is to find an explanation for the values represented above, looking at trends and errors in the data. The hypotheses and the assisting questions (formulated on page 10) will be used as guidelines so that the first research question can be answered in the end.

Comparison 1: Highway – Opposite lanes

This section compares reference data from the opposite driving directions of road stretch C (highway A12) surveyed in 2010. Due to the fact that the centerlines of the opposite directions of the highway lie quite far apart, a large difference between the SVF of the comparable points is expected here.

Before looking at the results of the analyses, a short explanation is required to interpret the figures that will be presented.

At least two graphs are presented for every road stretch. The first shows the relationship between the distance between two points in a point set and their SVF difference. This figure is used to prove or disprove the general hypothesis: if the hypothesis is valid the difference in SVF will increase as the distance increases.

The second graph compares the SVF of the point sets to see if there is a difference between them. If the SVF does not differ significantly between the reference datasets, the point cloud should be narrow and diagonal. This graph is used to prove or disprove the second and third hypotheses.

In both graphs a linear trend line is added, accompanied by its equation and the correlation coefficient. If the expected linear relationship between the datasets is present, the R^2 value will be close to 1 and in the equation $y = ax + b$ value a will be close to 1 and b close to 0.

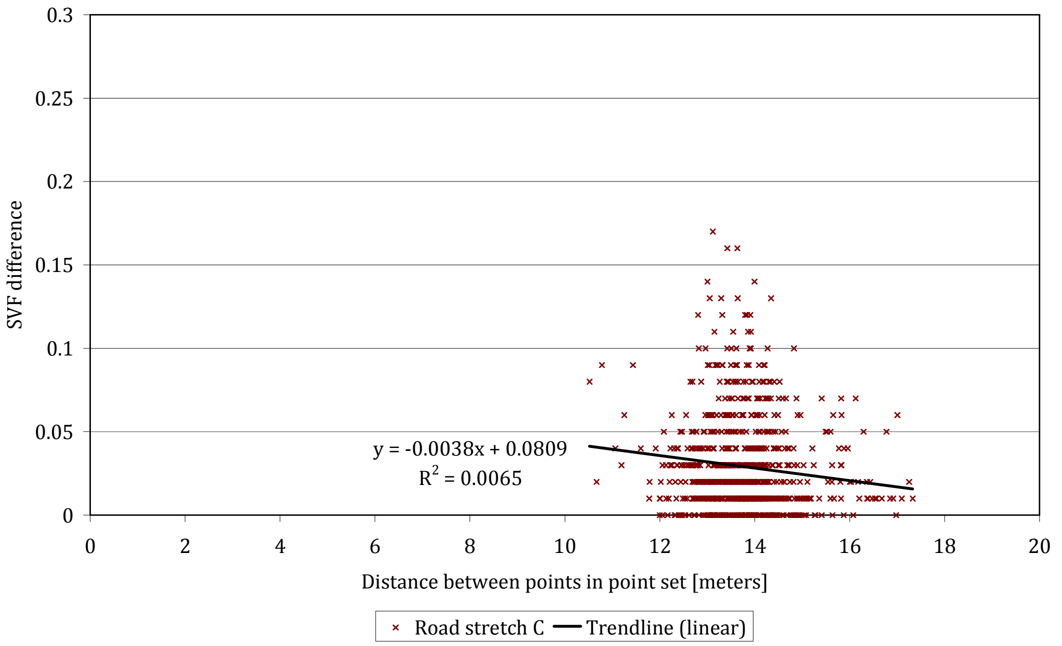


Figure 15: Road stretch C: Distance vs. difference in SVF between points in point sets. (Two points lie beyond the maximum SVF difference; these are marked as point set 1 and 2 in the figure below.)

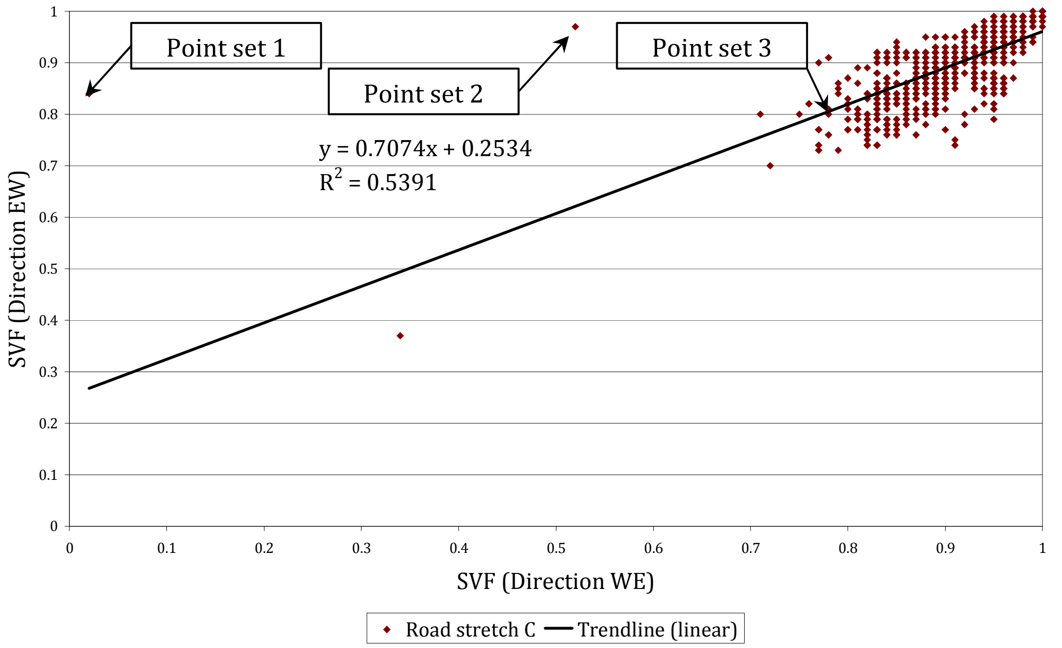
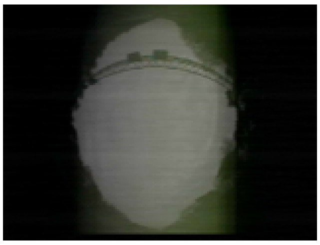


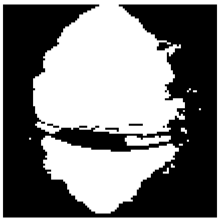
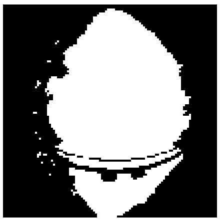
Figure 16: Road stretch C: Regression diagram comparing the SVF of point sets

According to Figure 15 there seems to be no relationship between the distance between the points and their SVF.

The large part of the points is very similar in SVF even though they lie on opposite sides of the highway (see Figure 16). Two point sets (marked 1 and 2) show a very high difference in their SVF. To see what causes this, it is necessary to observe the original fish-eye photographs in the table below (Table 5). A third point set is added as an example to show the photographs of a 'normal', non-deviating, point set looks like. The binary images (on which the SVF calculation is based) are added to see if the difference is caused by an incorrectly chosen threshold value.

Table 5: Photographs along road stretch C for point sets indicated in Figure 16. The fish-eye photographs in the second column (taken in the opposite driving direction) are rotated 180 degrees to facilitate comparison between the photographs.

Point set	**Driving direction WE**	**Driving direction EW**	**Remark**
1	  Image 1624 SVF = 0.02	  Image 10899 SVF = 0.84	Incorrect processing: the threshold for delineating sky and non-sky pixels is not set correctly.
2	  Image 1298 SVF = 0.52	  Image 11218 SVF = 0.97	Large change in surroundings over a short distance. (Distance between points = 3.7 m, but appears to be more.)

Point set	Driving direction WE	Driving direction EW	Remark
3	  Image 2143 SVF = 0.78	  Image 10402 SVF = 0.80	The SVF is very similar even though surroundings are not exactly the same.

As seen in the table above, incorrect processing can cause a very large error in the SVF. When the surroundings change a lot over a short distance, the SVF difference can be very high between two photographs.

When the two large outliers, point sets 1 and 2, are removed from the dataset, the correlation is much higher (see Table 6 and Figure 17). The standard deviation also shows a large improvement.

Table 6: Results of the analyses of comparison 1 including a dataset without large errors

Road stretch	Surroundings	R^2 (x = distance, y = SVF diff.)	R^2 (x = dir. EW, y = dir. WE)	Mean (SVF diff.)	S.d. (SVF diff.)
C	Mixed	0.007	0.539	0.02	0.042
C, excl. point sets 1 and 2	Mixed	0.004	0.695	0.02	0.028

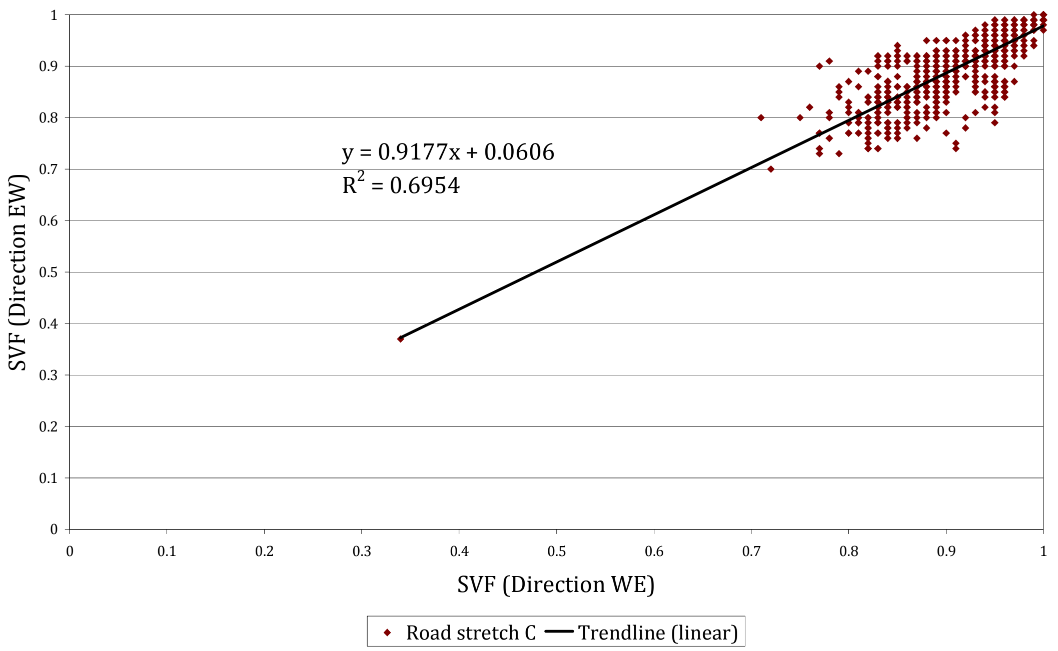


Figure 17: Regression diagram without deviating point sets

Comparison 2: Provincial road – Opposite lanes

Comparison 2 compares the reference datasets of provincial roads surveyed in both driving directions. There are four road stretches available for this comparison: two surrounded by tree rows (A and E) and two mostly surrounded by forest (B and D).

For this comparison the difference in SVF is expected to be lower than for comparison 1, as the lanes of a provincial road lie closer together than for a highway.

These results are presented in Table 7. Again it is clear that there is no correlation between the distance and the difference in SVF.

Table 7: Results of the analyses of comparison 2

Road stretch	Surroundings	R^2 (x = distance, y = SVF diff.)	R^2 (x = dir. EW, y = dir. WE)	Mean (SVF diff.)	S.d. (SVF diff.)
A	Tree rows	0.005	0.780	0.02	0.013
E	Tree rows	0.008	0.536	0.03	0.020
B	Forest	0.007	0.182	0.07	0.057
D	Forest	0.003	0.739	0.05	0.046

A surprising number is the correlation coefficient for road stretch B concerning the SVF of the two driving directions: it is very low compared to the other stretches. This issue is discussed later.

The mean and standard deviation of the point sets are significantly larger for the roads surrounded by forest (B and D) than for roads surrounded by tree rows (A and E). Therefore these two ‘situations’ are dealt with separately in the rest of the paragraph.

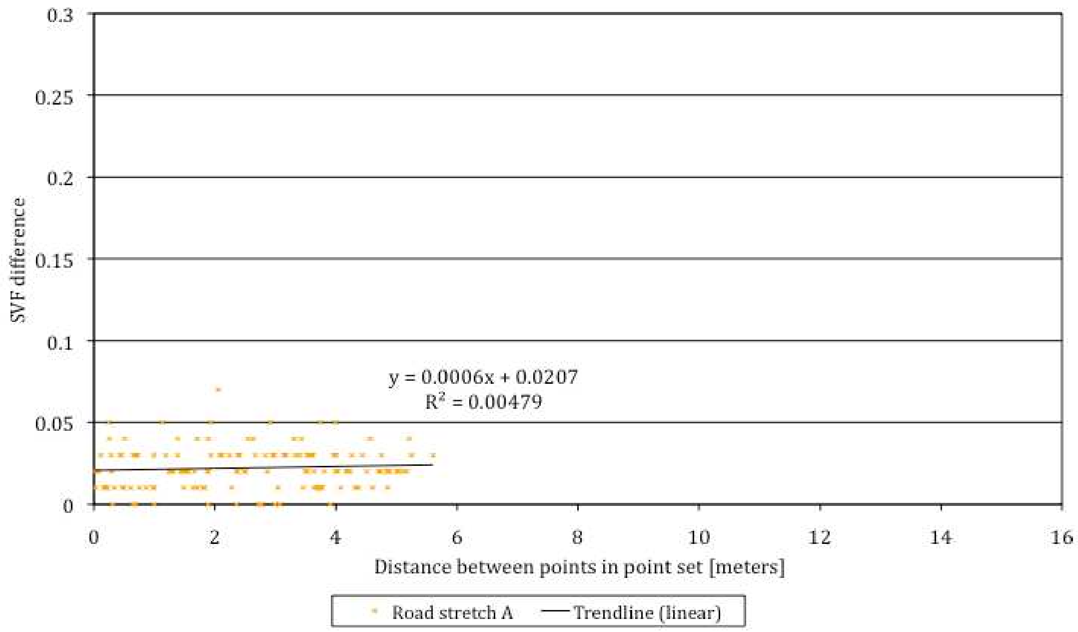


Figure 18: Road stretch A: Distance vs. difference in SVF between points in point sets

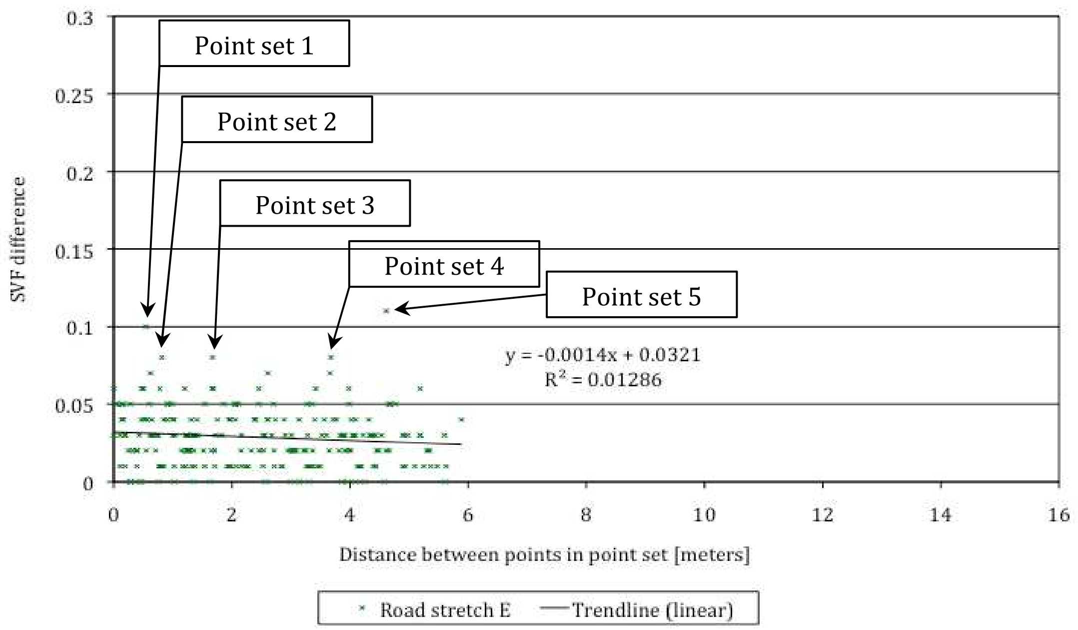
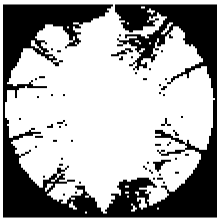
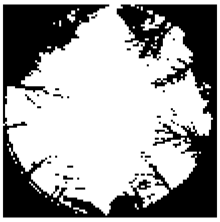
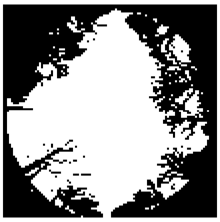
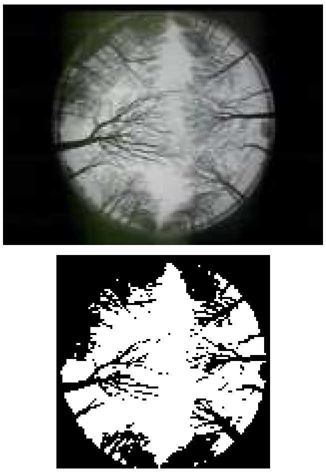


Figure 19: Road stretch E: Distance vs. difference in SVF between points in point sets

As seen in the both figures 18 and 19, there is slight difference in SVF for the two road stretches surrounded by tree rows: nearly all values lie below 0.05. For road stretch E the differences are higher. These differences are examined in detail by looking at the fish-eye photographs of the five largest outliers (see Table 8).

Table 8: Point set photographs of road stretch E as marked in Figure 19. The fish-eye photographs in the second column (taken in the opposite driving direction) are rotated 180 degrees to facilitate comparison between the photographs.

Point set	Driving direction EW	Driving direction WE	Remark
1	  Image 7210 SVF =0.92	  Image 8619 SVF = 0.82	The distance between the points is only 0.55m but it makes a lot of difference in what is captured. The light conditions are also different, enhancing the SVF difference in driving direction EW.
2	  Image 7356 SVF = 0.92	  Image 8475 SVF = 0.84	The opposite roadside seems denser in image 8475 than from close by in image 7356. (Distance = 0.81m)

Point set	Driving direction EW	Driving direction WE	Remark
3	 Image 7200 SVF = 0.88	 Image 8628 SVF = 0.80	The surroundings seem denser in image 8628 which is probably caused by a slight change in light conditions. (Distance = 1.68 m)
4	 Image 7292 SVF = 0.86	 Image 8534 SVF = 0.94	A distance of 3.68m makes a lot of difference in what is captured.
5	 Image 7205 SVF = 0.90	 Image 8624 SVF = 0.79	The distance between the points is 4.61m and the surroundings change significantly over that distance.

From the images above we can conclude that the surroundings in road stretch E are not homogeneous: they can change a lot over a short distance. The brightness of the photos in driving direction EW is structurally higher, but the homogeneous surroundings drown out this effect. This can also be seen in the regression diagram of the road stretch (Figure 21).

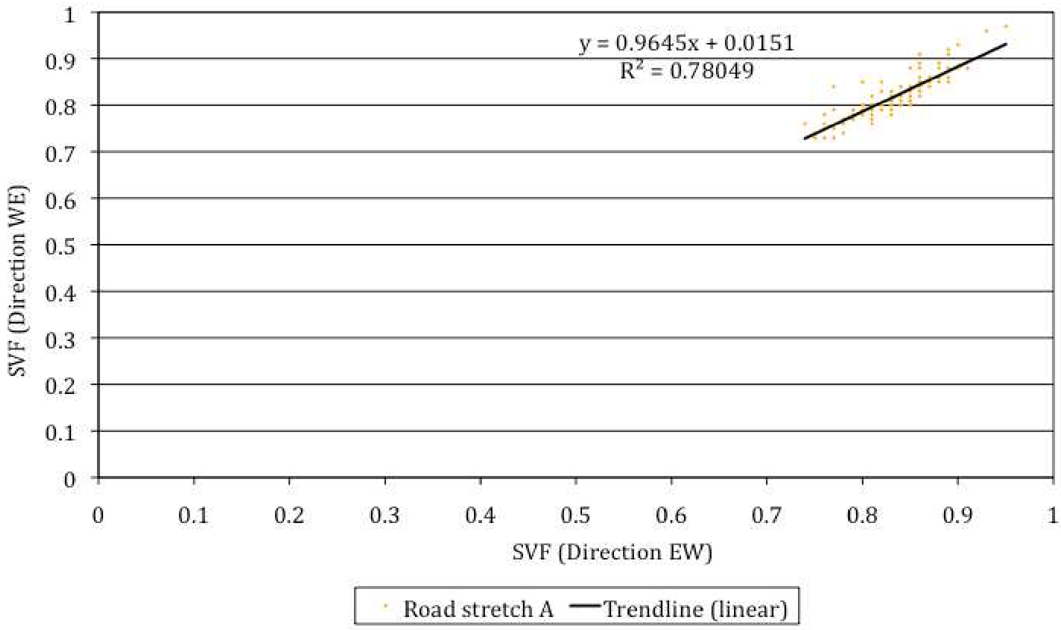


Figure 20: Road stretch A: Regression diagram comparing the SVF of point sets

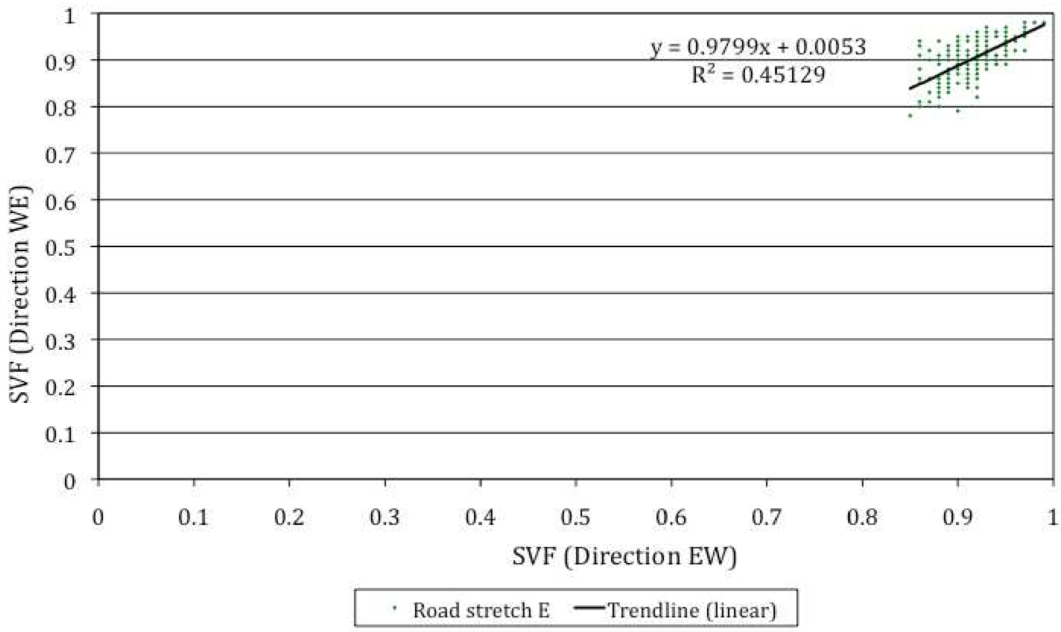


Figure 21: Road stretch E: Regression diagram comparing the SVF of point sets

The regression diagrams above show two striking differences. Firstly, the reference points of road stretch A have a very high correlation compared to E, suggesting very homogeneous surroundings (see Table 7 and Figure 20). Secondly, considering that both road stretches A and E are tree rows, it is striking that the SVF values of road stretch E are much higher than for road stretch A (see Figure 20 and Figure 21 above). Aerial photographs provide an explanation for both findings.

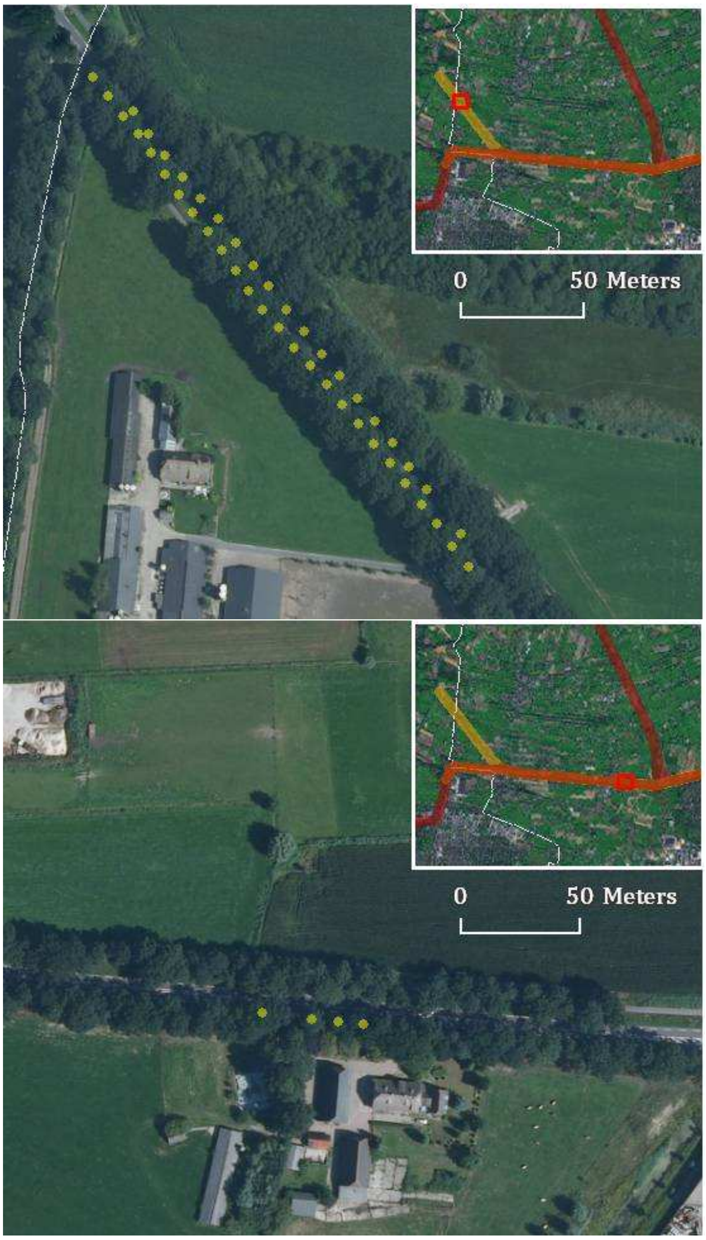


Figure 22: Parts of road stretches A (top) and E (bottom) containing photo locations where the SVF <= 0.80 (marked yellow) to see what causes the low SVF.

Figure 22 shows a canopy difference above the roads. Road stretch A has more homogeneous tree rows closer to the road which explains why the SVF values for this road stretch are lower and more similar to each other. The heterogeneous surroundings of road stretch E cause large differences in SVF, even though the compared points lie fairly close together.

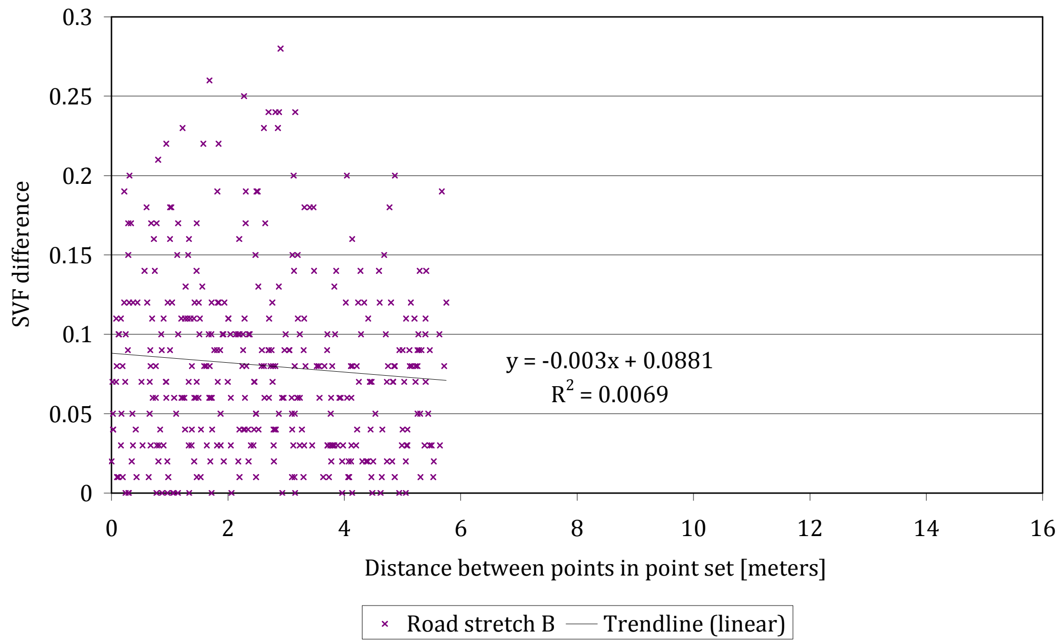


Figure 23: Road stretch B: Distance vs. difference in SVF between points in point sets

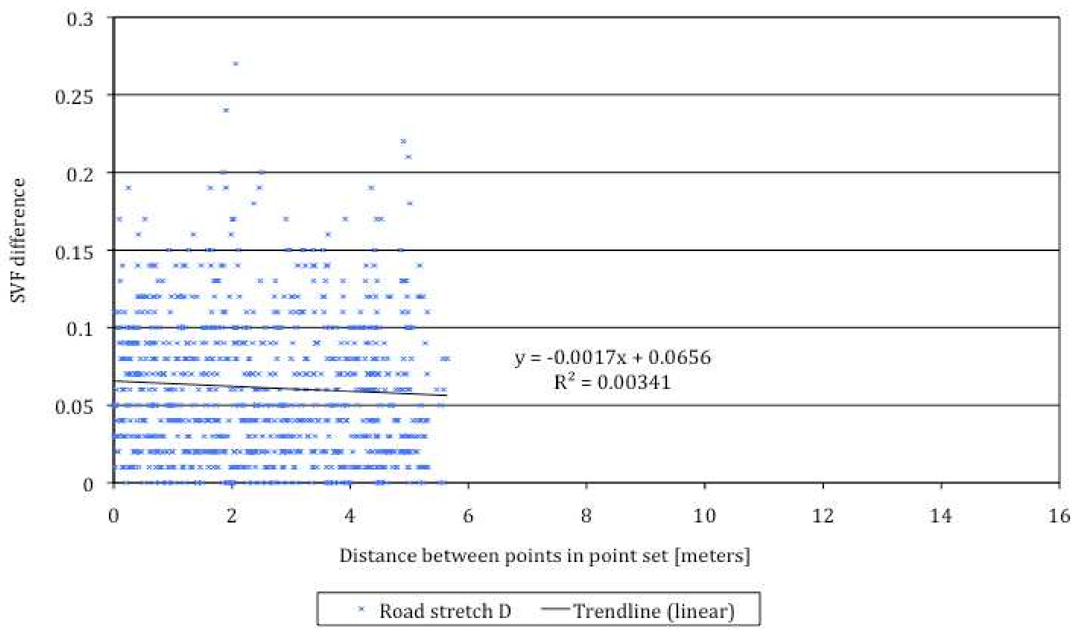
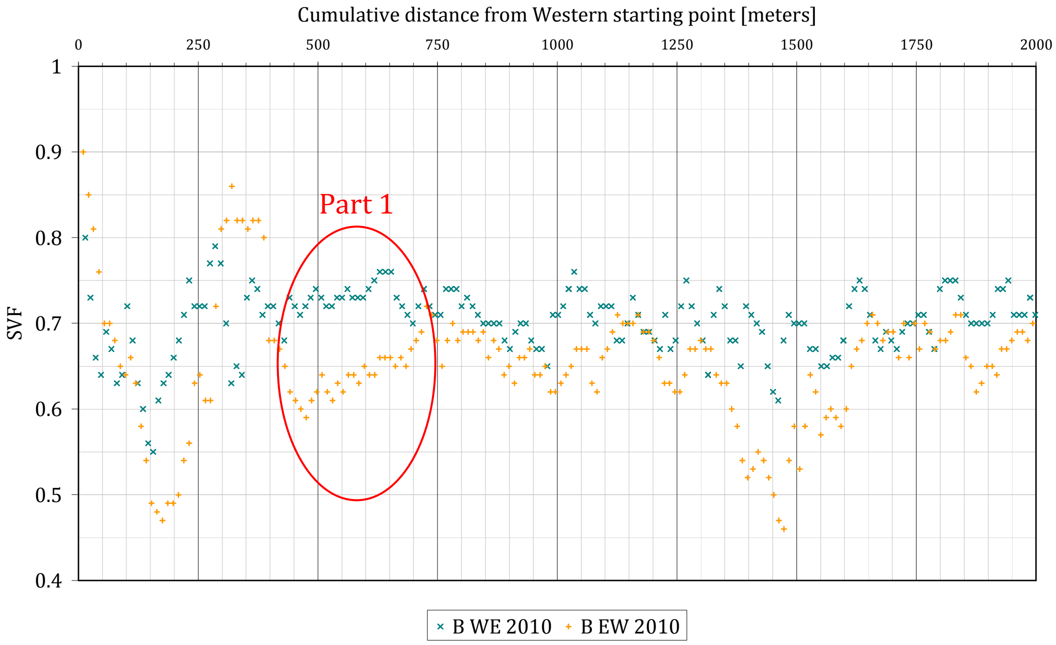


Figure 24: Road stretch D: Distance vs. difference in SVF between points in point sets

Roads surrounded by forest have very high SVF differences regardless of the distance, as shown in Figure 23 and Figure 24. Road stretch B even has several point sets with SVF differences higher than 0.2.

When looking at the SVF plots of both datasets more closely in Figure 25 below, the same trends in SVF are apparent; the SVF differences are very high. The values for direction WE are systematically higher (with some exceptions), suggesting that the northern side of the road provides more cover.

To validate this assumption four parts of the road stretch (as marked in Figure 25) are accompanied by an aerial photograph and fish-eye photographs. These parts are chosen due to their high SVF difference over a longer stretch. Part 4 is indicated because the trend is reversed here; the south side should provide more cover.



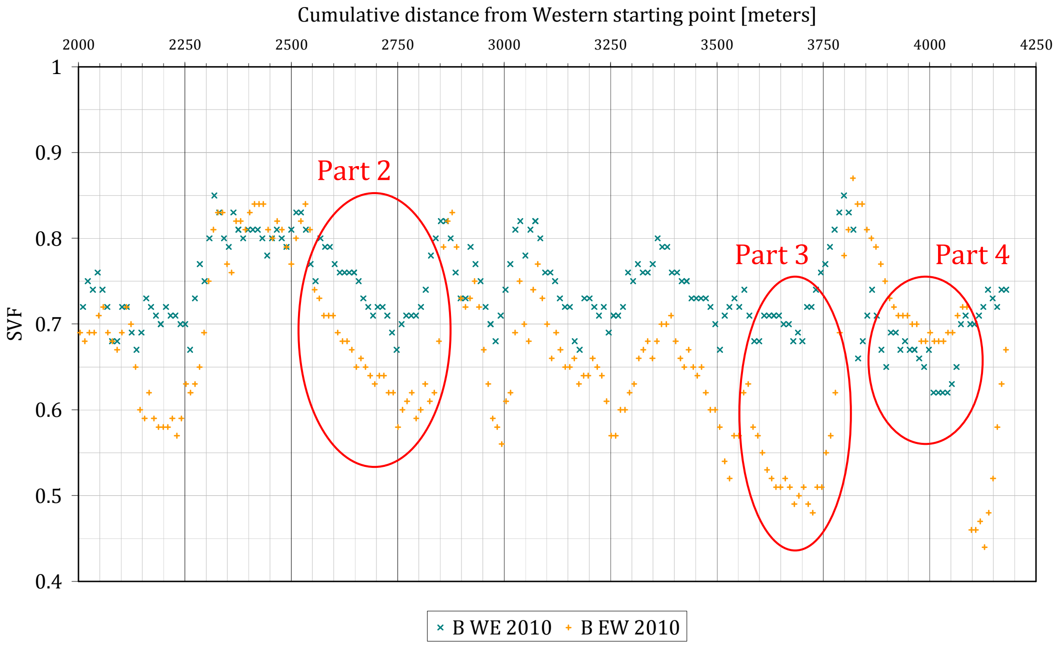


Figure 25: SVF for road stretch B plotted based on the cumulative distance from a common Western starting point. The aerial photographs of the four marked parts are included in Figure 26 through Figure 29.

For parts 1 and 2, the aerial photographs suggest that the north side is much more open than the south side, while they indicate the opposite for parts 3 and 4. The fish-eye photographs of one point set in each part are compared in Table 9, providing a better impression of each situation.

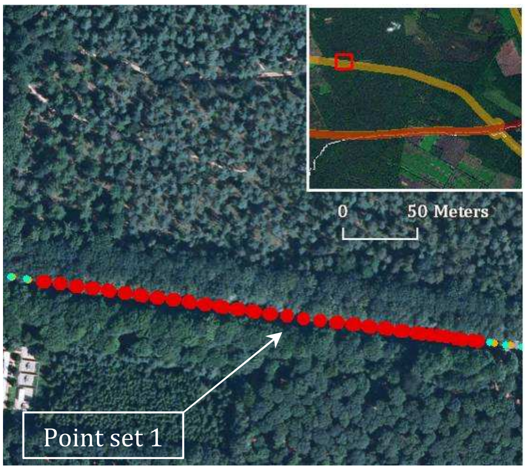


Figure 26: Part 1 – Cumulative distance between 450 and 750

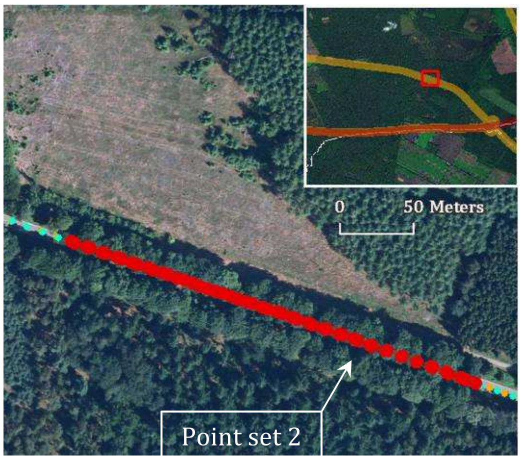


Figure 27: Part 2 – Cumulative distance between 2550 and 2850

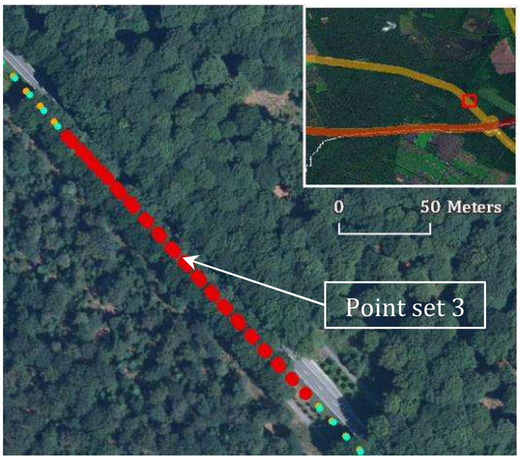


Figure 28: Part 3 – Cumulative distance between 3600 and 3800

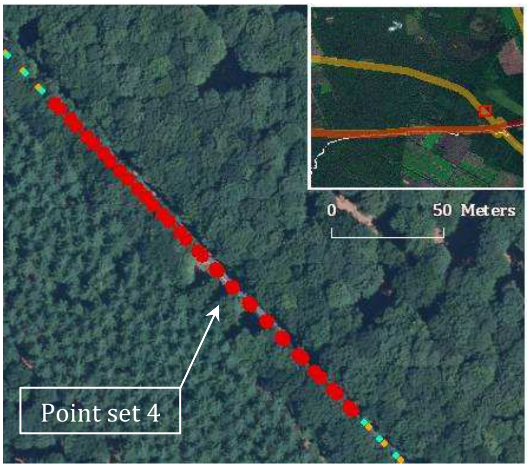
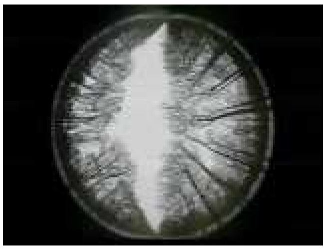
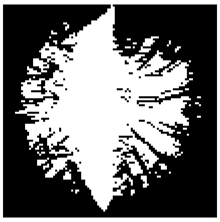
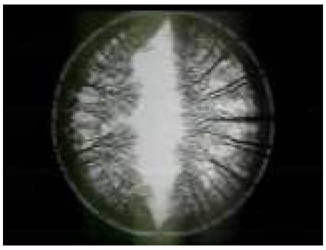
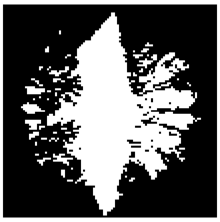
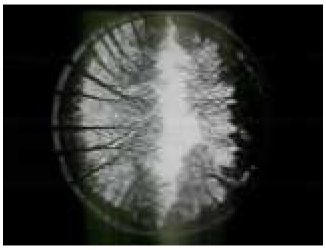
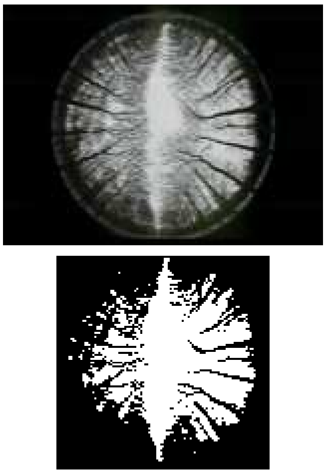
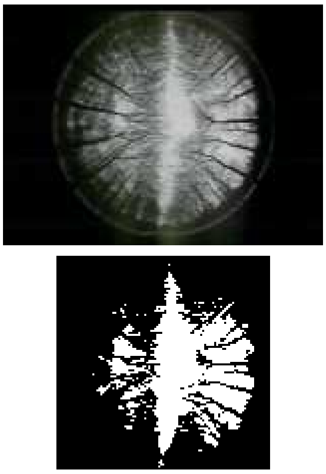
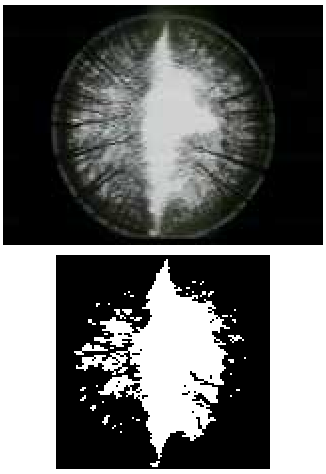
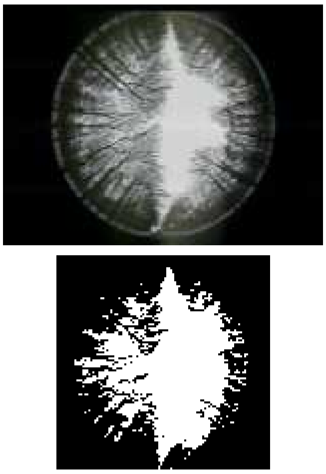


Figure 29: Part 4 – Cumulative distance between 3900 and 4100

Table 9: Point set photographs along road stretch B as indicated in Figure 26 through Figure 29. The fish-eye photographs in the second column (taken in the opposite driving direction) are rotated 180 degrees to facilitate comparison between the photographs.

Point set	**Driving direction WE**	**Driving direction EW**	**Remark**
1	  Image 9937 SVF = 0.75	  Image 5895 SVF = 0.64	The fish-eye photos indicate that the north side is more open than the south side, but the trees situated there obscure the sky more than those on the other side, reversing the effect. The light conditions are slightly different as well, enhancing the difference.
2	  Image 10127 SVF = 0.70	  Image 5697 SVF = 0.60	In part 2 the photos support the fact that the north side provides more cover. Again, the light conditions enhance the effect, lowering the SVF even more in driving direction EW.

Point set	Driving direction WE	Driving direction EW	Remark
3	 Image 10208 SVF = 0.69	 Image 5610 SVF = 0.50	Same situation as point set/part 2, but the photos are taken at nearly exactly the same location, eliminating the issue of the surroundings being different.
4	 Image 10238 SVF = 0.62	 Image 5579 SVF = 0.68	The difference in SVF seems to be caused by a different threshold, as the light conditions appear the same in the color photographs. The fact that the photos are not taken at the same location does not affect the outcome much, as the binary photographs show.

The large SVF differences over a longer distance are caused by the unequal road surroundings on either side of the road. As seen in the table above, the light conditions are different for the three first point sets. This is probably caused by a slight change in the weather: the diaphragm of the camera adjusts automatically to changes in light. If this is not spotted during processing (as is the case here) it causes a systematic error.

The regression diagram below (Figure 30) gives the impression that these errors have a large impact on the correlation coefficient. The point cloud is barely diagonal and very dispersed.

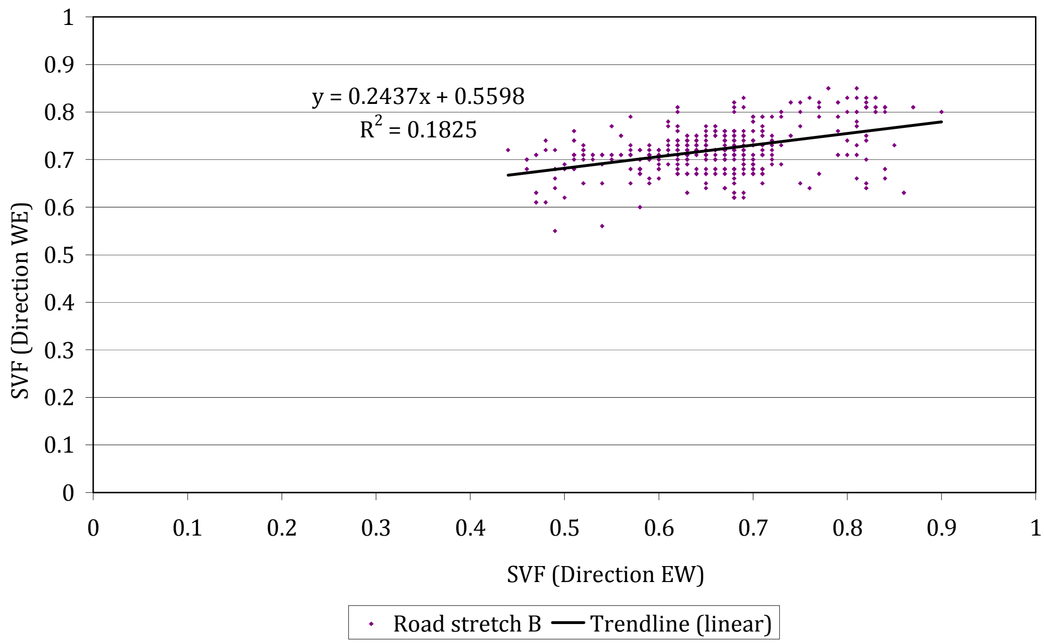


Figure 30: Road stretch B: Regression diagram comparing the SVF of point sets

The correlation between the point sets in road stretch D is much higher (see Figure 31). The point cloud even shows a pattern: narrow at high values and more dispersed as the SVF decreases.

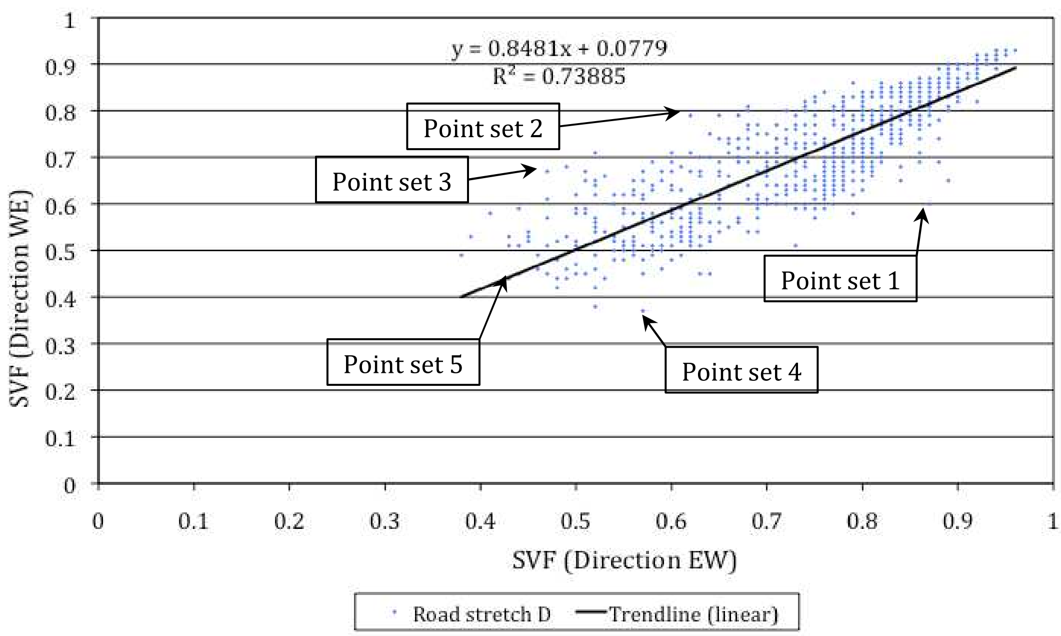
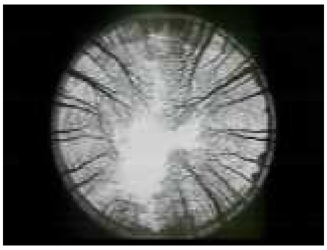
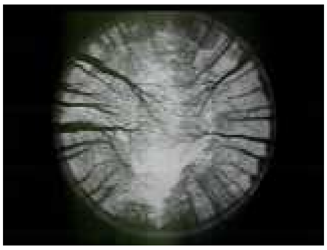
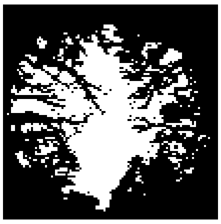
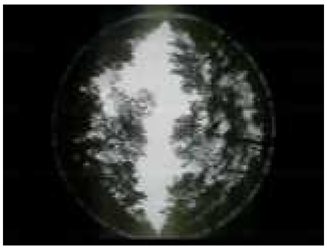
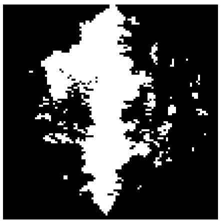
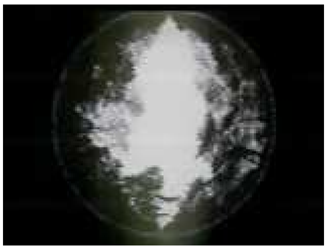
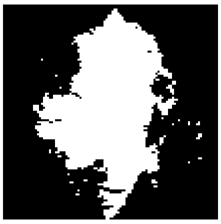
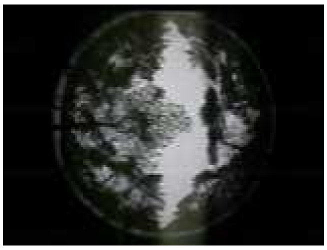
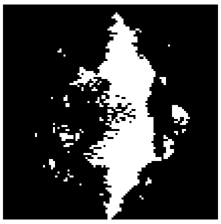
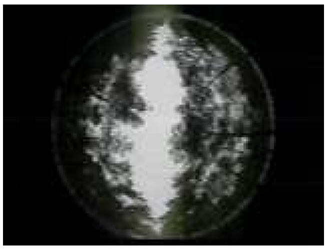
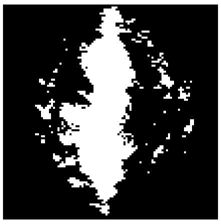
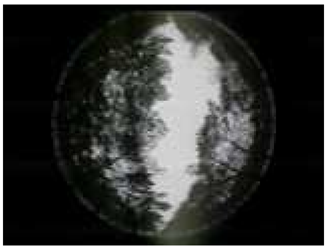


Figure 31: Road stretch D: Regression diagram comparing the SVF of point sets

Four outliers are chosen to see what causes these large deviations for this road stretch. A fifth point set is added as an example, to see how the photographs of a point set with low but equal SVF values look like.

Table 10: Point set photographs along road stretch B as indicated in Figure 31. The fish-eye photographs in the second column (taken in the opposite driving direction) are rotated 180 degrees to facilitate comparison between the photographs.

Point set	Driving direction EW	Driving direction WE	Remark
1	  Image 2884 SVF = 0.87	  Image 4828 SVF = 0.60	The tree row becomes much denser within a distance of 2.1 m. The light conditions also seem slightly different.
2	  Image 3170 SVF = 0.62	  Image 4588 SVF = 0.79	The content of the photographs is very different even though the distance between the points is indicated to be just 0.1 m. The GPS positioning error plays a role here.

Point set	Driving direction EW	Driving direction WE	Remark
3	  Image 3265 SVF = 0.47	  Image 4552 SVF = 0.67	Same situation as point set 2: the distance between the points is not that large (2.5 m), but the content of the photos is very different: a large GPS positioning error is probably the cause.
4	  Image 3552 SVF = 0.57	  Image 4327 SVF = 0.37	Again a large difference in the content is seen, so a distance of 1.9 m between the points seems to be incorrect: a positioning error accompanied by heterogeneous surroundings cause a large difference in SVF. The light conditions also seem different, lowering the SVF even more in photo 4327.
5	  Image 3287 SVF = 0.43	  Image 4508 SVF = 0.44	The captured surroundings are not the same, but they are still very similar thus yield a similar SVF. (Distance = 5.2 m.)

Based on the findings in the table above it seems that the large differences in SVF along this road stretch are mainly caused by heterogeneous surroundings in combination with a large positioning error. Different light conditions play a small role.

<u>Comparison 3: Highway – Same lane</u>

In this comparison the surveys of 2007 and 2010 along road stretch C (highway A12) are compared in driving direction West-East. The SVF values of the compared points are expected to be much more similar than in comparison 1 since the points lie much closer together. Reconstruction of the highway carried out in 2008 can contribute to differences in SVF.

Table 11: Results of the analyses of comparison 3

Road stretch	Surroundings	R^2 (x = distance, y = SVF diff.)	R^2 (x = 2010, y = 2007)	Mean (SVF diff.)	S.d. (SVF diff.)
C	Mixed	0.002	0.667	0.01	0.036

As can be seen in the table above, the correlation between the distance and the difference in SVF between the point sets is extremely low.

As expected, there is a higher correlation between the SVF in the same lane than between the SVF of opposite lanes; the correlation between the SVF in 2007 and 2010 is higher than the result for comparison 1 (see Table 6).

The mean and the standard deviation of the SVF difference are very low. Compared to comparison 1, the mean is just half as high, following the expectation that there is more similarity between the SVF in the same lane as opposed to in opposite lanes, surveyed in the same year.

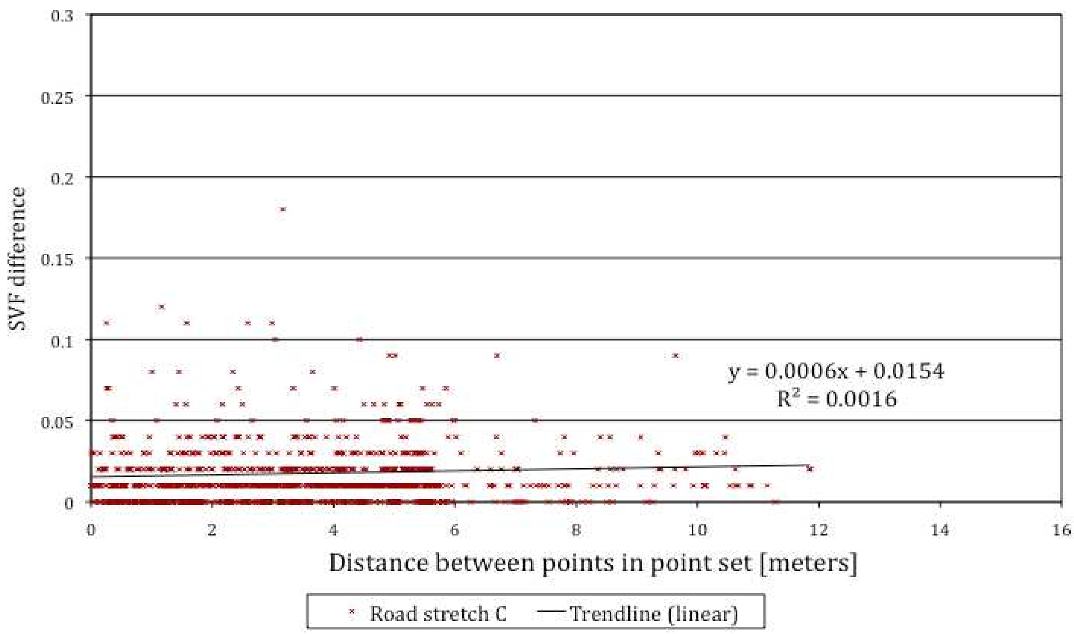


Figure 32: Road stretch C: Distance vs. difference in SVF between points in point sets. The large distances between the points are caused by the absence of erroneous photographs in the 2007 dataset. (Two values fall outside the maximum range set here: marked as point set 1 and 3 in next figure.)

Figure 32 reveals no relationship between the distance between the point sets and their SVF, even over larger distances. The regression diagram below (Figure 33) shows a very high correlation between the two datasets; the point cloud is very narrow and diagonal.

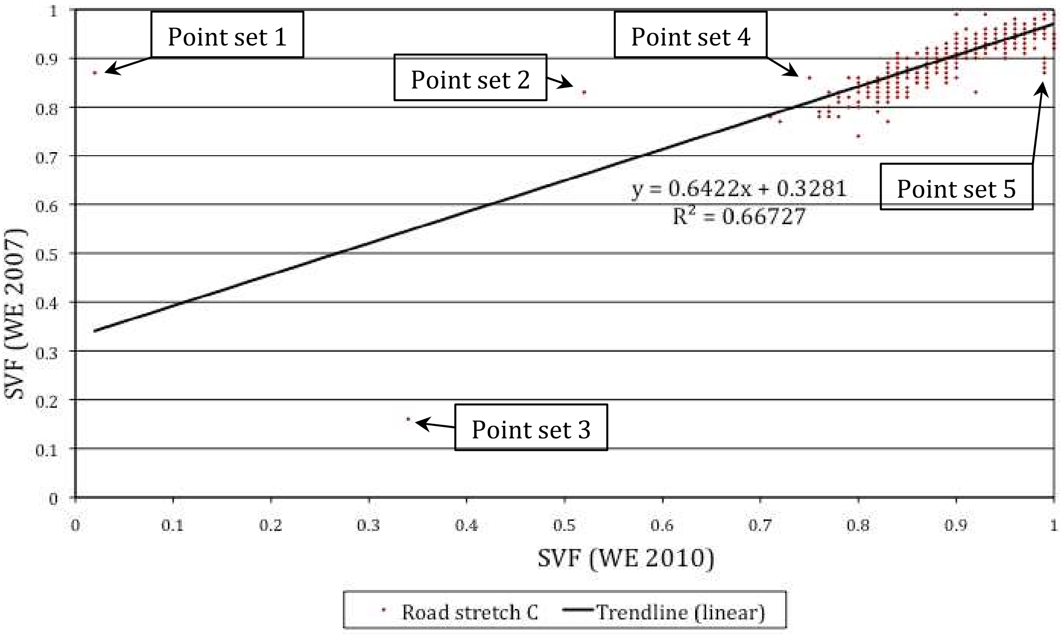


Figure 33: Road stretch C: Regression diagram comparing the SVF of point sets

The deviation in point sets 1 and 2 are identical to the ones dealt with in comparison 1, which were caused by a processing error and rapidly changing surroundings (an overpass). The same overpass influences point set 3. Therefore only the photographs of point sets 4 and 5 will be looked at more closely, in the table below.

Table 12: Photographs for point sets 4 and 5 as marked in Figure 33. A comparison using binary photographs is not possible here, as these were not stored during the 2007 processing section.

Point set	2007	2010	Remark
4	 Image 3629 SVF = 0.86	 Image 1999 SVF = 0.75	The photo location is nearly the same, but the light conditions are different. The possible effect of the tree's growth over 3 years cannot be judged as the photos are not taken at exactly the same location. The resolution may also be of some influence, but this is not analyzed further.

Point set	2007	2010	Remark
5	 Image 3026 SVF = 0.87	 Image 1318 SVF = 0.99	The reconstruction of the highway causes a large difference in the SVF here.

The changes caused by the reconstruction can easily be seen in the SVF plot accompanied by aerial photographs, as displayed below.

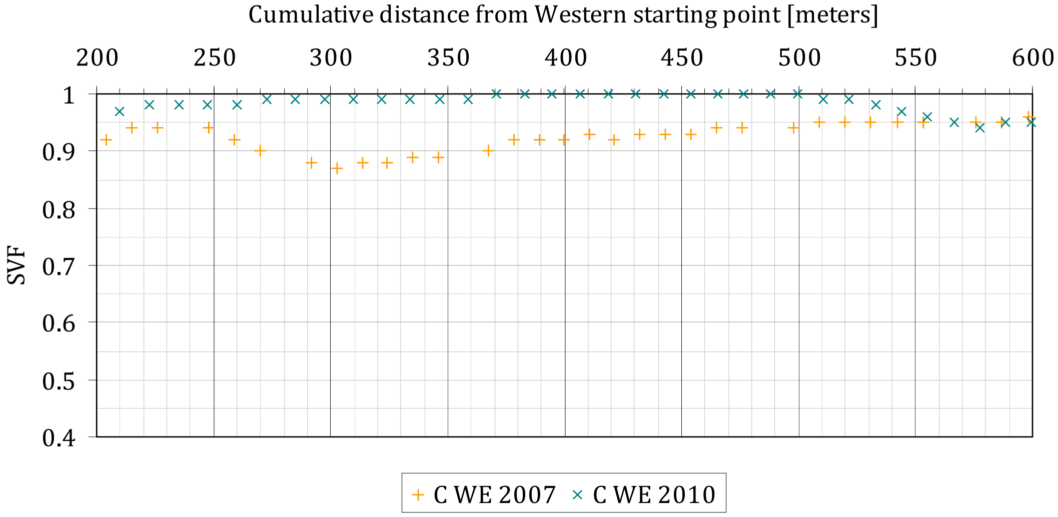


Figure 34: The SVF of the point sets surveyed before and after the construction in 2008

Figure 35: Aerial photographs from 2008, showing where construction takes place

The removal of the point sets containing errors (i.e. point sets 1 through 4) leads to a large improvement in the results. This is seen in the table and regression diagram below (Table 13 and Figure 36).

Table 13: Results of comparison 3 including and excluding outliers

Road stretch	Surroundings	R^2 (x = distance, y = SVF diff.)	R^2 (x = 2010, y = 2007)	Mean (SVF diff.)	S.d. (SVF diff.)
C	Mixed	0.002	0.667	0.01	0.036
C (excluding point sets 1 through 4)	Mixed	0.003	0.828	0.01	0.018

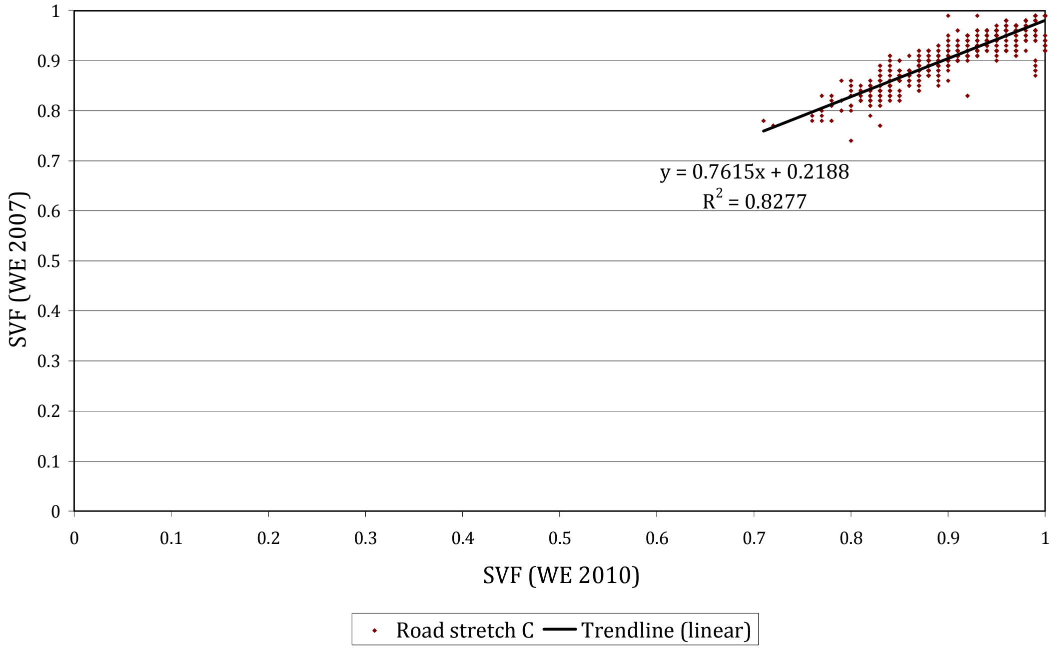


Figure 36: Regression diagram for road stretch C excluding point sets 1 through 4

<u>Comparison 4: Provincial road – Same lane</u>

In this comparison reference data from 2007 and 2010 are compared for road stretch E: a provincial road mostly surrounded by tree rows. The differences in SVF are expected to be lower (thus better) than for comparison 2 since the data is collected in the same lane.

Table 14: Results of comparison 4

Road stretch	Surroundings	R^2 (x = distance, y = SVF diff.)	R^2 (x = 2010, y = 2007)	Mean (SVF diff.)	S.d. (SVF diff.)
E	Tree rows	0.011	0.265	0.03	0.020

The correlation between the SVF of the different years is very low (see Table 14 and Figure 37), even lower than when comparing two lanes in the same year (R^2 = 0.536). The regression diagram shows a round point cloud with a slight diagonal tendency (see Figure 37) reflecting the effect of the heterogeneous surroundings along this road stretch.

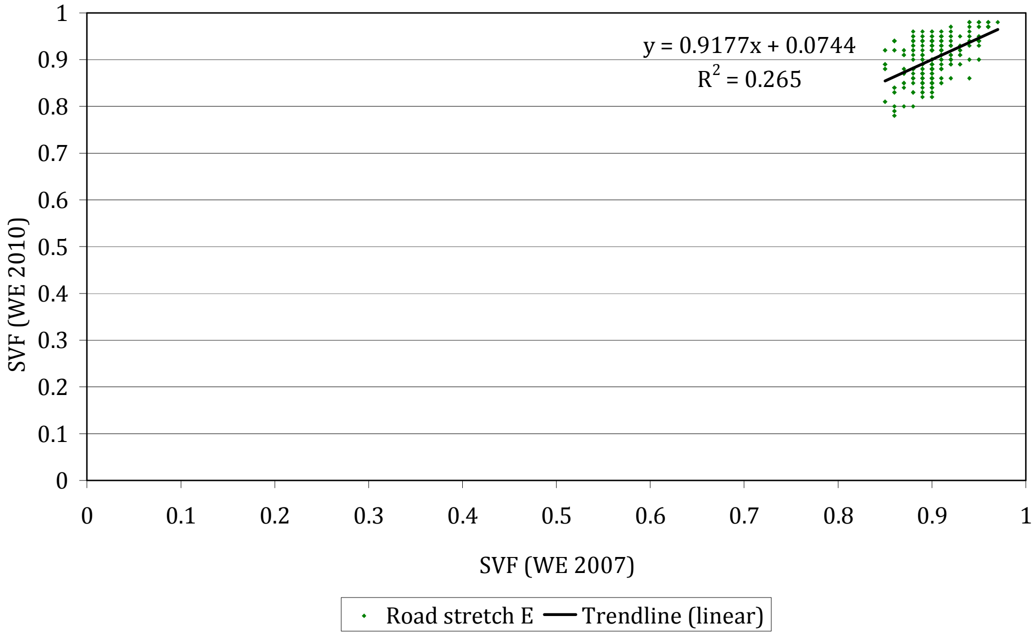


Figure 37: Road stretch E: Regression diagram comparing the SVF of point sets

There are many high SVF differences in this comparison. These lead to a lower correlation, but apparently do not affect the mean and standard deviation; these are exactly the same as in comparison 2. Four point sets with a large SVF difference are looked at more closely (in Table 15) to find a cause for these high SVF differences.

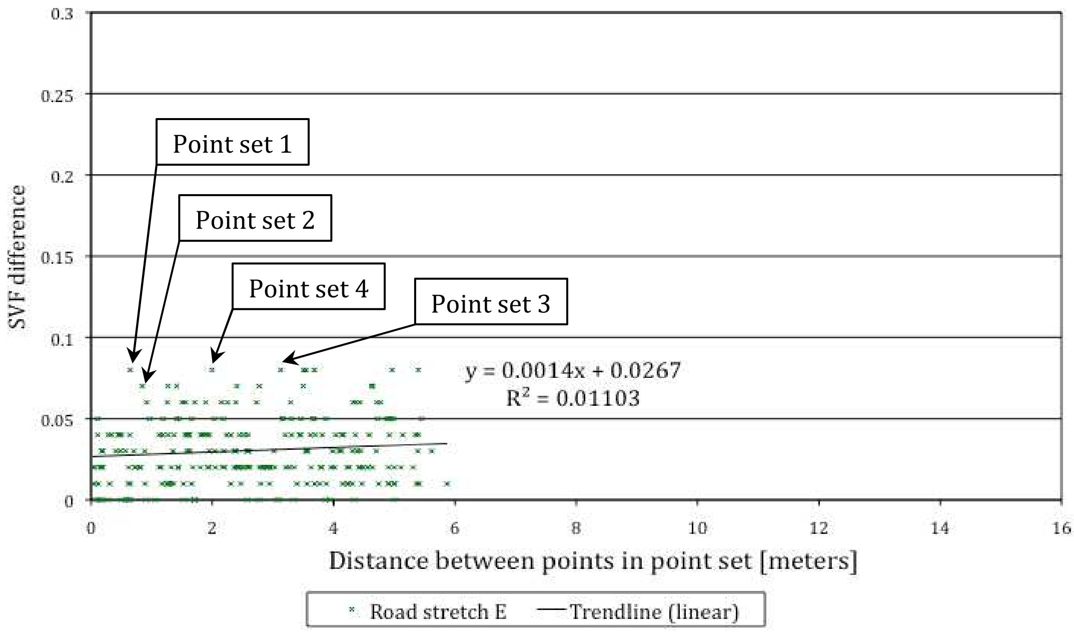


Figure 38: Road stretch E: Distance vs. difference in SVF between points in point sets

Table 15: Photographs belonging to point sets marked in Figure 38. A comparison using binary photographs is not possible here, as these were not stored during the 2007 processing section.

Point set	**2007**	**2010**	**Remark**
1	 Image 1490 SVF = 0.86	 Image 8578 SVF = 0.94	Distance between the points is just 0.65m, but the content on the photos is very different; the GPS positioning error plays a role here.
2	 Image 1387 SVF = 0.85	 Image 8485 SVF = 0.92	Again the distance between the points is very low (0.85m), but the content suggests that this distance is larger.
3	 Image 1386 SVF = 0.86	 Image 8484 SVF = 0.94	The captured location in image 1386 is more similar to image 8485 than 8484; the difference between these SVF values is 0.06 instead of 0.08.
4	 Image 1315 SVF = 0.88	 Image 8420 SVF = 0.96	The distance between the points is just 2.00m, but it seems larger. The different positioning has a large impact on the SVF.

As seen in the photographs above, the four outliers are caused by an inaccurate GPS position on one part and heterogeneous surroundings on the other. There are two systematic differences between the surveys, namely the cloud cover and the resolution of the pictures. Since the binary images of 2007 are not available to see how big of an impact this may have, these possible influences cannot be taken into consideration.

2.4 Conclusions and recommendations

As mentioned in the main introduction, the objective of this research is to create a GIS model that produces sky view factors of a quality that matches up to the photographic method. This chapter focused on assessing the quality of the photographic method so that the outcome can be used to develop the model and to judge the quality of the model later on. The reference data were subjected to four different comparisons (described in Table 1) to account for the following influences on the quality of the SVF:

- different road types;
- different types of road surroundings;
- different surroundings along opposite sides of the road;
- changes in surroundings over time;
- processing errors.

The research question that can be answered now is: *What is the quality of the reference data; is there a significant difference in the sky view factor when comparing fish-eye photos taken in the same area, at different moments in time (in the same season of the year)?*

Location of compared points

First of all it is important to note that the GPS positioning error complicated the comparisons of the surveys. The mean error ranged between 2.0m and 3.3m (see Table 2). Consequently, the calculated distance between two photo locations did not indicate the 'real' distance between them. Comparing two photographs in the same lane that seemed close by (according to the distance) but were not (according to the content of the photographs) could yield anything between:

- the same SVF due to the same kind of surroundings at both locations, despite the distance (as seen along the highway for example, in Figure 34);
- a very different SVF due to a large change in surroundings over that distance (like in forests and near overpasses).

Photographs taken at the same location were only found by coincidence (e.g. when selecting photographs to act as examples in the report). Only these photos provided a good opportunity to find subtle processing errors or minor changes in the surroundings.

In order to enable a better comparison, either the location of the photographs needs to be determined more accurately or comparable photographs should be selected manually.

Types of surroundings

The road surroundings can cause the SVF to fluctuate in a smaller or larger extent over smaller distances. When comparing point sets, it became clear that the SVF fluctuated most when the surroundings were both heterogeneous and close to the road. Large fluctuations were therefore common along provincial roads surrounded by forest and heterogeneous tree rows (as seen in comparison 2 for road stretches B, D, and E). On the highway large fluctuations in the SVF occurred when driving under an overpass (as noted in comparisons 1 and 3). Modeling these large fluctuations would require very precise data, as the fluctuations can be large over small distances.

Different lane or year

When comparing photographs taken in different years (comparisons 3 and 4) it is only possible to detect changes in the surroundings if they occur over a longer distance and cause a considerable trend shift in the SVF. The same goes for surveys done in opposite lanes, when spotting differences between the surroundings on either side of the road (comparisons 1 and 2). This is due to the relatively large processing error.

It was found that a survey in opposite lanes on provincial roads is only needed when the surroundings are very different on either side of the road, as seen in Figure 25 for road stretch B. This would mean that surveys in both directions are imperative when calibrating the model along a road stretch with asymmetrical surroundings.

Processing error

All photographs taken in 2010 were processed using the same threshold to create binary images (i.e. images used to calculate the SVF). The effect that this would have on the resulting SVF was severely underestimated during preliminary testing. In all comparisons examples were found of varying light conditions between photographs causing considerable differences in SVF. Point set 3 in Table 9 showed that just a minor difference in light conditions could cause a difference in SVF as large as 0.19. In other words: when the SVF is low, the threshold has a large impact on the results.

Quantifying quality

The quality of the reference data is quantified by several basic statistic analyses, summarized in Table 4. Given the errors encountered when inspecting the fish-eye photographs, these values should be interpreted cautiously. The values would be more useful if the point sets compared were surveyed on the same location and the deviating light conditions would be corrected for.

The mean and the standard deviation give a rough indication of the variations in SVF that can be expected under certain conditions, but for the purpose of modeling certain surroundings using segments with the same SVF, these values are too high. They are calculated for long stretches of road of which the surroundings seemed homogeneous during selection based on aerial photographs (i.e. all stretches along the provincial road) but proved quite diverse when looking at the SVF plots. Therefore these values should not be used as a guideline during calibration; the use of SVF plots of the reference data is preferred.

However, the method for creating the SVF plots is susceptible to errors, as discussed in step b. To make the use of SVF plots more reliable, it is recommended to find a new method to calculate the location of a SVF point along the road centerline.

3. Design and development of the GIS model

3.1 Introduction

Creating a new GIS model

This chapter describes the process of developing a model that calculates SVF values along roads in rural areas. The focus of the model is on modeling trees in 3D. As seen in the previous chapter, trees may cause very stable SVF values as well as very fluctuating ones, depending on the road surroundings. The aim is to create a model that can simulate the surroundings well enough to be considered a replacement for the photographic method. The developed GIS model is a workflow that describes how tree data and elevation data is used to calculate SVF values of an area.

There are several existing modeling methods for determining the SVF based on GIS data, but as mentioned in the main introduction, they focus on modeling urban – not rural – areas. When calculating the SVF in cities, buildings are the most influential surroundings, but along rural roads, trees are. So in order to be a good replacement for the photographic method, the model must be able to simulate trees as well as possible. Steenbergen has tried this: he mimicked trees by adding transparency to 2.5D extrusions of raster cells. This research will go a step further and try to model trees in 3D: this enables a more realistic representation which in turn is expected to lead to more accurate SVF values.

An important limiting factor of a detailed representation is the lack of detailed geo-data on trees, as Steenbergen also mentioned in his research. Initially, the proposed solution for this problem was to calculate the SVF along road segments: to calculate the SVF of portions of the road where the surroundings would be the same. While researching the quality of the reference data it became clear that the road surroundings fluctuate a lot, which makes classification of road stretches virtually impossible. Moreover, salting precision is expected to improve in the near future (Aebi Schmidt Nederland, September 2010), which also pleads for a more detailed modeling method. The new approach rendered research question 2 irrelevant, which is why this question is not included in this report.

The new method incorporates geo-data on trees supplied by the province of Gelderland. In this dataset the height of the trees is stored in more categories than is done in the municipal tree data of Ede (where 4 categories are used as opposed to 5), suggesting a higher accuracy. This accuracy is questioned in the next chapter, where the provincial height data is compared to the AHN 'vegetation data' (laser scanned elevation data containing all but the base ground reflectance).

Summarized, the new modeling method creates 3D trees based on provincial tree data that contain very limited information about the shape of the trees. In order to define 3D tree models, any additional information on the shape of the tree will have to be created based on the existing data. In this chapter, a tree model is defined (i.e. which elements make up a tree in which way) and a GIS model is developed that can calculate the SVF based on the input of these tree elements.

Development process

The chapter is structured according to the software development process of the GIS model, which consists of three main parts:

- Requirements
  - Defining the model criteria
- Design
  - Finding suitable software (in conjunction with the choice of tree model and choice of computational method)
  - Finding suitable input data
  - Creating a conceptual design (based on the model requirements and the software and data found)
- Implementation
  - Building a basic model

To start with, a list of criteria for the new model is compiled based on brief literature research and deliberation with Meteo Consult. Afterwards, this general list of criteria is supplemented by criteria for data handling in order to find suitable software for the model's execution. Concurrently, geo-data is collected that contain road surroundings that have a significant impact on the SVF, focusing on tree data in particular.

Based on the selected software and the found geo-data the basic model is built. This model is a conceptual workflow, a workflow that misses an essential part: the tree element parameters to define 3D tree elements. The values of these parameters can only be found by applying the model to a test location. This is done in the next chapter.

3.2 Requirements

In order to get an overview of modeling possibilities, a list was made of methods to choose from; some were proven by research and others existed in theory only.

Based on deliberation about strong and weak points of these methods, the requirements for the new model were set. Note that the criteria are formulated in order of importance, although <u>all</u> criteria, except for the last, must be met in order for the model to be a realistic option for the replacement of the photographic method. (The last criterion is only applicable to the execution of this research.)

The new model must meet the following requirements:

A. The model output should consist of viewsheds that can be used by Meteo Consult to calculate the sky view factor and sun view factor.

   Viewsheds are preferred above the direct output of the sky view factor and the sun view factor as it allows for complete transparency and freedom in the calculation of these values on Meteo Consult's side (see Figure 2).

B. To allow for comparison to the reference data, the viewsheds should be computed at the same location as the reference data. If this is not possible, the viewsheds may also be computed along the centerline of the road at an interval of max. 2 meters. This high viewshed resolution is necessary for two reasons: 1) it allows for comparison as the resulting values will lie close to the reference data; 2) as satellite positioning is improving, it's safe to assume that salting accuracy will also improve as the vehicles rely on satellite positioning (Aebi Schmidt Nederland, September 2010).

C. The trees are modeled in 3D as a simple representation of real trees.

   The greatest strength of the photographic method is that it captures everything that blocks the view of the sky realistically. By modeling trees in 3D it is expected that the model output is more realistic than what already has been done in previous research using 2.5D features, but it is not expected that the quality of the photographic method will be matched. Tree modeling is limited by the detail of the tree data available and the modeling method itself. Preferably the 3D models of the trees are scalable near-realistic replicas, containing many branches and twigs. However, a simpler model is preferred to reduce computing time and time needed for the creation of a base model for each species for example. It would be convenient to have one base tree model with several variable features (or 'tree elements') that can be scaled based on the tree data available. How this tree model is constructed and which features are chosen to be variables, is left open, as this largely depends on the capabilities of the modeling software.

   It is expected that the geo-data available does not contain a lot of data on the shape of the trees, which is important when modeling them in 3D. The conversion from tree data to 3D tree models will thus have to take as much relevant tree data into account as possible.

D. The model input will only consist of geo-data.

   This means that only geo-data that is produced or updated by a government institution on a regular basis will be used as input. These are datasets that are available to all companies in the Netherlands (although they might come with a price tag depending on the source). Using these datasets will benefit consistency of input data when using the model for different parts of the Netherlands, assuming that data is created and stored in the same way throughout the country.

   Fieldwork is avoided as a data collection method as it involves many man hours and means of transport. If an area was to be surveyed to collect high-quality data for the sole purpose of creating good 3D trees, it would probably be more efficient to take fish-eye photographs after all. This would deliver a more accurate result at about the same costs.

E. The execution of the model should be free of (additional) costs.

   As mentioned in the previous requirement, the costs of the execution of the model should not come close to – and certainly not surpass – the costs for the current method. A cost calculation is not a part of this research, but in general the costs will be kept low by:

- using software that is either free, available for a small fare, or available at government institutions (as they are possible clients);
- using geo-data available at government institutions to avoid extra costs for data acquisition (fieldwork or purchase);
- keeping the effort (= man hours) required to run the model to at a minimum.

F. The development time for the model is 1 – 1,5 months.

Considering the short development time, the model should be simple and straightforward: the main components should be developed but the aim is not to produce a fully functional end product for Meteo Consult. The lessons learned from this research are considered more valuable than the model itself.

Based on the requirements, the model's functions can be summarized into the flow chart below: the model will create 3D trees from available tree data and generate viewsheds based on the 3D trees and other road surroundings.

Figure 39: Summary of the model's functionality

3.3 Design

Software

<u>Methodology</u>

Once the requirements for the model were set, suitable software could be found to run the model. As Meteo Consult already has the software to calculate the SVF and sun view factor based on viewsheds, it is sufficient to find software that can project 3D trees (and other surroundings too, if possible) onto a viewshed.

Three programs were compared: ArcGIS Desktop, SkyHelios and Rayman. ArcGIS Desktop is enterprise software, while the other two programs are research prototype software. All programs have the ability to calculate the SVF based on 3D input. The comparison was done using the criteria in the table below.

Table 16: Software criteria

Area of concern	Requirement
Output	Viewshed image or SVF value and sun view value
Output	Same location as the reference data or Along the centerline of the road at an interval of max. 2 meters
Input	Shape and location of trees
Input	AHN5 as DEM
Licensing	Free or at a low price
Development time	1 – 1,5 month

<u>ArcGIS Desktop</u>

The ArcGIS Desktop suite by ESRI is the most common GIS program used in the Netherlands. It is used by most government institutions and is also available at Wageningen University.

Previously, Steenbergen calculated the SVF using 2,5D raster data as input. The new version of the program, ArcGIS10, provides the opportunity for the input of 3D features ('multipatches'). The 'Skyline Graph' tool of the 3D Analyst toolbox can calculate the SVF using these 3D features as input, but it has a major drawback which does not allow for detailed calculations using 3D trees: it does not support 'overhangs' (more information in the box below).

"The skyline does not support overhangs, even at full detail. It's as if, for each nonvertical edge in the feature, a vertical line is drawn down from each end of the edge until the skyline, and everything between those two vertical lines is obscured by the feature. The result is that if a building has just a narrow base and tower, with a much wider top, the wider top will cause the skyline to be generated as if from a cloth being draped over the wider top and hanging straight downward from the edge of the wider top."

Source: http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/ How_Skyline_3D_Analyst_works/00q90000008t000000/, Accessed December 20th, 2010

It was also found that it is very laborious to create a 'proper' multipatch feature that is closed and correctly positioned in space. In addition the program functioned very poorly on the available hardware.

As is it is safe to assume that most governmental data suppliers in the Netherlands possess ArcGIS version 9.3 or 10, the decision was made to use this program for standard data processing purposes only and to look for other programs to perform the SVF calculation.

SkyHelios and RayMan

SkyHelios and RayMan are two programs developed by prof. Matzarakis, researcher at the Meteorological Institute at the University Freiburg, Germany.

RayMan (Matzarakis, Rutz, & Mayer, 2010) focuses on the calculation of radiation indices defining the energy that a human body receives. In order to define the circumstances of the human body, it needs meteorological data, details about the human body, and spatial data on the surroundings as input. Two types of spatial data can be loaded into the program: the ground elevation (text file) and 'obstacles' on the surface (obstacle file). An obstacle file is a file format that can store three types of 3D data: buildings and two types of trees.

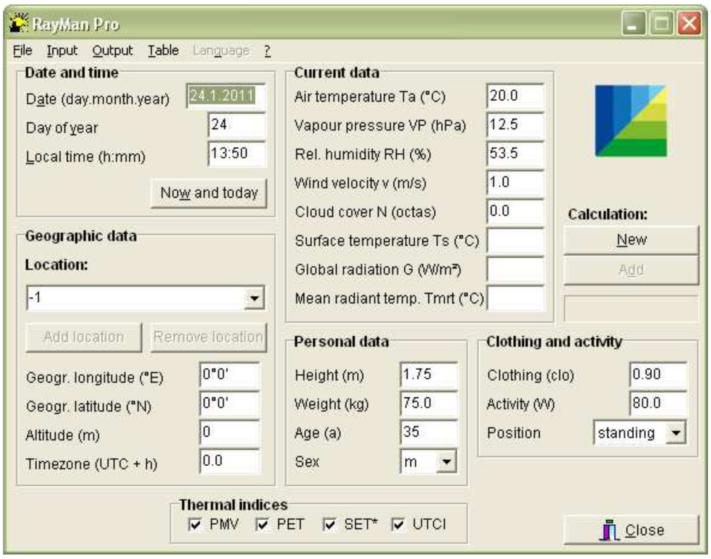


Figure 40: The interface of RayMan, version 2.1.

RayMan calculates a vast amount of climate values for one point in space, one of them being the SVF. It is possible to let the program calculate the conditions for several locations automatically by defining their locations and conditions in a specifically formatted text file. Manually, it is also possible to use a fish-eye photograph as input and to save the viewshed of a location as well as the sun view factor for a given moment in time.

SkyHelios (Matuschek & Matzarakis, 2010) is specialized in calculating the SVF of an area; it calculates the SVF of every point in this area according to a given resolution. In addition to a DEM and an obstacle file, shapefiles can also be used to define the surroundings (although this did not succeed during testing). The SVF can be calculated by two different methods (of which the difference is not documented). The program displays an image of the viewshed of any calculated spot, projecting the surrounding using the same projection used in a fish-eye photograph (Matzarakis & Matuschek, 2011). During SVF calculation this image is not saved along with the calculated SVF data; it can only be saved manually. The sun view can also only be prompted manually. SkyHelios maps all the spatial data in a window (see Figure 41), as opposed to RayMan where the data only can be loaded into memory and is not visible to the user.

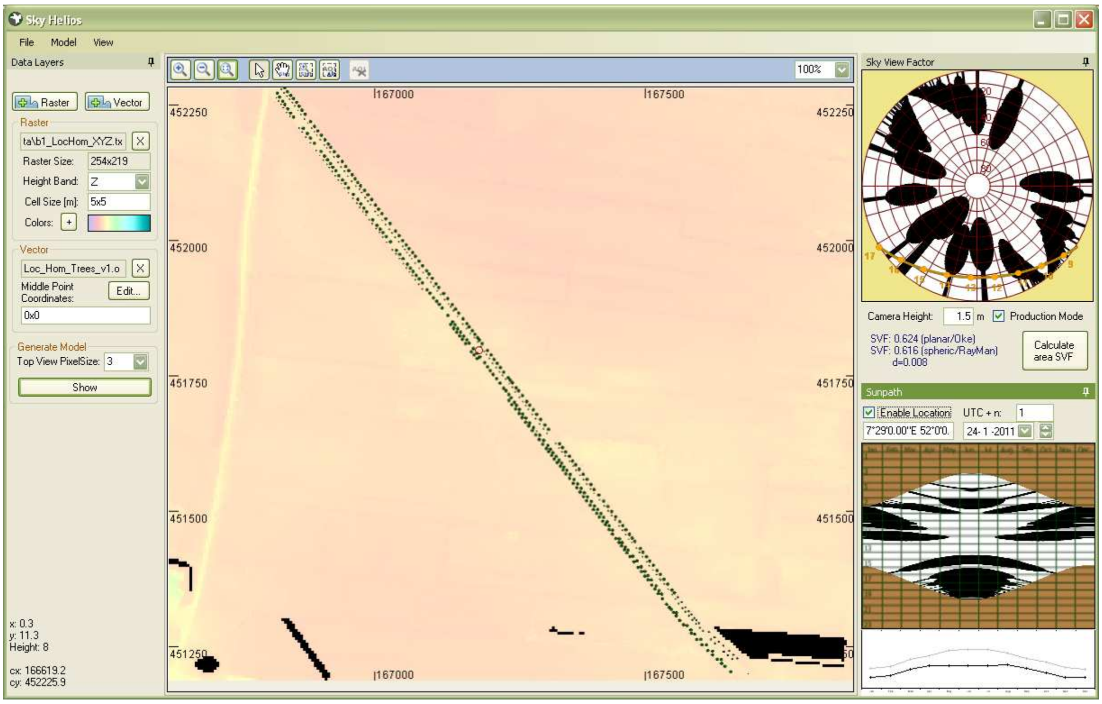


Figure 41: The interface of SkyHelios, version 1.2, build 3863

To load road surroundings into SkyHelios, a shapefile (containing polygons) and an obstacle file can be used. A shapefile is useful for displaying solid 2.5D features like buildings; the 2D polygons are extruded by a height attribute, stored in the shapefile's attribute table. The obstacle file has the ability to store solid features and two types of trees in 3D. The 3D trees are modeled by seven attributes, or 'tree elements' (see Figure 42):

- 'class' (deciduous or coniferous);
- X and Y coordinates;
- tree height ('h');
- trunk height ('l');
- trunk diameter ('d');
- crown radius ('r').

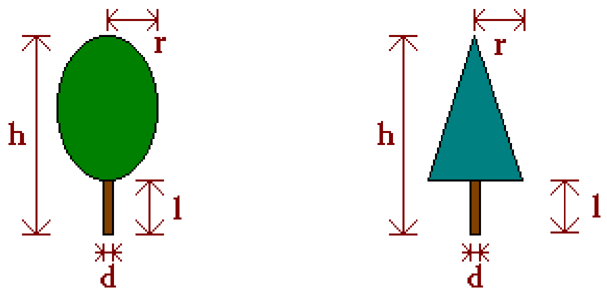


Figure 42: Tree models of a deciduous (left) and coniferous tree (right).
Source: RayManEditor.exe, distributed with SkyHelios version 1.2 build 3863

The tree elements need to be stored in this order in a text file (with the extension '.obs'). In this file, each row represents a 3D object. The unit of all metric values is meters.

Comparison and choice

Table 17: Requirements of the model and the capabilities of the available software

Area of concern	Requirement	ArcGIS 10	SkyHelios	RayMan
Output	Viewshed image	Y	N	N
	or			
	SVF value and	Y	Y	Y
	Sun view value	Y	N	N a)
Output	Same location as the reference data	Y	N	Y
	or			
	Along the centerline of the road at an interval of max. 2 meters	Y	Y	Y
Input	Shape and location of trees	N	Y	Y
Input	AHN5 as DEM	Y	Y	Y b)
Licensing	Free or at a low price	Y/N c)	Y	Y
Development time	1 – 1,5 month	N d)	Y	Y

Comments found in the table:

a) *Manually prompted calculation: per hour. Automatic calculation: per day.*
b) *Conversion into correct format theoretically possible, but did not succeed during tests.*
c) *Available at data suppliers, but not at Meteo Consult.*
d) *Making a 3D model in this program is too laborious.*

As can be seen in the table above, SkyHelios meets all criteria but one: it is not possible to calculate the sun view factor, neither directly within the program nor using viewsheds. This inability is a major disadvantage, as the calculated SVF has to be accompanied by the sun view factor in order to be used in Meteo Consult's network model. Without the enablement of one of these two functionalities Meteo Consult cannot use this program.

Fortunately, this drawback poses no danger for the continuation of the research. The focus is on creating a model that produces the same sky view factor as the photographic method; the sun view factor lies beyond the scope.

Preferably this drawback will be overcome by implementing a functionality that saves the viewshed for each calculated point. The viewsheds can then also be used to calculate the SVF using other formulas than those used in SkyHelios if desired. Chapter 3.6: 'Recommendations' elaborates on the points of improvement needed and desired in SkyHelios.

A large advantage of the use of SkyHelios is that it provides a very straightforward way of defining a tree and calculating the SVF. The trees are truly represented in 3D and the shape can be defined by many elements. The SVF calculation can take some time, but this is not seen as a drawback because it does not require human interference.

Assuming that a functionality to calculate the sun view factor can and will be implemented in the future, SkyHelios is chosen as the SVF calculation tool in the model.

Data

A lot of geo-data is used during the development of the model. A distinction can be made between data needed to run the model, and data used to refine it and to interpret trends and outliers.

<u>Geo-data for model input</u>

- Provincial road surroundings dataset (October 2010)

  This is a large-scale dataset provided by the municipality of Gelderland. It contains road features such as the road outline, centerline and kilometer marks as well as all road surroundings that fall within the road management zone. This zone is usually restricted to adjacent biking paths and vegetation zones, and is usually not wider than a few meters.

  The municipality of Ede was contacted first for more detailed data on road surroundings. As this data only covered the surroundings of roads managed by the municipality, the province of Gelderland was contacted for the geo-data along the provincial roads. Since only road stretches of provincial roads were selected for testing purposes, the Ministry of Public Works was not contacted for geo-data along the national highway within the research area.

  - Vegetation data (individual trees)

    For the model the subset containing vegetation data is used. This dataset contains points and polygons representing individual trees and areas with several trees respectively. The individual trees are used as input in the model.

    The tree data contains two attributes of relevance for creating a 3D tree: its species and its height (stored in 5 categories).

    On data delivery there was no join between the spatial features and their attributes as the unique identifier of the spatial features was not stored in the attribute table.

- AHN: Algemeen Hoogtebestand Nederland (AHN-1), 5x5m

  SkyHelios has the ability to use a DEM as an input, containing the base elevation of an area. The AHN with a resolution of 5 meters was used; the most detailed DEM available at the moment. It is provided for free by the GeoDesk (the geo-data service desk of Wageningen University and Research Centre) but is made on a national level by the AHN-organization – a collaboration between the Ministry of Public Works and the water boards.

- Top10NL

  The Top10NL is a national topographic vector dataset at the scale of 1:10.000. It is made by the Cadastre and was also acquired through the GeoDesk.

  Top10NL can be used as extra data source for road surroundings in order to include features that lie outside range of the provincial data.

- Forest polygons

  As SkyHelios accepts the input of shapefiles, an attempt was made to take patches of forest along the road into account during the application of the model (see chapter 4). However, this did not succeed. Since the conversion from shapefile to an obstacle file would require a lot of manual labor, these surroundings are not included in the model, but it is strongly advised to use them (as forests have a large impact on the SVF).

Geo-data for calibration, validation and interpretation

- Reference data (2007 and 2010)

  The reference data described in the previous chapter is used to calibrate and validate the model. The most important data for the comparison are the location and the SVF calculated from the picture taken at that location.

- Provincial road surroundings dataset
  - Road centerline and kilometer marks

    The road centerline is used to project the SVF locations on to enable a comparison between the calculated SVF and the SVF of the reference data. The road centerline of the provincial dataset is more accurate than the centerline provided in the Top10NL dataset and is therefore preferred.

    The kilometer marks are used to indicate the position of the test locations and the position of the SVF calculation points along the road.

- AHN vegetation dataset ('AHN-veg')

  The AHN is made based on a raw dataset containing all laser scanned elevation points. Not all of these points are used to create the AHN, because only the ground elevation is of interest. The AHN vegetation dataset contains all points that are not used to create the AHN. As the title suggests, it contains the elevation of vegetation as well as buildings and other objects that are not of interest to create the AHN.

  The AHN-veg is used to validate the tree heights of the provincial vegetation dataset in the next chapter.

Furthermore aerial photographs (2008, 25x25cm) and the Top10NL dataset were used for navigation and creation of overview maps. The aerial photographs also allowed for spotting flaws in the provincial tree dataset regarding the location of trees and the differences in crown size (a possible indication for differences in height).

Conceptual design of the GIS model

Based the software's capabilities of describing 3D trees, its calculation method, and the data available, it was clear that conceptual design of the model would consist of these steps:

1. Pre-processing the input data

   Extracting all data of an area of interest and transforming it into a format that can be handled in later steps

2. Calculation of 3D tree elements

   Calculating the values of the 3D tree elements based on the input data

3. SVF calculation

   Calculating the SVF of the whole area of interest based on ground elevation data and 3D trees

The diagram below shows the steps of the model supplemented with the datasets. This diagram forms the basis for the development of the model.

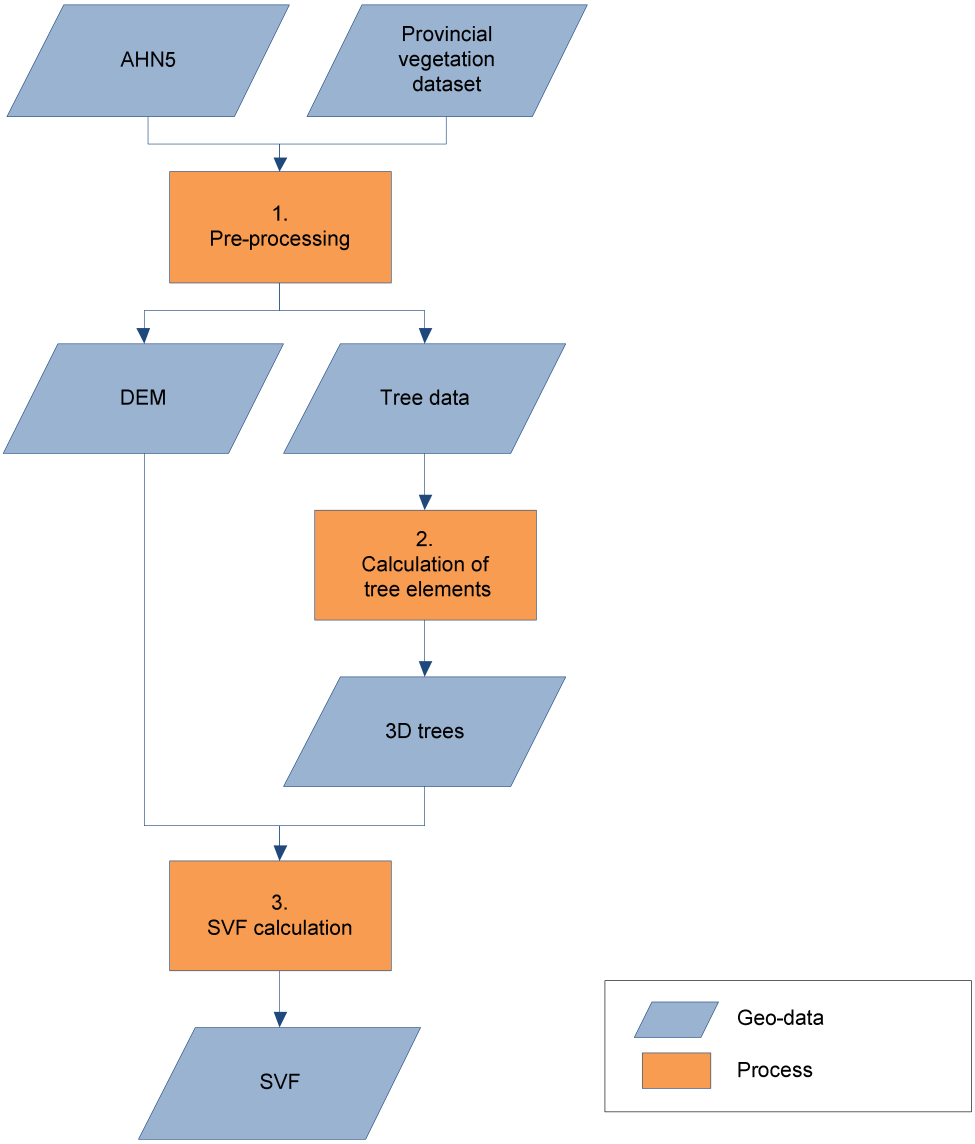


Figure 43: Flowchart of the basic model

3.4 Building a basic model

As mentioned earlier, the model was developed in two phases: by building a basic model first and refining it afterwards. This paragraph explains how the framework of the model was developed; it explains how the three steps of the conceptual design were implemented and why, but it is not applied on a test location yet. This is done in the following chapter: there, the values needed for the 3D tree conversion are found by calibration and the model is refined further after validation on two test locations.

The most important task of the model is step 2: to transform the provincial tree data into 3D shapes – the essential data source for SVF calculation. The other steps are handled in less detail as they have a supporting function to the working of the model.

The execution of the steps is described in detail in appendix III: 'GIS model processing in detail'.

Step 1. Pre-processing

The pre-processing involves acquiring the data needed and preparing it for handling in following steps. Two datasets had to be preprocessed: the DEM provided by the GeoDesk and the tree data provided by the province of Gelderland.

First the boundary box of the area of interest was created and used to clip data from the DEM and tree dataset.

For import in SkyHelios, the DEM has to be stored in a text file with a specific (but basic) column structure; each row in the text file represents a raster cell with X, Y and Z coordinates. The resolution is not stored (thus has to be remembered by storing it in the file name for example). As the AHN is in ESRI grid format, the conversion to text format was done using basic tools in ESRI's ArcGIS to start with, followed by the use of Excel and a text editor to complete the process. This conversion process was not automated since it is very basic in nature (thus does not rely on the use of specific software) and because it only had to be performed a few times during this research: once for every test location.

Step 2. Calculation of tree elements

In this step the provincial tree dataset is converted into 3D trees that can be used for SVF calculation in SkyHelios.

The provincial data only provides four of the seven elements required to define a 3D tree: the X and Y coordinates, the species, and the height. None of these four elements is 'ready for use':

- The coordinates are not stored as values in a table, but as spatial features (i.e. points). The values were extracted using ArcGIS.
- The tree heights are stored in five categories (5-10m, 10-15m, 15-20m, 20-25m, and >25m) described as a string (e.g. "25 20-25 METER"). The first two characters (which describe the maximum height of the category, except in the highest category) are stored in a new, numeral field. The height is changed to half of the category height during the conversion.
- The species are defined by an IMAG code[10] accompanied by a Dutch species name. This code can be matched to a list of species to determine the 'class', but during the research this conversion was done manually.

[10] The IMAG code is consists of maximum 8 characters and is based on the Latin name of a plant species. Dutch nurseries use the code for trade and communication.

Additionally, the 'missing' elements (trunk height, trunk diameter, and crown radius) need to be created. This was done by multiplying the tree height by a voluntary value. For the development of a functioning basic model, the exact values of these parameters did not matter; during calibration the values for a certain species were found. This is described in the next chapter.

The conversion from provincial data to 3D data was done using Excel. A workbook with three sheets was created:

- Input data sheet;
- Output data sheet;
- Tree element parameter sheet.

The output data sheet has the same column structure as the obstacle file. The table below shows which formulas and which data was used to create the tree elements in the output data sheet.

Table 18: Conversion sheet formulas

Tree element	Formula (pseudo code)
Class	*Assigned manually during development and testing. Proposed formula:* If InputSpecies is in ListOfConiferousSpecies Then Class = 'n' Else Class = 'l'
X	= InputX
Y	= InputY
Height	= (If InputHeight > 0 Then TempValue = InputHeight Else TempValue = 10)* -2.5**
Crown radius	= InputHeight · Parameter for crown radius
Trunk length	= InputHeight · Parameter for trunk length
Trunk diameter	= InputHeight · Parameter for trunk diameter

** This formula was used to assign height to trees of which the height was unknown in the provincial dataset. It was assumed that the values are unknown because the trees have just been planted; so the lowest height category is assigned.*
*** 2.5 meters is extracted from the height to transform it from the maximum of a category to the middle of a category. This should not be done for the maximum category.*

When the formulas are implemented, the output sheet is filled automatically when the input sheet is loaded with data. As an example: if the provincial data is stored in a shapefile, the .dbf can be imported in the input sheet and the output sheet will automatically be filled with the converted values. The output sheet has the same structure as an obstacle file, so the conversion is almost complete. Ultimately, the obstacle file is created by exporting the output sheet to a tab delimited text file and changing the extension to '.obs'.

Due to time restrictions, the species data is not used actively to determine the class of the 3D trees. During testing the class was set manually.

Step 3. SVF Calculation

SVF calculation is the main function of SkyHelios, therefore no development was required for this step.

After loading the data into the program, the SVF can be calculated by pressing the 'Calculate Area SVF' button. A window appears (see Figure 44) where the settings for the run can be set:

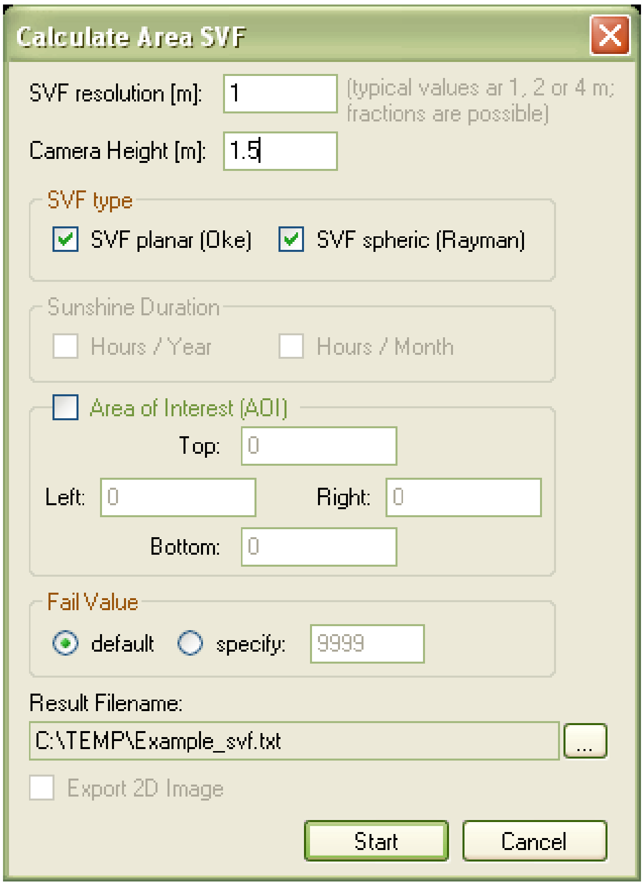


Figure 44: The 'Calculate SVF' window in SkyHelios where the preferences for a calculation session can be set.

- The resolution of the SVF sampling can be set as low as 1 meter, but was left at 2 or 4 meters during testing to decrease the processing time.
- The camera height was set at 1.5 meter as the camera is approximately at this height during photographic surveys. The effect that a change of this might have on the resulting SVF was not tested.
- It is possible to choose a method of SVF calculation ('spheric' or 'planar'). Both methods can be applied concurrently, yielding two sky view factors for each point location. The differences in SVF output of the calculation methods are documented in chapter 4 in section 'SVF calculation methods'.
- An area of interest can be set if needed.

The output of the program consists of a text file with the (X-, Y- and Z-) coordinates and the SVF per calculation method (with the camera height included in the heading) of all the points in the area of interest.

```
x;y;svf-spheric-1.5m;svf-planar-1.5m;z
166917.5;452117.5;0.86;0.90;7.3
166918.5;452117.5;0.86;0.90;7.2
166919.5;452117.5;0.85;0.89;7.2
166920.5;452117.5;0.84;0.88;7.2
166921.5;452117.5;0.83;0.87;7.1
166922.5;452117.5;0.84;0.87;7.1
166923.5;452117.5;0.83;0.86;7.2
166924.5;452117.5;0.82;0.85;7.4
166925.5;452117.5;0.82;0.85;7.5
166926.5;452117.5;0.81;0.84;7.7
166927.5;452117.5;0.81;0.84;7.8
```

Figure 45: A sample of the produced results by SkyHelios for a test location. It contains the coordinates and SVF for all points, calculated at a sampling resolution of 1 meter.

Due to the missing sun view factor, no effort was paid into transforming the results of this step into a format that Meteo Consult can input directly into their Network model. Therefore, this is the last step of the model.

3.5 Discussion and conclusions

This chapter has covered the development of a new GIS model that calculates the SVF of an area based on 3D trees and elevation data. The research question this chapter set to answer is: *Can the sky view factors of road segments be calculated using a GIS model?* This question can be answered affirmatively, even though it calculates the SVF of an area (as opposed to 'road segments') and has a few limitations.

Modeling method

The most important drawback of the model is the inability to pair the SVF with a sun view factor. The program that calculates the SVF, SkyHelios, only supports manual lookup of the sun view factor. As both the SVF as the sun view factor are needed as input for Meteo Consult's network model, the incorporation of a method to calculate the sun view factor is imperative. The proposed improvements concerning this issue can be found in the next paragraph, 'Recommendations'. It should be stressed that SkyHelios is a free program, but it is not open source; improvements to the program can only be done by the developer, prof. Matzarakis, researcher at the Meteorological Institute at the University of Freiburg, Germany.

On a more positive note, the model manages to calculate the SVF based on 3D trees to calculate the SVF in areas where trees are the main cause for sky view obstruction. As far as the author is aware of, this has not been done before. In the model, a 3D tree is defined the way SkyHelios can interpret its shape directly. SkyHelios can namely use 3D features like trees and buildings as input, stored a so called 'obstacle file'. In this file, a 3D tree is characterized by seven elements: five to describe the shape of a tree, and two for its location (X- and Y-coordinates). The conversion method used to create these elements is straightforward: each tree element is found by multiplying the tree height by a certain percentage. As the pros and cons of the chosen tree element definition and the conversion method can only be found after applying the model to a test site, these are discussed in the next chapter.

Only two datasets are needed as input to the model: a DEM and a provincial vegetation dataset containing data on individual trees. In order to create 3D trees, only the coordinates and the tree height category of the trees is used. Due to time restrictions, the tree species could not be taken into account. Ideas on how the species data could be put to use can be found in the 'Recommendations' below.

SkyHelios can calculate the SVF with two calculation methods: 'planar' and 'spheric'. The difference between the outcomes of these two methods can be seen in the next chapter, where the model is applied to test locations.

Road surroundings

The main focus when developing the model was on modeling trees, as they are the main cause of sky view obstruction in rural areas. While the model successfully takes the base elevation and individual trees into account, there are other relevant features surrounding the roads that are left out of the equation.

First of all, the model can only deal with individually surveyed trees (i.e. points) within the road management zone of the province of Gelderland. The dataset also contains polygons, representing areas with trees. As the model is made to handle individual trees only, these patches cannot be taken into account. Furthermore, it is uncertain if the model can process other province's vegetation datasets since the data structure of these datasets is not known.

Moreover, the model does not consider large features that may have a significant impact on the SVF, like forests and buildings adjacent to the province's management zone. More on the attempts to include these features and the road surroundings recommended for use can be found in the paragraph below. The next chapter shows how the SVF is affected when trees adjacent to the provincial road management zone are taken into account (see paragraph 4.3, section 'Additional surroundings').

3.6 Recommendations

SkyHelios

The model depends heavily on the use of SkyHelios. The functions implemented in this program can make or break the model as a whole. The program is not open-source, meaning that it is not possible to implement new features; this has to be done by the developer. This is a large drawback for the use of it.

The use of SkyHelios in the model can only be continued if two major improvements are made: the ability to calculate the sun view factor and a smooth handling of shapefiles.

There are (at least) two possible solutions for calculating the sun view factor. The first is to calculate it based on exported images of every viewshed of every point. During the calculation of the SVF SkyHelios shows the viewshed of the point it is processing. It would probably take just a minor effort to write a code that saves the viewshed as an image. The function already seems to have been developed, but has not been activated in the version used (see Figure 46).

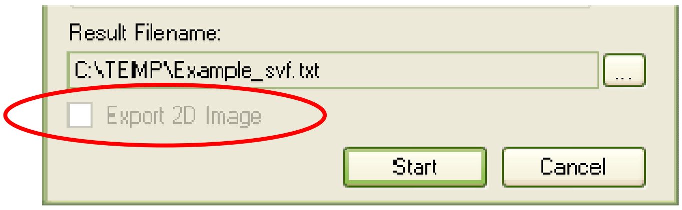


Figure 46: Inactive functionality for exporting 2D images (presumably viewsheds) in the 'Calculate Area SVF' window of SkyHelios.

A second possibility is to let SkyHelios output the sun view factor of every hour of a whole year for every location it calculates the SVF for. This should only require a small amount of effort as the position of SkyHelios is already capable of showing the sun's position per hour (as seen in Figure 47).

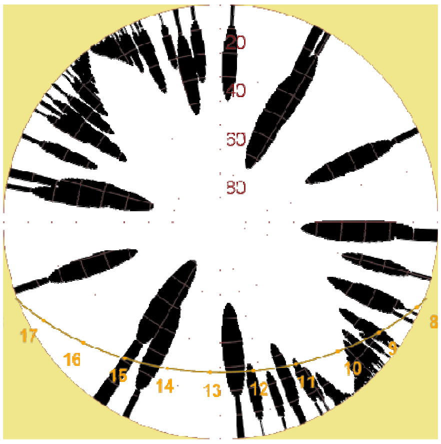


Figure 47: The sun view plotted per hour

The program was not able to render shapefiles if the DEM contained 'NoData' cells. As the maker of SkyHelios was not contacted concerning this issue, it may very well be easy to overcome. The use of shapefiles is a quick and convenient way to include 2.5D blocks of road surroundings in the SVF calculation, as it is a well-known format and will only be used to include simple features. If this issue is not resolved, the problem can be solved the other way around: by interpolating the DEM in order to fill the 'NoData' cells with values. This is of course not preferable at all, as it passes the problem (and the responsibility to solve it) on to the user.

Another feature that could be added to SkyHelios, but is not required for a well-functioning model, is the ability to limit the SVF calculation area. Currently, SkyHelios calculates the SVF of a large amount of points according to a set resolution. There is only one option to limit the area of interest, i.e. the area within the SVF is calculated, and that is by manually defining a rectangular area of interest in SkyHelios. It would save a considerable amount of processing time if the area of interest could be limited to the roads only. A proposed solution for this is to add the ability to import shapes that indicate the area(s) (or point locations) of interest; polygons can be used to indicate larger areas, and a point dataset can be used to indicate the exact positions where the calculations should take place.

Tree model

At the moment, the deciduous tree model provided in SkyHelios is solid, thus representing a tree with leaves. As the resulting SVF will be used to calculate the temperature of the road surface in the colder months of the year, coniferous trees and bare deciduous trees are the most important objects to model. The model mimics a bare tree by assigning artificial dimensions to the 3D trees. The transparency element should be variable to account for seasonality of the leaf cover.

The tree species provided in the provincial vegetation dataset are not put to use, although this is valuable information when defining the shape of a tree. As trees of different species have different shapes, the conversion parameters should be dependent on the species. It was considered out of scope to find out which parameters should be used for which species, as this requires extensive calibration. In the next chapter the parameters for just one species will be found as a proof of concept for the conversion.

In addition to including the species in the conversion process, a better conversion formula might be needed. The 3D trees are created by simply multiplying their parameter with the tree height, no attention paid to how a tree's shape actually changes as it grows.

Road surroundings

The model can take two types of road surroundings into account: individual trees and elevation data. In order to simulate the real situation better, more datasets can (and should) be applied.

Extra surroundings can easily be exported from the Top10NL dataset and used as 2.5D data in SkyHelios. Forests and buildings are especially well suited for this. 3D objects, like overpasses, can be included by modeling them in an obstacle file. The heights of these features, that are stored in 2D in the Top10NL dataset, can be derived from the AHN vegetation dataset, as this dataset includes the raw height of all objects that are not part of the ground based elevation. The next chapter contains an example showing the effect of additional surroundings on the SVF (in paragraph 4.3, section 'Additional surroundings').

The vegetation dataset provided by the province of Gelderland also contains polygons, representing areas with trees. These polygons can be used by extruding them in the same way suggested for 2.5D objects above. Another option is to derive the location of trees within these areas using the AHN vegetation dataset and to save these locations as points.

Concerning nationwide applicability, it is recommended to find out if the vegetation datasets of other provinces of the Netherlands have the same structure. Internationally, it is expected that the introduction of INSPIRE will accommodate seamless exchange of geo-data between countries within the EU, although it is not known if datasets that fall under this agreement will include the necessary geo-data for this model.

4. APPLICATION OF THE GIS MODEL

4.1 Introduction

Chapter content and structure

In the previous chapter the basic model was developed: a model with a workflow but without the right conversion values to create realistic 3D trees. In this chapter, the model is supplemented with these values first and validated on two locations afterwards.

The conversion values (or 'tree element parameters') are found through calibration of the model. The model is applied on a test location repeatedly, using different parameters every time. Only two tree element parameters are changed; the others are left constant in order to simplify the process. After each run, the SVF values are compared to the reference data to see how they match up. The best fitting values were found after nine iterations.

After the calibration session, the sensitivity of the model is studied. Changes are made to four parts of the model to see how these changes affect the output.

The chapter concludes with a validation of the model. It is applied on two other locations in order to see if the results match the reference data along a longer road stretch. Since the quality of the tree heights provided in the provincial data was questioned during the sensitivity test, the heights will be checked using AHN vegetation data. The model is run again using the heights derived from this dataset, using the same tree element parameters.

Adoptions to the model

The calibration and validation processes require an addition to the model in order to compare the output to the reference data (step 4 in Figure 48). In the previous step, SkyHelios calculates SVF according to a specified resolution for a whole area, but the only values of interest are the ones close to the road centerline. In the same way as was done with the reference data in the previous chapter (see step b on page 14), these points are transformed to points along the centerline, and afterwards the distance from a common starting point is calculated for each point. The starting point is based on the most western kilometer mark of the location in question. In the end, all SVF points are located along the provincial road according to the road kilometer marks, enabling comparison of all SVF values computed for that road stretch. Since this step is only added for the purpose of analysis and is not part of the final model, this process is not described in further detail in this report.

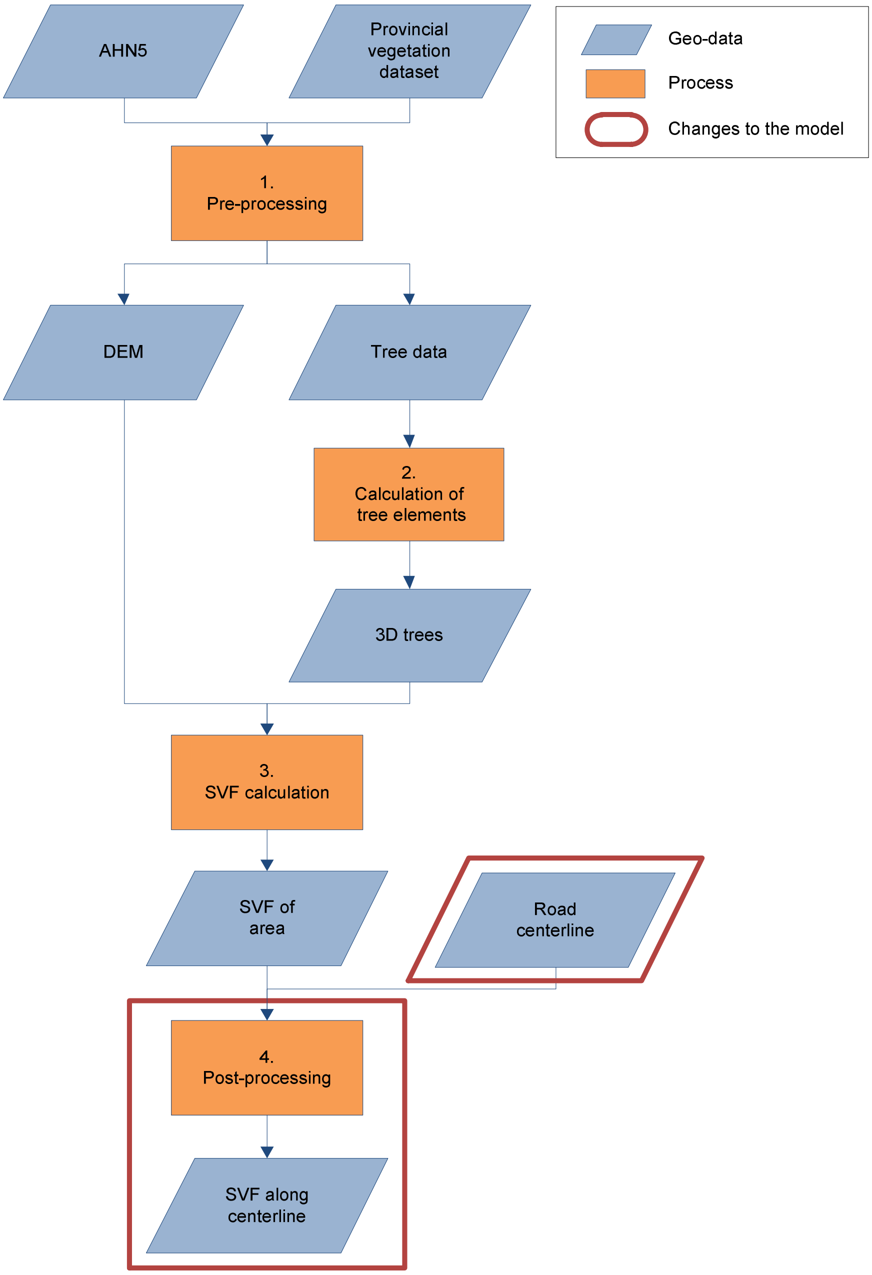


Figure 48: Changes made to the basic model to enable comparison to the reference data

4.2 Calibration

The calibration session focuses on one part of the model: the tree element parameters. The model is calibrated in order to find good values for these parameters, so that a conversion from provincial tree data into 3D trees will simulate the real situation as well as possible.

Location

The GIS model is calibrated along a road stretch where:

- trees are the main cause of sky view obstruction;
- the surroundings are relatively homogeneous throughout the road stretch;
- the SVF fluctuates at one location and is relatively constant at another;

- provincial tree data is available.

The chosen location for calibration is a road stretch of 300 meters along the provincial road N224 between Renswoude and Ede (HM 30.05 – HM 30.35, see Figure 49). It is surrounded by two rows of trees on each side and a small piece of forest at the beginning of the stretch.

Figure 49: Overview of the calibration location.

The location is chosen based on the consistency of the surroundings leading to a relatively steady SVF of 0.8 (see Figure 50). It is a part of road stretch 'A' analyzed in the previous chapter, where the comparison of the SVF of opposite driving directions yielded good values for correlation (0.78), mean SVF difference (0.02) and a low standard deviation (0.013) (as seen in Table 4).

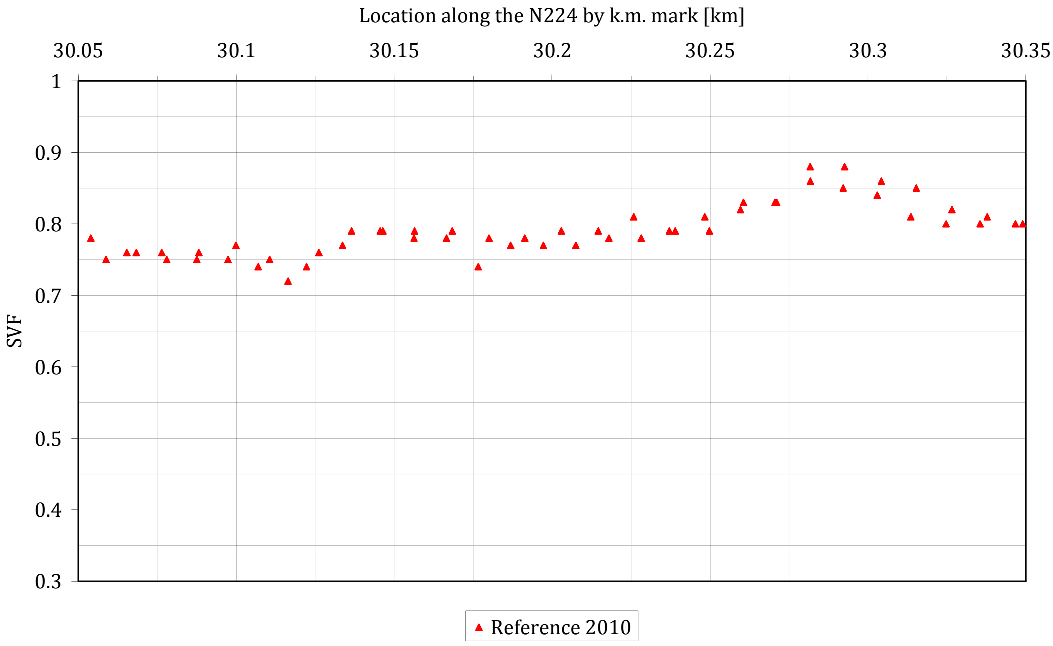


Figure 50: The SVF on the calibration location.

Approach

The calibration process requires a few additions to the basic model to enable comparison to the reference data and easy change of tree model parameters (see Figure 51). The only change needed in step 2 was to add a functionality to allow for an easy change in the parameters applied in the output sheet.

Another change to the model is mentioned in the introduction; step 4 (post-processing) is needed to transform the SVF output of SkyHelios to points along the road centerline. Afterwards, these points are plotted and compared to the reference data. Depending on the outcome, a new version of parameters is created and the model is run again from step 2 onwards.

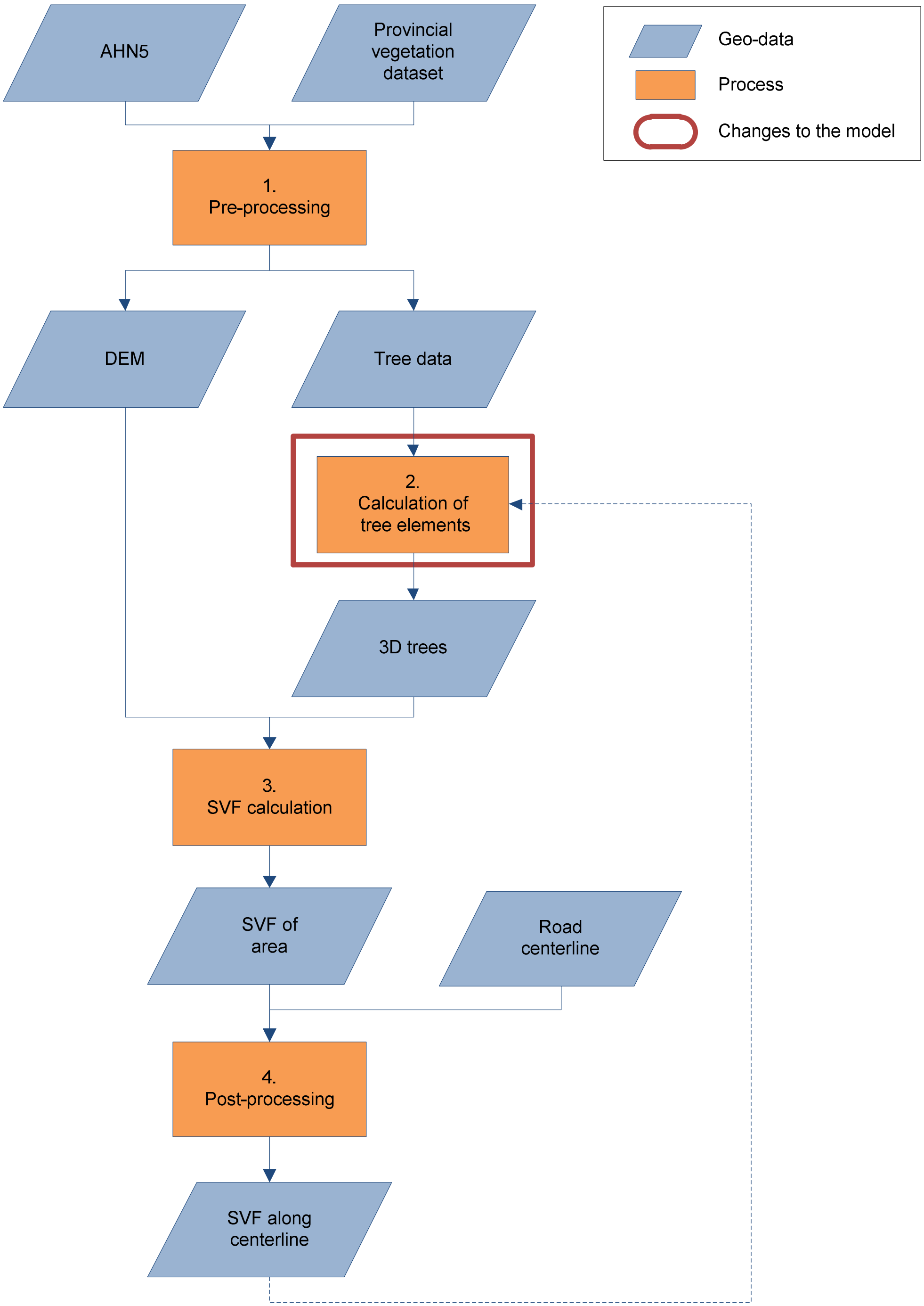


Figure 51: The workflow of the calibration process. The conversion parameters in step 2 are changed based on the outcome of step 4.

<u>Pre-processing</u>

During the execution of the first step of the model, the pre-processing, the provincial tree data provided some difficulties. The spatial features of the trees did not contain unique identifiers and had to be linked to the attribute table provided manually. This was done by looking at certain data from the attribute table: the distance along the road (referring to the kilometer marks), the distance from the road, and the side of the road (left or right).

The attributes needed for the conversion (the species, the X and Y coordinates and the tree height) were derived in the following manner:

- Only half of the trees contained data about the species. (This attribute did not need any pre-processing to be used in the next step.)
    - In case the species of a tree was unknown, it was compared visually to trees with a known species in the vicinity (using aerial photographs). As tree rows tend to be planted using only one species, the species of surrounding trees were assigned in all cases (see Figure 52).
- The X and Y coordinates were obtained from the spatial features of the provincial data on millimeter-level. The assumed accuracy for rural objects in general is approximately 50 cm.
- The height was acquired from the provincial data by extracting the first two numbers in the height string. This number indicates the maximum value of category, except for the highest category where the value indicates the minimum ("> 26").
    - During calibration, the extracted height value was not converted to indicate the middle of the category. This mistake was noticed and corrected during validation.
    - In case the height of a tree was unknown, it was assumed that the trees had been recently planted, so the lowest height category was assigned.

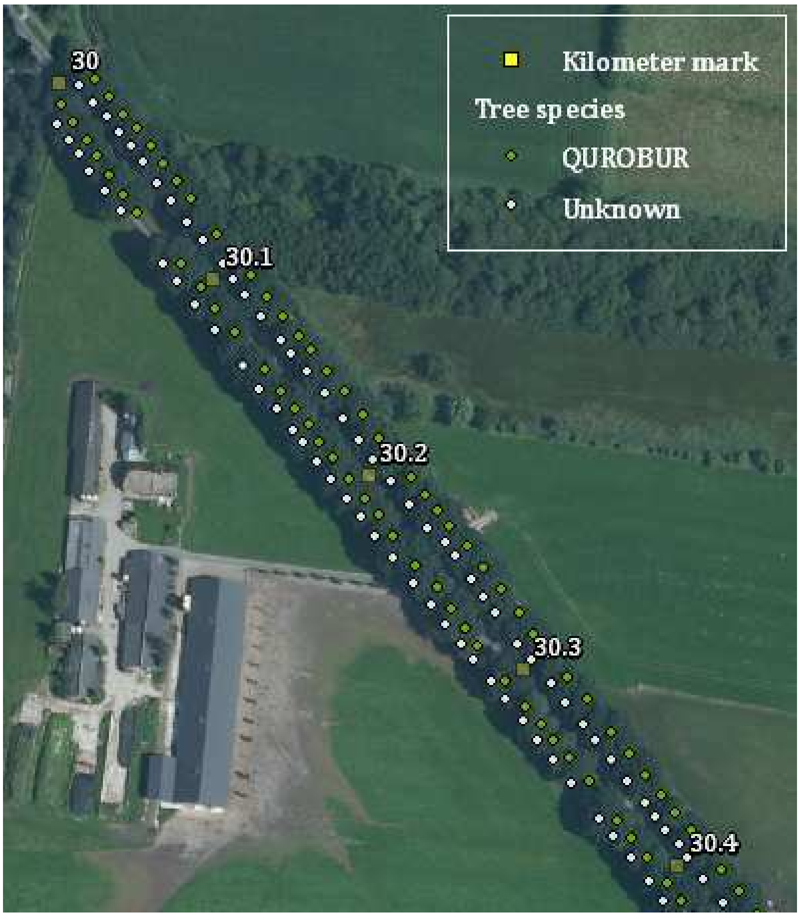


Figure 52: The provincial tree data lacked information on species and height on approximately half the trees on the calibration location.
('QUROBUR' is the IMAG code for Quercus robur*)*

Calibration of the tree elements

The variables chosen for the calibration are the parameters for the trunk length and the crown radius. The emphasis of the calibration is on the crown, as the crown is positioned closer to the center of the viewshed and therefore it has a larger effect on the SVF than the trunk. (This phenomenon is explained in paragraph 4.3, under the section 'Tree elements parameters'.)

In order to limit the amount of variables, the parameter for the trunk diameter is kept the same throughout the calibration session (0.025), as it is considered the least important.

The calibration started with arbitrary values for the parameters for the crown radius and the trunk length (being 10% and 25% respectively). For each run, these values were changed one parameter at a time, as can be seen in Figure 53.

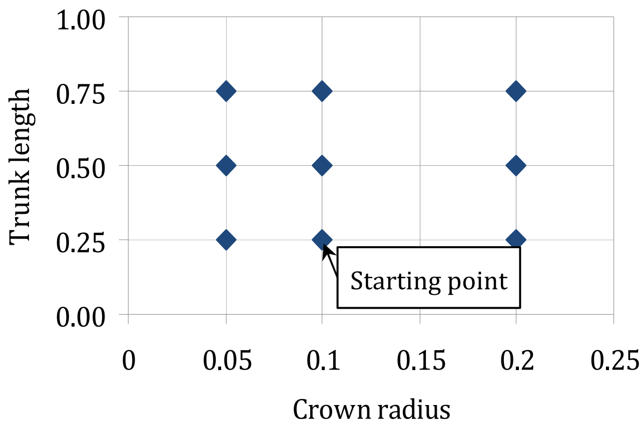


Figure 53: Combinations of tree element parameters used during calibration.

Each run included creating a new obstacle file (step 2), calculating the SVF in SkyHelios (step 3), transforming the SVF values to points along the road centerline (step 4), and finally plotting and comparing the values to the reference data. Both SVF calculation methods provided by SkyHelios are used during this process ('planar' and 'spheric'); the choice of calculation method is left up to Meteo Consult.

Results

After nine runs, two good combinations of tree element parameters were found; the combinations indicated as 'A' and 'B' in Figure 54. When applying the 'planar' SVF calculation method combination 'A' gives the best result (see Figure 55). Combination 'B' comes closest to the reference data when applying the 'spheric' calculation method (see Figure 56).

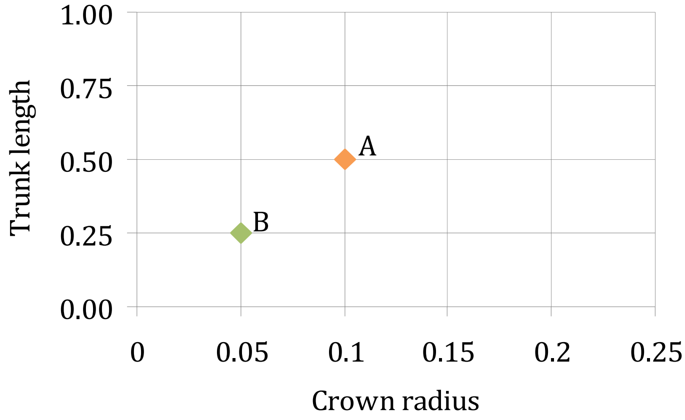


Figure 54: Combinations of tree element parameters that provide a well fitting result

The calculated SVF does not coincide with the reference data along the whole stretch. As can be seen both figures below, the similarity very strong along the 100 meters in the center of the road stretch (between 30.15 – 30.25 km), but not at the beginning (up to 30.15 km) and the end (around 30.28 km). The aerial photograph of the location (Figure 49) provides an explanation for the inconsistency at the beginning; the patch of forest causes it. The peak near the end cannot be explained directly. The section ‘Additional surroundings’ in paragraph 4.3 takes a closer look at how the inclusion of trees adjacent to the provincial road management zone affects the SVF.

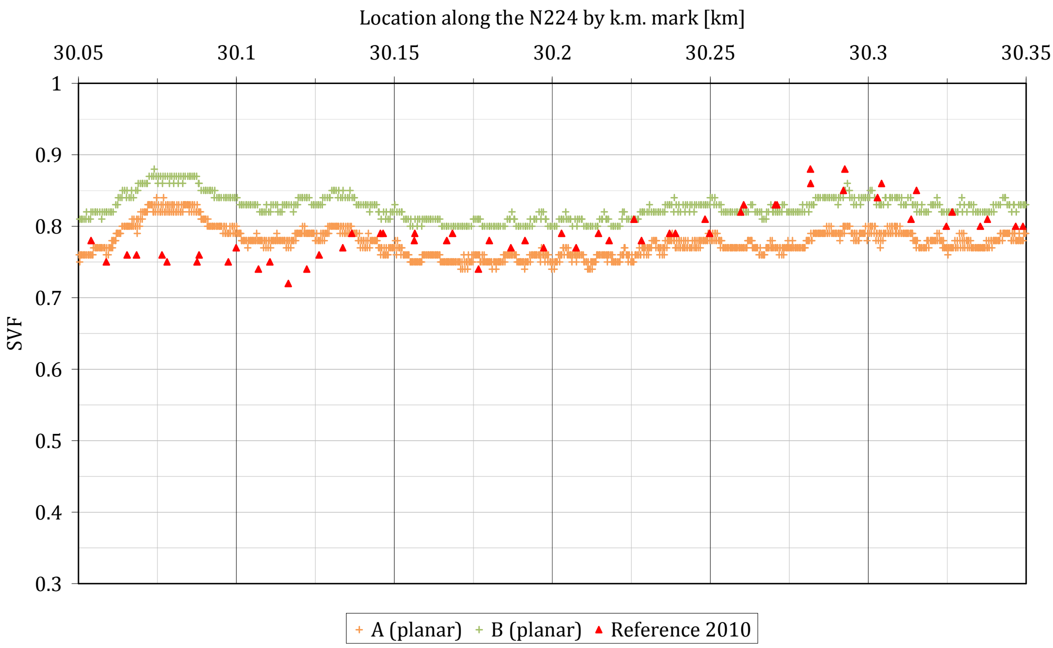


Figure 55: The SVF output of the best parameter combinations compared to the reference data. (SVF calculation method = ‘planar’)

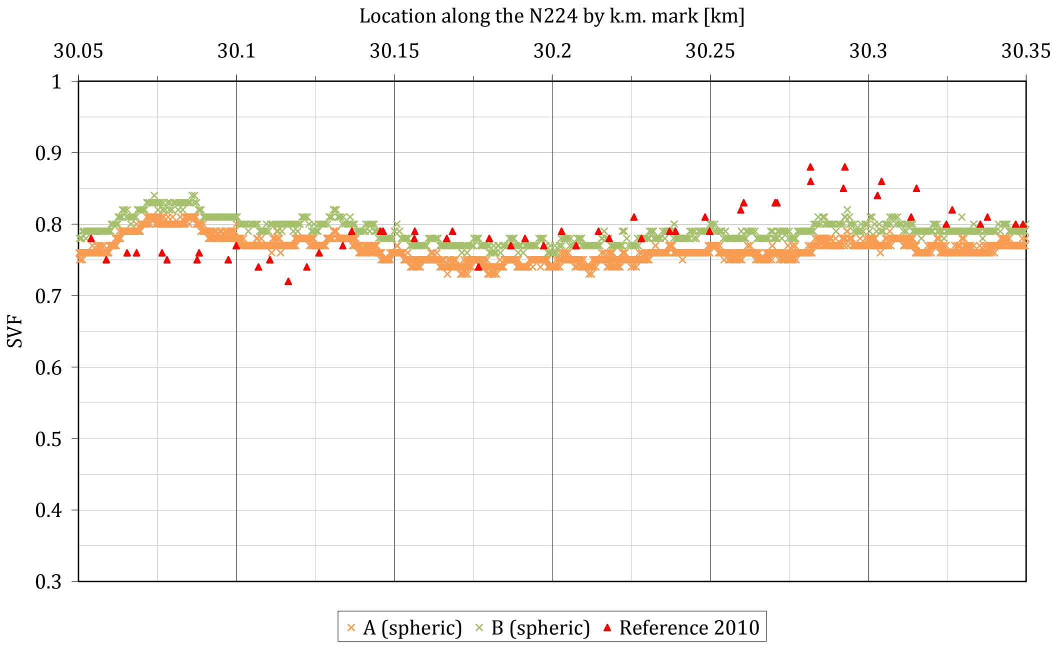


Figure 56: The SVF output of the best parameter combinations compared to the reference data. (SVF calculation method = 'spheric')

4.3 Model sensitivity

Tree elements parameters

During calibration, it became clear that the crown diameter had a large influence on the resulting SVF. Changing the trunk length did not have such a large effect on the outcome. This will be demonstrated by two examples.

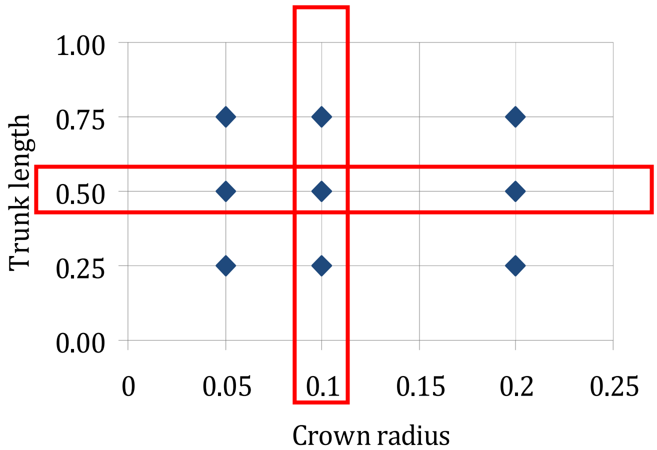


Figure 57: The combination of parameters used for the examples

The graph below shows the SVF when the same trunk length parameter is applied in combination with three different crown radius parameters. In the middle of the graph (30.2km) the SVF difference is already around 0.1 as the parameter changes from 0.05 to 0.1. As the parameter doubles from 0.1 to 0.2, the SVF difference also doubles (at this location).

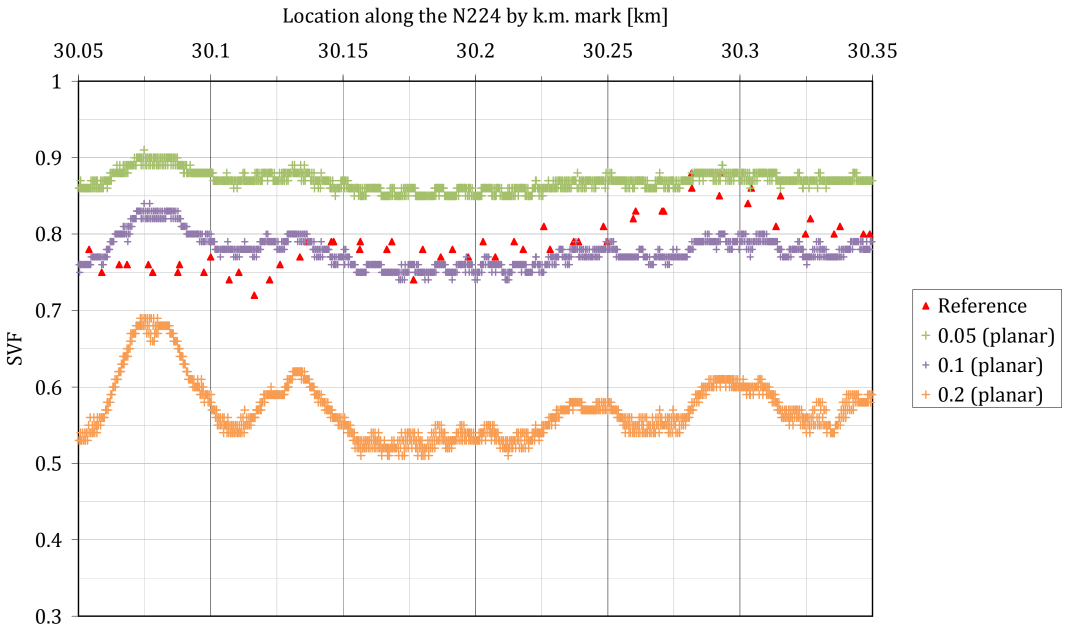


Figure 58: The SVF output when using three different crown radius parameters combined with a trunk length parameter of 50%.
(SVF calculation method = 'planar')

Figure 59 shows what happens if the trunk length is varied between 25 and 75% with a constant crown radius parameter of 10%. Although the trunk length changes considerably, the SVF does not.

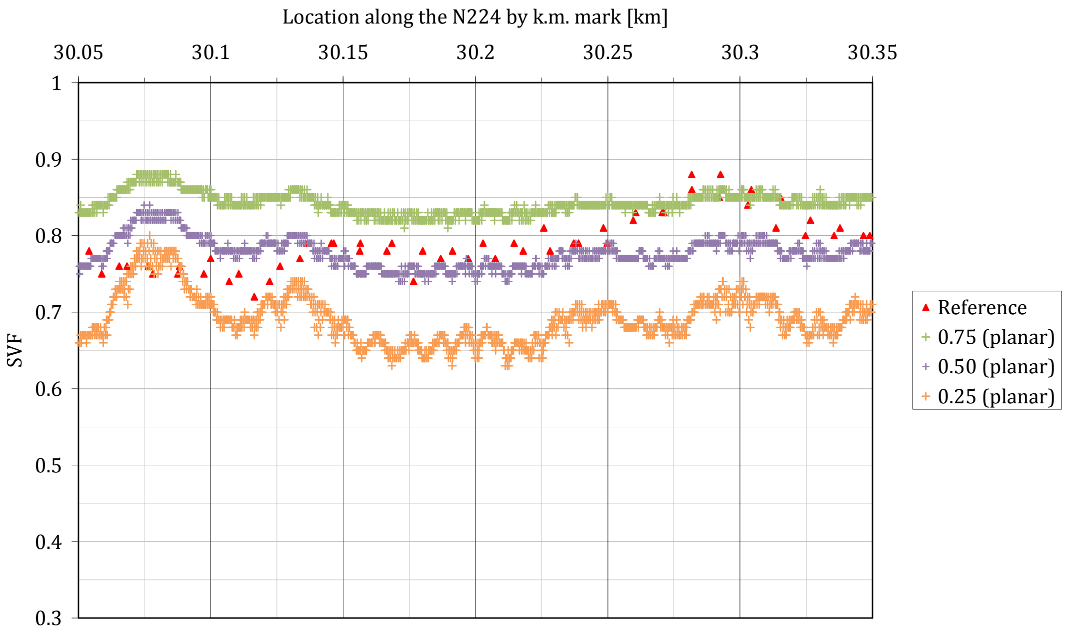


Figure 59: The SVF output when using three different trunk length parameters combined with a crown radius parameter of 10%.
(SVF calculation method = 'planar')

The large influence of the crown radius can be explained by what the sky view factor stands for in the first place. In the introduction the SVF is explained as a value that indicates what proportion of the sky is visible from a given point. Watson and Johnson (1987, p. 193) explain the SVF in more detail, namely as "the ratio of radiation received by a planar surface from the sky to that received from the entire hemispheric radiating environment." When calculating the sky view, the angle of the surrounding objects matter a lot, as Brown *et al.* (2001) illustrate: "For a flat surface at the ground, the incoming radiation from directly overhead spreads out over a smaller area, while radiation coming from near the horizon would spread out over a very large area, making the effective flux of radiation small." Increasing the trunk length leads to more sky obstruction closer to the horizon, as opposed to increasing the crown radius, where the sky is obscured more at an angle close to zenith, which has more influence on the SVF.

This test confirms the assumption that it was valid to leave the trunk diameter at a constant value; the effect of a change in the dimensions of that element would be minimal.

Height data accuracy

In the GIS model the height of a tree defines its shape, which makes it an important value for the creation of the 3D trees. The dimensions of the trunk height, crown radius, and trunk diameter are all derived from the tree height. This section takes a look at how sensitive the model is to changes in the tree height. Specifically, it displays the differences in SVF that can be expected if incorrect tree heights are applied.

At the end of the calibration session a flaw was found in the tree height conversion. According to the model, the maximum value of a category should be converted to the middle value of a category by extracting 2.5m; this had not been implemented correctly. When the error was corrected, the SVF output showed a steady increase of approximately 0.03, as can be seen in the graph below.

Considering that a category has a range of 5 meters, the error within a category is already up to 0.03 (the trees can be either 2.5m shorter or taller).

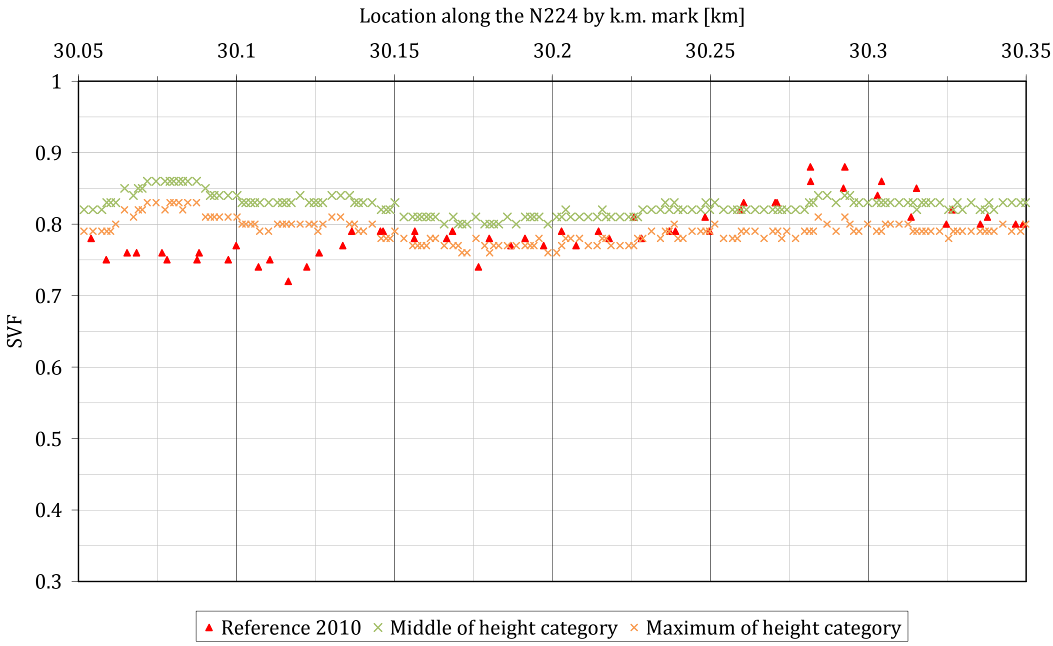


Figure 60: The SVF output increases after extracting 2.5m from all tree heights.
(SVF calculation method = 'spheric', crown radius parameter = 0.05, trunk length parameter = 0.25)

Although this adjustment undermined the parameters found during calibration, it was decided not to redo the calibration process, but rather to check if the correct tree heights were being used in the first place. It would be more interesting to see the effect of different tree heights on the SVF output than to find a new set of well-fitting parameters. Therefore, the heights of the trees were checked and given the height categories they 'should' have had in the provincial dataset. This was done for a number of reasons:

- The tree heights provided by the provincial vegetation dataset are very easy to input in the model; it is the preferred dataset for these data. For that reason it is interesting to see which impact an incorrectly categorized tree has on the results.
- The height categories are assumed to be 'wide' enough for the province to categorize the trees correctly.
- Given that the tree heights will be sampled and thus will not be retrieved for each individual tree, the applied heights will be approximate either way.

The tree heights were retrieved by making transects of the AHN vegetation dataset; several 20-meter wide polygons were made along both validation locations. The transect locations were selected based on the SVF results yielded after running the model with provincial height data. The most interesting locations were the ones where the resulting SVF was stable and deviated most from the reference data. The map in Figure 61 shows an example of a transect and the sampling points it contains.

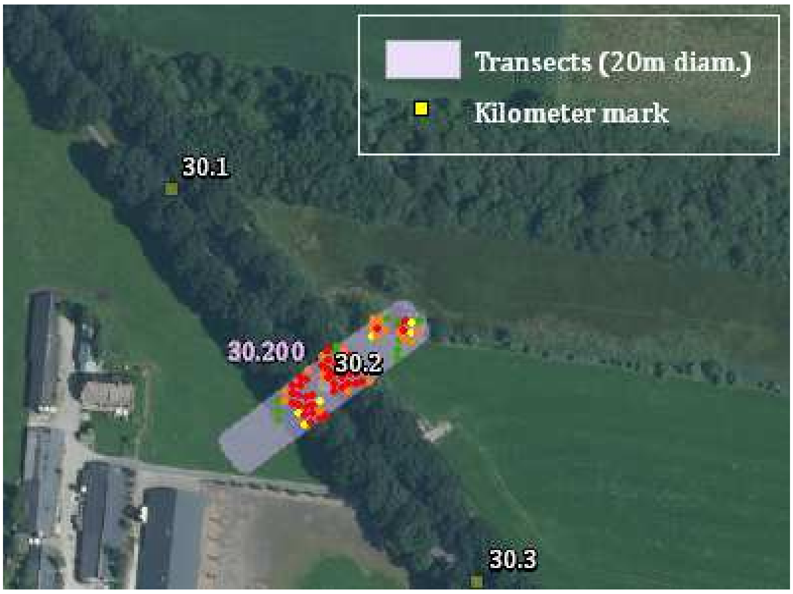


Figure 61: The transect on location 32.2 along the N224, containing AHN-veg elevation measurements

The height data within the transects were plotted according to their position in relation to the road centerline[11], as can be seen in Figure 62. It shows the shape of the trees and gives a good indication of their height. Plots of the transects on both validation locations can be found in appendix IV.

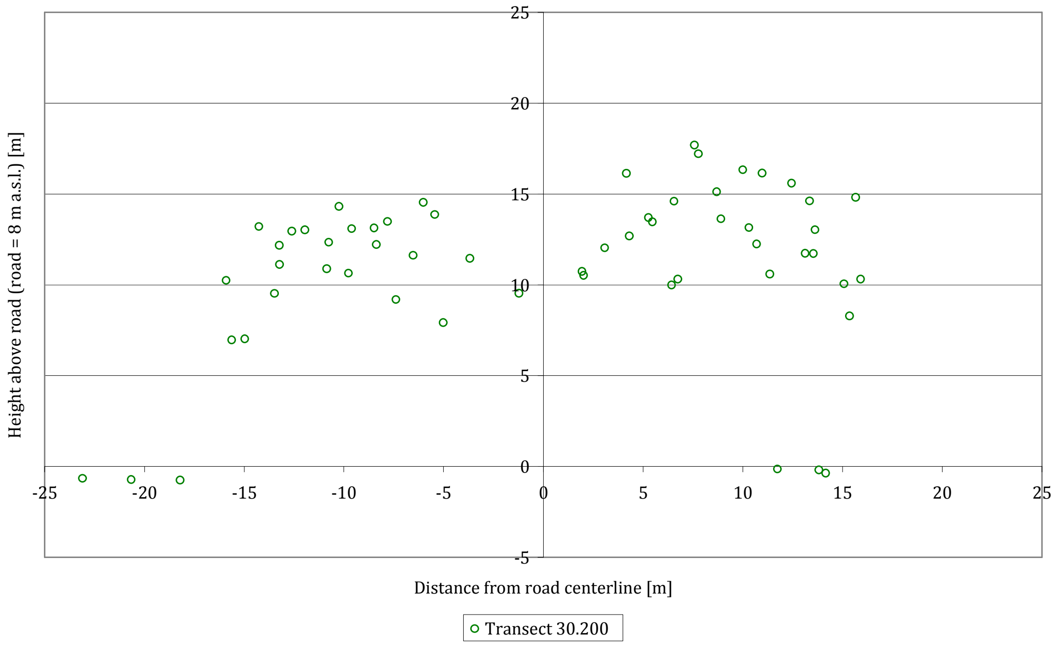


Figure 62: The height and position of the sampling points within the transect at 30.200km along the N224, limited to 25m distance from the road center.
(Right and left in the image correspond with the naming convention of 'right' and 'left' lanes of a road. Here, right is north-east and left is south-west.)

[11] The location of the points in relation to the centerline is calculated using the 'Near (Analysis)' tool in ArcGIS 9.3.

The 'correct' tree category was derived from the transect by looking at the maximum height value. If the maximum height was close to the maximum of a category, like on the left side of the road in Figure 62, the higher category was given. This was done under the assumption that the laser measurements are not intended to measure the height of vegetation, but of the ground; the laser beams are sparse and the small twigs at the top of a tree do not reflect the beams well.

The results obtained by the transects led to a revision of the height category for the vast majority of the trees, as can be seen in appendix V. The categories for the trees located outside the sampling locations were changed based on the plots of neighboring transects and aerial photographs.

For the calibration stretch this revision led to a peculiar outcome; the SVF output was nearly identical to the SVF calculated with the 'erroneous' maximum category values (see Figure 63).

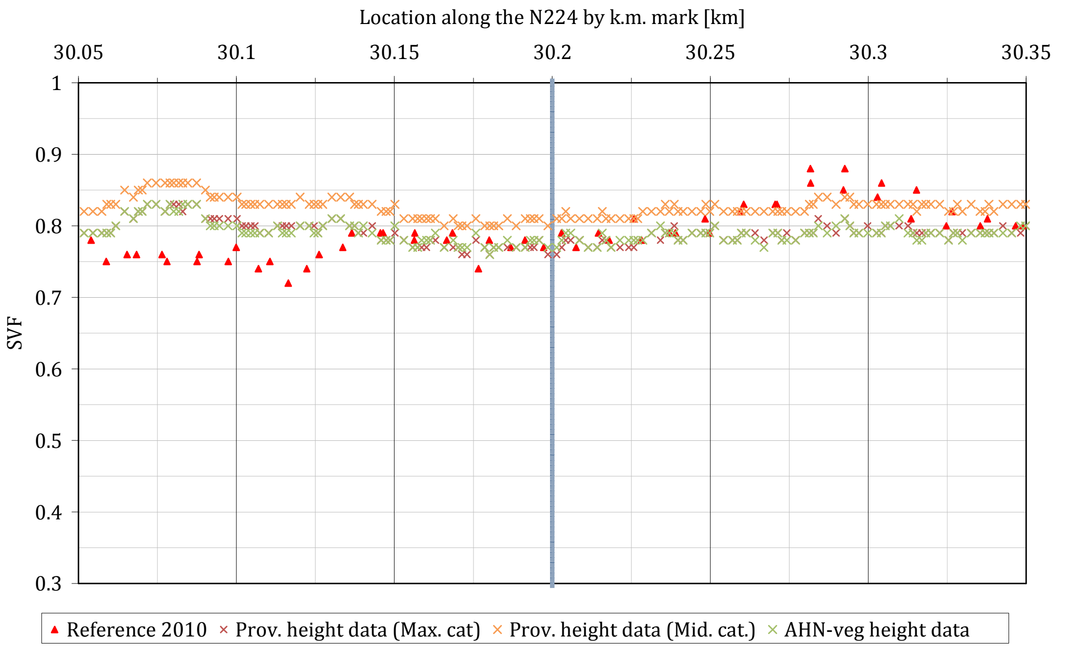


Figure 63: The SVF output using different sources and values for the tree height. The transect at 30.2km is indicated as a blue line.
(SVF calculation method = 'spheric', crown radius parameter = 0.05, trunk length parameter = 0.25)

This effect is caused by the faulty assumption done while creating the basic model: trees without height information are given the lowest category, assuming that they are newly planted and the information is not entered in the system yet. As can be seen in Figure 64, approximately half of the trees on the calibration location have no height data.

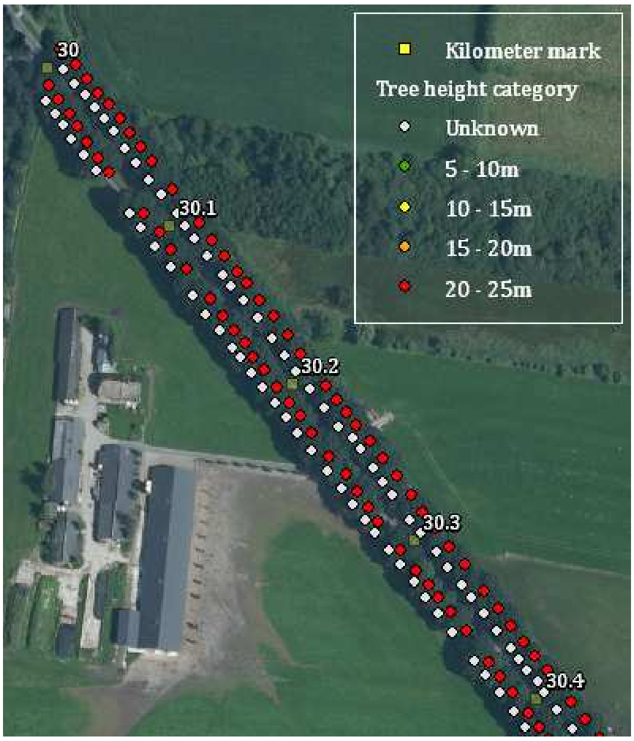


Figure 64: The tree height categories at the calibration location, provided by the province of Gelderland.

According to the AHN height data, the trees are about 15m high (as seen in Figure 62); all trees in the vicinity are assigned to the category 15 – 20m. This means a decrease in height for half of the trees, and an increase for the other half, resulting in approximately the same SVF.

This is good for the usability of the tree element parameter values found during conversion: they still produce SVF output similar to the reference data and do not need to be reassessed.

Revising the tree heights of the calibration location affirmed that correct conversion parameters had been found, but it did not provide a good impression of the model's sensitivity to changes in height. A second example is presented to get a better impression of the effect that concrete height changes have on the SVF.

Figure 65 shows the change in SVF after revising the height categories along the whole second validation location. This location was chosen to show as an example since all trees were considered to be in the category 20 – 25m according to the provincial vegetation dataset.

West of transect 32.496km (indicated as a blue line in the graph below), the height was reduced by two categories (i.e. 10 meters). This caused the SVF to rise by approximately 0.12. East of said transect, dropping the heights by one category (i.e. 5 meters) resulted in an SVF increase of around 0.6. The shifts are of the same magnitude when applying the planar calculation method.

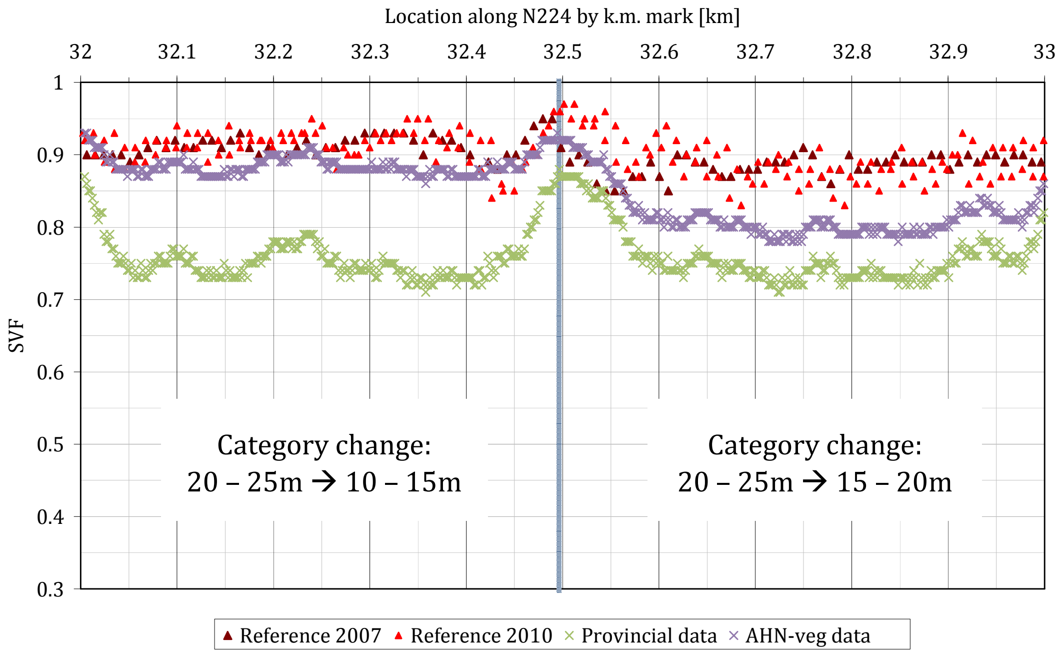


Figure 65: The SVF output using different sources for the tree height categories. The transect at 32.496km is indicated as a blue line.
(SVF calculation method = 'spheric', crown radius parameter = 0.05, trunk length parameter = 0.25)

The examples presented in this sensitivity test have shown that the model is quite sensitive to height changes; the SVF changes linearly by roughly 0.01 for every meter of change in the tree height. The tests also showed that the tree heights indicated in the provincial vegetation dataset are not correct. It seems that the data is stored more accurately than it is acquired.

Additional surroundings

The model only uses a DEM and provincial trees as input; other road surroundings that also might have an impact on the SVF are not taken into account. The calibration session showed that disregarding the presence of a forest has a great impact on the results. A small sensitivity test was done to see what the outcome of the model would be if this patch of forest was taken into account.

Preferably, a polygon from the Top10NL should be used to represent the forest. SkyHelios can import these polygons and extrude them to 2.5D blocks according to their height. However, the import of a shapefile in SkyHelios did not succeed successfully, so the trees were added manually (see Figure 66).

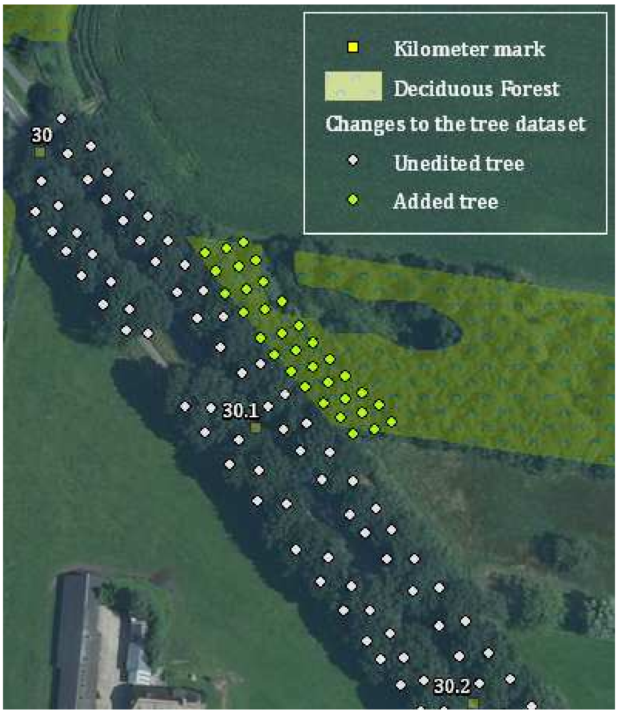


Figure 66: Changes made to the tree dataset.

The addition of extra trees led to a far better result than expected. As can be seen in the graphs below, the SVF even formed a perfect match to the reference data when calculated by the 'spheric' method. Based on this outcome it is strongly advised to extend the model to include relevant adjacent data.

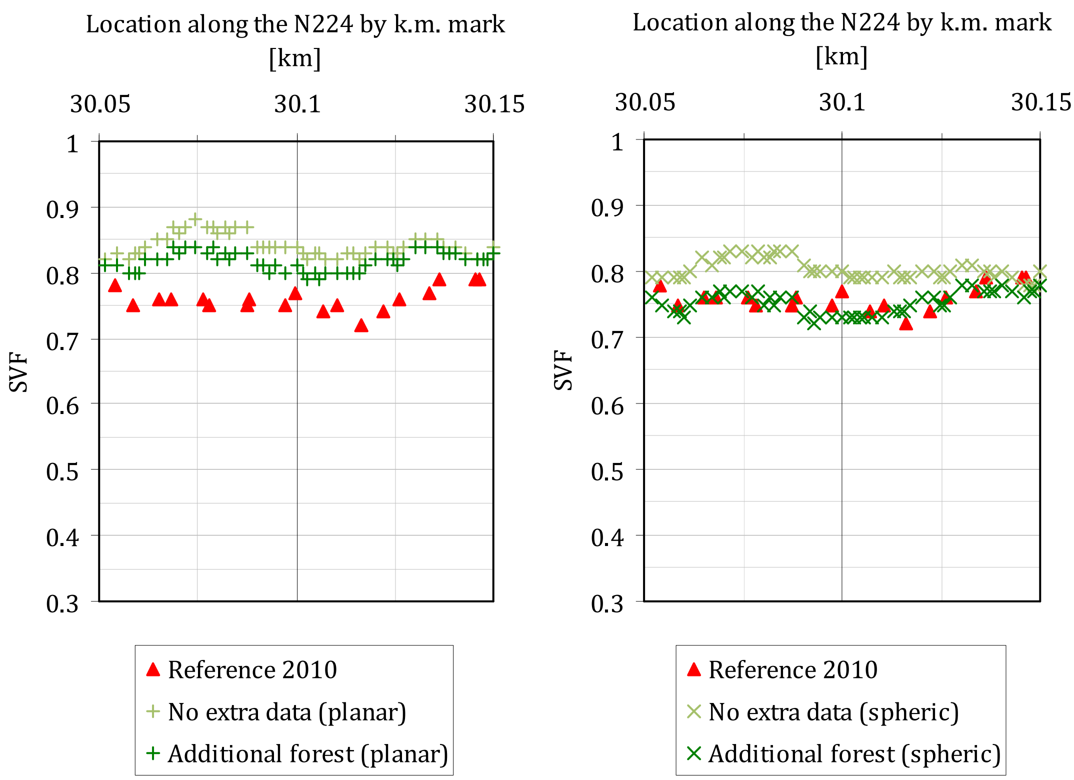


Figure 67: Effect of additional trees on the SVF output.
(All trees are in the category 15 - 20m. Crown radius parameter = 0.05, trunk length parameter = 0.25)

SVF calculation methods

As mentioned in the previous chapter, SkyHelios creates a viewshed with the same projection as is used in a fish-eye photograph. SkyHelios then uses this viewshed to calculate the SVF according to two methods, 'planar' and 'spheric'. Meteo Consult uses the SVF calculation method of 'Chapman' (Chapman & Thornes, 2004). The developer of SkyHelios does not go into detail on the differences between these methods, so the underlying formulas cannot be compared. To allow comparison of the three calculation methods anyway, Meteo Consult calculated the SVF based on images of manually exported viewsheds. The graph below shows that the 'Chapman' SVF on location 32.2 has a great resemblance to the 'planar' calculation method. The SVF on location 32.5 lies between 'planar' and 'spheric'.

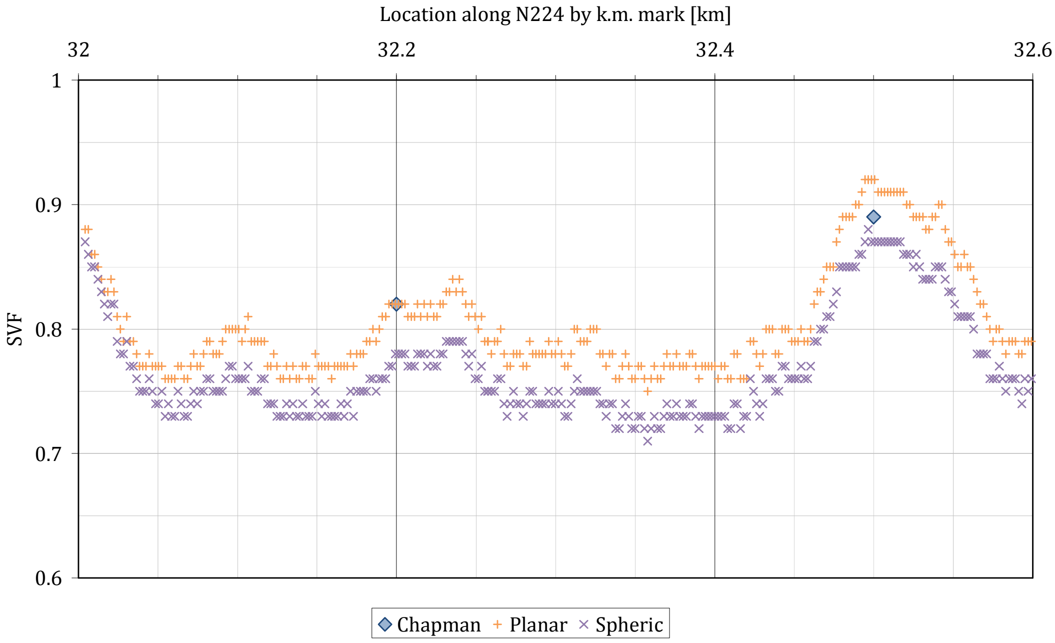


Figure 68: The SVF along a portion of the N224 calculated using three different methods. Only two locations were calculated using 'Chapman', at 32.200 and 32.500km.
(All trees are in the category 20-25m (according to provincial data), crown radius parameter = 0.05, trunk length parameter = 0.25)

To enable a better comparison, it is advised to study the underlying formulas, assuming that the developer is willing to release them. Collecting more sampling points is also a possibility, but this is strongly discommended due to the time it takes to pinpoint the exact position where the sample should be taken.

4.4 Validation

The aim of the validation is to see if the model can output sky view factors that are similar to the reference data along road stretches with similar surroundings to the calibration location. The model is applied using the tree height data derived by sampling the AHN vegetation data. The results are compared to the reference data by SVF plots and statistical analyses.

Locations

Two locations along a provincial road were chosen to validate the model. These locations have similar surroundings as the calibration location; double tree rows consisting of oak trees in a rural area. This choice was made to validate the assumption that the model works for road stretches with similar surroundings to which it is calibrated to work.

The first location is an extension of the road stretch used for the calibration. It has a total length of 1.4 kilometers along the provincial road N224 between Renswoude and Veenendaal de Klomp (HM 30.0 – HM 31.4) (see Figure 69). For the largest part of the stretch it is surrounded by two tree rows on each side. Extra tree rows and buildings are also present. The SVF varies quite a lot and it shows an upward trend from the beginning to the end (see Figure 70).

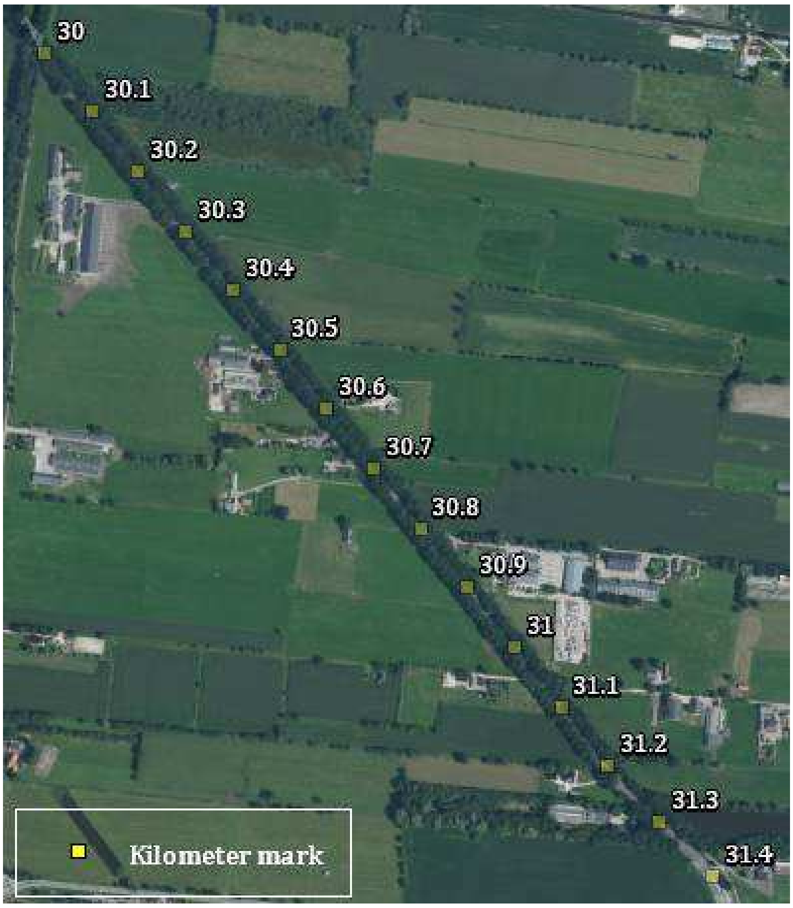


Figure 69: Overview of the first validation location

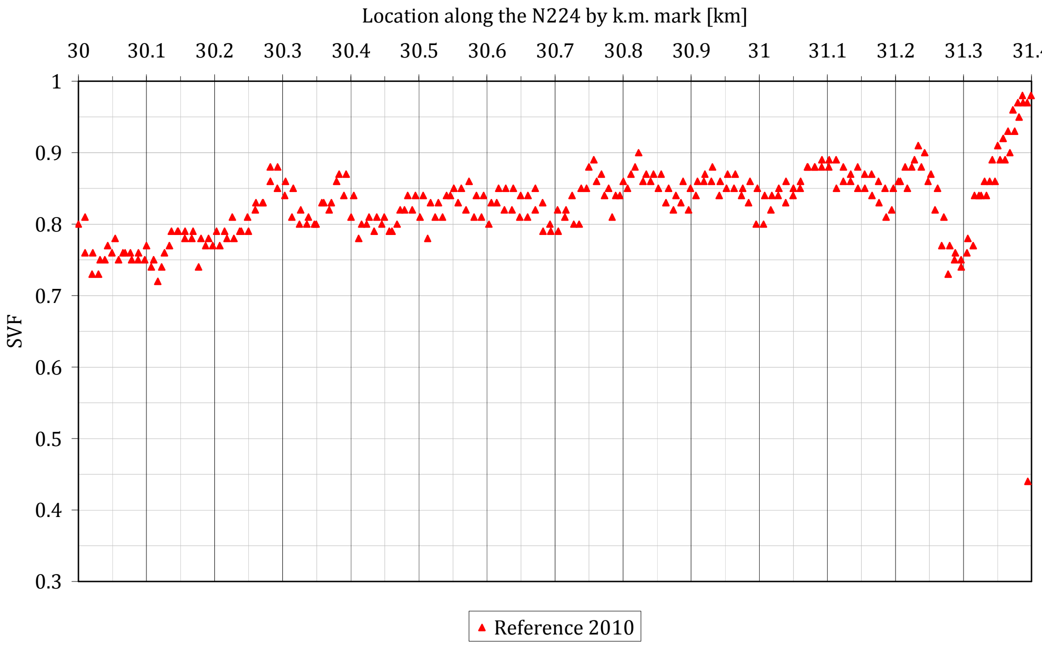


Figure 70: The SVF on the first validation location

The second validation location is a portion of the N224 between Veenendaal de Klomp and Ede (HM 32.0 – HM 33.0), just a few hundred meters from the first validation location (see Figure 71). It was chosen due to its regular tree rows, same tree species, and due to the presence of three different SVF situations: a regular trend in the beginning, an open area halfway, followed by an irregular trend until the end (see Figure 72). This made it an interesting case for testing the accuracy of the model.

Figure 71: Overview of the second validation location

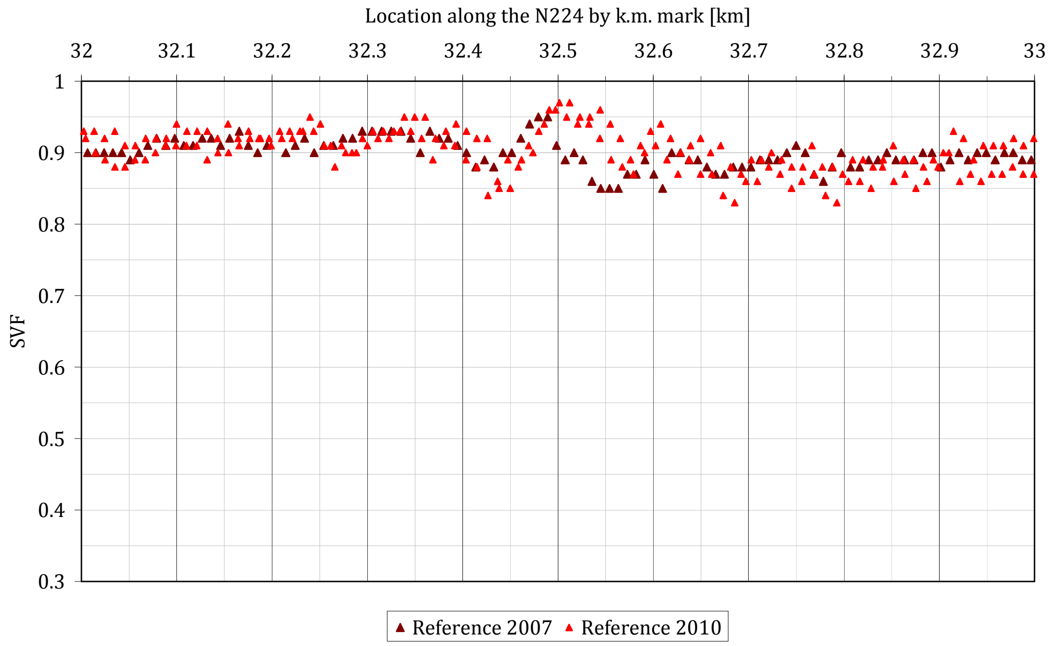


Figure 72: The SVF on the second validation location

Approach

For the validation, the model was executed once. The model was applied the same way as explained in the calibration session; however, a vital change was made with regard to the input data: a different source was used for the tree heights (see Figure 73). The tree heights were obtained by sampling the AHN, as addressed in section 'Height data accuracy' in paragraph 4.3. The transects used for the sampling can be found in appendix V. The tree heights derived from the transects are displayed in appendix IV.

The SVF is calculated using the 'spheric' calculation method, because this method provided the best fit during calibration. Due to the late discovery of the similarity between Meteo Consult's 'Chapman' method and the 'planar' method, the 'planar' calculation method is not applied, as that would require the use of a different combination of conversion parameters.

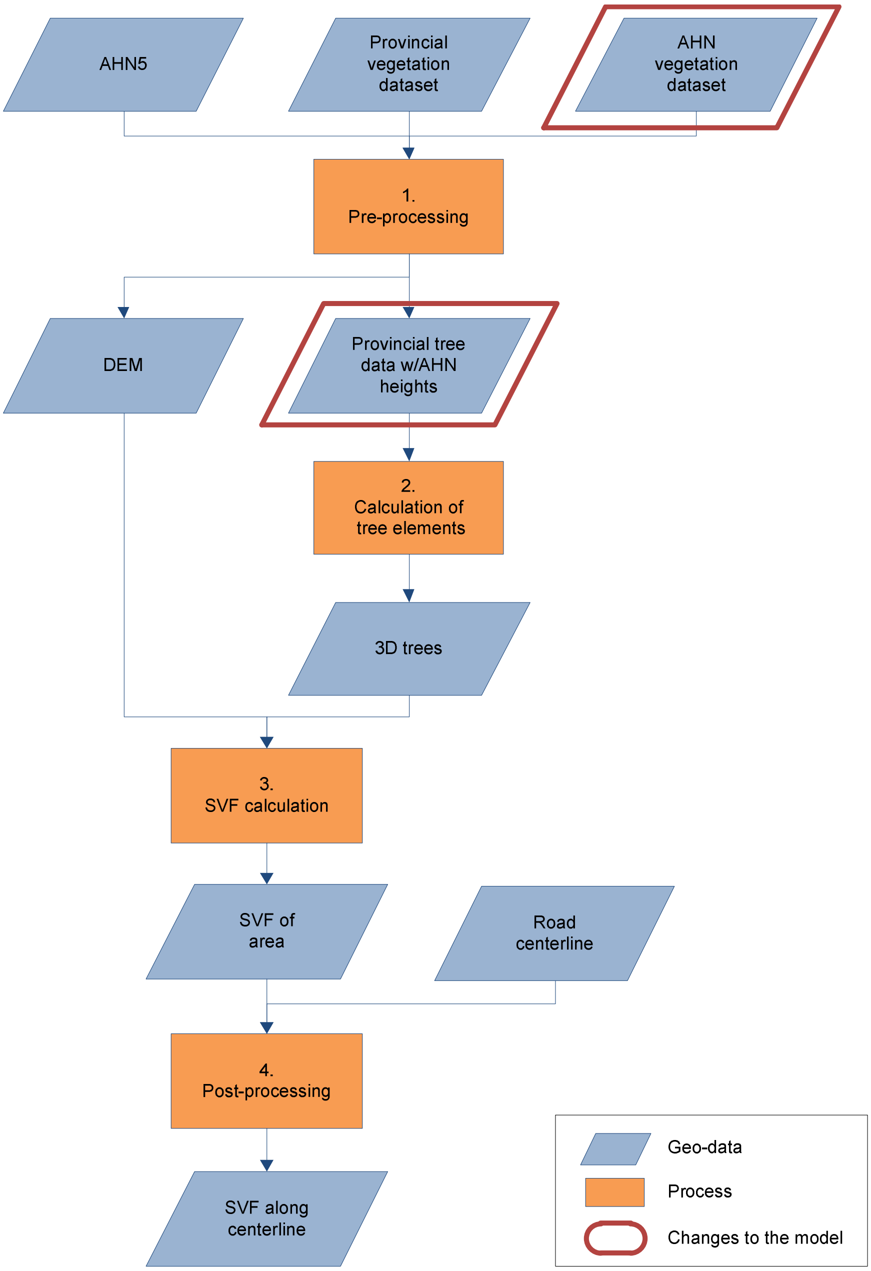


Figure 73: The workflow of the validation process. In addition to the changes made to the basic model during calibration, an extra data source is added and used to determine the tree height categories.

Results and discussion

When looking at the results, it should be kept in mind that the aim of developing the model is not to produce sky view factors that can simulate the road surroundings extremely well (in order to match the reference data). Rather, the model is developed to provide a proof-of-concept; it is more important to match the trends of the reference data. Furthermore, the anomalies in the output should be accounted for in order to identify which aspects of the model require extra attention if it is decided to put the model to use.

Anomalies will be looked at in detail using fish-eye photographs and sources presented earlier in this report if needed.

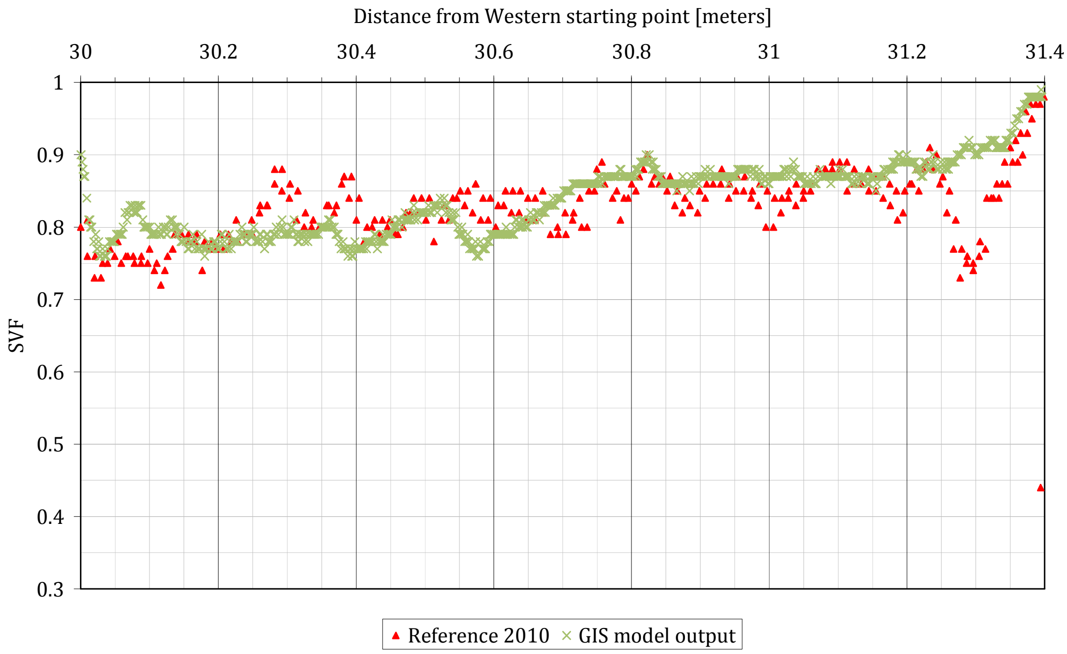


Figure 74: The SVF on the first validation location, based on tree heights derived from the AHN vegetation dataset.
(SVF calculation method = 'spheric', crown radius parameter = 0.05, trunk length parameter = 0.25)

At the first validation location the model output matches the reference data fairly well (see Figure 74). This is a pleasant result as the 'spheric' method also produced the best results during calibration (in combination with the applied conversion parameters). Generally, the SVF calculations follow the same trend as the reference data, but local variations in tree sizes cause several overestimations (at 31.0km for example) and underestimations (like at 30.3km, 30.38km, and 30.57km). As expected, the SVF is severely overestimated at locations where adjacent tree rows and forests are not taken into account (e.g. in the vicinity of 30.1km, 31.2km, and 31.3 km).

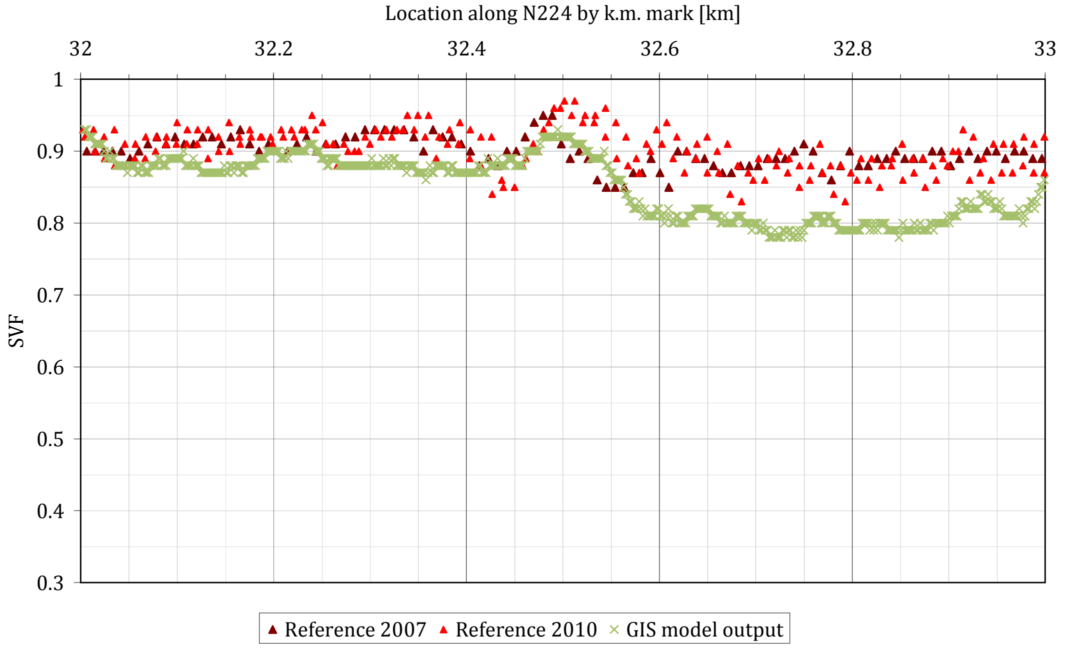


Figure 75: The SVF on the second validation location, based on tree heights derived from the AHN vegetation dataset.
(SVF calculation method = 'spheric', crown radius parameter = 0.05, trunk length parameter = 0.25)

The calculated sky view factors not fit well on the second validation location (see Figure 75). Considering the fact that the reference data from 2007 has been corrected for changing light conditions and the 2010 dataset has not, more importance should be given to how the output compares to the values from 2007. As this location is a part of road stretch E, the effect of not correcting for brightness is documented in comparison 2 and comparison 4 in chapter 2.3.

The result follows the same trend as the reference data, but it is constantly lower, especially from 32.55km onwards. The underestimation can be caused by incorrect appointment of new tree height categories. A look at this road stretch using the AHN vegetation data (using Figure 76), reveals that hardly any trees from 32.6km onwards are higher than 15 meters, while they have been categorized to be within 15 – 20 meters (see Figure 96). As seen in the sensitivity test 'Height data accuracy', wrongly assessing the tree heights by one category can result in an SVF difference of around 0.6, which indeed seems to be the case here.

The density of the points in the map in Figure 76 suggest that the trees along the first part of the road stretch (up to 32.4 km) have less dense tree crowns than the trees along the second part of the road stretch (32.6 onwards). This explains the higher SVF along the first part and the higher fluctuation of the SVF along the second part (as the SVF tends to fluctuate more as the SVF decreases).

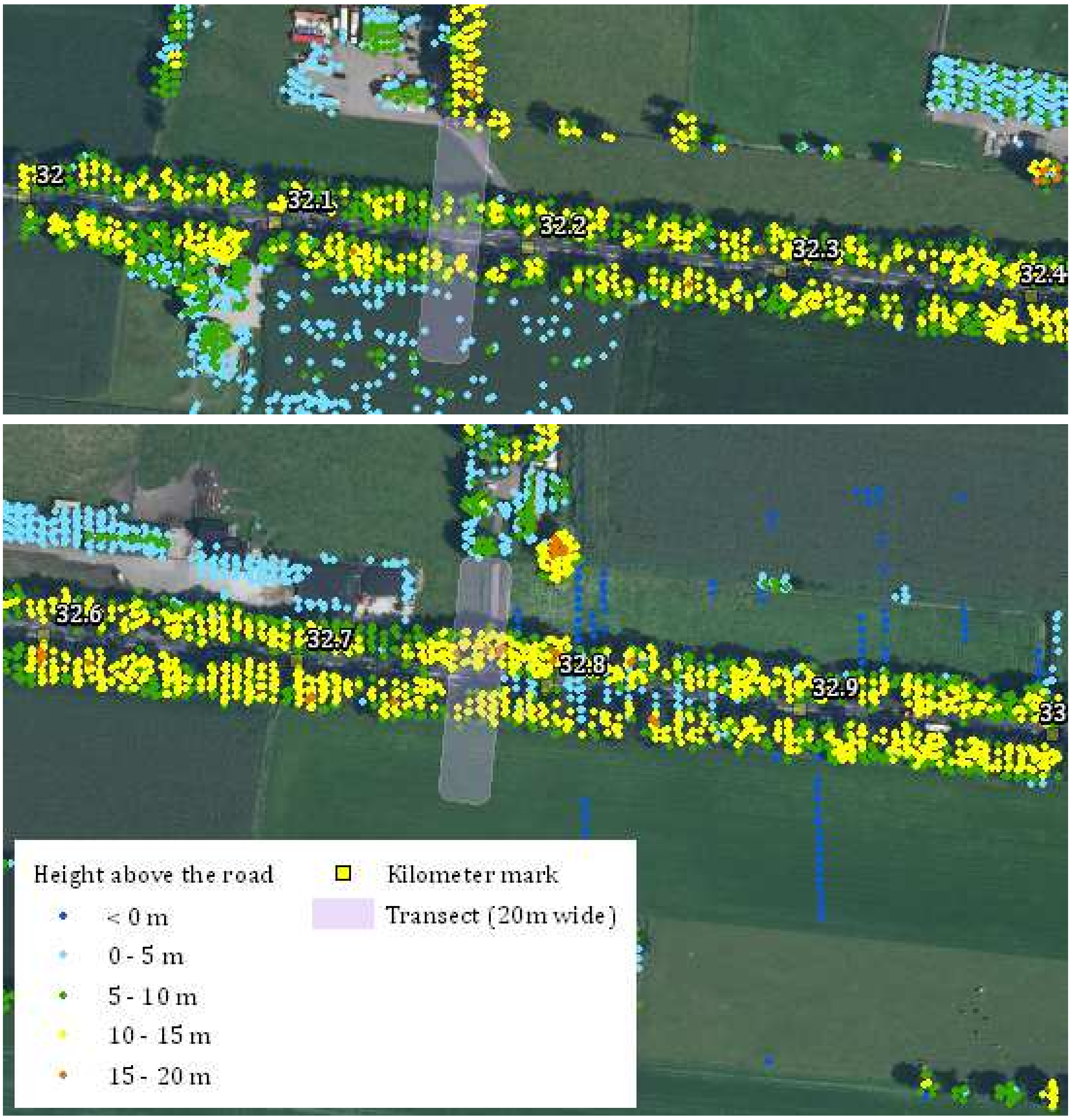


Figure 76: The heights of the AHN vegetation dataset along the beginning (top figure) and end (bottom figure) of the second validation location. The legend presented in the bottom figure also represents the features in the top figure.

4.5 Conclusions and recommendations

In this chapter the basic model was adapted to create trees that would simulate the actual sky obstruction at a certain location. The model was calibrated by changing the parameters for crown radius and trunk length until a good match to the reference data was found. Afterwards, the model's sensitivity was tested to see to what extent certain changes in the model would affect its output. The validation at the end proved that the model performs very well under the same conditions as it was calibrated for.

In reply to the research question "*What is the quality of the GIS model?*" it would be appropriate to say that the quality of the GIS model is quite high, but improvements are needed for the model to function more accurately and under different circumstances than it was calibrated for.

Applicability

The model was calibrated to mimic specific surroundings, namely a provincial road surrounded solely by double rows of oak trees. Since it was also tested along provincial roads with the same surroundings, it is not known how the model behaves when the surroundings are different.

As addressed in the previous chapter, the model only uses ground elevation and individual 3D trees as input; other surroundings that might impact the sky view are not taken into account. This chapter demonstrated the impact that surroundings adjacent to the direct road surroundings could have in sensitivity test 'Additional surroundings'. Since the impact proved to be high, it is strongly recommended to extend the model to include more surroundings.

Sensitivity

The model creates 3D trees based on one value: the tree height. The tree height determines the shape of the 3D tree as the dimensions for the crown diameter, trunk length and trunk diameter are derived from it. During the sensitivity tests it became clear that the three height impacts the SVF the most (see section 'Height data accuracy'). The model validation confirmed this. Considering the large effect that the height has on the SVF, the accuracy of the model could be improved by using the 'real' height of each tree as input, instead of its category. For now, it is advised use the AHN vegetation dataset as the source for the tree height.

The second most influential tree element on the SVF is the crown diameter. The AHN vegetation dataset can also be used to obtain this value, instead of letting it be dependant on the tree height. The current version of the AHN might not provide enough elevation points to derive these dimensions, but the new version of the AHN, the AHN-2, most probably will: it provides up to 30 points per square meter as opposed to one point per 25 square meters[12]. Terrestrial LIDAR might also provide a solution as it yields high resolution point clouds. However, this method requires expensive equipment and surveying on location.

For the purpose of validating the model, the tree heights were derived from the AHN vegetation dataset by taking samples on several locations. The heights found in the samples were categorized and assigned to the trees in the vicinity. This method was not only laborious; it also yielded very inaccurate trees heights. For these reasons a method should be developed that assigns the heights in AHN-veg dataset to the individual trees found in the provincial vegetation dataset.

The parameters used to create 3D trees are specifically suited to create solid 3D tree models that mimic leafless oak trees, but are probably not suited to represent trees with different shapes and transparencies. Further research is needed to confirm this.

As seen during calibration and sensitivity test 'SVF calculation methods' the choice of SVF calculation method also has a significant impact on the outcome. For that reason a decision should be made as to which calculation method to apply before calibrating the model on new locations with new surroundings.

---

[12] Het Waterschapshuis - Actueel Hoogtebestand Nederland. (Unknown publishing date). *Techniek van inwinning en productie van het AHN ('Method of acquisition and production of the AHN')*. Retrieved April 1, 2011, from http://www.ahn.nl/wat_is_het_ahn/techniek_van

5. DISCUSSION

This chapter summarizes the points of discussion raised in chapters 2 through 4. Consult chapters 2.3, 3.5, and chapters 4.2 through 4.4 for more detailed information.

The objective of this research was to improve GIS modeling of the SVF along roads. This was done by investigating the quality of reference data produced using the photographic method first (in chapter 2), followed by the development (chapter 3) and application (chapter 4) of a model that calculates sky view factors based on geo-data. Even though it was difficult to judge the quality of the reference data, it was possible to create a model that calculates sky view factors that resemble the values produced by the photographic method.

Quality of the reference data

The investigation of the reference data was impaired by several issues:

- In order to spot processing errors or changes in the surroundings, it would be ideal to compare photographs taken on the exact same spot on the road. Due to the large GPS positioning error (see Table 2), it was not possible to select photographs taken at the same location. These were found by coincidence.
- The fish-eye photographs taken in 2010 were processed differently than the ones taken in 2007. The 2007 dataset was processed meticulously, determining the threshold value to create binary images for small sections of the survey at a time. To lower the amount of processing time for the 2010 dataset, this threshold was set at the same value for all photographs, based on favorable results obtained during preliminary testing. However, the results of all comparisons done in 'Step e. Analyzing the differences in SVF using four comparisons' show that a slight change in light conditions between compared photographs could produce a large difference in SVF.
- The comparisons for the road surroundings were too general, due to the use of very long road stretches (like road stretch B, C and D, see 'Step a. Selecting comparable SVF data') and broad classes ('tree rows', 'forest' and 'mixed').

Due to these issues, the results of the statistical analyses (provided in Table 4) should be interpreted with caution; based on these values it is possible to draw conclusions regarding the extent of SVF fluctuation under certain general conditions, but not to draw direct conclusions on the quality of the reference data itself.

Road segments

Initially, the research aimed at calculating the SVF of road segments, as opposed to calculating it per point, like is done using the photographic method. While researching the quality of the reference data it became clear that it would be very hard to define classes or categories of road surroundings to which a SVF value could be assigned. Moreover, considering how salting precision is expected to improve in the future (Aebi Schmidt Nederland, September 2010), it seemed reasonable to abandon this direction of investigation. As a result, research question 2 was dropped and the research shifted in the direction of developing a model that calculates the SVF for point locations.

Sun view factor

The focus while developing a GIS model was to be able to calculate the SVF (and sun view factor) along rural roads, where trees are the main cause of sky obstruction. As mentioned in chapter 3.3, 'Design', the developed GIS model meets all design criteria but one: it is unable to calculate the sun view factor. The program selected to calculate the SVF, SkyHelios, only supports manual lookup of the sun view factor. Unfortunately, the program is not open source, so this function as well as other possible improvements, cannot be implemented by Meteo Consult themselves. All improvements required in SkyHelios can be found in chapter 3.6.

Model input

As mentioned above, the model calculates the SVF along rural roads. The 3D surroundings on which it bases its calculation is based on two types of surroundings: ground elevation and individual trees. Due to time restrictions other features that have a significant impact on the SVF, like forests and overpasses, were not taken into account. To get an impression of the effect of including additional surroundings nonetheless, a sensitivity test was conducted (see section 'Additional surroundings'). This test confirmed the assumption that it would be valuable to include additional data on the road surroundings.

During calibration and testing of the model (chapters 4.2 and 4.3 respectively) it was discovered that not only did many of the trees in the provincial vegetation dataset lack additional data (like height and species), the height categories provided were proven to be incorrect when compared to the AHN-veg dataset. Therefore, the heights derived from the AHN vegetation dataset were used during validation (chapter 4.4). This was done by sampling heights of a few trees by categories and assigning these height categories to trees in the vicinity. A more accurate method would have been to obtain the exact height of each individual tree, but this was considered too laborious. More on recommended use of datasets can be found in chapter 4.5, 'Conclusions and recommendations'.

Model performance and applicability

In chapter 4.2 the tree element parameters were calibrated to mimic oak trees along a provincial road as well as possible. As was also demonstrated during validation (in chapter 4.4), the model performs fairly well for roads with these conditions, but it is not known how the model behaves for different surroundings.

The sensitivity test of the tree element parameters in chapter 4.3 showed that the model is most sensitive to changes in the tree height and the crown radius. This can be explained by the fact that an object closer to zenith provides a spot on the road with more radiation than an object closer to the horizon does, because its radiation gets less dispersed. This sensitivity test confirms that it was valid to leave the trunk diameter at a constant value during calibration.

6. CONCLUSIONS

At the beginning of the thesis, the following research question was formulated:

Can a GIS model produce sky view factors for road segments of a quality equal to that of the photographic method?

This question will be answered at the end of this chapter, after answering the four questions below. As these questions have been answered in chapters 2 through 4, this chapter will provide a summary of these findings.

1. What is the quality of the reference data; is there a significant difference in the sky view factor when comparing fish-eye photos taken in the same area, at different moments in time (in the same season of the year)?
2. What is the definition of a road segment and how can it be created using geo-data?
3. Can the sky view factors of road segments be calculated using a GIS model?
4. What is the quality of the GIS model?

Question 1: What is the quality of the reference data?

Chapter 2 assessed the quality of the photographic method so that the outcome could be used during development and quality analysis of the GIS model. Reference datasets (fish-eye photos accompanied by their derived values, like SVF and sun view) made in 2007 and 2010 were subjected to different comparisons that revealed the influence of the following circumstances on the SVF:

- Road surroundings

  The SVF fluctuates most when the road surroundings are both heterogeneous and close to the road. Large SVF fluctuations are most common along provincial roads closely surrounded by forest and heterogeneous tree rows (as seen in 'Comparison 2: Provincial road – Opposite lanes' for road stretches B, D, and E). Overpasses were the main cause of large SVF fluctuations on the highway (see 'Comparison 1: Highway – Opposite lanes' and 'Comparison 3: Highway – Same lane').

  Due to their high influence on the SVF, these road surroundings should be modeled as precisely as possible.

- Different lane or year

  When comparing the SVF of fish-eye photographs taken in different years or in opposite lanes it is only possible to detect deviations if they cause a considerable trend shift. This is caused by the relatively large processing error.

  In view of this result, is necessary to use the SVF of both driving directions when calibrating the model along a road stretch with asymmetrical surroundings.

- Processing

  The photographs taken in 2010 were processed using the same threshold to create binary images (i.e. images used to calculate the SVF). The effect that this would have on the resulting SVF was severely underestimated during preliminary testing. During the comparisons in chapter 2.3 many examples were found where the large difference in SVF was caused by a small difference in light conditions (for example point set 3 in Table 9). Generally speaking, this processing error was larger when the SVF was lower.

In addition to the large processing error, two other issues impaired the comparison itself, being:

- GPS positioning error

  A major drawback encountered during the analysis was that it was very hard to find photographs taken close by (thus approximately respresenting the same location). Due to the large positioning error (see Table 2) comparable photographs could not be selected based on the calculated distance between them. Photographs taken nearby were found by coincidence and only these could be used to examine subtle processing errors or minor changes in the surroundings.

- Selection of road stretches

  The comparisons for the road surroundings were very general because of very long road stretches and broad classes (see 'Step a. Selecting comparable SVF data').

Due to these issues it was decided not to use the values of the statistical analysis during the development and quality analysis if the GIS model; SVF plots were used instead. However, the use of SVF plots may also not be a reliable way of displaying data, as the calculation method is susceptible to errors (as shown in Figure 10). Therefore it is recommended to develop a new method for the calculation of the location of SVF points along the road centerline.

Question 2: What is the definition of a road segment and how can it be created using geo-data?

This question was formulated based on the findings of Steenbergen which stated that a lack of geo-data would limit the ability of detailed representation of road surroundings. To circumvent this lack of detailed geo-data, the idea arose to calculate the SVF per road segment (i.e. a longer stretch of road with the same surroundings throughout) in stead of calculating it per point (like is done using the photographic method). During analysis of the reference data this method appeared to be to general as the road surroundings tend to fluctuate a lot and due to the expected improvements to salting precision in the future. For these reasons, the decision was made to attempt SVF calculation per point after all, rendering this question irrelevant.

Question 3: Can the sky view factors of road segments be calculated using a GIS model?

Yes, the developed GIS model can calculate the sky view factor of points in an area (instead of road segments) using geo-data on trees and ground elevation. However, the model has a few disadvantages that must be solved before it can be put to use.

The most important limitation of the model is the inability to calculate the sun view factor. In addition to the SVF, the sun view factor is required as input for Meteo Consult's network model. SkyHelios, the program used to calculate the SVF only supports manual lookup of the sun view factor. Recommendations on how this value can be derived using SkyHelios can be found in the next chapter, in addition to other proposed improvements to the program. Due to the fact that SkyHelios is not open source, changes to the program can only be made by the developer, prof. Matzarakis, researcher at the Meteorological Institute at the University of Freiburg, Germany.

As far as the author is aware of, it is the first time that the SVF is calculated based on 3D trees in areas where trees are the main cause for sky view obstruction. The model defines 3D trees in a way that SkyHelios can interpret them directly, i.e. by storing them in a so called 'obstacle file'. This is a file type native to SkyHelios in which the shape of a tree is defined by five elements (see 'SkyHelios and RayMan' in chapter 3.3). Only the location and the height of a tree are needed to create a 3D tree; during conversion the dimensions of the tree element are yielded by multiplied the height by a certain percentage (see 'Step 2. Calculation of tree elements' in chapter 3.4). Ideas on how the conversion method can be improved can be found in 'Conversion method' in chapter 7.

Question 4: What is the quality of the GIS model?

In chapter 4 the basic GIS model developed in chapter 3 was applied on test locations in order to identify the model's quality and applicability.

The sensitivity tests revealed that changes in the dimensions of the height and the crown radius of the 3D trees affect the SVF output the most. Furthermore, they demonstrated the necessity of including additional road surroundings and deciding on an SVF calculation method.

The validation showed that the model can produce sky view factors similar to those of the photographic method, but improvements are needed for the model to function more accurately and under different circumstances than it was calibrated for. Recommendations on how this can be achieved can be found in the next chapter, in section 'Increasing the model's accuracy'.

Main conclusion

All in all, the developed GIS model produces satisfactory results, but it cannot be used as a replacement for the photographic method.

Unfortunately the quality of the reference data and the GIS model cannot be compared using statistics, as the methodology used during the quality analysis of the reference data was deemed inadequate. Consequently, the model output was compared to the reference data by eye using SVF plots.

The developed GIS model uses an approach to SVF calculation that has not been applied before: it calculates the SVF along roads where only trees obstruct the sky view, simulating trees by a solid 3D tree model.

In order to be able to serve as a replacement for the photographic method, two major improvements are needed. The first is the possibility to calculate the sun view. The second is the development of method to automatically derive tree heights (and possibly other tree characteristics) of individual trees from the AHN vegetation dataset.

Should the two improvements be implemented, a closer look should be taken at the conversion method used to create the 3D trees. In addition, the creation of 3D trees representing leafless trees would probably be easier if the tree model would be extended with a transparency element.

Last but not least, more road surroundings should be taken into account in order to simulate the real situation better. These objects can be derived from the Top10NL and provided with their heights using the AHN.

7. Recommendations

This chapter summarizes the recommendations provided in chapters 2 through 4. For more elaborate information, chapters 2.4, 3.6, and 4.5 should be consulted.

The recommendations are provided for two main parts of the research; the quality analysis of the reference data and the developed GIS model.

7.1 Quality analysis of the reference data

Location of compared points

The quality analysis in chapter 2 was severely hampered by the fact that the photo locations had such a large GPS error that it was not possible to select two nearby points based on the distance between them. This can be solved by either using a surveying method that determined the photo location more accurately or by manually selecting photographs taken at approximately the same location.

Comparing points using SVF plots

Due to the fact that it the distance between two nearby points was not accurate, the results of the statistical analyses were not representable either. Instead of using these values while developing and testing the model, the SVF plots were used. However, the method for creating SVF plots is not always reliable either, as discussed in 'Step b. Enabling point comparison by distance from a common point'. Therefore, it is recommended to find a new method to calculate the location of a SVF point along the road centerline.

7.2 GIS model

As mentioned in the conclusion, the GIS model provides considerably good results, but many improvements are to be made before the model can be used as a replacement for the photographic method.

SkyHelios

The use of SkyHelios is essential in the model, but its use can only be continued if two important functions are implemented: sun view factor calculation and smooth handling of shapefiles.

It is recommended to implement sun view calculation in one of the following ways:

- by exporting the view shed image of every point of SVF calculation, so that sun view calculation can take place outside SkyHelios (a function that already seems to be implemented, but not activated, according to Figure 46);
- by letting SkyHelios calculate the sun view factor of every hour of a whole year for every location it calculates the SVF for (see Figure 46).

SkyHelios is not capable of displaying shapefiles correctly if any of the cells in the DEM has a cell that does not contain height data. This problem can be circumvented by interpolating the DEM to fill the 'NoData' cells, but it is recommended to contact the developer on this issue.

A third improvement suggested for SkyHelios, is to be able to limit the area of SVF calculation. Restricting the area of SVF calculation to only the surface of the roads or points on the road would save a considerable amount of processing time.

As the program is not open-source, the suggested improvements can only be implemented by the developer, prof. Matzarakis of the University of Freiburg.

Increasing the model's accuracy

The model's accuracy can be increased by making improvements in four areas, being:

- the input data;
- the conversion method when creating 3D trees;
- the tree model.

Input

The model relies heavily on accurate geo-data as input. By using the AHN to derive the dimensions of single trees and by including more road surroundings, the accuracy can be improved significantly.

The sensitivity tests in chapter 4.3 showed that the SVF is impacted most by changes in the tree height and crown diameter. During validation, the height of the trees was derived from the AHN vegetation dataset by sampling, assigning the found height to the surrounding trees using height categories. This method was laborious and yielded very inaccurate tree heights. Therefore it is recommended to develop a new method that derives the height of individual trees from the AHN vegetation dataset. It might also be possible to derive the dimensions of the crown using this dataset or its expected successor, the AHN-2.

The developed model takes two types of surroundings into account: the ground elevation and individual trees (point features in the provincial tree dataset). In order to account for the impact that other road surroundings have on the SVF, more datasets should be used as input for the model:

- The Top10NL and the AHN vegetation dataset can be used in conjunction to create 2.5 or even 3D objects. The Top10NL provides the 2D footprint of e.g. buildings and forests, while the AHN provides their height. SkyHelios can import shapefiles containing 2.5D objects directly, while true 3D objects need to be converted to the 'obstacle file' format ('.obs') first.
- The provincial vegetation dataset also contains polygons representing areas of trees. A method should be developed so that these data can also be taken into account.

Conversion method

In the GIS model the 3D trees are created based on the tree height only; the dimensions of the tree elements that make up the 3D tree are calculated by multiplying the tree height with the tree element's parameter (see 'Step 2. Calculation of tree elements' in chapter 3.4). Although the use of this conversion method in combination with certain tree element parameters provided fairly good results, it was tailored to work only for one kind of road surroundings, i.e. double rows of oak trees along a provincial road. Since the shape and transparency of a tree differs per species, it is to be expected that the conversion formulas and parameters have to be adapted for every species.

If different conversion methods are to be applied for different species, it is imperative that the species of the tree is derived from the provincial vegetation data first. This step was not added to the model since calibration was only performed on one tree species.

Before commencing on the quest of finding new conversion methods for different species, a choice should be made as to which SVF calculation method should be used. The affect of the choice of the calculation method in SkyHelios can be seen in chapter 4.2 'Calibration' and in section 'SVF calculation methods' of chapter 4.3.

Tree model

The GIS model calculates the SVF so that it can be used to calculate the temperature of the road surface in the months of the year when salting might be needed. Therefore coniferous trees and leafless deciduous trees are the most important objects to model. Since SkyHelios only provides solid tree models, a bare tree has to be mimicked by assigning artificial dimensions to a solid deciduous tree. It would make more sense to incorporate a transparency element that accounts for seasonality of the leaf cover. This adaption would have to be discussed with the developer of SkyHelios.

8. REFERENCES

Aebi Schmidt Nederland. (September 2010). Strooien gaat spannende toekomst tegemoet ('Salting is entering an exciting future'). *ASH magazine, 2*.

Barring, L., Mattsson, J. O., & Lindqvist, S. (1985). Canyon geometry, street temperatures and urban heat island in Malmo, Sweden. *Journal of Climatology, 5*(4), 433-444.

Brown, M., Grimmond, S., & Ratti, C. (2001). *Comparison of Methodologies for Computing Sky View Factor in Urban Environments.* Paper presented at the Proceedings of the International Symposium on Environmental Hydraulics.

Chapman, L., & Thornes, J. E. (2004). Real-time sky-view factor calculation and approximation. *Journal of Atmospheric and Oceanic Technology, 21*(5), 730-741.

Chapman, L., Thornes, J. E., & Bradley, A. V. (2001a). Modelling of road surface temperature from a geographical parameter database. Part 1: Statistical. *Meteorological Applications, 8*(4), 409-419.

Chapman, L., Thornes, J. E., & Bradley, A. V. (2001b). Rapid determination of canyon geometry parameters for use in surface radiation budgets. *Theoretical and Applied Climatology, 69*(1-2), 81-89.

Chapman, L., Thornes, J. E., & Bradley, A. V. (2002). Sky-view factor approximation using GPS receivers. *International Journal of Climatology, 22*(5), 615-621.

Eliasson, I. (1996). Urban nocturnal temperatures, street geometry and land use. *Atmospheric Environment, 30*(3), 379-392.

Fry, R., Slade, L., Taylor, G., & Davy, I. (2007). A GIS based approach to predicting road surface temperatures. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 4857 LNCS*, 16-29.

Lindberg, F. (2007). Modelling the urban climate using a local governmental geo-database. *Meteorological Applications, 14*(3), 263-273.

Matuschek, O., & Matzarakis, A. (2010). *Estimation of sky view factor in complex environment as a tool for applied climatological studies.* Paper presented at the Proceedings of the 7th Conference on Biometeorology, Freiburg.

Matzarakis, A., & Matuschek, O. (2011). Sky view factor as a parameter in applied climatology – rapid estimation by the SkyHelios model. *Meteorologische Zeitschrift, 20*(1), 039-045.

Matzarakis, A., Rutz, F., & Mayer, H. (2010). Modelling radiation fluxes in simple and complex environments: Basics of the RayMan model. *International Journal of Biometeorology, 54*(2), 131-139.

Meteo Consult, NMPO, Nieuwland Automatisering, NIDO, Provincie Gelderland, Rijkswaterstaat, et al. (2009). *RGI-168 Dynamische Gladheidsbestrijding Resultaten 2006-2008*. Wageningen.

Sass, B. H. (1997). A numerical forecasting system for the prediction of slippery roads. *Journal of Applied Meteorology, 36*(6), 801-817.

Souza, L. C. L., Rodrigues, D. S., & Mendes, J. F. G. (2003). A 3D-GIS extension for sky view factors assessment in Urban environment. *Proceedings of the 8th International Conference on Computers in Urban Planning and Urban Management*.

Thornes, J. E., Cavan, G., & Chapman, L. (2005). XRWIS: The use of geomatics to predict winter road surface temperatures in Poland. *Meteorological Applications, 12*(1), 83-90.

Unger, J. (2009). Connection between urban heat island and sky view factor approximated by a software tool on a 3D urban database. *International Journal of Environment and Pollution, 36*(1-3), 59-80.

Wassenaar, S. (2009, 12-01-2009). Strooimachine van de toekomst goed voor verkeersveiligheid en het milieu*. De Gelderlander*.

Watson, I. D., & Johnson, G. T. (1987). Graphical estimation of sky view-factors in urban environments. *Journal of Climatology, 7*(2), 193-197.

Wisse, J. S. P., Zuurendonk, I. W., & Wokke, M. J. J. (2008). Slimmer strooien door betere temperatuurverwachting (Salting smarter using a more accurate temperature forecast). *Land + Water*(10), 36-37.

Wokke, M. J. J., & Wisse, J. S. P. (2007). Meteorologen onderzoeken invloeden op wegdektemperatuur (Meteorologists investigate influences on Road Surface Temperature). *Land + Water*(12), 28-29.

APPENDICES

I. Step C: Correlation between GPS positioning error and SVF

The figures below show the SVF plotted against the positioning error of each photo location per reference dataset of 2010. There is one figure for every road stretch.

The point cloud of each reference dataset is accompanied by a trend line, the equation of the trend line and the correlation coefficient.

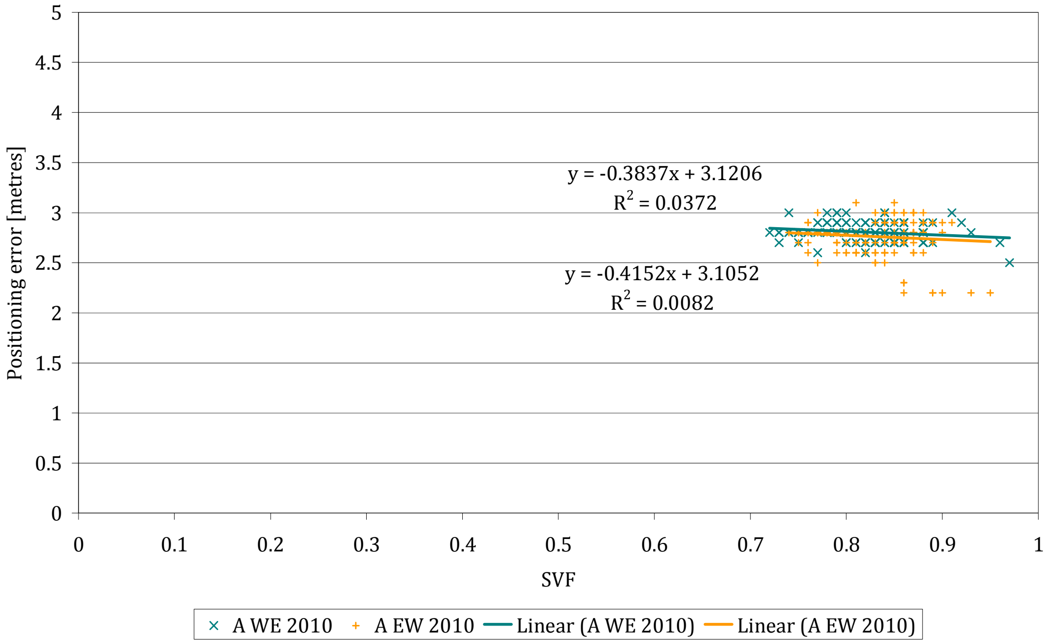


Figure 77: Relationship between GPS positioning error and SVF for both driving directions of road stretch A

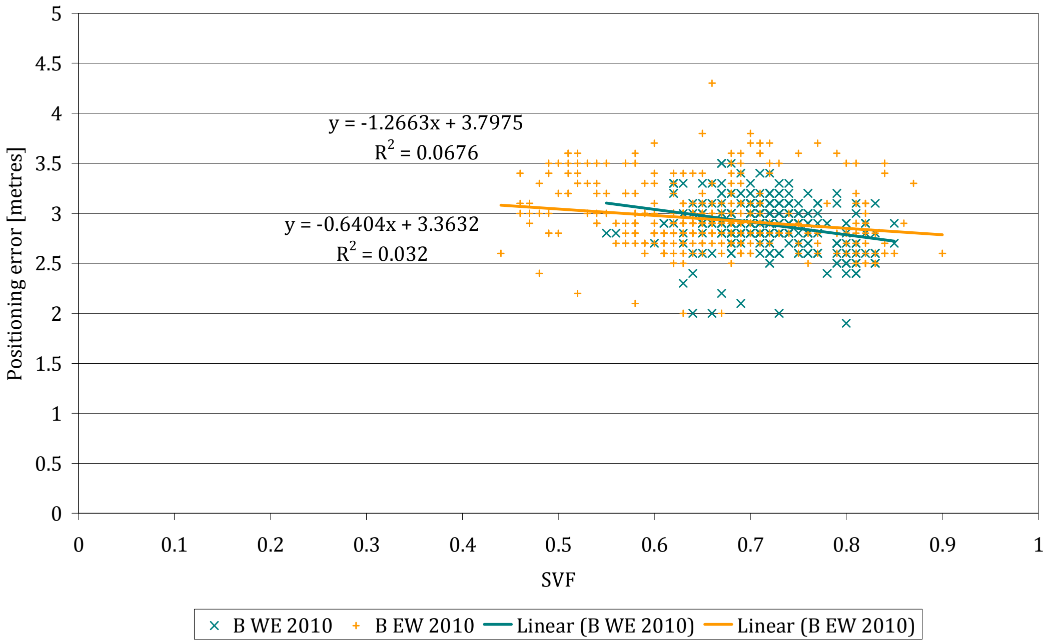


Figure 78: Relationship between GPS positioning error and SVF for both driving directions of road stretch B

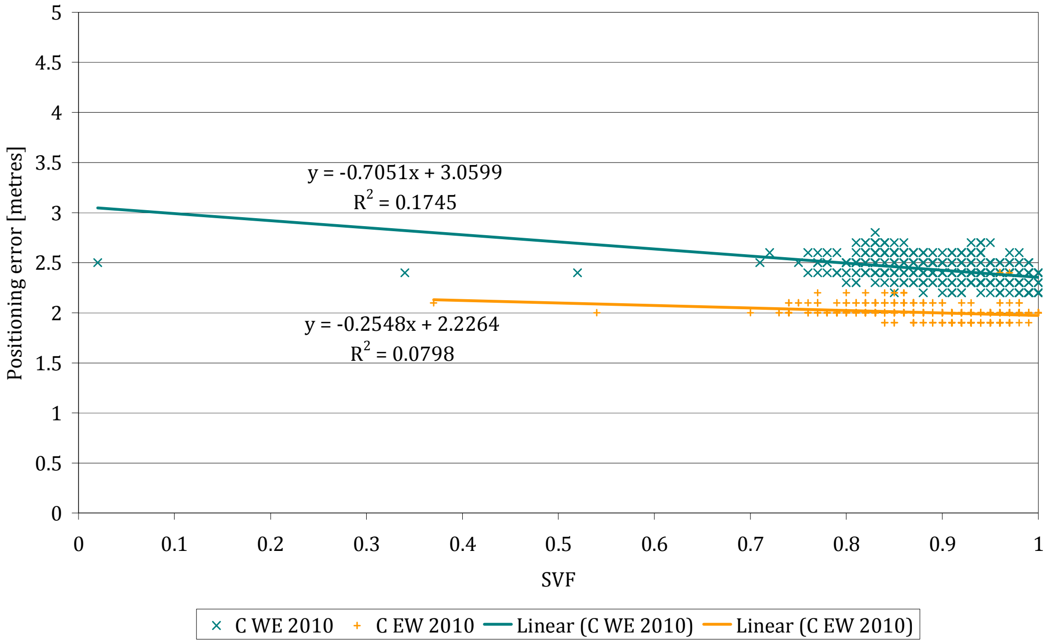


Figure 79: Relationship between GPS positioning error and SVF for both driving directions of road stretch C

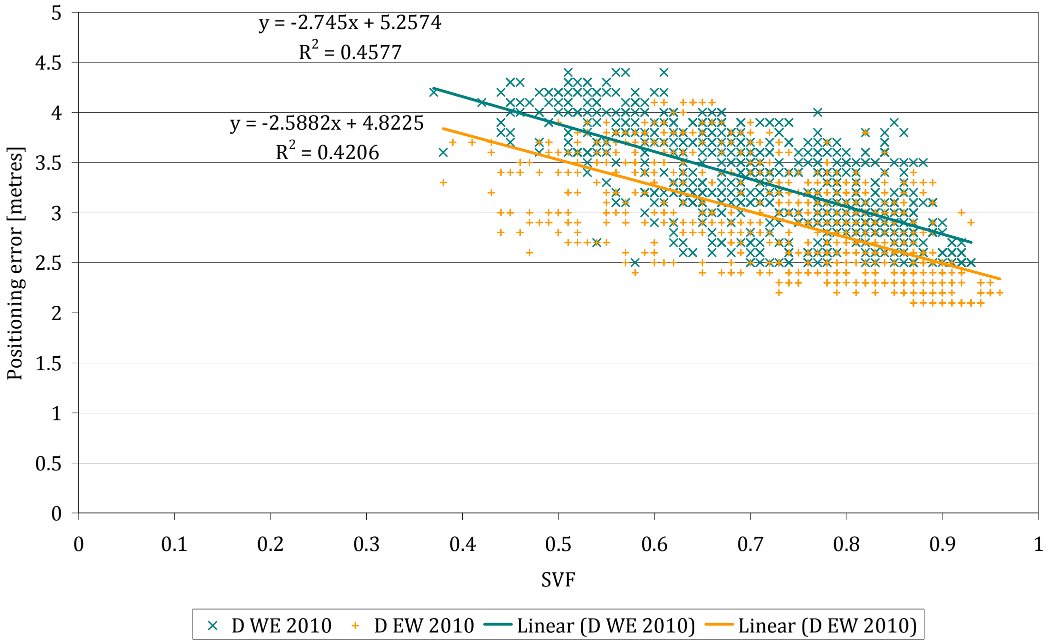


Figure 80: Relationship between GPS positioning error and SVF for both driving directions of road stretch D

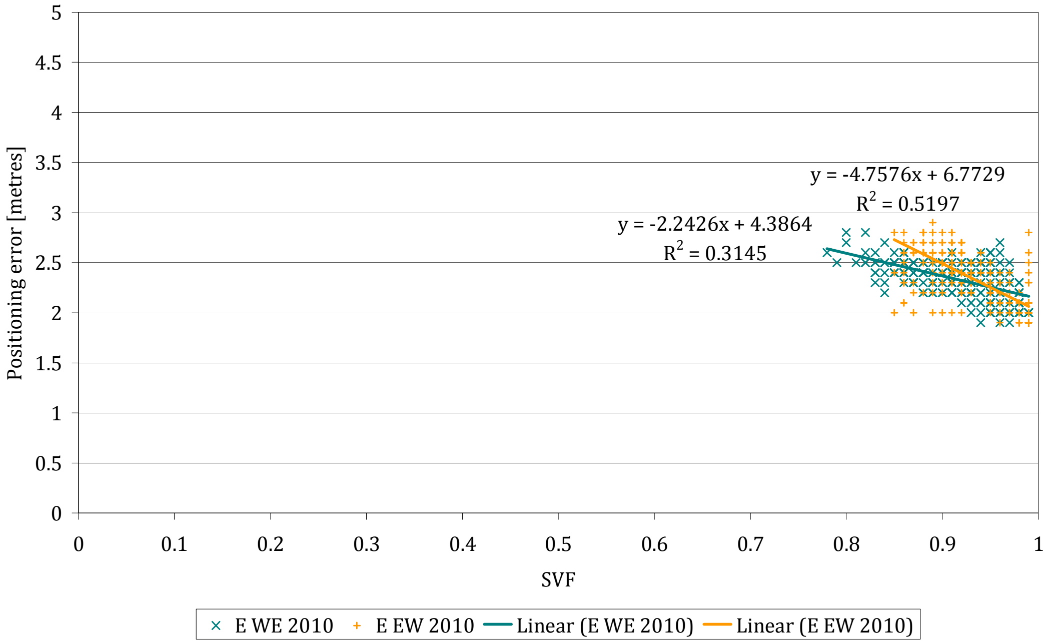


Figure 81: Relationship between GPS positioning error and SVF for both driving directions of road stretch E

II. STEP D: DIFFERENCES IN RESULTS DEPENDING ON THE INPUT DATASET

The 'input' and the 'near' datasets are reversed for two road stretches to see the effects that a different direction of comparison might have. As seen in the table below, the mean and the correlation coefficient are slightly affected, while the standard deviation is not.

Table 19: Results for road stretches D and E calculated for both directions of comparison

Road stretch	Direction / Year	Input dataset – Near dataset	n (Input)	R^2 (different directions / years)	Mean (SVF diff.)	S.d. (SVF diff.)
D	2010	EW – WE	830	0.739	0.05	0.046
		WE – EW	647	0.723	0.06	0.046
E	WE	2007 – 2010	237	0.265	0.03	0.020
		2010 – 2007	221	0.271	0.03	0.020

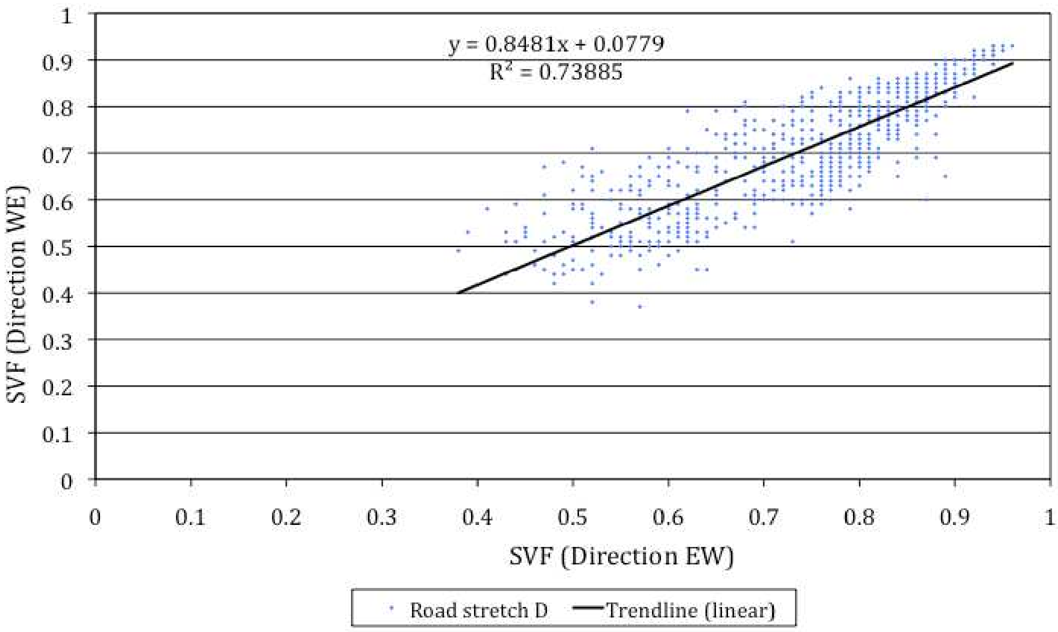


Figure 82: D EW-WE n = 830

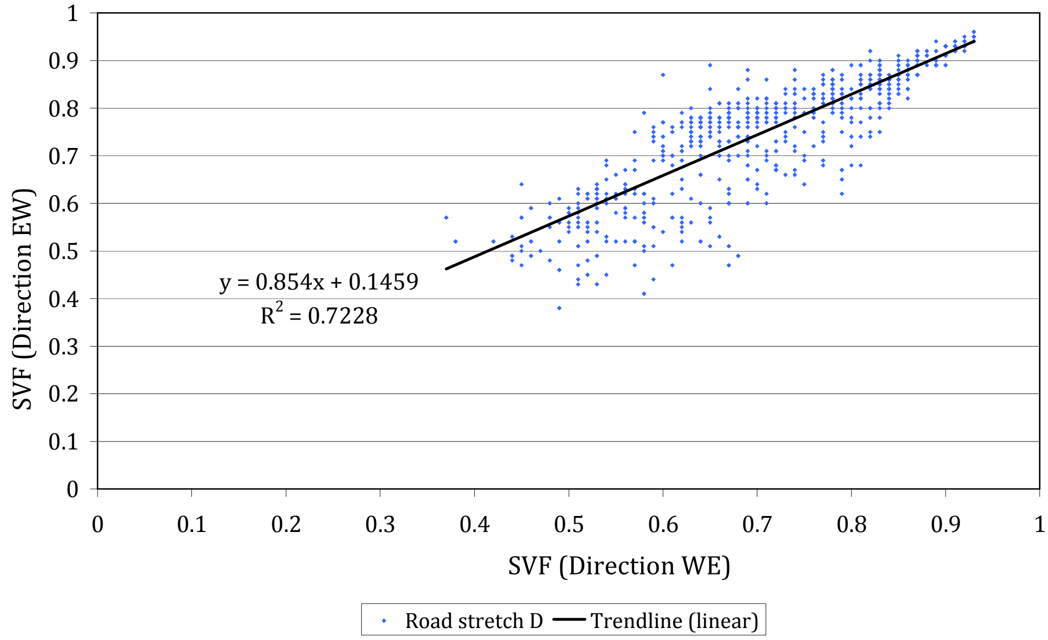


Figure 83: D WE-EW 2010, n = 647

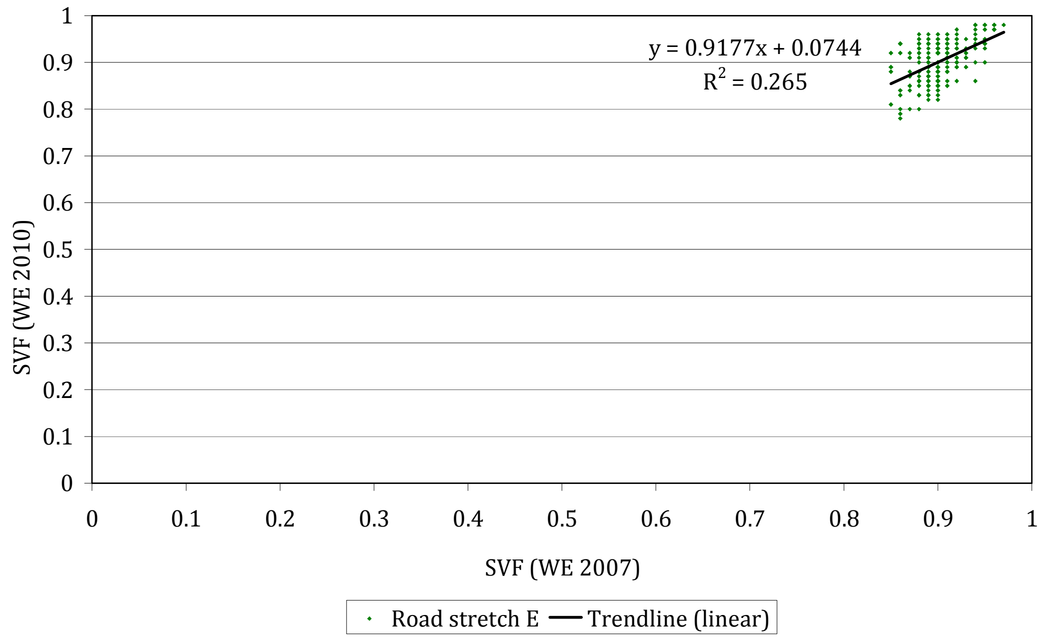


Figure 84: E 2007-2010 WE, n = 237

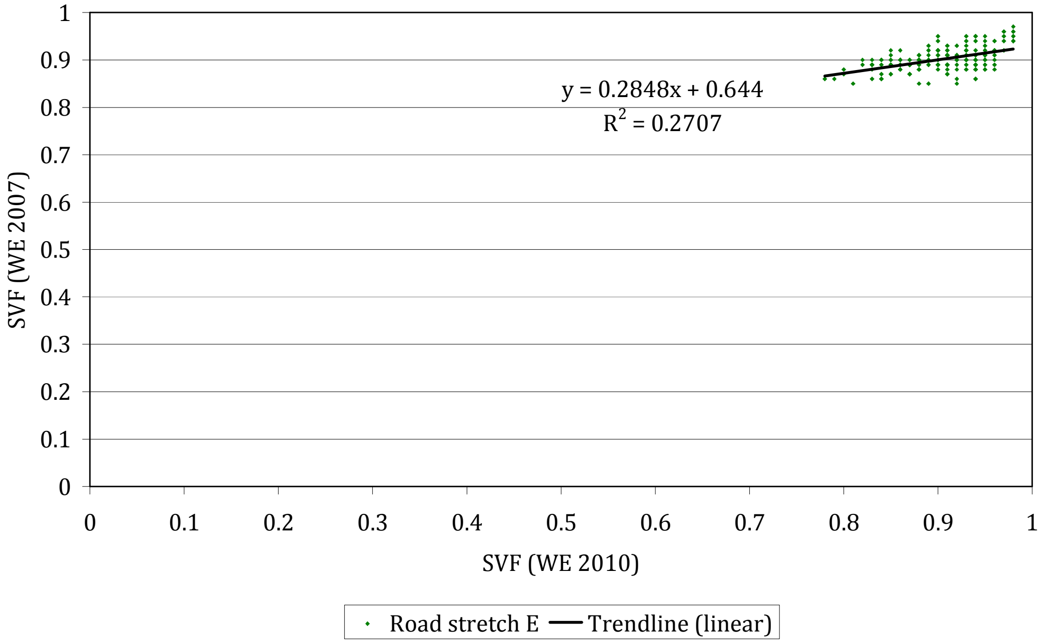


Figure 85: E 2010-2007 WE, n = 221

III. GIS MODEL PROCESSING IN DETAIL

Step 1: AHN → .txt

1. Clip DEM
2. Apply 'Raster To Point (Conversion)' tool on DEM, keep info of field [GRIDCODE] (= height in centimeters)
3. Add fields [X], [Y] and [Z] (Float)
4. Calculate geometry of [X] and [Y] fields
5. Convert height from centimeters to meters: [Z] = [GRIDCODE] * 0.01
6. Export table (to any format Excel can read)
7. Remove all columns but [X], [Y] and [Z]
8. Save to .txt format, separator ";" or tab

Step 2: .shp → .obs

1. Add fields [X], [Y] (Float) and [HEIGHT_M] (Integer) to tree data shapefile
2. Calculate geometry of [X] and [Y] fields
3. Derive height from [HOOGTE] and store in [HEIGHT_M] (First two numbers in the string equal the maximum value of category, except for the highest category ("> 26") where the value indicates the minimum.)
4. Delete all columns but [X], [Y], [SOORT], [HEIGHT_M]
5. (Copy and) import .dbf of shapefile in Excel into the conversion workbook, loading it into the input data sheet
6. Export output sheet as .txt, separator: tab
7. Delete header
8. Change extension to .obs

Step 3: Calculate SVF (SkyHelios)

- Import raster (from step 1), set resolution
- Import obstacle file (from step 2), set midpoint to '0x0'
- Select AOI if necessary
- Calculate area SVF: set camera height (1.5 m) and resolution

Step 4: .txt → Point feature

- Delete "." and "-" characters from header
- Change delimiter from ";" to tab
- Remove all extra "." from filename if present
- Create point dataset using 'Create Feature Class > From XY Table' (ArcGIS)
- Run script 'Bb5-7' in ArcGIS if visualization along centerline is needed (script available upon request)

IV. Transects: Tree heights sampled from the AHN vegetation dataset

Transects were made on both validation locations in order to check the height of the provincial trees. The graphs in this appendix show the AHN vegetation heights within each transect.

First validation location: N224 30.0 – 31.4km

Figure 86: The position of the transects at the first validation location

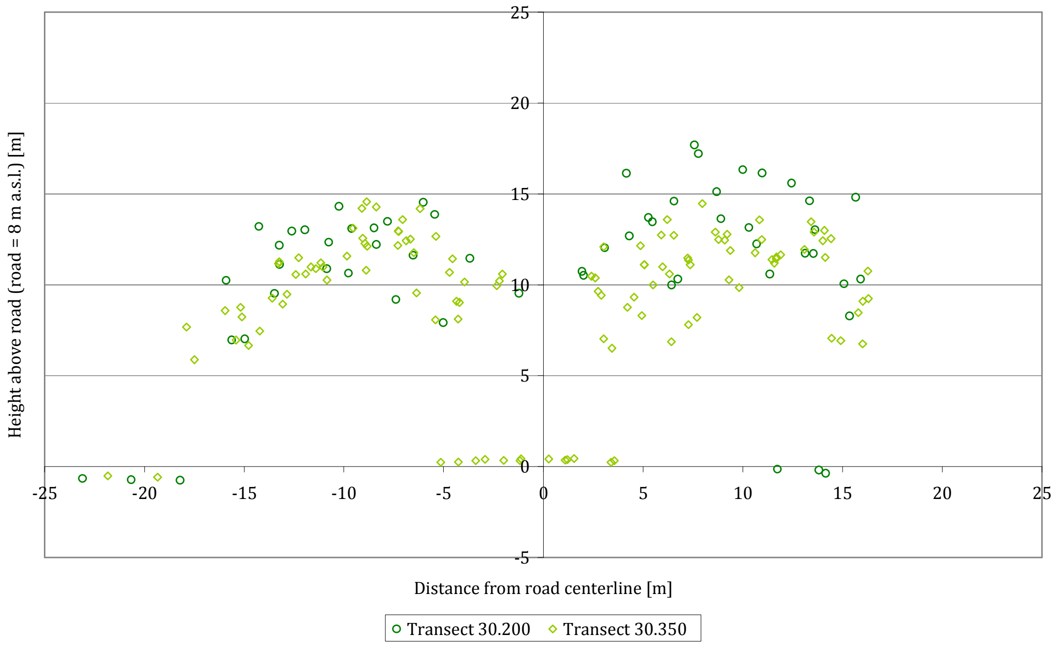


Figure 87: The height and position of the sampling points within the transects at 30.200 and 30.350km along the N224, limited to 25m distance from the road center

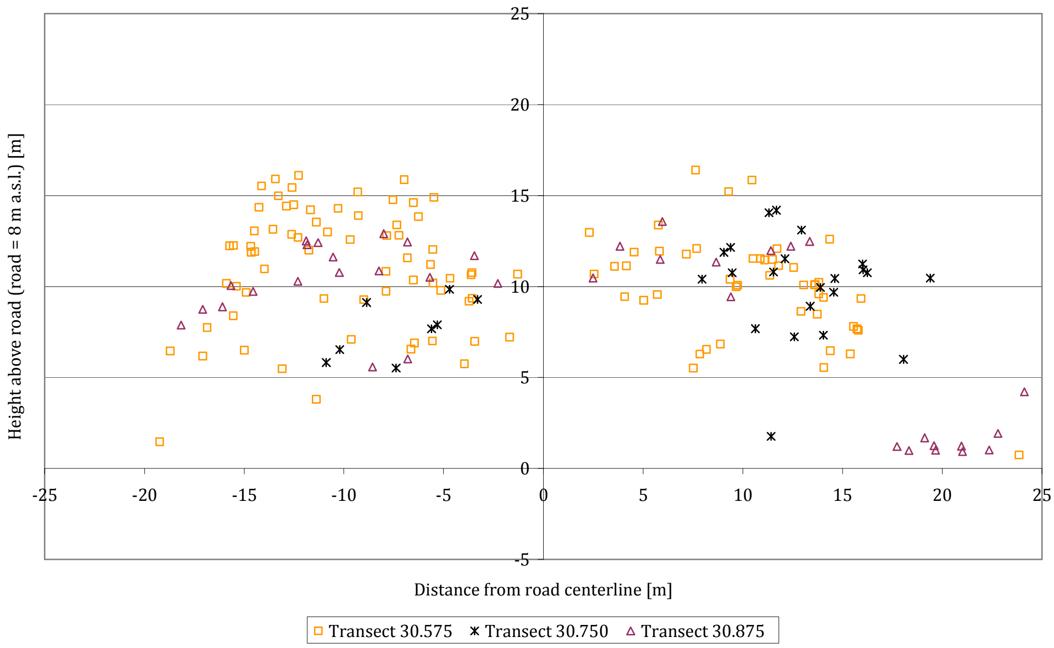


Figure 88: The height and position of the sampling points within the transects at 30.575, 30.750 and 30.875km along the N224, limited to 25m distance from the road center

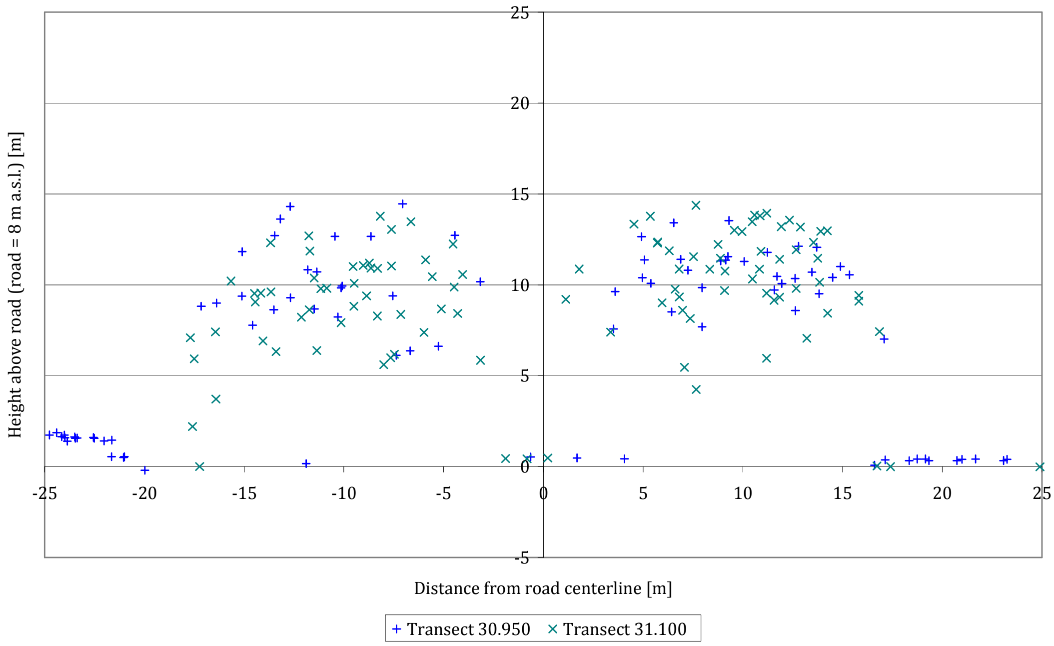


Figure 89: The height and position of the sampling points within the transects at 30.950 and 31.100km along the N224, limited to 25m distance from the road center

Second validation location: N224 32.0 – 33.0km

Figure 90: The position of the transects at the first validation location

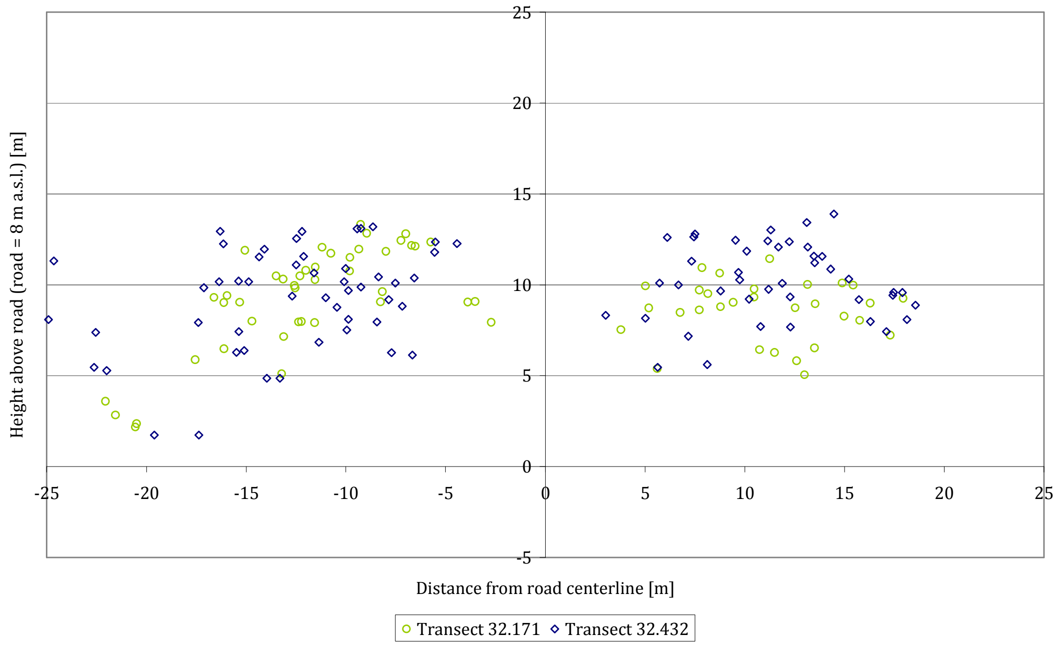


Figure 91: The height and position of the sampling points within the transects at 32.171 and 32.432km along the N224, limited to 25m distance from the road center

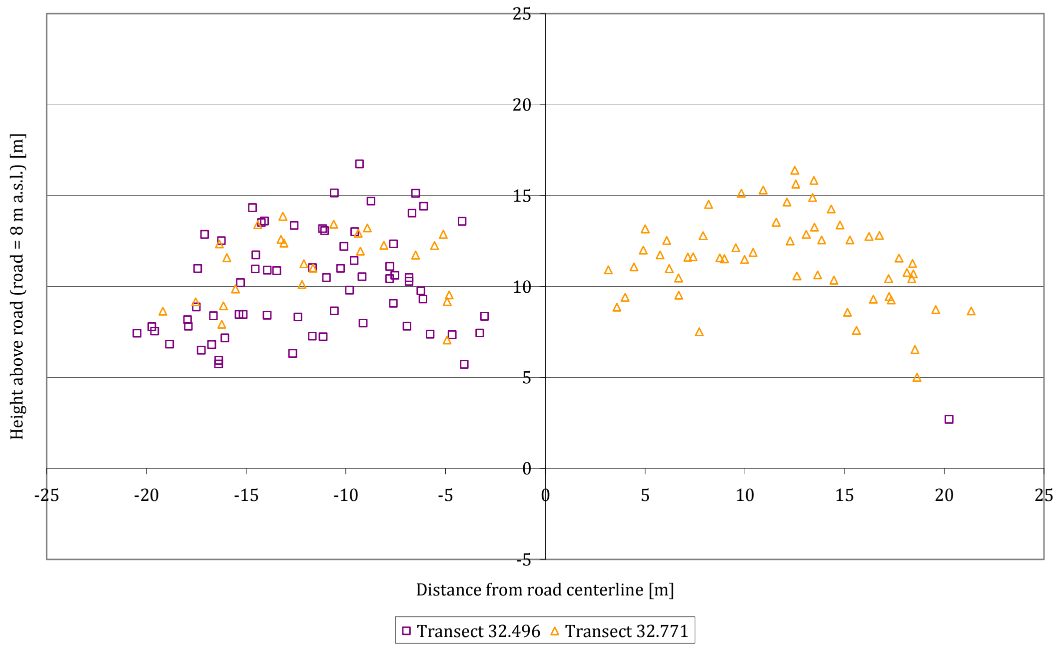


Figure 92: The height and position of the sampling points within the transects at 32.496 and 32.771km along the N224, limited to 25m distance from the road center

V. CHANGES MADE TO THE TREE HEIGHTS BASED ON THE AHN VEGETATION DATSET

First validation location: N224 30.0 – 31.4km

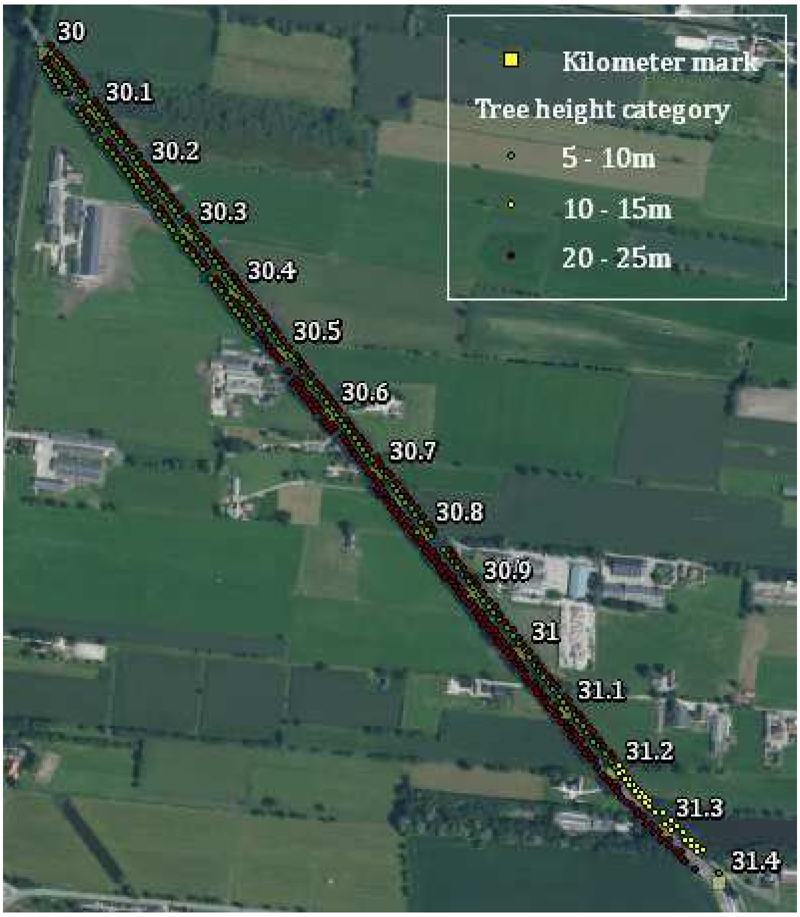


Figure 93: Tree heights based on provincial data. Trees with no height data were assigned the category 5 – 10m.

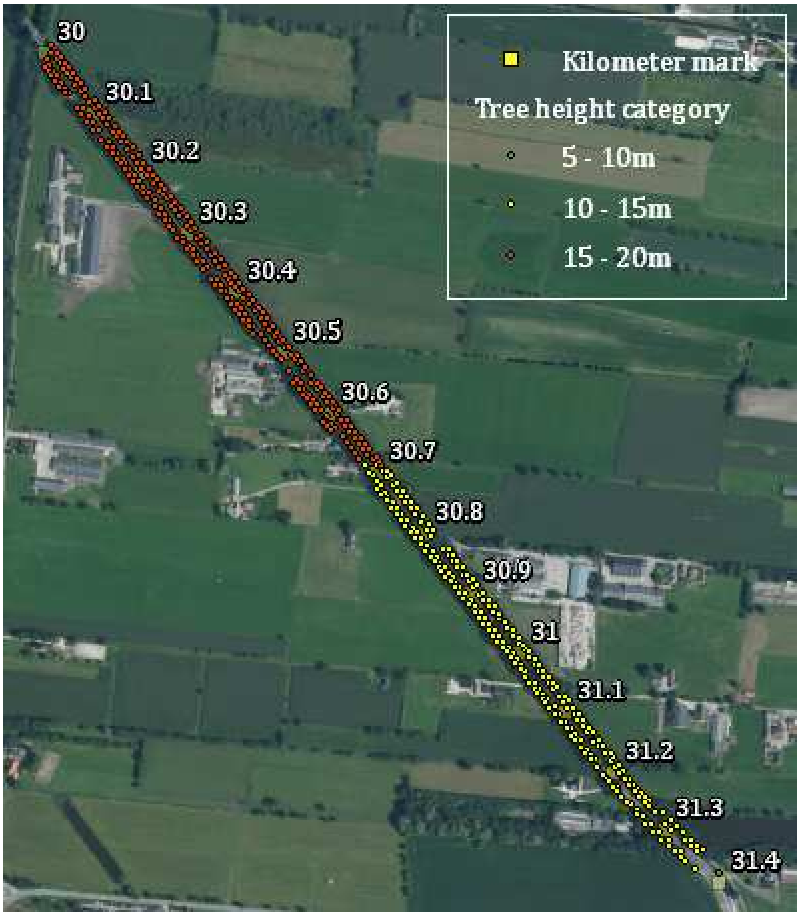


Figure 94: Tree heights based on the transects made

Second validation location: N224 32.0 – 33.0km

Figure 95: Tree heights based on provincial data. Trees with no height data were assigned the category 20 – 25m.

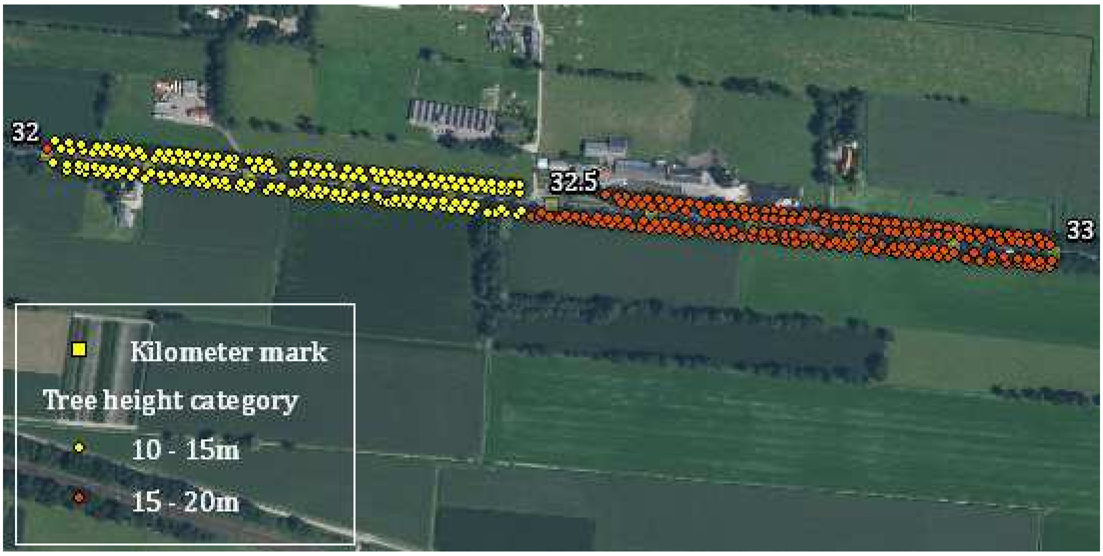


Figure 96: Tree heights based on the transects made