



UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvennoot • your knowledge partner

Extracting Tree Heights from Digital Elevation Model Images

Gerhardt Carl Schmidt
20733410

Report submitted in partial fulfilment of the requirements of the module
Project (E) 448 for the degree Baccalaureus in Engineering in the Department of
Electrical and Electronic Engineering at Stellenbosch University.

Supervisor: Dr RP Theart

November 2020

Acknowledgements

I would like to express my very great appreciation to Dr RP Theart for his guidance throughout the execution of this project. It was a wonderful experience to work with such an intelligent and kind person.

I would like to show my gratitude to the people at Aerobotics for making this project possible and going out of their way to assist me with this project.

My special thanks to Quintin Strydom and Ralf Hansen for being excellent study partners and even better friends.

Finally, to my loving parents, thank you for your constant support and letting me pursue my dream in becoming an engineer.



UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvennoot • your knowledge partner

Plagiaatverklaring / Plagiarism Declaration

1. Plagiaat is die oorneem en gebruik van die idees, materiaal en ander intellektuele eiendom van ander persone asof dit jou eie werk is.

Plagiarism is the use of ideas, material and other intellectual property of another's work and to present it as my own.

2. Ek erken dat die pleeg van plagiaat 'n strafbare oortreding is aangesien dit 'n vorm van diefstal is.

I agree that plagiarism is a punishable offence because it constitutes theft.

3. Ek verstaan ook dat direkte vertalings plagiaat is.

I also understand that direct translations are plagiarism.

4. Dienooreenkomsdig is alle aanhalings en bydraes vanuit enige bron (ingesluit die internet) volledig verwys (erken). Ek erken dat die woordelikse aanhaal van teks sonder aanhalingstekens (selfs al word die bron volledig erken) plagiaat is.

Accordingly all quotations and contributions from any source whatsoever (including the internet) have been cited fully. I understand that the reproduction of text without quotation marks (even when the source is cited) is plagiarism

5. Ek verklaar dat die werk in hierdie skryfstuk vervat, behalwe waar anders aangedui, my eie oorspronklike werk is en dat ek dit nie vantevore in die geheel of gedeeltelik ingehandig het vir bepunting in hierdie module/werkstuk of 'n ander module/werkstuk nie.

I declare that the work contained in this assignment, except where otherwise stated, is my original work and that I have not previously (in its entirety or in part) submitted it for grading in this module/assignment or another module/assignment.

20733410 Studentenommer / Student number	 Handtekening / Signature
GC Schmidt Voorletters en van / Initials and surname	09/11/2020 Datum / Date

Abstract

English

This project aimed to extract the height of trees located within orchards from Digital Elevation Models (DEM) and multispectral images. Latitude, longitude and altitude describing the ground surface are collected automatically from the theses images. The Universal Kriging method was the technique used to predict the altitude of the ground below trees from which the height of the trees could be estimated. The ground plane could then be removed from the DEM, resulting in a DEM that only indicates the height of trees relative to ground. Individual treetop heights could also be estimated from provided geographic coordinates.

Afrikaans

Die doel van hierdie projek was om gehad om die hoogte van bome in boorde te bepaal van uit *Digital Elevation Models* (DEM) en multispektrale fotos. Breedtegraad, lengte en hoogte wat die grondoppervlak beskryf, word outomaties uit die fotos versamel. Die *Universal Kriging*-metode was die tegniek wat gebruik is om die hoogte van die grond onder die bome te voorspel waaruit die hoogte van die bome geskat kon word. Die grondvlak kan dan van die DEM verwyder word, wat lei tot n DEM wat slegs die hoogte van bome in verhouding tot die grond aandui. Individuele boomtophoogtes kan ook geskat word vanaf geografiese koördinate.

Contents

Declaration	ii
Abstract	iii
List of Figures	vii
List of Tables	ix
Nomenclature	xi
1. Introduction	1
1.1. Background	1
1.2. Problem Description	1
1.3. Objectives	1
1.4. Summary of Work	2
1.5. Scope	2
1.6. Report Overview	3
2. Literature Study	4
2.1. Introduction	4
2.2. Remotely Sensed Data of Landscapes	4
2.2.1. Obtaining and Generating Digital Elevation Models	4
2.2.2. Detecting Vegetation Through Multispectral Images	6
2.3. Interpolation Techniques	7
2.3.1. Kriging	9
2.3.2. Inverse Distance Weighting	9
2.3.3. Spline and Polynomial Interpolation	10
2.3.4. Evaluating Interpolation Accuracy	11
2.4. Previous Work Done	12
2.4.1. Accurate Digital Elevation Modeling	12
2.4.2. Individual Tree Detection from Canopy Height Models	13
2.5. Conclusion	13
3. Software Design	15
3.1. Introduction	15

3.2. Inspecting the Data Set	16
3.2.1. The Raster Images	16
3.2.2. The GeoJSON Tree Features	17
3.3. Working Between Raster Images	18
3.4. Detecting Ground Points and Obtaining Their Geospatial Information	19
3.4.1. Detecting Ground Points from RGBA and NDRE Raster Images	19
3.4.2. Uniformly Reducing Detected Ground Points	20
3.4.3. Obtaining True Altitude of Ground Points Through a Local Minima Algorithm	21
3.5. Estimating the True Altitude of the Ground Surfaces Below Trees	22
3.5.1. Universal Kriging	22
3.5.2. Inverse Distance Weighting	23
3.5.3. Linear Spline	23
3.6. Estimating Tree Heights	23
3.6.1. Generating Canopy Height Models	24
3.6.2. Spatial Resolution of DTM and CHM	25
3.6.3. Estimating the Height of Individual Treetops	25
3.7. Metrics	26
3.7.1. Accuracy of Sampled Ground Points.	26
3.7.2. Measuring the Accuracy of the Tree Height Estimations	27
3.7.3. Manually Calculating Tree Heights	27
3.8. Conclusion	28
4. Detailed Design	29
4.1. Introduction	29
4.2. Determining the NDRE Classification Threshold	29
4.3. Selecting the Interpolation Method	29
4.3.1. Performance of UK, IDW and LS at Different Terrain Conditions .	29
4.3.2. Comparing Interpolation Methods through PDFs	31
4.4. The GitHub Repository	31
4.5. Conclusion	31
5. Results	33
5.1. Introduction	33
5.2. Accuracy of Ground Point Sampling	33
5.3. Accuracy of Treetop Heights and True Ground Altitude Estimations	34
5.4. Producing Canopy Height Models	35
5.5. Individual Tree Detection	36
5.6. Conclusion	36

6. Summary, Conclusion and Future Work	38
6.1. Summary	38
6.2. Conclusion	38
6.3. Future Work	39
Bibliography	40
A. Project Planning Schedule	44
B. Outcomes compliance	45
B.1. ELO 1. Problem Solving	45
B.2. ELO 2. Application of Scientific and Engineering Knowledge	45
B.3. ELO 3. Engineering Design	45
B.4. ELO 4. Investigations, Experiments and Data Analysis	46
B.5. ELO 5. Engineering Methods, Skills and Tools, Including Information Technology	46
B.6. ELO 6. Professional and Technical Communication	46
B.7. ELO 8. Individual Work	46
B.8. ELO 9. Independent Learning Ability	46
C. Comparing Estimated Tree Heights to Manually Calculated Tree Heights	48
D. Terrains	53
E. Visualisation of Steps to Produce a CHM	56
F. Contours of Digital Terrain Models for Terrains C, D and E	58
G. Comparing Probability Density Functions of Treetop Heights	59
H. Canopy Height Models of Terrains A,B,D and E	61
I. Interpolation Algorithms for UK, IDW and LS	62

List of Figures

2.1. Different types of Digital Elevation Models [1].	5
2.2. Grayscale heightmap of Terrain A.	6
2.3. Solar spectrum [2].	7
2.4. Interpolation of earth surfaces [3].	8
2.5. Predicted contours of the same earth surface, but the elevations were generated using different interpolation techniques [4].	8
2.6. Illustration of how polynomials of different degrees can be used to represent a surface. [5].	11
3.1. Software design flow diagram.	15
3.2. The various layers of Terrain A as viewed in QGIS.	16
3.3. An orange tree that has reached maturity [6].	17
3.4. Classifying pixels based on NDRE index values. Ground points would have very similar NDRE index values close to 0, whereas the majority of non-ground pixels that belong to trees would have NDRE values closer to 1.	19
3.5. Finding ground points by classifying pixels of the NDRE raster images. Note that, due to the size of the dots used for the plot, it appears that ground covers larger area than it does.	20
3.6. Uniformly spaced ground points that will be used in the interpolation process.	20
3.7. Distortion in ground surface elevation.	21
3.8. Sampled ground points with geospatial data determined by Algorithm 3.1.	22
3.9. 3-D elevation plots indicating the difference between the original DEM data and the predicted DTM from a portion of terrain A (a) DEM data that includes the true altitudes of trees. (b) The predicted DTM with only the true altitude of the ground surface.	24
3.10. Illustration explaining how PDFs can be used to determine accuracy of predicted tree heights. Blue represents the PDF of manually measured data. The green and red PDFs represent predicted data. Accurately predicted data would resemble that of the green PDF, while poorly predicted data would look like that of the red PDF. Note that this is an illustration and not plots representing actual data.	26

3.11. How measuring 3 points around a tree and the treetop in QGIS can be used to determine the tree's height. (a) Here the locations at which to measure is indicated through the black points (P1, P2, P3) and the one red point (treetop). (b) The black points are used to determine the plane on which the tree is on and the red dot represents its treetop.	28
4.1. RMSE at different terrain conditions (lower is better).	30
4.2. Comparing PDFs of the estimated tree heights to that of the manually calculated tree heights (blue). (a) The calculated tree heights from Table C.1. (b) The calculated tree heights from Table C.3.	32
5.1. Contours of DTMs.	35
5.2. CHMs of Terrain C with different spatial resolutions (a) The CHM with a spatial resolution 100 times lower than its DEM. (b) The CHM with a spatial resolution 25 times lower than its DEM.	36
5.3. Using DTM with high spatial resolution to produce CHM of Terrain A. .	37
5.4. Individual Tree Detection from CHMs.	37
A.1. Gantt chart of project.	44
D.1. Terrain A	53
D.2. Terrain B	54
D.3. Terrain C	54
D.4. Terrain D	55
D.5. Terrain E	55
E.1. The process of obtaining CHM from the DTM and DEM of Terrain A. .	57
F.1. Contours of DTMs	58
G.1. Green - 10 Calculated Treetop heights from the terrain. Blue - Treetop height obtained from low spatial resolution CHM. Yellow - Treetop heights estimated by predicting ground with UK. (a) Calculated tree heights from Table C.1. (b) Calculated tree heights from Table C.2. (c) Calculated tree heights from Table C.3. (c) Calculated tree heights from Table C.4. (d) Calculated tree heights from Table C.5.	60
H.1. CHMs Terrain A,B,D and E	61

List of Tables

4.1. RMSE of tree height estimations for different types of terrains (lower is better).	30
5.1. RMSE of tree height estimations for Terrains A-E.	34
C.1. Terrain A tree heights	48
C.2. Terrain B tree heights	49
C.3. Terrain C tree heights	49
C.4. Terrain D tree heights	50
C.5. Terrain E tree heights	50
C.6. Tree heights of Trees on Flat Ground Surface	51
C.7. Tree Heights of Trees on Slightly Sloped Ground Surface	51
C.8. Tree Heights of Trees on Sloped Ground Surface	52
C.9. Tree Heights of Trees on Bumpy Ground Surface	52

Nomenclature

Variables and functions

x	The column index of an array/raster band.
y	The row index of an array/raster band.
s	The location of a point.
s_0	The location at which interpolation occurs.
N	Total number of ground samples detected.
n	Number of ground samples used to interpolate at s_0 .
s_i	Location of the i^{th} sample point used for interpolation.
$\hat{Z}_K(s_0)$	The true altitude predicted through UK with respect to variable s_0 .
$\hat{Z}_{IDW}(s_0)$	The true altitude predicted through IDW with respect to variable s_0 .
$\hat{Z}_{LS}(s_0)$	The true altitude predicted through Linear Spline with respect to variable s_0 .
$Z(s_i)$	True altitude of sample point with respect to variable s_i .
$\lambda_K(s_i)$	The weight of the i^{th} sample point used for IDW interpolation with respect to variable s_i .
$\lambda_{IDW}(s_i)$	The weight of the i^{th} sample point used for IDW interpolation with respect to variable s_i .
p_{IDW}	The power factor used in IDW interpolation.
p_p	The degree of a polynomial function.
$d(s_0, s_i)$	Euclidean distance between s_0 and s_i
μ	Statistical mean.
σ	Statistical standard deviation.
res_h	The vertical spatial resolution of a raster image in degrees.
res_w	The horizontal spatial resolution of a raster image in degrees.
$raster_h$	Height of a raster image in pixels.
$raster_w$	Width of a raster image in pixels.
$\gamma(s_1, s_2)$	Variogram function with respect to s_1 and s_2 .
T	The number of detected trees within the area of interest.
μ_{NDRE}	The average of the values stored in the NDRE raster band.

Acronyms and abbreviations

TIF	Tagged Image File Format
ML	Machine Learning
DEM	Digital Elevation Model
DTM	Digital Terrain Model
DSM	Digital Surface Model
CHM	Canopy Height Model
NDRE	Normalized Difference Red Edge
UK	Universal Kriging
IDW	Inverse Distance Weighting
LS	Linear Spline
NDVI	Normalized Difference Vegetation Index
GIS	Geographic Information System
QGIS	Quantum Geographic Information System
RGB	Red, Green and Blue
RGBA	Red, Green, Blue and Alpha
PDF	Probability Density Function
LIDAR	Light Detection and Ranging
NIR	Near-Infrared
RE	Edge of Red
JSON	JavaScript Object Notation
IR	Infrared
ITD	Individual Tree Detection

Chapter 1

Introduction

1.1. Background

As we progress further into the Information Age, the value of data has become increasingly evident. Data can be processed to produce valuable information, and with information, decisions can be made with greater confidence. Consequently, there is a great need to develop technologies and methods of obtaining and analyzing data. One such data gathering method is that of remote sensing, which refers to the collection of data from locations that are inaccessible for humans. There are numerous applications of remote sensing, from monitoring climate change with the use of satellites to mapping the surface of the ocean floor from boats. The use of remote sensing has also become prevalent in the farming industry. Orchards and crops can be captured remotely, which provides access to data that would have otherwise been unobtainable. New methods are continuously being developed to improve the process of generating valuable information from data for agricultural purposes. This project aims to contribute to this development.

1.2. Problem Description

Aerobotics is an agricultural data analytics company that strives to find new and innovative ways of extracting useful data from crops or orchards. They have provided a data set consisting of Digital Elevation Models (DEM) and other multispectral images of orchards captured through the use of drones. The primary goal of this project is to determine the height of trees belonging to the orchards captured from the given data set. The secondary goal of this project is to remove the ground plane beneath the trees to produce a Canopy Height Model (CHM) that accurately shows the height of the tree canopies relative to the ground below them.

1.3. Objectives

The objectives of this project is therefore to design a software algorithm that is capable of:

1. Automatically detecting ground points located between the trees in an orchard, from which the geospatial data (latitude, longitude and true altitude) can be obtained.
2. Predicting the elevation of the ground below trees.
3. Estimating the height of trees within the captured orchards.
4. Generating accurate CHMs.
5. Identifying individual trees from the CHMs.

Furthermore, since orchards are planted on various types of terrains, such as hills and valleys, the algorithm must be robust enough to detect the height of trees on these different terrains accurately.

1.4. Summary of Work

Firstly, a ground detection algorithm was designed to identify pixels that belong to the ground surface of a terrain from NDRE raster images. Secondly, was the development of an algorithm that obtains the geospatial data of uniformly spaced ground points. An Inverse Distance Weighting (IDW) interpolation algorithm was created from base principles. The IDW algorithm and two other interpolation methods were implemented in the designed software to predict the elevation of ground below the trees. The best performing interpolation method was discovered to be the Universal Kriging (UK) method. An algorithm that estimates treetop heights given the geospatial data of the treetops was created. CHMs of varying spatial resolutions and extents were produced by removing the ground plane predicted through UK from DEMs. The accuracy of the tree height estimations and CMHs were determined. Individual tree detection through applying a local maxima algorithm to the CHMs was attempted.

1.5. Scope

The designed software algorithms aimed to estimate the height of trees as accurately as possible. The estimation of other parameters, such as tree volume could be possible from the same data set, but this was not an objective of this project. The method used to determine the elevation of the ground plane from the DEMs was that of spatial interpolation. Methods involving machine learning were not pursued. CHMs with varying extents and spatial resolutions were generated. The use of CHMs for individual tree identification is considered.

1.6. Report Overview

In Chapter 2, essential concepts regarding the provided data set are explained. This is followed by a discussion of three different interpolation methods and how the most accurate interpolation results can be achieved. Chapter 2 concludes with a description of how a local maxima algorithm can be used to detect trees. Chapter 3 contains the conceptual software design as well as the metrics that will be used to establish if the objectives of the project were met. Chapter 4 discusses the rationale behind the final software design decisions made. In Chapter 5, the results of the project is represented. Finally, a summary, conclusion and opportunities for future work for this project are given in Chapter 6.

Chapter 2

Literature Study

2.1. Introduction

In this chapter, essential concepts and techniques concerning this project will be researched and discussed to provide the necessary context for later chapters. First, the techniques that are used to obtain and generate the Digital Elevation Model (DEM) and Normalized Difference Red Edge (NDRE) data of a terrain are explained. This is followed by descriptions of different interpolation algorithms that are well suited to estimate values from geospatial data. This chapter concludes with research on topics relating to individual tree identification and generating accurate DEMs.

2.2. Remotely Sensed Data of Landscapes

2.2.1. Obtaining and Generating Digital Elevation Models

A Digital Elevation Model (DEM) is a digital representation of the surface elevation of a terrain which may or may not include objects on that surface. DEMs are used in Geographic Information Software (GIS) to indicate the elevation at locations within the area that is covered by the DEM [4, 7]. Various DEM types represent elevation differently by either removing some features located on the surface of the terrain or by changing the type of altitude that is used to represent the elevation. DEMs can either refer to a Digital Terrain Model (DTM) or a Digital Surface Model (DSM). In the case of this project, DEMs will refer to DSMs. DTMs indicates a terrain's ground plane elevation, where all non-ground objects are removed. These non-ground objects include buildings, bridges, tree canopies, etc. For DSMs, the elevation of features covering the surface of the terrain is retained when indicating the elevation over the surface of the terrain. Lastly, a Canopy Height Model (CHM) is a DEM which displays the height of the trees relative to the ground surface. The CHM of a terrain can be obtained by subtracting the DTM elevations from its DSM elevations [8]. Figure 2.1 illustrates the differences between these three DEM types.

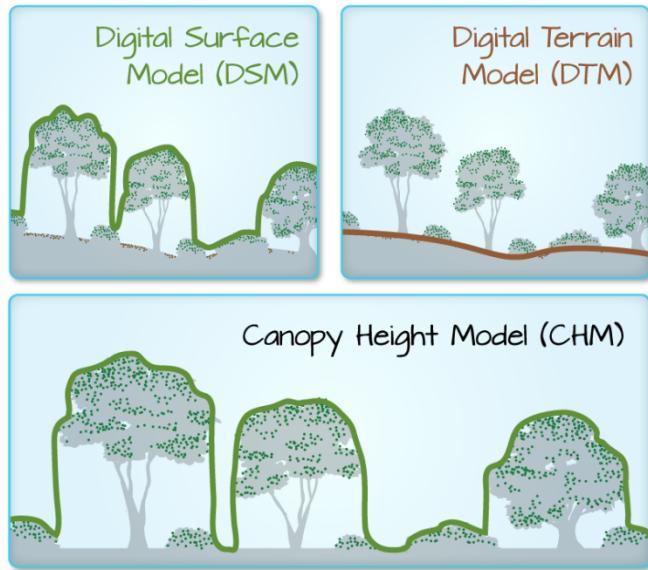


Figure 2.1: Different types of Digital Elevation Models [1].

There are two different types in which DEMs can be represented, either as raster images (also known as heightmaps) or as a vector-based Triangular Irregular Networks (TINs). An example of a heightmap DEM is shown in Figure 2.2. A raster image consists of a rectangular grid of pixels, in which the pixels of the raster image represent values stored inside 2-D data structures called raster bands. In general, heightmaps only store an elevation value for each pixel and the geospatial data identifying the location of the pixel as latitude and longitude. Heightmaps are often visualized as grayscale images, where the elevation of a point on the terrain is indicated by the intensity of a pixel. Grayscale images typically use 8-bit values to represent a pixel's intensity, resulting in only 256 unique elevations that can be indicated. The elevation resolution of the heightmap can be increased by using larger data types consisting of more bits. By using colour images for heightmaps, a wider range of elevations can be indicated. For example, using a Red-Green-Blue (RGB) image, 16,777,216 unique heights can be represented due to the three colour channel storing an 8-bit value [9].

TINs can approximate a surface by using a mesh consisting of connecting vertices (3-D points) to one another through edges forming non-overlapping triangles. The elevation of a point can be determined from the surface of the triangle it is located on. The triangles of the TIN model are created using interpolation techniques, such as Delaunay triangulation or distance ordering. TIN is a more complex representation of DEMs than raster images, as a result of the way in which data is obtained, stored and processed. The more complex data structure of TINs makes the processing of the data less efficient than that of raster data. Obtaining good source data to create a TIN model is also a more costly process. A benefit of using TIN is that points can be placed irregularly,

allowing higher elevation accuracy at locations where the surface is very irregular [9].

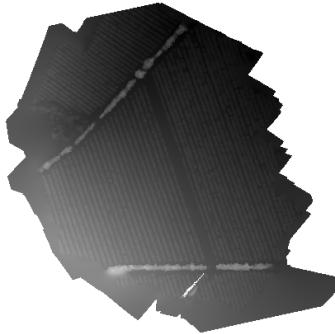


Figure 2.2: Grayscale heightmap of Terrain A.

DEMs are generated by interpolating points collected through remote sensing. The latitude, longitude and some type of altitude of points on a terrain's surface are collected during remote sensing. DEMs commonly use the true altitude of points, which is the absolute height above sea level. Some of the most popular remote sensing methods used to acquire DEM data points include stereo photogrammetry from aerial surveys and Light Detection and Ranging (LIDAR). LIDAR is a widely used due to its speed and high accuracy. The topographic LIDAR utilizes a near-infrared laser and a Global Position System (GPS) to establish the geospatial information at locations where the laser hits a surface [7]. Stereo photogrammetry processes and interprets various images of a landscape, usually from multiple aerial vantage points to generate a DEM [10]. Photogrammetry was the method used by Aerobotics to generate the DEMs processed in this project.

2.2.2. Detecting Vegetation Through Multispectral Images

As plants grow, the chlorophyll content in their leaves increases, causing the light of different wavelengths to be absorbed more than others. The normalized difference red edge index (NDRE) and Normalized Difference Vegetation Index (NDVI) are metrics used to identify green vegetation from multispectral images. These multi-spectral images are obtained through remote sensing techniques with the use of Red Edge (RE) and Infrared (IR) sensors. NDRE is very similar to the normalized difference vegetation index (NDVI), which uses visual-band red (visible red light) rather than the red edge band. Both are good indicators of plant health, where healthier plants have higher NDRE and NDVI values. The NDRE is calculated as a ratio using Near Infrared (NIR) and RE as shown in Equation (2.1) and NDVI is also calculated as a ratio shown in Equation (2.2). Therefore, the value of NDRE and NDVI can range from -1 to 1. The wavelengths of NIR and RE are 790 nm and 735 nm respectively, while visual-band red has a wavelength of 660 nm [11] as shown in Figure 2.3. Higher NDRE values relate to higher chlorophyll content within the leaves, and visual-band red light tends to be absorbed by large clumps of leaves having high

chlorophyll content. For these reasons, NDRE would be a better indicator of vegetation health than NDVI for fully developed plants and locations of dense vegetation [12].

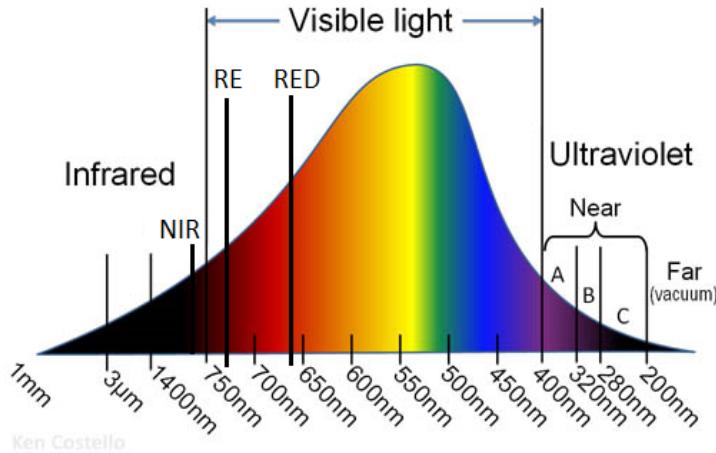


Figure 2.3: Solar spectrum [2].

$$NDRE = \frac{(NIR - RE)}{(NIR + RE)} \quad (2.1)$$

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \quad (2.2)$$

2.3. Interpolation Techniques

In this section, the process of generating DEMs by applying spatial interpolation techniques to three-dimensional (3-D) sampled ground points is discussed. Spatial interpolation is the process of estimating values of unknown data points from known sampled data points based on the principle of spatial autocorrelation, where measured points that are closer to the point being predicted will have a larger influence in determining its value [7]. This is due to the fundamental geostatistical assumption that densely packed points should have a higher similarity than points that are positioned further away from each other [13]. In Figure 2.4, the closer sampled data points (red) would have a larger influence in predicting the elevation of the yellow point than those further away (black).

Deterministic interpolation methods can be categorized into either local or global approaches. These approaches differ based on the type of measured data that is used to generate the predictions for a particular point. Local interpolators only use the measured data that is located within a specified area surrounding the point being calculated [14, 15]. On the other hand, global interpolators make use of the entire measured data set to estimate values [16]. Some common local interpolation methods that are used for earth surfaces are Inverse Distance Weighting (IDW), local polynomial, and Nearest Neighbor (NN) [13, 15]. An example of a global interpolator is global polynomial interpolation.

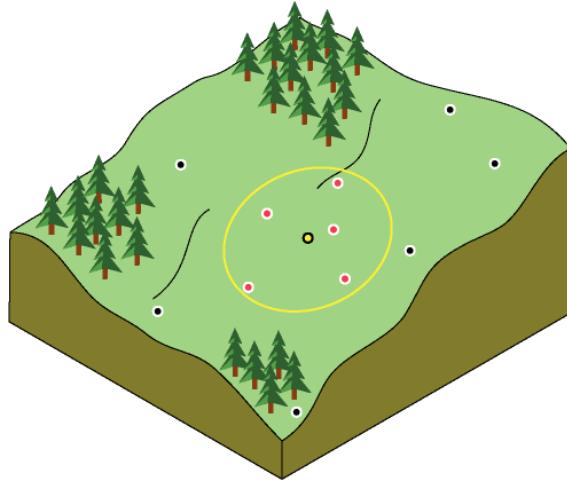


Figure 2.4: Interpolation of earth surfaces [3].

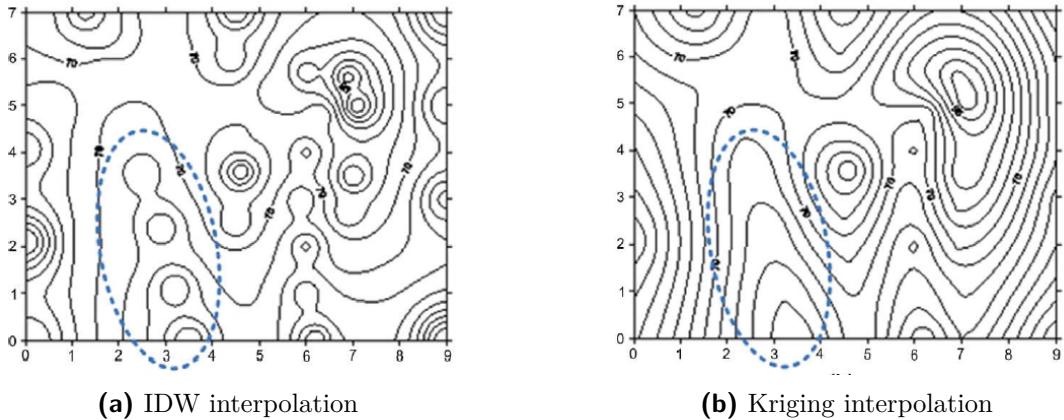


Figure 2.5: Predicted contours of the same earth surface, but the elevations were generated using different interpolation techniques [4].

Global interpolators are excellent at modelling trends in landscapes such as slopes of hills and valleys. Because of this, global methods describe the elevation of a terrain more accurately if its surface varies slowly and perform poorly for bumpy surfaces [16]. Geostatistical interpolation techniques, such as the Kriging method utilizes the spatial autocorrelation among the measured points to estimate values [15, 17, 18]. Additionally, these techniques can provide a measure of the variance for their predictions.

It is important to note that different interpolation techniques can produce significantly different results from the same set of sampled data [13]. Figure 2.5 shows contour maps of the same terrain for which elevations predicted through different interpolation techniques. Different interpolation methods will be compared to determine the optimal interpolation approach for this project.

2.3.1. Kriging

The Kriging method is a widely used geostatistical interpolation approach to predict the elevation of a terrain. Previous studies have concluded that the Kriging method is the best approach for interpolating earth surfaces [4, 17, 19]. Global trends can be modelled accurately if sampled points outside of the search window are used to calculate weights [4]. There are two main assumptions for the Kriging method, namely the sampled data is stationary, meaning that the sampled data's mean, variance and covariance does not change over time and, secondly, a sufficient number of sample points are available to produce a constant variogram function ($\gamma(s_1, s_2)$) [4]. Generating a Kriging model involves the following steps: examining the entire sampled data points, calculating covariance values, fitting a constant variogram function to the covariance values, generating matrices of Kriging equations and solving them. There are multiple Kriging equations, all of which use variations of the formula below [15]:

$$\hat{Z}_K(s_0) - \mu_{DEM} = \sum_{i=1}^n \lambda_K(s_i)[Z(s_i) - \mu_K(s_0)] \quad (2.3)$$

The stationary mean (μ_{DEM}) is the average of all the sampled data points, assumed to be constant over the entire area of interest. The parameter $\lambda_K(s_i)$ is the Kriging weight of the sample point located at s_i . n is the number of sampled points used to calculate the predicted value ($\hat{Z}_K(s_0)$) and $\mu_K(s_0)$ is the mean of the sample points within the search window [15].

The high accuracy of terrain models that are generated through Kriging is due to the data-driven weighting functions that are used. These functions calculate weights that reduce the bias towards input values [4, 13, 15]. The Kriging weights are determined through minimising the variance of the sampled data points as follows:

$$\begin{aligned} \text{var}[\hat{Z}_K(s_0)] &= E[(\hat{Z}_K(s_0) - Z(s_0))^2] \\ &= \sum_{i=1}^n \sum_{j=1}^n \lambda_{Ki} \lambda_{Kj} C(s_i - s_j) + C(s_0 - s_0) - 2 \sum_{i=1}^n \lambda_i C(s_i - s_0) \end{aligned} \quad (2.4)$$

Where $C(s_i - s_j) = \text{Cov}[Z(s_i), Z(s_j)]$ [20].

2.3.2. Inverse Distance Weighting

Inverse Distance Weighting (IDW) is an exact interpolator that only uses a selection of the sampled data to calculate the interpolated value at the desired position [21]. Sampled data points used for the IDW estimation are chosen either through a fixed or variable search radius. If a fixed search radius is applied all sampled points within a selected range from the point being predicted will contribute to its estimation. A variable search radius

will use a fixed number of the nearest sampled data points. The predicted elevation value \hat{Z}_{IDW} at position s_0 can be calculated using the following formula:

$$\hat{Z}_{IDW}(s_0) = \frac{\sum_{i=1}^n Z(s_i)\lambda_{IDW}(s_i)}{\sum_{i=1}^n \lambda_{IDW}(s_i)} \quad (2.5)$$

In Equation (2.5), the corresponding weights $\lambda_{IDW,i}$ of sampled data points are calculated as a function of the distance $d(s_0, s_i)$ between the location of the observation at s_i and the point being estimated at s_0 as shown in Equation (2.6) [4, 22]. Larger weights are assigned to points closer to s_0 and diminish as the distance increases. The default power value p_{IDW} that is used in Equation (2.6) is 2 [21]. However, the power value can be optimized for a specific terrain by determining the p_{IDW} that minimizes the Root Mean Square Error (RSME) in predictions through cross-validation discussed in section 2.3.4 [7].

$$\lambda_{IDW}(s_i) = \frac{1}{d(s_0, s_i)^{p_{IDW}}} \quad (2.6)$$

There are some important aspects to note when predicting values with IDW. The values that are calculated through IDW are guaranteed to be within the range of the sampled data points. The IDW algorithm is sensitive to clustering and the presence of outliers in the sampled data points. Therefore, the sampled data points should be homogeneously dispersed across the area of interest, as well as have any outliers removed [22]. This method only predicts a single value per location and does not provide any variability with the estimation at the location of estimation, as is the case with Kriging [23].

2.3.3. Spline and Polynomial Interpolation

Different types of surfaces can be represented by polynomial functions of different degrees. A first-order polynomial would fit a single plane through the sampled data; a second-order polynomial can match a single bending surface; a third-order global polynomial allows for two bends; and so forth. However, when a surface has a high varying shape, a single polynomial will not fit well. Through combining multiple polynomial planes, an accurate representation of a surface can be generated as shown in Figure 2.6. Spline interpolation is a non-geostatistical interpolation method that uses a special kind of piecewise polynomial interpolation called a spline. A single spline would consist of multiple polynomials with the same degree p_p . For degree $p_p = 1, 2$, or 3 , a spline is called linear, quadratic or cubic respectively. The polynomials of a spline are fitted together, ensuring a smooth transition across polynomials to describe an entire surface [15]. The locations of the discrete known data points are where the polynomials join and are called knots. The estimation of the spline interpolation is greatly affected by the selection of knots since they are used in establishing the various polynomial functions of the spline [16]. Furthermore, the spline estimations are highly susceptible to outliers in the knots, especially at the edges. Polyno-

mials with a degree of 3 are most commonly used for the interpolation of earth surfaces [15].

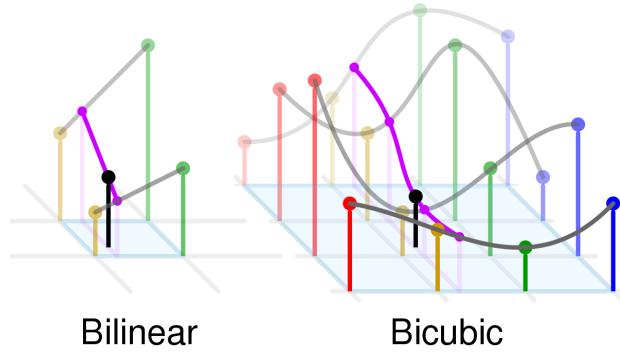


Figure 2.6: Illustration of how polynomials of different degrees can be used to represent a surface. [5].

Equation (2.7) shows the multivariate polynomial function with a degree of 1. Four observed points ($f(Q_{11}), f(Q_{12}), f(Q_{21})$ and $f(Q_{22})$) can be used in order to determine the constants (a_0, a_1, a_2 and a_3) of Equation (2.7) as shown in Equation (2.8). Using a multivariate spline interpolator with a degree of 1 would result in bilinear interpolation estimates. Bilinear interpolation is the extension of linear interpolation for data with two spatial dimensions. An estimation through bilinear interpolation is achieved by initially interpolating linearly in one direction, and then again in the other direction. The order in which the linear interpolation is done will result in the same bilinear prediction.

$$f(x, y) \approx a_0 + a_1x + a_2y + a_3xy \quad (2.7)$$

$$\begin{bmatrix} 1 & x_1 & y_1 & x_1y_1 \\ 1 & x_1 & y_2 & x_1y_2 \\ 1 & x_2 & y_1 & x_2y_1 \\ 1 & x_2 & y_2 & x_2y_2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} f(Q_{11}) \\ f(Q_{12}) \\ f(Q_{21}) \\ f(Q_{22}) \end{bmatrix} \quad (2.8)$$

2.3.4. Evaluating Interpolation Accuracy

The mean error (ME), the mean relative error (MRE) and the root mean square error (RMSE) are commonly used metrics that are used to determine the accuracy of interpolation techniques. MRE is an important metric because, unlike RMSE and ME, it provides a relative indication about the actual data. The RMSE is the most commonly used metric for interpolation accuracy [7, 24, 25]. Equations (2.9) - (2.11) can be used to calculate these metrics [26]:

$$RMSE = \sqrt{\frac{1}{N} \sum_{k=1}^N (z_p - z_o)^2} \quad (2.9)$$

$$ME = \frac{1}{N} \sum_{k=1}^N (z_p - z_o) \quad (2.10)$$

$$MRE = \frac{RMSE}{\Delta} \quad (2.11)$$

Where z_p and z_o is the predicted and observed values respectively. Δ is the range and equals the difference between the maximum and minimum observed data.

2.4. Previous Work Done

2.4.1. Accurate Digital Elevation Modeling

With the construction of DEMs, it is essential to have enough samples to accurately depict the terrain. Numerous studies have conducted research in determining the number of sampled data points required for accurate DEM interpolation , which allows a large data set to be reduced into a more manageable data sets that do not cause a significant decrease in accuracy [25, 27, 28]. Some techniques that are used to reduce data are uniform, curvature, grid and random data reduction.

Uniform data reduction uniformly divides the area of interest into equally sized square cells and only uses a single sampled data point from each cell. The selection of the sampled data used in uniform data reduction can be random or based on the values.

Curvature data reduction retains sampled data points based on the curvature caused by surrounding points. Accuracy of surface curves is maintained by retains sampled data points within high curvature regions. Observations located in flatter areas are more likely to be removed.

Grid data reduction divides the area of interest into a grid structure. Observations are assigned to the individual grids and the median point of each grid is selected to represent data points belonging to that grid.

Random data reduction removes data points randomly, each with an equal chance of being removed until the desired number of points have been reached [25].

A research study was conducted to compare the four data reduction algorithms

mentioned above. The interpolation methods that were used to test the accuracy of the reduced data sets were IDW, Kriging, Trilinear interpolation (TLI) and Thin Plat Spline (TPS). The study concluded that the best reduction algorithm that produces the most accurate elevation predictions was the uniform algorithm. It was better than the other three algorithms at each percentage of observations retained. The grid and curvature algorithm were second and third respectively. The random algorithm resulted in the worst elevation accuracy. The study also found that the accuracy of the interpolated terrain only started to significantly decrease when the data set was reduced to below 50 percent [25]. These results are supported by other previous studies [29,30]. Thus, the best elevation model is constructed if the reduced data points are uniformly positioned, with at least 50 percent of the original data set included.

2.4.2. Individual Tree Detection from Canopy Height Models

It is most effective to remove the ground plane from a DSM image to produce a CHM before attempting individual tree detection. The procedure to achieve this is by first identifying and sampling the ground points from the DSM. These ground points would form a 3-D point cloud that is used to create the DTM through interpolation. The area of interest's CHM is then acquired by subtracting the DTM from the DSM [8]. It has previously been shown that individual trees can be identified from a CHM by using a local maximum algorithm. This algorithm would identify the point with the maximum elevation within a specific region as the top of a tree. The maximum is only calculated within a fixed-sized sliding window that is moved across the CHM. This window is generally chosen to be slightly larger than the average tree in the image. Identifying all the trees present in the CHM would require that the local maximum algorithm be applied to the entire area of interest. A challenge with this method is to find the optimal shape and size of the sliding window. Windows covering a smaller area are more accurate in identifying trees. However, if a window becomes too small the local maximum algorithm will erroneously detect trees. If the window becomes too large it increases the possibility of missing some trees [8,31].

2.5. Conclusion

In this chapter, several techniques and concepts that will be utilized throughout this project have been researched and discussed. This included digital elevation models, metrics used for vegetation detection, interpolation methods and how to identify individual trees from CHMs.

In Chapter 3, the data set provided by Aerobotics is described based on observations made in a geographic information system. The process of how the data from different

files are combined at various stages in the software is also discussed. This chapter will also discuss how this data is processed to remove the ground plane and detect trees using various algorithms.

Chapter 3

Software Design

In this chapter, the data set that is analyzed in this project will be examined, and the methods that will be used to process it are explained.

3.1. Introduction

For this project, data from multiple terrains was provided by Aerobotics, with aerial photos of each shown in Appendix D. Every terrain has four files containing some useful characteristics of the terrain's landscape and the vegetation covering it. This chapter starts by explaining what data is contained in the provided files and how the files were utilized to extract tree heights from the DEMs. The flow diagram in Figure 3.1 depicts the major steps executed by the software, and in-depth explanations of each step will follow in this chapter.

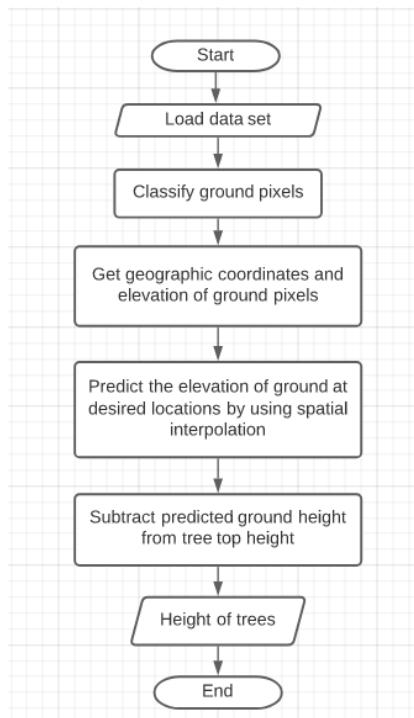


Figure 3.1: Software design flow diagram.

3.2. Inspecting the Data Set

Quantum Geographic Information System (QGIS) is an open-source desktop geographic information system application that can be used to view, edit and analyse geospatial data. The DEM, NDRE and RGBA (Red, Green, Blue and Alpha) images are all represented by raster images in the Tag Image File Format (TIFF). Aerobotics also provided additional data of some automatically detected trees in the orchard which was supplied as Geographic JavaScript Object Notation (GeoJSON) files. QGIS was used to inspect the various TIFF images and GeoJSON files, where files covering the same location are aligned with one another in a single QGIS project as separate layers. In Figure 3.2, the four files describing Terrain A as separate layers in QGIS are shown.

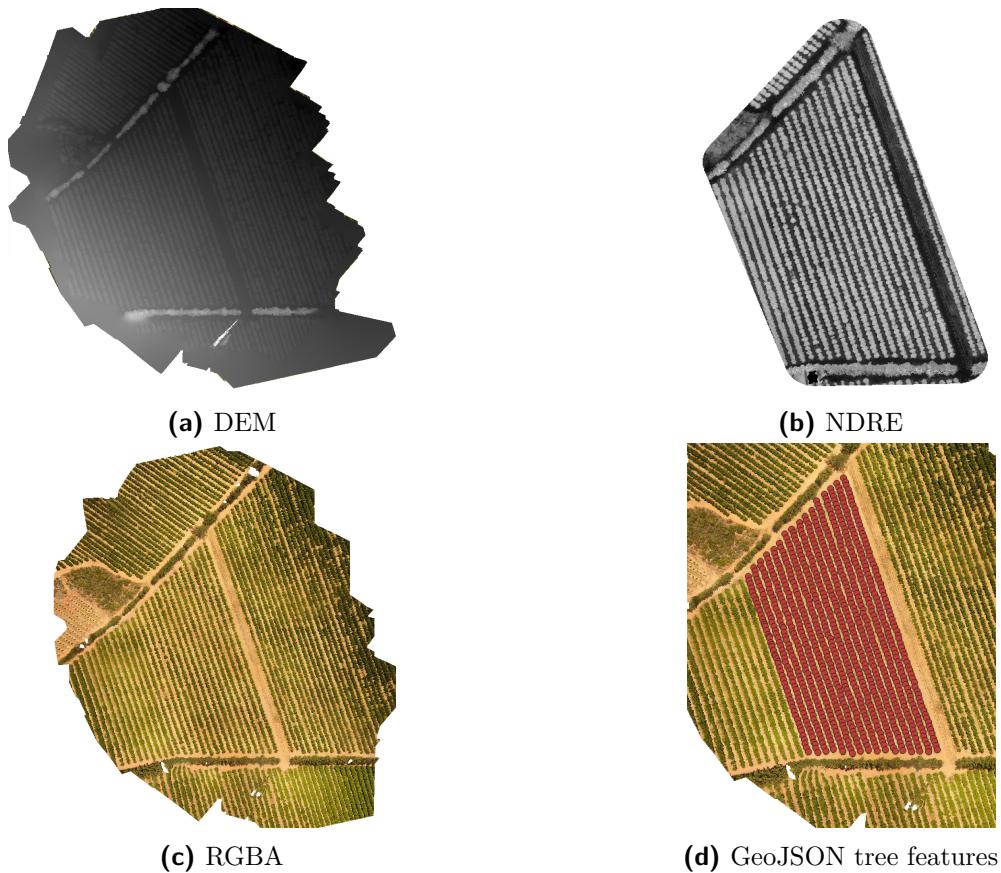


Figure 3.2: The various layers of Terrain A as viewed in QGIS.

3.2.1. The Raster Images

The DEM images that were provided stored the absolute height above sea level (true altitude) in meters in the raster bands. In QGIS the DEMs are displayed as grayscale heightmaps, with a very high elevation resolution. The raster band of an NDRE image contains the measured NDRE index values, and similar to the DEM heightmaps, the NDRE index value is represented by the intensity of the pixels. The RGBA raster images

have four raster bands, one for each colour channel and one for the alpha value. The shape of a raster image relates to the height ($rasertr_h$) and width ($rasertr_w$) of the raster image in pixels. Each raster image has a property that specifies the total area the image covers, called the extent. The latitude and longitudes (geographic coordinates) of the bottom-left and top-right pixels of a raster image is used to specify its extent. The bottom-left pixel has the minimum latitude ($Latitude_{min}$) and longitude ($Longitude_{min}$), while the maximum coordinates ($Latitude_{max}$ and $Longitude_{max}$) of the raster image is associated to the top-right pixel. Locations that were not measured and are within the extent of the raster image are represented by pixels that have a default value. The default value for a DEM image is a very small negative number, which varies for each DEM image, while the unmeasured pixels of NDRE and RGBA images have NAN value assigned to them. The raster images have many unmeasured locations at their edges, with the majority of the measured locations at the centre of the raster image. The spatial resolution of a raster image is defined by the width (res_w) and height (res_h) of its pixels. The higher the spatial resolutions of a raster image, the smaller the area a pixel covers and the more accurately it will represent terrain.



Figure 3.3: An orange tree that has reached maturity [6].

3.2.2. The GeoJSON Tree Features

GeoJSON is a type of JSON (JavaScript Object Notation) format which uses non-spatial attributes such as points, line strings and polygons to represent a single or a collection of geographical features. GeoJSON points only provide the geographical coordinates of a feature by default, but additional information of the feature can be stored as a property of the point. In the case of the data that was received from Aerobotics, the identified trees of the given terrains are stored as GeoJSON points having a single property that indicates

the diameter of the tree's canopy in meters. These GeoJSON points representing trees can be seen in Figure 3.2d. The coordinates of the GeoJSON points representing the trees are assumed to be the coordinates of the highest point of the tree canopy (treetop). All of the trees in the captured orchards are assumed to consist of only orange trees. A fully grown (mature) orange tree typically has a lollipop shape, which is characterised by having a spherical canopy on top of a narrow tree trunk as seen in Figure 3.3.

3.3. Working Between Raster Images

The raster images are loaded using the rasterio python library [32], after which the data in the raster band elements are accessible using array-like indexing. It is important to note that even though the images are of the same terrain, their extent and spatial resolution may vary, leading to differences in raster band dimensions ($raster_h$, $raster_w$). For this reason, pixels that correspond to the same location will have a different row (y) and column (x) indices in the different raster bands. Processing the data data, therefore, require obtaining values stored in the raster bands from different raster images taken of the same area of interest. By being able to convert from geographic coordinates (latitude and longitude) to array indices (x and y) of a raster band, all possible data relating to the same location can be obtained. A function that converts between geographic coordinates and array indices was created. The principals used in creating this function are explained next. The data that is stored at the 1st element (at $y = 0$ and $x = 0$) in a raster band will relate to the most North-West pixel of the raster image. An increase in y corresponds to a decrease in longitude, while an increase in x corresponds to an increase in latitude. Converting coordinates to array indices of a specific raster image requires the spatial resolution and extent of that raster image. Determining and using the spatial resolution in converting between coordinates and array indices is shown in the Equations (3.1) - (3.4).

$$res_w = \frac{(Latitude_{max} - Latitude_{min})}{raster_w} \quad (3.1)$$

$$res_h = \frac{(Longitude_{max} - Longitude_{min})}{raster_h} \quad (3.2)$$

$$Latitude = Latitude_{min} + res_w \times x \quad (3.3)$$

$$Longitude = Longitude_{max} - res_h \times y \quad (3.4)$$

3.4. Detecting Ground Points and Obtaining Their Geospatial Information

In order to establish which geographic coordinates have the elevation of the ground plane in the DEM images, the ground points must be detected. Ground points can be represented in 3-D space once the geographic coordinates and true altitude are obtained. The locations within the terrain that belong to ground can be classified by observing a combination of the data available in the different raster images that are available.

3.4.1. Detecting Ground Points from RGBA and NDRE Raster Images

The data contained within the raster band of the NDRE images are excellent for identifying ground points, as pixels belonging to ground have considerably lower NDRE index values compared to that of pixels belonging to the tree canopies. The data in the RGBA images could also possibly be used to identify the geographic coordinates of ground points, but due to the initial success with classifying ground through only NDRE index values, it was not pursued. A one-dimensional (1-D) classification algorithm was used to classify pixels by comparing the NDRE index value of a pixel to a classification threshold and then to change its raster band element value to either 0 (ground) or 1 (not ground). Figure 3.4 shows how using a classification threshold can be used to distinguish between ground and not ground pixel based on their NDRE values. Pixels that have NAN NDRE index values are classified as not ground. Figure 3.5 shows two examples of ground points that were detected by using this method.

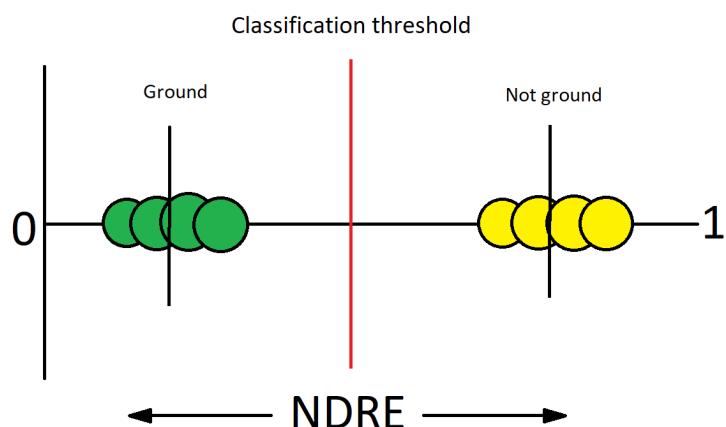


Figure 3.4: Classifying pixels based on NDRE index values. Ground points would have very similar NDRE index values close to 0, whereas the majority of non-ground pixels that belong to trees would have NDRE values closer to 1.

3.4.2. Uniformly Reducing Detected Ground Points

Based on the finding in Section 2.4.1, uniformly spaced sample points achieve the best interpolation results. An algorithm to produce uniformly spaced ground points was created. The algorithm iteratively locates a NDRE raster image pixel classified as ground (0), then sets all surrounding pixels to one if they are within a square with a fixed shape. This filtering process reduces the sample set of ground points greatly, but does considerably affect the accuracy of the interpolation, because the ground pixels are likely to be surrounded by other ground pixels that have very similar elevations. The resulting uniformly spaced sampled ground points of Terrain A and C are shown in red in Figure 3.6.

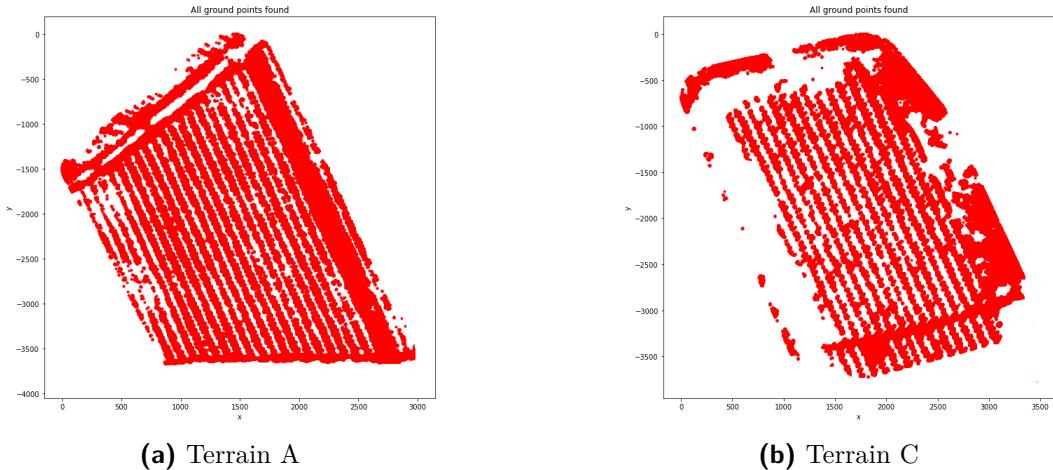


Figure 3.5: Finding ground points by classifying pixels of the NDRE raster images. Note that, due to the size of the dots used for the plot, it appears that ground covers larger area than it does.

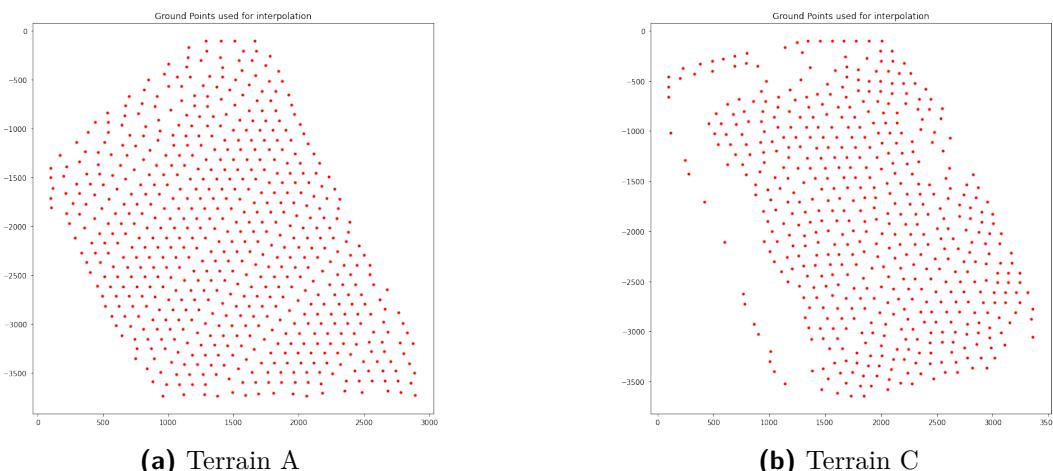


Figure 3.6: Uniformly spaced ground points that will be used in the interpolation process.

3.4.3. Obtaining True Altitude of Ground Points Through a Local Minima Algorithm

A problem with only using the NDRE images for ground identification is that ground surfaces might be covered by some vegetation like grass, freshly fallen leaves or weeds. This may erroneously be classified as not ground due to them having high NDRE index values. These locations do describe the elevation of bare earth and neglecting them might lead to less accurate results. Furthermore, there might also be some inaccuracies in the DEM data. The elevation of some locations of the ground might be slightly distorted due to being located close to higher objects, as illustrated in Figure 3.7. Therefore, the true altitude of locations should also be used to ensure that the most accurate spatial data of ground points are obtained. To achieve this, a local minima algorithm that uses the data of a DEM raster images to find the lowest ground point surrounding the NDRE classified ground points was created. Algorithm 3.1 is used to obtain the geographic coordinates and true elevation of ground points that are most likely to belong to the ground based on the DEM raster band. As mentioned previously in Section 3.2.1, unmeasured DEM pixels have a very small negative value assigned to those pixels. Thus pixels in the DEM raster band that have elevations representing unmeasured pixels must be ignored. The result of applying Algorithm 3.1 to the classified ground locations is a sample set of ground points described by their geographic coordinates and true altitude (geospatial data). This can be represented as a 3-D point cloud. In Figure 3.8 are the sampled ground points of Terrains A and C, achieved by applying Algorithm 3.1 to the data in Figures 3.6a and 3.6b.

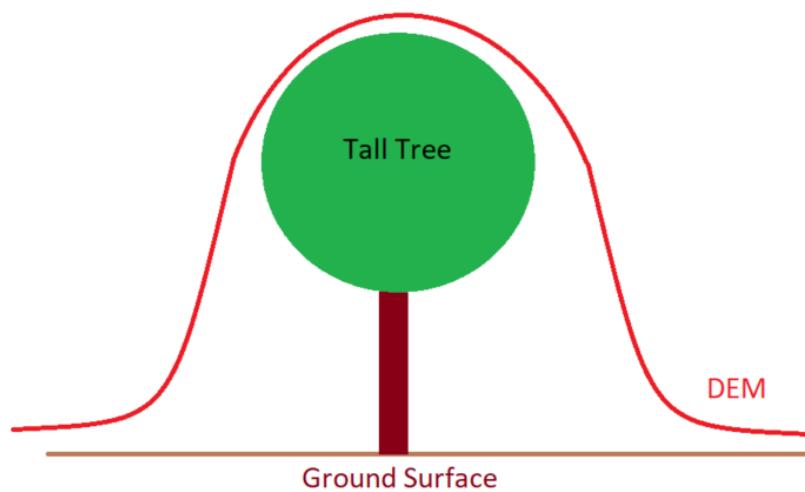


Figure 3.7: Distortion in ground surface elevation.

Algorithm 3.1: Local minima

```
for The number of ground points classified through NDRE do
    Get  $y$  and  $x$  of pixel classified as ground in NDRE image
    Convert  $y$  and  $x$  of ground pixel to geographic coordinates
    Convert geographic coordinates of ground pixel to  $y$  and  $x$  of DEM raster band
    Apply a fixed square search window centered at  $y$  and  $x$  in the DEM raster band
    with size (100,100).
    Locate the pixel with the lowest elevation within search window
    Obtain the geographic coordinates and true altitude (geospatial data) of lowest pixel
    Store the geospatial data in array as [latitude, longitude, true altitude]
end for
```

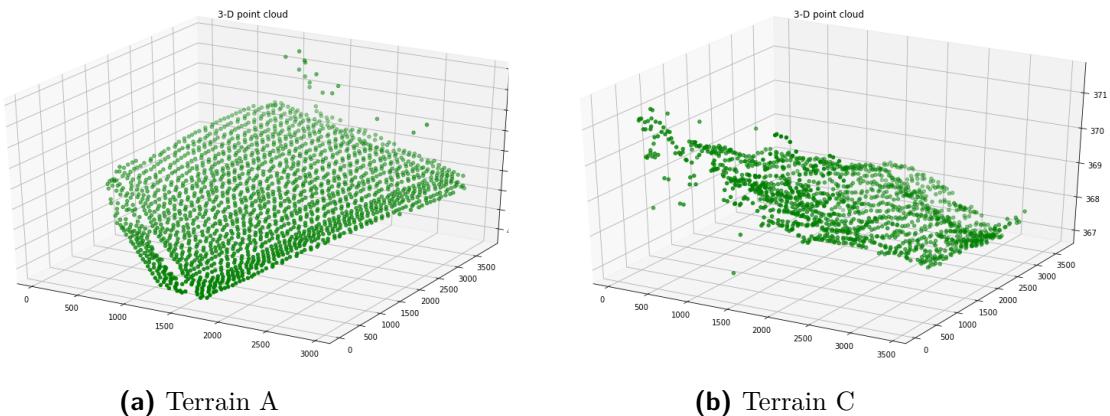


Figure 3.8: Sampled ground points with geospatial data determined by Algorithm 3.1.

3.5. Estimating the True Altitude of the Ground Surfaces Below Trees

The true altitude of the ground below trees cannot be directly obtained from the DEM raster images. Therefore the true altitude of the ground surface below the tree canopies must be predicted. The geospatial data of all the sampled ground points describes the DTM of a terrain as a 3-D point cloud. The elevation of ground points hidden below the tree canopies is predicted through spatial interpolation with the geospatial information of the sampled ground point. Universal Kriging (UK), Inverse Distance Weighting (IDW) and Linear Spline (LS) were the spatial interpolation methods used for predicting the true altitude of ground surfaces at a given geographic coordinate located within the extent of an NDRE raster image.

3.5.1. Universal Kriging

A power function for the variogram function model 2.3.1 was used for the interpolation of the ground surfaces through UK. PyKrig [33] is a python library that contains four classes, each class is capable of generating a variogram function model for pre-

dicting values and associated variances at given geographic coordinates implementing a different variation of Kriging interpolation. The UK algorithm is implemented in the `Pykrige.uk.UniversalKriging` class and the algorithm can predict values and variances at locations by utilizing the variogram function generated from the spatial information of sampled ground points. The procedure that is used to predict the true altitude of the ground through UK is explained by Algorithm 9.3.

3.5.2. Inverse Distance Weighting

Algorithm 9.4 was created and is capable of interpolating 2-D data through IDW. The algorithm uses a fixed search area and a power value of 2 for calculating the weight of a sample. The power value of 2 was selected as it is the most common power value that is used for interpolating earth surfaces. There are two special cases that must be handled by the algorithm. One of which is if s_0 is equal to s_i . In this case, the algorithm must return the elevation of the sample point at s_i . The second is where there are no sample point within the search area. In this case the algorithm must output a default value, which is the mean true altitude of sampled ground points.

3.5.3. Linear Spline

The LS interpolation was performed by using the `scipy.interpolate.interp2d` class that is part of the SciPy Python library [34]. Initializing this class correctly will create an object with a spline function capable of predicting the elevation of ground. The degree of the function must be set to 1 if ground points are to be interpolated through a linear spline function. The linear spline algorithm uses the geospatial information of sampled ground points in determining the constants of the polynomial functions that make up the spline.

3.6. Estimating Tree Heights

Estimating the height of a tree at s_{tree} ($h_{tree}(s_{tree})$) requires the true altitude of both the tree's canopy and the ground surface below its canopy. The true altitude of the canopy Z_{tree} is obtained from the corresponding DEM raster image. The true altitude of the ground ($\hat{Z}(s_{tree})$) is predicted by interpolating at the geographic coordinates of the tree being estimated. h_{tree} is estimated as follows:

$$h_{tree}(s_{tree}) = Z_{tree}(s_{tree}) - \hat{Z}(s_{tree}) \quad (3.5)$$

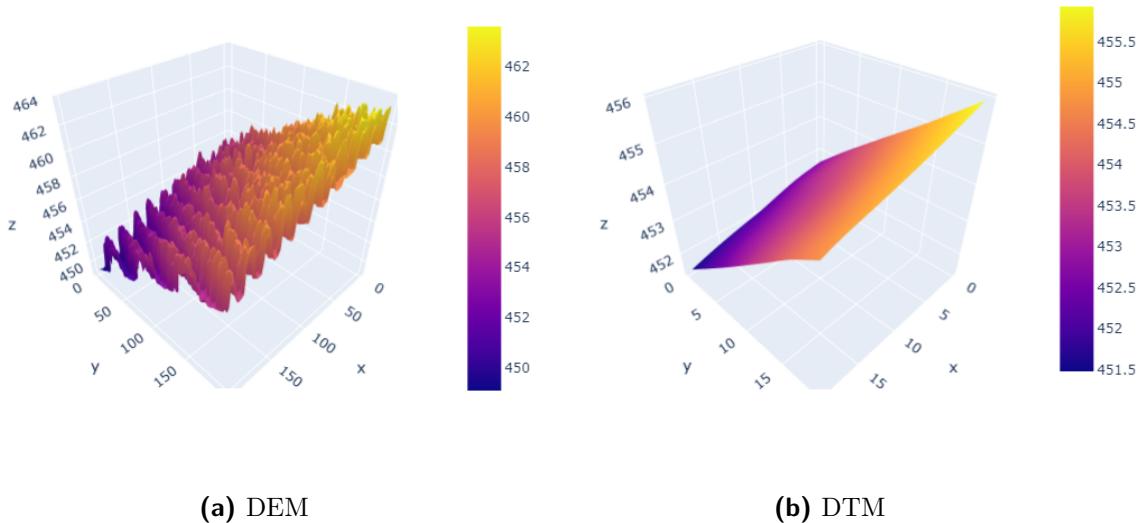


Figure 3.9: 3-D elevation plots indicating the difference between the original DEM data and the predicted DTM from a portion of terrain A (a) DEM data that includes the true altitudes of trees. (b) The predicted DTM with only the true altitude of the ground surface.

3.6.1. Generating Canopy Height Models

A DTM raster image of a terrain can be generated by using any of the three interpolation algorithms discussed up to this point Algorithms 9.3 - 9.5. The DTM covers the same extent as the NDRE raster image and have a width of DTM_w and a height of DTM_h . Figure 3.9 shows the differences between a DTM generated through Algorithm 9.3 and the original DEM data covering the same terrain from Terrain A. Once the DTM of the terrain is generated through the desired interpolation algorithm, the CHM can be obtained. Firstly, the DEM is aligned to the DTM by reducing the extent of the DEM raster band to match the extent of the DTM. Secondly, the spatial resolution of these two raster bands are matched. The spatial resolution of all the DEMs are very high. Due to the computationally-intensity involved in interpolating the DTM to match the spatial resolution of the DEM, it was decided to reduce the spatial resolution of DEM. This is not ideal, since this would result in a less accurate representation of the CHM. After this reduction in spatial resolution of the DEM, the raster bands of the DEM and the DTM would have the exact same dimensions. The CHM is then generated by simply subtracting the DTM raster band from the reduced DEM raster band. The progression of this process is shown in Figure E.1 where a CHM of Terrain A was produced with a spatial resolution 100 times smaller than that of its DEM raster image.

3.6.2. Spatial Resolution of DTM and CHM

Raster images that have a higher spatial resolution describe images of terrains in more detail because the images would consist of more pixels covering smaller areas. All of the DEM and NDRE raster images have very high spatial resolutions. If the spatial resolution of the DTM is to match the corresponding DEM, it will increase the number of points to interpolate. Increasing DTM_w and DTM_h both by a factor of k to increase the spatial resolution of the DTM requires k^2 times more values to be interpolated. It is for this reason that interpolating the DTM to the spatial resolution of its corresponding DEM would become a more computationally-intensive process. The software is designed to produce CHMs that cover the entire extent of the NDRE raster images at the best spatial resolution possible and to also produce CHMs at the exact same spatial resolution of the corresponding DEM at the cost of a smaller extent. It should be possible to generate a DTM and CHM of the full extent and at the same spatial resolution as the original DEM by interpolating multiple segments separately and stitching them together.

3.6.3. Estimating the Height of Individual Treetops

In order to estimate the height of the treetops of the trees that were labelled in the GeoJSON files, Algorithm 3.2 was developed. It is assumed that the geographic coordinates of all the tree features belongs to the treetops. Algorithm 3.2 takes the geographic coordinates of the identified tree features from a GeoJSON files and predicts the true altitude of ground directly at that geographic coordinates with any one of Algorithms 9.3 - 9.5 to estimate the height of the tree $h_{tree}(s_{tree})$. Algorithm 3.2 only predicts the height of a single point in the canopy of a tree for each identified tree.

Algorithm 3.2: Estimating tree heights

```
for The number of detected trees ( $T$ ) do
    Get geographic coordinates of treetop from the GeoJSON tree feature.
     $s_{tree} \leftarrow$  geographic coordinates of treetop.
    Convert  $s_{tree}$  treetop to DEM raster band indices.
    Get true altitude of treetop from DEM raster image.
    Use one algorithm from Algorithms 9.3- 9.5 to estimate the true altitude of the
    ground below the treetop ( $\hat{Z}(s_{tree})$ ).
     $h_{tree}(s_{tree}) \leftarrow Z_{tree}(s_{tree}) - \hat{Z}(s_{tree})$ .
    Save tree height.
end for
```

3.7. Metrics

In this section the metrics that will be used to determine the accuracy of the interpolation algorithms are discussed.

3.7.1. Accuracy of Sampled Ground Points.

To achieve the most accurate interpolation results for the ground surface under the trees, it is essential that the sampled ground points were correctly classified and that they have the correct geospatial information. Due to a large number of sample points and the 3-D nature of the data, it is easier to identify possible errors in the sampled ground points from viewing them as 3-D point clouds as seen Appendix D. A 3-D point cloud of the sampled ground points can be compared to the heightmap containing the original DEM data. Any sampled points that vary significantly from the DEM data should be investigated to determine what might have caused such a difference. It is more important that points located within the orchard are sampled correctly, as these points will have a larger influence when applying the interpolation Algorithms 9.3-9.5 that are used to estimate the true ground altitudes below the trees.

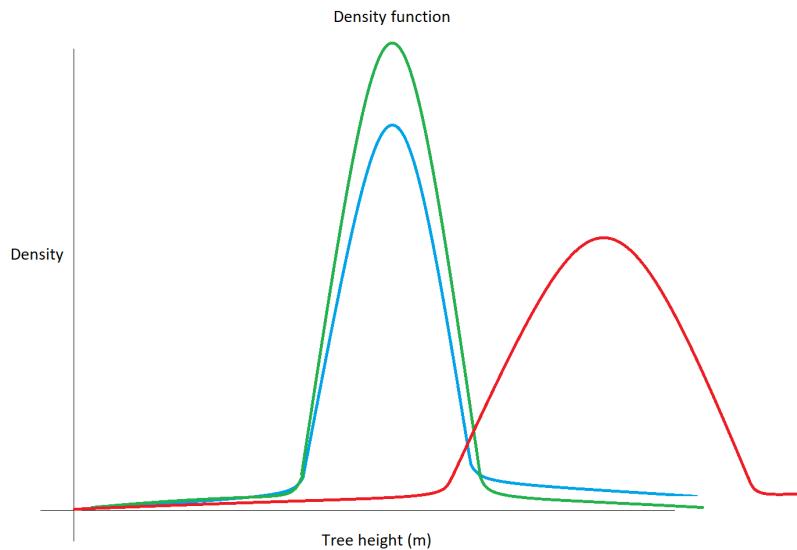


Figure 3.10: Illustration explaining how PDFs can be used to determine accuracy of predicted tree heights. Blue represents the PDF of manually measured data. The green and red PDFs represent predicted data. Accurately predicted data would resemble that of the green PDF, while poorly predicted data would look like that of the red PDF. Note that this is an illustration and not plots representing actual data.

3.7.2. Measuring the Accuracy of the Tree Height Estimations

The accuracy of the estimated tree heights is dependent on how well Algorithms 9.3 - 9.5 predict the altitude of the ground below the tree canopies. The RMSE of the tree height estimations that were made by each interpolation algorithm can be determined by comparing the estimated tree heights with tree heights calculated through the process described in Section 3.7.3. The predicted tree heights are obtained by using Algorithm 3.2 with one of the three interpolation algorithms. Since there are multiple types of terrains, it can also be established which interpolation algorithms perform better at estimating tree heights for specific types of terrains. The interpolation algorithm that performs best overall is then selected for the final design. Lower RSME values relate to better estimations. The software is designed to achieve the lowest RSME for tree height estimations. The RSME of tree height estimations is the same RSME of the true altitude of ground predictions, because of how the tree heights were estimated through Algorithm 3.2.

It is assumed that trees in the same orchard should have relatively similar heights due to them being exposed to very similar conditions throughout their growth and the high likelihood that they would have been planted at the same time. The accuracy of estimated tree heights can be tested by comparing the probability density functions (PDFs) of a large number of tree heights estimated through Algorithm 3.2 to that of a small sample of manually calculated tree heights. The calculated tree heights are obtained through the process described in Section 3.7.3. The better the PDF of estimated tree heights match the PDF of the calculated tree heights, the more accurate the predictions are assumed to be. Figure 3.10 depicts how the PDFs of estimated tree heights will be compared to the PDF of calculated tree heights. Lastly, tree heights that are estimated to be negative or higher than 6 m are considered invalid estimations.

3.7.3. Manually Calculating Tree Heights

Calculating the true altitude of ground below the tree is done, by solving Equation 3.6. This Equation 3.6 is known as the scalar plane equation that contains nearby ground points surrounding the tree being calculated. The latitude (x_{lat}), longitude (y_{long}) and true altitude (z_{el}) of at least 3 ground points are required to establish the constants (a , b , c and d) of Equation 3.6. The z_{el} of the ground below the treetop can be calculated from this equation by setting x_{lat} and y_{long} to the latitude and longitude of the treetop. Figure 3.11 shows roughly where measurements are made and how they are used to establish the height of a tree with Equation 3.6. The true altitude of the treetops that are used for the tree height calculations is the same as those used for the tree height estimations made by Algorithm 3.2, meaning that the only difference between calculated and predicted tree

heights is the way in which the true altitude of the ground was determined.

$$ax_{lat} + by_{long} + cz_{el} = d \quad (3.6)$$

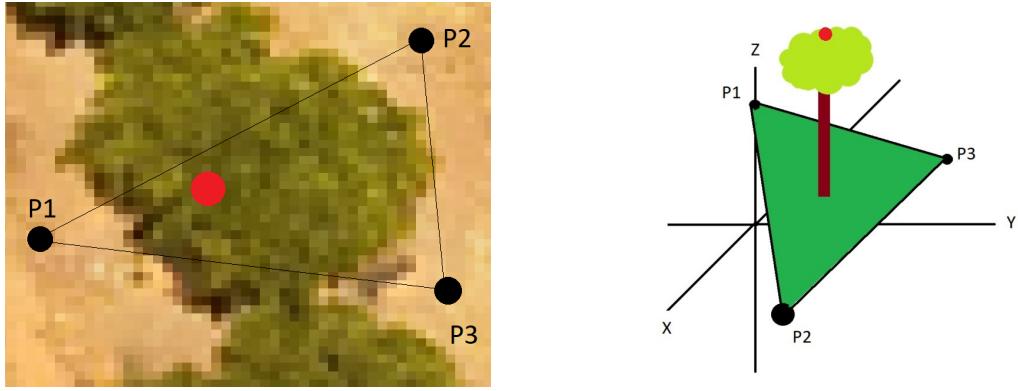


Figure 3.11: How measuring 3 points around a tree and the treetop in QGIS can be used to determine the tree's height. (a) Here the locations at which to measure is indicated through the black points (P1, P2, P3) and the one red point (treetop). (b) The black points are used to determine the plane on which the tree is on and the red dot represents its treetop.

3.8. Conclusion

In this chapter, the final software design was presented to collect data from one DEM and one NDRE raster image representing the same terrain to establish the 3-D spatial information of ground points. The ground points are identified through applying a 1-D threshold classifier to the NDRE raster band and a local minima algorithm to the DEM raster band. The elevations of ground surfaces that are covered by tree canopies are predicted through the process of spatial interpolation from the sampled ground points. By predicting all unknown ground surfaces, a DTM of the terrain is generated. Furthermore, a CHM that has the same extent as the NDRE raster image that is used for ground point classification is produced by subtracting the DTM elevations from the elevations in the DEM.

In Chapter 4 the design decisions made to optimize tree height estimations are discussed in detail.

Chapter 4

Detailed Design

4.1. Introduction

In this chapter, the rationale behind the final software design decisions are discussed. First, an explanation of how to determine the classification threshold that is used to classify ground pixels based on NDRE index values is given. The accuracy of tree height estimations from the different interpolation algorithms (UK, IDW and LS) are compared to establish which is the best for the data that is used in this project.

4.2. Determining the NDRE Classification Threshold

Tests were conducted to establish the optimal NDRE index classification threshold. These tests involved interpolating from sampled ground points that were classified using different threshold values. The threshold value that is used is determined by the average of the values in the NDRE raster band (μ_{NDRE}). When the threshold was reduced from half the μ_{NDRE} to a third of the μ_{NDRE} , the number of points that are used for interpolation drops by more than 50%, this decrease in sample points resulted in less accurate estimations from the IDW and LS algorithms. The estimation accuracy of the UK algorithm did not perform worse with lower threshold values but did perform worse at thresholds closer to μ_{NDRE} . It was decided that using a classification threshold of a value equal to $\frac{\mu_{NDRE}}{2}$ produces good results.

4.3. Selecting the Interpolation Method

4.3.1. Performance of UK, IDW and LS at Different Terrain Conditions

The ground estimations accuracy of Algorithms 9.3 - 9.5 were tested in order to determine how the methods perform under different terrain conditions. These terrain condition were bumpy, slightly sloped, sloped and flat. For each of the aforementioned terrains, the elevation of ground directly below 10 treetops were predicted using the Algorithms 9.3 -

9.5 from which the treetop heights were estimated. Estimated and manually calculated tree heights are given in Tables C.6, C.7, C.8 and C.9. The resulting RMSE values are summarised in Table 4.1 and Figure 4.1.

Table 4.1: RMSE of tree height estimations for different types of terrains (lower is better).

type	RMSE (m)		
	UK	IDW	LS
Bumpy	0.731	0.664	0.685
Sloped	0.380	0.815	0.433
Slightly sloped	0.176	0.207	0.445
Flat	0.051	0.063	0.606

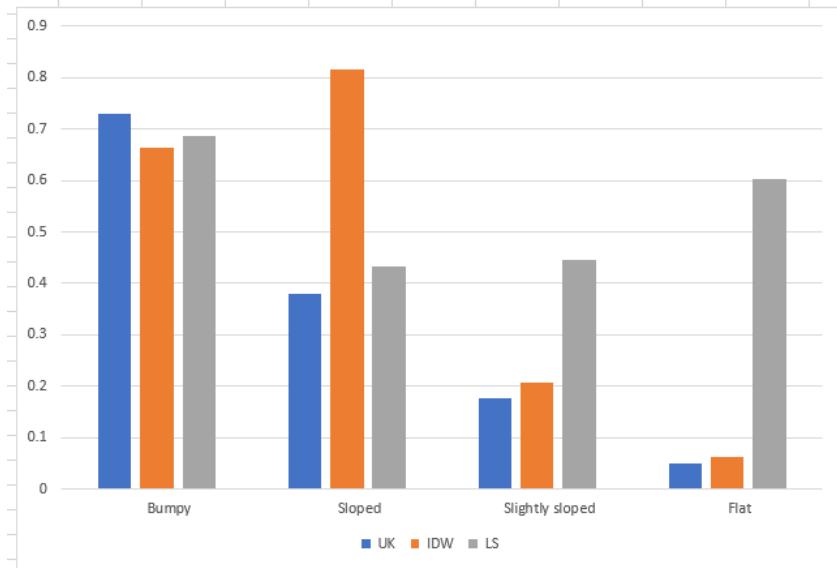


Figure 4.1: RMSE at different terrain conditions (lower is better).

UK achieved the best results for sloped, slightly sloped and flat terrains with RMSE values of 0.388 m, 0.176 m and 0.0505 m respectively. Tree height predictions for trees on the bumpy surface performed poorly with each interpolation algorithm achieving RMSEs larger than 0.66 m. The IDW method performed the best at interpolating the bumpy terrain, obtaining an RMSE of 0.665 m. IDW and UK performed very well at estimating tree heights for flat surfaces, while LS performed the worst with an RMSE roughly ten times larger than the others. LS is also the only interpolation algorithm that had tree height estimations below 0 m and higher than 6 m. On inspection of these invalid tree height estimations, it was discovered that there were very few sample points surrounding the locations at which the true altitude of ground was predicted.

4.3.2. Comparing Interpolation Methods through PDFs

For both Terrain A and C, the height of 100 treetops were estimated using Algorithms 9.3-9.5 to predict the true ground altitude below the treetops. Resulting in a total of 600 treetop height estimations, with each algorithm being used to make 100 treetop height estimation in both Terrain A and C. The Probability Density Function of the estimated treetop heights are compared to that of the calculated heights as seen in Figure 4.2.

For Terrain A, the PDFs of the UK and LS estimated treetop heights matches the calculated tree heights distribution function very well, as seen in Figure 4.2a. The mean of the estimated treetop heights for both UK and LS are within 0.03 m of the manually calculated tree height mean and almost all the treetop estimated heights from using UK and LS were within the PDF of the calculated heights. The PDF of treetop heights estimated using IDW for ground prediction had the worst resemblance to the manually calculated treetop height PDF. From the 300 treetop height estimations that were made for Terrain A not a single estimation was invalid.

For Terrain C, IDW interpolation seems to have achieved the best true ground altitudes below treetops based on how well the estimated treetop height PDF in Figure 4.2b resembles that of manually calculated treetop heights. The treetop heights that were estimated through interpolating the ground points with IDW had a mean and standard deviation closest to that of the measured data with values of 2.82 m and 0.399 m respectively. LS was the the only method that predicted true ground altitudes resulting in invalid tree height estimations, with a total of eight invalid treetop height estimations. All of the invalid predictions made by the LS, was at treetops located where there were very few sampled ground points.

4.4. The GitHub Repository

The code that was written for this project can be accessed at: <https://github.com/GCSchmidt/Skripsie-2020>. The IDW.ipynb, Kriging.ipynb and Linear_Interpolation.ipnb Jupyter notebooks contains the IDW, UK and LS algorithms respectively.

4.5. Conclusion

The classification threshold for classifying ground pixels based on NDRE index value was chosen to be $\frac{\mu_{NDRE}}{2}$. The UK interpolation method outperformed the IDW and LS methods for predicting true ground elevation below treetops which resulted in better

tree height estimations. For this reason, Algorithm 9.3 is selected to be used in the final software design.

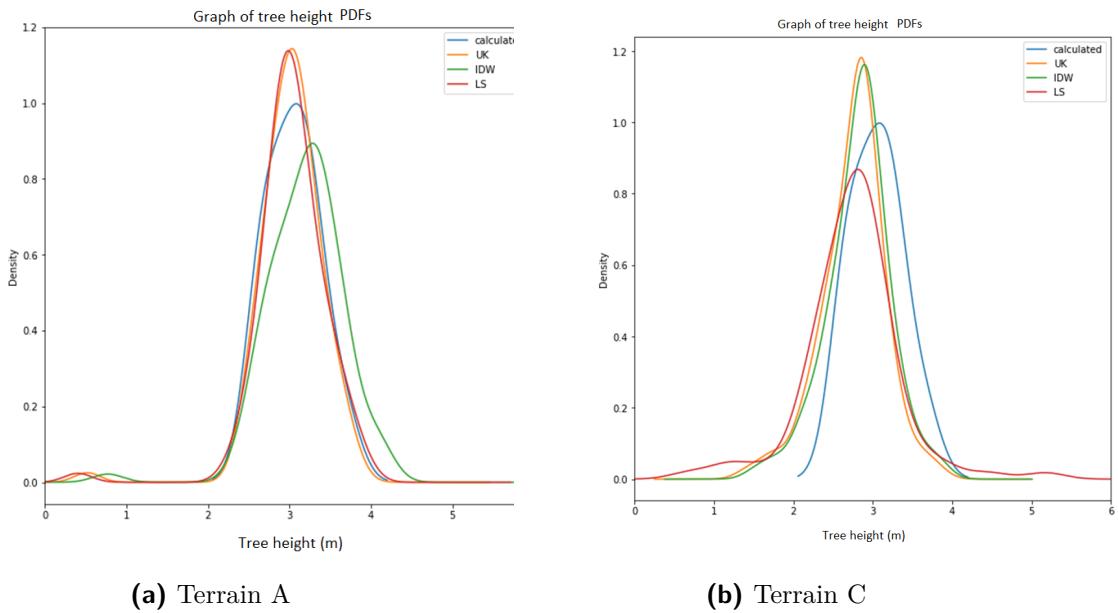


Figure 4.2: Comparing PDFs of the estimated tree heights to that of the manually calculated tree heights (blue). (a) The calculated tree heights from Table C.1. (b) The calculated tree heights from Table C.3.

Chapter 5

Results

5.1. Introduction

In this chapter the results of estimating tree heights by generating DTMs outperformed the IDW and LS methods for estimating the tree heights UK are examined. The effect of CHMs with different spatial resolutions and extents are discussed. Firstly the 3-D point clouds that represent the ground surfaces of Terrains A-E, are analysed to spot anomalies in the sampled ground points. The accuracy of the DTMs and tree height estimations by interpolating the ground surface across the entire extent of the NDRE raster band images with Algorithm 9.3 are presented. The chapter concludes with the results of CHMs at different spatial resolutions and extents of Terrains A-E.

5.2. Accuracy of Ground Point Sampling

The sampled data of Terrains A and B had some points that did not resemble points on their ground surfaces as is shown in Figures D.1b and D.2b. For Terrain A, the error was caused by elements in the NDRE raster image that had a value of zero, but was located within a large tree, which resulted in a point being incorrectly classified as ground. The incorrectly sampled ground points from terrain B was due to a large bush for which some low NDRE index values appeared in its leaves. This led to true altitude of the bush to be used and not that of the ground. The 3-D point cloud of Terrain C is shown in Figure D.3b) and appeared to have no incorrectly sampled ground points, but there are relatively large areas with no sampled ground points. On inspection, it was determined that some ground pixels covered by the shade of trees were not classified as ground because of the high NDRE index values. The lack of direct sunlight on the ground surface causes an increase in NDRE index values of ground. The 3-D point clouds of Terrains D and E showed no signs of having any inaccurately sampled ground points and are shown in Figures D.5b and D.4b respectively. Both of these terrains had very little vegetation other than the orchards covering their surface and plenty of exposed ground surfaces. Overall the ground points located within the orchards were sampled correctly.

5.3. Accuracy of Treetop Heights and True Ground Altitude Estimations

For Terrains A-E, the treetop height of 10 trees were estimated using Algorithm 3.2 with the three different interpolation Algorithms 9.3-9.5. The 50 estimated tree heights of each interpolation algorithm and the 50 manually calculated tree heights are summarised in Tables C.1-C.5. Based on these results, the RMSE values shown in Table 5.1 were determined. The poor estimation accuracy of Terrain C is due to the large ground areas covered by shade from large trees surrounding the orchard. The shade causes the ground pixel to have higher NDRE index values, which lead to them not being sampled as ground points. UK outperformed IDW and LS for every terrain except for Terrain C, where IDW was the most accurate.

Due to of how the tree heights were estimated, the RMSE values in Table 5.1 also confirms that the DTMs that are generated through the UK method were the most accurate representations of the true ground altitude located within the orchards. The contour plots of the two terrains that achieved the lowest RMSE values for the UK method are shown in Figure 5.1. The other contour plots are seen in Figures F.1.

The PDFs of 100 estimated treetop heights using UK (Yellow) for Terrains A-E are shown in Figures G.1a - G.1e. PDFs of estimated heights showed similarities to the small set of manually calculated values. The PDFs of estimated treetop heights for Terrains A, B and C fit almost completely in the manually calculated PDFs, indicating high accuracy in the estimations. From the 500 treetop heights estimated by using the UK method to predict ground, not a single invalid estimation was made.

Table 5.1: RMSE of tree height estimations for Terrains A-E.

Terrain	RMSE (m)		
	UK	IDW	LS
A	0.133	0.254	1.439
B	0.187	0.318	7.323
C	0.692	0.591	0.652
D	0.265	0.308	0.250
E	0.288	0.406	0.312

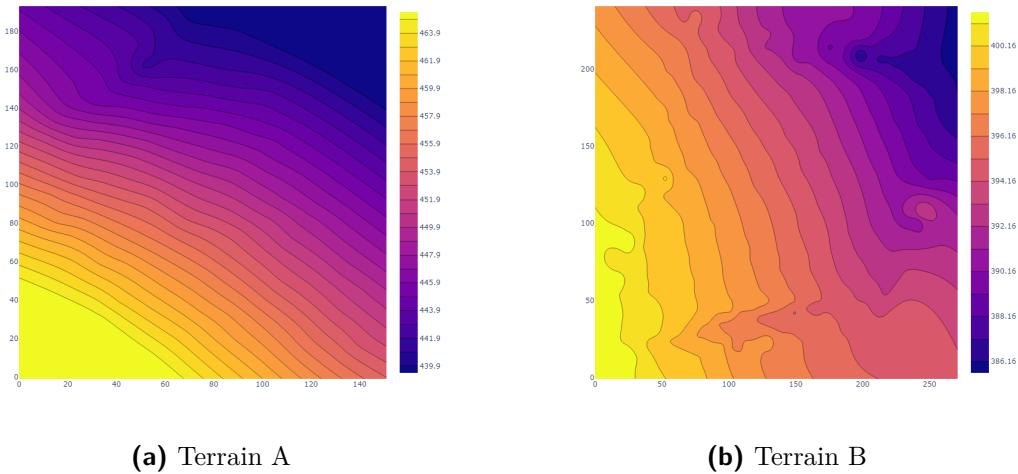


Figure 5.1: Contours of DTMs.

5.4. Producing Canopy Height Models

CHMs covering the full extent of Terrains A-E were produced according to the steps explained in Section 3.6.1 and are shown in Figure H.1. The loss in tree height estimation accuracy can be seen in Figures G.1a - G.1e that show the PDFs of treetop heights collected from CHMs. CHMs with a spatial resolution 100 times lower than the corresponding DEM was produced for Terrains A-D. The highest spatial resolution CHM for Terrain E was 200 times lower than the corresponding DEM. The treetop heights obtained from the CHMs are from the exact same treetops used to produce the UK estimated treetop heights. For this reason, the PDFs of treetop heights from the CHMs should be the exact same as the PDFs of treetop heights estimated by applying UK directly below the treetop. The DEM of Terrain E had the highest spatial resolution of all the terrains and therefore resulted in the least accurate CHM as seen in Figure H.1d. This is also seen in Figure G.1e, where the CHM PDF contains several invalid measurements and does not resemble the manually calculated PDF or the UK estimated PDF. A CHM covering the full extent of Terrain C was able to be produced with a spatial resolution closest to its corresponding DEM. This was possible because Terrain C has a DEM with the lowest spatial resolution of all the terrains that were considered. Figure 5.2 shows the CHMs of Terrain C at different spatial resolutions. Generating CHMs at the same spatial resolution of the corresponding DEMs required a reduction in the extent covered due to computational intensity involved with interpolating the DTM at such high spatial resolution. Figure 5.3c shows the resulting CHM of Terrain A at the same spatial resolution of the corresponding DEM.

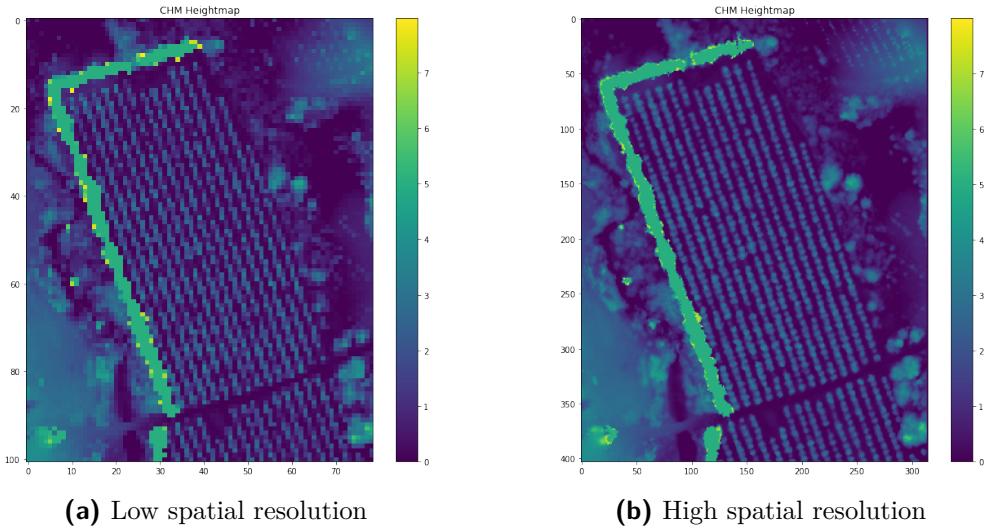


Figure 5.2: CHMs of Terrain C with different spatial resolutions (a) The CHM with a spatial resolution 100 times lower than its DEM. (b) The CHM with a spatial resolution 25 times lower than its DEM.

5.5. Individual Tree Detection

Individual Tree Detection (IDT) was attempted by applying a local maxima algorithm across CHMs at different spatial resolutions. Many trees were missed by the detection algorithm when using the CHMs with low spatial resolution. For CHMs with high spatial resolutions, the same tree would sometimes be detected several times. Multiple results are represented visually in Figure 5.4. The local maxima algorithm was able to locate the treetops accurately but to achieve IDT, the algorithm would require further optimisation. Optimising the local maxima algorithm would involve adjusting the size of the search window based on the spatial resolution of the CHM. Since this was not the main aim of this project, no further optimisation was attempted

5.6. Conclusion

The ground points that are located between the trees belonging to the orchards were sampled using NDRE and a local minima algorithm with no clear anomalies, except for situations where low NDRE values were erroneously in the middle of a tree. The PDFs of treetop heights were estimated by using Algorithm 3.2 that predicted the ground altitudes with the UK method (Algorithm 9.3) showed similarities with the smaller samples of manually calculated tree heights. The CHMs generated by subtracting the DTMs from the lower resolution DEMs resulted in much less accurate representations of the tree heights as expected. CHMs with high spatial resolution were generated at the cost of covering smaller areas, due to the computational difficulty that were experienced when interpolating too many locations.

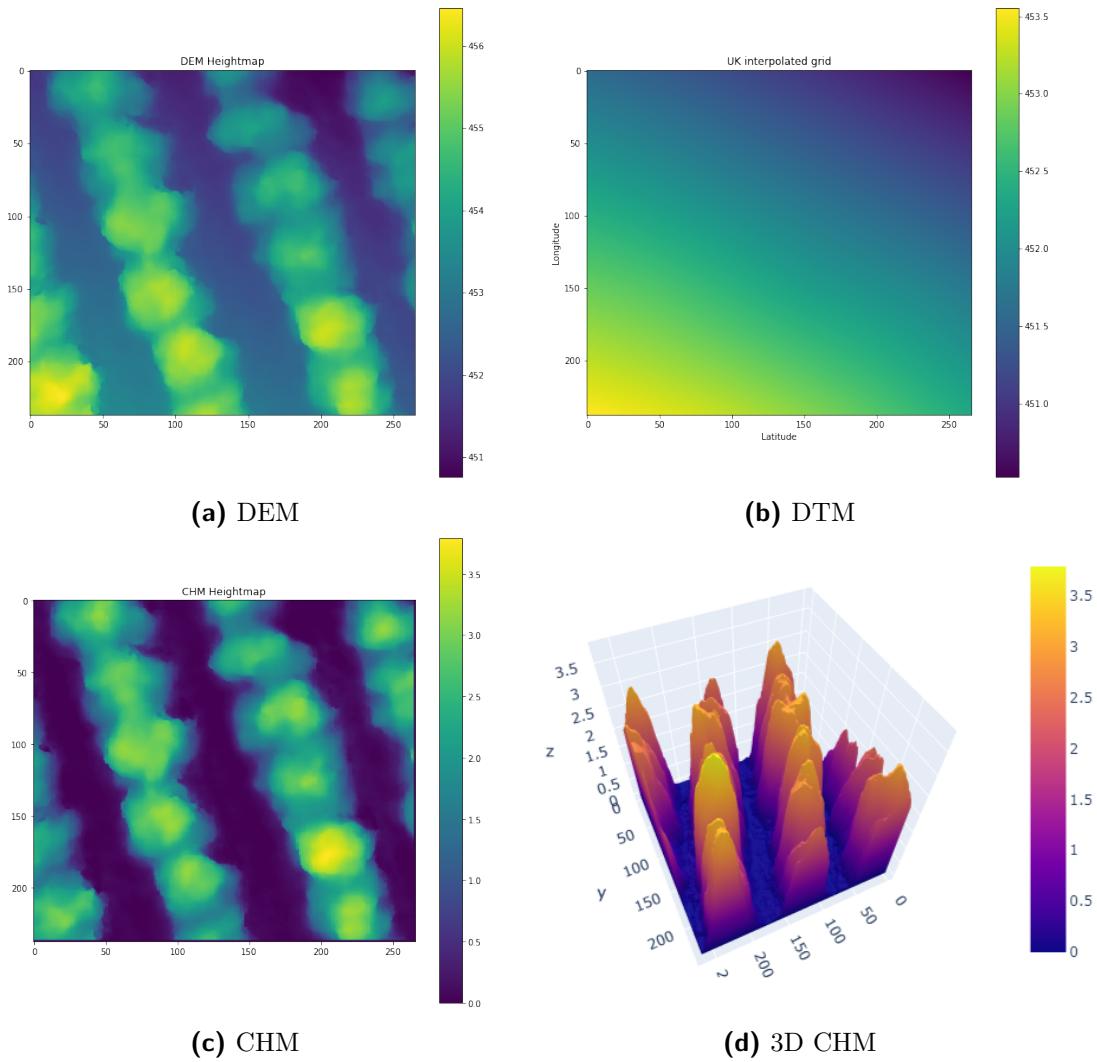


Figure 5.3: Using DTM with high spatial resolution to produce CHM of Terrain A.

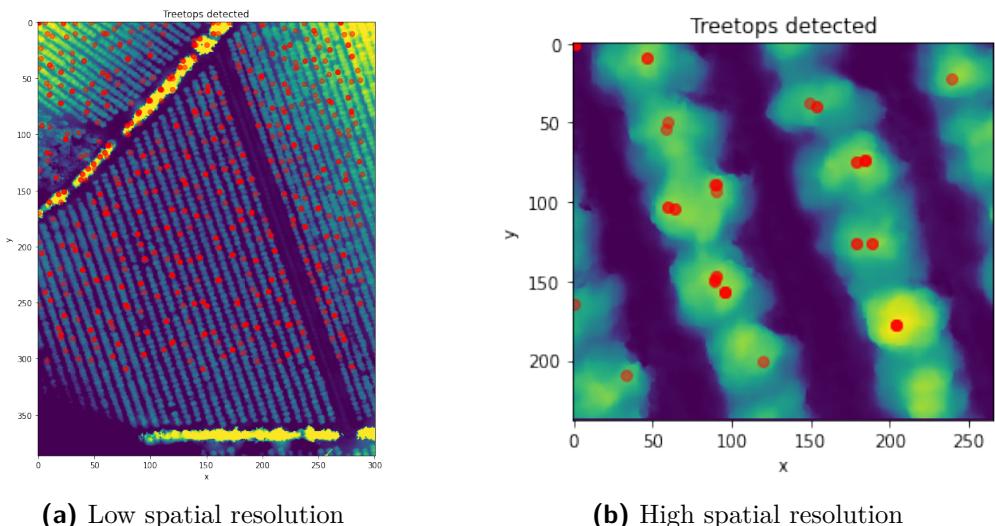


Figure 5.4: Individual Tree Detection from CHMs.

Chapter 6

Summary, Conclusion and Future Work

6.1. Summary

Software algorithms were designed to determine tree heights in orange tree orchards from raster images that contain geospatial data. The software automatically detects ground surfaces within and surrounding the orchards using a threshold classifier and a local minima algorithm. The 3-D spatial data of detected ground points are obtained from the NDRE and DEM raster images of the area of interest. The Universal Kriging (UK) method is then applied to predict the altitude of the ground below the trees from which the tree heights can be estimated using the digital elevation model raster image

A collection of treetop heights can be determined based on their geographical coordinates resulting in a PDF of estimated results. Entire orchards can be represented by a Canopy Height Model (CHM) of varying extents, where the CHMs with lower extents can represent the height of trees more accurately due to having higher spatial resolutions. This was as a result of the need to downscale the images due to computational limitations.

6.2. Conclusion

The aim of this project was to extract tree heights from Digital Elevation Models and other multispectral images. This required automatically collecting geospatial data of ground points that are used to predict the elevation of the ground below trees. Once the ground below trees are predicted, the ground plane could be removed resulting in a CHM. The height of treetops could be accurately estimated for Terrains A, B, D and E, but not for Terrain C. Inaccurate tree estimations were due to the misclassification of ground points that were covered by shade from large trees surrounding the orchard. By supplying the software with a set of geographic coordinates of treetops a PDF of the estimated treetop heights could be plotted. The ground altitude predictions that were obtained by using the UK through the UK interpolation method (Algorithm 9.3). DTMs within the extent of the corresponding NDRE raster image could be generated with varying extents and spatial resolutions. DTMs covering the full extent of the NDRE image had low spatial

resolutions. DTMs that have spatial resolutions equal to the corresponding DEM could be generated at a lower extent. The DTMs can be viewed as contours as well as heightmaps. Producing DTMs with large extents were only possible at low spatial resolution due to the computational intensity that is involved in interpolating at many locations. The desired spatial resolution for DTMs was achieved at the cost of having a smaller extent. With the success in predicting the true altitude of the ground surface below the trees the ground plane was able to be removed from the DEM to produce a CHM. Similarly to the DTM, CHMs with high spatial resolutions could only be achieved if it covers a small area that does not need to many ground points to be interpolated. Individual tree detection from the CMS was attempted, but trees were either missed or detected multiple times.

6.3. Future Work

As mentioned in Section 3.6.1 is preferred to produce a DTM that covers the entire area of the orchard and that has the same spatial resolution as the corresponding DEM. This could possibly be achieved through a different approach than the one used in this project. Rather than trying to interpolate the entire DTM from a 3-D point cloud that is made from known ground points, the elevation of all ground points could be used to interpolate the ground below the trees. A better classification method could be used for ground identification to ensure ground surfaces that are covered by shade are correctly classified. The RGBA raster images could possibly be utilized in improving the classification of ground as this would add three more dimension from which ground can be distinguished by using a method such as HSV thresholding which can identify colours independent of brightness. IDT and estimating other parameters of the trees could be done in future projects.

Bibliography

- [1] EarthLab, “Lesson 4. canopy height models, digital surface models digital elevation models - work with lidar data in python,” [Online; accessed November 1, 2020]. [Online]. Available: <https://www.earthdatascience.org/courses/use-data-open-source-python/data-stories/what-is-lidar-data/lidar-chm-dem-dsm/>
- [2] [Online]. Available: http://www.chemistryland.com/CHM107Lab/Exp08_UV/Lab/LightSpectrum.jpg
- [3] esri, “Analyzing the surface properties of nearby locations,” [Online; accessed November 1, 2020]. [Online]. Available: <https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/analyzing-the-surface-properties-of-nearby-locations.htm>
- [4] Z. R. Detweiler and J. B. Ferris, “Interpolation methods for high-fidelity three-dimensional terrain surfaces,” *Journal of Terramechanics* 47, vol. 47, no. 4, pp. 209–217, 2010.
- [5] , “Bicubic interpolation,” [Online; accessed November 1, 2020]. [Online]. Available: https://en.wikipedia.org/wiki/Bicubic_interpolation
- [6] The garden of Eden, “How to grow an orange tree from seed,” [Online; accessed November 1, 2020]. [Online]. Available: <https://1.bp.blogspot.com/-ntes9Kl5oqs/UStdphMp3mI/AAAAAAAQPo/NP-Gy1sHi0s/s1600/Orange-Tree.jpg>
- [7] L. Li, M. A. Nearing, M. H. Nichols, V. O. Polyakov, D. P. Guertin, and M. L. Cavanaugh, “The effects of dem interpolation on quantifying soil surface roughness using terrestrial lidar,” *Soil and Tillage Research*, vol. 198, no. 1, pp. 1–10, 2020.
- [8] M. Mohan, C. A. Silva, C. Klauberg, P. Jat, G. Catts, A. Cardil, A. T. Hudak, and M. Dia, “Individual Tree Detection from Unmanned Aerial Vehicle (UAV) Derived Canopy Height Model in an Open Canopy Mixed Conifer Forest,” *Photogrammetric Engineering Remote Sensing*, vol. 8, no. 9, p. 340, 2017.
- [9] Z. Li, C. Zhu, and C. Gold, *Digital Terrain Modeling: Principles and Methodology*. Boca Raton: CRC press, 2005.
- [10] G. Geography. (2020) What is photogrammetry? [Online]. Available: <https://gisgeography.com/what-is-photogrammetry/>

- [11] M. A. Hassan, M. Yang, A. Rasheed, G. Yang, M. Reynolds, X. Xia, Y. Xiao, and Z. He, “A rapid monitoring of ndvi across the wheat growth cycle for grain yield prediction using a multi-spectral uav platform,” *Plant Science*, vol. 282, no. 1, pp. 95–103, 2019.
- [12] J. Rubin. (2019) Ndre vs ndvi: What’s the difference? [Online]. Available: <https://blog.aerobotics.co/ndre-vs-ndvi-whats-the-difference-4f507662823>
- [13] P. Arum, “A comparative analysis of different dem interpolation methods,” *The Egyptian Journal of Remote Sensing and Space Science*, vol. 16, no. 2, pp. 133–139, 2013.
- [14] Esri. (2020) Deterministic methods for spatial interpolation. [Online]. Available: <https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/deterministic-methods-for-spatial-interpolation.htm>
- [15] J. Li and A. D. Heap, *A Review of Spatial Interpolation Methods for Environmental Scientists*. Australia: Geoscience Australia, 2008.
- [16] P. A. Burrough and R. A. McDonnell, *Principles of Geographical Information Systems*, 1998.
- [17] T. H. Meyer, “The discontinuous nature of kriging interpolation for digital terrain modeling,” *Cartography and Geographic Information Science*, vol. 31, no. 4, pp. 209–216, 2004.
- [18] M. Georges, “Principles of geostatistics,” *Economic Geology*, vol. 58, no. 8, pp. 1246–1266, 1963.
- [19] D. Zimmerman, C. Pavlik, A. Ruggles, and M. P. Armstrong, “An Experimental Comparison of Ordinary and Universal Kriging and Inverse Distance Weighting,” *Mathematical Geology*, vol. 31, no. 1, pp. 114–126, 1999.
- [20] E. Isaaks and M. Srivastava, *Applied geostatistic*. New York: Oxford University Press, 1989.
- [21] Esri. (2020) How inverse distance weighted interpolation works.
- [22] B. I. Harman, H. Koseoglu, and C. O. Yigit, “Performance evaluation of IDW, Kriging and multiquadric interpolation methods in producing noise mapping: A case study at the city of Isparta, Turkey,” *Applied Acoustics*, vol. 112, no. 8, pp. 147–157, 2016.

- [23] Esri. (2020) Classification trees of the interpolation methods offered in geostatistical analyst. [Online]. Available: <https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/classification-trees-of-the-interpolation-methods-offered-in-geostatistical-analyst.htm>
- [24] H. Peng, L. Xiaohang, and H. Hai, “Accuracy Assessment of Digital Elevation Models based on Approximation Theory,” *Photogrammetric Engineering Remote Sensing*, vol. 1, no. 1, pp. 49–56, 2009.
- [25] M. Yilmaz and M. Uysal, “Comparison of data reduction algorithms for LiDAR-derived digital terrain model generalisation,” *Area, Royal Geographical Society*, vol. 48, no. 4, pp. 521–532, 2016.
- [26] G. S. Bhunia, P. K. Shit, and R. Maiti, “Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon (SOC),” *Journal of the Saudi Society of Agricultural Sciences*, vol. 17, no. 2, pp. 114–126, 2018.
- [27] E. Anderson, J. Thompson, and R. Austin, “LIDAR density and linear interpolator effects on elevation estimates,” *INTERNATIONAL JOURNAL OF REMOTE SENSING*, vol. 26, no. 18, pp. 3889–3900, 2005.
- [28] X. Liu, “Progress in Physical Geography: Earth and Environment,” *INTERNATIONAL JOURNAL OF REMOTE SENSING*, vol. 32, no. 1, pp. 31–49, 2008.
- [29] E. Anderson, J. Thompson, D. Crouse, and R. Austin, “Horizontal resolution and data density effects on remotely sensed LIDAR-based DEM,” *INTERNATIONAL JOURNAL OF REMOTE SENSING*, vol. 132, no. 3, pp. 406–415, 2006.
- [30] J. Immelman and L. G. C. Scheepers, “The effects of data reduction on lidar-based digital elevation models,” *2011 4th International Congress on Image and Signal Processing*, vol. 3, no. 1, pp. 1694–1698, 2011.
- [31] D.-A. Kwak, W.-K. Lee, J.-H. Lee, G. S. Biging, and P. Gong, “Detection of individual trees and estimation of tree height using LiDAR data,” *The Japanese Forest Society and Springer*, vol. 12, no. 1, pp. 425–434, 2017.
- [32] S. Gillies *et al.*, “Rasterio: geospatial raster i/o for Python programmers,” Mapbox, 2013–. [Online]. Available: <https://github.com/mapbox/rasterio>
- [33] MuellerSeb, “Pykrige 1.5.1,” 2020. [Online]. Available: <https://github.com/GeoStat-Framework/PyKrig>

- [34] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python,” *Nature Methods*, vol. 17, pp. 261–272, 2020.
- [35] S. Van Der Walt, S. C. Colbert, and G. Varoquaux, “The numpy array: a structure for efficient numerical computation,” *Computing in Science & Engineering*, vol. 13, no. 2, p. 22, 2011.

Appendix A

Project Planning Schedule

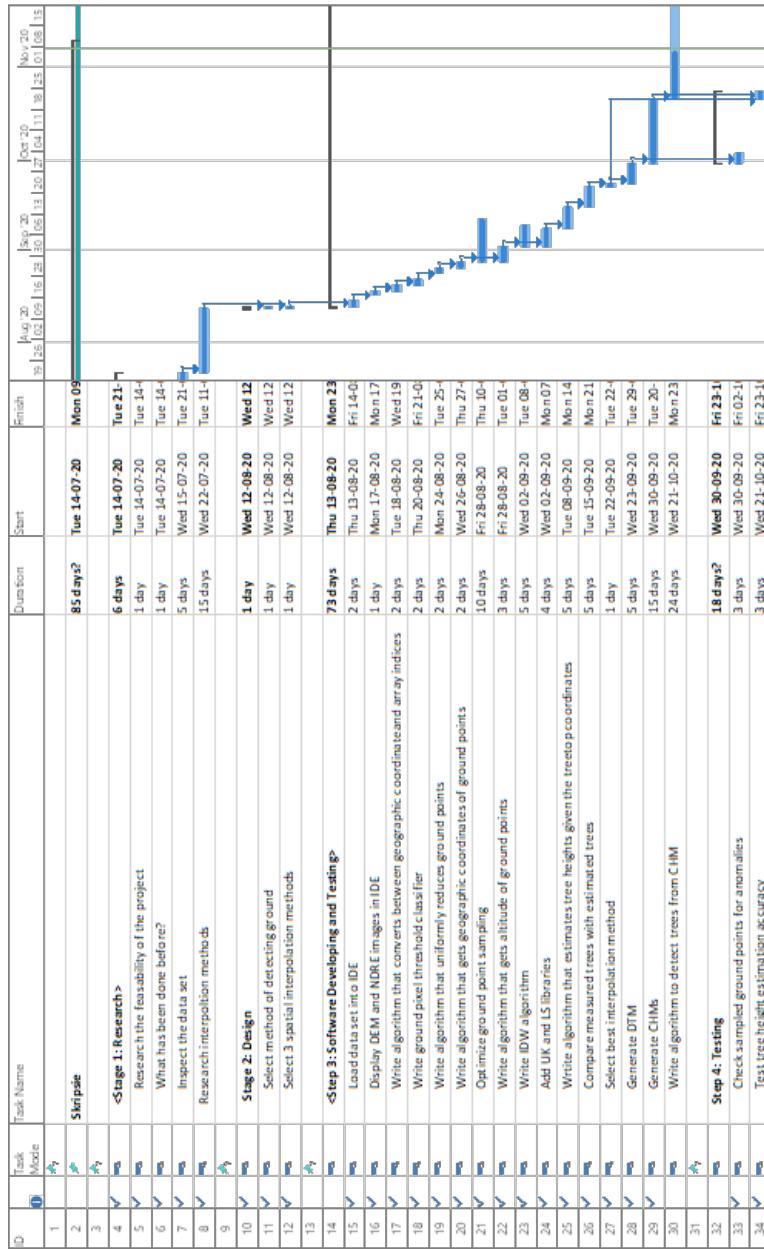


Figure A.1: Gantt chart of project.

Appendix B

Outcomes compliance

B.1. ELO 1. Problem Solving

The problem statement that was provided was further refined into measurable objectives as well as defining the scope in Chapter 1. The primary objectives of this project were achieved by using the steps discussed in Chapter 3. Multiple possible solutions were considered to establish the most appropriate approach. The solution was then optimized in Chapter 4 to enhance the results.

B.2. ELO 2. Application of Scientific and Engineering Knowledge

A scientific and engineering approach was taken throughout the entire project execution. One example of the application of engineering knowledge is in Chapter 4, where the best performing interpolation algorithm could be determined by evaluating different probability density functions produced by these different algorithms. Another would be the use of Applied mathematics to develop the several algorithms for this project.

B.3. ELO 3. Engineering Design

The success of this project relied on well-designed software. The IDW, local minima, ground classifier and individual treetop height estimation algorithms discussed in Chapter 3, were all developed from the base principles and underwent multiple changes that lead to the final design. The software design was approached systematically to verify that at each stage the desired output was achieved.

B.4. ELO 4. Investigations, Experiments and Data Analysis

The provided data set was inspected to understand what it contains and how it can be processed. This process is explained in detail in Chapter 3. QGIS was used to analyse the data set and compare results obtained from the software developed in this project. The results that were used to compare the interpolation methods are shown in Appendix C. Conclusions were made by analysing the data from plots and images similar to those in Appendix F and G.

B.5. ELO 5. Engineering Methods, Skills and Tools, Including Information Technology

Several engineering methods, skills and tools were used during the execution of this project. Applied mathematics was essential in developing and understanding the spatial interpolation techniques. The software of the project was written in Python using Jupyter notebooks that were run in Google Colab. Several popular scientific python libraries such as Numpy [35], SciPy [34] and PyKrig [33] were used in this project. The majority of the information used for developing the software was obtained from the internet.

B.6. ELO 6. Professional and Technical Communication

A meeting was held with professional data analysts and software developers from Aerobatics to gain insight into the topic of the project. This report, as well as a oral presentation were completed to explain how this project was accomplished. Therefore, this outcome was achieved.

B.7. ELO 8. Individual Work

This project was completed through the work of an individual student. Everything accomplished in this project was done independently with very little input from others. For this reason, this outcome is reached.

B.8. ELO 9. Independent Learning Ability

This project required a thorough understanding of several techniques and concepts. Therefore, independent research was conducted in order to complete the project. This report

should be proof that this outcome was achieved.

Appendix C

Comparing Estimated Tree Heights to Manually Calculated Tree Heights

Table C.1: Terrain A tree heights

	Tree height (m)			
	UK	IDW	LS	Calculated
1	2.772	3.075	2.879	2.685
2	3.229	3.392	3.939	3.047
3	3.129	3.224	3.233	3.247
4	3.409	3.591	3.453	3.363
5	3.222	3.416	3.216	3.126
6	2.982	3.038	-1.572	2.873
7	2.795	2.809	2.762	2.595
8	3.167	3.228	3.188	2.977
9	3.336	3.563	3.339	3.264
10	2.835	2.796	2.787	2.697

Table C.2: Terrain B tree heights

	Tree height (m)			
	UK	IDW	LS	Calculated
1	4.033	4.04	4.039	3.713
2	3.485	3.813	8.355	3.177
3	2.932	2.908	2.951	2.835
4	3.841	4.175	3.803	3.688
5	4.09	4.314	-18.256	3.981
6	2.871	2.864	2.862	2.91
7	4.094	4.047	4.098	3.934
8	3.817	3.814	3.821	3.587
9	4.488	4.339	0.732	4.575
10	3.69	3.705	3.702	3.547

Table C.3: Terrain C tree heights

	Tree height (m)			
	UK	IDW	LS	Calculated
1	2.794	2.775	2.415	2.549
2	2.573	2.527	2.594	1.544
3	2.359	2.119	2.084	1.07
4	1.749	1.499	1.39	1.341
5	2.697	2.972	2.253	2.484
6	1.512	1.785	2.322	1.43
7	3.321	3.094	2.594	2.305
8	2.504	2.682	2.55	2.645
9	3.316	3.192	1.955	2.997
10	3.191	2.959	2.12	2.395

Table C.4: Terrain D tree heights

	Tree height (m)			
	UK	IDW	LS	Calculated
1	1.736	1.563	1.789	1.777
2	2.466	2.476	2.463	2.183
3	2.839	2.905	2.882	2.896
4	3.272	3.388	3.301	2.91
5	1.709	1.741	1.708	2.072
6	2.935	2.956	2.916	3.1
7	1.122	1.246	1.235	1.338
8	2.542	2.564	2.574	2.669
9	2.721	2.886	2.65	2.377
10	2.59	2.635	2.58	2.211

Table C.5: Terrain E tree heights

	Tree height (m)			
	UK	IDW	LS	Calculated
1	2.397	2.490	2.415	2.201
2	2.579	2.652	2.594	2.048
3	2.097	2.227	2.084	1.960
4	1.484	1.460	1.390	1.371
5	2.265	2.359	2.253	1.851
6	2.325	2.310	2.322	2.231
7	2.589	2.772	2.594	2.331
8	2.539	2.686	2.550	2.513
9	1.787	2.064	1.955	1.568
10	2.068	2.263	2.120	1.644

Table C.6: Tree heights of Trees on Flat Ground Surface

	Treetop true altitude	UK	IDW	LS	Calculated
1	370.837	2.705	2.786	2.656	2.064
2	370.295	2.493	2.507	2.46	1.949
3	370.686	2.859	2.901	2.834	2.559
4	370.679	2.97	3.002	2.964	2.464
5	370.451	2.856	2.906	2.88	2.42
6	371.504	2.451	2.532	2.47	1.808
7	371.490	2.499	2.545	2.525	2.084
8	371.760	2.842	2.889	2.846	1.878
9	371.762	2.948	2.977	2.919	2.034
10	371.57	2.878	2.92	2.85	2.474

Table C.7: Tree Heights of Trees on Slightly Sloped Ground Surface

	Treetop true altitude	UK	IDW	LS	Calculated
1	443.668	2.652	3.139	2.544	2.195
2	443.306	2.699	2.689	2.613	2.361
3	442.454	1.719	1.85	1.744	2.786
4	443.231	2.011	1.938	2.006	1.65
5	444.12	2.55	2.565	2.522	2.179
6	444.153	2.444	2.592	2.432	2.378
7	443.823	2.096	2.199	2.107	1.895
8	443.625	2.096	1.967	2.085	1.933
9	443.89	2.27	2.267	2.256	1.874
10	444.045	2.747	2.791	2.746	2.248

Table C.8: Tree Heights of Trees on Sloped Ground Surface

	Treetop true altitude	UK	IDW	LS	Calculated
1	467.162	1.788	1.99	2.019	1.176
2	467.0158	2.346	2.802	2.502	1.945
3	466.599	2.624	3.376	2.529	1.933
4	464.791	1.657	2.346	1.615	1.243
5	464.698	1.927	2.288	1.867	1.774
6	463.321	1.972	2.180	2.008	1.987
7	463.194	2.353	2.866	2.372	2.213
8	462.680	2.308	2.925	2.434	1.885
9	461.849	2.201	2.079	2.122	2.008
10	462.056	2.118	2.452	2.2	2.112

Table C.9: Tree Heights of Trees on Bumpy Ground Surface

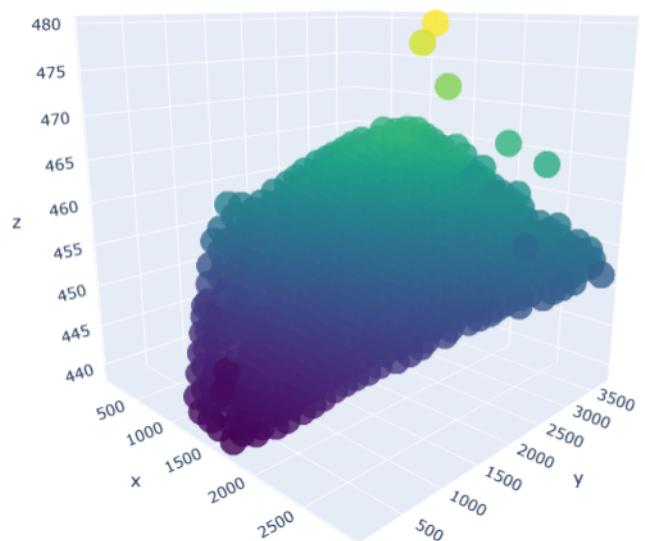
	Treetop true altitude	UK	IDW	LS	Calculated
1	397.001	4.025	4.055	4.018	3.899
2	397.565	3.891	3.863	3.88	3.385
3	396.020	3.738	3.868	3.775	3.739
4	398.521	3.965	4.045	4.011	3.912
5	397.529	4.424	4.421	4.436	4.26
6	397.049	3.667	3.722	3.589	2.08
7	397.639	3.989	3.602	3.864	2.529
8	398.053	3.921	3.755	3.963	3.652
9	398.577	4.122	4.086	4.059	3.594
10	398.663	4.168	4.035	4.137	3.981

Appendix D

Terrains



(a) Visual

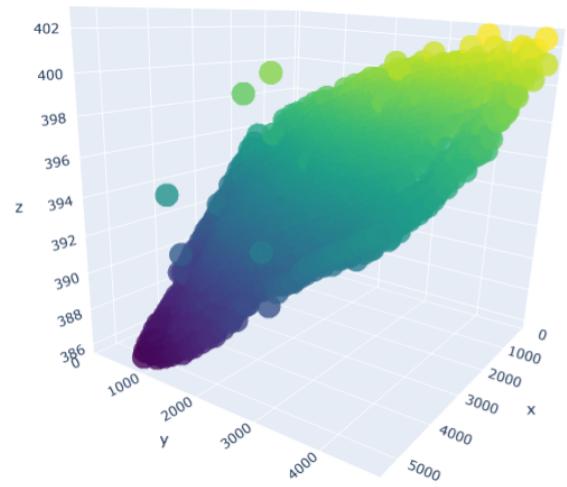


(b) 3-D point cloud

Figure D.1: Terrain A



(a) Visual

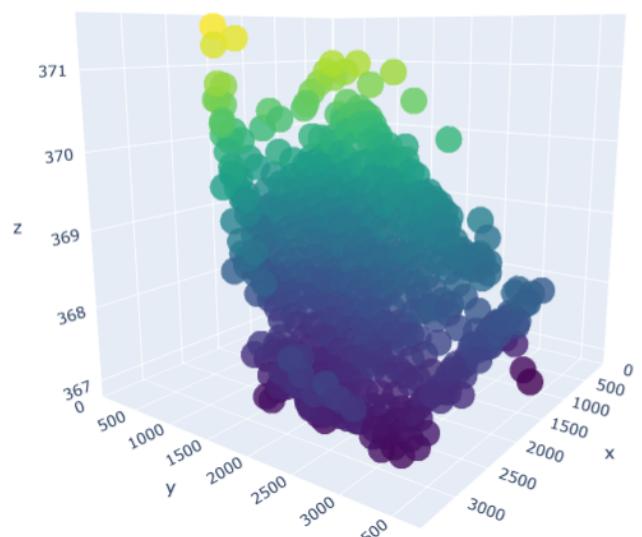


(b) 3-D point cloud

Figure D.2: Terrain B



(a) Visual

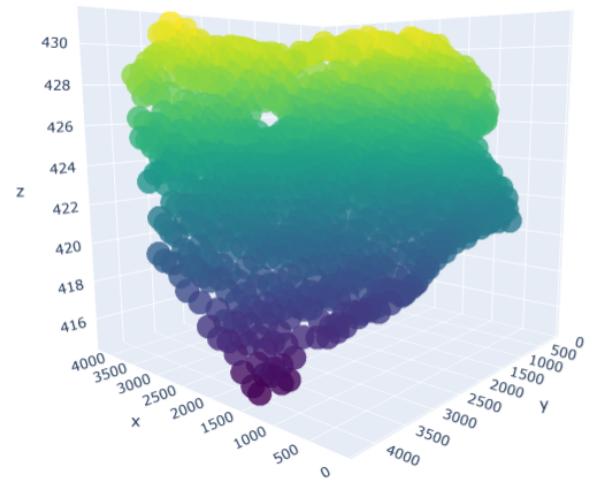


(b) 3-D point cloud

Figure D.3: Terrain C



(a) Visual

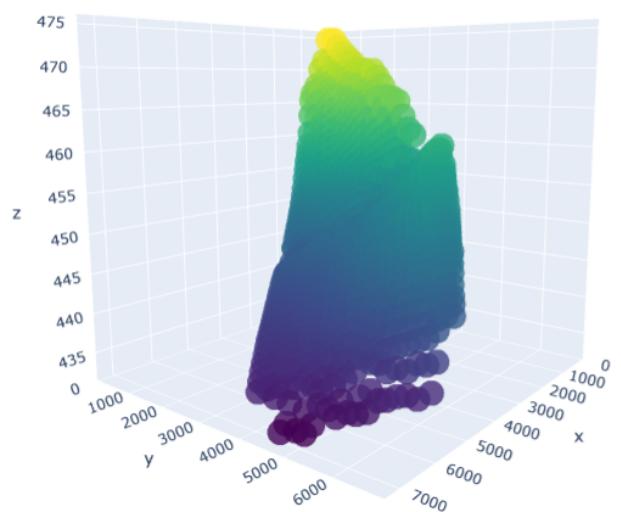


(b) 3-D point cloud

Figure D.4: Terrain D



(a) Visual



(b) 3-D point cloud

Figure D.5: Terrain E

Appendix E

Visualisation of Steps to Produce a CHM

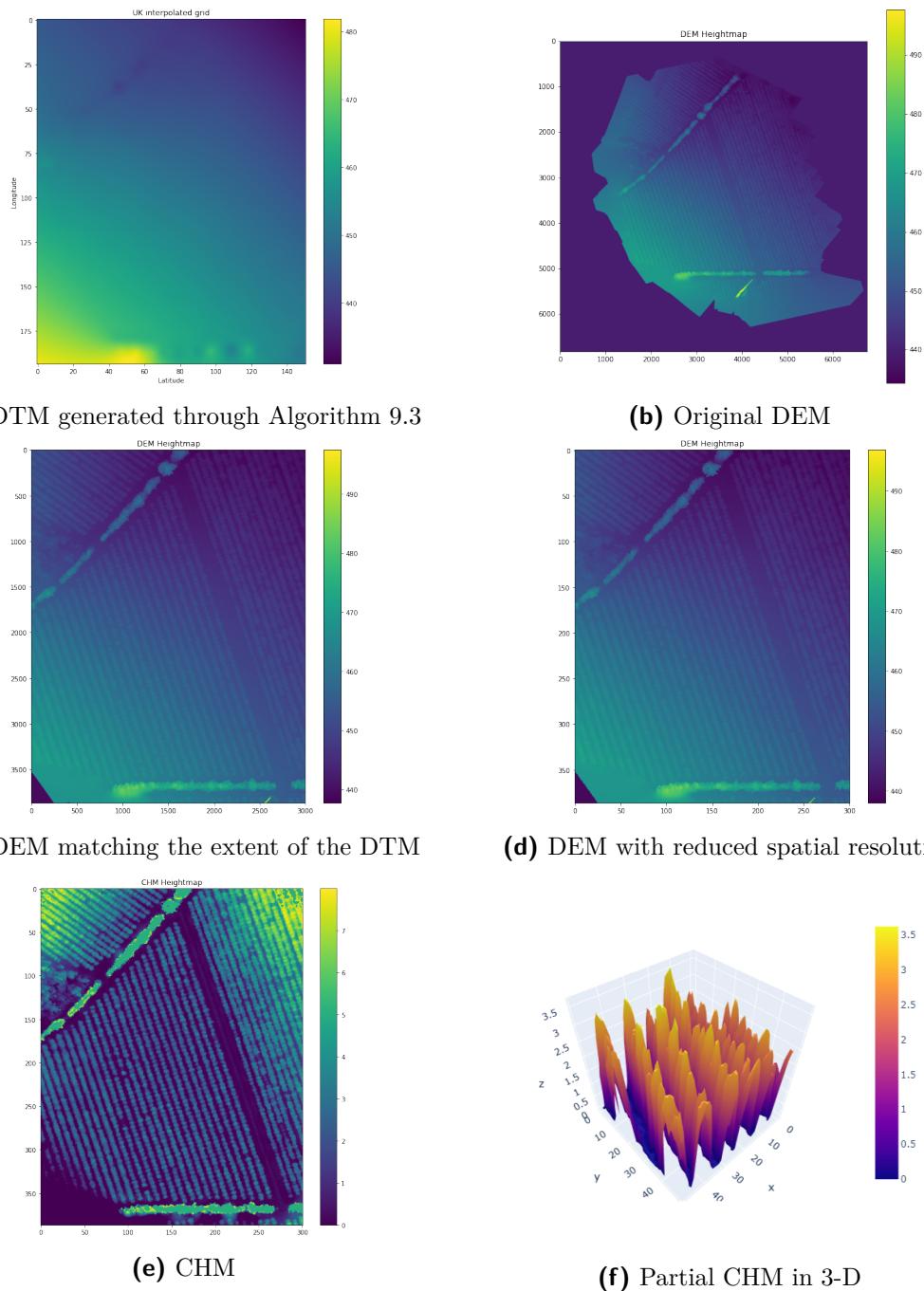


Figure E.1: The process of obtaining CHM from the DTM and DEM of Terrain A.

Appendix F

Contours of Digital Terrain Models for Terrains C, D and E

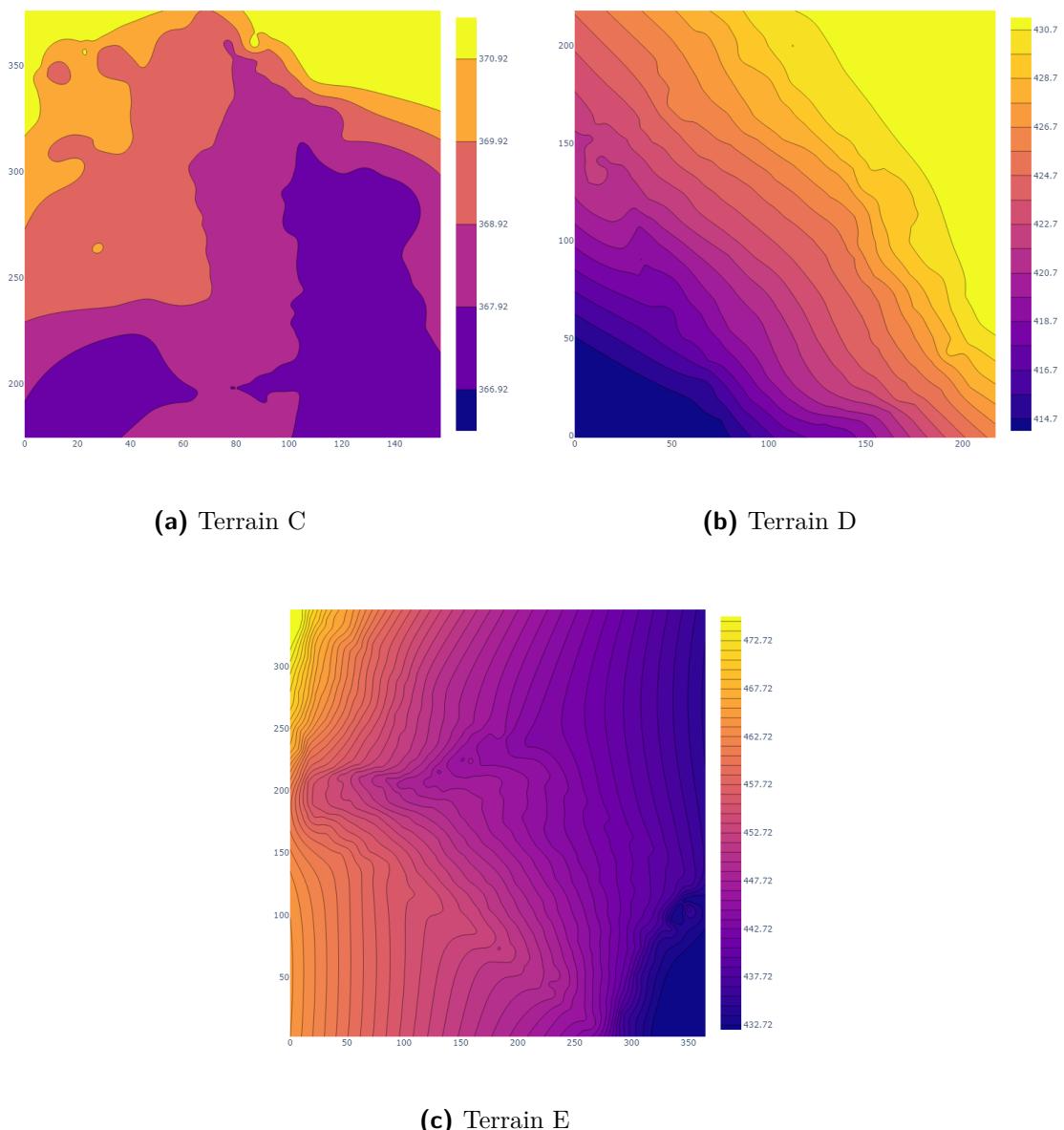


Figure F.1: Contours of DTMs

Appendix G

Comparing Probability Density Functions of Treetop Heights

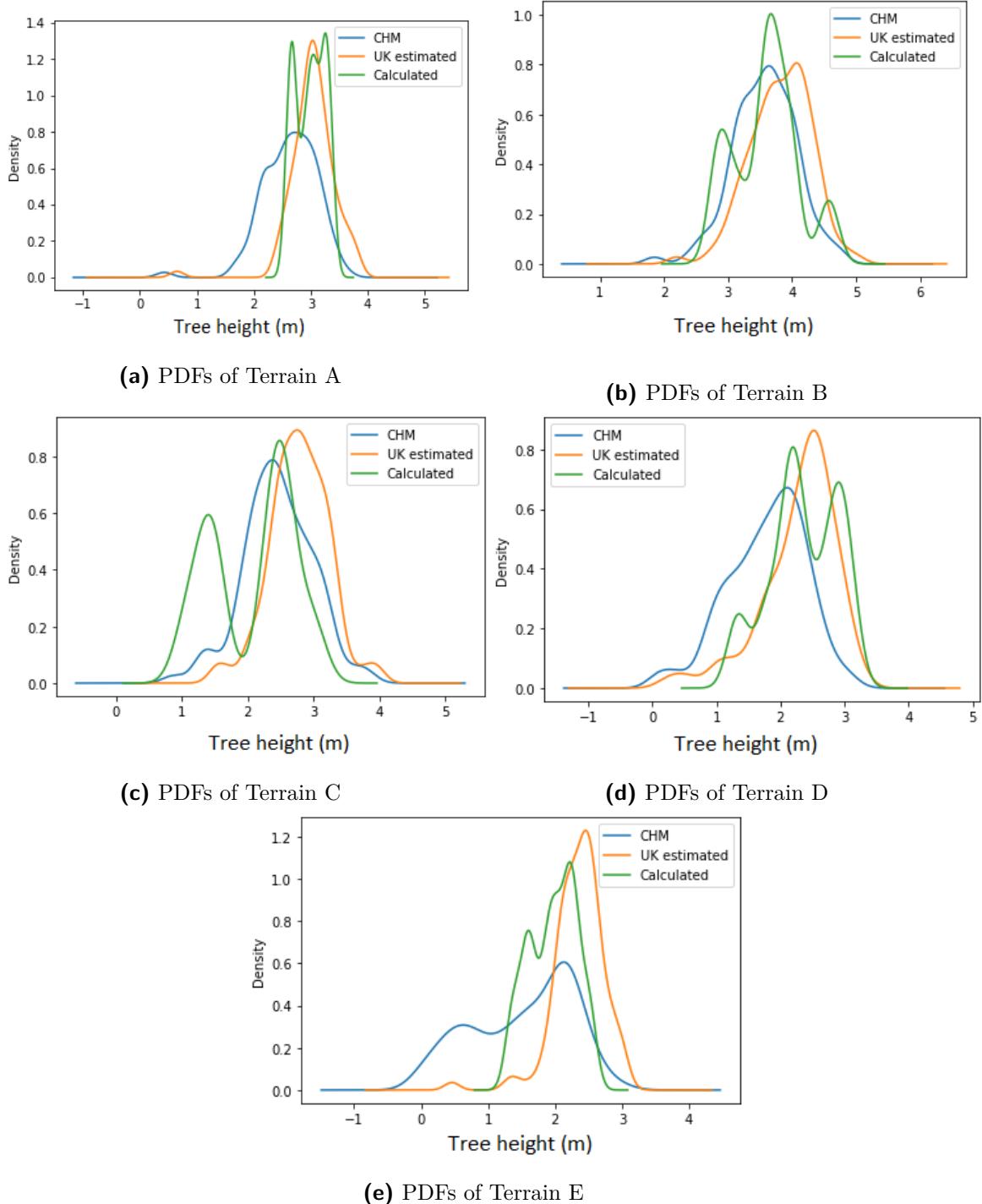


Figure G.1: Green - 10 Calculated Treetop heights from the terrain. Blue - Treetop height obtained from low spatial resolution CHM. Yellow - Treetop heights estimated by predicting ground with UK. (a) Calculated tree heights from Table C.1. (b) Calculated tree heights from Table C.2. (c) Calculated tree heights from Table C.3. (c) Calculated tree heights from Table C.4. (d) Calculated tree heights from Table C.5.

Appendix H

Canopy Height Models of Terrains A,B,D and E

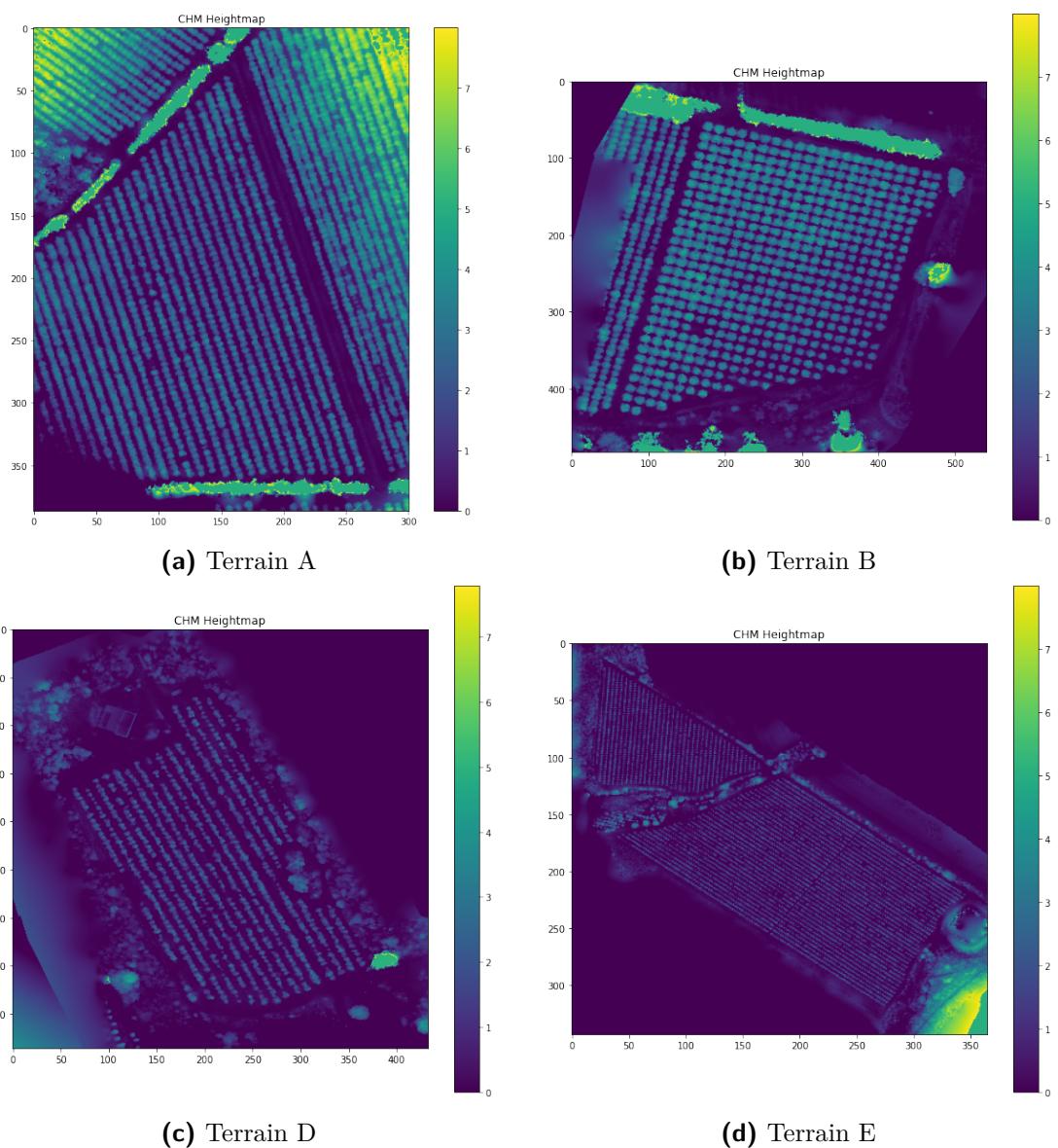


Figure H.1: CHMs Terrain A,B,D and E

Appendix I

Interpolation Algorithms for UK, IDW and LS

Algorithm 9.3: UK

Determine spatial autocorrelation of sampled data points
Fit variogram function model ($\gamma(s_1, s_2)$) to spatial autocorrelation
Determine Kriging equations matrices
for points to interpolate **do**
 Get location at which to interpolate s_0
 for sample points used to interpolate at s_0 **do**
 Get sample point location s_i
 Get elevation of sample point $Z(s_i)$
 Calculate weight of sample point as $\lambda_K(s_i)$
 Calculate $\hat{Z}_K(s_0)$ and $\sigma_K(s_0)$ through UK equation
 end for
end for

Algorithm 9.4: IDW

for points at which to interpolate **do**
 Get geographic coordinates of point being estimated (s_0)
 numerator $\leftarrow 0$
 denominator $\leftarrow 0$
 for sample points within the search area **do**
 Get geographic coordinates of sample point (s_i)
 Get true altitude of sample point ($Z(s_i)$)
 Calculate distance between sample point and point being estimated ($d(s_0, s_i)$)
 Calculate the weight of sample point as $\lambda_{IDW}(s_i)$
 numerator \leftarrow *numerator* + $Z(s_i) \times \lambda_{IDW}(s_i)$
 denominator \leftarrow *denominator* + $\lambda_{IDW}(s_i)$
 end for
 $\hat{Z}_{IDW}(x_0) \leftarrow \frac{\text{i numerator}}{\text{i denominator}}$
 Store estimated true altitude ($\hat{Z}_{IDW}(s_0)$).
end for

Algorithm 9.5: LS

for points at which to interpolate **do**

 Get geographic coordinates of point being estimated (s_0).

 Locate the 4 nearest sample points.

 Use 4 nearest sample points to determine the multivariate polynomial function $f(x, y)$.

 Calculate $\hat{Z}_{LS}N(s_0)$ by using $f(x, y)$

end for
