An early study on the potential of landscape and geographical variables to reduce bias in forest forecast planning in Ireland

– A view on the value of data mining in an industry and era rich in information.

Charles Socrates Judd

Independent project • 30 credits
Swedish University of Agricultural Sciences, SLU
Southern Swedish Forest Research Centre
Alnarp 2022

An early study on the potential of landscape and geographical variables to reduce bias in forest forecast planning in Ireland

– A view on the value of data mining in an industry and era rich in information.

Charles Socrates Judd

Supervisor:	**Per Christer Oden, Swedish University of Agricultural Sciences, Southern Swedish Forest Research Centre**
Assistant supervisor:	Mulualem Tigabu, Swedish University of Agricultural Sciences, Southern Swedish Forest Research Centre
Examiner:	Henrik Böhlenius, Swedish University of Agricultural Sciences, Southern Swedish Forest Research Centre

Credits:	30 credits
Level:	Second cycle, AE2
Course title:	Master’s thesis in Forest Management
Course code:	EX0630
Course coordinating dept:	Southern Swedish Forest Research Centre
Place of publication:	Alnarp
Year of publication:	2022
Cover picture:	Charles S. Judd

Keywords: Landscape, Logistic Regression, Strategic, Tactical, Forecast, Elevation

Swedish University of Agricultural Sciences
Faculty of Forest Sciences
Southern Swedish Forest Research Centre

Publishing and archiving

Approved students' theses at SLU are published electronically. As a student, you have the copyright to your own work and need to approve the electronic publishing. If you check the box for **YES**, the full text (pdf file) and metadata will be visible and searchable online. If you check the box for **NO**, only the metadata and the abstract will be visible and searchable online. Nevertheless, when the document is uploaded it will still be archived as a digital file.

If you are more than one author you all need to agree on a decision. Read about SLU's publishing agreement here: https://www.slu.se/en/subweb/library/publish-and-analyse/register-and-publish/agreement-for-publishing/.

☒ YES, I/we hereby give permission to publish the present thesis in accordance with the SLU agreement regarding the transfer of the right to publish a work.

☐ NO, I/we do not give permission to publish the present work. The work will still be archived and its metadata and abstract will be visible and searchable.

Abstract

At a time when data are an integral part of many industries and the world as a whole, driven by a digital dominant era, there is an emerging discussion surrounding the level of value captured. Using data accumulated in the forest management system and the surrounding databases of the forestry semi state Coillte, my aim was to research if there was any type of data that could potentially be used to identify and understand inaccuracy in long- and short-term forecasting. The Company's Tactical and Strategic forecast volumes for 2018 were used in conjunction with the actual harvest volume from weigh-bridge measurements and roadside stocks in order to understand the current extent of over- and under-estimation. To achieve this, the methods of linear and stepwise backwards logistic regressions were used. The linear regressions based on the percentage of difference of the forecast volumes towards the actual harvested volumes were inconclusive. The logistic regressions were produced using eight binary response variables based on over-and under-estimation. For each forecast type they included, a dataset with all species and product volumes, a dataset with only the dominant species volume (Sitka Spruce/*Picea Sitkensis),* and datasets using the most valuable product (large sawlog) with total volumes and volumes of the primary species only. The predictors consisted of landscape and geographic variables; namely elevation, slope, aspect, country segment, distance from coast, latitude, soil type and roughness. The results showed that over-estimation is the most common form of forecast bias with the tactical forecast models being the most accurate using the predictor variables, Elevation, Roughness and Soil Type. The variables, Aspect, Segment within country and Slope were shown to be the least valuable for prediction. Thus, these aspects should be taken into account when researching forecast bias at the planning level.

Keywords: Landscape, Logistic Regression, Data mining, Elevation, Strategic forecast, Tactical forecast, Forecast bias

Preface

The idea for this thesis shifted several times as difficulties presented themselves. In the early days of designing the method by which the research would take place, it became evident that finding an institution to receive a student overseas can be a difficult task. As time passed and no institution was found, an attempt was made to make the research a meta study developed from existing research since fieldwork was not possible. This was revealed to have its own complexities and difficulties, but change came to provide opportunity. This change came in the form of employment by the Irish semi-state forestry company, Coillte.

My focus hence became to facilitate a research question that would benefit both myself and Coillte as many questions exist that need answering surrounding the operation of any company. Through discussions with the strategic resource lead, and an understanding that forest planning was to be the main focus, a decision was made to research accuracy in forecasts by using the abundance of data the Company already possessed. Initially it would have been to compare several sources of forecasting and estimation including a new technology called SATMODO, which consisted of sensors placed on harvester heads, that calculated the length and size of cuts. Unfortunately, after gaining access to this dataset and progressing the analysis with it, a critical flaw was found 8 months later in the software that made it unusable. This was a major setback, as a central part of the research was tracking the flow of the data through the forecast steps and various estimation steps, like SATMODO and random samplings, in order to identify where losses and shortcomings were most pronounced.

Fieldwork inaccessibility and hardware challenges were followed by the challenges of working a full-time job and trying to reserve the time to research and write a thesis. I felt it should be stated, for all who know and those who do not and wish to undertake this path, that it is a very challenging one that requires great commitment and sacrifice of time outside of every work day and weekend. The final challenge was the pandemic, and of course I refer to Covid-19. The psychological effects of being house bound did not free up time, as one would anticipate, but instead diluted productive time. There was no time to decompress and the will to get up and do things became a heavier burden every week.

Having written this preface and remembering the road that brought me here, I am grateful to have seen it to the end.

Table of contents

1. Introduction 15

1.1. Technology and Forestry 15
1.2. Geography and landscape of Irish forestry 15
1.3. Inspiration behind the variables 16
1.4. An introduction to regressions 17
1.5. What are the tactical and strategic forecasts? 17
1.6. Data and Forestry 18

2. Objectives 19

3. Materials and Methods 20

3.1. Identifying the list of forest sites 20
3.2. Predictor Variables 21
3.2.1. Roughness of the ground's surface condition 23
3.2.2. Soil type of the selected sites 23
3.2.3. Elevation from sea level 24
3.2.4. Aspect from compass direction 24
3.2.5. Slope degree angle of forest sites 25
3.2.6. Buffered distance from coast 25
3.2.7. Segment within the country (Quadrant) 26
3.2.8. Degrees of Latitude 27
3.3. Response Variables 27
3.3.1. The sources of the data 27
Strategic and Tactical Forecast adaptation for thesis 27
Actual volume from harvest operations 27
3.3.2. The response variables made from the source data 28
Linear regression response variables 28
Logistic regression response variables 28
3.4. Data analysis 29
3.4.1. Linear regression 29
3.4.2. Binomial Logistic Regression 29

4. Results 31

4.1. A macroscopic view 31

Forest stand normalcy 31
4.2. Linear Regression 32
4.3. Logistic Regression 33
4.3.1. Evaluation tools for regressions 33
4.3.1.1. The omnibus test 33
4.3.1.2. The Pseudo R^2 33
4.3.1.3. The Hosmer & Lemeshow Test 34
4.3.1.4. Classification Tables 34
4.3.2. Detailed description of regressions 36
4.3.2.1. Strategic estimation of sawlog volume of *Sitka spruce* 36
4.3.2.2. Strategic estimation of all products of *Sitka spruce* 37
4.3.2.3. Strategic estimation of sawlog volume of all species 39
4.3.2.4. Strategic estimation of all species and products 40
4.3.2.5. Tactical estimation of sawlog volume of *Sitka spruce* 41
4.3.2.6. Tactical estimation of all products of *Sitka spruce* 43
4.3.2.7. Tactical estimation of sawlog volume of all species 45
4.3.2.8. Tactical estimation of all species and products 47
4.3.3. Summary of results 49
5. Discussion 52
5.1. Evaluation of the Linear Regression 52
5.2. Evaluation of the Logistic Regression 53
5.2.1. The story of variables Removed 54
6. Conclusion 56

List of Tables

Table 1: Predictor variables, their categories and coding 22
Table 2: Full list of soil types 23
Table 3: The response variables used in the logistic regression analysis 29
Table 4:Strategic forecast Over/Under-estimation ratio 31
Table 5:Tactical forecast Over/Under-estimation ratio 31
Table 6: Example of regression summary for linear regression 33
Table 7:Omnibus compilation of results – Backwards stepwise method 33
Table 8: Pseudo R^2 tests – Backwards stepwise method 34
Table 9: Hosmer & Lemeshow Test – Backwards stepwise method 34
Table 10: Classification table - Backwards stepwise method...............................35
Table 11: Backwards Stepwise Regression – Full model – Response variable: Strategic estimation of sawlog volume of Sitka spruce............................37
Table 12: Backwards Stepwise Regression – Final model – Response variable: Strategic estimation of sawlog volume of Sitka spruce............................37
Table 13: Backwards Stepwise Regression – Full model – Response variable: Strategic estimation of all products of Sitka spruce 38
Table 14: Backwards Stepwise Regression – Final model – Response variable: Strategic estimation of all products of Sitka spruce 38
Table 15: Backwards Stepwise Regression – Full model – Response variable: Strategic estimation of sawlog volume of all species...............................40
Table 16: Backwards Stepwise Regression – Final model – Response variable: Strategic estimation of sawlog volume of all species...............................40
Table 17: Backwards Stepwise Regression – Full model – Response variable: Strategic estimation for all products and species 41
Table 18: Backwards Stepwise Regression – Final model – Response variable: Strategic estimation for all products and species 41
Table 19: Backwards Stepwise Regression – Full model – Response variable: Tactical estimation of sawlog volume of Sitka spruce.............................42
Table 20: Backwards Stepwise Regression – Final model – Response variable: Tactical estimation of sawlog volume of Sitka spruce.............................43
Table 21: Backwards Stepwise Regression – Full model – Response variable: Tactical estimation of all products of Sitka spruce 44

Table 22: Backwards Stepwise Regression – Final model – Response variable: Tactical estimation of all products of Sitka spruce 44
Table 23: Backwards Stepwise Regression – Full model – Response variable: Tactical estimation of sawlog volume of all species 46
Table 24: Backwards Stepwise Regression – Final model – Response variable: Tactical estimation of sawlog volume of all species 46
Table 25: Backwards Stepwise Regression – Full model – Response variable: Tactical estimation of all species and products 48
Table 26: Backwards Stepwise Regression – Final model – Response variable: Tactical estimation of all species and products 49
Table 27: Tables of variables removed per model..50
Table 28: Summary of variables in final models..51

List of figures

Figure 1: Snapshot of query used to export data from the FMS software.............20
Figure 2: To the left, a map showing the sample sites used in this thesis 20
Figure 3: Exported image of site polygons layered onto the digital elevation model (DEM)...21
Figure 4: Aspect Bracket Range 24
Figure 5:DEM slope calculation explanatory image taken from ArcGIS pro documentation website (pro.arcgis.com)..25
Figure 6: To the left, a map showing distance to coastal buffer zones..................26
Figure 7: To the right, a map showing the visualisation of the country segments (Quadrants). 26
Figure 8: Histogram of average tree 32
Figure 9: AD tests for Aspect before and after transformation of response variable to natural log (LN)..33

Abbreviations

AO	Actual Out-turn
BAU	Business Area Unit
CF	Clear Fell
Co	County
CRS	Coordinate Reference System
DfM	Distance from Mean
DEM	Digital Elevation Model
FMS	Forest Management System (Cengea)
GIS	Geographic Information System
HU	Harvest Unit
ID	Identification
JL	Japanese Larch
LPP	Lodge-Pole Pine
NF	Noble Fir
OLS	Ordinary Least Squares
PoD	Percentage of Difference
PSM	Pre-Sale Measurement
RSS	Road Side Stock
SS	Sitka Spruce
STR	Strategic Forecast
TAC	Tactical Forecast
VB	Visual Basic
WD	Weight Docket
WS	Weather Station

1. Introduction

1.1. Technology and Forestry

Forestry has made great technological advances over the last century. With an expanding scale that has been driven and facilitated by the amalgamated use of mechanical and digital assistance, thus greatly increasing the capacity and accuracy of the industry, a demand and supply dynamic formed that will continue to drive the scale of forestry into the future (McEwan et al. 2020). Modern forestry enterprises have embraced this technology widely as evident by the constantly increasing demand for forest software globally (Global Forestry Software Market 2020-2024 2021) and the contractual demand for newer harvesting machinery to achieve the fastest pace of operation technology can provide. With software becoming indispensable to the short and long-term functionality of the forest industry, data collection has become increasingly substantial. This has led to a belief within the industry that a focus should be directed at getting more value out of this sizeable pool of information (Stojanova et al. 2006; Zhang 2014; Istomin et al. 2019; Rossit et al. 2019), usually coined under the term *“data mining”*(Clifton 2009).

1.2. Geography and landscape of Irish forestry

Large-scale forest planting began over the last century in Ireland to re-forest the island after the forest cover was reduced to the historic low of 1.41%. A focus was placed on coniferous timber, especially Sitka Spruce, because it was considered to grow more successfully on inferior classes of land and was believed to be more commercially viable (O’Carroll 2004). Many cases of inferior for forestry land such as peatlands were planted with coniferous trees as better-quality sites were retained for agriculture, which has retained a prominent position in Irish culture (Gray 1964). Therefore, the central commercial species to the Irish forest industry became Sitka Spruce and plantations were often placed in areas where there was availability rather than site specific preference to the species. Consequently, understanding the landscapes still being used today and examining the combination of factors pertaining to them that could influence forecasting bias and error could in turn potentially yield valuable information. Such information could affect the decisions of how we manage mid- to long-term forest planning as well as question where we have placed out forests and why.

1.3. Inspiration behind the variables

Together with the goal of getting more value out of an existing pool of information and a coupling with the well-known, in the field of forestry, topic of landscape and geography, an avenue to potentially improve forecast accuracy presented itself. The central premise of the thesis was to search within the Company databases for variables whose data are regularly collected and linked to the landscape but were not at the time of writing used in any direct capacity in the forecasting process.

The choice of dependant variables was inspired from information gathered during my employment by the semi-state Irish Forestry Company from various discussions with fellow foresters and from personal observation during field-work in the Irish counties of Wicklow in the east, Cork in the south/southwest and Kerry and Limerick in the West.

In the eastern part of the country, the evidence of better-quality stands was thought to be related, at least partially, to the abundance of mineral soils as opposed to the dominance of peatlands in the west which made for poorer site conditions and stability (Fay et al. 2007). The stability and site conditions could be further inferred from the harvester machinery used in forest operations which were obligated to install caterpillar tracks due to the soil conditions in most of the western parts of the country, however, tyres were able to be used in several eastern counties.

The strongest weather patterns in the country are commonly known to arrive from the southwest, originating in the Atlantic Ocean (Met Eireann, as cited by Rohan 1986). This increases the exposure on the western and southern parts of the country that leads to poorer site conditions and the prominent in Irish forestry, issue of windblow (Dhubháin 1998). I observed cases of sites with serious deformities from breakage, windblown sections and in some cases overall stunted growth when monitoring several forest stands in the Dingle, Iveragh and Beara peninsulas of west Cork and Kerry.

A picture was starting to form surrounding landscape and geography that potentially characterises forestry in Ireland. The next step was to review the Company databases for these landscape and geographical variables that could potentially be used to evaluate the area effects on forecast accuracy. The selection was limited to variables that met robust methods of measurement with adequate definition and specificity to meet the criteria of academic research.

1.4. An introduction to regressions

Dobson (2013) defines a model as signal and noise, where the signal is a mathematical description of the main features of the data and noise is the characteristics not explained by the model. The goal is to derive the maximum amount of information from the signal in the form of variability that is not attributed to noise (Dobson 2013).

The selection of candidate variables for a regression is determined by the balancing of experience and theoretical study with the information acquired from data analysis. The goal thereafter is to select the variables for the model based on an equilibrium of complexity and necessity by leveraging the need towards the bias of added regressors (Hintze 2007). There are several methods at the disposal of statistical software for the machine selection of variables.

Stepwise selection is a good technique to observe the effects predictor variables have on the response variables. The method of Backward stepwise selection is preferable in many situations, in comparison to the Forward stepwise selection, because it does not produce suppression effects where a predictor variable may be uncorrelated with the response variable but related to other predictors (Thompson & Levine 1997). The backwards stepwise selection method is one of many good methods of observing the value and interactions of variables, as is the scope of this study.

The criteria for a good model can rely on many factors depending on the nature of the research. Generally, a good model is one that explains a large amount of variability (Dobson 2013). However, there is actually no agreed upon criteria for a "good" model, thus lending interpretation in a way to personal preference. A good way of generalizing the pillars of a good model according to Moody and Shanks (1994) are the selection of desirable properties (qualities), a way to measure them (metrics), a description of their relative importance (weightings) and the ways in which a model can be improved (strategies) (Moody & Shanks 1994).

1.5. What are the tactical and strategic forecasts?

The tactical and strategic forecasts are the mid- and long-term forecasting methods used by many private and public forestry entities in order to get information on estimated future resource volumes.

The strategic forecast used by Coillte foresees two standard rotations into the future, for Sitka Spruce that is 80 years approximately. It is customarily based on variables available at this stage of planning like yield class, planting density, seedling type and others. Given the variable types used and the fact that it is not possible to have more accurate variables at this stage such as site, climate, growth and tree characteristics, only a broad accuracy is expected. The actual volume recoverable from the stand at the time of harvest may be quite different from the strategically forecasted one. This is because it is unlikely for the trees, the site, the weather patterns and the silviculture applied throughout the life of the forest to be perfectly predicted. The world is unpredictable and the further ahead you set a forecast the least predictable it will be as a general rule.

The Tactical forecast used by Coillte is mid-term and is usually carried out 3-5 years prior to the expected harvest of a forest. At this stage, additional variables become available such as stand characteristics measured from sampling sites (e.g., top height, diameter at breast height, etc...) or remote sensing. The effects from the climate and weather patterns are now visible for their influence on the overall health of the stand (e.g., crown breakage, windfall, etc…). The additional variables provide a more accurate and current view of the forest's health, quality and growth.

1.6. Data and Forestry

The integrated use of data permeates every aspect of the forestry process from forecasting and resource planning to the management of harvested stock and the supply chains that handle it thereafter. Focusing on the first two, forecasting and resource planning, both provide the information that drive major key decisions for the industry, and data mining can be used to discover and extract patterns to improve their accuracy. As with most forestry operations, if not all, a large amount of data is created and stored. This includes the data surrounding long-term strategic forecasts, mid-term tactical forecasts and actual measured volumes that are used to evaluate and appreciate potential stock or keep track of its commercial performance. Many key decisions hinge on the information of these forecasts, from operational to financial and even further to a national level, when assessing the Country's future timber stocks from bodies like the Council for Forest Research and Development (COFORD Wood Mobilisation and Production forecasting Group 2018). Therefore, there is a strong incentive to research ways of improving the accuracy of these forecasts using the resources at hand. The collection of data used in this research was sourced from the Irish semi-state forestry company, Coillte, a company established in 1989 from the Irish Forest Service and became the custodian of state-owned forests in Ireland (*www.coillte.ie (Coillte Website)* n.d.).

2. Objectives

The main objective of the study in this thesis work was to understand the variables that contribute to a greater or lower level of estimation bias using data already available in Coillte’s forest management system and surrounding digital infrastructure. For this purpose, the effects of landscape features and geography on the bias in estimation of harvest volume were examined with the use of regression models.

The study is expected to provide insight into the landscape factors that are the most and least indicative of estimation bias during the two planning stages. Considering that this is one of the first studies of its kind for Coillte, it will examine the results for valuable directions to advise future research aimed at identifying difficult and less accurately predictable landscapes.

3. Materials and Methods

3.1. Identifying the list of forest sites

The sample data consisted of a number of forests exported from a Forest Management System (FMS) using a custom query within its internal database. The sites were all chosen using a specific list of parameters, which included (Figure 1):

- Sites from the complete 2018 harvest schedule
- Where felling activities were successfully carried out
- Consisting of only clear-felled sites
- And the primary species was Sitka Spruce.

Figure 1: Snapshot of query used to export data from the FMS software

	Show	Table	Column	Operator	Criteria
	✓	Harvest Unit	Harvest Unit Number [D		
And	✓	Harvest Unit	Activity Status [Detail]	Not In	('X')
And	✓	Harvest Unit	Activity Type [Detail]	=	CF
And	✓	Harvest Unit	Plan Year [Detail]	=	2018
And	✓	Harvest Unit	Standing Or Harvested	=	Harvested
And	✓	Harvest Unit	Slope [Site Conditions]		
And	✓	Harvest Unit	Roughness [Site Conditi		
And	✓	Harvest Unit	Total Volume [Detail]		
And	✓	Harvest Unit	Net Volume [Detail]		
And	✓	Harvest Unit	Stake Volume [Detail]		
And	✓	Harvest Unit	Average Tree [Detail]		
And	✓	Harvest Unit	Total Volume [Detail]		

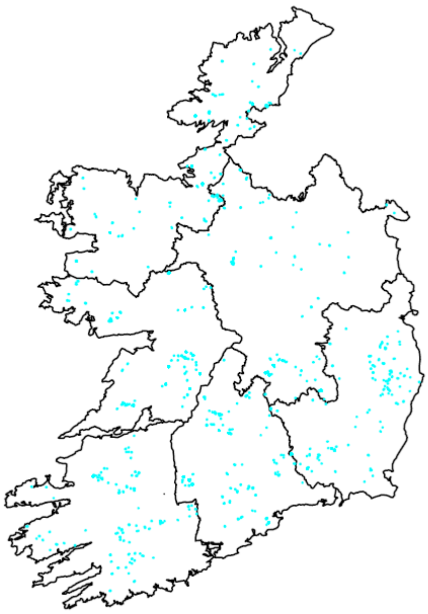


Figure 2: To the left, a map showing the sample sites used in this thesis

The query's output gave a list of 475 sites (Figure 2) that became the list of sample sites for this Thesis. Following the export of the tabular information, a multi-part polygon was created for the 475 sites using the spatial tool in FMS and was then exported to ArcMap for processing.

3.2. Predictor Variables

For this thesis work, eight predictor variables were considered; namely roughness, elevation, aspect, slope, distance to the coast, segment within the country, soil type and latitude (Table 1).

Figure 3: Exported image of site polygons layered onto the digital elevation model (DEM)

For the ground conditions, data were retrieved regarding the soil type for each site and the corresponding ground roughness. The forest management system included a vast array of spatial data in the form of forest site spatial polygon objects and various types of underlying maps. This offered potential for the development of variables from the available spatial information. From use in unrelated projects within the Company, the land observation team was in possession of digital elevation models that had been acquired from the Copernicus project database (*EU-DEM v1.1 — Copernicus Land Monitoring Service* n.d.). With some development using the ArcGIS and QGIS software this provided opportunity for the development of additional variables (Figure 3).

The variables of aspect, elevation and slope have often been linked in literature with forest growth and productivity (Stage 1976; Stage & Salas 2007). Consequently, utilising GIS, these variables were developed and exported for each of the 2018 harvest sites. Furthermore, there was a particular interest to examine the potential for effects moving throughout the country according to the aforementioned noted oceanic influences. Therefore, the variable latitude was formed with GIS to examine if there is an affect moving linearly up the length of the nation. Distances from the coast were developed in three 10 km buffer increments to evaluate the coastal affects. Finally, a split of the sites in the country by quadrat was made to evaluate the broad scale merit of the Eastern to Western or Northern to Southern effects that had been observed by fellow foresters and my own personal observation from the southwest of the country.

Table 1: Predictor variables, their categories and coding

Predictor	***Category***	***Coding***
Roughness	*unknown*	*0*
	Even	*1*
	Uneven	*2*
	Rough	*3*
Elevation	*10- 60 m*	*1*
	61-120 m	*2*
	121-180 m	*3*
	181-240	*4*
	241-300 m	*5*
	301-360	*6*
	361-505	*7*
Aspect	*N-NE*	*1*
	E-SE	*2*
	S-SW	*3*
	W-NW	*4*
	N-NE	*5*
Slope	*Gentle*	*0*
	Intermediate	*1*
	Steep	*2*
	Very steep	*3*
Distance from the coast	*0-10 km*	*0*
	10-20 km	*1*
	20-30 km	*2*
	> 30 km	*3*
Segment within the country	*NE*	*0*
	NW	*1*
	SE	*2*
	SW	*3*
Soil types	*Deep Acidic Mineral Non-Calcareous*	*1*
	Shallow Acidic Mineral Non-Calcareous	*2*
	Blanket Peat	*3*
	Alkaline Mineral Calcareous	*4*
Latitude	*535995-615996 degree*	*1*
	615997- 659997 degree	*2*
	659998-695996	*3*
	695997-775996	*4*
	775997-935996	*5*

3.2.1. Roughness of the ground's surface condition

Roughness was determined by exporting the data from the Forest Management System that the Company uses, with a custom query for the parameters. It is defined as a categorical variable taking the values of Even, Uneven, Rough and Unknown. These values describe the evenness of the ground, size and frequency of obstacles.

- **Even**: No Obstacles – ground with some small stones, field ditches present and machinery can move unhindered.
- **Uneven**: Obstacles quite frequent – furrows, high stumps, small stones frequent with occasional large stones (> 40 cm), ditches and drains present making travel hindered.
- **Rough**: Obstacles frequent – wide drains, deep ploughed furrows, large stones, small stones, rocks (> 60 cm) and banks.
- **Unknown**: Cases where data is unavailable.

Roughness was coded into 4 numerical categories as shown in (Table 1).

3.2.2. Soil type of the selected sites

Soil type was identified by exporting the data from the Forest Management System that the Company uses with a custom query for the thesis site parameters in the database. The data were then sorted by site code in alphabetical order and by intersected area descending. A check was then made to see if the soil type with the largest overall coverage also had the majority percentage from the other soil types that may be found in the same site. This was concluded to be the case and therefore the soil type with the largest coverage was attributed as the soil type for each site. The sites in total produced a total of 16 soil types as seen below (Table 2):

Table 2: Full list of soil types

Soil Type code	***Description***
AlluvMIN	Mineral Alluvium
AminDW	Acidic - Deep - Well Drained Mineral - non calcareous
AminPD	Acidic - Deep - Poorly Drained Mineral - non calcareous
AminPDPT	Acidic - Deep - Poorly Drained Mineral - non calcareous - Peaty Topsoil
AminSP	Acidic - Shallow - Poorly Drained Mineral - non calcareous
AminSPPT	Acidic - Shallow - Poorly Drained Mineral - non calcareous - Peaty Topsoil
AminSRPT	Acidic - Shallow - Lithosolic or Podzolic - non calcareous - Peaty Topsoil
AminSW	Acidic - Shallow - Well Drained Mineral - non calcareous
BktPt	Blanket Peat
BminDW	Alkaline - Deep - Well Drained Mineral
BminPD	Alkaline - Deep - Poorly Drained Mineral - Calcareous
BminPDPT	Alkaline - Poorly Drained Mineral - Calcareous - Peaty Topsoils
BminSRPT	Alkaline - Shallow - Lithosolic/Podzolic - Calcareous - Peaty Topsoils
BminSW	Alkaline - Shallow - Well Drained Mineral - Calcareous
Cut	Cutaway Peat
Scree	Scree

The soil type, when reviewed as a pivot table with a column revealing the number of sites that comprised each type, yielded the conclusion that some soil types are represented by a very small number of sites. For this reason and to reduce the number of categories from the initial 15, a grouping was applied to make 4 categories as shown in Table 1.

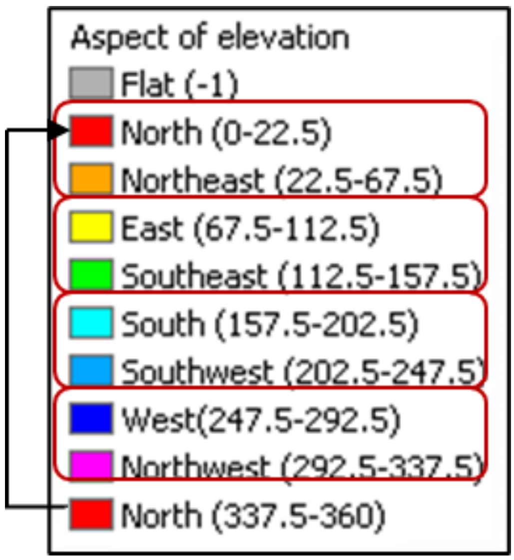
3.2.3. Elevation from sea level

The multi-polygon shape file of the 475 sites was used for this variable and the DEM (Digital Elevation Model) from the Copernicus project. Ireland as a whole does not exist within any single rectangle in the DEM so 2 separate ones were merged using the “Mosaic to new raster” tool (E20N30 & E30N30). Furthermore, the Clip tool was employed to isolate the raster area of the polygons of the sites. The raster was converted from Float point to Signed Integer with the “Copy Raster” tool and then converted to polygon feature using the “Raster to Polygon” tool. This output now carrying an attribute table with the segmented elevation data was intersected with the site polygons. This new feature’s attribute table data were exported to excel and the average elevation was determined for each site with the development of Visual Basic Script (VB) to create a function that concatenates pixel group measurements and then calculates the averages.

This variable was continuous at this stage and was further coded into categories as shown in Table 1.

3.2.4. Aspect from compass direction

Figure 4: Aspect Bracket Range

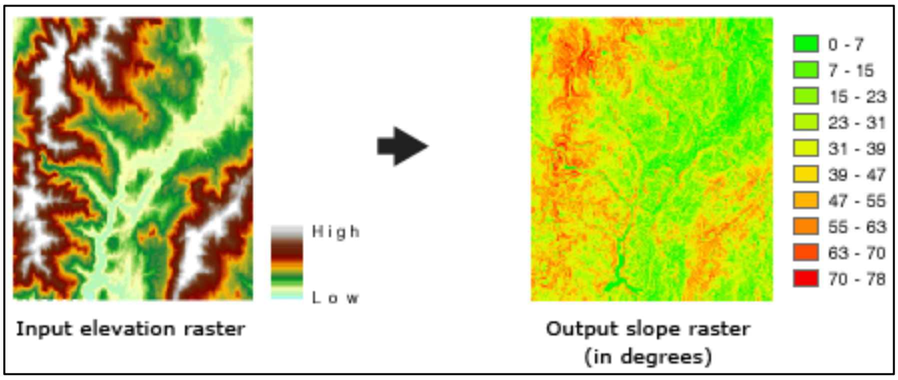


The spatial analyst tool was used on the clipped site raster files from the DEM, as developed in the previous variable, to calculate the aspect using the aspect tool which assigns compass direction. The raster output was converted from Float point to Signed Integer with the “**Copy Raster**” tool and then converted to polygon feature using the “**Raster to Polygon**” tool. This output now carrying an attribute table with the segmented Aspect data were **intersected** with the site polygons. This new feature’s attribute table data was exported to excel and the average aspect was determined for each site using VB Script to concatenate pixel group measurements and the average function. Furthermore, a bracket range was used to define parameters for direction (Figure 4) and functions were used to identify which degree direction each site was in. This variable is characterized as categorical as shown in Table 1.

3.2.5. Slope degree angle of forest sites

To calculate slope, the spatial analyst tool (Figure 5) was used on the clipped site raster files. The raster output was converted from Float point to Signed Integer with the "Copy Raster" tools and then converted to polygon feature using the "Raster to Polygon" tool. This output now carrying an attribute table with the segmented slope data was intersected with the site polygons. This new feature's attribute table data was exported to excel and the average slope was determined for each site using VB Script to concatenate pixel group measurements and the average function. This variable is characterized as continuous, but was later further coded into categories (Table 1).

Figure 5:DEM slope calculation explanatory image taken from ArcGIS pro documentation website (pro.arcgis.com)



3.2.6. Buffered distance from coast

Coastal buffers were made from 0-10km, 10-20km, 20-30km using a buffer tool on a polygon of the republic of Ireland and then intersected with the site layer (Figure 6). This variable is categorical and can take four distinct values (Table 1).

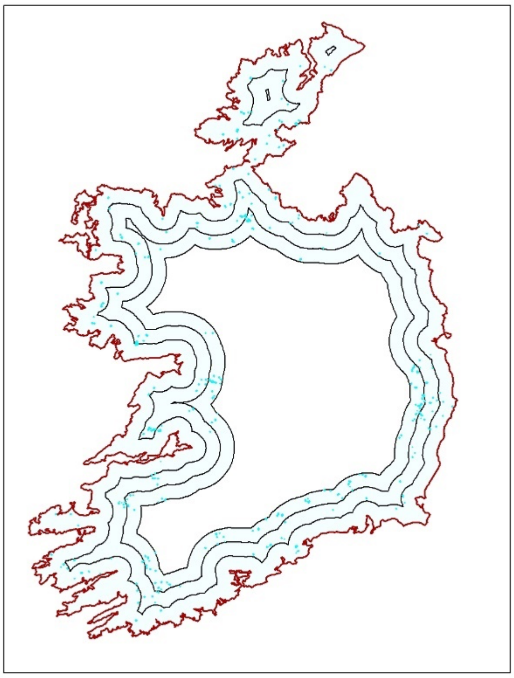


Figure 6: To the left, a map showing distance to coastal buffer zones

There was a consideration explored to treat this variable as ordinal and consequently whether it would violate the assumption of equal intervals leading to type 1 & 2 errors. Since the last category is not of equal interval to the previous three which are equal between each other, the assumption of equal intervals was indeed violated and this variable was be treated as categorical.

3.2.7. Segment within the country (Quadrant)

The centre point for the island of Ireland was calculated from its polygon feature and a cross section from the centre-point, with the vertical line pointing north to south, that separated the country into four compartments (Figure 7). These compartments were intersected with the site layer to get data on which segment of the country the sites fall into. The segments are categorical and can take values from 0 to 3 (Table 1).

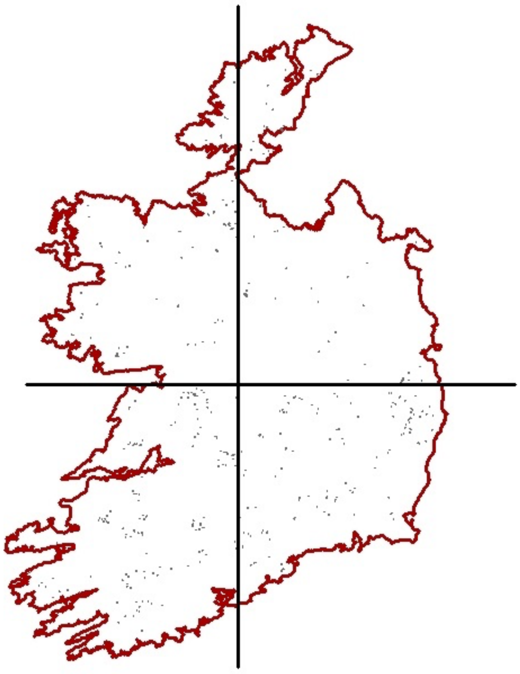


Figure 7: To the right, a map showing the visualisation of the country segments (Quadrants).

3.2.8. Degrees of Latitude

The latitude was calculated for every site by producing the Y Centroid in ArcMAP with the ITM IRENET95 coordinate system. This was added for each site to the polygon feature attribute table and exported to excel. The values ranged from approximately 53.6 degrees for the most southern site to approximately 93.6 degrees for the most northern. The variable, which was initially continuous, was coded into 5 categories of roughly equal sample sites (Table 1).

3.3. Response Variables

3.3.1. The sources of the data

The response variables were sourced from volumetric data taken from the Remsoft tactical optimiser software. This data was provided by the manager of Coillte's central resource team of which I was also a member at the time of collection. The data concerned the harvest year of 2018 and included data from the strategic forecast volumes and tactical forecast volumes. The actual harvest volumes were exported from Coillte's Forest Management System (FMS) and were comprised by the amalgamated volumes from digital weight dockets and roadside stock reports.

The data were provided under agreement that all Company code designations and actual volume values would be used with discretion and not published. Therefore, the data required transformation to protect the privacy policy that was agreed.

Strategic and Tactical Forecast adaptation for thesis

The Strategic & Tactical forecast volume data were assigned to each site using a shapefile containing polygon features of the forest sites that were then intersected with the polygon export from the Remsoft software that was tasked with producing the forecast models. This was necessary because forecasts and harvest sites are two different spatial units. This thesis is using the spatial perspective of harvested forest sites, therefore comparatively, several forecast blocks from Remsoft were amalgamated into each harvest site as a general rule.

Actual volume from harvest operations

To calculate the actual harvest volume, two elements were required, both exported from FMS. The first element, the digital weight-docket (WD) data, which were derived from datasets formed from the weight-dockets which are digital records of lorry loads from each forest. The dockets get their values after the lorries pass over a weigh-bridge and the measured weight is then converted to volume with a volume/weight factor that is calibrated by actual random sampling of lorry loads daily. The second item was the roadside stock (RSS) which is the remaining timber volume stacked roadside at each of the forest sites, for which the recorded volumes are stored in databases on the forest management system. Similar to previous variables the results were aggregated using Visual Basic script in excel to make a unique list of forest sites with the summarized volumes of WD's and RSS comprising the actual volumetric retrieval of timber after harvesting.

3.3.2. The response variables made from the source data

Linear regression response variables

For this type of regression, in order to use the data in any comparable capacity, the values needed to represent a linear form. Considering that all sites had various sizes, a simple subtraction of forecasted volumes from actuals would be inadequate. Therefore, the subtracted value was divided by the actual volume to become a representative percentage of difference (PoD) from the total of 100% volume of each site.

$$PoD = \frac{Forcast\ volume - Harvested\ volume}{Harvested\ volume}$$

If the actual volume was higher than the forecasted volume then it was under-estimation and was given a negative value, conversely if it was the opposite, it was over-estimation and was left as a positive value. The PoD was calculated for each site for both the strategic and tactical forecast against the actual harvested volumes.

Logistic regression response variables

For this type of regression, the response variable must have a dichotomous form. The response variables for the logistic regression applied in this thesis work were overestimation and underestimation of harvest volume, which were defined as binary variables with a value 1 for overestimation and 0 for underestimation.

Using the dichotomous principal, 8 variations of the response variable were created (Table 3). These where 4 variations for each forecast type (strategic & tactical) that were separated according to the species and product. The product in question was large sawlog as it is the most valuable log cut and of particular importance in Irish forecasts. Therefore the 4 response variables for each forecast type, included the overestimation or underestimation per forest site for:

- the full set for all products and species,
- a set for large sawlog across all species
- a set for all products with only volumes of the primary species of *Sitka Spruce* (SS)
- a set for large sawlog with only volumes of the primary species of *Sitka Spruce* (SS)

Table 3: The response variables used in the logistic regression analysis

Codes	Description
STR.Lsawlog.SS	Strategic estimation of sawlog volume of *Sitka spruce*
STR.SS	Strategic estimation of all products of *Sitka spruce*
STR.Lsawlog.Total	Strategic estimation of sawlog volume of all species
STR.Total	Strategic estimation of all species and products
TAC.Lsawlog.SS	Tactical estimation of sawlog volume of *Sitka spruce*
TAC.SS	Tactical estimation of all products of *Sitka spruce*
TAC.Lsawlog.Total	Tactical estimation of sawlog volume of all species
TAC.Total	Tactical estimation of all species and products

3.4. Data analysis

3.4.1. Linear regression

Linear regressions were utilised to examine the variability of the response variables explained by the predictor variables. The predictor variables that had originally a continuous form were used as such and dummy variables were created for the strictly categorical. The regressions were carried out using the RegressIT addon in excel, developed by Robert Nau, Professor Emeritus in the Fuqua School of Business at Duke University. To understand the interactions of each predictor variable with the response variable and given the manageable number of predictor variables, one-on-one regressions were selected to identify the statistical relationships and challenges. Multivariate linear regressions were then used to evaluate if a suitable model from a combination of variables could be produced.

3.4.2. Binomial Logistic Regression

Using binomial logistic regressions, eight stepwise binary logistic regression models were developed through the backward elimination procedure to assess the significance of landscape features in explaining the bias of volume estimation.

The logistic regression model is an appropriate statistical tool to determine the influence of predictor variables on response variables, when the latter have dichotomous outcomes and the former are continuous, categorical or dummy variables. Essentially, the logistic model predicts the logit of the response variable (Y) from the predictor variables (X). The logit is the natural logarithm (ln) of odds of Y, and odds are ratios of probabilities (π) of Y happening to probabilities ($1-\pi$) of Y not happening. The logistic model is specified as:

$$ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki}$$

where β_0 is the intercept and β_1, β_2 ... β_k are the coefficients of the predictor variables x_1, x_2 … x_k. Using this dichotomous principle, the datasets were regressed using the SPSS Statistical Package (SPSS 19.0, Chicago, IL, USA).

The tools used to access the model were the following:

- The omnibus test: a likelihood-ratio chi-squared test of the current model against the null model.
- The Pseudo R Squared Tests, though not directly indicative of the variance explained like in a linear regression that utilizes ordinary least squares (Montgomery et al. 2021), can be used as a measure of comparison between the models (Portl 2021). Nagelkerke's Pseudo R^2 was used, as a modification on Cox & Snell's pseudo R^2 to permit values up to 1.
- The Hosmer and Lemeshaw test was used to assess the goodness-of-fit of the models where a significant p-value would indicate a poor fit.
- The classification table results examine the practical results of using each model.

Furthermore, after quantifying the usability of the results and their relative importance using the above tools, the regression results were examined for the individual variable characteristics and the interactions between variables from the full model, through the stages of removal, and in the final model.

4. Results

4.1. A macroscopic view

A macroscopic glance reveals that the Strategic and Tactical models tend to overestimate in roughly 65% of the cases with the contrast being that the average Percentage of Difference is moderately lower by approximately 5% between the forecasts (Tables 4 & 5). The Tactical forecast is derived from additional variables that are measured with sample plots (E.g., DBH, stand density, Top Height, etc…) and therefore more accuracy would typically be expected than from the Strategic forecast which is based largely on yield class projections and initial spacing.

Table 4:Strategic forecast Over/Under-estimation ratio

Strategic Forecast	***No. of HU***	***Average PoD***	***% of Total***
Under-estimation	155	-20%	33%
Over-estimation	307	34%	65%

Table 5:Tactical forecast Over/Under-estimation ratio

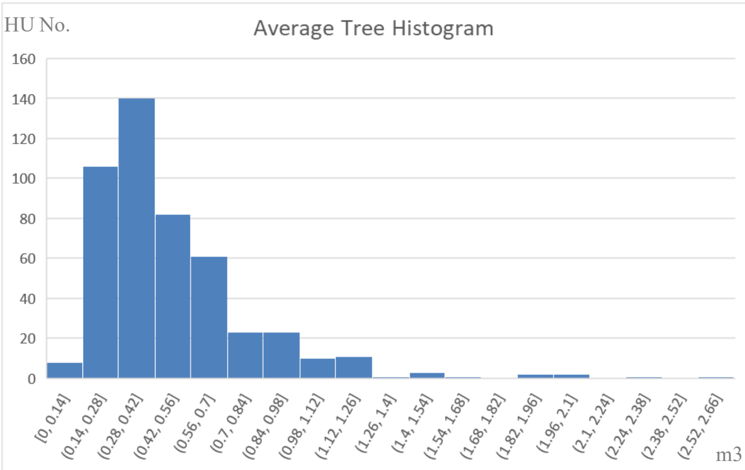
Tactical Forecast	***No. of HU***	***Average PoD***	***% of Total***
Under-estimation	154	-22%	32%
Over-estimation	305	29%	64%

Forest stand normalcy

Using data on the average tree volume, insight on the normalcy of the stands can be observed (Figure 8). According to the rotation period for Sitka Spruce used by the Company and accounting for the occasions of early and late harvests, the expected values for most stands should fall approximately between 0.25 and 0.70 m^3. This approximation is based on average stands YC12+ with a rotation cycle of 30+ years (Matthews et al. 2016). Above or below this point, the stand can be assumed to have some atypical size characteristic to various extents. Examples of such characteristics could be, an extended rotation due to poorer growth than expected, challenges with access resulting in delaying harvesting until more favourable access can be secured, storm damage, windblow, poor drainage or damages from pests and decease. As Figures 1 and 2 demonstrate, the majority of the stands (approx. 71%) do appear to be within expected ranges for an average stand. It is also demonstrated that a relatively moderate number of stands are outside the expected ranges (approx. 29%) of which there is a similar percentage of stands falling below (43%) and above (57%) the expected upper and lower range thresholds.

The above indicates, that given a harvest year, nature is often unpredictable and a number of forests will be subject to atypical patterns.

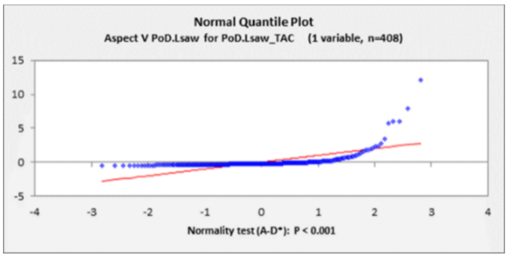
Figure 8: Histogram of average tree



4.2. Linear Regression

The one-on-one linear regressions that were produced for each of the predictor variables revealed a poor fit with issues surrounding the residuals which were not normally distributed and the A-D normality tests were skewed (see figure 9, left). After a natural log transformation of the response variable, the skewedness improved (see figure 9, right). However, the significance of the variables and any attempt at a multivariate linear regression model were unsuccessful with adjusted R^2 values being slightly improved but still insignificant (Table 6). It became apparent that the method of linear regression was likely not the appropriate tool for the data at hand and could not explain sufficient variation in the response variables or the relationships between the predictor variables.

Figure 9: AD tests for Aspect before and after transformation of response variable to natural log (LN)



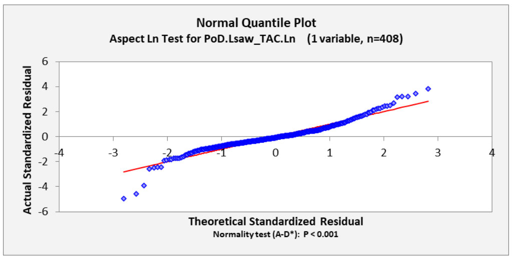


Table 6: Example of regression summary for linear regression

Summary of Regression Model Results				
Linear Model For PoD.CF_STR.Lsaw	Aspect V PoD.Lsaw	Elevation V PoD.Lsaw	Latitude V PoD.Lsaw	Dist.Coast V PoD.Lsaw
R-squared	0.002	0.011	0.001	0.009
Adjusted R-squared	-0.001	0.008	-0.002	0.001

Summary of Regression Model Results			
Linear Model For PoD.CF_STR.Lsaw	Roughness V PoD.Lsaw	Seg.Country V PoD.Lsaw	Soil V PoD.Lsaw
R-squared	0.003	0.011	0.011
Adjusted R-squared	-0.005	0.003	-0.006

4.3. Logistic Regression

4.3.1. Evaluation tools for regressions

4.3.1.1. The omnibus test

The backwards stepwise regression indicated all models showed a significant improvement over the base model with p<0.05 (see Table 7).

Table 7:Omnibus compilation of results – Backwards stepwise method

LOG. REGRESSION – RESPONSE VARIABLE	OMNIBUS TEST RESULTS
Strategic estimation of sawlog volume of *Sitka spruce*	0.015
Strategic estimation of all products of *Sitka spruce*	0.000
Strategic estimation of sawlog volume of all species	0.022
Strategic estimation of all species and products	0.032
Tactical estimation of sawlog volume of *Sitka spruce*	0.000
Tactical estimation of all products of *Sitka spruce*	0.001
Tactical estimation of sawlog volume of all species	0.001
Tactical estimation of all species and products	0.004

4.3.1.2. The Pseudo R^2

Nagelkerke's Pseudo R^2 was used to compare the models between response variables by descending value (Table 8). The Tactical forecast of large sawlog for

Sitka Spruce was revealed to be the best performing. The Strategic forecast for Sitka Spruce was comparatively the second best and the Tactical forecast for all species for large sawlog was found to be the third. The rest follow in a declining order (Pseudo $R^2 < 0.1$). The worst performing models were generally those using the Strategic forecast.

Table 8: Pseudo R^2 tests – Backwards stepwise method

LOG. REGRESSION – RESPONSE VARIABLE	COX & SNELL	NAGELKERKE
Tactical estimation of sawlog volume of *Sitka spruce*	0.080	0.113
Strategic estimation of all products of *Sitka spruce*	0.077	0.106
Tactical estimation of sawlog volume of all species	0.070	0.101
Tactical estimation of all species and products	0.059	0.081
Tactical estimation of all products of *Sitka spruce*	0.059	0.080
Strategic estimation of sawlog volume of all species	0.043	0.065
Strategic estimation of sawlog volume of Sitka spruce	0.036	0.053
Strategic estimation of all species and products	0.029	0.041

4.3.1.3. The Hosmer & Lemeshow Test

In all cases the observed event group matched the expected event group (Table 9).

Table 9: Hosmer & Lemeshow Test – Backwards stepwise method

LOG. REGRESSION – RESPONSE VARIABLE	HOSMER & LEMESHOW TEST
Strategic estimation of sawlog volume of *Sitka spruce*	0.329
Strategic estimation of all products of *Sitka spruce*	0.980
Strategic estimation of sawlog volume of all species	0.084
Strategic estimation of all species and products	1.000
Tactical estimation of sawlog volume of *Sitka spruce*	0.849
Tactical estimation of all products of *Sitka spruce*	0.686
Tactical estimation of sawlog volume of all species	0.641
Tactical estimation of all species and products	0.582

4.3.1.4. Classification Tables

The classification table results were compiled to examine the practical results of using each model (Table 10). The outcome varied from 60.2% to 76.6% in overall percentage of prediction. The percentage of accurately predicting over-estimation was at first glance high, ranging from 79.1% to 100%. However, the success in predicting under-estimation was quite poor ranging from 0% to 35.7%. Taking a closer look, the four models that appeared to predict over-estimation in 100% of cases (all Strategic estimations except the one for all products for *Sitka Spruce* volumes and the Tactical estimation of sawlog volume of all species) were also the ones that failed to predict under-estimation in any capacity resulting in significant Type 1 errors with false positive predictions. To various extents, Type 1 errors were apparent in all models whereas Type 2 errors of false negative predictions were smaller in value and only in half the models with the most observed in the Tactical estimation of all products of Sitka spruce model. The models that revealed from their classification tables to be an improvement on their respective base models were the following:

- Strategic estimation of all products of *Sitka spruce*
- Tactical estimation of sawlog volume of *Sitka spruce*
- Tactical estimation of all products of *Sitka spruce*
- Tactical estimation of all species and products

Models whose response variable was based on data from the tactical forecast data were overall better in their predictive capacity. Amongst the Tactical forecast-based models and considering the Type 1 errors, some of the better models were the Strategic estimation of all species and products and Tactical estimation of all products of Sitka spruce in overall predictive success being better at predicting under-estimation. The Tactical estimation of sawlog volume of *Sitka spruce* had a high overall score due to a high predictive capacity with over-estimation, but when considering the much lower predictive capacity for under-estimation it was seen as a less optimal model than its counterparts. The only suitable model using data from the Strategic forecast was Strategic estimation of all products of *Sitka spruce* with all others failing to predict any under-estimation.

Table 10: Classification table - Backwards stepwise method

LOG. REGRESSION – RESPONSE VARIABLE	**% UNDER**	**% OVER**	**% OVERALL**
Strategic estimation of sawlog volume of *Sitka spruce*	0	100	74.3
Strategic estimation of all products of *Sitka spruce*	25	88.4	65.5
Strategic estimation of sawlog volume of all species	0	100	76.6
Strategic estimation of all species and products	0	100	72.6
Tactical estimation of sawlog volume of *Sitka spruce*	10.6	96.7	71.2
Tactical estimation of all products of *Sitka spruce*	35.7	79.1	60.2
Tactical estimation of sawlog volume of all species	0	100	73
Tactical estimation of all species and products	19.8	92.2	66.7

4.3.2. Detailed description of regressions

4.3.2.1. Strategic estimation of sawlog volume of *Sitka spruce*

This model was ranked 7th out of the 8 that were run according to the comparison using the Negelkerke pseudo R^2 (Table 8). The backwards stepwise regression, that completed over 7 steps, resulted in a model using the predictor variables of roughness and soil type.

In the full model, with all the variables included, it could be seen that elevation, latitude and soil type were significant predictors (Table 11). Specifically, looking at their estimates and starting with elevation, categories 2 through 6 corresponding with 61-360 m were related significantly with the response variable. The negative coefficient for Elevation-Cat(2) (B=-2.09) reveals that sites at elevations 61-120 m were less likely to over-estimate compared to sites at 361-505 m. Similarly, the rest of the categories followed the same patterns having smaller negative coefficients showing that they were less likely to over-estimate compared to sites at 361-505 m though this was less the case than elevation at 61-120 m. Latitude categories of 1 and 4, corresponding with approximately 53 to 61 degrees and 69 to 77, were related significantly with the response variable. The coefficients for Latitudes 1 and 4 were similar (B1=1.053 & B4=0.965) and indicated that sites at latitudes of 53 to 61 degrees were more likely to over-estimate compared to sites at 77 to 93 degrees. Sites at 69 to 77 degrees were also more likely to over-estimate compared to sites at 77 to 93 degrees but this was less than the case that Latitudes at 53 to 61 degrees. Soil category 1, corresponding with deep acidic mineral non calcareous soils, was related significantly with the response variable. The coefficient for Soil(1) (B=-1.223) indicated that deep acidic mineral non calcareous soils were less likely to over-estimate than alkaline calcareous mineral and Cut Soils.

The first variable to be removed from the model was slope ($p = 0.948$), which affected elevation positively; slightly increasing its significance. The removal of segment of the country mildly adversely affected latitude but revealed an overall significance in roughness. Elevation continued to improve within its categories, but lost marginal overall significance with every removed variable. The removal of aspect furthered the adverse effects on latitude and elevation. The removal of distance to the coast had a substantial negative affect on elevation and latitude. The decline in the significance in elevation lead to its removal in step 5 which had a negative overall effect on Latitude leading to its removal in step 6.

The final model included soil type with overall significance, which remained largely unaffected throughout most steps. Nevertheless, the significance within its categories was affected, as was Roughness which became overall significant with the removal of segment in the country (Table 12).

Table 11: Backwards Stepwise Regression – Full model – Response variable: Strategic estimation of sawlog volume of Sitka spruce

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
Elevation-Cat			9.451	6	.150			
Elevation-Cat(1)	-1.612	.852	3.583	1	.058	.199	.038	1.059
Elevation-Cat(2)	-2.091	.731	8.171	1	.004	.124	.029	.518
Elevation-Cat(3)	-1.499	.699	4.606	1	.032	.223	.057	.878
Elevation-Cat(4)	-1.548	.697	4.929	1	.026	.213	.054	.834
Elevation-Cat(5)	-1.373	.705	3.792	1	.051	.253	.064	1.009
Elevation-Cat(6)	-1.634	.704	5.379	1	.020	.195	.049	.776
Latitude-EqDis(1)	1.053	.480	4.805	1	.028	2.867	1.118	7.350
Latitude-EqDis(2)	.617	.475	1.685	1	.194	1.853	.730	4.704
Latitude-EqDis(3)	.148	.469	.100	1	.752	1.160	.463	2.908
Latitude-EqDis(4)	.956	.453	4.440	1	.035	2.600	1.069	6.324
Soil(1)	-1.223	.509	5.782	1	.016	.294	.109	.798
Soil(2)	-.773	.544	2.019	1	.155	.462	.159	1.341
Soil(3)	-.524	.552	.900	1	.343	.592	.201	1.747

Table 12: Backwards Stepwise Regression – Final model – Response variable: Strategic estimation of sawlog volume of Sitka spruce

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
Roughness(1)	.468	.413	1.282	1	.258	1.596	.710	3.586
Roughness(2)	.109	.426	.065	1	.799	1.115	.484	2.569
Roughness(3)	-.365	.370	.976	1	.323	.694	.336	1.433
Soil			9.815	3	.020			
Soil(1)	-.770	.412	3.482	1	.062	.463	.206	1.040
Soil(2)	-.283	.429	.436	1	.509	.753	.325	1.747
Soil(3)	.048	.444	.012	1	.914	1.049	.440	2.503

4.3.2.2. Strategic estimation of all products of *Sitka spruce*

This model was ranked second best according to the Negelkerke pseudo R^2 (Table 8). In the full model, only elevation appeared to have an overall significance (Table 13). Categories 2 through 6 corresponding to 61-360 m was related significantly with the response variable. The negative coefficient for Elevation-Cat(2) (B=-2.396) revealed that sites at elevations 61-120 m were less likely to over-estimate compared to sites at 361-505 m. Consequently, the rest of the categories in the variable followed the same patterns having smaller negative coefficients. This showed that they were less likely to over-estimate compared to sites at 361-505 m though this was less the case than elevations at 61-120 m.

The removal of variables in the ensuing steps showed that the removal of the variable distance to the coast corresponded with an increase in the significance of soil type overall. The significance of soil category 3 (blanket peat) was lost ($p > 0.05$) when segment within the country was removed. Roughness had a sizeable positive effect from the removal of latitude but not enough to become significant, the other variables had a small effect on it. Elevation was a strong significant variable throughout the regression with marginal effects from the other variables.

In the final step, the variable of elevation had similar descriptions as the full model with mildly increased coefficients (B). The variable of soil type remained with overall significance and the variable of roughness with no improvement to the model estimated with any further removals (Table 14).

Table 13: Backwards Stepwise Regression – Full model – Response variable: Strategic estimation of all products of Sitka spruce

							95% C.I.for EXP(B)	
	B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Elevation-Cat			20.321	6	.002			
Elevation-Cat(1)	-1.077	.779	1.913	1	.167	.341	.074	1.567
Elevation-Cat(2)	-2.396	.659	13.234	1	.000	.091	.025	.331
Elevation-Cat(3)	-1.812	.626	8.385	1	.004	.163	.048	.557
Elevation-Cat(4)	-1.492	.631	5.587	1	.018	.225	.065	.775
Elevation-Cat(5)	-1.968	.627	9.860	1	.002	.140	.041	.477
Elevation-Cat(6)	-1.628	.635	6.577	1	.010	.196	.057	.681

Table 14: Backwards Stepwise Regression – Final model – Response variable: Strategic estimation of all products of Sitka spruce

							95% C.I.for EXP(B)	
	B	S.E	Wald	df	Sig.	Exp(B)	Lower	Upper
Elevation-Cat			18.896	6	.004			
Elevation-Cat(1)	-.985	.695	2.005	1	.157	.374	.096	1.460
Elevation-Cat(2)	-2.156	.605	12.700	1	.000	.116	.035	.379
Elevation-Cat(3)	-1.704	.589	8.370	1	.004	.182	.057	.577
Elevation-Cat(4)	-1.350	.597	5.118	1	.024	.259	.081	.835
Elevation-Cat(5)	-1.878	.600	9.806	1	.002	.153	.047	.495
Elevation-Cat(6)	-1.628	.616	6.978	1	.008	.196	.059	.657
Roughness			6.808	3	.078			
Roughness(1)	.700	.380	3.397	1	.065	2.013	.957	4.236
Roughness(2)	.377	.395	.909	1	.340	1.458	.672	3.164
Roughness(3)	.071	.350	.042	1	.838	1.074	.541	2.134

Soil			9.753	3	.021			
Soil(1)	-.138	.378	.134	1	.714	.871	.416	1.825
Soil(2)	-.251	.392	.411	1	.522	.778	.361	1.678
Soil(3)	.570	.405	1.981	1	.159	1.769	.799	3.913

4.3.2.3. Strategic estimation of sawlog volume of all species

This model was ranked 5th out of the 8 that were run compared by the Negelkerke pseudo R^2 (Table 8). In the full model as well as the final model, elevation and soil type were significant (Table 15). Elevation categories 2, 3 and 6 corresponding to 61-180 m and 301-360 m were related significantly with the response variable. The negative coefficient for Elevation-Cat(2) (B=-1.836) reveals that sites at elevations 61-120 m were less likely to over-estimate compared to sites at 361-505 m. Consequently, categories 3 and 6 followed the same patterns having smaller negative coefficients showing that 120-180 m and 301-360 m elevations were less likely to over-estimate compared to sites at 361-505 m though this was less the case than elevations at 61-120 m and Elevation-Cat(3) less the case from Elevation-Cat(6). Soil type showed an overall significance as a variable (p=0.017). Sites with Soil(1) and Soil(2), corresponding to Acidic, well to poorly drained, Shallow or Deep, non-calcareous, with or without peaty top soil sites were less likely to over-estimate then alkaline calcareous mineral soils and cut. For Soil(2) it was less the case than Soil(1) (B_{s1}=-1.7 and B_{s2}=-1.4).

The removal of variables in the ensuing steps showed that with the removal of aspect, elevation lost significance in its 3rd category ($p > 0.05$) but this effect was reversed by the removal of slope. The removal of distance to the coast also reduced the significance of elevation categories 3 & 4, but this was partially reversed by the removal of latitude. Soil type was beneficially affected by segment within the country bringing overall significance to the variable. The removal of aspect reduced the significance of soil type's 3rd category ($p > 0.05$), this trend continued until the last step until the removal of latitude which increased significance but not to $p < 0.05$.

The final step after the removal of latitude, which had a minimal effect on the model, the regression was concluded with the variables of elevation and soil type with similar significance as with the full model (Table 16).

Table 15: Backwards Stepwise Regression – Full model – Response variable: Strategic estimation of sawlog volume of all species

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
Elevation-Cat			10.450	6	.107			
Elevation-Cat(1)	-.830	.891	.868	1	.351	.436	.076	2.498
Elevation-Cat(2)	-1.836	.729	6.348	1	.012	.159	.038	.665
Elevation-Cat(3)	-1.359	.694	3.828	1	.050	.257	.066	1.002
Elevation-Cat(4)	-1.296	.695	3.478	1	.062	.274	.070	1.068
Elevation-Cat(5)	-.938	.706	1.763	1	.184	.392	.098	1.563
Elevation-Cat(6)	-1.507	.701	4.623	1	.032	.222	.056	.875
Soil			10.211	3	.017			
Soil(1)	-1.725	.594	8.444	1	.004	.178	.056	.570
Soil(2)	-1.434	.628	5.208	1	.022	.238	.070	.817
Soil(3)	-1.166	.633	3.397	1	.065	.312	.090	1.077

Table 16: Backwards Stepwise Regression – Final model – Response variable: Strategic estimation of sawlog volume of all species

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
Elevation-Cat(1)	-.851	.786	1.174	1	.279	.427	.091	1.991
Elevation-Cat(2)	-1.675	.665	6.338	1	.012	.187	.051	.690
Elevation-Cat(3)	-1.287	.654	3.872	1	.049	.276	.077	.995
Elevation-Cat(4)	-1.279	.656	3.807	1	.051	.278	.077	1.006
Elevation-Cat(5)	-.944	.675	1.959	1	.162	.389	.104	1.459
Elevation-Cat(6)	-1.527	.680	5.040	1	.025	.217	.057	.824
Soil			9.375	3	.025			
Soil(1)	-1.526	.535	8.133	1	.004	.217	.076	.620
Soil(2)	-1.253	.547	5.239	1	.022	.286	.098	.835
Soil(3)	-1.049	.552	3.614	1	.057	.350	.119	1.033

4.3.2.4. Strategic estimation of all species and products

According to Nagelkerke's pseudo R^2, this was the worst performing model (Table 8). The full model, as with all previous, produced some significance with the variable of elevation (Table 17). Elevation categories (5) and (6) corresponding to 241-360 m were related significantly with the response variable. The negative coefficient for elevation categories (5) and (6) (B=-1.274/-1.562) revealed that sites at elevations 241-360 m were less likely to over-estimate compared to sites at 361-505 m.

Elevation was negatively affected by the removal of latitude losing significance. The removal of segment within the country returned significance, which remained to the final step.

The final step of the model included only the variable Elevation, which was significant for category 5 and 6, similar to the full model (Table 18).

Table 17: Backwards Stepwise Regression – Full model – Response variable: Strategic estimation for all products and species

							95% C.I.for EXP(B)	
	B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Elevation-Cat(1)	.083	.832	.010	1	.920	1.087	.213	5.551
Elevation-Cat(2)	-.976	.635	2.360	1	.124	.377	.108	1.309
Elevation-Cat(3)	-1.048	.592	3.129	1	.077	.351	.110	1.120
Elevation-Cat(4)	-1.099	.594	3.422	1	.064	.333	.104	1.068
Elevation-Cat(5)	-1.274	.594	4.603	1	.032	.280	.087	.896
Elevation-Cat(6)	-1.562	.597	6.847	1	.009	.210	.065	.676

Table 18: Backwards Stepwise Regression – Final model – Response variable: Strategic estimation for all products and species

							95%C.I.for EXP(B)	
	B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Elevation-Cat			12.201	6	.058			
Elevation-Cat(1)	.156	.719	.047	1	.828	1.169	.286	4.784
Elevation-Cat(2)	-.711	.541	1.727	1	.189	.491	.170	1.418
Elevation-Cat(3)	-.853	.531	2.583	1	.108	.426	.151	1.206
Elevation-Cat(4)	-.962	.534	3.245	1	.072	.382	.134	1.088
Elevation-Cat(5)	-1.071	.542	3.906	1	.048	.343	.119	.991
Elevation-Cat(6)	-1.453	.563	6.665	1	.010	.234	.078	.705

4.3.2.5. Tactical estimation of sawlog volume of *Sitka spruce*

According to Nagelkerke's pseudo R^2 this was the better performing model (Table 8). The full model and the final model found elevation, latitude and roughness to be related significantly with the response variable (Tables 19 and 20). The Coefficients for elevation were negative with the values following a descending order with the close exception of Elevation-Cat(3) and (4). Elevation-Cat(1) (B=-3.004) revealed that sites at elevations 10-61m were less likely to over-estimate compared to sites at 361-505 m. Every category thereafter, with the exception of the aforementioned (3) and (4) which were reverse, followed the same patterns having descending lower negative coefficients. This indicates that as elevation decreases there is less likelihood for over-estimation compared to sites at 361-505 m and with every category being less the case from the one above it. In other words, as elevation decreases so does the intensity of over-estimation.

The variable latitude had overall significance with Latitudes (1) and (2) having similar coefficients (B_1=1.176; B_2=1.158) indicating that sites at latitudes of 54 to 61 degrees were more likely to over-estimate compared to sites at 77 to 93 degrees. Sites at 61 to 65 degrees were also more likely to over-estimate compared to sites at 77 to 93 degrees but this is less the case than latitudes at 54 to 61 degrees. The variable roughness showed overall significance but without significance between its categories and the reference category.

The removal of variables in the ensuing steps showed that in step 3 roughness was negatively affected by the removal of the distance to the coast variable but remained significant overall. Roughness regained a positive effect in the next step with the removal of slope.

The model concluded after six steps with the significant variables of elevation, latitude and roughness with no improvement to the model estimated by any further variables removed (Table 20). The significance and coefficients were similar to the full model, who's key variables were not strongly impacted by the regression steps.

Table 19: Backwards Stepwise Regression – Full model – Response variable: Tactical estimation of sawlog volume of Sitka spruce

							95% C.I.for EXP(B)	
	B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Elevation-Cat			11.126	6	.085			
Elevation-Cat(1)	-3.004	.925	10.556	1	.001	.050	.008	.304
Elevation-Cat(2)	-2.537	.846	8.994	1	.003	.079	.015	.415
Elevation-Cat(3)	-2.511	.815	9.487	1	.002	.081	.016	.401
Elevation-Cat(4)	-2.526	.818	9.531	1	.002	.080	.016	.398
Elevation-Cat(5)	-2.416	.816	8.756	1	.003	.089	.018	.442
Elevation-Cat(6)	-2.396	.819	8.565	1	.003	.091	.018	.453
Latitude-EqDis			17.229	4	.002			
Latitude-EqDis(1)	1.176	.456	6.663	1	.010	3.242	1.327	7.918
Latitude-EqDis(2)	1.158	.460	6.326	1	.012	3.182	1.291	7.844
Latitude-EqDis(3)	.025	.440	.003	1	.955	1.025	.433	2.427
Latitude-EqDis(4)	.757	.411	3.392	1	.065	2.133	.953	4.774
Roughness			8.768	3	.033			
Roughness(1)	.187	.437	.183	1	.668	1.206	.512	2.839
Roughness(2)	.605	.471	1.651	1	.199	1.831	.728	4.606
Roughness(3)	-.321	.396	.656	1	.418	.726	.334	1.577

Table 20: Backwards Stepwise Regression – Final model – Response variable: Tactical estimation of sawlog volume of Sitka spruce

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
Elevation-Cat			10.649	6	.100			
Elevation-Cat(1)	-2.639	.846	9.740	1	.002	.071	.014	.375
Elevation-Cat(2)	-2.115	.790	7.162	1	.007	.121	.026	.568
Elevation-Cat(3)	-2.177	.779	7.816	1	.005	.113	.025	.522
Elevation-Cat(4)	-2.270	.781	8.440	1	.004	.103	.022	.478
Elevation-Cat(5)	-2.246	.790	8.092	1	.004	.106	.023	.497
Elevation-Cat(6)	-2.382	.801	8.850	1	.003	.092	.019	.444
Latitude-EqDis			13.899	4	.008			
Latitude-EqDis(1)	.753	.333	5.100	1	.024	2.122	1.105	4.078
Latitude-EqDis(2)	.782	.348	5.041	1	.025	2.186	1.104	4.326
Latitude-EqDis(3)	-.158	.327	.233	1	.629	.854	.450	1.622
Latitude-EqDis(4)	.577	.335	2.964	1	.085	1.781	.923	3.434
Roughness			7.959	3	.047			
Roughness(1)	.172	.414	.173	1	.677	1.188	.528	2.673
Roughness(2)	.551	.439	1.575	1	.209	1.735	.734	4.101
Roughness(3)	-.261	.375	.483	1	.487	.770	.369	1.607

4.3.2.6. Tactical estimation of all products of *Sitka spruce*

This model was of average performance, ranking 5th of the 8 in total according to Negelkerke's pseudo R^2 (Table 8). The full model, with all the variables entered, found elevation to be related significantly with the response variable (Table 21). Elevation was found to have overall significance and according to the negative coefficients, category 2 corresponding to 61-120 m had the highest negative coefficient for Elevation (B=-2.085) which revealed that those sites at elevation 61-120m were less likely to over-estimate compared to sites at 361-505 m. Consequently, categories 1 and 3 to 6 followed a similar pattern but with smaller negative coefficients showing that 10-61 m and 120-360 m elevations were less likely to over-estimate compared to sites at 361-505 m though this was less the case than elevations at 61-120 m and between the categories according to the descending order of their respective coefficients.

The removal of variables in the ensuing steps showed that elevation continued to be significant and largely unaffected through to the final step. Roughness in the full model was not significant, but was close to significance (p=0.057) as well as the first two of its categories. The removal of the variables of aspect, distance to the coast and soil type appeared to have had a negligible effect on roughness. However, the removal of latitude in step 4 achieved a positive effect making the variable significantly related with the variable. This significance continued to improve with the removal of segment of the country leading to an overall significance in the variable.

The final model included elevation and roughness (Table 22). Elevation was very similar in significance and coefficients to the full model. Roughness gained significance with Categories 1 and 2, corresponding to unknown and even soils, more likely to over-estimate compared to sites that were rough, with similar coefficients (B_1=0.828; B_2=0.841).

Table 21: Backwards Stepwise Regression – Full model – Response variable: Tactical estimation of all products of Sitka spruce

							95% C.I.for EXP(B)	
	B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Elevation-Cat			14.209	6	.027			
Elevation-Cat(1)	-1.704	.687	6.163	1	.013	.182	.047	.699
Elevation-Cat(2)	-2.085	.589	12.546	1	.000	.124	.039	.394
Elevation-Cat(3)	-1.588	.551	8.306	1	.004	.204	.069	.602
Elevation-Cat(4)	-1.573	.555	8.015	1	.005	.207	.070	.616
Elevation-Cat(5)	-1.861	.554	11.294	1	.001	.156	.053	.460
Elevation-Cat(6)	-1.546	.563	7.539	1	.006	.213	.071	.642

Table 22: Backwards Stepwise Regression – Final model – Response variable: Tactical estimation of all products of Sitka spruce

							95% C.I.for EXP(B)	
	B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Elevation-Cat			15.425	6	.017			
Elevation-Cat(1)	-1.628	.583	7.799	1	.005	.196	.063	.616
Elevation-Cat(2)	-1.919	.513	14.020	1	.000	.147	.054	.401
Elevation-Cat(3)	-1.385	.501	7.631	1	.006	.250	.094	.669
Elevation-Cat(4)	-1.314	.508	6.685	1	.010	.269	.099	.728
Elevation-Cat(5)	-1.619	.518	9.781	1	.002	.198	.072	.546
Elevation-Cat(6)	-1.422	.539	6.968	1	.008	.241	.084	.693
Roughness			12.076	3	.007			
Roughness(1)	.828	.364	5.183	1	.023	2.289	1.122	4.669
Roughness(2)	.841	.384	4.793	1	.029	2.318	1.092	4.922
Roughness(3)	.203	.339	.360	1	.549	1.225	.631	2.381

4.3.2.7. Tactical estimation of sawlog volume of all species

This model had a good performance, ranking 3rd of the 8 in total according to Negelkerke's pseudo R^2 (Table 8), but when taking into consideration the classification table results (Table 10), there are substantial type 1 errors.

In the full model (Table 23) the variables elevation, latitude and soil type were found to be related significantly with the response variable. The performance of these variables was enhanced moderately through the backwards stepwise process and due to the similarities, the variables will be described in detail in the final model section further down.

The initial removal of the variables aspect and slope showed no significant effect on the model. The removal of the segment in the country in step 4 increased the significance of latitude and brought an overall significance into roughness. The removal of the Distance to the coast variable in step 5 decreased the significance of the model by affecting roughness and soil type negatively. The removal of soil type and the next least significant variable further depreciated the model into its final form with the variables of elevation, latitude and roughness with the latter no longer significant.

The model could be seen to perform best at step 4 (Table 24) with elevation, latitude, roughness and soil type controlling for distance from the coast. So this was selected as the final model. In the final model, Elevation was observed to be related significantly with the response variable. The coefficients for this variable were very similar with a slight discernible increase from categories 1 to 4 and then a decrease in 5 and then an increase in 6, similar but lower than categories 3 and 4. This indicates that sites in elevations between 10-240 m were progressively less likely to over-estimate in comparison to sites with an elevation of 361-505 m. The above over-estimation is less the case for sits at 241-360 m, 241-300 m and 301 to 360 m. Latitude was found to have overall significance, with sites within categories 1 and 2 corresponding to 53-65 degrees being more likely to over-estimate than sites in latitudes of 77-93 degrees. For sites in category 1 with a coefficient of B=0.862 (approx. 53-61 degrees) was less the case than sites in category 2 (approx. 61-65 degrees) with coefficient B=1.371. Roughness was found to have only overall significance with its reference category and soil type had one category significant towards the reference category. Soil 1, corresponding with to Deep Acidic Mineral Non-Calcareous sites were less likely to over-estimate then alkaline calcareous mineral soils and cut.

Table 23: Backwards Stepwise Regression – Full model – Response variable: Tactical estimation of sawlog volume of all species

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
Elevation-Cat			8.841	6	.183			
Elevation-Cat(1)	-1.997	.944	4.475	1	.034	.136	.021	.864
Elevation-Cat(2)	-2.214	.843	6.899	1	.009	.109	.021	.570
Elevation-Cat(3)	-2.292	.810	8.011	1	.005	.101	.021	.494
Elevation-Cat(4)	-2.308	.811	8.095	1	.004	.099	.020	.488
Elevation-Cat(5)	-2.169	.810	7.181	1	.007	.114	.023	.558
Elevation-Cat(6)	-2.194	.813	7.273	1	.007	.112	.023	.549
Latitude-EqDis			13.975	4	.007			
Latitude-EqDis(1)	.813	.458	3.154	1	.076	2.255	.919	5.532
Latitude-EqDis(2)	1.356	.487	7.754	1	.005	3.881	1.494	10.078
Latitude-EqDis(3)	.055	.450	.015	1	.903	1.056	.438	2.550
Latitude-EqDis(4)	.572	.420	1.859	1	.173	1.773	.778	4.036
Soil			6.145	3	.105			
Soil(1)	-1.027	.518	3.937	1	.047	.358	.130	.987
Soil(2)	-.575	.558	1.065	1	.302	.562	.189	1.677
Soil(3)	-.539	.551	.955	1	.329	.583	.198	1.719

Table 24: Backwards Stepwise Regression – Final model – Response variable: Tactical estimation of sawlog volume of all species

	B	S.E.	Wald	df	Sig.	Exp(B)	95%C.I.for EXP(B)	
							Lower	Upper
Elevation-Cat			8.419	6	.209			
Elevation-Cat(1)	-2.062	.910	5.137	1	.023	.127	.021	.757
Elevation-Cat(2)	-2.162	.824	6.884	1	.009	.115	.023	.579
Elevation-Cat(3)	-2.214	.793	7.798	1	.005	.109	.023	.517
Elevation-Cat(4)	-2.210	.791	7.808	1	.005	.110	.023	.517
Elevation-Cat(5)	-2.083	.794	6.881	1	.009	.125	.026	.591
Elevation-Cat(6)	-2.185	.804	7.392	1	.007	.112	.023	.543
Latitude-EqDis			16.278	4	.003			
Latitude-EqDis(1)	.862	.352	6.000	1	.014	2.368	1.188	4.720
Latitude-EqDis(2)	1.371	.402	11.619	1	.001	3.938	1.791	8.662
Latitude-EqDis(3)	.174	.360	.233	1	.629	1.190	.587	2.412
Latitude-EqDis(4)	.642	.371	2.996	1	.083	1.900	.919	3.931
Dist.Coast			2.994	3	.392			
Dist.Coast(1)	.569	.386	2.168	1	.141	1.766	.828	3.766
Dist.Coast(2)	.253	.341	.551	1	.458	1.288	.660	2.515

Dist.Coast(3)	-.044	.305	.021	1	.884	.957	.527	1.738
Roughness			8.085	3	.044			
Roughness(1)	-.003	.433	.000	1	.995	.997	.427	2.331
Roughness(2)	.510	.476	1.148	1	.284	1.665	.655	4.228
Roughness(3)	-.403	.395	1.038	1	.308	.669	.308	1.450
Soil			6.361	3	.095			
Soil(1)	-1.145	.504	5.155	1	.023	.318	.118	.855
Soil(2)	-.784	.522	2.253	1	.133	.457	.164	1.271
Soil(3)	-.730	.534	1.874	1	.171	.482	.169	1.371

4.3.2.8. Tactical estimation of all species and products

This model was of moderate performance ranking 4th of the 8 in total according to Negelkerke's pseudo R^2 (Table 8). The full model revealed a significant relationship between the response variable and elevation, slope and segment of the country (Table 25). Elevation was significant overall as well as within categories 2-6, corresponding with 61-360 and Slope showed overall significance for its reference category. The performance of these variables was enhanced moderately through the backwards stepwise process and due to the similarities, the variables will be described in detail in the final model section further down.

The removal of variables in the ensuing steps showed that elevation and slope remained unaffected by the removal of other variables to the final step. Roughness became significant in category 2 (Even) after the removal of latitude and retained this significance until its removal in step 5. Segment in the country became significant in category 3 (SE) after the removal of aspect and retained it until its removal in step 6. Soil type was close to being significant but never quite achieved it, however no benefit was gained by its removal at any stage of the regression. At step 4, the model had the most variables with various degrees of significance with no positive gain by the removal of any other variables. The further removals of roughness and segment within the country in the next steps, had they been accepted, had little effect on the model.

The 4th step of the backwards stepwise regression was selected as the final model with the variables of elevation, slope, roughness, segment within country and soil type. During the evaluation of the final model (Table 26), elevation had an overall significance and was significantly related with the response variable. Elevations 2 to 6, corresponding with 61-360 m were less likely to over-estimate in comparison to sites with an elevation of 361-505 m. For category 3 (121-180 m) this was less the case than category 2 (61-120 m) and for both it was less the case than the categories 4 (181-240 m), 5 (241-300 m) and 6 (301-360 m) which were progressively less likely to over-estimate as seen from categories 4 towards 6.

The variable of slope appeared to have an overall significance but no specific indication of significance between the categories and the reference category. The variable of roughness, though not significant overall, revealed sites in its 2nd category, corresponding with even sites, were more likely to over-estimate than rough sites. However, the coefficient indicates that this effect was not very strong. The Segment within the country variable, though not significant overall, had sites in category 3, corresponding with the south-eastern quadrant, more likely to over-estimate than sites in the reference category, corresponding with the south-western quadrant. Soil type was retained in the final model but was not significant, however, its presence was maintained to the last step as no improvement to the model was achieved with its removal.

Table 25: Backwards Stepwise Regression – Full model – Response variable: Tactical estimation of all species and products

							95% C.I.for EXP(B)	
	B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Elevation-Cat			14.834	6	.022			
Elevation-Cat(1)	-1.234	.725	2.898	1	.089	.291	.070	1.205
Elevation-Cat(2)	-1.560	.615	6.424	1	.011	.210	.063	.702
Elevation-Cat(3)	-1.292	.581	4.944	1	.026	.275	.088	.858
Elevation-Cat(4)	-1.733	.582	8.852	1	.003	.177	.056	.554
Elevation-Cat(5)	-1.883	.584	10.382	1	.001	.152	.048	.478
Elevation-Cat(6)	-1.954	.593	10.873	1	.001	.142	.044	.453
Slope-Cat			10.736	3	.013			
Slope-Cat(1)	.156	.594	.069	1	.793	1.168	.364	3.746
Slope-Cat(2)	1.002	.591	2.873	1	.090	2.723	.855	8.670
Slope-Cat(3)	.075	.594	.016	1	.899	1.078	.337	3.451
Seg.Country			4.951	3	.175			
Seg.Country(1)	.435	.398	1.196	1	.274	1.545	.708	3.369
Seg.Country(2)	.126	.389	.104	1	.747	1.134	.529	2.431
Seg.Country(3)	.584	.279	4.372	1	.037	1.792	1.037	3.097

Table 26: Backwards Stepwise Regression – Final model – Response variable: Tactical estimation of all species and products

	B	S.E.	Wald	df	Sig.	Exp(B)	95%C.I.for EXP(B) Lower	Upper
Elevation-Cat			13.921	6	.031			
Elevation-Cat(1)	-1.153	.670	2.960	1	.085	.316	.085	1.174
Elevation-Cat(2)	-1.464	.574	6.504	1	.011	.231	.075	.713
Elevation-Cat(3)	-1.219	.552	4.880	1	.027	.296	.100	.872
Elevation-Cat(4)	-1.656	.555	8.908	1	.003	.191	.064	.566
Elevation-Cat(5)	-1.753	.561	9.775	1	.002	.173	.058	.520
Elevation-Cat(6)	-1.796	.569	9.972	1	.002	.166	.054	.506
Slope-Cat			10.373	3	.016			
Slope-Cat(1)	.211	.576	.134	1	.714	1.235	.400	3.816
Slope-Cat(2)	1.013	.572	3.139	1	.076	2.754	.898	8.448
Slope-Cat(3)	.168	.580	.084	1	.772	1.183	.379	3.689
Roughness			5.398	3	.145			
Roughness(1)	.704	.378	3.476	1	.062	2.022	.965	4.237
Roughness(2)	.927	.408	5.173	1	.023	2.528	1.137	5.621
Roughness(3)	.620	.357	3.023	1	.082	1.859	.924	3.739
Seg.Country			5.319	3	.150			
Seg.Country(1)	.333	.332	1.005	1	.316	1.395	.728	2.676
Seg.Country(2)	.036	.315	.013	1	.908	1.037	.559	1.923
Seg.Country(3)	.569	.263	4.672	1	.031	1.766	1.054	2.958
Soil			5.873	3	.118			
Soil(1)	.076	.413	.034	1	.854	1.079	.481	2.422
Soil(2)	-.594	.438	1.841	1	.175	.552	.234	1.302
Soil(3)	-.112	.426	.069	1	.793	.894	.388	2.060

4.3.3. Summary of results

The machine decision to remove a variable from the model is based on the probability of the likelihood-ratio statistic based on the maximum partial likelihood estimates (*Backwards Regression (LR)* 2021).

What can be seen in Table 27 below, is the removed variables in every step across as regressions. The shaded area covers the first 3 steps and helps with the amalgamation into the top 3 results in the overall section on the bottom right side. Looking at the variables that were removed in stepwise order from every model, we can see that aspect, slope and segment within the country tend to be the least favoured by the regressions. These three variables were consistently found to be selected for removal within the first 3 steps in the majority of the regressions. The following 3 variables to be removed most often starting from step 4 were latitude, distance from coast and soil type.

Table 28 provides a summary view of the variables found in the final models after no more benefit can be gained by any further removals. The most common variable to be found in the final models is that of elevation (7 of 8) followed by roughness (5 of 8) and soil type (5 of 8). The only other variable to be found in a final step more than once was latitude. Latitude appeared to be a good candidate in tactical forecast volume datasets for large sawlog and regardless of species. The only regression that didn't include elevation in its final step was for the strategic forecast volumes for large sawlog for *Sitka spruce*.

Table 27: Tables of variables removed per model

STR.Lsawlog.SS

Steps	Variable
1	Slope
2	Seg.Country
3	Aspect
4	Dist.Coast
5	Elevation
6	Latitude

STR.SS

Steps	Variable
1	Dist.Coast
2	Slope
3	Seg.Country
4	Latitude
5	Aspect

STR.Lsawlog.Total

Steps	Variable
1	Seg.Country
2	Aspect
3	Slope
4	Roughness
5	Dist.Coast
6	Latitude

STR.Total

Steps	Variable
1	Roughness
2	Aspect
3	Slope
4	Latitude
5	Dist.Coast
6	Seg.Country
7	Soil

TAC.Lsawlog.SS

Steps	Variable
1	Aspect
2	Dist.Coast
3	Slope
4	Seg.Country
5	Soil

TAC.SS

Steps	Variable
1	Aspect
2	Dist.Coast
3	Soil
4	Latitude
5	Seg.Country
6	Slope

TAC.Lsawlog.Total

Steps	Variable
1	Aspect
2	Slope
3	Seg.Country
4	Dist.Coast
5	Soil

TAC.Total

Steps	Variable
1	Latitude
2	Aspect
3	Dist.Coast
4	Roughness
5	Seg.Country

Overall

Variable	**Top 3**
Aspect	7
Slope	6
Seg.Country	4
Variable	**Low 3**
Latitude	5
Dist.Coast	4
Soil Type	3

Table 28: Summary of variables in final models

STR.Lsawlog.SS

No.	Variable
1	Roughness
2	Soil type

STR.SS

No.	Variable
1	Elevation
2	Roughness
3	Soil type

STR.Lsawlog.Total

No.	Variable
1	Elevation
2	Soil type

STR.Total

No.	Variable
1	Elevation

TAC.Lsawlog.SS

No.	Variable
1	Elevation
2	Latitude
3	Roughness

TAC.SS

No.	Variable
1	Elevation
2	Roughness

TAC.Lsawlog.Total

No.	Variable
1	Elevation
2	Latitude
3	Dist.Coast
4	Roughness
5	Soil Type

TAC.Total

No.	Variable
1	Elevation
2	Slope
3	Roughness
4	Seg.Country
5	Soil Type

Top 3

Variable	**No.**
Elevation	7
Roughness	5
Soil Type	5

5. Discussion

5.1. Evaluation of the Linear Regression

The outputs of the linear regression, after various corrections and transformations of the response variables, revealed that it was unlikely for the current data to yield any significant results with this particular method. Estimation bias was clearly evident in the forecasts and especially focused around over-estimation (Tables 4 & 5).

If a relationship between the predictor variables and the response variables could not be found, it is probable that in order to use a linear regression, the data needed to be more targeted to clear the noise and with a larger population of forests. Several methods could accomplish this such as a timescale in order to have a more structured and longer scope of the effects between the variables. A method to approach this could utilise data from several harvest years, gathered specifically around sites meeting specific characteristics, such as using only sites that over-estimate.

A larger pool of data could bring a clearer statistical view to the subcategories that are formed. To see the value in this, we could turn our attention to the fact that some categories in the predictor variables were poorly represented due to using only a single harvest year (e.g., there were very few sites with a Northern Aspect in the 2018 schedule) and others that became poorly represented when further divided into the categories of another variable (e.g., The roughness per segment of the country). Considering this research was targeted at examining the effectiveness of data already possessed for additional value, expanding further in data already available by adding additional harvest years is an easy way to broaden each variable. The use of a timescale could potentially provide evidence that criteria such as stand maturity, might not be the only suitable decision maker in harvesting and that an amalgamation of sites that are more difficult to forecast with sites that are more accurate could improve the overall accuracy of the forecast for a given harvest year.

5.2. Evaluation of the Logistic Regression

The logistic regression revealed a greater amount of information about the relationships between the predictor variables and the response variable. It was most evident that elevation was by far the best predictor variable having presence in all full models and their subsequent final steps (Table 28). There was an overall negative relationship (negative coefficients) with over-estimation bias across all regressions. The nature of this negative relationship was not always consistent but some possible patterns emerged. Elevation was found to have a much higher number of significant categories in datasets based on tactical forecast volumes. Overall significance amongst the better performing regressions was apparent in both tactical and strategic forecast datasets for *Sitka spruce* volume only, as well as with tactical forecast volume for all species and products set. This would indicate it is a poorer candidate for large sawlog alone. In the datasets for all species and products, elevation appeared to have an order of descending values of coefficients from higher elevations to lower against the reference category which corresponds with the highest elevation. This could indicate an expected hypothesis that lower elevations behave in a more predictable manner or have less adverse site conditions like exposure, water runoff, rougher soils and a correlation with higher slopes. However, for most other regressions this is not the case as elevation was not as ordered with its coefficients.

Elevation 1, corresponding to 10-60 m, was found to have the lowest or second lowest coefficients in 6 out of the 8 regression models and in 5 of 8 models no overall significance. Elevation 1 yielded better results in the tactical volume datasets generally. Specifically, it became significant ($P < 0.05$) but with the lowest coefficient in the tactical volume for large sawlog across all species set. It became more significant in the tactical volume for all products of *Sitka Spruce* set and was the most significant in the tactical volume for large sawlog for *Sitka spruce* set. This revealed that potentially for *Sitka Spruce* large sawlog volumes, lower elevations had the least over-estimation in both the full models and the final steps.

Elevation 2, corresponding to 61-120 m, had very strong coefficients in the *Sitka spruce* only regressions for all products. It performed the poorest in *Sitka spruce* datasets for large sawlog. Therefore, amongst the better performing regressions it is an indicator of least over-estimation for Sitka Spruce overall. The reason for elevations 10-60 m behaving so much worse than those at 61-120 m could be related to the fact that the very low elevations, which are also flatter, commonly have trouble with drainage in Ireland and therefore are less predictable if the sites are not tended to during their rotation. After Elevations 2 (61-120 m) the patterns become less evident and the remaining categories do not exhibit any strong pattern between regressions.

Among the better performing regressions, roughness and soil type are seen as good predictors even though the effect of roughness is not always visible through typical significance in the full models (Table 28).

Overall, soil type has a negative relationship with over-estimation similar to elevation. Soil type was found to be significant in several regressions and amongst the ones that have significant values within their categories it appears that acidic soils overestimate more so than alkaline and shallow acidic more so than deep acidic. Even though the initial hypothesis based on discussions amongst foresters was that Blanket Peat should be the most unpredictable, the regressions cannot confirm this view.

Roughness primarily had overall significance. In a single case, roughness achieved significance between its categories and the reference category, founding even sites to overestimate more than rough sites. This is difficult to interpret in a vacuum such as this as it may be related to stand stability, drainage or the distribution of rough and even sites in the specific annual plan this thesis was based on.

Finally, latitude, which is only present in one of the better regressions final steps indicates that the southern latitudes of the country are more likely to overestimate than the northern. This can explain the overall over-estimation being high nationwide as the majority of forests are located in the South.

5.2.1. The story of variables Removed

Much can be said about the variables that remain in a model but the variables that are removed also reveal useful information especially when there is consistency (Table 27). The variables of aspect, slope and segment within the country were the least favoured by the regressions. This would suggest that given the data used, these variables were not suitable to predict bias to any significant rate.

The variable of aspect had a strong influence with its removal in 3 of the 8 regressions. This influence was mostly positive after removal to the significance of other variables. It would appear that the relationship of directional aspect with that of ground roughness has some importance and one has an effect on the other. It is speculative to provide any explanation and it would definitely be a relationship worth examining further in the future. It’s possible that erosion and soil depth is related to directional aspect as some directions face harsher weather than others on a more consistent basis.

In 6 of the 8 regressions, the variable of segment within the country had a strong influence on the model by its removal. In 4 of those 6 regressions the effect was positive giving significance to other variables, especially roughness. In 2 of the 6 the effects were negative on other variables and concerned only the strategic forecast volume datasets for the species of *Sitka spruce.* It is most likely that the number of sample sites was not large enough for segment within the country to be fully developed. The tactical forecast volumes for all species and product sets had the only regression that found any significance in this variable with no overall significance and south-eastern sites over-estimating more compared with south-western sites. At the beginning of this thesis, it was stated that an overall belief exists amongst the foresters in Coillte that the eastern sites are better than the western sites and perhaps that belief can be seen as overconfidence in the forecasts. Furthermore, the removal of segment within the country almost always benefitted the remaining variables making it a poor candidate for the purposes of this thesis and inferior to its large-scale geographical counterpart, latitude.

The variable of slope had a relatively small contribution to the models. Slope was found to be the second most removed variable in the first three steps of the regressions. Regressions with total species values were found to gain significance with the removal of this variable. Conversely, the regression using the set of tactical forecast volumes for large sawlog in *Sitka spruce* was the only regression to lose significance by the removal of slope. The only regression that found slope to be overall significant was the tactical forecast volumes for all species and products, though no inferences can be made about the categories. Slope may not have been a significant contributor due to the site selection process which would limit the inclusion of a large number of sites with significant slopes from accumulating in the annual harvest schedule. Subsequent substitutions may have occurred with sites on the harvest schedule for 2018 due to the complexities surrounding road access, maintenance, logging and space for roadside timber placement. Therefore, a longer period with additional harvest years could express the value of this variable better.

6. Conclusion

In this thesis work, factors affecting the estimation bias were studied. Overall, the results of the research showed that most landscape features are not good predictors of accuracy in estimation of harvest volume in Ireland. Therefore, the interpretation of the meaning of the results has to be handled with care.

What we can do is see the suggestion and direction that has resulted from this research to understand which key areas are valuable for future focus if the pursuit of this kind of accuracy through data mining is further developed. There is evidence to support the notion that further research into estimation bias and any attempt at quantifying it for the possibility of forecast correction will include the variable of elevation. Ground roughness is also a strong candidate along with understanding its relationship to directional aspect. Attempts at understanding large scale geographical patterns with forecast bias are more likely to be successful using latitude rather than splitting the country into quadrats. Finally, soil type is also a good candidate and will likely need to be examined more closely with various tree growth variables to fully understand its potential.

Reflecting on the results, it appears that both the strategic and tactical forecast data predictions are better when using data for Sitka Spruce only. Furthermore, using only the volume of the most valuable product (large sawlog) is better for the tactical forecast but not for the strategic forecast. This could be expected as the tactical forecast uses sample measurements which increase its accuracy to the level of products as the size of the trees becomes more certain, whereas the strategic forecast, which is based primarily on growth yield class predictions, species type and provenance, would be less accurate. The overall predictive value of tactical forecast-based datasets is better than that of the strategic as 3 of the 4 tactical volume-based datasets were able to predict an amount of under-estimation whereas only 1 of the 4 strategic volume-based datasets was able to do the same.

When considering the most valuable predictor variables, elevation had the best results while it was clear that aspect, slope and segment within the country were the least suitable to use. Other research also confirms, that for Sitka spruce and in the northern country conditions, elevation is a key indicator of growth quality and therefore accuracy in forecasts while aspect on its own and slope are not strong predictors (Blyth & Macleod 1981).

In future research, there would be great value in repeating these regressions for other harvest years to ascertain the consistency of the findings. If they are confirmed, additional research on the results using a larger pull of data with more targeted characteristics is needed. There is potential to broaden the approach on how to structure a harvest year plan with less over-estimation by selecting sites with a mix of sites that are expected to produce less accurate results with sites that are known to have less variance.

References

Backwards Regression (LR) (2021). https://prod.ibmdocs-production-dal-6099123ce774e592a519d7c33db8265e-0000.us-south.containers.appdomain.cloud/docs/en/spss-statistics/27.0.0?topic=regression-logistic-variable-selection-methods [2021-06-26]

BLYTH, J.F. & MacLEOD, D.A. (1981). Sitka Spruce (Picea sitchensis) in North-East Scotland I. Relationships between Site Factors and Growth. *Forestry: An International Journal of Forest Research*, 54 (1), 41–62. https://doi.org/10.1093/forestry/54.1.41

Clifton, C. (2009). *Data mining*. *Encyclopedia Britannica - Definition of Data Mining*. https://www.britannica.com/technology/data-mining [2021-05-15]

Dobson, A.J. (2013). *Introduction to Statistical Modelling*. Springer.

Fay, D., Environmental Protection Agency, & Teagasc (2007). Soil geochemical atlas of Ireland Dublin: Teagasc and the Environmental Protection Agency.

Global Forestry Software Market 2020-2024 (2021). *Global Forestry Software Market 2020-2024*. https://www.researchandmarkets.com/reports/5136951/global-forestry-software-market-2020-2024 [2021-05-15]

Gray, H.J. (1964). The economics of Irish forestry. *Journal of The Statistical and Social Inquiry Society of Ireland*, 21 (2), 18–24

Hintze, J.L. (2007). *NCSS User's Guide III*. *NCSS Statistical Software*. https://www.ncss.com/wp-content/uploads/2012/09/NCSSUG3.pdf [2022-01-17]

Istomin, E., Stepanov, S., Petrov, Y., Kolbina, O. & Sidorenko, A. (2019). Model of optimum integration of diverse geodata for the benefit of management of forestry. *IOP Conference Series: Earth and Environmental Science*, 316, 012013. https://doi.org/10.1088/1755-1315/316/1/012013

Matthews, R.W., Jenkins, T.A.R., Mackie, E.D., Dick, E.C., Great Britain, & Forestry Commission (2016). *Forest yield: a handbook on forest growth and yield tables for British forestry*. S.l.: Forestry Commission.

McEwan, A., Marchi, E., Spinelli, R. & Brink, M. (2020). Past, present and future of industrial plantation forestry and implication on future timber harvesting technology. *Journal of Forestry Research*, 31 (2), 339–351. https://doi.org/10.1007/s11676-019-01019-3

Met Eireann & As cited by Rohan, P.K. (1986). *The Climate of Ireland*. Stationery Office. https://www.met.ie/climate/climate-of-ireland [2021-10-18]

Montgomery, D.C., Peck, E.A. & Vining, G.G. (2021). *Introduction to Linear Regression Analysis*. Page 36: John Wiley & Sons.

Moody, D.L. & Shanks, G.G. (1994). What makes a good data model? Evaluating the quality of entity relationship models. In: Loucopoulos, P. (ed.), Berlin, Heidelberg, 1994. 94–111. Berlin, Heidelberg: Springer. https://doi.org/10.1007/3-540-58786-1_75

O'Carroll, N. (2004). *Forestry in Ireland: a concise history*. Dublin, Ireland: COFORD.

Portl, S.U. (2021). Logistic Regression. *Categorical Data Analysis*, 13

Rossit, D.A., Olivera, A., Viana Céspedes, V. & Broz, D. (2019). A Big Data approach to forestry harvesting productivity. *Computers and Electronics in Agriculture*, 161, 29–52. https://doi.org/10.1016/j.compag.2019.02.029

Stojanova, D., Panov, P., Kobler, A., Džeroski, S. & Taškova, K. (2006). *Learning to Predict Forest Fires with Different Data Mining Techniques*.

Thompson, F. & Levine, D.U. (1997). Examples of Easily Explainable Suppressor Variables in Multiple Regression Research. *Multiple Linear Regression Viewpoints*, 24, 11–13

www.coillte.ie (Coillte Website) (n.d.). *Coillte*. https://www.coillte.ie/about-us/our-story/ [2021-06-11]

Zhang, W.Q. (2014). Application Research of Data Mining Technology on Growth Management of Forestry. *Advanced Materials Research*, 846–847, 995–998. https://doi.org/10.4028/www.scientific.net/AMR.846-847.995

Acknowledgements

Dedicated to the memory of my dad who guided and saw me through many adventures and who will always be a part of me and them.

Many thanks to Coillte's:

Liam Malone,
Director of Forest Planning and IT, for sitting with me to discuss what thesis subject would be valuable for the company and to advance my view on the current challenges of a modern forestry company.

John Landy,
Central Resource Planning Manager, for sitting with me to discuss the nature of the data that I could use, what could be provided as well as producing the forecast data for the Thesis.

Many thanks to SLU's:

Per Christer Oden, Professor in the Faculty of Forest Sciences,

And

Mulualem Tigabu, Associate Professor in the Faculty of Forest Sciences,

for supervising the Thesis