

Bernhard Petautschnig, BSc

**Entwicklung eines logistischen Regressionsmodelles zur  
Baumartendifferenzierung auf Basis von LiDAR- und  
Hyperspektral-Fernerkundungsdaten**

**MASTERARBEIT**

zur Erlangung des akademischen Grades

Master of Science

Masterstudium Geospatial Technologies

eingereicht an der

**Technischen Universität Graz**

Betreuer

Univ. Prof. Dr. Mathias Schardt

Institut für Geodäsie

Graz, August 2016

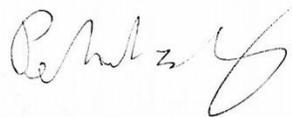
## EIDESSTÄTTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Das in TUGRAZonline hochgeladene Textdokument ist mit der vorliegenden Masterarbeit identisch.

01.08.2016

---

Datum

A handwritten signature in black ink, appearing to be 'P. Mathias', written over a light grey rectangular background.

---

Unterschrift

## DANKSAGUNG

Ich möchte mich bei meinen Eltern und meiner Lebensgefährtin Angelika bedanken, die mich während des Studiums auf jegliche Art und Weise unterstützt haben. Dank gebührt auch meinem Betreuer Univ. Prof. Dr. Mathias Schardt für die geduldige und sehr gute Begleitung während des gesamten Arbeitsverlaufes.

## **Zusammenfassung**

Diese Arbeit erörtert das Potential der Kombination von hochauflösenden LiDAR- und Hyperspektral-Fernerkundungsdaten zur Baumartendifferenzierung. Dazu werden in Feldbegehungen Referenzbäume für die untersuchten Baumarten Buche, Eiche, Fichte, Kiefer, Lärche und Tanne im Projektgebiet St. Margarethen an der Raab erhoben.

Basierend auf den Referenzbäumen wird ein multinomiales logistisches Regressionsmodell mit Variablen aus LiDAR- und Hyperspektraldaten erstellt und über eine Modelanpassung werden die aussagekräftigsten Prädiktoren zur Trennung von allen sechs Baumarten determiniert. Das Regressionsmodell mit der besten Performance zur Baumartendifferenzierung setzt sich aus Reflexions-Mittelwerten von sieben Kanälen des Hyperspektraldatensatzes zusammen. Mit diesem Modell können 100% der Referenzbäume der korrekten Zielklasse zugeordnet werden und bei der LOOCV-Kreuzvalidierung wird ein sehr geringer Anteil von 3,18% der Referenzbäume fehlklassifiziert. Zur statistischen Trennung aller sechs untersuchten Baumarten sind ausschließlich Hyperspektralvariablen signifikant und LiDAR-Variablen rufen keine Verbesserung des Modells hervor. Dies ist darauf zurückzuführen, dass durch die sehr hohe räumliche sowie spektrale Auflösung der Hyperspektraldaten das baumartenspezifische Reflexionsverhalten sehr gut quantifiziert werden kann und es schwierig ist, mit LiDAR-Variablen einzelne Nadelbaum- und Laubbaum-Spezies zu differenzieren.

Dennoch wird die grundsätzliche Eignung der aus dem nDSM abgeleiteten LiDAR-Variablen zur Baumartendifferenzierung über eine paarweise Trennung von Baumarten evaluiert. Mit einer Erstlergenauigkeit von 88% und einem Kreuzvalidierungsfehler von 10,15% liefert die Trennung von Buche/Lärche mit der Standardabweichung der Höhenwerte innerhalb eines Baumpolygons als Prädiktor das beste Ergebnis.

## **Abstract**

This study analyses the potential of the combination of high-resolution LiDAR- and hyperspectral remote sensing data for tree species classification. Therefore ground truth data for the tree species beech, oak, spruce, pine, larch and fir was acquired in the project area St. Margarethen an der Raab.

A multinomial logistic regression model was developed and the most significant predictors for tree species classification were determined through a model adjustment. The logistic regression model with the best tree species classification performance is based on seven bands of the hyperspectral data. With this model, 100% of the ground truth data could be classified in the correct target class. The low cross-validation error of 3,18% approves the excellent classification output.

Furthermore the significance of variables derived from the LiDAR dataset was evaluated through a pairwise comparison. With an accuracy of 88% and an cross validation error of 10,15% the separation of beech and larch reaches the best result.

# Inhaltsverzeichnis

<b>1. Einleitung</b>	7
1.1. Motivation	7
1.2. Problemstellung und Ziel der Arbeit	9
1.3. Struktur der Arbeit	11
<b>2. State of the Art</b>	12
2.1. Methoden der Baumartendifferenzierung	12
2.1.1. Logistische Regression	12
2.1.2. Random Forest	13
2.1.3. Support Vector Machines (SVM)	14
2.2. Kombination von LiDAR- und Hyperspektraldaten	14
<b>3. Theoretischer Hintergrund</b>	17
3.1. Waldinventur	17
3.2. Spektrales Reflexionsverhalten von Bäumen	17
3.3. Normalized Differenced Vegetation Index (NDVI)	19
3.4. Für die Arbeit relevante Aufnahmesysteme	20
3.4.1. Hyperspektralfernerkundung	20
3.4.2. LiDAR	22
<b>4. Untersuchungsgebiet und verwendete Daten</b>	24
4.1. Beschreibung des Projektgebiets	24
4.2. Untersuchte Baumarten	25
4.3. Verwendete Geodaten	26
4.3.1. Hyperspektraldatensatz Leica Aisa EAGLE	26
4.3.2. LiDAR-Datensatz Leica ALS50-II	27
<b>5. Verwendete Software und Methoden</b>	29
5.1. Verwendete Software	29
5.1.1. ArcGIS 10	29
5.1.2. Erdas Imagine 9.1	30
5.1.3. LASTools	30
5.1.4. Statistikpaket R	30
5.2. Verwendete Methoden	31
5.2.1. Lineare Regression	31
5.2.2. Logistische Regression	32
5.2.2.1. Grundlagen	32
5.2.2.2. Generalized Linear Models (GLM)	33
5.2.2.3. Model Selection	34
5.2.2.4. Multinomiale logistische Regression	36
<b>6. Referenzdatenerhebung</b>	37
6.1. Feldarbeit	37
6.2. Baumhöhen–Brusthöhendurchmesser Plot	39
6.3. Digitalisierung der Referenzbäume	40

<b>7. Datenaufbereitung</b> .....	42
7.1. Import der Streifen in Erdas Imagine .....	42
7.2. Mosaikierung .....	42
7.3. Radiometrische Datenvorverarbeitung .....	42
7.4. Streifenproblematik und geometrischer Versatz .....	44
7.5. Berechnung eines normalisierten digitalen Oberflächenmodells .....	46
<b>8. Ergebnisse</b> .....	47
8.1. Signaturanalyse.....	47
8.1.1. Umsetzung .....	47
8.1.2. Mean Plot aller Referenzbäume.....	47
8.1.3. Verteilung der Referenzbäume im spektralen Merkmalsraum .....	48
8.2. Verwendete LiDAR-Variablen.....	49
8.3. Baumartendifferenzierung .....	51
8.3.1. Multinomiales logistisches Regressionsmodell zur Trennung aller Baumarten.....	51
8.3.1.1. Ermittlung der Modellparameter .....	52
8.3.1.2. Modellparameter und Koeffizienten .....	53
8.3.1.3. Kontingenzmatrix.....	54
8.3.1.4. Kreuzvalidierung .....	55
8.3.1.5. Analyse der Modellanpassung .....	56
8.3.1.6. Conditional Density Plots des multinomialen logistischen Regressionsmodells.....	56
8.3.2. Generalized Linear Models(GLM) zur paarweisen Baumartendifferenzierung .....	60
8.4. Diskussion der Trennbarkeit der Baumarten .....	69
<b>9. Zusammenfassung</b> .....	71
<b>10. Literaturverzeichnis</b> .....	74
<b>11. Abbildungsverzeichnis</b> .....	77
<b>12. Tabellenverzeichnis</b> .....	78

## Abkürzungsverzeichnis

<i>AIC</i>	<i>Akaike Information Criterion</i>
<i>AOI</i>	<i>Area of interest</i>
<i>ASTER</i>	<i>Advanced Spaceborne Thermal Emission and Reflection</i>
<i>BFW</i>	<i>Bundesamt für Wald</i>
<i>BHD</i>	<i>Brusthöhendurchmesser</i>
<i>BIC</i>	<i>Bayesian Information Criterion</i>
<i>CIR</i>	<i>Color-infrared</i>
<i>DTM</i>	<i>Digital Terrain Model</i>
<i>DSM</i>	<i>Digital Surface Model</i>
<i>ETM+</i>	<i>Enhanced Thematic Mapper</i>
<i>FODIS</i>	<i>Fiber-Optic Downwelling Irradiance Sensor</i>
<i>GIS</i>	<i>Geographisches Informationssystem</i>
<i>GLM</i>	<i>Generalized Linear Model</i>
<i>GPS</i>	<i>Global Positioning System</i>
<i>GUI</i>	<i>Graphical User Interface</i>
<i>IFSAR</i>	<i>Polarimetric Interferometric Synthetic Aperture Radar</i>
<i>INS</i>	<i>Inertialnavigationssystem</i>
<i>LAI</i>	<i>Leaf Area Index</i>
<i>LiDAR</i>	<i>Light Detection and Ranging</i>
<i>LOOCV</i>	<i>Leave-One-Out Cross Validation</i>
<i>NASA</i>	<i>National Aeronautics and Space Administration</i>
<i>NDSM</i>	<i>Normalized Digital Surface Model</i>
<i>NDVI</i>	<i>Normalized Differenced Vegetation Index</i>
<i>NIR</i>	<i>Near-infrared</i>
<i>ÖWI</i>	<i>Österreichische Waldinventur</i>
<i>RADAR</i>	<i>Radio Detection and Ranging</i>
<i>RF</i>	<i>Random Forest</i>
<i>SAR</i>	<i>Synthetic Aperture Radar</i>
<i>SVM</i>	<i>Support Vector Machines</i>
<i>UTM</i>	<i>Universal Transverse Mercator</i>
<i>VNIR</i>	<i>Visible and near-infrared</i>
<i>WGS84</i>	<i>World Geodetic System 1984</i>

# 1. Einleitung

## 1.1. Motivation

Mit ca. 4 Millionen Hektar Fläche bedeckt Österreichs Wald fast die Hälfte des Bundesgebietes und ist nicht nur ein wichtiger Lebensraum für die Tierwelt, sondern auch Naherholungsraum und Rückzugsgebiet für den Menschen. Umfragen zufolge verbringen rund 70% der Menschen ihre Freizeit bevorzugt in Wäldern. Die Schutzfunktion des Waldes, zum Beispiel vor Lawinenabgängen, Muren und Steinschlag ist nicht zu vernachlässigen.

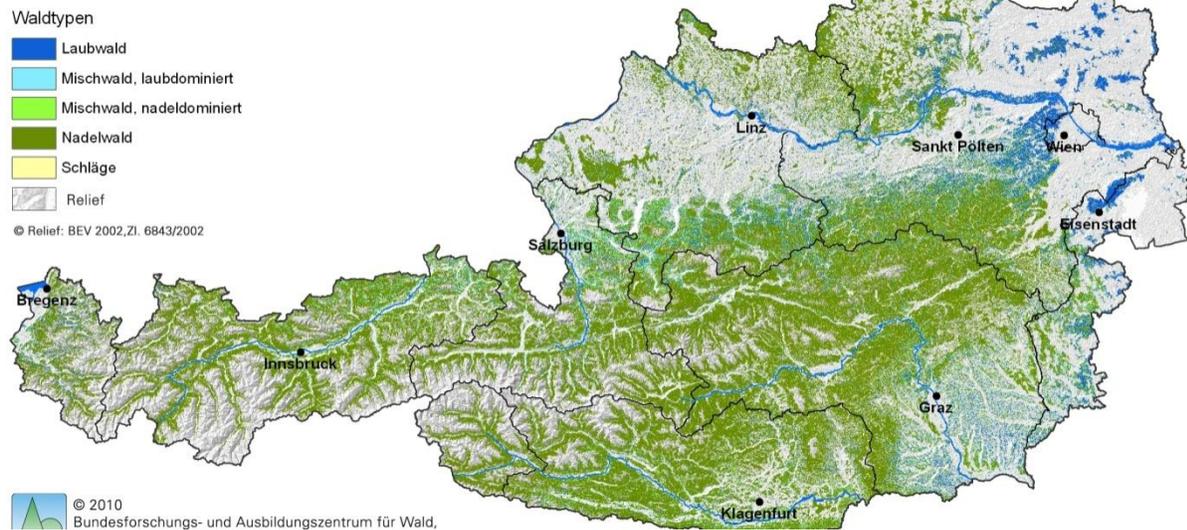
In Zeiten, wo Klimaerwärmung und CO<sub>2</sub>-Ausstoß in aller Munde sind, ist die ökologische Funktion des Waldes ein wichtiger Stabilisator für Klima und Umwelt. Der Wald als Sauerstofflieferant und Wasserspeicher ist überlebensnotwendig für alle Lebewesen und deswegen besonders schützenswert.

Auch der ökonomische Wert des Waldes, vor allem für ein walddreiches Land wie Österreich, ist nicht zu vernachlässigen. Der Großteil der Wertschöpfung basiert nach wie vor auf der Ressource Holz, der Wald als Kulturlandschaft hat jedoch auch einen enormen Wert für den Tourismus.

Um nachhaltige Forstwirtschaft betreiben zu können und das Ökosystem Wald zu verstehen, sind flächendeckende Bestandsaufnahmen von Waldgebieten unumgänglich. Waldinventuren sind nicht nur eine wichtige Entscheidungsgrundlage im Forstmanagement, sondern stellen auch eine wichtige Datengrundlage für ökologische Fragestellungen dar. So kann die Biodiversität eines Waldgebietes und deren Veränderung über einen bestimmten Zeitraum über hochauflösende Waldinventuren quantifiziert werden (IMMITZER et al. 2012).

Die österreichische Waldinventur (ÖWI) wird vom Waldforschungszentrum BFW durchgeführt und erhebt periodisch Daten zum Zustand des österreichischen Waldes. Das Hauptaugenmerk der ÖWI liegt darauf, die Anzahl der Bäume und deren räumliche Verteilung, die entsprechenden Baumarten, den Gesundheitsstatus sowie das Holzvolumen der Waldgebiete in Österreich zu erheben. Abbildung 1 zeigt ein Ergebnis der ÖWI, eine Karte der Waldtypen des Bundesgebietes.

## Waldtypenkarte Österreich



Die Unterscheidung von Baumarten ist eine fundamentale Aufgabe in der Forstwirtschaft. Diese Aufgabe wird hauptsächlich von Förstern in situ auf Basis von optischen Merkmalen durchgeführt. Für eine Waldinventur werden auf Stichprobenflächen Baumart, Brusthöhendurchmesser und Baumhöhe erhoben und über statistische Methoden die Verteilung für den gesamten Wald geschätzt. Diese Aufgabe erweist sich bei größeren Waldgebieten als schwierig, da der Aufwand für die Stichprobenerhebung im Feld enorm ist. Oft besteht das Problem, dass Waldgebiete schwer zugänglich sind und damit der Kostenaufwand und das Risiko für eine Waldinventur auf Basis von Feldbegehungen enorm sind. Über traditionelle Waldinventuren ist es nicht möglich, eine flächendeckende Erhebung von Baumarten auf Einzelbaumbene durchzuführen (IMMITZER et al. 2012). Solche Inventuren liefern jedoch keine flächenhaften Ergebnisse sondern Flächenstatistiken, folglich können daraus keine räumlichen Zusammenhänge abgeleitet werden.

Eine attraktive Alternative zu terrestrischen Waldinventuren ist die Fernerkundung, da sie zum einen eine bessere Abdeckung und kosteneffizientere Datenerhebung bietet, zum anderen eine höhere temporale Auflösung und damit hohe Wiederholungsraten gewährleistet. Ein weiterer Vorteil der Fernerkundung ist der synoptische Überblick über das Untersuchungsgebiet sowie die Möglichkeit, Daten über sehr große Gebiete mit einem hohen Detailgrad zu erheben (IMMITZER et al. 2012).

Die Fernerkundung bietet ein breites Spektrum an Methoden zur Durchführung einer Waldinventur. Die semiautomatische Unterscheidung von Nadelwald und Laubwald auf Basis hochauflösender Fernerkundungsdaten ist in der Forstwirtschaft bereits State of the Art (IMMITZER et al 2012, DALPONTE 2013).

Vor allem die Entwicklung von modernen Sensoren, welche über eine hohe räumliche als auch spektrale Auflösung verfügen, schafft neue Möglichkeiten der Baumartendifferenzierung.

Neueste Entwicklungen in der Fernerkundung ermöglichen die Erhebung der dreidimensionalen Struktur von Waldgebieten, zum Beispiel über das aktive Aufnahmesystem LiDAR. Aus LiDAR-Daten können Formparameter wie Baumhöhe und vertikaler Aufbau der Baumkrone auf Einzelbaumebene abgeleitet werden. Die Detektierung von Einzelbäumen auf „individual tree level“ stellt mit aktuellen LiDAR-Systemen auch aufgrund der hohen Abtastrate kein Problem mehr dar. Baumarten ausschließlich auf Basis von LiDAR-Daten zu differenzieren, gestaltet sich allerdings als schwierig, da eine hohe räumliche Heterogenität im vertikalen Aufbau von Einzelbäumen besteht (ZHANG & QIU 2011).

Hier setzt die Hyperspektralfernerkundung an. Durch die hohe Anzahl an Bändern kann das spektrale Reflexionsverhalten von Baumarten sehr gut quantifiziert werden. Aufgrund der hohen Kosten und der damit verbundenen limitierten Verfügbarkeit findet Hyperspektralfernerkundung für praktische Anwendungen in der Forstwirtschaft jedoch wenig Anklang (JAN et al 2008). Die Kombination von hochauflösenden Hyperspektral-Fernerkundungsdaten mit der aktiven Fernerkundungsmethode LiDAR ist ein vielversprechender Ansatz für eine moderne Waldinventur und wird in dieser Arbeit untersucht.

## **1.2. Problemstellung und Ziel der Arbeit**

Es gilt zu erörtern, ob auf Basis von hochauflösenden LiDAR- und Hyperspektraldaten, Baumarten statistisch getrennt und in weiterer Folge klassifiziert werden können. LiDAR- und Hyperspektraldaten sollten dabei gleichberechtigt in der Untersuchung behandelt werden. Dieser Ansatz zur Baumartendifferenzierung ist relativ neu und innovativ, eine Evaluierung der Performance ist ein wichtiger Punkt dieser Arbeit.

Zur Referenzdatenerhebung müssen mehrere Feldbegehungen im Projektgebiet St. Margarethen an der Raab durchgeführt werden. Um statistische Aussagekraft zu gewährleisten, sollten mindestens 30 Referenzbäume pro Baumart erhoben werden. Das Hauptaugenmerk wird auf Baumarten gelegt, welche am häufigsten im Projektgebiet vorkommen und für die Forstwirtschaft relevant sind.

Um Baumarten statistisch trennen zu können, ist es essentiell, Variablen aus dem LiDAR- und Hyperspektraldatensatz abzuleiten, welche zum einen die baumartenspezifische Form und den Aufbau der Baumkrone, zum anderen die spektrale Charakteristik von Baumarten wiedergeben.

Damit stellt sich die Frage, wie interoperabel die vorliegenden LiDAR- und Hyperspektraldaten, die simultan von einem Kleinflugzeug aufgenommen wurden, für das Anwendungsgebiet Baumartendifferenzierung sind. Können ausgewählte Baumarten über die Kombination aus Parametern, welche die Form der Baumkrone wiedergeben und Parametern, welche die spektrale Charakteristik einer Baumart beschreiben, statistisch getrennt werden?

Dazu sollte die Trennbarkeit der ausgewählten Baumarten über die Methode der logistischen Regression evaluiert werden. Des Weiteren sollte ermittelt werden, ob zwischen einzelnen Laub- und Nadelwaldspezies differenziert werden kann. Bei der Untersuchung der Hyperspektraldaten soll eruiert werden, inwieweit der Informationsgehalt der einzelnen Kanäle quantifiziert werden kann. Das heißt, die Eignung von ausgewählten Kanälen zur Baumtypendifferenzierung wird verglichen und statistisch beschrieben. Bei den LiDAR-Daten stellt sich die Frage, welche Parameter generell zur Beschreibung der charakteristischen Kronenform einer Baumart verwendet werden können. Können statistisch aussagekräftige Formparameter aus den LiDAR-Daten abgeleitet werden, so gilt es herauszufinden, ob damit eine paarweise statistische Trennung ausgewählter Baumarten möglich ist.

Ein weiteres Ziel dieser Arbeit ist es, die Eignung der Methode der logistischen Regression zur Baumartendifferenzierung zu untersuchen. Als Datengrundlage steht ein hochauflösender Hyperspektraldatensatz sowie LiDAR-Daten zur Verfügung. Für die sechs Zielklassen Buche, Eiche, Fichte, Kiefer, Tanne und Lärche wird ein logistisches Regressionsmodell mit sieben Prädiktoren ( $n+1$ ) gesucht.

### **1.3. Struktur der Arbeit**

In Kapitel 1 wird die Motivation für diese Arbeit geschildert und ein konkretes Projektziel definiert.

Kapitel 2 beschreibt den aktuellen Stand der Forschung in Form einer „State of the Art“ Analyse. Dieses Kapitel vergleicht Arbeiten zum Thema Baumartendifferenzierung auf Basis von Fernerkundungsdaten und beschreibt gängige Methoden. Außerdem werden Studien, welche auf die Kombination von LiDAR- und Hyperspektraldaten setzen, vorgestellt.

Kapitel 3 befasst sich mit dem theoretischen Hintergrund der Arbeit. Dieses Kapitel beschreibt fundamentale Konzepte aus der Fernerkundung, stellt die Methodik einer Waldinventur vor und veranschaulicht das spektrale Reflexionsverhalten von Bäumen. Ein weiterer Aspekt ist die Beschreibung der Aufnahmesysteme LiDAR und Hyperspektralsensor.

Kapitel 4 gibt einen Überblick über das Projektgebiet St. Margarethen an der Raab und beschreibt die verwendeten Fernerkundungsdaten.

Kapitel 5 befasst sich mit den verwendeten Softwarepaketen ArcGIS 10, Erdas Imagine 9.1, LASTools und der Statistikumgebung R. Ein weiterer Aspekt in diesem Kapitel ist die Vorstellung von statistischen Methoden, welche in dieser Studie Anwendung finden.

Kapitel 6 bespricht die Referenzdatenerhebung, das heißt die Vorgehensweise bei den Feldbegehungen und die anschließende Digitalisierung der Referenzbäume.

Die Datenaufbereitung wird in Kapitel 7 beschrieben. Die Umsetzung der Mosaikierung sowie des Histogrammausgleichs sind Inhalte, welche in diesem Kapitel behandelt werden. Darauf folgend wird die Erstellung des normalisierten Geländehöhenmodells nDSM erörtert.

Kapitel 8 veranschaulicht und diskutiert die Ergebnisse der Projektarbeit.

Zu guter letzt wird die Projektarbeit in Kapitel 9 zusammengefasst und ein Resümee gebildet.

## **2. State of the Art**

### **2.1. Methoden der Baumartendifferenzierung**

Die Unterscheidung von Baumarten ist eine essentielle Aufgabe bei der Analyse von Wald-Ökosystemen, nachfolgend werden gängige Methoden der Fernerkundung vorgestellt.

#### **2.1.1. Logistische Regression**

Der theoretische Hintergrund für die Methode der logistischen Regression wird in Kapitel 5.2.2 dargelegt.

Baumarten und Forstparameter über die Methode der logistischen Regression zu klassifizieren, wird in einer Vielzahl von Studien untersucht, zwei Anwendungsbeispiele werden in den folgenden Abschnitten vorgestellt.

IOST (2006) untersucht die Eignung logistischer Regressionsmodelle zur Klassifizierung der Forstparameter Waldfläche, Waldmischung und Totholzvorkommen auf Basis von Landsat TM, Landsat ETM+ und Quickbird Daten. Die Güte des Klassifikationsergebnisses ist abhängig vom Trainingsdatenumfang pro Merkmalsklasse. Eine unausgewogene Ausrichtung der Koeffizientenschätzer in der Modellanpassung resultiert in einer schlechteren Vorhersagegenauigkeit der unterrepräsentierten Klassen.

Das Merkmal Waldstatus mit den binären Ausprägungen „Wald“ und „Nicht-Wald“ konnte auf Basis der Landsat-Daten mit einer Modellgesamtgenauigkeit zwischen 85,1% und 95,4% klassifiziert werden. Als Modellprädiktor wird hierbei der Grauwert des Kanals 2 verwendet. Bei welchem Grauwert die Wahrscheinlichkeitsschwelle von 0,5 unterschritten wird und das Pixel somit als Nicht-Wald klassifiziert wird, ist ausschlaggebend für das Ergebnis.

Das Merkmal Waldmischung mit den Ausprägungen „Nadelwald“, „Laubwald“ und „Mischwald“ konnte auf Basis von Landsat ETM+ Daten mit Erstellergenauigkeiten von ca. 35% nicht zuverlässig über ein multinomiales logistisches Regressionsmodell getrennt werden.

PANT et al. (2014) entwickeln ein logistisches Regressionsmodell zur Selektion von Hyperspektralbändern zur Baumartendifferenzierung. Datengrundlage ist ein AisaEAGLE II Hyperspektraldatensatz mit 64 Kanälen und einem Wellenlängenbereich von 400-1000 nm. Auf Basis von 8-11 ausgewählten Bändern des Hyperspektraldatensatzes wird eine pixelbasierte „Support Vector Machines“-Klassifikation(SVM) über einen sehr kleinen Trainingsdatensatz von 0,5-1,5% des gesamten Datensatzes durchgeführt. Für die untersuchten Baumarten Kiefer, Fichte und Birke wird eine Gesamtgenauigkeit von 93,50% erzielt.

Um ein bestmögliches Ergebnis bei der Baumartendifferenzierung zu gewährleisten, wird der Hyperspektralsensor Leica AisaEAGLE II aufbauend auf einem besonders signifikanten Wellenlängenbereich über ein simuliertes Sensormodell neu kalibriert. Das simulierte Sensormodell weist eine um 3-5% Prozentpunkte bessere Klassifikationsgenauigkeit als die Standard-Sensoreinstellung des Leica ADS40 Sensors auf.

### **2.1.2. Random Forest**

Das Verfahren Random Forest basiert auf mehreren verschiedenen, unkorrelierten Entscheidungsbäumen. Ein Anwendungsgebiet in der Fernerkundung ist die Klassifikation und die Regressionsanalyse. Im üblichen Klassifikationsszenario sollen über Beobachtungen für zwei oder mehr Zielklassen Regeln erstellt werden um neue Beobachtungen einer Zielklasse zuordnen zu können. Beim Random Forest Ansatz werden Regeln über Klassifikationsbäume aufgestellt, welche über rekursive Partitionen die homogene Regionen („nodes“) relativ zu den Klassenvariablen quantifizieren. Bei der Anpassung des Klassifikationsbaumes wird eine Optimierung durchgeführt um einen Knoten und einen entsprechenden Prädiktor so auszuwählen, dass die Homogenität der Untergruppen im untersuchten Datensatz maximal ist. Indikator dafür ist der sogenannte Gini-Index (BREIMAN et al. 1984). Der Klassifikationsbaum ist ausgewachsen, wenn weitere Aufteilungen den Gini-Index nicht mehr senken. Random Forest verknüpft mehrere solcher Klassifikationsbäume zu einem Datensatz und kombiniert die Vorhersagen aller Klassifikationsbäume (CUTLER et al. 2007).

Folgende Anwendungsbeispiele bauen auf die Random Forest-Methode zur Baumartendifferenzierung:

IMMITZER et al. (2012) wenden die Random Forest-Methode zur Klassifikation von Baumarten basierend auf WorldView-2 Daten an. Über eine Random Forest Klassifikation wird sowohl pixel- als auch objektbasiert die Trennung von zehn Baumarten eines Waldgebietes in Österreich untersucht. Die Random Forest Klassifikation wird über die spektralen Signaturen manuell eingezeichneter, sonnenbeschienener Kronensegmente durchgeführt.

Beim objektbasierten Ansatz werden die Pixelwerte der einzelnen Kanäle pro Baumpolygon gemittelt um den negativen Einfluss von Mischpixel, zum Beispiel aufgrund von Schatteneffekten, zu senken. Dadurch konnte bei der objektbasierten Klassifikation eine Gesamtgenauigkeit von 82% erzielt werden. Die pixelbasierte Klassifikation liefert mit 77% Gesamtgenauigkeit ein etwas schlechteres Ergebnis.

DALPONTE (2013) analysiert das Potential der hochauflösenden Hyperspektralsensoren HySpex-VNIR 1600 und HySpex-SWIR 320i zur Klassifikation von Baumarten in einem borealen Waldgebiet. Verglichen werden dabei die Methoden Random Forest (RF), Support Vector Machines (SVM), und Maximum Likelihood für die vier Zielklassen Fichte, Kiefer, Birke und andere Laubbäume. Zur Klassifikation verwendet werden zum einen alle spektralen Bänder, zum anderen eine ausgewählte Teilmenge.

Für die Baumarten Fichte und Kiefer konnte auf Basis des HySpex-VNIR 1600 Sensors eine Klassifikationsgenauigkeit von über 95% erzielt werden. Die Klassifikatoren SVM und RF liefern dabei keine signifikant unterschiedlichen Ergebnisse.

### **2.1.3. Support Vector Machines (SVM)**

Support Vector Machines (SVM) ist eine Methode, die von VAPNIK (1995) vorgestellt wird und zur Klassifikation von Fernerkundungsdaten angewendet wird. Dabei werden über eine Optimierung Ausprägungen im Referenzdatensatz („Support Vectors“) gesucht, über welche eine Trennfläche zwischen den Zielklassen im spektralen Merkmalsraum gebildet werden kann. Es werden nicht alle Referenzdaten verwendet, sondern nur Trainingsgebiete, die im spektralen Merkmalsraum nahe zur Trennfläche, über welche eine Klassifikationsfunktion gebildet wird, liegen (GUALTIERI & CROMP 1998).

## **2.2. Kombination von LiDAR- und Hyperspektraldaten**

Die Kombination von LiDAR-Daten und digitalen Luftbildern zur Ableitung von Waldparametern findet in folgenden beiden Untersuchungen Anwendung:

WASER et al. (2008) untersuchen in einer Studie das Potential der logistischen Regression zur Ableitung von Waldparametern wie Baumfläche und Baumart auf Basis von LiDAR-Daten und digitalen Luftbildern. Baumbestände werden dabei semiautomatisch über den LiDAR-Datensatz erhoben. Die Unterscheidung von Nadel- und Laubbäumen ergibt eine höhere Genauigkeit als eine Unterscheidung aller Hauptbaumarten im Projektgebiet. Insgesamt werden 20 Modell-Prädiktoren verwendet: 18 aus dem digitalen Luftbild und dem Leica ADS40 Datensatz abgeleitete Parameter (ganze Kanäle und ausgewählte Kanäle dividiert durch die Summe von drei zugehörigen Kanälen), der NDVI und ein geglätteter Leica ADS40 Kanal (blau). Aus den LiDAR-Daten abgeleitete Formparameter werden im logistischen Regressionsmodell nicht als Prädiktoren verwendet. Das beste Klassifikationsergebnis wird erzielt, wenn man Modellkoeffizienten beider Sensoren (Leica ADS40 und CIR Luftbild) kombiniert und als Input für das logistische Regressionsmodell verwendet.

KANTOLA et al. (2010) erörtern die Kombination von LiDAR-Daten und digitalen Luftbildern zur Klassifikation des Entlaubungsgrades von Einzelbäumen. Basierend auf aus den LiDAR-Daten abgeleiteten Formparametern und dem Luftbild werden die Methoden Logistische Regression, Random Forest (RF) und Most Similar Neighbor (MSN) angewendet und verglichen. Abhängig von der Methode können Klassifikationsgenauigkeiten zwischen 83,7% und 88,1% erzielt werden, am besten schnitt der Random Forest Klassifikator mit den Formparametern „Baumkrone als konvexe Hülle“ und „Prozentsatz der LiDAR>Returns unter 50% der Baumhöhe“ aus den LiDAR-Daten und Mittelwerten der drei Kanäle des digitalen Luftbilds (NIR) ab. Werden die beschriebenen Methoden nur auf Basis eines Datensatzes (entweder LiDAR oder Luftbild) angewendet, so sinkt die Klassifikationsgenauigkeit auf 80,7%-87,4%. Summa summarum wird gezeigt, dass die Kombination von LiDAR-Daten und digitalen Luftbildern zur Detektion der Nadelverluste von Einzelbäumen vielversprechende Ergebnisse liefert.

Der Ansatz LiDAR- und Hyperspektraldaten zur Klassifikation von Baumarten und zur Ableitung von Forstparametern zu kombinieren, ist relativ neu und Forschungsobjekt in folgenden Studien:

ZHANG & QIU (2011) analysieren die Kombination von LiDAR- und Hyperspektraldaten zur Klassifikation von Baumarten im urbanen Wald. Der LiDAR-Datensatz weist eine Punktdichte von 3,5 p/m<sup>2</sup> auf, die Hyperspektralaufnahmen wurden simultan über einen Leica AISA dual hyperspectral sensor (492 Kanäle, 1,6 m Bodenauflösung) erhoben. Methodisch wird ein neurales Netzwerk gebildet und damit eine Baumartenklassifikation auf „individual tree level“ durchgeführt. Die Einzelbaumdetektion wird dabei auf Basis der LiDAR-Daten durchgeführt, die Baumartentrennung basiert auf einem „Adaptive Gaussian Fuzzy Learning Vector Quantization“-Algorithmus, welcher die spektrale Variabilität der Hyperspektraldaten quantifiziert. Mit einer Gesamtgenauigkeit von 68,8% und einem Kappa-Wert von 0,66 konnte ein durchaus gutes Ergebnis für 40 Zielklassen erzielt werden. Es wird gezeigt, dass die Integration von LiDAR- und Hyperspektraldaten zur Waldinventur großes Potential für die Ablöse von traditionellen Feldstudien aufweist.

DALPONTE et al. (2012) untersuchen verschiedene Multisensor-Setups zur Klassifikation von sieben Baumarten in den Südalpen: Zum einen die Kombination von LiDAR mit Multispektralaufnahmen, zum anderen LiDAR mit hochauflösenden Hyperspektraldaten. Bei den LiDAR-Daten werden verschiedene Punktdichten (0,48 points/m<sup>2</sup> und 8,6 points/m<sup>2</sup>) und deren Einfluss auf das Klassifikationsergebnis verglichen. Verwendete Formparameter aus den LiDAR-Daten sind die Baumhöhe, die „Interquartile Range“, die Standardabweichung der Höhenwerte und verschiedene Baumhöhen-Perzentile. Bei der Untersuchung werden die beiden Methoden RF und SVM angewendet. Die Kombination von hochauflösenden

Hyperspektraldaten und LiDAR-Daten mit hoher Punktdichte (8,6 points/m<sup>2</sup>) liefert mit Kappa-Genauigkeiten von 76,5% das beste Ergebnis zur Trennung von Baumarten. Setups mit Multispektraldaten in Kombination mit LiDAR führen zu einer Minderung der Klassifikationsgenauigkeit von ca. 5-10 Prozentpunkten. LiDAR-Daten mit hoher Punktdichte verbessern das Klassifikationsergebnis und eignen sich besser zur Baumartendifferenzierung als Datensätze mit niedriger Punktdichte.

GOSH et al. (2014) vergleichen die Eignung von HyMap-Hyperspektraldaten mit 4 m und 8 m Bodenaufklärung mit satellitenbasierten Hyperion-Hyperspektralaufnahmen mit 30 m Bodenaufklärung zur Baumartendifferenzierung. Die jeweiligen Hyperspektraldatensätze werden dabei mit einem nDSM, abgeleitet aus LiDAR-Daten kombiniert. Angewendet werden die beiden Klassifikatoren Random Forest und Support Vector Machines mit sechs Kombinationen von Prädiktoren. Als Prädiktoren dienen der Reflexionswert von signifikanten Kanälen, ausgewählte Minimum Fraction Noise Bands, Vegetationsindices und die Baumhöhe. Die Baumhöhe, abgeleitet aus dem nDSM, liefert keinen signifikanten Beitrag zur Baumartendifferenzierung und wird als Prädiktor verworfen. Das beste Klassifikationsergebnis wird mit den HyMap-Daten mit 8 m Bodenaufklärung erzielt (kappa-Wert von über 0,83). Es wird jedoch aufgezeigt, dass es keine eindeutig beste Kombination von Klassifikatoren (RF und SVM), Prädiktoren und Auflösung der Hyperspektraldaten gibt, da mehrere Ansätze vergleichbare Ergebnisse erzielen.

PUTTONEN et al. (2011) testen die Kombination von LiDAR- und Hyperspektraldaten, welche simultan mit dem finnischen Sensei-System aufgenommen werden, zur Klassifikation von Baumarten in einer Parkanlage in Finnland. Als LiDAR-Variable wird unter anderem der „Height coefficient of variation“, die Standardabweichung der LiDAR-Punktwolke eines Baumpolygons dividiert durch die mittlere Baumhöhe, getestet und liefert eine gute Performance zur Baumartendifferenzierung. Laub- und Nadelbäume konnten durch die Kombination von Hyperspektral- und LiDAR-Variablen mit 95,8%, einzelne Baumarten mit 83,5% Klassifikationsgenauigkeit getrennt werden. Die Kombination von LiDAR- und Hyperspektraldaten liefert beträchtlich bessere Ergebnisse als eine Referenzklassifikation mit Einzelsensor-Setups.

### **3. Theoretischer Hintergrund**

Im folgenden Teil wird der theoretische Hintergrund der Arbeit erörtert, das heißt Begriffe beziehungsweise Konzepte, welche in der Projektarbeit eine wichtige Rolle spielen, werden diskutiert.

#### **3.1. Waldinventur**

Unter einer Waldinventur versteht man die periodische Erhebung des Waldzustandes. Ziel einer Waldinventur ist ein ökologisch und ökonomisch orientiertes Waldmonitoring, welches Parameter wie Biodiversität, Nachhaltigkeit und Biomasse eruiert. Waldinventuren werden als Entscheidungsgrundlage bei verschiedensten Problemen wie Naturgefahren, Ressourcenmanagement und Forstpolitik berücksichtigt (HAUK & SCHADAUER 2009).

#### **3.2. Spektrales Reflexionsverhalten von Bäumen**

Für die Untersuchung von Baumarten mittels Fernerkundung müssen die spektralen Eigenschaften der untersuchten Baumarten beleuchtet werden. Elektromagnetische Strahlung sowie das elektromagnetische Spektrum werden in Abschnitt 3.4.1 erörtert.

Grundsätzlich müssen Reflexions-, Absorptions- und Transmissionseigenschaften von Vegetation berücksichtigt werden. Faktoren, welche die Reflektion des Waldes in den jeweiligen Spektralbereichen beeinflussen sowie die Form des reflektierten Signals müssen bestimmt werden, um quantitative Beziehungen zwischen Fernerkundungsdaten und variierenden Waldeigenschaften aufstellen zu können (MIRGORODSKY 2005).

Elektromagnetische Strahlung wird von Blatt- und Nadeloberflächen reflektiert, transmittiert oder absorbiert, die optischen Eigenschaften können dabei als Funktion von Blattstruktur, Wassergehalt und biochemische Zusammensetzung modelliert werden. Der physiologische Zustand der Blattoberfläche beeinflusst die Reflexions- und Streuungseigenschaften von Vegetation maßgeblich. Die Blattmorphologie ist für die Ausbreitung des Lichts innerhalb des Blattes verantwortlich, die Diffusion im Blatt wird dabei von homogenen Strukturen des Mesophylls und deren interzelluläre Lufträume gesteuert (KNEUBÜHLER 2002).

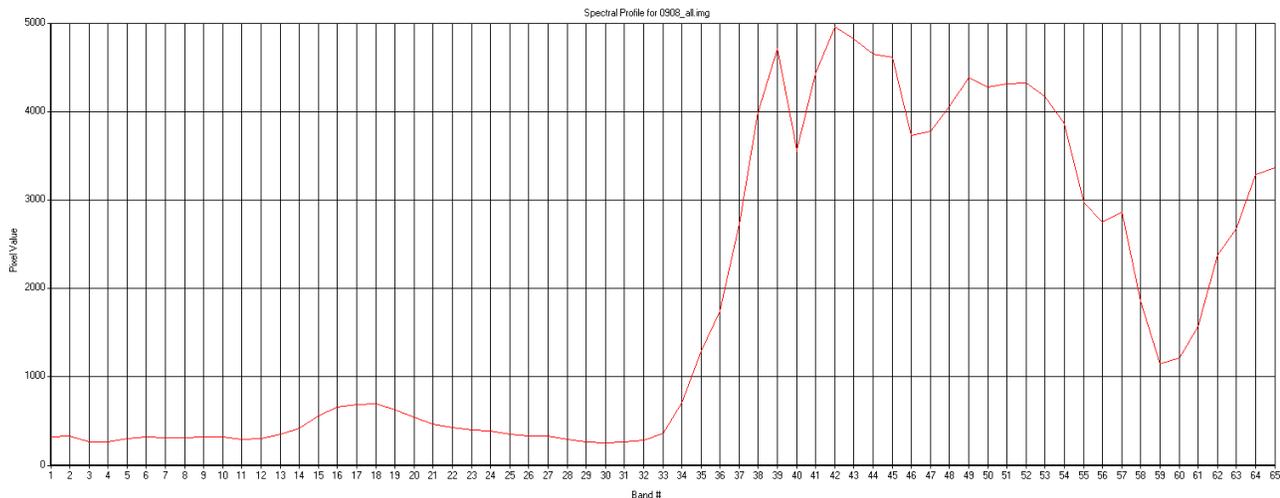


Abbildung 2: *Spektrales Reflexionsverhalten von Vegetation*

Abbildung 2 zeigt das Reflexionsverhalten von Vegetation in Form einer Kurve. Die Y-Achse ist der Grauwert, die X-Achse bildet die Kanäle des in dieser Studie verwendeten Hyperspektraldatensatzes ab.

Die Spektralregionen zwischen  $0,4 \mu\text{m}$  und  $2,5 \mu\text{m}$  des elektromagnetischen Spektrums sind für die optischen Eigenschaften von Vegetation ausschlaggebend und werden näher erläutert:

#### ❖ **Sichtbarer Bereich ( $0,4\text{-}0,7 \mu\text{m}$ )**

Der sichtbare Bereich des elektromagnetischen Spektrums wird hauptsächlich durch Chlorophyll und Carotenoide absorbiert, die Maxima der Absorption liegen bei  $0,43 \mu\text{m}$  (Kanal 3 in Abbildung 2) und  $0,65 \mu\text{m}$  (Kanal 29 in Abbildung 2). Aufgrund der hohen Absorption des Lichtes ist der Grauwert in diesen Bereichen niedrig.

Ein Anstieg der Reflexion ist im sogenannten „Red edge“-Bereich des Vegetationsspektrums zwischen  $0,67 \mu\text{m}$  und  $0,78 \mu\text{m}$  (Kanal 31 bis Kanal 43 in Abbildung 2) erkennbar (NILSON et al. 2003).

#### ❖ **Nahes Infrarot ( $0,7\text{-}1,3 \mu\text{m}$ )**

Beim Übergang vom Sichtbaren Bereich des Lichtes zum Nahen Infrarot, also bei ca.  $0,7 \mu\text{m}$ , kommt es zu einem signifikanten Anstieg der Reflexion. Dieser Anstieg der Reflexion ist in Abbildung 2 vom Kanal 33 aufwärts erkennbar. Die Absorption des Lichtes ist in diesem Wellenlängenbereich gering, da die einfallende Strahlung die Blattstruktur tief durchdringt. Hauptverantwortlich dafür ist die interne Blattstruktur, dieser Bereich wird jedoch auch von Wasserversorgung und Vitalität der Vegetation beeinflusst und kann dadurch zur Erkennung von Pflanzenschädigungen und Stresssituationen verwendet werden (MIRGORODSKY 2005).

### ❖ Mittleres Infrarot (1,3-2,5 µm)

Auch in diesem Wellenlängenbereich bestimmen Wassergehalt und die biochemische Komponenten der Pflanze die Strahlungseigenschaften. Die Hauptabsorption des Wassers findet bei 1,45 µm, 1,9 µm und 2,5 µm statt (ALBERTZ 2001), die Reflexion und Transmission des Lichtes wird durch eine Abnahme des Wassergehalts in der Pflanze erhöht.

Um die spektralen Eigenschaften von Bäumen ganz zu erfassen, sollte nicht nur die Reflexionscharakteristik des Vegetationsdaches betrachtet werden. Strukturelle Eigenschaften der Vegetationsdecke, optische Eigenschaften der Äste und des Stammes sowie der Einfallswinkel der Sonnenstrahlung müssen berücksichtigt werden (USTIN et al, 1999). Die optischen Eigenschaften von Vegetation in Hinblick auf Reflexion werden nach KNEUBÜHLER (2002) mit folgender Funktion beschrieben:

$$P_{(\text{canopy})} = f(\text{Geometry, Structure, Biochemistry, Geochemistry})$$

Der Parameter „Geometry“ umfasst dabei die Aufnahmegeometrie zwischen Sonne, Oberfläche und Beobachter sowie den Einfallswinkel der Strahlung von Sonne und Sensor. „Struktur“ beschreibt die Architektur der Vegetation und das Bedeckungsmaterial. Parameter wie Baumart, Baumhöhe und Blattflächenindex (LAI) fallen in diese Kategorie.

### 3.3. Normalized Differenced Vegetation Index (NDVI)

Für Fragestellungen in der Forstwirtschaft ist in der Fernerkundung der Vegetationsindex NDVI nicht wegzudenken und wird deswegen näher beleuchtet. Der NDVI beruht auf der Tatsache, dass es bei gesunder Vegetation bei 0,7 µm (Übergang vom sichtbaren Licht ins Nahe Infrarot) zu einem sprunghaften Anstieg der Reflexion kommt. Unbewachsener Boden weist in diesem Wellenlängenbereich einen eher geradlinigen Verlauf der Reflexionskurve auf. Grundsätzlich korreliert der Anteil des Chlorophylls mit der Steigung der Reflexion im nahen Infrarot zwischen 0,78 µm und 1 µm. Dadurch lässt sich nicht nur Vegetation von Nicht-Vegetation trennen, sondern es kann auch die Stärke der Vegetation modelliert werden. Der NDVI wird über folgende Formel berechnet:

$$\text{NDVI} = \text{NIR} - \text{Rot} / \text{NIR} + \text{Rot}$$

Der NDVI nimmt über eine Normierung nur Werte zwischen -1 und +1 an, wobei +1 gesunde und -1 erkrankte Vegetation repräsentiert (ALBERTZ 2001). Der Großteil der auftretenden Strahlung wird im nahen Infrarot vom Blattgewebe reflektiert, während das rote Licht weitgehend durch in den Blättern enthaltene Pigmente absorbiert wird. Die photosynthetische Aktivität sowie Dichte und Vitalität korrelieren stark mit dem NDVI und können damit mit Methoden der Fernerkundung erfasst werden.

### 3.4. Für die Arbeit relevante Aufnahmesysteme

#### 3.4.1. Hyperspektralfernerkundung

Hyperspektrale Fernerkundungssysteme sind eine Entwicklung der Fernerkundung, welche mit dem Aufkommen hochauflösender Sensoren einhergeht. Prinzipiell wird bei der Hyperspektralfernerkundung das elektromagnetische Spektrum über sehr viele, relativ enge Kanäle mit einer Bandbreite von ca. 1-10 nm abgebildet. Das elektromagnetische Spektrum (Abbildung 3) beschreibt die Frequenz- und Wellenlängenbereiche elektromagnetischer Strahlung. Wichtig für die Fernerkundung in forstwirtschaftlichen Anwendungen sind dabei das sichtbare Licht, der Infrarotbereich, sowie der Radarbereich. Passive Fernerkundungssensoren nehmen zum Beispiel elektromagnetische Strahlung im sichtbaren Bereich und im Infrarotbereich auf, aktive Sensoren senden selbst ein „Mess-Signal“ über Laser- oder Radarwellen aus (ALBERTZ 2001).

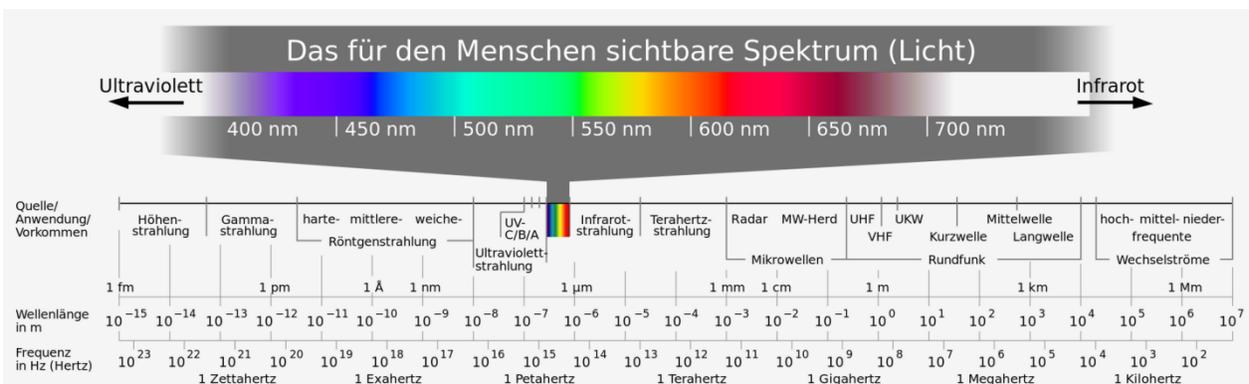


Abbildung 3: Das elektromagnetische Spektrum (©Wikipedia, 2013)

Hyperspektralsensoren sind in der Lage, Bilder über sehr viele Kanäle vom sichtbaren Bereich des Lichtes über das nahe Infrarot bis ins mittlere Infrarot aufzunehmen. Der große Wellenlängenbereich von 400-2500 nm und die damit verbundene hohe spektrale Auflösung erlaubt eine genaue quantitative Erfassung von spezifischen Spektralkurven (JAN et al. 2008).

Abbildung 4 vergleicht die Spektralkurven einer Landsat ETM+ Multispektralaufnahme mit einem Hyperspektraldatensatz. Das elektromagnetische Spektrum wird beim Hyperspektraldatensatz viel genauer abgebildet. Die Grafik links in Abbildung 4 veranschaulicht die spektrale Tiefe von Hyperspektraldaten anhand eines Kubus.

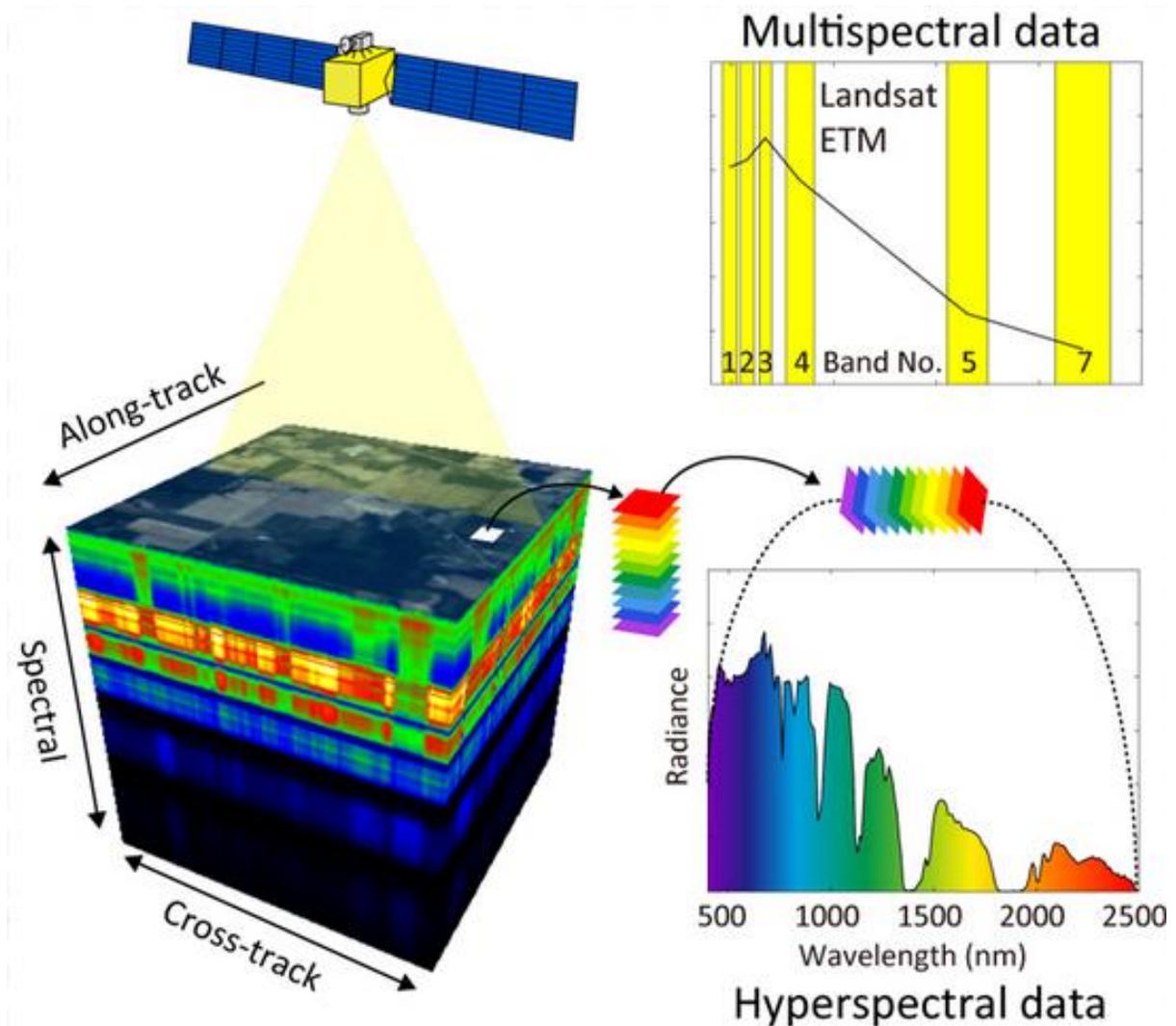


Abbildung 4: Das Prinzip der Hyperspektralfernerkundung(© <http://opticalnanofilter.com/2013/02/hyperspectral-imaging-in-der-fernerkundung/>)

### 3.4.2. LiDAR

Passive Aufnahmesysteme mit optischen Sensoren sind aufgrund ihrer Zweidimensionalität für gewisse Anwendungen in der Forstwirtschaft nur beschränkt geeignet. Nach MIRGORODSKY (2005) bietet „LiDAR-Fernerkundung für forstwirtschaftliche Anwendungen die Möglichkeit, dreidimensionale Waldstrukturen wie Vegetationshöhe, vertikale Verteilung des Pflanzenmaterials, Kronenvolumen, Biomasse und Baumdichte, zu charakterisieren“.

Aktive LiDAR-Fernerkundungssysteme bestimmen über Laserstrahlen die Distanz zwischen Aufnahmesensor und Objekt auf der Erdoberfläche. Dazu wird die Zeitspanne, welche der Laserstrahl vom Aufnahmesensor zum Objekt auf der Erdoberfläche, auf welcher der Laserstrahl reflektiert wird, und wieder zurück zum Sensor gemessen und in eine Distanz über „laser altimetry“ umgerechnet (LIM et al. 2003).

LiDAR-Systeme für terrestrische Anwendungen arbeiten in einem Wellenlängenbereich von 900-1064 nm, da Vegetation in diesem Bereich des elektromagnetischen Spektrums eine hohe Reflexion aufweist (LEFSKY et al. 2002). LiDAR-Aufnahmeplattformen verfügen über eine GPS-Einheit zur Ermittlung der genauen Position der Einheit sowie ein Inertialnavigationssystem (INS) zur Orientierung und Steuerung des LiDAR Sensors (HYPPÄÄ et al. 2004).

Abbildung 5 veranschaulicht das Prinzip von LiDAR-Laserscanning.

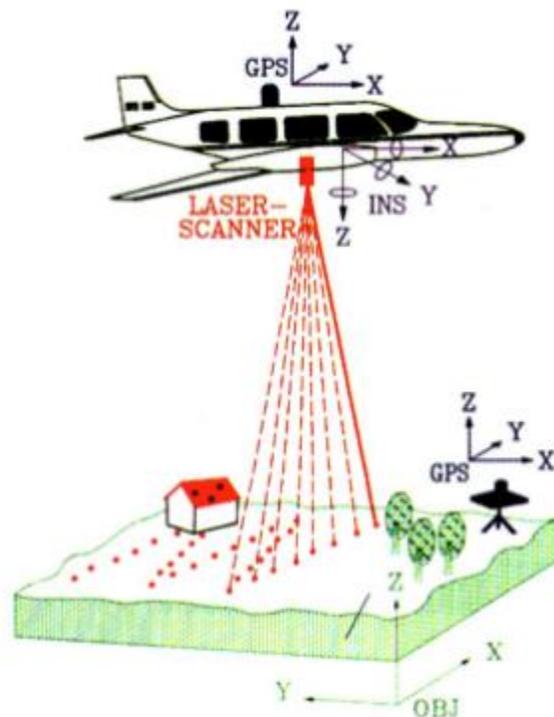


Abbildung 5: Schematisches Prinzip von LiDAR

(© <http://www2.geog.ucl.ac.uk/~plewis/lidarforvegetation/lidarRS.pdf>)

Grundsätzlich unterscheidet man zwischen zwei LiDAR-Aufnahmemethoden, „large-footprint“-LiDAR und „small-footprint“-LiDAR:

„Large-footprint“-LiDAR-Sensorsysteme arbeiten mit einem Aufnahmedurchmesser von 10–25 m und werden im Anwendungsbereich Vegetationsanalyse *„hauptsächlich zur Ermittlung spezifischer Vegetationsparameter wie Vegetationshöhe, vertikale Vegetationsstruktur und Biomasse herangezogen (MIRGORODSKY 2005)“*. Die Vegetationshöhe wird über die Differenz des ersten empfangenen LiDAR-Signals (Spitze der Vegetation) und aus dem Mittelpunkt der letzten LiDAR-Signale (Bodenoberfläche) errechnet.

Mittels „small-footprint“-LiDAR wird die Oberfläche nicht kontinuierlich abgebildet sondern mittels 1-5 Messwerten des reflektierten Signals mit einem Aufnahmedurchmesser von 1 m aufgenommen. „Small-footprint“-LiDAR Datensätze eignen sich zur hochauflösenden Kartierung von Geländeoberflächen (EVANS et al. 2001).

Technologisch unterscheidet man zwischen „Full waveform“-LiDAR und „first/last-pulse“-LiDAR. „Full waveform“-Aufnahmesysteme können das gesamte ausgestrahlte und wieder reflektierte Lasersignal aufnehmen. Damit ist es mit „Full waveform“-LiDAR möglich, Information über die innere Struktur der Vegetation zu erhalten. „First/last-pulse“-Systeme können hingegen nur den jeweils ersten und letzten Laserimpuls rekonstruieren (STILLA et al. 2009).

LiDAR-Daten haben eine hohe räumliche Auflösung von ~10-15 cm vertikal und ~100 cm horizontal. Weitere Vorteile sind die höhere Datendichte, schnelle Datenakquisition und Datenverarbeitung, relative Unabhängigkeit von der Wetterlage, sowie die Möglichkeit, Daten unter der Vegetationsdecke zu erheben, da die Laserstrahlen diese durchdringen (BEHERA & ROY 2002).

## 4. Untersuchungsgebiet und verwendete Daten

### 4.1. Beschreibung des Projektgebiets

Das Projektgebiet umfasst Waldgebiete westlich und südwestlich von St. Margarethen an der Raab, einer Marktgemeinde im Bezirk Weiz in der Oststeiermark. St. Margarethen an der Raab liegt in einem breiten Flachmuldental des Flusses Raab. Das Untersuchungsgebiet befindet sich in einer sanften Hügellandschaft mit einer Seehöhe von 319-499 m. Dieses Projektgebiet wurde deswegen ausgewählt, weil dafür Hyperspektral- und LiDAR-Daten verfügbar sind und viele verschiedene Baumarten in den Mischwäldern vorkommen.

Referenzbäume wurden in zwei sich überlappenden Streifen einer mosaikierten und kontrastverstärkten 7-Kanal Aufnahme erhoben, welche als Datengrundlage für die Referenzdatenerhebung verwendet wird. Abbildung 6 bietet einen Überblick über die Testgebiete (in Gelb) im Projektgebiet, welche im Großraum St. Margarethen an der Raab erhoben wurden. Gut ersichtlich sind die Übergänge der Streifen in der Aufnahme.

### Referenzbäume - St. Margarethen

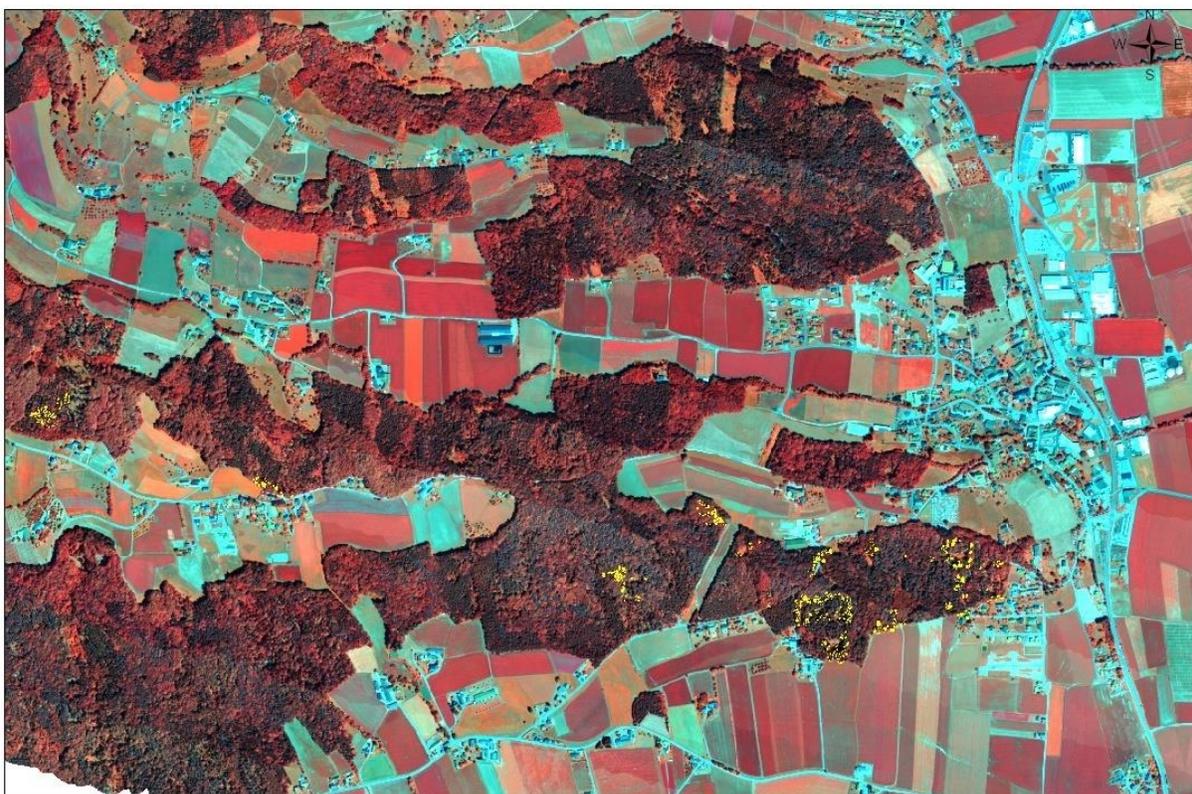


Abbildung 6: Überblickskarte des Projektgebietes St. Margarethen an der Raab

## 4.2. Untersuchte Baumarten

### ❖ Fichte (*Picea abis*)

Mit einem Anteil von 56% ist die Fichte die häufigste Baumart in Österreich. Der immergrüne Nadelbaum kommt natürlicherweise in einer Seehöhe von 600-800 m vor, durch künstliche Verbreitung wurde die Fichte jedoch auch in tieferen Lagen heimisch. Sie wird meist 30-50 m hoch, hat eine kegelförmige Krone und liebt lockeren, humosen Boden. In künstlich angelegten Fichten-Monokulturen herrscht ein großes Schadensrisiko für den Baumbestand, vor allem durch Borkenkäfer und Sturmschlag (SCHADAUER et al. 2003).

### ❖ Lärche (*Larix decidua*)

Die Lärche ist eine bis zu 40 m hohe Nadelbaumart, welche zur Familie der Kieferngewächse zählt. Die Baumkrone ist kegelförmig mit dichter Beastung, ältere Äste sind herabhängend mit aufsteigenden Spitzen. Lärchen kommen in Österreich von Tieflagen wie dem Wienerwald bis ins Hochgebirge über 2000 m vor und werfen als einzige heimische Nadelbaumart im Winter ihre Nadeln ab. Der Anteil der Lärchen am österreichischen Wald liegt bei 4%, damit ist die Lärche die dritthäufigste Nadelbaumart nach Fichte und Kiefer. Lärchenholz wird aufgrund seiner Härte vor allem als Bau- und Möbelholz geschätzt (SCHADAUER et al. 2003).

### ❖ Kiefer (*Pinus sylvestris*)

Die Weißkiefer ist mit einem Anteil von 6% die zweithäufigste Nadelbaumart in Österreich. Sie ist oft bestandsbildend und kommt am Alpenrand bis ca. 1300 m, in den Zentralalpen bis 2000 m vor. Weißkiefer sind sehr anpassungsfähig, gegen Frost und Hitze unempfindlich und kommen deswegen in ganz Europa vor (SCHADAUER et al. 2003).

### ❖ Tanne (*Abies alba*)

Die Weißtanne ist ein immergrüner Nadelbaum, welcher bis zu 50 m hoch und in Naturwäldern bis zu 500 Jahre alt werden kann. Mit einem Anteil von 2% am österreichischen Wald ist die Tanne eine wichtige bestandsbildende Nadelbaumart. Die Kronenform ist dadurch charakterisiert, dass Jungbäume zunächst kegelförmig sind und mit dem Alter der Baum immer abgerundeter wird. Weißtannen kommen in Nadel- und Laub-Nadelmischwäldern in ganz Mitteleuropa vor und sind gegen Luftverschmutzung sehr empfindlich (SCHADAUER et al. 2003).

#### ❖ **Buche (*Fagus sylvatica*)**

Die Rotbuche ist ein bis zu 35 m hoher sommergrüner Laubbaum und mit 9% Anteil die häufigste und für die Forstwirtschaft wichtigste Laubbaumart in Österreich. Die Kronenform bei Jungbäumen ist schlank, ausgewachsene Bäume sind kuppelförmig gewölbt und entwickeln sich vor allem im Freiland prächtig. Sie gilt als Laubbaum der Ebene und Mittelgebirge, bevorzugt frische nährstoffreiche Böden und kommt bis ca. 1500 m Seehöhe vor (SCHADAUER et al. 2003).

#### ❖ **Eiche (*Quercus robur*)**

Die Eiche ist mit einem Anteil von 2% die zweithäufigste Laubbaumart in Österreich und kann bis zu 45 m hoch werden. Die weit ausladende Krone ist breit und hoch und ist im Freiland meist kugelförmig gewölbt. Die Eiche kommt auf grundfeuchten Böden in weiten Teilen Europas, vor allem in den Flachländern, vor. Eichenholz ist hart und schwer und ist wegen seiner hohen Elastizität als Möbelholz geschätzt (SCHADAUER et al. 2003).

#### ❖ **Ahorn (*Acer pseudoplatanus*)**

Die Ahorn ist ein typischer Laubbaum mittlerer Berglagen und kommt vom Hügelland bis in subalpine Regionen vor. Sie wird bis zu 35 m hoch und hat eine hoch gewölbte, aber nicht sehr weit ausladende Krone. Ahornholz ist ein ausgesprochen wertvolles Edellaubholz, damit ist der Ahorn sowohl forstlich als auch für Parkanlagen von großer Bedeutung (SCHADAUER et al. 2003).

### **4.3. Verwendete Geodaten**

#### **4.3.1. Hyperspektraldatensatz Leica Aisa EAGLE**

Der vorliegende Hyperspektraldatensatz wurde mit dem Sensorsystem Leica Aisa EAGLE und mit einem Kleinflugzeug als Plattform aufgenommen. Der Aisa EAGLE Sensor basiert auf dem Pushbroom-Prinzip, das heißt, die Pixel einer gesamten Zeile werden gleichzeitig erfasst und durch Aneinanderfügen der einzelnen Zeilen entsteht eine flächenhafte Szene (JAN et al. 2008). Aufgebaut ist das Sensorsystem aus einem kompakten Hyperspektralsensor, einer Datenverarbeitungseinheit, einer GPS/INS-Einheit und einem FODIS-Bestrahlungssensor. Position und Flughöhe der Aufnahmeplattform wird dabei von GPS/INS-Einheit in Form einer drei-axialen Inertialnavigation aufgenommen.

Wie in Abbildung 7 ersichtlich liegt der Aufnahmebereich dieses Sensors zwischen 400-970 nm und deckt damit das sichtbare Licht sowie einen Teil des Nahen Infrarots ab.

Als „Spectral binning option“ wird mit „8x“ gearbeitet, das heißt, es stehen 65 Kanäle mit einer Bandbreite von 8,62 nm – 9,49 nm zur Verfügung. Bei einer Flughöhe von 385 m wird eine Bodenauflösung von 40 cm pro Pixel erreicht.

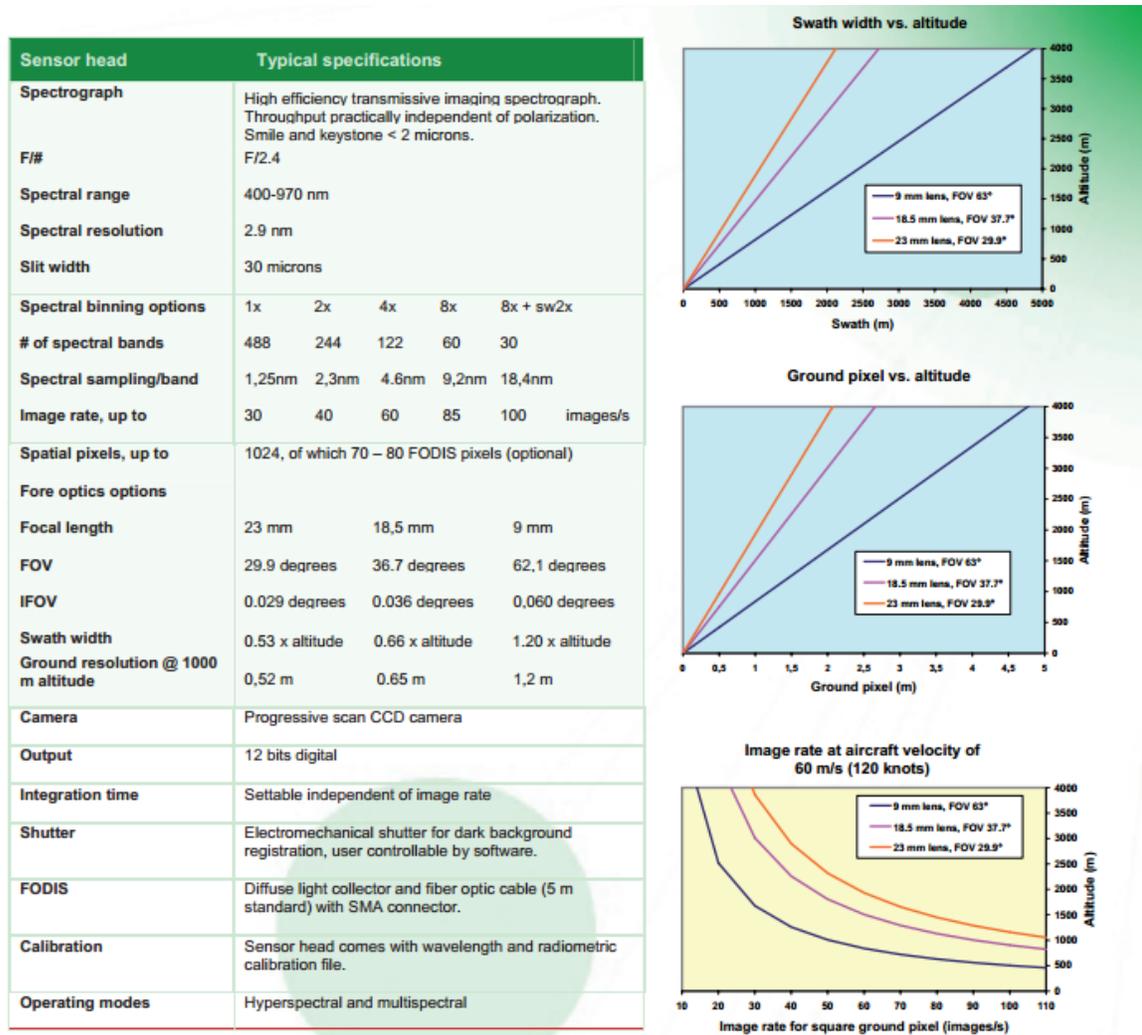


Abbildung 7: Sensorspezifikation Aisa EAGLE  
 (© [http://www.ufz.de/export/data/1/18870\\_aisa\\_Eagle\\_datasheet.pdf](http://www.ufz.de/export/data/1/18870_aisa_Eagle_datasheet.pdf))

#### 4.3.2. LiDAR-Datensatz Leica ALS50-II

Der Laserscanningdatensatz wurde im .las-Format über einen Leica ALS50 Sensor simultan mit den Hyperspektraldaten aufgenommen und beruht auf dem „first/last-pulse“-Prinzip. Laut der Sensorspezifikation sind Pulsraten von bis zu 150 kHz und Bodenauflösungen von bis zu 11 cm möglich. Die Punktdichte hängt von der Flughöhe ab und beläuft sich auf ~5 Punkte/m<sup>2</sup>. Daraus lässt sich das in Abbildung 8 dargestellte normalisierte digitale Oberflächenmodell (nDSM) mit 50 cm Bodenauflösung ableiten. Die Datenaufbereitung und nDSM-Erstellung erfolgt in der Software-Umgebung LAsTools und wird in Kapitel 7.5 erläutert.

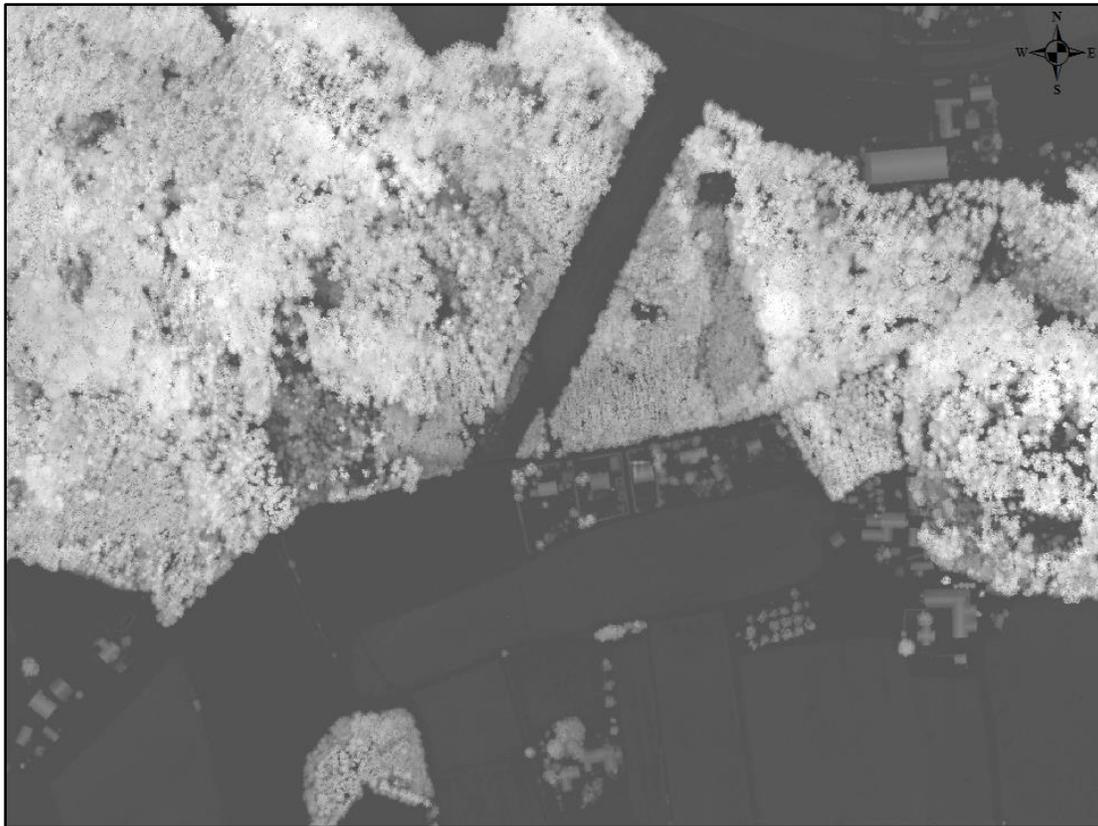


Abbildung 8: *Normalisiertes digitales Oberflächenmodell (nDSM)*

Zusätzlich zu den LiDAR-Daten wird ein digitales Geländehöhenmodell(DTM) zur nDSM-Berechnung (siehe Kapitel 7.5) verwendet. Dieser Datensatz stammt aus der gleichen Laserscanner-Befliegung und hat eine Bodenauflösung von 0,5 m. Abbildung 9 zeigt das digitale Geländehöhenmodell (DTM) des Projektgebietes.

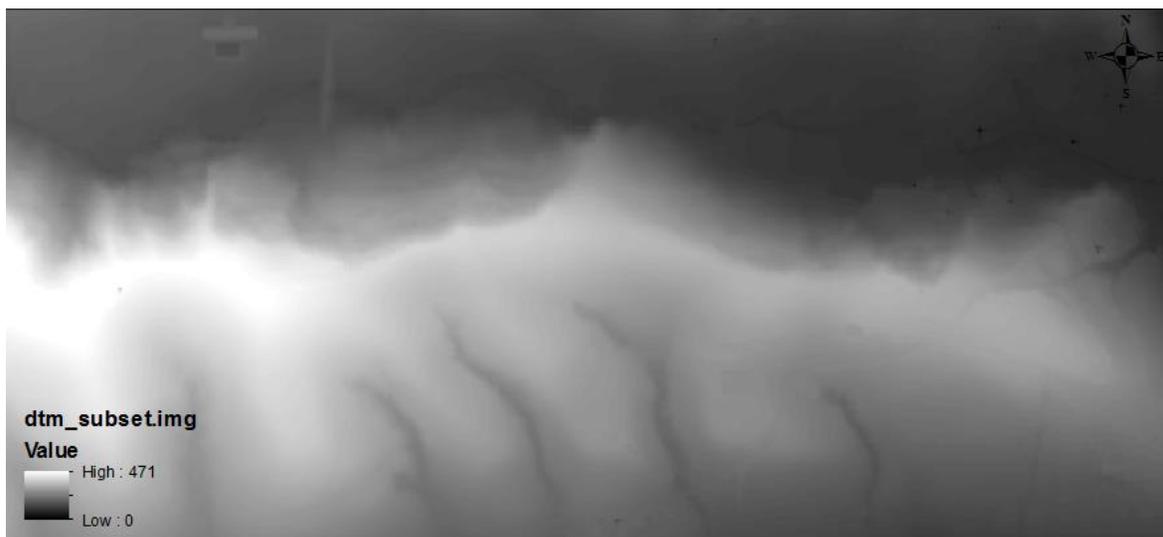


Abbildung 9: *Digitales Geländehöhenmodell (DTM)*

## 5. Verwendete Software und Methoden

### 5.1. Verwendete Software

#### 5.1.1. ArcGIS 10

ArcGIS 10 ist ein geographisches Informationssystem der Firma ESRI, welches eine Vielzahl von professionellen GIS-Anwendungen bietet. ArcGIS 10 eignet sich zur Kartenerstellung und Datenkompilierung und verfügt über eine große Anzahl von Methoden zur räumlichen Analyse von Geodaten, welche von GIS-Fachleuten über räumliche Operatoren genutzt werden können. Räumliche Datenbanken können über Geodatabases erstellt, verwaltet und manipuliert werden. Abbildung 10 zeigt die Infrastruktur des ArcGIS Systems. Geographische Informationen werden über die Client-Arten „Web“, „Mobile“ und „Desktop“ an die GIS-Community zur Verfügung gestellt. Verschiedenste GIS-Funktionen werden über folgende Services, welche die technische Grundlage für die Verwaltung und Organisation von freigegebenen geographischen Informationen bilden, gehostet (ArcGIS Resource Center, 2013, <http://resources.arcgis.com/de/help/getting-started/articles/026n00000014000000.htm>):

- ❖ Cloud Services: Funktionen, welche in einer Cloud veröffentlicht beziehungsweise freigegeben sind
- ❖ Enterprise Services: Als GIS-Web-Service zur Verwendung innerhalb eines Unternehmens
- ❖ Local Services: Über GIS-Ressourcen auf einem lokalen Computer

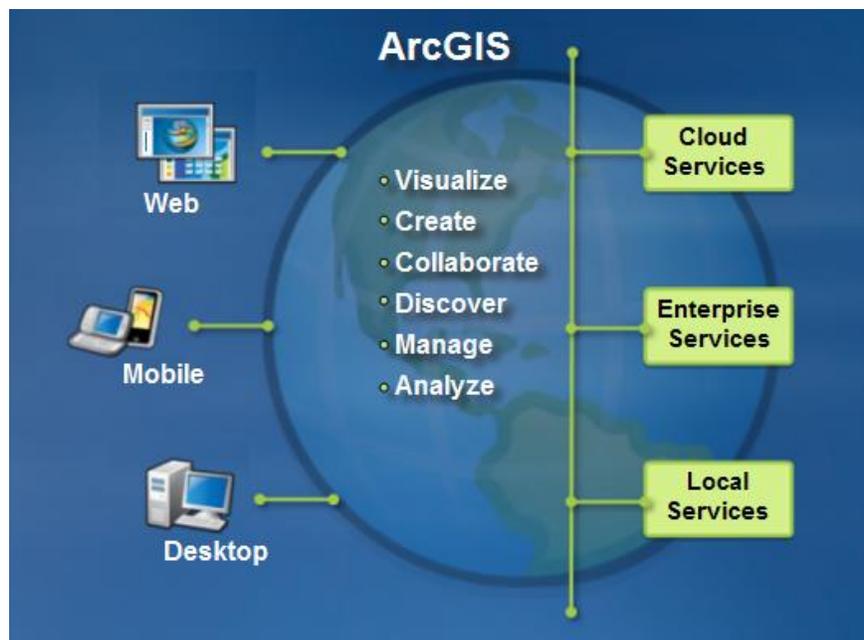


Abbildung 10: Komponenten des ArcGIS Informationssystems  
(© ArcGIS Resource Center 2013)

### **5.1.2. Erdas Imagine 9.1**

Erdas Imagine 9.1 ist eine Software Umgebung der Firma Intergraph, welche zur Datenvorverarbeitung, Analyse und Auswertung von rasterbasierten Bild- und Fernerkundungsdaten verwendet wird. Über geostatistische und geoanalytische Anwendungen können digitale Fernerkundungsdaten für ein GIS aufbereitet werden. Eine weitere Anwendung von Erdas Imagine 9.1 ist die Geokodierung, das heißt, Grafiken und Photoaufnahmen werden entzerrt und mit geographischen Koordinaten versehen. Über räumliche Rechenoperationen ist es möglich, pixelbasierte Klassifizierungen von Fernerkundungsdaten durchzuführen und in Form von Karten darzustellen.

(Erdas Imagine Field Guide, 2013

[http://geospatial.intergraph.com/Libraries/Tech\\_Docs/ERDAS\\_Field\\_Guide.sflb.ashx](http://geospatial.intergraph.com/Libraries/Tech_Docs/ERDAS_Field_Guide.sflb.ashx)  
)

### **5.1.3. LASTools**

LASTools ist ein Softwarepaket, welches zur Datenaufbereitung und Prozessierung von Laserscanning-Daten verwendet wird. Aus LiDAR-Rohdaten, aufgenommen im .las-Format, können geokodierte DTM's und DSM's generiert und in gängige GIS-Formate (ASCII, LAS, Terrascan, Shapefile, GeoTiff usw.) exportiert werden. Über ein GUI erfolgt der Zugriff auf verschiedenste Tools zur LiDAR-Prozessierung, wie zum Beispiel Modelanpassung und Systemkalibrierung des LiDAR-Sensorsystems. Um aus LiDAR-Rohdaten ein DTM oder DSM zu erhalten, erfolgt eine Klassifizierung der einzelnen Laserpunkte der reflektierten Oberfläche in drei Schritten:

1. Interaktive Parametrisierung
2. Automatische Klassifizierung
3. Interaktive Verfeinerung

LASTools verfügt über einen Data Viewer, welcher LiDAR-, Raster-, Volumes- und Vektordaten darstellen kann. Der objektorientierte und modulare Aufbau von LASTools gewährleistet Interoperabilität und Integration von neuen LiDAR-Technologien in der Zukunft (HUG et al. 2013).

### **5.1.4. Statistikpaket R**

R ist eine open-source Softwareumgebung für statistische Berechnungen und Visualisierungen, welche auf der statistischen Programmiersprache S basiert. Folgende Anwendungen, welche für eine statistische Analyse unumgänglich sind, können über ein GUI durchgeführt werden (ROSSITER 2010):

- effektive Datenmanipulation und Speicherung von Datensätzen
- eine Vielzahl von Operatoren für Matrizen- und Arrayberechnungen
- eine große Ansammlung von Tools zur statistischen Datenanalyse
- Visualisierung von statistischen Zusammenhängen
- eine hochentwickelte aber simple Programmiersprache mit Bedingungen, Schleifen und rekursiven Funktionen
- eine Vielzahl von statistischen Paketen für ein breit gefächertes Spektrum von statistischen Anwendungen

Die Erweiterung R Commander bietet zusätzliche Visualisierungsfunktionen und ein verbessertes GUI.

## **5.2.      Verwendete Methoden**

Zur Evaluierung von statistischen Zusammenhängen in der Fernerkundung gibt es eine Vielzahl an Methoden. Die gängigsten sind Regressionsanalysen, welche statistische Zusammenhänge über eine mathematische Funktion beschreiben. Für Regressionsanalysen wird meist die Methode der kleinsten Quadrate verwendet. Weitere Methoden sind die lineare Regression und die logistische Regression, beide Modelle verwenden als Schätzmethode das „Maximum-Likelihood“-Verfahren (IOST 2006).

In dieser Arbeit wird die Methode der logistischen Regression verwendet, zur Einführung in die Methodik wird auch die lineare Regression vorgestellt.

### **5.2.1.     Lineare Regression**

Die lineare Regression beruht auf dem Prinzip der kleinsten Abweichungsquadrate, das heißt, die Summe der quadrierten Abweichungen zwischen vorhergesagten und beobachteten Werten wird über die Regressionsgleichung minimiert. Das lineare Regressionsmodell ist einfach und intuitiv interpretierbar, zur sinnvollen Anwendung der linearen Regression sollten folgende Bedingungen erfüllt werden (IOST 2006):

- Linearität der Zusammenhänge
  - o es sollte eine additive Wirkung der Regressoren vorliegen
- Stetigkeit der Regressanden
  - o die kontinuierliche abhängige Variable sollte absolut- oder intervallskaliert sein
- Unabhängigkeit der Regressanden
  - o die Beobachtungen der abhängigen Variable sollten unabhängig voneinander sein

- Normalverteilung der Regressanden
  - o Normalverteilung der Beobachtungen der abhängigen Variable muss gegeben sein
  
- Homoskedasizität
  - o die Residuen sollten in etwa eine gleiche Varianz aufweisen und den Mittelwert  $\mu = 0$  haben
  
- Beobachtungsfehlerfreiheit der Regressoren
  - o die erklärenden Variablen der Regressionsgleichung sollten keine Messfehler aufweisen und keine Zufallsvariablen sein

## **5.2.2. Logistische Regression**

### **5.2.2.1. Grundlagen**

Nominal- und ordinal-skalierte Daten eignen sich nicht für ein lineares Regressionsmodell, hier setzt die logistische Regression an. Der grundsätzliche Anwendungsbereich der logistischen Regression gleicht der linearen Regression: *„Man hat eine abhängige Variable und möchte diese mittels einer oder mehrerer unabhängiger Variablen vorhersagen (HUPFELD 1999)“*.

Über eine Gewichtungsfunktion werden Prädiktoren, welche die abhängige Variable beeinflussen, determiniert. Die Mindestanforderung an Bedingungen für die Anwendungen einer logistischen Regression ist geringer als bei der linearen Regression, die Daten können nominal-, ordinal- und metrisch-skaliert sein (IOST 2006).

Die Stärke des Einflusses sowie die Zusatzbedingungen multivariate Normalverteilung, Homoskedasizität und die Unabhängigkeit der Beobachtungen müssen aber berücksichtigt werden um eine optimale Gewichtungsfunktion zu erstellen. Die unabhängigen linearen Einflüsse der einzelnen Prädiktoren lassen sich anhand der Regressionsgewichte bestimmen (HUPFELD 1999).

Bei der binären logistischen Regression hat die abhängige Variable genau zwei Ausprägungen, zum Beispiel „Wald“ oder „Nicht-Wald“. Hat die abhängige Variable mehr als zwei Ausprägungen, so spricht man von einem multinomialen logistischen Regressionsmodell. Als Beispiel für eine Variable mit mehr als zwei Ausprägungen wären verschiedene Baumarten (Fichte, Kiefer, Tanne oder Lärche?) zu nennen. Da bei der logistischen Regression eine Wahrscheinlichkeit modelliert wird, ist der Erwartungswert der abhängigen Variable auf Werte zwischen 0 und +1 beschränkt.

Die Kurvenform der logistischen Regression wird in Abbildung 11 am Beispiel Waldstatus veranschaulicht. Die Variable Waldstatus hat zwei Ausprägungen („Wald“ für alle Waldgebiete und „Nicht-Wald“ für alle anderen Flächen) und weist damit eine binäre Ausprägung auf.

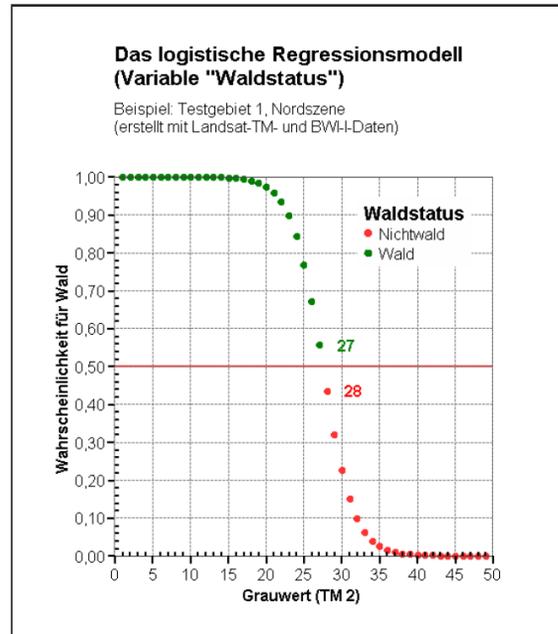


Abbildung 11: *Das logistische Regressionsmodell*  
(© IOST 2006)

Die Kurvenform der logistischen Regression ist so zu interpretieren, dass bis zu einem gewissen Schwellwert, in Abbildung 11 der Bereich um Grauwert 27-28, die Steigerung der Einflussgröße (Grauwert) auf die abhängige Variable (Wahrscheinlichkeit für Wald) wenig bewirkt. Bei der Überschreitung des Schwellwertes reagiert die abhängige Variable deutlich, außerhalb des Schwellwertbereiches ist die Veränderung wieder vernachlässigbar (IOST 2006).

#### 5.2.2.2. Generalized Linear Models (GLM)

Die logistische Regression gehört zur Familie der linearen Modelle (GLM), welche durch folgende drei typische Elemente charakterisiert sind (LANE 2002):

- zufällige Komponente
- systematische Komponente
- Verbindung zwischen diesen beiden Bestandteilen

Die zufällige Komponente steckt in der abhängigen Variable und muss eine exponentielle Wahrscheinlichkeitsverteilung aufweisen. Die erklärenden Variablen, welche als Prädiktoren für die abhängige Variable dienen, bilden die systematische Komponente eines GLM's. Die Verknüpfung zwischen dem Erwartungswert der abhängigen Variable und dem linearen Prädiktor wird über eine sogenannte Link-Funktion, welche von der Wahrscheinlichkeitsverteilung der Zielgröße determiniert wird, hergestellt. Neben der üblichen Normalverteilung sind auch andere exponentielle Verteilungen wie Poisson-, Gamma-, oder Binomialverteilungen bei der Aufstellung eines GLM's zulässig (LANE 2002).

$$\mathbf{I)} \quad P(Y = 1|GW) = \pi(x) = \frac{e^{\beta X}}{1 + e^{\beta X}}.$$

Formel I) zeigt die Grundformel der logistischen Regression am Beispiel Grauwert als Prädiktor für die bedingte Wahrscheinlichkeit, dass ein Pixel den Waldstatus „Wald oder „Nicht-Wald“ aufweist. Die Komponente  $\beta X$  ist dabei die Linearkombination aller zur Vorhersage herangezogenen Prädiktoren beziehungsweise Parameter. Die abhängige Variable Waldstatus nimmt Werte zwischen 0 und 1 an, wobei 0 für „Nicht-Wald“ und 1 für „Wald“ steht (IOST 2006). Die abhängige Variable Waldstatus ist damit binomialverteilt, da sie genau zwei mögliche Ausprägungen aufweist.

Zur Modelanpassung der logistischen Regression werden die Koeffizientenschätzer über das Maximum-Likelihood Verfahren eruiert. Dazu wird jene Parameter-Kombination ermittelt, welche *„das Auftreten der beobachteten Daten als Gesamtheit am Wahrscheinlichsten macht (IOST 2006)“*.

In statistischen Softwarepaketen erfolgt die Modelanpassung eines logistischen Regressionsmodells über ProgrammROUTINEN wie zum Beispiel die „Stepwise Model Selection“.

### **5.2.2.3. Model Selection**

Da logistische Regressionsmodelle immer mathematisch-abstrakte Abbildungen der Realität sind, welche in ihrer vollen Komplexität nie gänzlich erfasst werden kann, gibt es per Definition kein optimales Modell für eine konkrete Fragestellung. Das „beste“ Modell sollte eine abhängige Variable möglichst genau erklären und gut ausbalanciert sein (GERNAND 2009).

Die Regressionsparameter werden auf Basis des Maximum-Likelihood-Verfahrens geschätzt, zur Bewertung der Charakteristika eines logistischen Regressionsmodells kann das Akaike Informationskriterium (AIC) oder das Bayesschen Informationskriterium (BIC) herangezogen werden.

Während beim AIC Prädiktion und Validierung möglichst allgemeiner Aussagen im Vordergrund steht, ist beim BIC eine genaue Theorieprüfung die oberste Prämisse. Bei großem Stichprobenumfang wird meist der AIC verwendet, da das Modellselektionsproblem minimiert wird.

Über den AIC ausgewählte Modelle neigen zu „Overfitting“ und es wird kein „wahres“ Modell unterstellt, sondern nur das am besten „gefittete“ Modell. Über den BIC wird ein quasi-wahres Modell selektiert, bei welchem der Informationsverlust bei steigender Stichprobenanzahl sinkt. Aufgrund der Tatsache, dass die Prädiktoren von auf Basis des BIC bewerteten Modellen sehr sparsam ausgewählt werden, neigen diese Modelle zu „Underfitting“ und haben einen hohen Bias (GERNAND 2009).

Die drei am häufigsten genutzten automatisierten Methoden zur Auswahl der besten Prädiktoren für das logistische Regressionsmodell lauten wie folgt (AUSTIN 2004):

1. Backward elimination
2. Forward selection
3. Stepwise selection

Ad 1) Bei der „Backward elimination“ werden zuerst alle möglichen Modellprädiktoren beleuchtet und schlechte Variablen sequentiell eliminiert bis eine gewisse Abbruchbedingung erfüllt ist. Während des Eliminierungsprozesses wird die Variable, welche das Gesamtergebnis (Devianz oder  $R^2$ ) am wenigsten beeinflusst, eliminiert. Die häufigste Abbruchbedingung ist, dass alle Prädiktoren bis zu einem vordefinierten Level signifikant sind (AUSTIN 2004).

Ad 2) Eine „Forward selection“ startet mit dem leeren logistischen Regressionsmodell und Variablen werden sequentiell zum Modell hinzugefügt, bis eine Abbruchbedingung erfüllt ist. Während des Selektierungsprozesses wird immer die Variable mit dem signifikantesten Einfluss auf das Gesamtergebnis zum Modell hinzugefügt. Der Selektierungsprozess wird dann abgebrochen, wenn eine hinzugefügte Variable ein vordefiniertes Signifikanzniveau unterschreitet (AUSTIN 2004).

Ad 3) Bei der „Stepwise model selection“ handelt es sich um eine Abwandlung der „Forward selection“. Nachdem eine neue Variable hinzugefügt wird ist es möglich, dass Variablen, welche über einem vordefinierten Signifikanzniveau liegen, wieder aus dem logistischen Regressionsmodell entfernt werden. Dieser iterative Prozess wird beendet, wenn eine Abbruchbedingung erfüllt ist (AUSTIN 2004).

#### 5.2.2.4. Multinomiale logistische Regression

Die multinomiale logistische Regression gehört zur Gruppe der Generalized Linear Models (GLM) und wird angewendet, wenn die Zielvariable mehr als zwei kategorische Ausprägungen aufweist. Die abhängige Variable „Baumart“ hat sechs kategorische Ausprägungen, da in dieser Arbeit die Trennbarkeit von insgesamt sechs Baumarten untersucht wird. Deswegen wird die Methode der multinomialen logistischen Regression ausgewählt. Die mathematische Formulierung der multinomialen logistischen Regression gestaltet sich komplizierter als die allgemeine Formulierung der logistischen Regression in Formel 5.2.2.2.

Grundsätzlich geht es um die Wahrscheinlichkeit  $\pi_{ij}$ , also der Wahrscheinlichkeit, dass die Beobachtung  $i$  der  $j$ -en Klasse der Zielklassen  $j = 1 \dots m$  angehört. Die erste Zielklasse  $j = 1$  wird als Basisklasse definiert, die Basiswahrscheinlichkeit  $\pi_{i1}$  wird als Residuenwahrscheinlichkeit festgelegt nachdem die anderen Zielklassen  $\pi_{i2} \dots \pi_{im}$  modelliert wurden. Folglich hat das multinomiale logistische Regressionsmodell  $k + 1$  Koeffizienten für jede der  $j = m - 1$  Zielklassen: einen intercept-Koeffizienten  $\alpha_j$  und eine Steigung für jeden Prädiktor  $\beta_{lj}$  (ROSSITER 2012).

Formel II) zeigt die Grundformel eines multinomialen logistischen Regressionsmodells, welches über die Maximum-Likelihood-Methode angepasst wird (ROSSITER 2012).

$$\text{II)} \quad \pi_{ij} = \frac{e^{(\alpha_j + \beta_{1j}x_{i1} + \dots + \beta_{kj}x_{ik})}}{1 + \sum_{l=2}^m e^{(\alpha_l + \beta_{1l}x_{i1} + \dots + \beta_{kl}x_{ik})}}, j = 2, \dots, m$$

$$\pi_{i1} = 1 - \sum_{j=2}^m \pi_{ij}$$

Die Log-Odds einer Beobachtung für eine konkrete Zielklasse relativ zur Basisklasse werden über die angepassten  $\alpha$  und  $\beta$  berechnet. Man beachte, dass  $\pi_{ij}$  die Wahrscheinlichkeit einer Beobachtung für eine konkrete Klasse relativ zur Basisklasse abbildet (ROSSITER 2012).

## 6. Referenzdatenerhebung

### 6.1. Feldarbeit

Zusätzlich zu den vorliegenden Hyperspektral- und LiDAR-Daten müssen für eine Untersuchung der Baumarten Referenzbäume im Feld erhoben werden. Als Vorbereitung für die Feldarbeit wird im Zuge des Mission Plannings ein kontrastverstärktes Bild aus sieben zusammengeführten Kanälen der Hyperspektralaufnahme erstellt. Aufbauend aus Ausdrucken dieser Aufnahme werden besonders heterogene und interessant aufgebaute Waldgebiete für Feldbegehungen ausgewählt.

Die Erhebung der Referenzbäume im Feld ist eine essentielle Aufgabe um gute Ergebnisse zu erzielen. Es wurde versucht, die Ausprägungen der Individuen der untersuchten Baumarten möglichst gut abzudecken, das heißt, es müssen für eine Baumart alle Ausprägungen repräsentativ erfasst werden. An weiteren Merkmalen wurden das Baumalter, die Kronenform und Kronengröße sowie der Gesundheitsstatus erhoben.

Kontrastverstärkte Ausdrücke der Hyperspektralaufnahme von ausgewählten Gebieten mit markanter Verteilung der Baumarten sind die Datengrundlage der Referenzdatenerhebung. Ist ein Referenzbaum identifiziert, wird dieser im Ausdruck markiert, mit einem Index versehen und die Baumart sowie der Brusthöhendurchmesser erhoben. Die Zuweisung eines Baumes zu einem Objekt im Hyperspektralbild erfolgt über die Kronenform und deren Farbverteilung. Räumliche Orientierung wird durch die Einbeziehung von markanten Landmarken wie zum Beispiel Wegkreuzungen oder Waldlichtungen gewährleistet. Da dies bei einer Probe-Feldbegehung gut funktionierte, wurde auf die Verwendung eines GPS-Systems verzichtet. Die Baumart wird über die Nadel- beziehungsweise Laubform, die Rinde und das allgemeine äußere Erscheinungsbild bestimmt.

Insgesamt wurden im Projektgebiet St. Margarethen an der Raab in fünf Feldbegehungen 451 Referenzbäume erhoben. Die Testgebiete wurden so gewählt, dass sie über zwei Streifen des Hyperspektraldatensatzes verteilt sind. Die zwei Streifen liegen nebeneinander, überlappen sich teilweise und wurden deshalb verwendet, um bei Bedarf radiometrische Vergleiche durchführen zu können. Tabelle 1 zeigt eine Zusammenfassung aller erhobenen Referenzbäume in den beiden untersuchten Hyperspektraldatenstreifen „0908“ und „0909“.

**Tabelle 1: Anzahl der erhobenen Referenzbäume**

<b>Baumart</b>	<b>Streifen 0908</b>	<b>Streifen 0909</b>
Ahorn	14	2
Buche	56	39
Eiche	28	29
Fichte	50	25
Kiefer	46	41
Lärche	44	3
Tanne	43	31
<b>Summe:</b>	<b>281</b>	<b>170</b>

Wie man in Tabelle 1 erkennen kann, wurden im Streifen „0908“ 281 Referenzbäume und im Streifen „0909“ 170 Referenzbäume erhoben. Die anvisierte Sample Size von 30 Referenzbäumen pro untersuchte Baumart konnte bei der im Projektgebiet selten vorkommenden Ahorn in keinem der beiden Hyperspektralstreifen erreicht werden. Referenzbäume der Baumart Ahorn werden deswegen für die Baumartendifferenzierung über ein logistisches Regressionsmodell verworfen.

Viele Referenzbäume sind im Überlappungsbereich der beiden Hyperspektralstreifen erhoben worden. Es wurde angedacht, diese Referenzbäume für die statistische Datenanalyse beider Hyperspektralstreifen zu verwenden. Die im Überlappungsbereich erhobenen Referenzbäume können aufgrund eines nicht systematischen geometrischen Versatzes jedoch nur für den Streifen „0908“ verwendet werden. Diese Streifenproblematik und der geometrischen Versatz werden in Kapitel 7.4 genauer erläutert. Die nicht gegebene Verwendbarkeit der im Überlappungsbereich erhobenen Referenzbäume für Streifen „0909“ erklärt die im Vergleich zu Streifen „0908“ geringere Anzahl an Referenzbäumen im Streifen „0909“.

Die Baumartendifferenzierung wird aufgrund der größeren Zahl an verwendbaren Referenzbäumen für den Streifen „0908“ und die sechs Baumarten Buche, Eiche, Fichte, Kiefer, Lärche und Tanne durchgeführt.

## 6.2. Baumhöhen-Brusthöhendurchmesser Plot

Die Ausprägungen der Referenzbäume lassen sich im Baumhöhen-Brusthöhendurchmesser Plot in Abbildung 12 über den Zusammenhang aus Baumhöhe (MAXIMUM) und Brusthöhendurchmesser (BHD) pro Baumart ablesen. Die Punktwolken der Baumarten Buche, Eiche, Tanne und Fichte sind sehr gestreut, das heißt, es handelt sich um Referenzbäume mit unterschiedlichsten Ausprägungen. Dies ist für eine repräsentative Datenerhebung essentiell. Die Baumarten Kiefer und Lärche sind sehr geclustert. Dies ist darauf zurückzuführen, dass im Projektgebiet nur Individuen in einem bestimmten Höhen- und Altersbereich anzufinden sind.

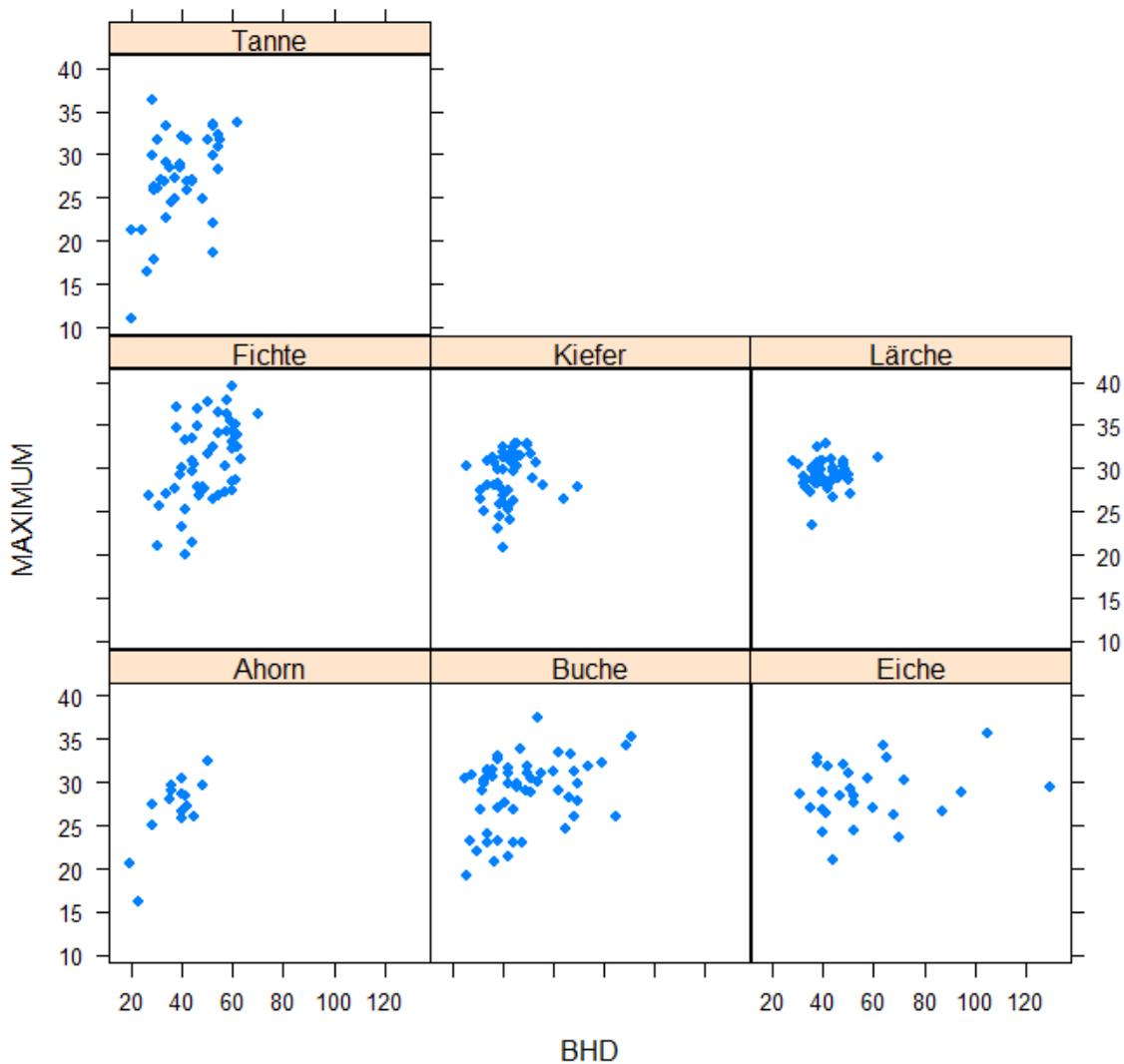


Abbildung 12: Baumhöhen-Brusthöhendurchmesser Plot

### 6.3. Digitalisierung der Referenzbäume

Die Digitalisierung der in den Ausdrucken eingezeichneten Referenzbäume erfordert besonderes Augenmerk. Mischpixel können klarerweise nicht vermieden werden, durch eine möglichst präzise Erfassung der Kronenform wird aber versucht, die Anzahl der Mischpixel möglichst gering zu halten.

Bei der Digitalisierung der Referenzbäume wird zusätzlich zur Hyperspektralaufnahme das nDSM zur Ermittlung der Kronenform verwendet. Abbildung 13 zeigt die kontrastverstärkte 7-Kanal-Aufnahme und einen Teil der erhobenen Referenzbäume.

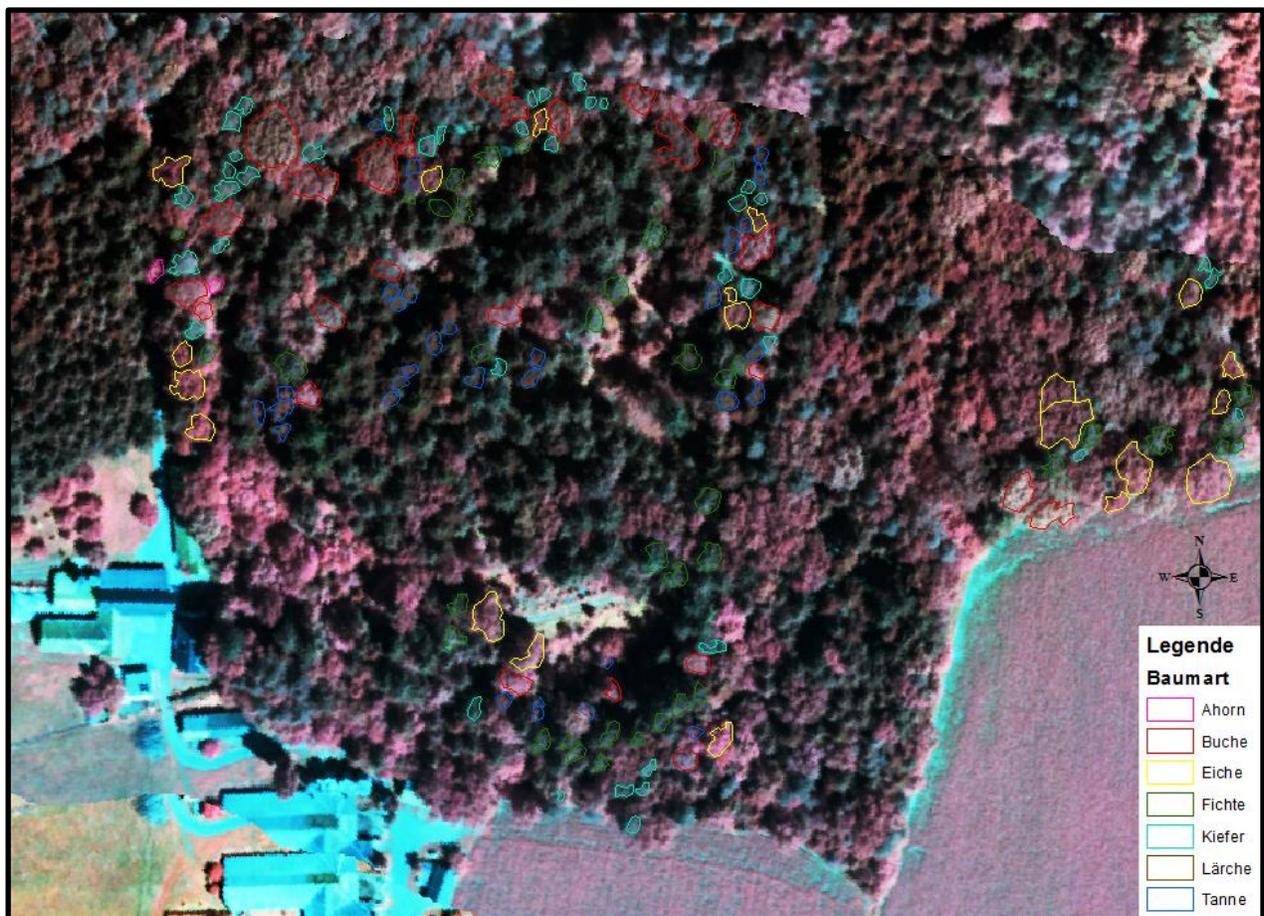


Abbildung 13: Referenzdatenerhebung

Man erkennt, dass eine genaue Abgrenzung der Individuen beziehungsweise eine exakte Digitalisierung der Kronenform nicht möglich ist. Besonders im dichten Laubwald ist eine Abgrenzung der Individuen de facto nicht realisierbar, da die Baumkronen sich überschneiden beziehungsweise ineinander verschachtelt sind. Dies kann die 3d-Form der untersuchten Individuen und damit die Analyse der

LiDAR-Daten negativ beeinflussen. Die digitale Kronenform ist demnach immer eine Abstraktion der physikalischen Kronenform eines konkreten Baumes.

Die Digitalisierung der im Feld erhobenen Referenzbäume erfolgt in ArcGIS 10. Als Datengrundlage dient die kontrastverstärkte 7-Kanal-Aufnahme sowie das normalisierte Oberflächenmodell nDSM. Bei der Auswahl der Vertices wird versucht, die Kronenform möglichst gut zu erfassen um die Anzahl der Mischpixel so klein wie möglich zu halten. Der erstellte Datensatz sieht folgendermaßen aus:

FID	Shape *	Baumart	BHD	BaumID	Streifen	Area	MAXIMUM	MINIMUM	MEAN	STD DEV	B ART	BART ID
0	Polygon	Buche	67	249	908	122,571	33,3	5,51	26,277	3,799	5	Buche_249
1	Polygon	Tanne	50	250	908	8,113	31,69	9,16	26,046	4,684	3	Tanne_250
2	Polygon	Kiefer	38	251	908	11,793	29,92	9,68	24,263	4,169	2	Kiefer_251
3	Polygon	Tanne	34	252	908	12,398	33,23	12,91	25,953	3,929	3	Tanne_252
4	Polygon	Tanne	34	253	908	8,092	29,03	16,42	24,217	3,137	3	Tanne_253
5	Polygon	Fichte	39	254	908	5,903	29,34	3,23	21,229	6,756	1	Fichte_254
6	Polygon	Eiche	50	255	908	30,48	31,01	8,74	22,568	5,124	6	Eiche_255
7	Polygon	Kiefer	50	256	908	20,23	32,88	12,78	26,739	3,699	2	Kiefer_256
8	Polygon	Kiefer	40	257	908	7,904	29,79	3,45	19,795	7,891	2	Kiefer_257
9	Polygon	Buche	34	258	908	11,754	31,45	9,65	27,017	3,341	5	Buche_258
10	Polygon	Buche	34	259	908	21,538	31,24	15,31	26,494	3,384	5	Buche_259
11	Polygon	Fichte	46	260	908	21,828	34,86	0,1	18,203	10,374	1	Fichte_260
12	Polygon	Fichte	60	261	908	20,714	33	0,09	17,282	9,311	1	Fichte_261
13	Polygon	Fichte	58	262	908	20,603	34,33	5,15	26,031	4,349	1	Fichte_262
14	Polygon	Fichte	58	264	908	24,148	36,36	0,06	21,896	9,478	1	Fichte_264
15	Polygon	Fichte	41	266	908	9,419	33,26	2,34	21,656	6,258	1	Fichte_266
16	Polygon	Kiefer	47	267	908	8,485	31,49	9,6	25,757	4,269	2	Kiefer_267
17	Polygon	Eiche	38	268	908	18,129	32,93	1,22	24,471	6,848	6	Eiche_268
18	Polygon	Kiefer	40	269	908	10,207	31,95	1,43	26,268	6,278	2	Kiefer_269
19	Polygon	Buche	47	270	908	33,666	33,95	3,67	27,155	5,261	5	Buche_270
20	Polygon	Buche	42	271	908	66,532	31,09	4,78	25,048	5,065	5	Buche_271
21	Polygon	Buche	38	272	908	42,992	32,68	9,17	28,298	3,394	5	Buche_272
22	Polygon	Kiefer	45	274	908	7,381	32,93	16,16	28,73	4,171	2	Kiefer_274
23	Polygon	Kiefer	44	275	908	7,802	29,63	19,36	24,971	2,387	2	Kiefer_275
24	Polygon	Kiefer	39	276	908	9,638	27,95	17,35	23,357	2,232	2	Kiefer_276
25	Polygon	Kiefer	44	277	908	7,404	30,56	19,64	25,752	2,274	2	Kiefer_277
26	Polygon	Kiefer	34	278	908	5,313	30,82	1,88	23,539	6,383	2	Kiefer_278
27	Polygon	Buche	38	279	908	46,263	33,11	7,16	26,988	4,312	5	Buche_279
28	Polygon	Fichte	52	280	908	12,015	32,49	6,29	26,098	5,775	1	Fichte_280
29	Polygon	Buche	91	281	908	110,333	35,32	0,09	27,451	4,084	5	Buche_281
30	Polygon	Fichte	50	282	908	28,487	37,73	7,83	26,393	5,12	1	Fichte_282
31	Polygon	Buche	54	283	908	46,141	30,12	6,09	24,144	3,977	5	Buche_283
32	Polygon	Buche	62	284	908	66,189	33,46	0,14	24,962	6,719	5	Buche_284
33	Polygon	Kiefer	46	285	908	11,313	31,41	0,05	22,444	9,976	2	Kiefer_285
34	Polygon	Kiefer	51	286	908	22,277	31,62	0,06	18,157	12,394	2	Kiefer_286
35	Polygon	Eiche	48	287	908	17,11	32,12	5,59	26,268	5,721	6	Eiche_287
36	Polygon	Tanne	40	288	908	10,319	32,17	20,35	28,461	2,796	3	Tanne_288
37	Polygon	Tanne	28	44	908	10,71	36,24	8,98	27,924	7,041	3	Tanne_44
38	Polygon	Tanne	52	290	908	10,527	33,61	1,1	23,713	8,047	3	Tanne_290
39	Polygon	Tanne	39	291	908	11,929	28,45	4,6	20,028	6,086	3	Tanne_291
40	Polygon	Tanne	29	292	908	11,707	25,85	0,2	15,984	5,907	3	Tanne_292

Abbildung 14: Digitalisierung der Referenzbäume

Die im Feld erhobenen Attribute Baumart, BaumID, Brusthöhendurchmesser und Streifen werden für jeden Baum definiert. Die Fläche jedes Baum-Polygons wird mit ArcGIS 10 berechnet und als Attribut „Area“ festgelegt. „B\_Art“ ist ein Index für die Baumart, welcher die weitere Datenverarbeitung erleichtert. „BART\_ID“ stellt eine Stringkombination aus Baumart und BaumID dar. Durch Einbeziehung des nDSM's werden darüber hinaus folgende Attribute über eine räumliche Abfrage eruiert:

- MAXIMUM: Höchster Punkt des nDSM's innerhalb des Baumpolygons; wird als Baumhöhe angenommen
- MINIMUM: Niedrigster Punkt des nDSM's innerhalb des Baumpolygons
- MEAN: Mittelwert der Höhenwerte des nDSM's innerhalb des Baumpolygons
- ST\_DEV: Standardabweichung der Höhenwerte des nDSM's innerhalb des Baumpolygons; Indikator für vertikalen Aufbau und Form der Baumkrone

## **7. Datenaufbereitung**

### **7.1. Import der Streifen in Erdas Imagine**

Es liegen insgesamt 30 Streifen der Hyperspektraldaten, welche den Großraum St. Margarethen an der Raab abdecken, vor. Über das Modul „Import“ im Programmpaket Erdas Imagine 9.1 wird jeder Streifen importiert und im img.-Format abgespeichert. Alle 30 Streifen werden deswegen importiert, um eine Übersichtskarte für die Datenerhebung erstellen zu können.

### **7.2. Mosaikierung**

Für die Referenzdatenerhebung werden flächendeckende, kontraststarke Ausdrücke der Hyperspektraldaten benötigt. Dazu werden die 30 Streifen des Hyperspektraldatensatzes über eine Mosaikierung in Erdas Imagine 9.1 zusammengeführt. Das erfolgt im Modul „Data Preparation“ über den Mosaic Wizard. Um die Datengröße zu reduzieren, werden für das Mosaik sieben Kanäle der Hyperspektralaufnahme, welche gleichmäßig über das elektromagnetische Spektrum verteilt sind, ausgewählt. Ausdrücke der erstellten 7-Kanal-Aufnahme werden als Datengrundlage für die Referenzdatenerhebung verwendet.

### **7.3. Radiometrische Datenvorverarbeitung**

Zur Kontrastverstärkung wird eine „Histogram Equalization“ für die mosaikierte 7-Kanal-Aufnahme, welche als Datengrundlage für die Feldarbeit und die Referenzdatenerhebung dient, durchgeführt. Dabei werden die Pixelwerte einer Aufnahme so transformiert, dass die Graustufen in etwa gleich werden und ein möglichst flaches Histogramm entsteht. „Histogram Equalization“ eignet sich damit nicht zur radiometrischen Anpassung benachbarter Hyperspektralstreifen, sondern wird für die unabhängig von den Nachbarstreifen durchzuführende Kontrastoptimierung angewendet. Um die statistische Aussagekraft der Ergebnisse nicht zu verfälschen, werden die Pixelwerte des Hyperspektraldatensatzes des Untersuchungsstreifens „0908“ nicht verändert.

Abbildung 15 zeigt die mosaikierte 7-Kanal Aufnahme in der Kanalkombination 5-3-2 (entspricht Kanalkombination 52-32-22 im Hyperspektraldatensatz) und das zugehörige Histogramm vor der „Histogram Equalization“. Wie in Abbildung 15 erkennbar, ist die Grauwertverteilung sehr spitz, es wird nur ein Teil der zur Verfügung stehenden Grauwertskala genutzt und die Aufnahme ist folglich relativ kontrastarm.

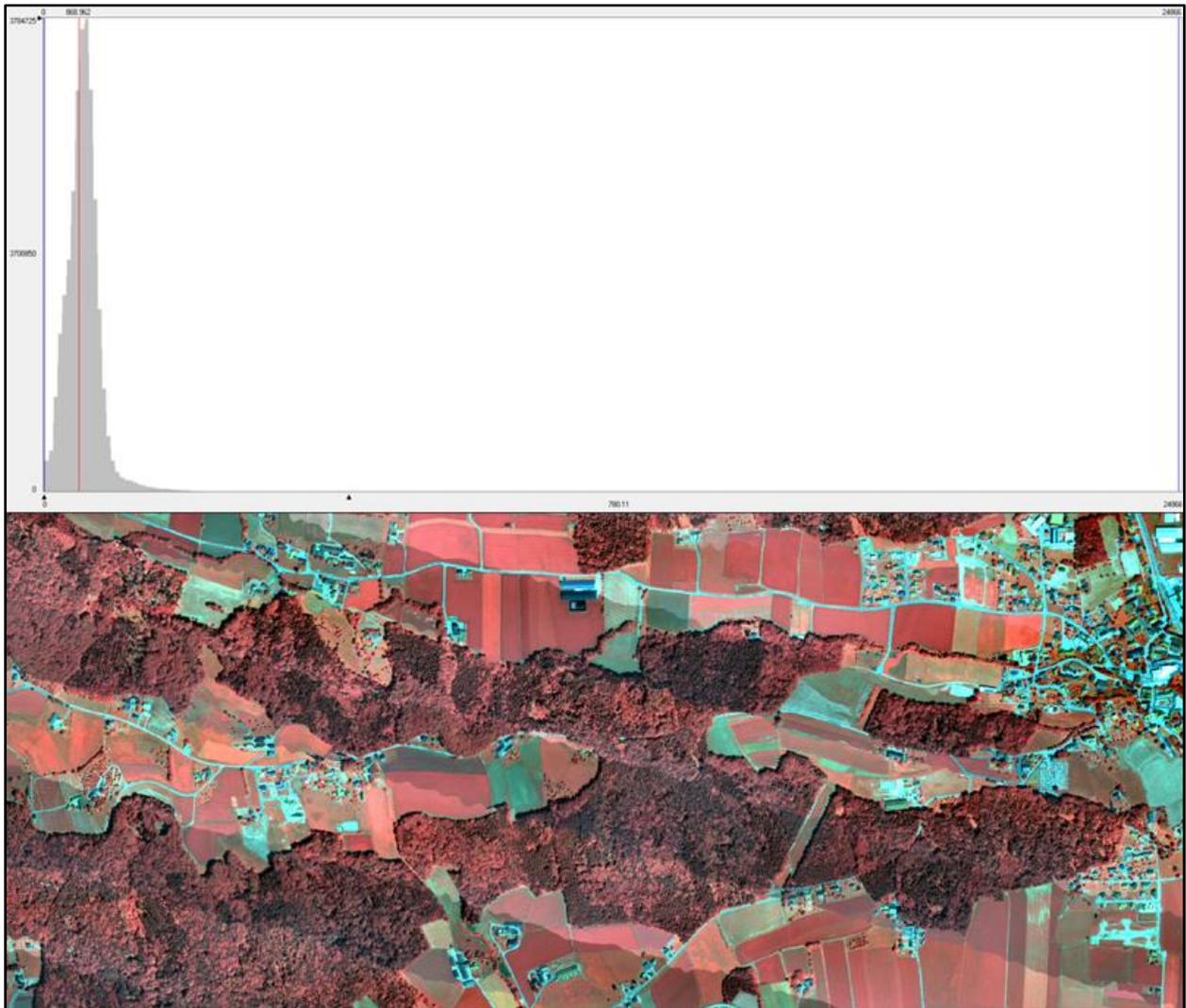


Abbildung 15: 7-Kanal-Aufnahme vor der Histogram Equalization

Abbildung 16 zeigt die 7-Kanal-Aufnahme und das zugehörige Histogramm nach dem Histogrammausgleich. Die Intensitätsverteilung der Grauwerte ist über die gesamte Grauwertskala gleichmäßig verteilt und der zur Verfügung stehende Wertebereich wird optimal ausgenutzt. Häufige Grauwerte werden quasi „auseinandergezogen“ und die Grauwertskala wird in diesen Bereichen gestreckt, weniger häufige Grauwerte werden „zusammengeschoben“ und die Grauwertskala wird gestaucht. Der Kontrast wird dadurch gestärkt und einzelne Individuen können in der Referenzdatenerhebung besser identifiziert werden.

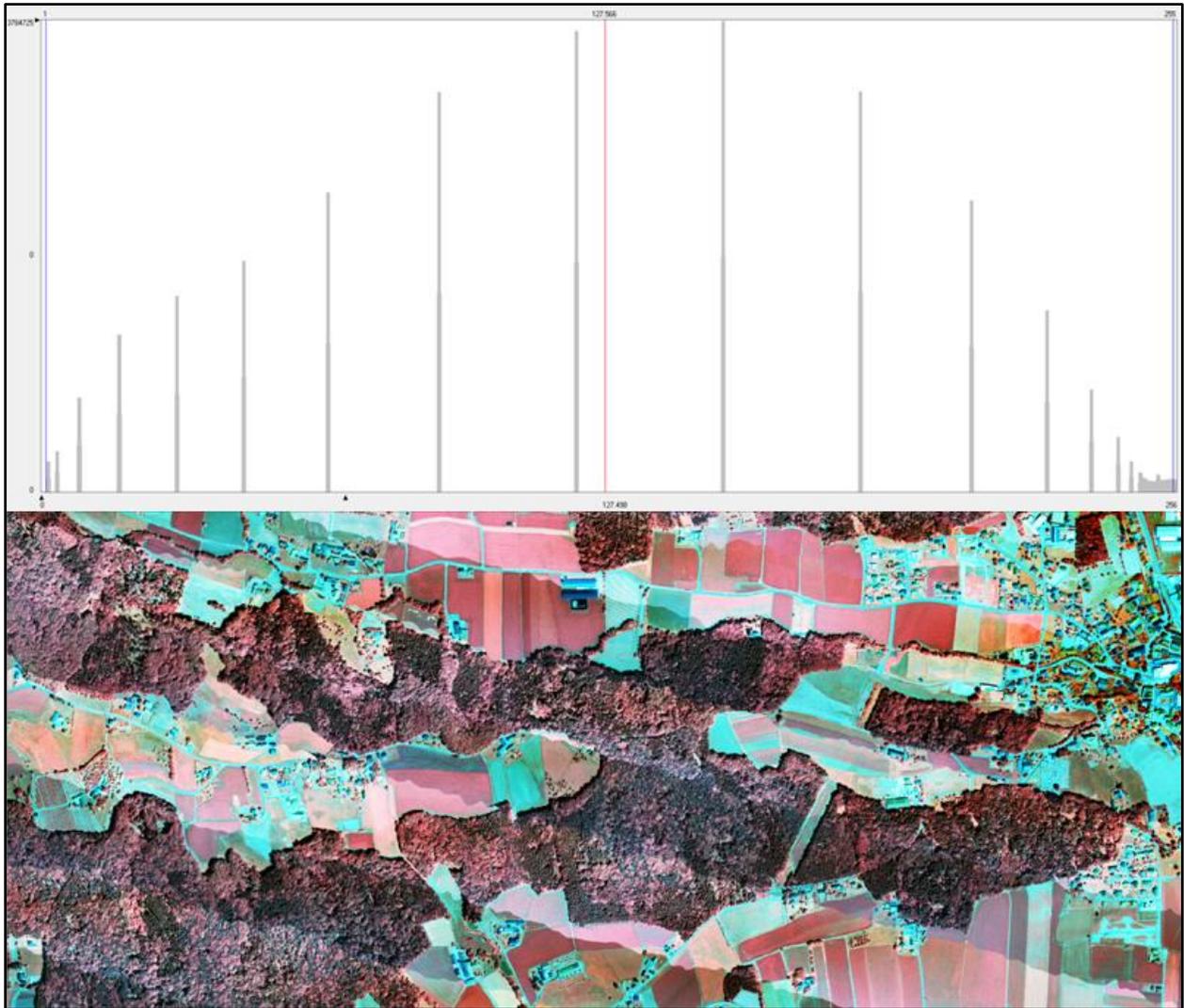


Abbildung 16: Kontrastverstärkte 7-Kanal-Aufnahme nach der Histogram Equalization

#### 7.4. Streifenproblematik und geometrischer Versatz

Benachbarte Streifen des Hyperspektraldatensatzes weisen eine Überlappung von ca. 20% auf. Betrachtet man das Überlappungsgebiet genauer, so erkennt man einen nicht systematischen geometrischen Versatz von bis zu 12 m. Der Verlauf der Waldgrenze und die Position der Referenzbäume in den zwei überlappenden Streifen werden in Abbildung 17 veranschaulicht. Man erkennt, dass der Versatz zwischen 6-12 m variiert, was möglicherweise auf Fehler bei der geometrischen Entzerrung durch den Datenanbieter zurückzuführen ist. Eine Korrektur des nicht systematischen geometrischen Versatzes ist für die Signaturanalyse nicht erforderlich, würde die Arbeit sprengen und wurde deswegen verworfen. Zu beachten ist, dass benachbarte Streifen mit entgegengesetzter Flugrichtung aufgenommen wurden und dadurch gleiche Bäume im Überlappungsgebiet ein völlig anderes Erscheinungsbild in den zwei beobachteten Streifen haben können.

Der Grund dafür ist Aufnahmegeometrie, Sonnenstand und die daraus resultierenden optischen Eigenschaften.

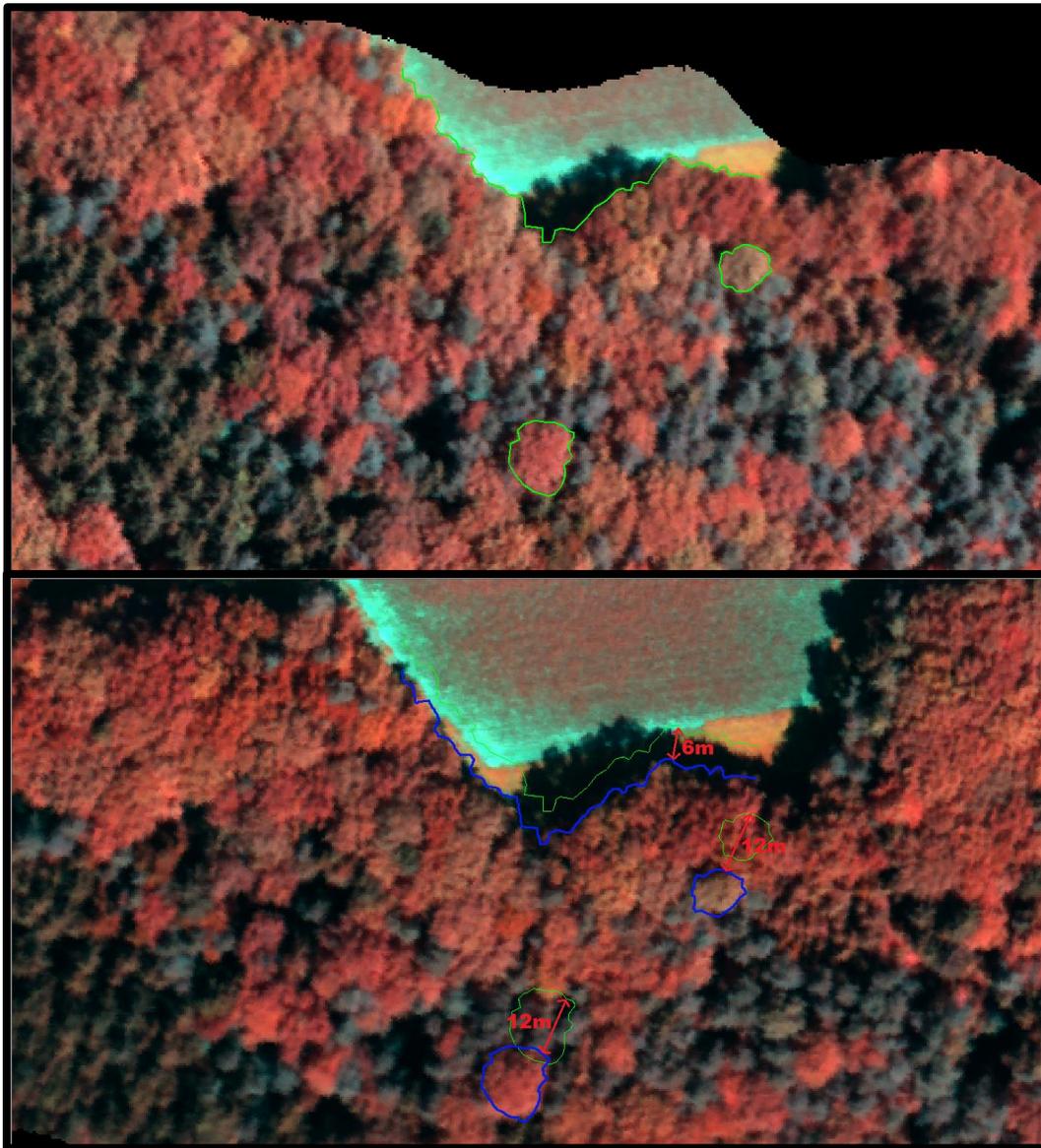


Abbildung 17: Geometrischer Versatz

Die Digitalisierung der Referenzbäume erfolgt auf Basis der kontrastverstärkten 7-Kanal-Aufnahme, welche im Überlappungsgebiet aufgrund der Mosaikierung aus dem Streifen „0908“ besteht. Wegen dem nicht systematischen geometrischen Versatz können die im Überlappungsgebiet digitalisierten Referenzbäume nicht direkt für den Nachbarstreifen „0909“ übernommen werden. Die Tatsache, dass ein solcher Referenzbaum im Streifen „0909“ ein völlig anderes Erscheinungsbild haben kann, verhindert eine Nachdigitalisierung. Die im Überlappungsgebiet der beiden Hyperspektralstreifen erhobenen Referenzbäume können deswegen ausschließlich für den Streifen „0908“ verwendet werden.

## 7.5. Berechnung eines normalisierten digitalen Oberflächenmodells

Die Berechnung eines digitalen Oberflächenmodells (DSM) erfolgt mit dem Programmpaket LAStools. Der Import der Punktwolke im .las-Format ist der erste Schritt. Daraufhin wird mit dem LAS2DEM-Tool ein DSM mit einer Gridsize von 0,5m errechnet. Abbildung 18 veranschaulicht die Umgebung LAStools und den geclippten Projektbereich bei der DSM-Erstellung.

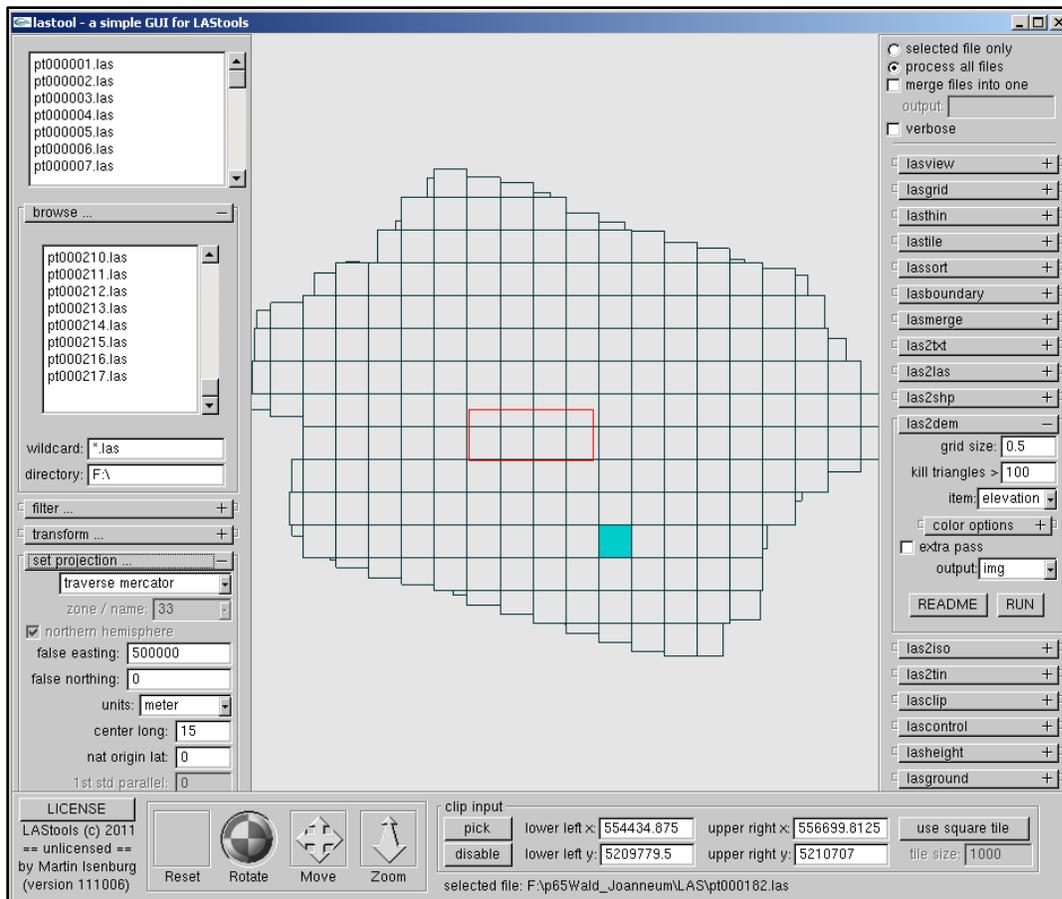


Abbildung 18: DSM-Erstellung in LAStools

Als Output erhält man ein Höhenmodell im img.-Format, welches auf das Projektgebiet geclipped wird. Die Georeferenzierung des DSM's erfolgt in Erdas Imagine über „Map Model -> Add Projection“. Ausgewählt wird UTM, WGS84 North und Streifen 33. Ein DTM aus derselben Befliegung liegt vor und wird auf das Projektgebiet geclipped. Ein normalisiertes digitales Oberflächenmodell mit relativen Baumhöhen wird im Erdas Imagine „Model Maker“ über die Gleichung  $nDSM = DSM - DTM$  berechnet. Das nDSM wird mit einem „3x3 Low-Pass-Filter“ gefiltert und fehlende Werte werden rekodiert. Das berechnete nDSM wird eingesetzt um Formparameter wie zum Beispiel die Baumhöhe, die Standardabweichung der Höhenwerte oder die durchschnittliche Steigung zu ermitteln.

## **8. Ergebnisse**

### **8.1. Signaturanalyse**

#### **8.1.1. Umsetzung**

Um eine Signaturanalyse in Erdas Imagine 9.1 durchzuführen, wird aus dem digitalisierten Referenzdatensatz für jede Baumart ein eigenes Shape-File erstellt und als AOI im Signature Editor importiert. Die einzelnen Trainingsgebiete werden nach Baumart über einen „Merge“ zusammengeführt und aus den gemittelten Spektralkurven wird ein Mean Plot erstellt. Um die Signaturdaten für eine statistische Analyse in R verwenden zu können, werden die entsprechenden Statistiken als Textfile exportiert und mit dem Referenzdatensatz über einen „Join“ vereint.

Der Import in R erfolgte über „Datenmanagement -> Importiere Daten -> from textfile“. Dabei war darauf zu achten, dass das Datenfeldtrennzeichen sowie das Kommazichen korrekt definiert werden. Der Datensatz liegt nach dem Import im Tabellenformat vor und kann somit für die statistische Auswertung verwendet werden.

#### **8.1.2. Mean Plot aller Referenzbäume**

Der Mean Plot der Referenzbäume in Abbildung 19 veranschaulicht die spektralen Charakteristika der untersuchten Baumarten im untersuchten Hyperspektralstreifen „0908“. Die Spektralkurven setzen sich aus den gemittelten Werten der einzelnen Testgebiete für die jeweilige Klasse zusammen. Die X-Achse ist der Kanal im Hyperspektraldatensatz, die Y-Achse gibt den mittleren Reflexionswert des Baumpolygons wieder.

In der Abbildung erkennt man, dass das spektrale Reflexionsverhalten der Baumarten im sichtbaren Bereich des Lichts (bis Kanal 33) sehr ähnlich ist, im nahen Infrarotbereich sind die Spektralkurven differenzierter. Im Nahen Infrarot ab Kanal 33 findet der typische Anstieg der Reflexion von Vegetation statt. Die Absorption des Lichtes ist in diesem Wellenlängenbereich gering, da die einfallende Strahlung die Blattstruktur tief durchdringt.

Das Reflexionsverhalten der Baumarten Buche und Eiche sind im Red-Edge Bereich (Kanal 31 – Kanal 43) und im nahen Infrarot nahezu ident, im sichtbaren Bereich jedoch deutlich differenzierter. Gleiches gilt für die beiden Nadelbaumarten Kiefer und Tanne. Laubbaumarten weisen im nahen Infrarot einen deutlich höheren Reflexionswert als Nadelbaumarten auf.

Die Trennbarkeit der Baumarten kann man in Abbildung 19 nicht ablesen, da dafür die Varianz berücksichtigt werden muss. So können große Differenzen der Mittelwerte mit sehr hoher Varianz schlechter trennbar als kleine Differenzen mit sehr kleiner Varianz.

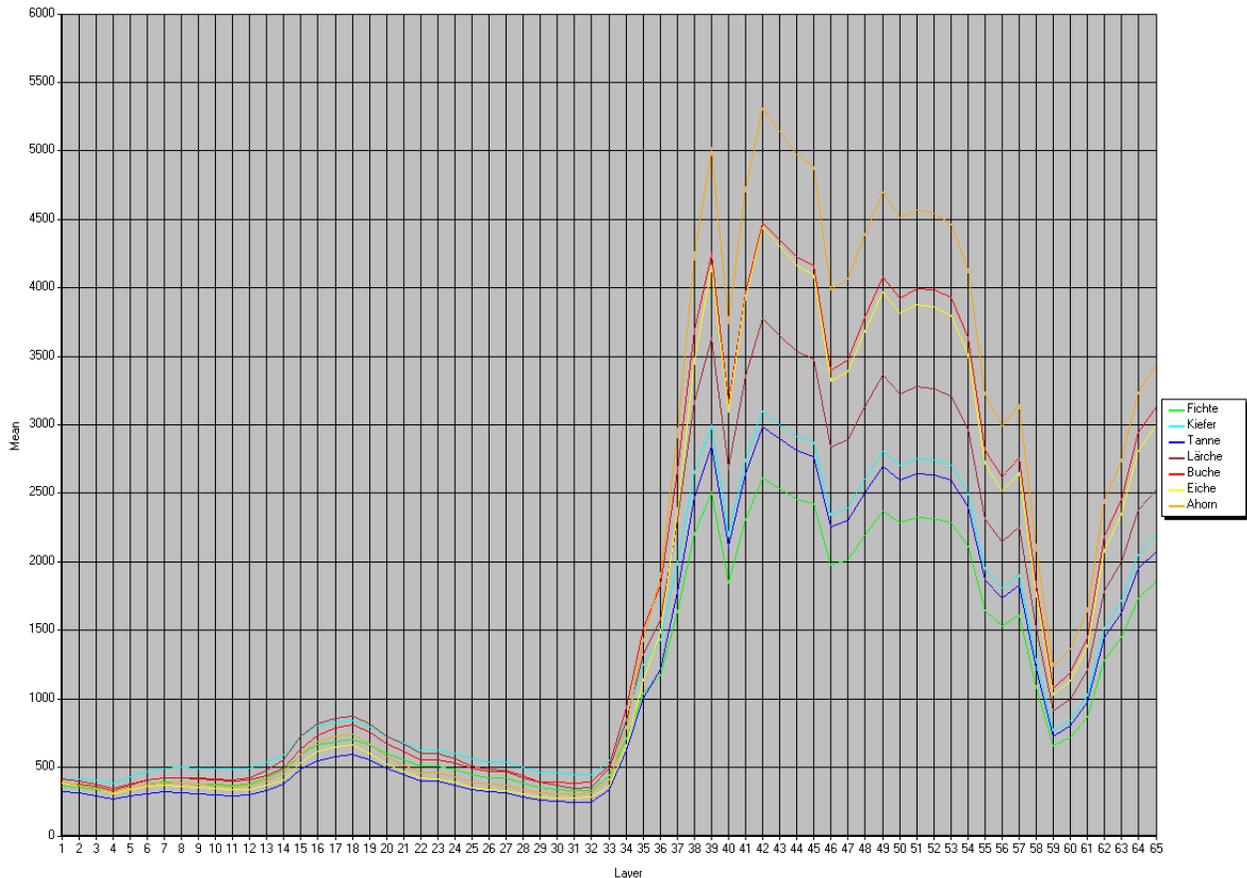


Abbildung 19: Mean Plot der Referenzbäume

### 8.1.3. Verteilung der Referenzbäume im spektralen Merkmalsraum

Abbildung 20 zeigt einen 3d-Plot der Referenzbäume mit drei aussagekräftigen Prädiktoren (Kanal 7, Kanal 33 und Kanal 62) des logistischen Regressionsmodells als Achsen. Die Punkte stellen dabei die Mittelwerte der Referenzbäume dar. Kanal 7 deckt einen Teil des blauen Lichtes ab, Kanal 33 befindet sich am Anfang des Red-Edge-Bereichs und Kanal 62 ist im nahen Infrarot angesiedelt. Um die Baumarten ist ein 50% Konzentrationsellipsoid gelegt, die Ellipsoide der Laubbaumarten Buche (grün), Eiche (orange) und Ahorn (blau) liegen sehr nahe beieinander beziehungsweise überlappen sich leicht. Die Ellipsoide der Nadelbaumarten Fichte (pink), Kiefer (türkis), Lärche (rot) und Tanne (gelb) berühren sich nicht.

Nadelbaumarten lassen sich mit dem erhobenen Referenzdatensatz und der Hyperspektralaufnahme als Datengrundlage folglich einfacher trennen als Laubbaumarten. Interessanterweise berühren sich die Ellipsoide der Nadelbaumart Tanne und der Laubbaumart Eiche. Diese Berührung ist darauf zurückzuführen, dass Eichen und Tannen im sichtbaren Bereich des Lichtes ein ähnliches Reflexionsverhalten aufweisen.

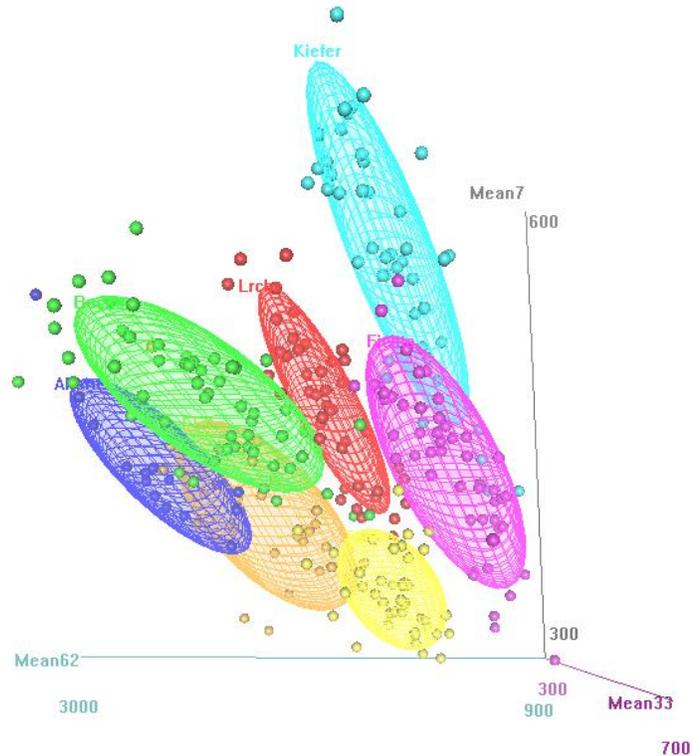


Abbildung 20: Referenzbäume im spektralen Merkmalsraum

## 8.2. Verwendete LiDAR-Variablen

Die folgenden LiDAR-Variablen dienen als Indikatoren für den baumartenspezifischen vertikalen Aufbau der Baumkrone. Vier Variablen, welche aus den LiDAR-Daten abgeleitet sind, werden auf ihre Eignung zur statistischen Trennung von Baumarten untersucht:

- ❖ **STDV**: Standardabweichung der Höhenwerte pro Baumpolygon, ermittelt in ArcGIS über zonale Statistiken. Dazu wurde ein Raster aus dem nDSM erstellt, welches auf die Baumpolygone geclippt ist. Danach ist es möglich über zonale Statistiken im ArcGIS Spatial Analyst die Standardabweichung der Höhenwerte innerhalb des entsprechenden Baumpolygons zu berechnen.

- ❖ **STDV\_filter**: Standardabweichung der Höhenwerte aus einem gefilterten nDSM, selbe Vorgehensweise wie bei Variable STDV
- ❖ **Slope\_mean**: Mittlere Steigung des Baumpolygons, berechnet über ArcGIS Spatial Analyst – Methode „Slope“. Dieses Werkzeug berechnet für jede Rasterzelle des nDSM's die Steigung, also den maximale Höhenunterschied jeder Rasterzelle zu seinen unmittelbaren Nachbarzellen. Als Ergebnis erhält man ein neues Raster mit den Steigungswerten. Darauf wird der Mittelwert der Steigungswerte innerhalb jedes Baumpolygons berechnet.
- ❖ **Slope\_stdv**: Standardabweichung der Steigung – aus dem Raster mit den Steigungswerten der Baumpolygone wird über zonale Statistiken die Standardabweichung berechnet.

Tabelle 2 zeigt eine Zusammenfassung der Variablen Slope\_mean, Slope\_stdv und STDV.

**Tabelle 2: Zusammenfassung numerischer Variablen**

```

Variable: slope_mean
      mean      sd      IQR      0%      25%      50%      75%      100%  n
Ahorn  76.15073  3.233374  5.402450  70.7618  74.20325  76.09990  79.60570  80.2863  16
Buche  68.29804  5.688552  6.298975  49.8960  65.66875  68.76570  71.96773  78.1507  56
Eiche  73.97764  5.127158  7.174725  63.3220  70.50290  75.32260  77.67762  83.6835  28
Fichte 75.19062  3.879874  5.914825  66.0906  72.26772  76.20370  78.18255  80.7298  50
Kiefer 72.55888  6.044852  9.804250  59.5673  67.88760  72.43360  77.69185  82.4738  46
Lärche 78.17864  4.109138  5.058575  67.0760  75.98855  79.66885  81.04713  83.8891  44
Tanne  73.84811  5.635027  8.208300  59.8656  69.86000  75.46230  78.06830  83.3326  43

Variable: slope_stdv
      mean      sd      IQR      0%      25%      50%      75%      100%  n
Ahorn  12.61719  2.952763  4.228975  7.71063  10.756725  12.12890  14.98570  17.9566  16
Buche  14.74082  2.561577  2.902425  8.43968  13.301100  14.57010  16.20352  23.4044  56
Eiche  12.17725  3.285189  3.954638  4.22021  10.367513  12.18930  14.32215  20.3533  28
Fichte 10.99831  3.431151  3.284560  5.48999  9.202565  10.84370  12.48713  26.0846  50
Kiefer 12.17880  3.924444  4.609932  3.57915  10.040567  11.90935  14.65050  24.7735  46
Lärche 11.89318  4.933744  5.986120  4.32776  8.475655  11.95075  14.46177  25.8528  44
Tanne  11.45576  3.505515  5.120540  4.06272  9.437810  11.09440  14.55835  18.0703  43

Variable: STDV2
      mean      sd      IQR      0%      25%      50%      75%      100%  n
Ahorn  7.093294  1.875141  2.418637  3.61363  6.088915  7.574510  8.507552  9.94764  16
Buche  4.491742  1.397622  1.598323  1.58546  3.641090  4.371815  5.239413  8.22600  56
Eiche  5.938240  1.548985  1.864297  2.93964  5.017688  5.994080  6.881985  9.10974  28
Fichte 6.070068  2.173003  3.227257  2.53919  4.539808  5.499960  7.767065  11.16870  50
Kiefer 4.936192  2.138862  2.961000  2.11675  3.456870  4.686145  6.417870  11.73980  46
Lärche 8.382105  2.132973  2.651645  3.26314  7.432880  8.753240  10.084525  11.36750  44
Tanne  5.161508  1.850609  2.971835  1.65031  3.736285  5.009690  6.708120  8.55039  43

```

Der statistische Wert „mean“ ist der Mittelwert, „sd“ die Standardabweichung und „IQR“ die interquartile Range. Daneben werden ausgewählte Quantile von 0-100% sowie die Stichprobenanzahl „n“ dargestellt.

Die beiden aus der Steigung abgeleiteten Variablen Slope\_mean und Slope\_stdv werden bezugnehmend auf eine Arbeit von BALTSAVIAS et al. (2007) als Indikatoren für den Kronenaufbau und die Neigung der Baumkrone angenommen.

Betrachtet man die Variable Slope\_mean genauer so stellt man fest, dass Buchen eine eindeutig geringere mittlere Steigung als andere Baumarten haben. Dies ist auf den laubbaumspezifischen kugelförmigen Kronenaufbau und das dichte Blätterdach zurückzuführen. Bei der Standardabweichung der Steigung Slope\_stdv haben Fichten die geringsten und Buchen die höchsten Werte.

Die Variable STDV ist die Standardabweichung der Höhenwerte aus dem nDSM, welche über zonale Statistiken in ArcGIS berechnet wird. Laubbäume haben in der Regel aufgrund des dichten Blätterdachs eine tendenziell niedrigere Standardabweichung als Nadelbäume. Buchen haben mit einem Wert von 4,49 die niedrigste Standardabweichung der Höhenwerte, Lärchen mit 8,38 die höchste. Die aus einem gefilterten NDSM abgeleitete Standardabweichung STDV\_filter weist eine ähnliche Charakteristik wie die Variable STDV auf und ist in Tabelle 2 nicht abgebildet.

### **8.3. Baumartendifferenzierung**

#### **8.3.1. Multinomiales logistisches Regressionsmodell zur Trennung aller Baumarten**

Wegen dem nicht zu lösenden geometrischen Versatz, welcher die Verwendung von Referenzbäumen im Überlappungsgebiet für beide Streifen nicht ermöglicht und der damit zu geringen Anzahl an Referenzbäumen im Streifen „0909“, beschränkt sich die Baumartendifferenzierung auf den Streifen „0908“.

Über ein multinomiales logistisches Regressionsmodell wird die Trennbarkeit von 267 Referenzbäumen der Baumarten Buche, Eiche, Fichte, Kiefer, Lärche und Tanne untersucht. Für die sechs Baumarten wird ein multinomiales logistisches Regressionsmodell mit sieben Prädiktoren( $n+1$ ) gesucht.

Ein multinomiales logistisches Regressionsmodell wird in R Commander über „Statistik-> Regressionsmodelle -> Multinomiales Logitmodell“ erstellt. Wie in Abbildung 21 ersichtlich, werden in der Modellgleichung die Baumart als abhängige Variable und Prädiktoren, welche über die Stepwise Model Selection ausgewählt werden, eingegeben. Als Ergebnis erhält man Koeffizienten, Standardfehler und Wald-Statistiken für alle untersuchten Baumarten.

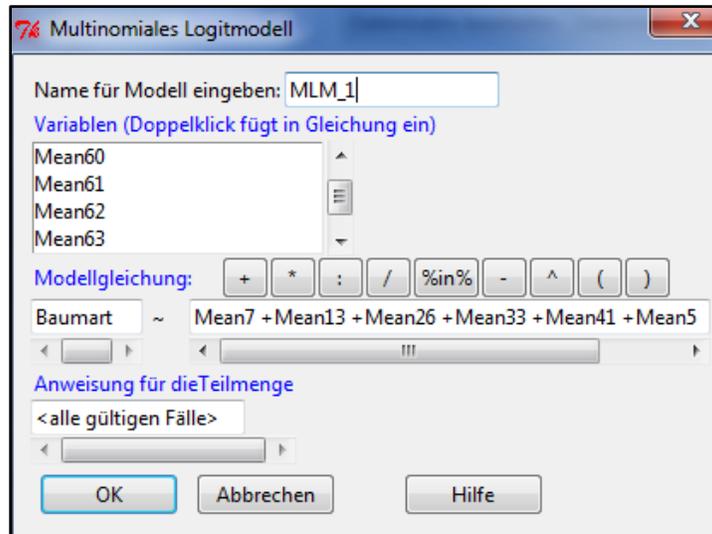


Abbildung 21: Erstellung eines multinomialen Logitmodells

### 8.3.1.1. Ermittlung der Modellparameter

Die besten Prädiktoren für das multinomiale logistische Regressionsmodell werden über eine „Stepwise model selection“ (Kapitel 5.2.2.3) ausgewählt. Die statistische Signifikanz aller getesteten Modell- Prädiktoren zur Baumartendifferenzierung der Referenzbäume wird dabei beurteilt. Hierzu wird ein multinomiales logistisches Regressionsmodell mit den Reflexions-Mittelwerten aller Kanäle des Hyperspektraldatensatzes sowie mit den in Abschnitt 8.2 beschriebenen LiDAR-Variablen erstellt. LiDAR- und Hyperspektralvariablen werden bei der statistischen Differenzierung aller Baumarten über die Methode der „Stepwise model selection“ gleichberechtigt behandelt.

Für das Gesamtmodell wird eine forward/backward selection mit Abbruchkriterium AIC zur Ermittlung der sieben besten Prädiktoren(n+1) durchgeführt.

Die signifikantesten Prädiktoren zur Trennung aller sechs untersuchten Baumarten sind Mittelwerte der Reflexionswerte aus folgenden sieben Kanälen des Hyperspektraldatensatzes:

Start: Baumart ~ 1	AIC=1081,25
+ Kanal 62	AIC=694,83
+ Kanal 7	AIC=325,10
+ Kanal 33	AIC=220,49
+ Kanal 13	AIC=154,94
+ Kanal 57	AIC= 132,47
+ Kanal 26	AIC= 104,50
+ Kanal 41	AIC=98,773

Residuendevianz = 2,772589

### 8.3.1.2. Modellparameter und Koeffizienten

Tabelle 3 zeigt eine Zusammenfassung des multinomialen logistischen Regressionsmodells mit den Modellkoeffizienten und den zugehörigen Standardfehlern. Die Spalte ganz links beinhaltet die Intercepts, die Modellkonstanten des multinomialen logistischen Regressionsmodells. Daneben sind die Modellkoeffizienten der einzelnen Prädiktoren für alle Ausprägungen der Zielvariable abgebildet.

Der Log-Odds-Wert, also die modellierte Wahrscheinlichkeit einer Beobachtung für eine konkrete Zielklasse relativ zur Basisklasse, kann über Formel 5.2.2.4 und die Modellparameter in Tabelle 3 berechnet werden. Die Beobachtung ist in dem Fall ein Baum-Polygon, welches einer Zielklasse zugeordnet wird. Dazu wird für jede Zielklasse ein Log-Odds-Wert berechnet. Darauf wird das untersuchte Baumpolygon der Zielklasse mit dem höchsten Log-Odds-Wert zugewiesen.

Wichtig für die Lesbarkeit von Tabelle 3 sind die Vorzeichen der Modellkoeffizienten. Ist das Vorzeichen eines Modellkoeffizienten positiv, so erhöht ein hoher Kanalmittelwert des untersuchten Baumpolygons (z.B. „Mean7“ bei Zielklasse Buche) die Wahrscheinlichkeit für die Zugehörigkeit zur entsprechenden Zielklasse. Ist das Vorzeichen des Modellkoeffizienten negativ, so mindert der Prädiktor den Log-Odds-Wert der Zielklasse. Zum Beispiel mindert ein sehr hoher Reflexionsmittelwert in Kanal 13 („Mean13“) die Wahrscheinlichkeit, dass ein Baumpolygon der Zielklasse Buche zugewiesen wird, da der Wert des negativen Modellkoeffizienten „Mean13“ bei der Zielklasse Buche am höchsten ist.

#### **Tabelle 3: Modellparameter des logistischen Regressionsmodells**

Call:

```
multinom(formula = Baumart ~ Mean62 + Mean7 + Mean33 + Mean13 +  
  Mean57 + Mean26 + Mean41, data = SamplesFinal, trace = FALSE)
```

Coefficients:

	(Intercept)	Mean62	Mean7	Mean33	Mean13	Mean57	Mean26	Mean41
Buche	-13179.197	-37.339868	515.0170	-236.20392	-777.4488	43.75864	585.81043	-10.798654
Eiche	-2117.251	5.933034	377.6097	-162.70961	-579.9685	-11.52234	398.79186	5.130543
Fichte	17167.744	-83.822485	275.2437	112.15153	-387.7292	34.36789	77.11252	8.703592
Kiefer	-21191.609	-83.958229	580.2521	-36.91802	-578.4750	60.46336	209.15244	-7.250736
Lärche	6873.315	-28.849924	157.9443	-192.19314	-275.7279	21.90014	331.26268	-1.103609
Tanne	22182.139	-61.317401	179.6450	40.03971	-362.5013	51.22499	152.68743	-7.760896

Std. Errors:

	(Intercept)	Mean62	Mean7	Mean33	Mean13	Mean57	Mean26	Mean41
Buche	1.078857e-04	2.092106e-01	5.115151e-02	6.242718e-02	5.796989e-02	2.621236e-01	6.109469e-02	3.894829e-01
Eiche	1.051148e-13	2.191079e-10	4.347684e-11	4.462070e-11	4.434604e-11	2.728945e-10	4.257622e-11	3.983451e-10
Fichte	2.665888e-08	5.165568e-05	9.667603e-06	1.018095e-05	1.006167e-05	6.741553e-05	9.502531e-06	1.060606e-04
Kiefer	1.956857e-85	3.301541e-82	1.055104e-82	1.299604e-82	1.185730e-82	4.036188e-82	1.278205e-82	5.709911e-82
Lärche	1.078857e-04	2.092106e-01	5.115151e-02	6.242718e-02	5.796989e-02	2.621236e-01	6.109469e-02	3.894829e-01
Tanne	2.665888e-08	5.165568e-05	9.667602e-06	1.018095e-05	1.006167e-05	6.741553e-05	9.502531e-06	1.060606e-04

Residual Deviance: 2.772589

AIC: 98.77259

Tabelle 4 zeigt die Wald-Statistiken des multinomialen logistischen Regressionsmodells. Die Wald-Statistiken setzen sich aus dem Quotient der Modellkoeffizienten mit dem zugehörigen Standardfehler zusammen. Die Wald-Statistik gibt einen Überblick über die Signifikanz jedes Prädiktors für die Zielklasse, je größer der Absolutwert der Wald-Statistik, desto signifikanter ist der Prädiktor. Ist das Vorzeichen der Wald-Statistik negativ, so hat der entsprechende Prädiktor einen negativen Effekt auf die Wahrscheinlichkeit für die Zielklasse.

**Tabelle 4: Wald-Statistiken des logistischen Regressionsmodells**

```
Value/SE (Wald statistics):
(Intercept)      Mean62      Mean7      Mean33      Mean13
Buche  -1.221589e+08 -1.784798e+02 1.006846e+04 -3.783671e+03 -1.341125e+04
Eiche  -2.014228e+16  2.707814e+10 8.685306e+12 -3.646505e+12 -1.307825e+13
Fichte  6.439785e+11  -1.622716e+06 2.847073e+07  1.101583e+07 -3.853525e+07
Kiefer  -1.082941e+89  -2.543002e+83 5.499478e+84 -2.840714e+83 -4.878640e+84
Lärche  6.370924e+07  -1.378990e+02 3.087774e+03 -3.078677e+03 -4.756399e+03
Tanne   8.320732e+11  -1.187041e+06 1.858217e+07  3.932808e+06 -3.602793e+07
      Mean57      Mean26      Mean41
Buche   1.669390e+02  9.588566e+03 -2.772562e+01
Eiche  -4.222268e+10  9.366540e+12  1.287964e+10
Fichte  5.097918e+05  8.114946e+06  8.206241e+04
Kiefer  1.498031e+83  1.636298e+84 -1.269851e+82
Lärche  8.354891e+01  5.422119e+03 -2.833523e+00
Tanne   7.598396e+05  1.606808e+07 -7.317414e+04
```

### 8.3.1.3. Kontingenztmatrix

Das errechnete multinomiale logistische Regressionsmodell aus Tabelle 3 wird getestet, indem die erhobenen Referenzbäume als Input für eine Klassifikation herangenommen werden. Dieser Schritt ist im Analyseprogramm R über die predict()-Methode möglich, dabei werden für jeden Referenzbaum die Log-Odds-Werte für alle untersuchten Baumarten berechnet. Als Prädikatoren werden Mittelwerte aus den Kanälen 7, 13, 26, 33, 41, 57 und 62 des Hyperspektraldatensatzes verwendet.

Tabelle 5 zeigt die Kontingenztmatrix der Klassifikation aller Referenzbäume auf Basis des logistischen Regressionsmodells mit Prädiktoren aus dem Hyperspektral-Datensatz. Alle 267 Referenzbäume können der korrekten Zielklasse zugeordnet werden. Die Kombination von sieben ausgewählten Kanälen aus dem Hyperspektraldatensatz zur Trennung aller sechs Zielklassen liefert damit eine sehr gute Performance. Die Kontingenztmatrix ist jedoch keine echte Verifizierung, da ausschließlich Bäume klassifiziert werden, welche auch für die Modellanpassung verwendet werden.

**Tabelle 5: Kontingenztmatrix**

Baumart	Samplesize	klassifiziert als						Genauigkeit
		Buche	Eiche	Fichte	Kiefer	Tanne	Lärche	
Buche	56	56	0	0	0	0	0	100%
Eiche	28	0	28	0	0	0	0	100%
Fichte	50	0	0	50	0	0	0	100%
Kiefer	46	0	0	0	46	0	0	100%
Tanne	43	0	0	0	0	43	0	100%
Lärche	44	0	0	0	0	0	44	100%

Gesamtgenauigkeit: **100%**

### 8.3.1.4. Kreuzvalidierung

Aufgrund der Tatsache, dass im Feld erhobene Referenzdaten immer fehlerbehaftet sind, muss der Messfehler für das statistische Vorhersagemodell berücksichtigt werden. Laut WOLLSCHLÄGER (2010) orientiert sich das Vorhersagemodell „zu stark an den zufälligen Besonderheiten einer konkreten Stichprobe aus der selben Population“, was in der Regel zu schlechten Vorhersagen und sogenannten „Overfitting“ führt. Die Generalisierbarkeit der Parameterschätzungen lässt sich mit einer Kreuzvalidierung quantifizieren, das heißt inwieweit das Modell für gering abweichende Ausprägungen von den tatsächlichen Werten zu sinnvollen Vorhersagen führt.

Methodisch wird bei der Kreuzvalidierung die gesamte Stichprobe in eine Trainingsstichprobe und in eine Validierungsstichprobe aufgeteilt. Die Trainingsstichprobe ist die Datenbasis für die Parameterschätzung der Validierungsstichprobe. Die daraus abgeleitete Modelanpassung wird auf die verbleibende Validierungsstichprobe angewendet, indem in die Vorhersagegleichung die Prädiktorwerte der Teststichprobe eingesetzt werden. Dieser Prozess ist iterativ, das heißt die Anpassung wird für wechselnde Trainings- und Validierungsstichproben durchgeführt (WOLLSCHLÄGER 2010).

Das erstellte Regressionsmodell wird über eine Leave-One-Out-Kreuzvalidierung (LOOCV) getestet. Dabei wird jede Beobachtung (Referenzbaum) separat als Teststichprobe für ein Modell verwendet, welches über die verbleibenden n-1 Stichproben angepasst wird. Über den Parameter K, welcher bei einer LOOCV-Kreuzvalidierung der Stichprobenanzahl entspricht, wird die Aufteilung der Trainings- und Validierungsstichprobe in K-disjunkte Teilmengen determiniert. Wichtiger Indikator für die Modelanpassung ist der Kreuzvalidierungsfehler „delta“ und seine korrigierte Anpassung. Der Kreuzvalidierungsfehler wird zum Vergleich von Modellen herangezogen und gibt Prozentsatz der Fehlklassifikationen wieder. Ein Kreuzvalidierungsfehler „delta“ von 0,0318 bedeutet dass 3,18% der getesteten Referenzbäume der falschen Zielklasse zugeordnet werden.

### **8.3.1.5. Analyse der Modellanpassung**

Für die optimale Güte des multinomialen logistischen Regressionsmodells werden Prädiktoren beziehungsweise eine Kombination von Prädiktoren gesucht, die für eine bestmögliche Trennbarkeit der untersuchten Baumarten sorgen.

Zur Ermittlung der Prädiktoren wird eine „forward/backward selection“ mit Abbruchkriterium AIC verwendet. Laut GERNAND (2009) neigen über den AIC bestimmte Regressionsmodelle zu einem sogenannten „overfitting“, das Modell enthält dabei zu viele erklärende Variablen (Prädiktoren). Damit ist das logistische Regressionsmodell zwar auf den bestehenden Datensatz gut angepasst, lässt sich jedoch auf die Grundgesamtheit schwer übertragen. Es besteht die Gefahr, dass statistisch irrelevante Variablen aufgrund von Zufallseffekten als signifikant erscheinen und damit die tatsächliche Anpassungsgüte verschleiert wird.

Wichtig für eine gute Modellanpassung ist die Auswahl der Stichproben und damit einhergehend möglichst repräsentative Referenzdaten. Die Residuendevianz („goodness-of-fit“) vergleicht das geschätzte Modell mit dem größtmöglichen Modell und ist im multinomialen logistischen Regressionsmodell zur Trennung aller Baumarten mit einem Wert von 2,77259 sehr niedrig. Auch der niedrige AIC, ein in Abschnitt 5.2.2.3 beschriebenes Gütemaß des logistischen Regressionsmodells weist mit dem niedrigsten Wert (96,00001) aller Kombinationen von Prädiktoren auf eine gute Modellanpassung hin. Je kleiner der AIC, desto besser ist das logistische Regressionsmodell „gefittet“. Keine andere Kombination der erklärenden Variablen führt zu einem niedrigeren AIC.

### **8.3.1.6. Conditional Density Plots des multinomialen logistischen Regressionsmodells**

Die folgenden Plots zeigen die Verteilung der kategorischen Zielvariable Baumart in Hinblick auf die unabhängigen Modellvariablen, die Mittelwerte von ausgewählten Kanälen der Hyperspektralaufnahme. Die Wahrscheinlichkeit, dass ein Baum mit gegebenem Reflexionswert im entsprechenden Kanal einer Baumart zugeordnet werden kann, lässt sich in diesen Plots auf der Y-Achse ablesen. Auf der X-Achse ist der Reflexionswert im jeweiligen Kanal der Hyperspektralaufnahme abgebildet.

Die Kurvenform repräsentiert den grundsätzlichen Aufbau der multinomialen logistischen Regression, über ausgewählte Prädiktoren wird die Zugehörigkeits-Wahrscheinlichkeit für ein Objekt zu einer gewissen Gruppe, in dieser Studie die Baumart, modelliert. Vorgestellt werden drei Prädiktoren, Kanal 7 aus dem sichtbaren Bereich des Lichts, Kanal 33 aus dem Red-Edge-Bereich des Lichts und Kanal 57 aus dem nahen Infrarot.

## ❖ Kanal 7

Kanal 7 ist ein Prädiktor des multinomialen logistischen Regressionsmodells, welcher im sichtbaren Bereich des Lichtes angesiedelt ist. Mit einer Wellenlänge von 464,7 nm ist Kanal 7 dem blauen Licht zuzuordnen. Ein Baumpolygon mit einem mittleren Reflexionswert von 400 in Kanal 7 ist Abbildung 22 zu Folge zu ~10% eine Ahorn, zu ~30% eine Buche, zu ~10% eine Eiche, zu ~25% eine Fichte, zu ~5% eine Kiefer, zu ~18% eine Lärche und zu ~2% eine Tanne. Ein Baumpolygon mit einem mittleren Reflexionswert von 500 ist zu ~3% eine Ahorn, zu ~15% eine Buche, zu ~1% eine Eiche, zu ~2% eine Fichte, zu ~60% eine Kiefer, zu ~20% eine Lärche und zu ~0% eine Tanne. Man vergleiche dazu den Mean Plot der Referenzbäume (Abbildung 19) in Abschnitt 8.1.2.

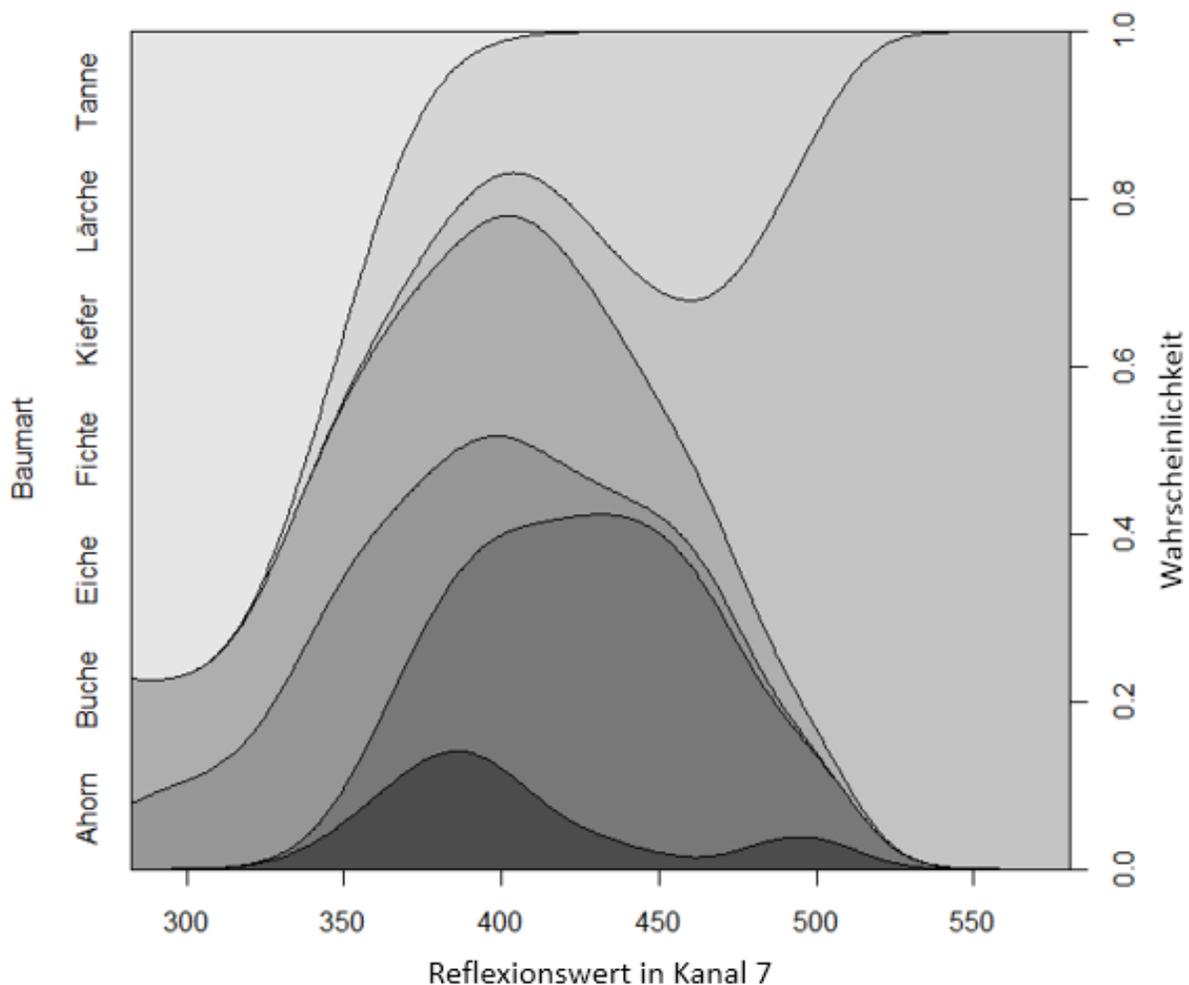


Abbildung 22: Conditional Density Plot - Kanal 7

## ❖ Kanal 33

Kanal 33 ist mit einer Wellenlänge von 703,6 nm im Red-Edge-Bereich (Übergang Rot -> nahes Infrarot) des Lichtes angesiedelt. Das Erscheinungsbild des Conditional Density Plots von Kanal 33 in Abbildung 23 unterscheidet sich merklich vom Plot des Kanals 7 in Abbildung 22, die Kurvenform der logistischen Regression ist trotzdem präsent. Baumpolygone mit einem mittleren Reflexionswert von 300 in Kanal 33 sind zu ~0% eine Ahorn, zu ~0% eine Buche, zu ~38% eine Eiche, zu ~12% eine Fichte, zu ~0% eine Kiefer, zu ~0% eine Lärche und zu ~50% eine Tanne. Tannen- und Eichen-Referenzbäume weisen Abbildung 19 (Kapitel 8.1.2) zu folge auch die niedrigsten Reflexionswerte in Kanal 33 auf. Man erkennt, dass das Reflexionsverhalten von Bäumen Nadelbaumarten im Red-Edge-Bereich nicht nadelbeziehungsweise laubwaldspezifisch ist.

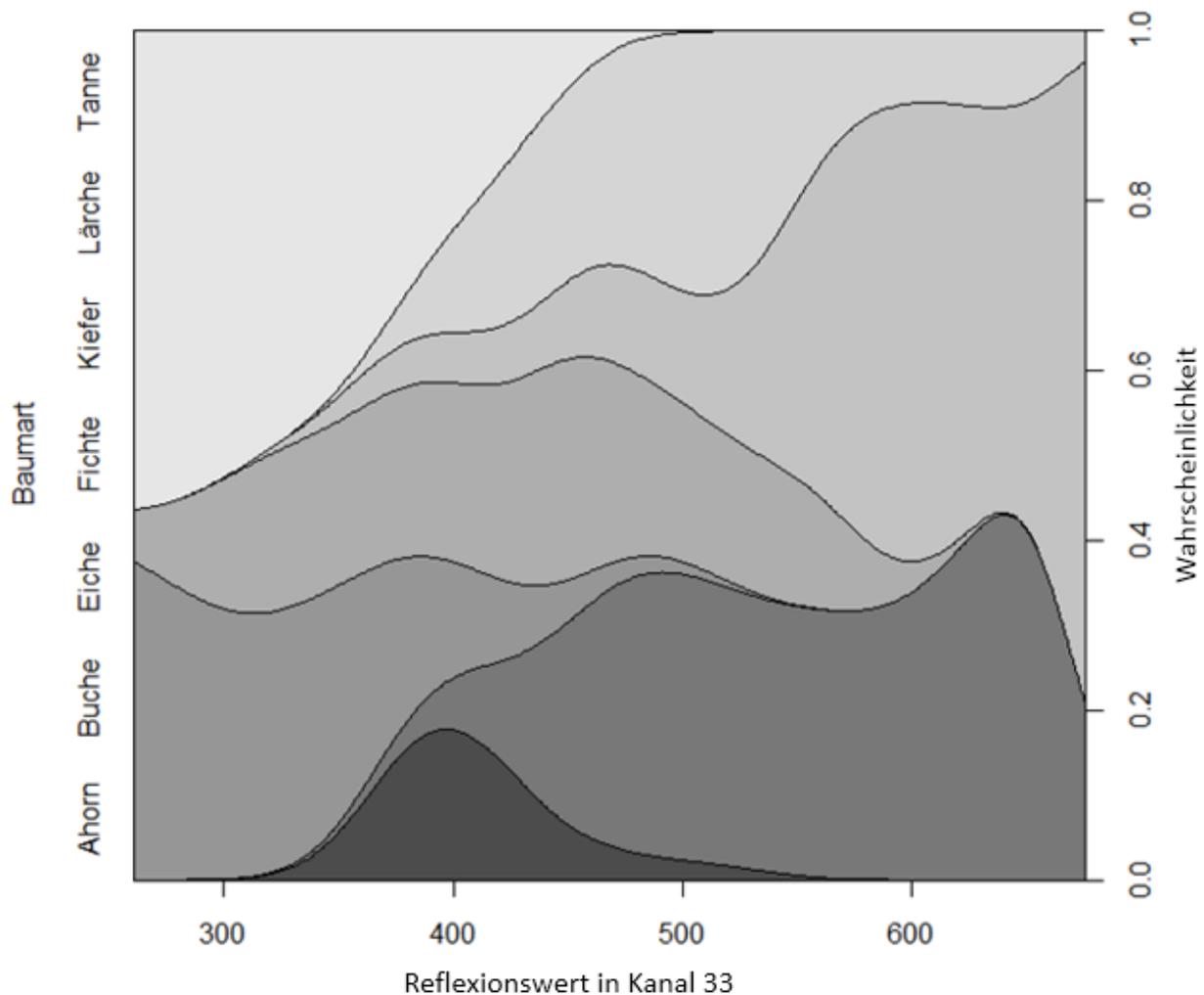


Abbildung 23: Conditional Density Plot - Kanal 33

## ❖ Kanal 62

Mit einer Wellenlänge von 970,4 nm befindet sich Kanal 62 im nahen Infrarot des elektromagnetischen Spektrums. Wie in Abbildung 19 in Kapitel 8.1.2 deutlich erkennbar, weisen die Laubbaumarten Buche, Eiche und Ahorn eine höhere Reflexion des Lichtes in Kanal 62 als die Nadelbaumarten Fichte, Kiefer, Lärche und Tanne auf. Ein Baum mit einem mittleren Reflexionswert von 1500 in Kanal 62 ist laut dem Conditional Density Plot in Abbildung 24 zu ~0% eine Ahorn, zu ~2% eine Buche, zu ~1% eine Eiche, zu ~20% eine Fichte, zu ~30% eine Kiefer, zu ~10% eine Lärche und zu ~37% eine Tanne. Reflexionswerte über 2500 in Kanal 62 kommen nur bei Laubbaumarten vor.

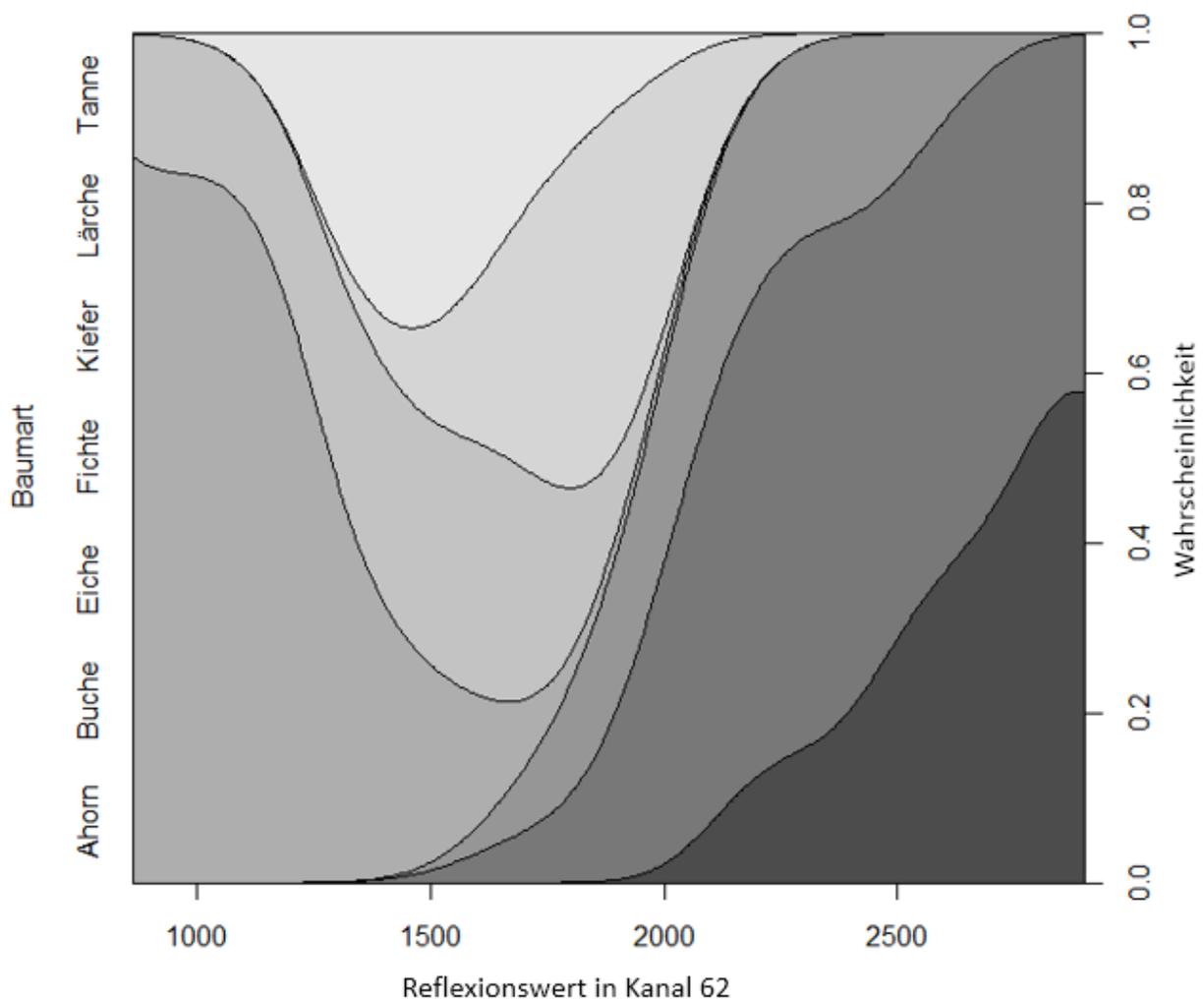


Abbildung 24: Conditional Density Plot - Kanal 62

### **8.3.2. Generalized Linear Models (GLM) zur paarweisen Baumartendifferenzierung**

Um die direkte Trennbarkeit von ausgewählten Baumarten zu evaluieren, werden aus den Referenzbäumen des Streifens „0908“ Subsets mit jeweils zwei Baumarten erstellt. Darauf aufbauend wird die paarweise Trennung ausgewählter Baumarten zum einen mit ausschließlich LiDAR-Variablen, zum anderen über eine Kombination von LiDAR- und Hyperspektralvariablen untersucht.

Dazu wird analog zur Methodik in Abschnitt 8.3.1 ein GLM mit allen ausgewählten Prädiktoren(LiDAR oder LiDAR+Hyperspektral) erstellt und über eine „Stepwise model selection“ wird die beste Kombination von Prädiktoren zur Trennung der jeweils untersuchten Baumarten gesucht.

Zur Auswertung wird das Klassifikationsergebnis der Referenzbäume in Form einer Kontingenzmatrix dargestellt. Über eine LOOCV- Kreuzvalidierung wird der Kreuzvalidierungsfehler „delta“ bestimmt, ein wichtiger Parameter für die Güte des logistischen Regressionsmodells.

Auf den folgenden Seiten werden die Ergebnisse der paarweisen Baumartendifferenzierung für folgende Baumarten-Paare erörtert:

- Fichte – Kiefer
- Fichte – Lärche
- Fichte – Tanne
- Kiefer – Lärche
- Buche – Eiche
- Buche – Fichte
- Buche – Kiefer
- Buche – Lärche

## Fichte - Kiefer

### ❖ Datengrundlage LiDAR

Modellgleichung: Baumart ~ STDV

Kontingenztabelle

Baumart	Fichte	Kiefer	Genauigkeit
Fichte	30	20	60,00%
Kiefer	18	28	60,87%

Gesamtgenauigkeit: 60,41%

LOOCV-Kreuzvalidierungsfehler: 24,22%

Der signifikanteste Prädiktor zur statistischen Trennung von Fichte und Kiefer auf Basis von LiDAR-Variablen ist die Standardabweichung der Höhenwerte innerhalb eines Baumpolygons(STDV). Damit können Fichten zu 60% und Kiefer zu 60,87% der richtigen Klasse zugeordnet werden. Dies ergibt eine Gesamtgenauigkeit von 60,41% und das GLM-Regressionsmodell hat einen Kreuzvalidierungsfehler von 24,22%.

### ❖ Datengrundlage LiDAR- und Hyperspektraldaten

Modellgleichung: Baumart ~ Mean7 + Mean22

Kontingenztabelle

Baumart	Fichte	Kiefer	Genauigkeit
Fichte	50	0	100%
Kiefer	0	46	100%

Gesamtgenauigkeit: 100%

LOOCV-Kreuzvalidierungsfehler: ~0%

Die Trennung von Fichte und Kiefer auf Basis von LiDAR- und Hyperspektraldaten ist mit einer Gesamtgenauigkeit von 100% und einem Kreuzvalidierungsfehler von ~0% sehr gut möglich. Die Modellgleichung besteht aus einer Linearkombination der Hyperspektralvariablen Mean7 und Mean22, welche bei der statistischen Differenzierung zwischen den Baumarten Fichte und Kiefer die höchste Signifikanz aufweisen. Mean7 und Mean22 sind Kanal-Mittelwerte im sichtbaren Bereich des elektromagnetischen Spektrums.

## Fichte – Lärche

### ❖ Datengrundlage LiDAR

Modellgleichung: Baumart ~ STDV + STDV\_filter

Kontingenztmatrix

<b>Baumart</b>	Fichte	Lärche	<b>Genauigkeit</b>
Fichte	42	8	84,00%
Lärche	12	32	72,73%

Gesamtgenauigkeit: 78,72%

LOOCV-Kreuzvalidierungsfehler: 18,08%

Die statistische Trennung von Fichte und Lärche auf Basis von LiDAR-Variablen wird über eine Linearkombination der Variablen STDV und STDV\_filter vollzogen. Fichten können damit zu 84% und Lärchen zu 72,73% der richtigen Zielklasse zugeordnet werden. Dies ergibt eine Gesamtgenauigkeit von 78,72% und einen Kreuzvalidierungsfehler von 18,08%.

### ❖ Datengrundlage LiDAR- und Hyperspektraldaten

Modellgleichung: Baumart ~ Mean59 + Mean65 + Mean62

Kontingenztmatrix

<b>Baumart</b>	Fichte	Lärche	<b>Genauigkeit</b>
Fichte	50	0	100%
Lärche	0	44	100%

Gesamtgenauigkeit: 100%

LOOCV-Kreuzvalidierungsfehler: 0,28%

Auf Basis von Variablen aus LiDAR- und Hyperspektraldaten können Fichten und Lärchen mit einer Gesamtgenauigkeit von 100% und einem Kreuzvalidierungsfehler von 0,28% statistisch getrennt werden. Dieses Ergebnis wird über eine Linearkombination der drei Hyperspektralvariablen Mean59, Mean65 und Mean62 erzielt. Dies sind Kanal-Mittelwerte aus dem nahen Infrarot-Bereich des elektromagnetischen Spektrums.

## Fichte - Tanne

### ❖ Datengrundlage LiDAR

Modellgleichung: Baumart  $\sim$  STDV\_filter

Kontingenzmatrix

<b>Baumart</b>	Fichte	Tanne	<b>Genauigkeit</b>
Fichte	33	17	66,00%
Tanne	19	24	55,81%

Gesamtgenauigkeit: 61,29%

LOOCV-Kreuzvalidierungsfehler: 24,63%

Der beste Prädiktor zur statistischen Trennung von Fichten und Tannen auf Basis von LiDAR-Variablen ist STDV\_filter, die gefilterte Standardabweichung der Höhenwerte innerhalb eines Baumpolygons. Damit können Fichten zu 66% und Kiefer zu 55,81% der richtigen Klasse zugeordnet werden. Dies ergibt eine Gesamtgenauigkeit von 61,29% und einen Kreuzvalidierungsfehler von 24,63%.

### ❖ Datengrundlage LiDAR- und Hyperspektraldaten

Modellgleichung: Baumart  $\sim$  Mean62 + Mean49 + Mean55 + Mean1

Kontingenzmatrix

<b>Baumart</b>	Fichte	Tanne	<b>Genauigkeit</b>
Fichte	50	0	100%
Tanne	0	43	100%

Gesamtgenauigkeit: 100%

LOOCV-Kreuzvalidierungsfehler: 3,36%

Wie bei den vorherigen paarweisen Untersuchungen, bei welchen Variablen aus LiDAR- und Hyperspektraldaten verwendet werden, sind in der Modellgleichung ausschließlich Hyperspektralvariablen zu finden. Die Prädiktoren Mean62, Mean49 und Mean55 sind im nahen Infrarot angesiedelt. Mean1 ist im blauen Bereich des sichtbaren Lichts. Mit der Linearkombination der beschriebenen Hyperspektralvariablen können alle untersuchten Fichten und Tannen der korrekten Zielklasse zugewiesen werden.

## Kiefer - Lärche

### ❖ Datengrundlage LiDAR

Modellgleichung: Baumart  $\sim$  STDV + STDV\_filter

Kontingenztmatrix

<b>Baumart</b>	Kiefer	Lärche	<b>Genauigkeit</b>
Kiefer	38	8	82,61%
Lärche	7	37	84,09%

Gesamtgenauigkeit: 83,33%

LOOCV-Kreuzvalidierungsfehler: 14,77%

Kiefer und Lärchen werden auf Basis von LiDAR-Variablen über eine Linearkombination der Variablen STDV und STDV\_filter am besten statistisch getrennt. Damit können 82,61% aller Kiefer und 84,09% aller Lärchen der richtigen Klasse zugeordnet werden. Dies ergibt eine Gesamtgenauigkeit von 83,33% und einen Kreuzvalidierungsfehler von 14,77%.

### ❖ Datengrundlage LiDAR- und Hyperspektraldaten

Modellgleichung: Baumart  $\sim$  Mean41 + Mean4

Kontingenztmatrix

<b>Baumart</b>	Kiefer	Lärche	<b>Genauigkeit</b>
Kiefer	46	0	100%
Lärche	0	44	100%

Gesamtgenauigkeit: 100%

LOOCV-Kreuzvalidierungsfehler:  $\sim$ 0%

Kiefer und Lärchen werden auf Basis von LiDAR- und Hyperspektralvariablen über die Linearkombination der Hyperspektralvariablen „Mean41“ und „Mean4“ statistisch getrennt. Mean41 ist im Red-Edge-Bereich des elektromagnetischen Spektrums, Mean4 im blauen Bereich des sichtbaren Lichts. Mit dieser Modellgleichung können 100% der untersuchten Kiefer und Lärchen der richtigen Klasse zugeordnet werden.

## Buche - Eiche

### ❖ Datengrundlage LiDAR

Modellgleichung: Baumart ~ STDV\_filter

Kontingenztabelle

<b>Baumart</b>	Buche	Eiche	<b>Genauigkeit</b>
Buche	50	6	89,29%
Eiche	16	12	42,86%

Gesamtgenauigkeit: 73,81%

LOOCV-Kreuzvalidierungsfehler: 21,56%

Als signifikantester Prädiktor für die Trennung von Buche und Eiche auf Basis von LiDAR-Variablen wurde die gefilterte Standardabweichung der Höhenwerte innerhalb eines Baumpolygons errechnet. Buchen werden damit zu 89,29% richtig klassifiziert, Eichen zu 42,86%. Dies ergibt eine Gesamtgenauigkeit von 73,91% und einen Kreuzvalidierungsfehler von 21,56%.

### ❖ Datengrundlage LiDAR- und Hyperspektraldaten

Modellgleichung: Baumart ~ Mean33 + Mean14 + Mean1

Kontingenztabelle

<b>Baumart</b>	Buche	Eiche	<b>Genauigkeit</b>
Buche	56	0	100%
Eiche	0	28	100%

Gesamtgenauigkeit: 100%

LOOCV-Kreuzvalidierungsfehler: 3,31%

Die Trennung von Buche und Eiche auf Basis von LiDAR- und Hyperspektraldaten ist mit einer Gesamtgenauigkeit von 100% und einem Kreuzvalidierungsfehler von 3,31% gut durchführbar. Als signifikantesten Prädiktoren werden die drei Hyperspektralvariablen Mean33, Mean14 und Mean1 identifiziert. Dies sind Kanal-Mittelwerte im sichtbaren Bereich des elektromagnetischen Spektrums.

## Buche - Fichte

### ❖ Datengrundlage LiDAR

Modellgleichung: Baumart ~ Slope\_mean

Kontingenztabelle

<b>Baumart</b>	Buche	Fichte	<b>Genauigkeit</b>
Buche	45	11	80,36%
Fichte	14	36	72,00%

Gesamtgenauigkeit: 76,42%

LOOCV-Kreuzvalidierungsfehler: 16,97%

Die LiDAR-Variable Slope\_mean in der Modellgleichung ist die durchschnittliche Steigung innerhalb eines Baumpolygons. Slope\_mean erweist sich zur Trennung von Buchen und Fichten auf Basis von LiDAR-Variablen als am signifikantesten. Buchen werden damit zu 80,36% Fichten zu 72% der richtigen Klasse zugewiesen. Daraus ergibt sich eine Gesamtgenauigkeit von 76,42% und der Kreuzvalidierungsfehler beträgt 16,97%.

### ❖ Datengrundlage LiDAR- und Hyperspektraldaten

Modellgleichung: Baumart ~ Mean37 + Mean12

Kontingenztabelle

<b>Baumart</b>	Buche	Fichte	<b>Genauigkeit</b>
Buche	56	0	100%
Fichte	0	50	100%

Gesamtgenauigkeit: 100%

LOOCV-Kreuzvalidierungsfehler: ~0%

Auch die statistische Differenzierung von Buche und Fichte auf Basis von LiDAR- und Hyperspektraldaten führt mit einer Gesamtgenauigkeit von 100% und einem Kreuzvalidierungsfehler von 3,31% zu einem sehr guten Ergebnis. Als signifikantesten Prädiktoren sind die beiden Hyperspektralvariablen Mean37 und Mean12 errechnet worden. Mean12 ist im grünen Bereich des sichtbaren Lichts angesiedelt, Mean37 befindet sich im Red-Edge-Bereich.

## Buche - Kiefer

### ❖ Datengrundlage LiDAR

Modellgleichung: Baumart  $\sim$  Slope\_stdv

Kontingenzmatrix

<b>Baumart</b>	Buche	Kiefer	<b>Genauigkeit</b>
Buche	43	13	76,79%
Kiefer	14	32	69,57%

Gesamtgenauigkeit: 73,53%

LOOCV-Kreuzvalidierungsfehler: 22,57%

Die LiDAR-Variable Slope\_stdv in der Modellgleichung ist die Varianz der einzelnen Steigungswerte innerhalb eines Baumpolygons. Slope\_stdv erweist sich zur Trennung von Buchen und Kiefer auf Basis von LiDAR-Variablen als bester Prädiktor. Buchen werden damit zu 76,79% und Kiefer zu 69,57% der richtigen Klasse zugewiesen. Daraus ergibt sich eine Gesamtgenauigkeit von 73,53% und der Kreuzvalidierungsfehler beträgt 22,57%.

### ❖ Datengrundlage LiDAR- und Hyperspektraldaten

Modellgleichung: Baumart  $\sim$  Mean41 + Mean17

Kontingenzmatrix

<b>Baumart</b>	Buche	Kiefer	<b>Genauigkeit</b>
Buche	56	0	100%
Kiefer	0	46	100%

Gesamtgenauigkeit: 100%

LOOCV-Kreuzvalidierungsfehler:  $\sim$ 0%

Auf Basis von Variablen aus LiDAR- und Hyperspektraldaten können Buchen und Kiefer mit einer Gesamtgenauigkeit von 100% und einem Kreuzvalidierungsfehler von  $\sim$ 0% statistisch differenziert werden. Dieses Ergebnis wird über eine Linearkombination der beiden Hyperspektralvariablen Mean41 und Mean17 erzielt. Mean17 ist im grünen Bereich, Mean41 im Red-Edge-Bereich des elektromagnetischen Spektrums angesiedelt.

## Buche - Lärche

### ❖ Datengrundlage LiDAR

Modellgleichung: Baumart ~ STDV

Kontingenztabelle

<b>Baumart</b>	Buche	Lärche	<b>Genauigkeit</b>
Buche	52	4	92,86%
Lärche	8	36	81,82%

Gesamtgenauigkeit: 88,00%

LOOCV-Kreuzvalidierungsfehler: 10,15%

Die statistische Trennung der Baumarten Buche und Lärche liefert mit einer Gesamtgenauigkeit von 88% das beste Ergebnis aller paarweisen Differenzierungen auf Basis von LiDAR-Variablen. Der signifikanteste Modellprädiktor zur Trennung der beiden Baumarten ist die Standardabweichung der Höhenwerte innerhalb eines Baumpolygons STDV. Der LOOV-Kreuzvalidierungsfehler ist im Vergleich zu allen anderen paarweisen Untersuchungen mit Datengrundlage LiDAR mit einem Wert von 10,15% relativ gering.

### ❖ Datengrundlage LiDAR- und Hyperspektraldaten

Modellgleichung: Baumart ~ Mean26 + Mean7

Kontingenztabelle

<b>Baumart</b>	Buche	Lärche	<b>Genauigkeit</b>
Buche	56	0	100%
Lärche	0	44	100%

Gesamtgenauigkeit: 100%

LOOCV-Kreuzvalidierungsfehler: ~0%

Eine Linearkombination der beiden Hyperspektralvariablen Mean26 und Mean7 liefert zur Trennung der Baumarten Buche und Lärche das beste Ergebnis. Mean26 ist im roten Bereich und Mean7 befindet sich im blauen Bereich des sichtbaren Lichts. Mit dieser Kombination aus Hyperspektralvariablen werden 100% der untersuchten Buchen und Lärchen der korrekten Zielklasse zugewiesen.

#### **8.4. Diskussion der Trennbarkeit der Baumarten**

Die Signifikanz von LiDAR- und Hyperspektralvariablen zur statistischen Trennung aller sechs untersuchten Baumarten wird über eine „Stepwise model selection“ untersucht. Die besten Prädiktoren zur Differenzierung aller Baumarten sind sieben Mittelwerte von Kanälen des Hyperspektraldatensatzes. Diese sieben Variablen werden als Modelprädiktoren für das multinomiale logistische Regressionsmodell verwendet. Mit einer Erstellergenauigkeit von 100% und einem sehr geringen Kreuzvalidierungsfehler von 3,18% liefert dieses Modell ein sehr gutes Ergebnis.

Es stellt sich die Frage, warum ausschließlich Prädiktoren aus dem Hyperspektraldatensatz und keine Variablen aus den LiDAR-Daten wie zum Beispiel die Standardabweichung der Höhenwerte oder die mittlere Steigung zur Trennung aller Baumarten signifikant sind. Dies ist darauf zurückzuführen, dass durch die hohe räumliche sowie spektrale Auflösung der Hyperspektraldaten Baumarten über ihre spektrale Charakteristik sehr gut differenziert werden können und eine Trennung einzelner Nadel- und Laubbaum-Spezies über LiDAR-Variablen schwierig ist. Die beiden Nadelbaumarten Fichte und Tanne haben zum Beispiel einen sehr ähnlichen Kronenaufbau und konnten paarweise mit einer Erstellergenauigkeit von 61,29% nicht erfolgreich über LiDAR-Variablen getrennt werden. Eine statistische Differenzierung der beiden Baumarten über die spektrale Charakteristik, abgeleitet aus den hochauflösenden Hyperspektraldaten, ist jedoch sehr wohl möglich.

Nadelbäume und Laubbäume haben eine spezifische Kronenform, was sich in den variablen Steigungen und Standardabweichungen der Höhenwerte niederschlägt. Nadelbäume weisen aufgrund der spitzen Kronenform eine tendenziell höhere durchschnittliche Steigung und eine höhere Standardabweichung der Höhenwerte als Laubbäume auf. Dies ist in Tabelle 2 in Kapitel 8.2 vor allem zwischen der Baumart Buche und allen anderen Nadelbaumarten zu beobachten. Für die Laubbaumart Eiche trifft dies nur bedingt zu, was möglicherweise auf die knappe Stichprobenanzahl von 28 Individuen zurückzuführen ist.

Um die grundsätzliche Eignung der vier LiDAR-Variablen zur Baumartendifferenzierung zu evaluieren, wird die statistische Differenzierung ausgewählter Baumarten paarweise untersucht. Dazu wird über ein GLM die Signifikanz der vier bereits beschriebenen LiDAR-Variablen STDV, STDV\_filter, Slope\_mean und Slope\_stdv für jedes Baumarten-Paar berechnet. Die LiDAR-Variablen werden dabei nicht nur separat betrachtet, sondern es sind auch Linearkombinationen der Variablen zulässig.

Das beste Ergebnis lieferte die Trennung von Buche/Lärche über die Variable STDV, die Standardabweichung der Höhenwerte aus dem LiDAR-Datensatz. Mit einer Gesamtgenauigkeit von 88% bei der Klassifikation der Referenzbäume und einem LOOCV-Kreuzvalidierungsfehler von 10,15% konnten die beiden Baumarten erfolgreich über eine LiDAR-Variable statistisch getrennt werden.

Überraschend lieferte auch die Trennung der beiden Nadelbaumarten Kiefer/Lärche über eine Linearkombination der beiden LiDAR-Variablen STDV und STDV\_filter mit einer Erstellergenauigkeit von 83,3% und einem Kreuzvalidierungsfehler von 14,77% ein gutes Ergebnis. Betrachtet man die LiDAR-Variablen der Baumart Lärche genauer, so erkennt man in Tabelle 2 in Kapitel 8.2, dass Lärchen eine besonders hohe Standardabweichung der Höhenwerte aufweisen und damit gut von anderen Baumarten trennbar sind. Die Baumarten Fichte/Kiefer können mit einer Gesamtgenauigkeit von 60,41% und einem Kreuzvalidierungsfehler von 24,22% nicht erfolgreich über LiDAR-Variablen getrennt werden. Dies ist auf den ähnlichen Kronenaufbau der beiden Nadelbaumarten zurückzuführen.

Bei der paarweisen Trennung ausgewählter Baumarten über eine Kombination von LiDAR- und Hyperspektralvariablen konnten für alle Baumarten-Paare Erstellergenauigkeiten von 100% erzielt werden. Nimmt man als Input für die GLM's ausschließlich Hyperspektralvariablen, so erhält man die gleichen Modellgleichungen und Erstellergenauigkeiten für die jeweils untersuchten Baumarten-Paare. Die vier LiDAR-Variablen liefern demnach keinen statistischen Mehrwert zur paarweisen Trennung der ausgewählten Baumarten. Trotzdem konnte gezeigt werden, dass über ein logistisches Regressionsmodell eine statistische Trennung von Baumarten über Formparameter, abgeleitet aus dem LiDAR-Datensatz, grundsätzlich möglich ist und bei ausgewählten Baumarten gute Ergebnisse liefert.

## 9. Zusammenfassung

Diese Arbeit erörtert die Eignung der Methode der logistischen Regression zur Differenzierung von Baumarten auf Basis von Hyperspektral- und LiDAR-Fernerkundungsdaten. In fünf Feldbegehungen wurden 451 Referenzbäume im Projektgebiet St. Margarethen an der Raab im Bezirk Weiz erhoben und anschließend in einem GIS digitalisiert. Dazu wurde eine kontrastverstärkte 7-Kanal-Aufnahme verwendet. Ein aus dem LiDAR-Datensatz erstelltes nDSM ist bei der Digitalisierung der Referenzbäume sehr hilfreich, da damit die Abgrenzung von Individuen und deren Baumkrone erleichtert wird.

Über eine Signaturanalyse wird das spektrale Reflexionsverhalten der Baumarten im Untersuchungsstreifen „0908“ erörtert. Aus den Hyperspektraldaten werden die Reflexionsmittelwerte aller 65 Kanäle, bezogen auf das jeweilige Baumpolygon, als Variablen zur Baumartendifferenzierung abgeleitet. Die vier aus dem LiDAR-Datensatz abgeleiteten Variablen sind zwei verschiedene Standardabweichungen der Höhenwerte (STDV und STDV\_filter), der Mittelwert der Steigung (Slope\_mean) und die Standardabweichung der Steigung pro Baumpolygon (Slope\_stdv). Diese Variablen sind Indikatoren für den baumartenspezifischen vertikalen Aufbau der Baumkrone.

Untersucht wird die Trennbarkeit der Nadelbaumarten Fichte, Kiefer, Lärche und Tanne und der Laubbaumarten Buche und Eiche. Basierend auf 65 Hyperspektral-Variablen und vier LiDAR-Variablen wird im Kapitel 8.3.1 ein multinomiales logistisches Regressionsmodell erstellt, über welches die besten Prädiktoren zur Trennung aller untersuchten Baumarten gesucht werden. Als beste Prädiktoren zur Differenzierung aller sechs Baumarten werden folgende sieben Hyperspektral-Variablen über eine „Stepwise model selection“ eruiert:

- ❖ Kanal 7: 463-472 nm, blaues Licht
- ❖ Kanal 13: 517-526 nm, grünes Licht
- ❖ Kanal 26: 634-643 nm, rotes Licht
- ❖ Kanal 33: 697-706 nm, Red-Edge-Bereich
- ❖ Kanal 41: 778-787 nm, Nahes Infrarot
- ❖ Kanal 57: 913-922 nm, Nahes Infrarot
- ❖ Kanal 62: 958-967 nm, Nahes Infrarot

Auf Basis des logistischen Regressionsmodells werden die Zugehörigkeits-Wahrscheinlichkeiten der Referenzbäume für alle Zielklassen berechnet. Mit dem logistischen Regressionsmodell, welches aus sieben Kanälen des Hyperspektraldatensatzes aufgebaut ist, können 100% der Referenzbäume der korrekten Zielklasse zugeordnet werden. Über eine LOOCV-Kreuzvalidierung wird der Kreuzvalidierungsfehler „delta“ bestimmt. Der sehr geringe Wert von 3,18% bedeutet, dass nur 3,18% der Referenzbäume der falschen Baumart zugeordnet werden. Die statistische Differenzierung der sechs Baumarten über ein multinomiales logistisches Regressionsmodell mit sieben ausgewählten Kanälen des Hyperspektraldatensatzes ist folglich sehr gut möglich.

Da Variablen aus dem LiDAR-Datensatz, welche den baumartenspezifischen Kronenaufbau beschreiben, im multinomialen logistischen Regressionsmodell keine Signifikanz zeigen, wird die paarweise Trennbarkeit von Baumarten über ausschließlich LiDAR-Variablen in Abschnitt 8.3.2 untersucht. Dazu wird für Referenzbäume der beiden untersuchten Baumarten ein GLM mit allen vier LiDAR-Variablen erstellt und über eine „Stepwise model selection“ der beste Prädiktor, beziehungsweise die beste Kombination von Prädiktoren gesucht. Tabelle 6 zeigt eine Zusammenfassung der paarweisen Baumartendifferenzierung auf Basis der LiDAR-Variablen. Man erkennt, dass von der Kronenform sehr ähnliche Baumarten wie Fichte/Tanne sehr schwer trennbar sind. Ausgewählte Baumarten mit sehr unterschiedlichem Kronenaufbau, wie zum Beispiel Buche/Lärche, können mit einer Genauigkeit von 88% und einem Kreuzvalidierungsfehler von 10,15% gut über LiDAR-Variablen differenziert werden.

**Tabelle 6: Paarweise Baumartendifferenzierung mit LiDAR-Variablen**

Baumarten	Prädiktor	Erstellergenauigkeit	Kreuzvalidierungsfehler
Fichte/Kiefer	STDV	60,41%	24,22%
Fichte/Lärche	STDV + STDV_filter	78,72%	18,08%
Fichte/Tanne	STDV_filter	61,29%	24,63%
Buche/Fichte	Slope_mean	76,42%	16,97%
Buche/Kiefer	Slope_stdv	73,53%	22,57%
Buche/Lärche	STDV	88,00%	10,15%
Buche/Eiche	STDV_filter	73,81%	21,56%
Kiefer/Lärche	STDV + STDV_filter	83,33%	14,77%

Zum Vergleich wird eine paarweise Baumartendifferenzierung mit Variablen aus LiDAR- und Hyperspektraldaten durchgeführt. Alle Referenzbäume werden dabei der korrekten Zielklasse zugewiesen und bei keiner Kombination von Baumarten überschreitet der Kreuzvalidierungsfehler einen Wert von 4%. Wie bei der Trennung aller sechs Baumarten über ein multinomiales logistisches Regressionsmodell zeigen Hyperspektralvariablen eine höhere Signifikanz als LiDAR-Variablen.

Nimmt man als Input für die Regressionsmodelle ausschließlich Hyperspektralvariablen, so erhält man die gleichen Modellgleichungen und Prädiktoren für die jeweils untersuchten Baumarten-Paare. Die vier LiDAR-Variablen liefern demnach keinen statistischen Mehrwert zur paarweisen Trennung der ausgewählten Baumarten.

Zusammenfassend wird in dieser Arbeit aufgezeigt, dass es auf Basis der vorliegenden LiDAR- und Hyperspektraldaten möglich ist, einzelne Baumarten im Projektgebiet statistisch zu trennen. Mittelwerte von ausgewählten Kanälen des Hyperspektraldatensatzes eignen sich zur statistischen Trennung der erhobenen Referenzbäume besser als LiDAR-Variablen. Dies ist darauf zurückzuführen, dass über die sehr hohe räumliche sowie spektrale Auflösung der Hyperspektraldaten das baumartenspezifische Reflexionsverhalten sehr gut quantifiziert werden kann. Bei der paarweisen Baumartendifferenzierung wird gezeigt, dass eine statistische Trennung ausgewählter Baumarten, welche einen unterschiedlichen Kronenaufbau haben, über LiDAR-Variablen möglich ist.

## 10. Literaturverzeichnis

ALBERTZ, *"Einführung in die Fernerkundung"*, Darmstadt, 2001

AUSTIN & TU, *"Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality"*, Journal of Clinical Epidemiology, 2004

BALTSAVIAS, EISENBEISS, AKCA, WASER, KÜCHLER, GINZLER, THEE, *"Modeling fractional shrub/tree cover and multitemporal changes using high-resolution digital surface models and CIR-aerial images"*, DGPF Tagungsband 16/2007

BEHERA & ROY, *"Lidar remote sensing for forestry applications: The Indian context"*, Current Science, Vol.82 No.11, 2002

DALPONTE & BRUZZONE & GIANELLE, *"Tree species classification the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data"*, Remote Sensing of Environment Vol. 123, 2012

DALPONTE, *"Tree Species Classification in boreal forests with Hyperspectral Data"*, Geoscience and Remote Sensing Vol. 51, 2013

EVANS, ROBERTS, MCCOMBS, HARRINGTON, *"Detection of regulary spaced targets in small-footprint LIDAR data"*, Photogrammetric Engineering & Remote Sensing, Vol.67, No.10, 2001

GERNAND, *"Understanding AIC and BIC in Model Selection"*, Ludwig-Maximilians-Universität München: Institut für Statistik, 2009

GINZLER et al., *"Luftbildinterpretation LFI3 – Interpretationsanleitung zum dritten Landesforstinventar"*, Eidgenössische Forschungsanstalt für Wald, Schnee und Landschaft, 2004

GOSH et al., *"A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales"*, International Journal of Applied Earth Observation and Geoinformation, Vol 26, Page 49-63

GUALTIERI & CROMP, "*Support Vector Machines for Hyperspectral Remote Sensing Classification*", Applied Information Sciences and Global Science and Technology Code 935, 1998

HAUK & SCHADAUER, "*Instruktion für die Feldarbeit der Österreichischen Waldinventur 2007 – 2009*", Bundesforschungszentrum für Wald 2009  
HYYPÄ J., HYYPÄ H., YUI, MALTAMO, "*Effects of flight altitude on tree height estimation using airborne laser scanning*", International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol XXXVI, 2004

IMMITZER, ATZBERGER, KOUKAL, "*Tree Species Classification with Random Forest using very high spatial resolution 8-Band WorldView-2 Satellite Data*", Journal of Remote Sensing 2012/4

IOST, "*Untersuchung der Eignung logistischer Regressionsmodelle zur Kartierung forstlicher Merkmale mit Satelliten-Fernerkundungsdaten*", Universität Hamburg – Fakultät für Mathematik, Informatik und Naturwissenschaften, 2006

JAN et al., "*Potentials of the VNIR airborne hyperspectral system Aisa EAGLE*", GIS Ostrava 2008

KANTOLA et al, "*Classification of defoliated trees using tree-level airborne Laser Scanning data combined with Aerial Images*", Remote Sensing ISSN 2072-4292, 2010

KNEUBÜHLER, "*Remote Sensing of Agriculture*", Remote Sensing Series 38., Zürich, 2002

LANE, "*Generalized linear models in soil science*", European Journal of Soil Science 53, 2002

LEFSKY, COHEN, PARKER, HARDING, "*Lidar Remote Sensing for Ecosystem Studies*", Bioscience Vol.52, 2002

LIM, TREITZ, WULDER, ST-ONGE, FLOOD, "*LiDAR remote sensing of forest structure*", Progress in Physical Geography 27, 2003

MIRGORODSKY, "Die Bedeutung der Fernerkundung in der Forstwirtschaft", Grin, 2006

NILSON, KUUSK, LANG, LÜKK, "Forest reflectance Modeling: Theoretical aspects and applications", *Ambio* Vol. 32, 2003

PANT, HEIKKINEN, KORPELA, HAUTA-KASARI, TOKOLA, "Logistic Regression-based spectral band selection for tree species classification: Effects of spatial scale and balance in Training samples"; *IEEE Geoscience and Remote Sensing Letters*, 2014

PUTONNEN, JAAKKOLA, LITKEY, HYYPPÄ, "Tree Classification with Fused Mobile Laser Scanning and Hyperspectral Data", *Sensors* (Basel) 2011

ROSSITER & LOZA, "Analyzing land cover change with logistic regression in R", University of Twente, Faculty of Geo-Information Science & Earth Observation (ITC), 2012

SCHADAUER, NIESE, MATZIK, "Österreichische Waldinventur Baumartenatlas", Website Bundesamt und Forschungszentrum für Wald:  
[http://bfw.ac.at/700/2092\\_1.html](http://bfw.ac.at/700/2092_1.html) , letzter Zugriff 11/2013

STILLA et al., "Full Waveform Laserscanning – Auswertemethoden und Anwendungen", Technische Universität München, 2009

USTIN, SMITH, JAQUEMOUD, VERSTRAETE, "Geobotany: vegetation Mapping for earth Sciences", *Remote sensing of the Earth Sciences*, 1999

WASER, EISENBEISS, KUECHLER, BALTSAVIAS, "Potential and limits of Airborne Remote Sensing data for extraction of fractional canopy cover and forest stands and detection of tree species", *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* Vol. XXXVII, 2008

WOLLSCHLÄGER, „Grundlagen der Datenanalyse mit R: eine anwendungsorientierte Einführung“; Springer Verlag, 2010

ZHANG & QIU, „Mapping individual tree species in an urban forest using airborne Lidar data and Hyperspectral Imagery“, *Photogrammetric Engineering & Remote Sensing* Vol. 78, 2012

## 11. Abbildungsverzeichnis

Abbildung 1: <i>Waldtypenkarte Österreich</i> .....	8
Abbildung 2: <i>Spektrales Reflexionsverhalten von Vegetation</i> .....	18
Abbildung 3: <i>Das elektromagnetische Spektrum</i> .....	20
Abbildung 4: <i>Das Prinzip der Hyperspektralfernerkundung</i> .....	21
Abbildung 5: <i>Schematisches Prinzip von LiDAR</i> .....	22
Abbildung 6: <i>Überblickskarte des Projektgebietes St. Margarethen an der Raab</i> ...	24
Abbildung 7: <i>Sensorspezifikation Aisa EAGLE</i> .....	27
Abbildung 8: <i>Normalisiertes digitales Oberflächenmodell (nDSM)</i> .....	28
Abbildung 9: <i>Digitales Geländehöhenmodell (DTM)</i> .....	28
Abbildung 10: <i>Komponenten des ArcGIS Informationssystems</i> .....	29
Abbildung 11: <i>Das logistische Regressionsmodell</i> .....	33
Abbildung 12: <i>Baumhöhen-Brusthöhendurchmesser Plot</i> .....	39
Abbildung 13: <i>Referenzdatenerhebung</i> .....	40
Abbildung 14: <i>Digitalisierung der Referenzbäume</i> .....	41
Abbildung 15: <i>7-Kanal-Aufnahme vor der Histogram Equalization</i> .....	43
Abbildung 16: <i>Kontrastverstärkte 7-Kanal-Aufnahme nach der Histogram Equalization</i> .....	44
Abbildung 17: <i>Geometrischer Versatz</i> .....	45
Abbildung 18: <i>DSM-Erstellung in LAStools</i> .....	46
Abbildung 19: <i>Mean Plot der Referenzbäume</i> .....	48
Abbildung 20: <i>Referenzbäume im spektralen Merkmalsraum</i> .....	49
Abbildung 21: <i>Erstellung eines multinomialen Logitmodells</i> .....	52
Abbildung 22: <i>Conditional Density Plot - Kanal 7</i> .....	57
Abbildung 23: <i>Conditional Density Plot - Kanal 33</i> .....	58
Abbildung 24: <i>Conditional Density Plot - Kanal 62</i> .....	59

## 12. Tabellenverzeichnis

Tabelle 1: Anzahl der erhobenen Referenzbäume .....	38
Tabelle 2: Zusammenfassung numerischer Variablen .....	50
Tabelle 3: Modellparameter des logistischen Regressionsmodells.....	53
Tabelle 4: Wald-Statistiken des logistischen Regressionsmodells .....	54
Tabelle 5: Kontingenzmatrix .....	55
Tabelle 6: Paarweise Baumartendifferenzierung mit LiDAR-Variablen .....	71