Universität für Bodenkultur Wien
Studienrichtung Kulturtechnik und Wasserwirtschaft

Diplomarbeit

Landbedeckungskartierung aus Satellitenbildern durch kNN-Klassifizierung am Beispiel Oberösterreich

ausgeführt am
Department für Raum, Landschaft und Infrastruktur
Institut für Vermessung, Fernerkundung
und Landinformation (IVFL)

betreut von
Univ.Prof.i.R. Dipl.-Ing. Dr.techn. Werner Schneider

Unter Mitbetreuung von
Dipl.-Ing. Dr.nat.techn. Tatjana Koukal

eingereicht von
Stefan Wegschaider
9940374

Wien, August 2011

INHALT

1. Einleitung 7
   1.1 Motivation 7
   1.2 Ziel der Arbeit 7
2. Datengrundlagen 8
   2.1 Satellitendaten 8
   2.2 Trainingsdaten 11
      2.2.1 Generierung der Trainingsdaten 12
      2.2.2 Verbesserung der Trainingsdaten 14
   2.3 Vergleichsdatensätze 18
      2.3.1 Österreichische Waldkarte 18
      2.3.2 Invekos Agrarförderflächen 19
      2.3.3 Aufbereitung der Daten auf Bezirksebene 20
3. kNN (k-Nearest-Neighbour) Klassifizierung 21
   3.1 kNN Methode 21
   3.2 kNN Programm 22
      3.2.1 Datenaufbereitung 22
      3.2.2 Read Image 23
      3.2.3 Cross Validation 23
      3.2.4 Map Maker 25
   3.3 Optimierung der kNN Klassifizierung 25
      3.3.1 Outlier (Ausreißer) 25
      3.3.2 Werte von k 27
      3.3.3 Suchradien (Lage und Höhe) 29
      3.3.4 Bandgewichtungen 30
      3.3.5 Zusammenfassende Betrachtung der Optimierungsschritte 34
   3.4 Klassifizierung des Satellitenbildes 38
      3.4.1 Bezirksweise Flächenstatistik 38
      3.4.2 Fuzzy Klassifizierung 39
4. Ergebnisse der kNN Klassifikation 42

4.1 Ergebnisse für das gesamte Bundesland 42
4.2 Ergebnisse der bezirksweisen Flächenstatistik für Wald und Ackerland 44
4.2.1 Vergleich der Klassifizierungsergebnisse mit der Österreichischen Waldkarte 45
4.2.2 Vergleich der Klassifizierungsergebnisse mit den Invekos Daten 49
4.2.3 Detaillierte Betrachtung einzelner wichtiger Bezirke 50
5. Landbedeckungskarte Oberösterreichs 57
6. Diskussion und Schlussfolgerung 59
Literaturverzeichnis / Quellenangabe 61
Abbildungsverzeichnis 62
Tabellenverzeichnis 63

Danksagung

Ich möchte mich besonders bei meinen Eltern bedanken. Sie haben meinen Entschluss zu Studieren nie in Frage gestellt und mich in dieser Zeit immer nach besten Kräften unterstützt.

Vielen Dank an das Institut für Vermessungswesen, Fernerkundung und Landinformation der Universität für Bodenkultur (BOKU). Im Besonderen an Prof. Dr. Werner Schneider und Dr. Tatjana Koukal für die Geduld bei der Betreuung meiner Diplomarbeit.

Erwähnen möchte ich auch meinen Kollegen Dipl.-Ing. Georg Schweidler, mit dem ich viele Vorbereitungen zu Prüfungen, Projektarbeiten und sonstige wichtige Momente des Studiums teilen durfte.

Besten Dank!
Stefan Wegschaider

Kurzfassung

Ziel dieser Arbeit ist es, eine Landbedeckungskarte von Oberösterreich mittels kNN-Klassifikation (k Nearest Neighbour) aus einem Landsat-Satellitenbild zu erstellen. Dabei sollen besonders die Fuzzy-Klassifikation und die daraus resultierende Möglichkeit für eine Genauigkeitssteigerung der automatischen Interpretation von Satellitenbildern betrachtet werden.

Die Ergebnisse aus den automatischen Klassifikationen werden mit Referenzdatensätzen aus der Österreichischen Waldkarte und der INVEKOS Datenbank landwirtschaftlicher Förderflächen verglichen, um ihre Genauigkeit zu prüfen und eine Aussage über die Qualität der Klassifikation zu erlauben.

Der Vergleich der Ergebnisse der automatischen Klassifikationen mit den Referenzdatensätzen zeigt eine gute Übereinstimmung. Zur Erstellung einer herkömmlichen Landbedeckungskarte erzielt man mit der Fuzzy Klassifikation keine Verbesserung der Genauigkeit, da die harte Klassifikation für diese Anwendung bereits sehr gute Ergebnisse liefert.

Abstract

The goal of this diploma thesis is to create a land cover map of Upper Austria by using kNN (k nearest neighbours) classification from a Landsat satellite image. In this context, the focus is set on fuzzy classification and the possibilities of increasing accuracy in automatic interpretation of satellite pictures.

The results of the classifications are compared to reference data consisting of Austria's forest map and Invekos datasets of agricultural areas, to verify the accuracy and evaluate the quality of classification.

Comparing the results of automatic classification to the reference data shows a good correlation. Creating a standard land cover map by using fuzzy classification does not improve the accuracy, because standard kNN classification already gives good results.

1. Einleitung

1.1 Motivation

Geoinformation hat in der Informations gesellschaft von heute eine besondere Bedeutung. Gründe für den st eigenden Bedarf an Geoinformation liegen in den konkurrenzierenden Nutzungsansprüchen der unterschiedlichen Wirtschaftssektoren an Grund und Boden, sowie in der Umwelt problematik. Der Geodatenmarkt zeigt aber dem gegenüber nur eine sehr beschei dene Entwicklung. Sowohl bestehende als auch neue Anwendungen wer den dadurch stark eingesc hränkt. Diese Situation der begrenzten Dat enverfügbarkeit betrifft auch Geodaten z ur Landbedeckung, welche für zahlreic he Anwendungen eine wichtige Datengrundlage darstellen. (Grillmayer 2004)

1.2 Ziel der Arbeit

Ziel dieser Arbeit ist die Erstellung ei ner Landbedeckungskarte von Oberösterreich mittels kNN (k Nearest Neighbour) Kla ssifikation. Besonders Augenmerk wird dabei auf die Fuzzy-Klassifikation gelegt. Es soll die Möglichkeit untersucht werden, durch Verwendung der Fuzzy Klas sifikation die Genauigk eit der automatischen Klassifizierung von Satellitenbildern zu st eigern. Zur Verifizierun g der Ergebniss e sollen die Daten aus den aut omatischen Klassifikationen mit Referenzdatensätzen aus der Österreichischen Wald karte und der INVEKOS-Datenban k landwirtschaftlicher Förderflächen verglichen werden.

2. Datengrundlagen

2.1 Satellitendaten

Als Ausgangsbasis für diese Arbeit dient eine Landsat 5 Szene aus den Beständen des Instituts für Vermessung, Fernerkundung und Landinformation (IVFL) vom 16.Juni 2003, welche beinahe ganz Oberösterreich und weite Teile von Salzburg abdeckt. Bei dieser Szene fehlt ein Stück westlich von Braunau. Es wird auf das Anfügen einer weiteren Szene für diesen kleinen Bereich verzichtet, da für eine weitere Szene neue Trainingsdaten benötigt würden. Bei der Interpretation der Ergebnisse ist das Fehlen dieses Stücks zu berücksichtigen. Es können daher für den Bezirk Braunau keine absoluten Flächenvergleiche stattfinden.

Die Georeferenzierung der Landsat 5 Szene wird im UTM System WGS84 Abschnitt 33N durchgeführt. Zur Anwendung kommen neben dem Landsat Bild die ÖK50 und ein digitales Höhenmodell.

Zuerst müssen Passpunkte sowohl in der ÖK50 als auch im Satellitenbild identifiziert und markiert werden. Dazu werden aus der Karte die korrekten Lagekoordinaten der Punkte herausgesucht und auf der Szene mittels ERDAS GCP Tool Punkt für Punkt manuell eingegeben. Am Ende der Passpunktsuche sind 35 Passpunkte über das gesamte Bild verteilt. Abbildung 1 zeigt die verwendete Landsat 5 Szene mit allen identifizierten Passpunkten. Die Genauigkeit der Passpunkte erreicht die folgenden Werte:

Control Piont Error

- (X) – 0,2861 Pixel
- (Y) – 0,2226 Pixel
- (Total) – 0,3626 Pixel

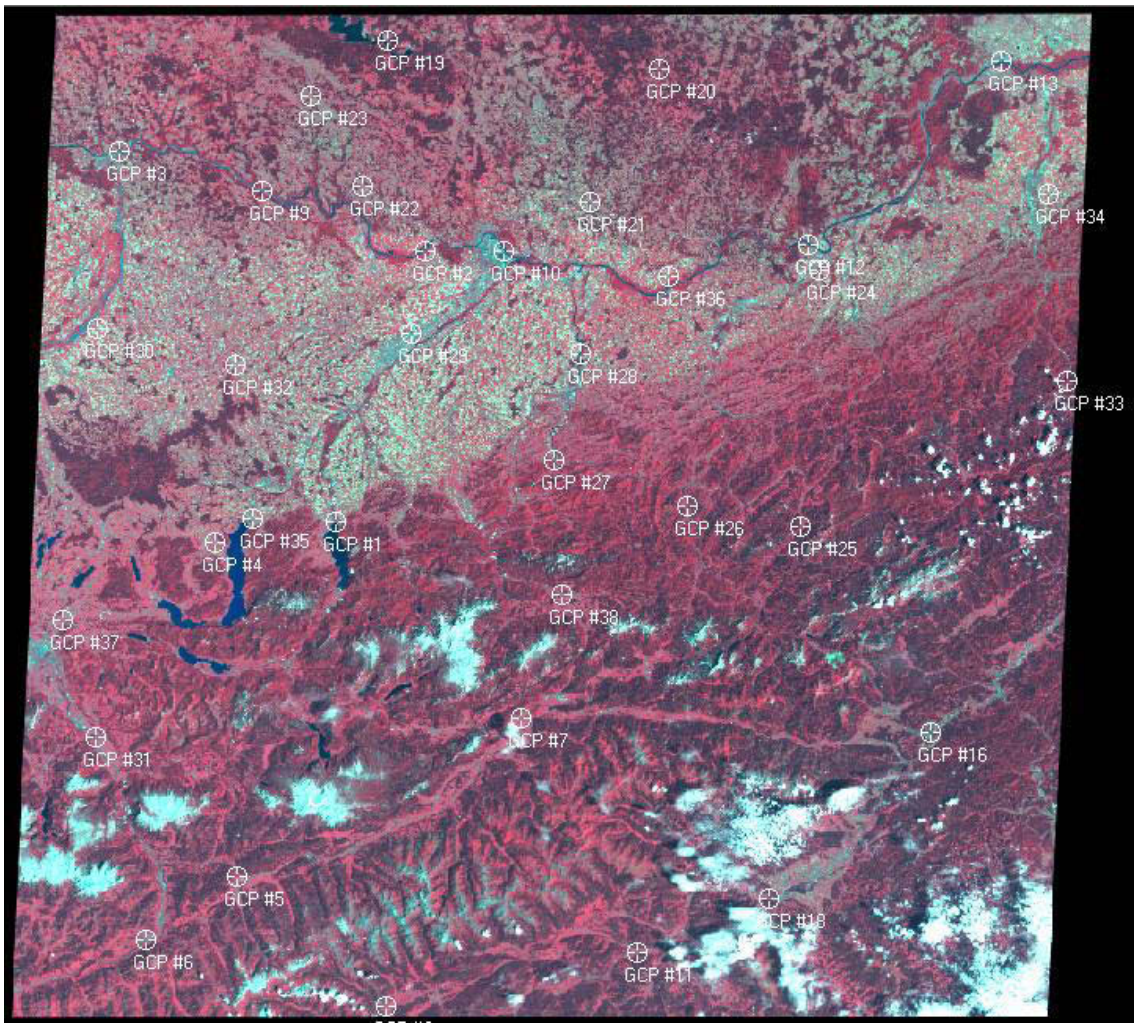


Abb. 1: Landsat 5 Szene mit allen Passpunkten

Als Resamplingmethode wird die Near est Neighbour Methode gewählt, um eine Veränderung der Pixelfarbwerte durch die Umrechnung zu verhindern, da dies eine weitere spektrale Klassifizierung erschweren könnte.

Zur visuellen Kontrolle der durchgeführten Georeferenzi erung wird mittels ArcMap über das georeferenzierte Lan dsat Bild ein Aussc hnitt der ÖK50 gelegt. Die Ergebnisse sind in den Abbildungen 2 und 3 zu sehen.

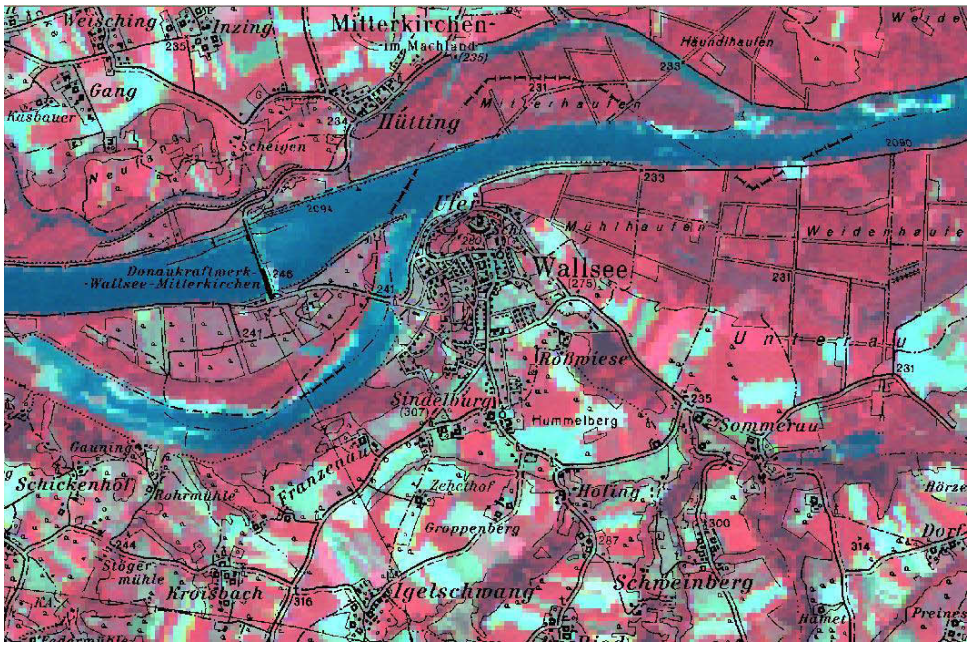


Abb. 2: Georeferenzierte Landsat 5 Szene mit ÖK 50

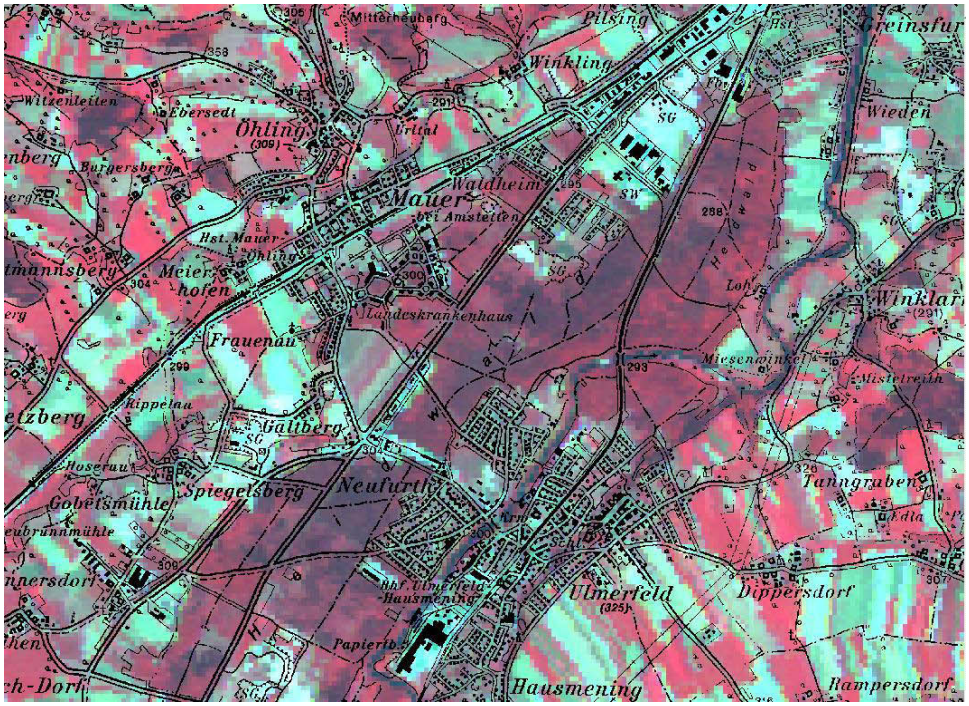


Abb. 3: Georeferenzierte Landsat 5 Szene mit ÖK 50

Begriffsbestimmungen:

<u>*Landsat 5:*</u> Die Erderkundungssatelliten der LANDSAT Serie nehmen seit 1972 Bilder der Erdoberfläche auf. Die Satelliten überflogen bzw. überfliegen die Erde polumlaufend in sonnensynchronen Bahnen . Die Bildelementgr öße der TM-Bilddaten ist 30m, es werden 6 Spektralka näle im Sichtbaren, im nahen Infrarot und im mittleren Infrarot aufgenommen. Zusä tzlich gibt es einen Thermalkanal

mit 120m Bildelementgröße (Schneider 2002. S.57) Landsat 5 ist der fünfte Satellit des Landsat Programms. Er wurde am 1. März 1984, mit dem Ziel ein globales Archiv von Satellitenbildern zu gewährleisten, in Betrieb genommen. Das Landsat Programm wird vom Zentrum für Erdbeobachtung des USGS betrieben. (Wikipedia am 29.8.2011)

Digitales Höhenmodell: Als Digitales Höhenmodell bezeichnet man die Menge der digital gespeicherten Höhenwerte von regelmäßig oder unregelmäßig verteilten Geländepunkten die als Funktion der Lage der Punkte die Höhenstruktur des Objektes, z.B. des kontinuierlich verlaufenden Geländes, hinreichend repräsentieren. Die Anordnung der Stützpunkte wird häufig in Form eines Quadratgitters gewählt, weshalb man dann auch den Begriff Gitterhöhenmodell verwendet. (Geoinformatik am 29.8.2011) Die Höhen von beliebigen zwischen den gespeicherten Rasterpunkten liegenden Punkten können aus den 4 umliegenden Rasterpunkten interpoliert werden. (Schneider 2002. S.101)

2.2 Trainingsdaten

Zur Generierung von Trainingsdaten wird ein Punktraster über die Landsat Szene gelegt und diese Punkte werden danach einzeln visuell nach ihrer Zugehörigkeit zu einer Landbedeckungsklasse interpretiert.

Begriffsbestimmungen:

Trainingsdaten: Trainingsdaten sind repräsentative Beispielpixel mit bekannter Landbedeckung. Diese werden bei der automatischen Klassifikation als Interpretationsschlüssel, der die spektralen Eigenschaften aller Landbedeckungsklassen beschreibt, verwendet. (Lillesand; Kiefer 2000. S.533)

Landbedeckung (Land Cover): Der Begriff Landbedeckung bezieht sich auf jene Objekte, welche die Oberfläche der Erde bedecken. Maisfelder, Seen, Ahornbäume und betonierte Strassen sind alles Beispiele für Landbedeckungsklassen. (Lillesand; Kiefer 2000 S.208)

Landnutzung: (Land Use): Der Begriff Landnutzung ist abhängig von der Nutzung durch den Menschen oder der wirtschaftlichen Nutzung, welche mit einem Stück Land verbunden ist. (Lillesand; Kiefer 2000. S.208)

2.2.1 Generierung der Trainingsdaten

Erstellung des Punktrasters für die visuelle Interpretation

Es wird in ArcGis ein Punktraster erzeugt, welcher die gesamte Landsat Szene umfasst. Die dazu nötigen Ausdehnungen werden aus den Layereigenschaften des Landsat Bildes aus gelesen. Das generierte Raster erstreckt sich mit einer Maschenweite von 2000m über das komplette Bild.

Klassifizierung der Gitterpunkte

Der Raster definiert auf dem Satellitenbild Punkte, welche mit Hilfe von Orthofotos aus dem digitalen oberösterreichischen Rauminformationssystem (DORIS) einzeln visuell interpretiert werden. (DORIS; http://doris.ooe.gv.at/geoinformation)

Zur Interpretation muss man die Punkte des Gitterrasters sowohl auf der Landsat Szene als auch auf den Orthofotos eindeutig identifizieren und jedem einzelnen Pixel eine Landbedeckungsklasse zuordnen. Es werden die in Tabelle 1 aufgelisteten Landbedeckungsklassen gemäß Dipl.-Ing. Mattiuzzi verwendet. (Mattiuzzi 2007)

Landbedeckungsklassen:

11 Fels, Schotter, Gestein
12 Fels, Schotter, Gestein mit Vegetationsanteil
13 Schnee

20 Wasser (See, Fluss)

31 Nadelwald dicht (Kronenschluss min. 90%)
32 Nadelwald schütter (sobald nicht mehr ganz dicht, wird diese Klasse zugewiesen)
33 Nadelwald sehr schütter (hoher Grasanteil zw. den Bäumen. Max. 50% Kronenschluss)

34 Mischwald dicht
35 Mischwald schütter
36 Mischwald sehr schütter

37 Laubwald dicht
38 Laubwald schütter
39 Laubwald sehr schütter

41 Latsche
42 Erle

61 Acker grün
62 Acker vegetationslos, anderer vegetationsloser Boden (nicht Fels)
63 Acker abgeerntet
64 Golfplatz
65 Grünland
66 Grünland gemäht
67 Hochalpine Vegetation
68 Moor

70 versiegelte Flächen

80 Wolke
81 Wolkenschatten

91 Fels – Nordhang Schatten
92 Wald – Nordhang Schatten
93 Acker/Grünland – Nordhang Schatten
94 Versiegelte Flächen – Nordhang Schatten

Tab.1: Landbedeckungsklassen

Insgesamt werden 3014 Punkte manuell interpretiert. Abbildung 4 zeigt alle interpretierten Rasterpunkte. In den aufgelisteten Landbedeckungsklassen ist wegen ihrer eindeutigen spektralen Erkennbarkeit auch die Landnutzungsklasse „Golfplatz“ enthalten.

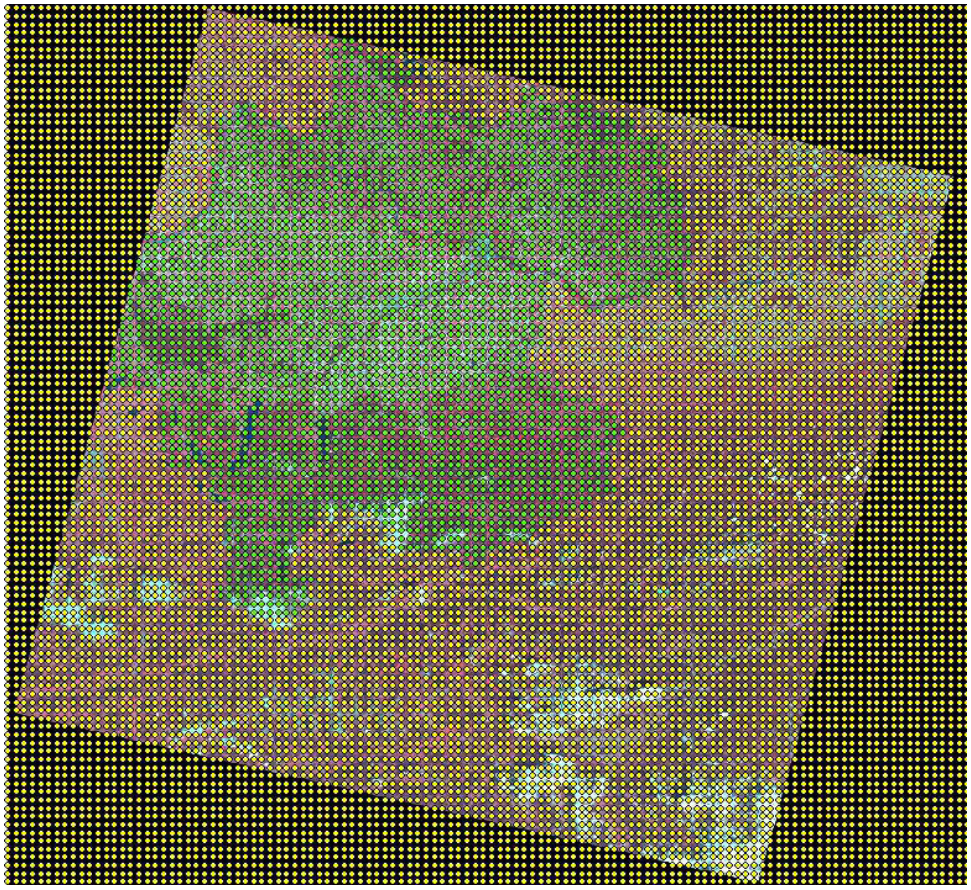


Abb. 4: Satellitenbild mit allen interpretierten Rasterpunkten

2.2.2 Verbesserung der Trainingsdaten

Korrektur der Punkte, welche an Pixelgrenzen liegen

Einige der automatisch gesetzten Rasterpunkte liegen an Pixelgrenzen. Es könnte also vorkommen, dass keine eindeutige Zuordnung eines Punktes zu einem bestimmten Pixel gewährleistet ist. Daher werden alle Punkte auf ihre Lage überprüft und wenn notwendig in das Zentrum eines Pixels verschoben.

Zusätzliche Trainingsdatenpunkte für unterrepräsentierte Klassen

Ziel ist es, für jede Klasse eine minimale Referenzpunktzahl von 10 klassifizierten Punkten zu bekommen. Dies dient dazu, die kleinen Klassen zu stärken, so dass sie nicht durch die größeren Klassen überstimmt werden. Beim Setzten der neuen Punkte wird darauf geachtet, dass sie in einer der gewünschten kleinen Landbedeckungsklassen liegen. Diese Punkte werden dann wie die automatisch

gesetzten Rasterpunkte visuell Interpretiert. Nach der Suche nach neuen Referenzpunkten sind die „kleinen“ Klassen folgendermaßen vertreten:

- Schnee (17 Punkte)
- Laubwald sehr schütter (14 Punkte)
- Latsche(15 Punkte)
- Golfplatz (16 Punkte)
- Hochalpine Vegetation (15 Punkte)
- Wolke(12 Punkte)
- Wald Nordhang Schatten(15 Punkte)
- Fels Nordhang Schatten (6 Punkte)

Einzig die Kategorie „Fels Nordhang Schatten“ ist unverändert bei unter 10 Referenzpunkten. Hier wurden nicht genug Referenzpunkte gefunden.

Es wird davon ausgegangen, dass die zusätzlichen Referenzpunkte die kleinen Klassen stärken und eine Erhöhung der Genauigkeit der einzelnen Klassen erreicht wird. Nach diesem Arbeitsschritt sind insgesamt 3048 Referenzpunkte vorhanden.

Anschließend werden Ausreißer (Outlier) unter den Trainingsdaten eliminiert. Dieser Arbeitsschritt ist in Kapitel 3.3.1 näher beschrieben. Danach sind noch 2962 Referenzpunkte gegeben.

Die gesamten Trainingsdaten für die Klassifizierung nach der Ausreißereliminierung sind, wie in Tabelle 2 ersichtlich, über die Landbedeckungsklassen verteilt.

Nr.	Landbedeckungsklasse	Punktanzahl	%
11	Fels, Schotter, Gestein	41	1,38
12	Fels, Schotter, Gestein mit Vegetationsanteil	30	1,01
13	Schnee	20	0,68
20	Wasser (See, Fluss)	51	1,72
31	Nadelwald dicht (Kronenschluss min. 90%)	492	16,61
32	Nadelwald schütter	192	6,48
33	Nadelwald sehr schütter (Max. 50% Kronenschluss)	65	2,19
34	Mischwald dicht	193	6,52
35	Mischwald schütter	55	1,86
36	Mischwald sehr schütter	19	0,64
37	Laubwald dicht	122	4,12
38	Laubwald schütter	35	1,18
39	Laubwald sehr schütter	14	0,47
41	Latsche	15	0,51
61	Acker grün	275	9,28
62	Acker vegetationslos, anderer vegetationsloser Boden	444	14,99
63	Acker abgeerntet	44	1,49
64	Golfplatz	17	0,57
65	Grünland	391	13,2
66	Grünland gemäht	224	7,56
67	Hochalpine Vegetation	13	0,44
70	versiegelte Flächen	178	6,01
80	Wolke	11	0,37
91	Fels – Nordhang Schatten	6	0,2
92	Wald – Nordhang Schatten	15	0,51

Tab.2: Punkteverteilung der Trainingsdaten (Anzahl und Prozentanteil) in den Klassen

In Abbildung 5 wird die gesamte Punkt eanzahl der Trainingsdaten nach ihrer Klassenzugehörigkeit dargestellt. So lasse n sich auf den ersten Blick einige sehr große Landbedeckungsklassen erkennen. Erwartungsgemäß sind das dichter Nadelwald, vegetationsloses Ackerland, und Grünland mit einem Landbedeckungsanteil von jeweils über 10% sowie schütterer Nadelwald, dichter Mischwald, grünes Ackerland, gemähtes Grünland und Siedlung/Versiegelte Flächen. Diese Klassen decken einen Großteil des Landesgebiet es von Oberösterreich ab und sich daher auch in den zufällig verteilten Trainingsdaten sehr stark repräsentiert.

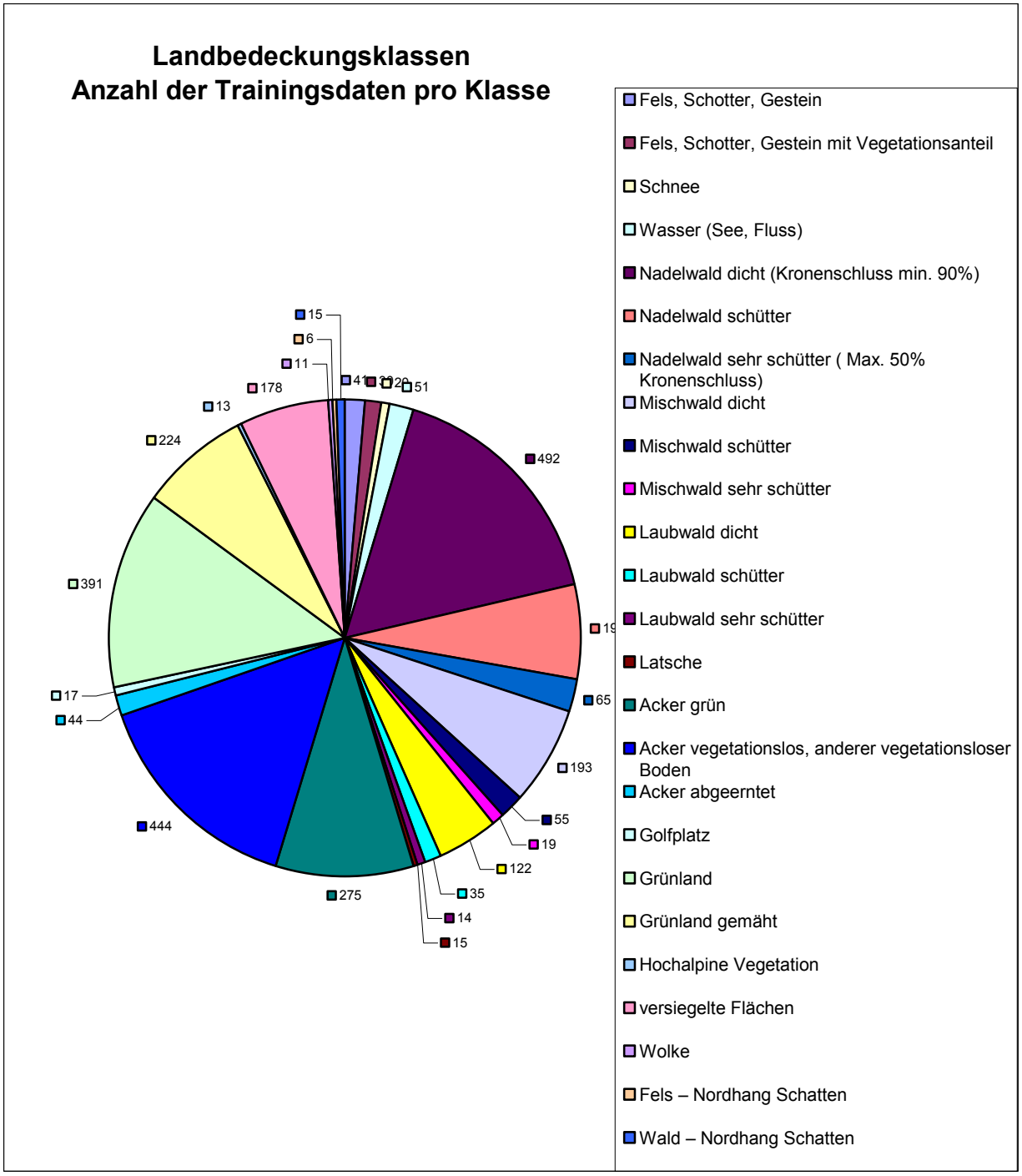


Abb.5: Verteilung der Trainingsdaten - Anzahl der Trainingsdaten pro Landbedeckungsklasse

Mit diesen Trainings daten, die von bei der visuellen Interpretation auf tretenden Ausreißern bereinigt und durch zu sätzliche Stärkung der kleinen Landbedeckungsklassen an die folgenden Prozesse angepassten sind, wer den alle weitern Berechnungen und Klassifikationen durchgeführt.

2.3 Vergleichsdatensätze

Als Vergleichsdatensätze für die Kontrolle der automatischen Klassifikation werden für die Walddaten die Daten der Österreichischen Waldkarte und für die Agrarflächen die landwirtschaftlichen Förderflächen aus der Invekos-Datenbank verwendet.

2.3.1 Österreichische Waldkarte

Zuerst muss der Datensatz der Österreichischen Waldkarte aufbereitet werden, um ein Vergleichen mit den Ergebnissen aus der automatischen Klassifizierung zu ermöglichen. In diesem in Abbildung 6 dargestellten Datensatz wird auf Bezirksebene in folgende Klassen unterschieden:

- Value 0 = Kein Wald (Grau)
- Value 1 = Laubwald (Violett)
- Value 2 = LW-dom. Mischwald (Rot)
- Value 3 = NW-dom. Mischwald (Rot)
- Value 4 = Nadelwald (Grün)
- Value 5 = Schlagfläche (Schwarz)

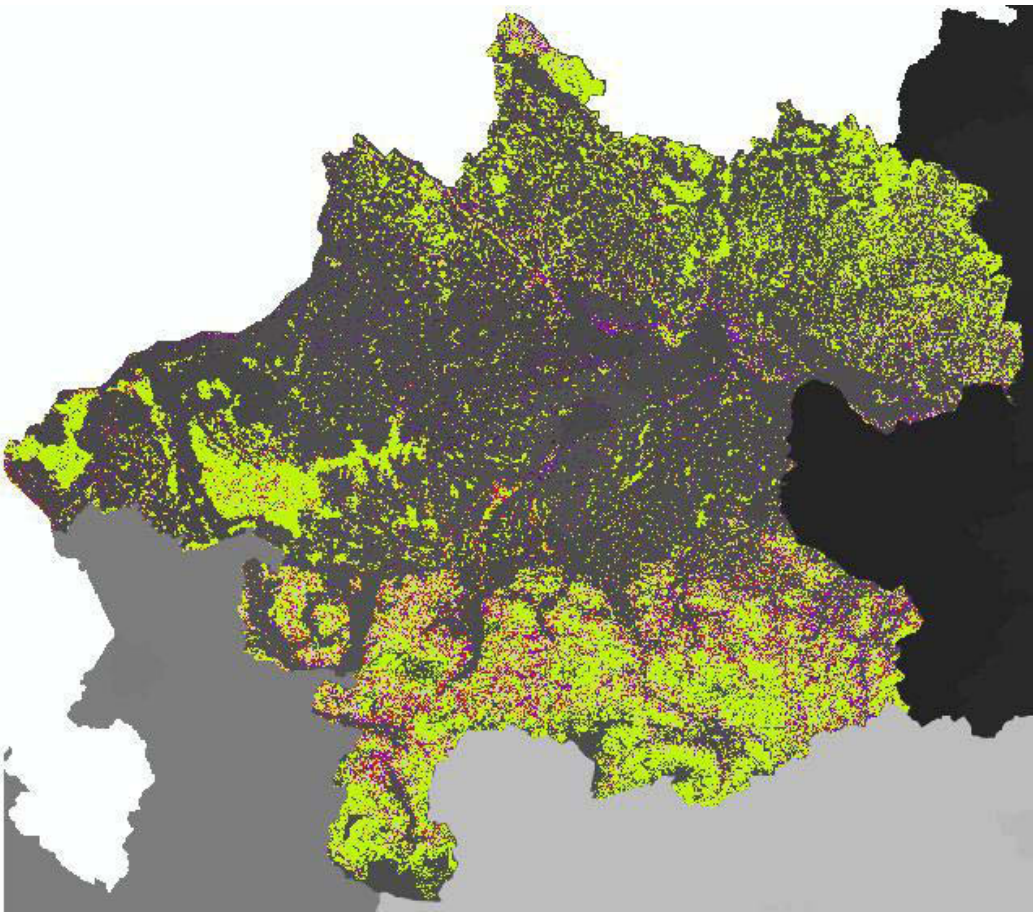


Abb. 6: Waldkarte Oberösterreich

Die beiden Klassen des Mischwaldes werden aus Vergleichsgründen zu einer Klasse Mischwald zusammengefasst. Daraus ergeben sich für ganz Oberösterreich folgende Referenzflächen:

Laubwald:	77.133 ha	6,4%
Mischwald:	78.308 ha	6,5%
Nadelwald:	326.485 ha	27,3%
Gesamt:	501.646 ha	41,9% (incl. Schlagflächen)

Die Prozentwerte sind an der Gesamtfläche Oberösterreichs zu messen.

2.3.2 Invekos Agrarförderflächen

Bei der Invekos-Datenbank handelt es sich um Agrarförderflächen des Landes Oberösterreich. Um einen Vergleich mit den Ergebnissen der automatischen Klassifikation zu ermöglichen, müssen aus dem Invekos Datensatz Flächen entfernt werden, welche in der hier verwendeten automatischen Klassifikation nicht als Landwirtschaftliche Fläche definiert sind. Die Daten stehen auf Gemeindeebene zur Verfügung und werden für den Vergleich auf Bezirksebene zusammengefügt.

Es werden folgende Flächen von der Summe der Förderfläche abgezogen: Einjährige Baumschulen; Mehrjährige Baumschulen; Einmähdige Wiesen; Mehrmähdige Wiesen; Kultur weiden; Streuwiesen; Einmähdige Streuobstwiesen; Mehrmähdige Streuobstwiesen; Hutweiden; Bergmähder aus ÖPUL; Sonstige Grünlandfläche; Streuobst/Kulturweide; Streuobst/Hutweide; Landschaftselement G; Mehrj. Baumschulen S; Kirschen; Marillen; Pfirsiche; Tafeläpfel; Tafelbirnen; Weichseln; Zwetschken; Quitten; Pflaumen; Strauchbeeren; Holunder; Wein; Wein Junganlagen; Wein Schnittweingärten; Christbäume auf Ackerland; Christbäume auf Grünland; Aufforstung auf Ackerland; Aufforstung auf Grünland; Neuaufforstungspflege; Pflege von Wald; Haus/Gemüsegärten; Energieholzflächen; Sonst. Aufforstungsflächen; Teichflächen; SL Grünbrache keine Beihilfe; SL Grünbrache mit Beihilfe; SLG Gewässerstreifen GL keine Beihilfe; SLG Gewässerstreifen GL mit Beihilfe; Sonstige Fläche; Landschaftselement A keine Beihilfe; Landschaftselement A mit Beihilfe; Dauergrünland; Weiden aus Almauftriebsliste; Ackerland (Summe); Almfutterfläche aus Almauftriebsliste;

Landwirtschaftlich genutzte Fläche (FL) ohne Almen, Landwirtschaftlich genutzte Fläche.

Damit sind alle geförderten Agrarflächen entfernt, welche bei der automatischen Klassifikation als Wald oder Grünland zu identifizieren sind.

2.3.3 Aufbereitung der Daten auf Bezirksebene

Die Kontrolle der automatischen Klassifikation mit der Waldkarte und den Invekosdaten wird auf Bezirksebene durchgeführt. Dazu wird eine Bezirks-Maske benötigt. Mit dieser Maske werden die Landbedeckungsklassen der einzelnen Pixel dem diesem Pixel zugehörigen Bezirk zugeordnet. Diese automatische Zuordnung durch das kNN Programm ermöglicht einen Vergleich dieser Bezirksdaten mit den Vergleichsdatensätzen.

Die Maske für das kNN Programm wird mittels ArcMap erstellt. Es ist darauf zu achten, dass die erstellte Maske dieselbe Pixelgröße (Cell Size) und dieselben Ausdehnungen wie das Satellitenbild aufweist. Pixel innerhalb Oberösterreichs erhalten den Wert 10 und Pixel außerhalb von OÖ den Wert 0. Die so erzeugte Maske muss wegen des fehlenden Bildteils westlich von Braunau noch entlang des Randes der Landsat Szene abgeschnitten werden. Dadurch entsteht die in Abbildung 7 dargestellte Maske, welche von den Bildrändern und den Grenzen Oberösterreichs begrenzt wird.

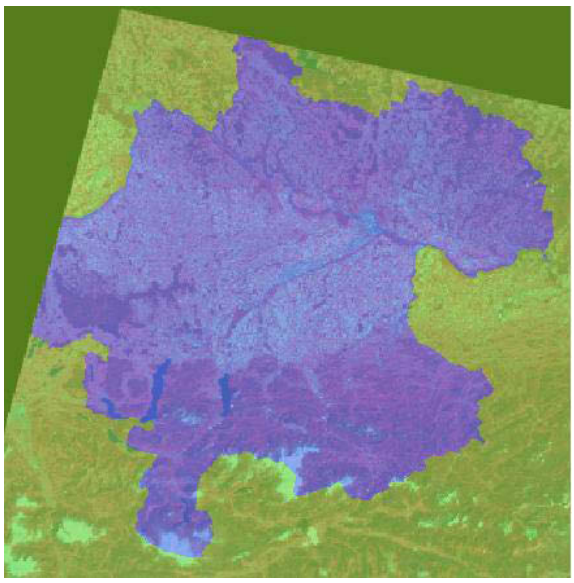


Abb. 7: fertige OÖ Maske über dem Landsat Bild

3. kNN (k-Nearest-Neighbour) Klassifizierung

3.1 kNN Methode

Der kNN Algorithmus ist eine Instanz-basierte Klassifikation. Verfahren dieser Art trachten nicht danach, eine optimale Repräsentation jeder Klasse aufzubauen, sondern orientieren sich an den Klassen, die jene Trainingsdaten aufweisen, die dem zu klassifizierenden Pixel am ähnlichsten sind. Beim Klassifizieren eines neuen Pixels untersucht der Algorithmus die k diesem Pixel ähnlichsten Trainingspixel und teilt das Pixel der unter diesen k Trainingspixel am häufigsten vertretenen Klasse zu (Oberhauser 2005. S.30) Im Vergleich zu parametrischen Verfahren wird bei dieser Methode eine größere Menge an Referenzpunkten benötigt, damit auch alle Klassen mit einem großen Streubereich gut repräsentiert sind. (Popovici 2008. S.11)

Als Ähnlichkeitsmaß wird üblicherweise die Euklidische Distanz in einem Merkmalsraum gewählt, dessen Koordinaten die (gegebenenfalls mit Gewichten versehenen) multispektralen Pixelwerte und allenfalls weitere Merkmale (wie etwa Höhenlage, Hangneigung usw.) sind.

Bei der Evaluierung eines automatischen Klassifizierungsverfahrens wird üblicherweise dessen Wirksamkeit gemessen, das heißt die Fähigkeit, die richtigen Klassifizierungsentscheidungen zu treffen (Oberhauser 2005. S.31) Ein Maß für die Klassifizierungsgüte ist die Genauigkeit (Accuracy). Die Genauigkeit drückt die Wahrscheinlichkeit aus, dass ein zufällig ausgewählter Pixel korrekt klassifiziert wird. (Oberhauser 2005. S.33)

Begriffsbestimmung:

Euklidische Distanz: Die Euklidische Distanz zweier Punkte x und y im n-dimensionalen (Merkmals-) Raum ist gegeben durch:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$ (Mitchell 1997. S231)

3.2 kNN Programm

Das kNN-Programm wurde auf dem Institut für Vermessung, Fernerkundung und Landinformation von Dr. Koukal während des Projektes „Untersuchung zur flächigen Verdichtung der Waldinformation aus der Österreichischen Waldinventur mit Satellitendaten“ entwickelt. (Tatjana Koukal. User Guide to the kNN-Programm S.2) Es setzt sich aus 3 verschiedenen Programmenteilen mit den Funktionen „Read Image“, „Crossvalidation“ und „Map Maker“ zusammen.

Mit dem Programmteil „Read Image“ wird die Bildinformation der einzelnen Trainingspixel ausgelesen, „Crossvalidation“ dient zum Parametertuning und „Map Maker“ klassifiziert den gesamten Datensatz und erzeugt die Resultatskarte.

Das kNN-Programm bietet verschiedene Möglichkeiten, die automatische Klassifikation zu beeinflussen und zu optimieren. Dazu ist es möglich, bei der Klassifikation die Spektralbänder der Landsat-Szene mit verschiedenen Gewichtungen in die Berechnung einzubeziehen. Ebenfalls ist es möglich, die Suche nach spektral k Nächsten Nachbarn geografisch horizontal und vertikal mit einem Suchradius um das zu klassifizierende Pixel zu begrenzen. Diese Optimierungsmöglichkeiten werden im Verlauf dieser Arbeit genauer untersucht.

Der Eingangsdatensatz ist eine Tabelle mit den Koordinaten der einzelnen Trainingspixel und den dazu gehörenden Landbedeckungsklassen.

3.2.1 Datenaufbereitung

Während der Aufbereitung des Digitalen Geländemodells in ArcMap wird eine Pixelgröße 30 Metern errechnet. Aus dem aufbereiteten digitalen Geländemodell wird in ArcMap mittels Oberflächenanalyse eine Slope-, Aspect- und Hillshade-Datei erzeugt. Die Bearbeitung des Satellitenbildes wird mit Erdas Imagine 9.1 durchgeführt. Der Sonnenstand während der Aufzeichnung des Satellitenbildes wird aus dem Header File der Landsat-Szene entnommen und durch folgende Parameter berücksichtigt: Azimuth: 133,66°

Elevation: 56,86°

3.2.2 Read Image

In diesem Programmteil werden Lage- und Spektralinformationen der Trainingspixel in einer Tabelle zusammengefasst, welche in den folgenden Modulen des Programms verwendet wird. Im Detail werden Pixelwerte ausgelesen und wenn gewünscht Bandkombinationen berechnet (Koukal 2005. S.5)

3.2.3 Cross Validation

In diesem Modul werden Fehlerstatistiken berechnet und zur Optimierung der Parameter und der Abschätzung der Genauigkeit verwendet. (Koukal 2005. S.5) Die Kreuzvalidierung dient der Feststellung der „inneren Genauigkeit“, welche durch die Parameter Kappa und Overall Accuracy angegeben wird. Damit wird überprüft, ob die interpretierten Punkte auch mittels automatischer Klassifikation die zugeordnete Klasse erhalten.

Als Eingangsdatensatz dient die in Modul „Read Image“ erstellte Tabelle. Das kNN Programm ermöglicht es, den einzelnen Bändern des Landsat Bildes verschiedene Gewichtungen für die weiteren Berechnungen zuzuweisen. Diese Gewichtung entscheidet über die „Wichtigkeit“ des einzelnen Bandes bei der automatischen Interpretation der Pixel, das ist bei der Berechnung der Euklidischen Distanzen. Im vorliegenden Fall sind 7 Bänder und 1 Bandkombination zu berücksichtigen. Auch Hillshade, Slope und Aspect werden mit einbezogen. Damit werden für die Optimierung der inneren Genauigkeit sämtliche zur Verfügung stehenden Datensätze verwendet, was eine bestmögliche Feineinstellung der Gewichtungsparameter erlaubt.

Begriffsbestimmungen:

Fehlermatrix: Die Fehlermatrix ist eine Kreuztabelle, in der die Spalten die in der Realität gegebenen Kategorien (Soll-Kategorien) und die Zeilen die bei der Klassifizierung erhaltenen Kategorien (Ist-Kategorien) angeben. Die einzelnen Tabellewerte sind die Häufigkeiten der Pixel für die verschiedenen Kombinationen von Soll- und Ist-Kategorien. In der Hauptdiagonale der Fehlermatrix sind dementsprechend die korrekt klassifizierten Pixel angegeben, für die Soll- und Ist-Kategorie identisch sind.

Gesamt-Genauigkeit (Overall Accuracy): Die Gesamtgenauigkeit lässt sich aus der Fehlermatrix als Summe der Werte in der Hauptdiagonale (richtig klassifizierte Pixel), dividiert durch die Gesamtanzahl der Pixel, errechen (Lillesand; Kiefer 2000. S.570) Bei der Betrachtung der Gesamtgenauigkeiten soll bedacht werden, dass sogar eine völlig zufällige Zuweisung eines Pixels zu einer Landbedeckungsklasse eine korrekte Klassifizierung dieses Pixels ergeben kann. Solche Zufallsübereinstimmungen können zu einer überraschend guten Gesamtgenauigkeit führen. (Lillesand; Kiefer 2000. S.573)

Kappa: Kappa ist ein weit verbreitetes Maß für die Übereinstimmung und ermöglicht häufig einfache Vergleiche. (Sachs; Hedderich 2009. S.626) Zur Berechnung von Kappa werden die Zufallsübereinstimmungen berücksichtigt und abgezogen.

Producers Accuracy: Die „Genauigkeit aus der Sicht des Herstellers“ errechnet sich für jede Spalte/ Klasse (der Fehlermatrix) mit der Anzahl der korrekt definierten Pixel in einer Klasse geteilt durch die insgesamt dieser Klasse zugehörenden Pixel der Referenzdaten. Diese Genauigkeit zeigt an, wie gut die Pixel der Referenzdaten einer Landbedeckungsklasse klassifiziert sind. (Lillesand; Kiefer 2000. S.570)

Users Accuracy: Die „Genauigkeit aus der Sicht des Anwenders“ wird für jede Zeile/Klasse (der Fehlermatrix) berechnet. Dabei wird die Anzahl der korrekt einer Klasse zugewiesenen Pixel durch die Gesamtzahl der durch den Klassifikationsalgorithmus dieser Klasse zugewiesenen Pixel dividiert. Es wird dadurch angegeben, mit welcher Wahrscheinlichkeit ein einer Klasse zugewiesenes Pixel dieser Klasse auch in Wirklichkeit angehört. (Lillesand; Kiefer 2000. S.570)

Bei ersten Probedurchläufen wird festgestellt, dass bei steigenden k nächsten Nachbarn die Genauigkeitsmaße Kappa und Overall Accuracy stiegen. Dieser scheinbare Genauigkeitsgewinn geht auf Kosten von kleinen Klassen mit wenigen identifizierten Referenzpunkten, welche von der hohen Anzahl der nächsten Nachbarn aus anderen spektral ähnlichen Klassen „überstimmt“ werden.

3.2.4 Map Maker

Das dritte Modul des kNN Programms errechnet aus sämtlichen Eingangsdaten für jedes einzelne Pixel des Satellitenbildes die Landbedeckungsklasse und schreibt diese in einen oder mehrere Rasterdateien (Maps). (Koukal 2005. S.6) Diese Ergebnisdateien lassen sich zur weiteren Verarbeitung mit ArcGis statistisch und für die visuelle Betrachtung aufbereiten.

3.3 Optimierung der kNN Klassifizierung

Zunächst werden die Trainingsdaten mit Hilfe des kNN-Algorithmus von Ausreißern bereinigt. Es gibt verschiedene weitere Möglichkeiten, die automatische Klassifizierung durch den kNN- Algorithmus zu beeinflussen. Untersucht werden die Auswirkung von verschiedenen Werten von k (Anzahl der nächsten Nachbarn) bei der Klassifizierung, sowie das Einbeziehen der Lageinformationen (geografischer Suchradius) der zu klassifizierenden Pixel. Auch eine Optimierung der Gewichtungen der einzelnen Landsat-Bänder sowie von Hillshade, Slope und Aspect als Eingangsparameter in die Berechnung wird durchgeführt.

3.3.1 Outlier (Ausreißer)

Outlier sind Bestandteile eines Datensatzes, die sich vom allgemeinen Muster der restlichen Daten unterscheiden. Outlier werden auch als Rauschen oder „Noise“ bezeichnet und zeigen einen Fehler im Prozess der Datenerstellung an. Outlier können die Datenanalyse erschweren oder die Vorhersagewahrscheinlichkeit reduzieren. (Koukal 2005. S.26) Sie können durch Änderungen in der am betrachteten Pixel vorhandenen Landbedeckung zwischen den Aufnahmezeitpunkten von Satellitenbild und Orthofoto entstehen, oder aus untypischen spektralen Eigenschaften einzelner Pixel hervorgehen.

Die Outlier-Suche wird mit k=1 durchgeführt, da der kNN Algorithmus bei k=1 besonders anfällig gegenüber Ausreißern ist und somit Ausreißer leichter auffindbar sind (Popovici 2008. S.42) Dabei wird eine Liste erstellt, welche darüber Auskunft gibt, welcher Punkt wie oft zur Klassifizierung verwendet wird und wie oft dieses Ergebnis mit der vorgegebenen Klassifikation übereinstimmt. Mit diesen Daten lassen sich jene Punkte finden, welche die Ergebnisse eventuell ungünstig beeinflussen. Die Suche nach Outlier-Punkten und deren Entfernung aus dem Trainingsdatensatz wird so lange wiederholt, bis sich keine Änderung mehr in der Genauigkeit erkennen lässt. In dieser Arbeit wird nach dem fünften Durchlauf keine Verbesserung mehr erreicht und die Suche damit als abgeschlossen angesehen.

Bei der Entfernung der Outlier werden also jene Punkte entfernt, welche bei der Crossvalidation oft zu Falschklassifizierungen führen. Es wird nun weiter überprüft, ob eine zusätzliche Verbesserung der Trainingsdaten durch Punkte, welche bei der Crossvalidation überdurchschnittlich oft zur richtigen Klassifizierung führten, möglich ist.

Ein einfaches Reduzieren der Trainingsdaten nur auf positiv aufgefallene („gute“) Punkte wäre möglicherweise nicht sinnvoll. Würde man nur die „Guten“ als Referenzpunkte für die Klassifikation verwenden, so wäre zwar die mit Hilfe der Crossvalidation geschätzte Klassifikationsgenauigkeit sehr hoch, jedoch würden die Referenzpunkte die tatsächliche spektrale Verteilung der Klassen möglicherweise nicht mehr ausreichend repräsentieren (Popovici 2008. S.41) Zur Überprüfung dieser Annahme wird überprüft, ob Punkte, welche in der Crossvalidation positiv auffallen, häufiger zur Klassifizierung anderer Punkte herangezogen werden als Punkte, welche nicht positiv auffallen. Wenn dem so ist, sollte man mit diesen Punkten durchaus die Klassifizierungswahrscheinlichkeit positiv beeinflussen können.

Um eine Liste der als „gut“ zu bewertenden Punkte zu bekommen, werden während der Outlier Suche nicht nur die tatsächlichen Ausreißer entfernt, sondern es werden auch jene Punkte zusammengefasst, welche bei der Crossvalidation in mehr als 75% der Klassifizierungen anderer Punkte zum richtigen Ergebnis führen. 1497 solcher Punkte werden identifiziert.

Die indifferenten Punkte, welche bei der Klassifikation weder positiv noch negativ auffielen, werden für die weitere Klassifikation der neuen Punkte ebenfalls herangezogen. Der Referenzdatensatz besteht aus 51% guten und 49% indifferenten Punkten. Anhand der beiden Datensätze lässt sich prüfen, ob es Unterschiede zwischen guten und indifferenten Referenzpunkten im Auftreten als „nächste Nachbarn“ bei der Klassifizierung gibt.

Dazu wird ein neues Raster über die Landsat Szene gelegt. Die dadurch entstandenen neuen Punkte werden mit der Crossvalidation mittels des vorhandenen „alten“ Datensatzes klassifiziert. Dadurch lässt sich erkennen, welche Punkte aus dem Referenzdatensatz wie häufig zur Klassifizierung der neuen verwendet werden.

Die Auswertung bei k=1 ergibt, dass jene Punkte, welche bei der Outlier Suche als „gut“ bewertet werden, bei der Klassifikation nicht häufiger verwendet werden als indifferente Punkte. Die indifferenten Punkte werden bei der Klassifikation neuer Punkte 721mal verwendet. „Gute“ Punkte werden bei der Klassifikation 627 mal verwendet. Damit werden die für gut befundenen Punkte sogar geringfügig weniger zur Klassifizierung neuer Punkte verwendet.

Es lässt sich also anhand der Klassifizierungswahrscheinlichkeit der einzelnen Punkte keine tatsächliche Verbesserung des Referenzdatensatzes oder der Klassifizierung neuer Punkte erreichen.

3.3.2 Werte von k

Um die Auswirkungen der verschiedenen Werte von k bei der automatischen Klassifikation darzustellen, wird eine Berechnung der Genauigkeiten jeder einzelnen Landbedeckungsklasse mit k nächsten Nachbarn von k=1 bis k=10 durchgeführt. Die Abbildungen 8 und 9 zeigen einen Auszug aus den wichtigsten Landbedeckungsklassen und ihre Anwendergenauigkeit (User’s Accuracy) sowie ihre Herstellergenauigkeit (Producer’s Accuracy) in Abhängigkeit von k. Jede Linie des Diagramms stellt eine Landbedeckungsklasse und ihren Genauigkeitsverlauf in Abhängigkeit von k dar.

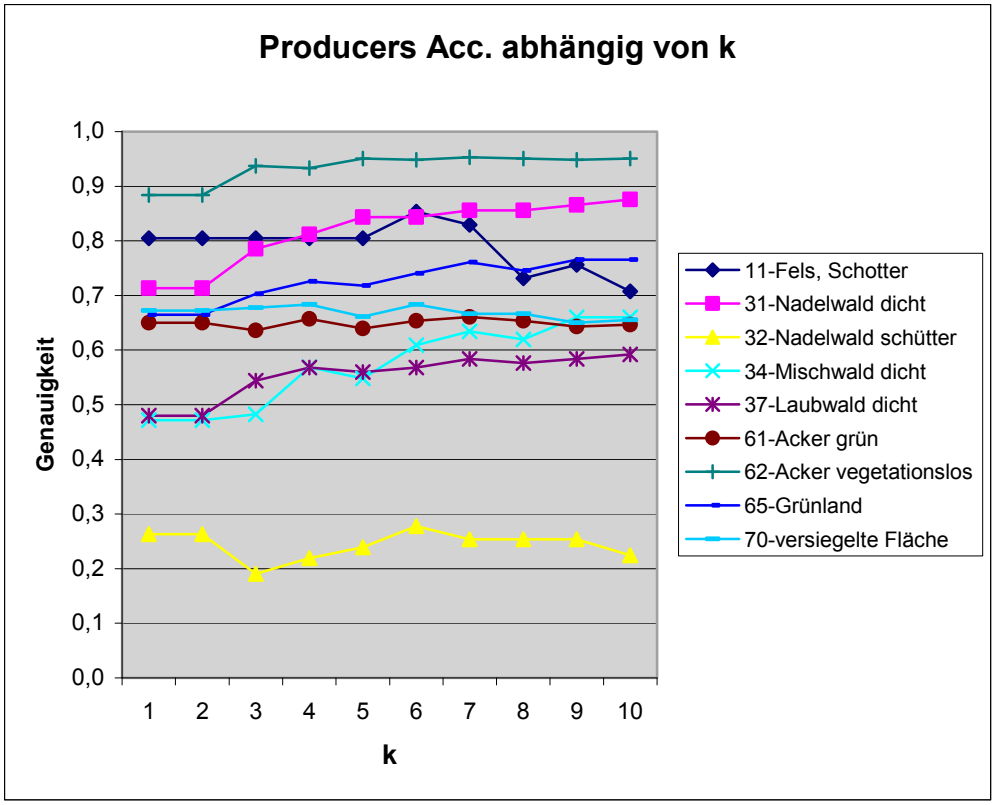


Abb. 8:Producers Accuracy abhängig von k

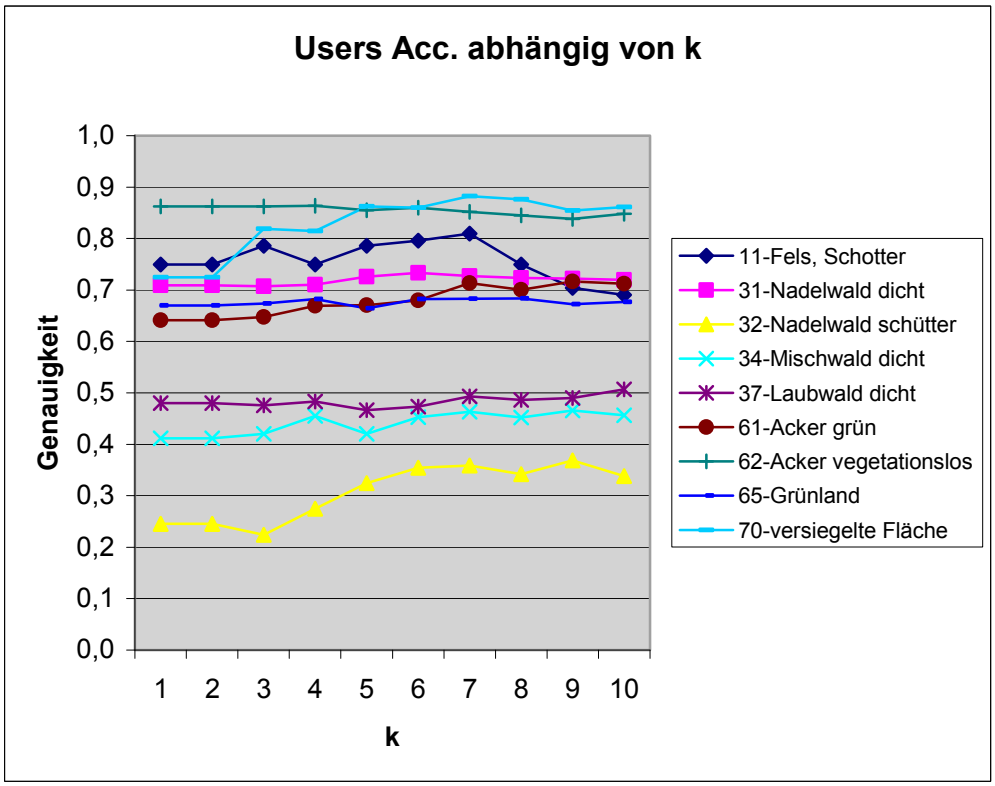


Abb. 9:Users Accuracy abhängig von k

Es ist im Allgemeinen ein leichter Anstieg der Genauigkeiten ab k=3 erkennbar. Bei manchen Landbedeckungsklassen fällt die Genauigkeit ab einem k von 8 wieder leicht ab. Viele der wichtigen Klassen reagieren kaum auf eine Änderung von k. Es ist dennoch sinnvoll, den Bereich von k zwischen 4 und 8 einzugrenzen, da hier einige Klassen positiv beeinflusst werden.

3.3.3 Suchradien (Lage und Höhe)

Es wird von der Annahme ausgegangen, dass eine Landbedeckungsklasse in ähnlichen Höhenlagen auch ähnliche spektrale Eigenschaften aufweist. Daher wird im kNN-Programm die Möglichkeit verwendet, einen Suchradius betreffend der Höhenlage der Trainingsdaten zu definieren. Dieser wird in den folgenden Schritten so eingeschränkt, dass nur Trainingsdaten zur Klassifizierung verwendet werden, welche sich innerhalb dieses Suchradius, also in einer ähnlichen Höhenlage wie das betrachtete Pixel befinden. Auch wird von der Annahme ausgegangen, dass eine Landbedeckungsklasse in ihrer näheren Umgebung mehr Übereinstimmungen als in weiterer Entfernung findet. Daher wird der horizontale Suchradius ebenfalls eingeschränkt. Die Auswirkungen dieser Maßnahmen auf die Genauigkeiten der einzelnen Klassen wird betrachtet und in die Optimierung der kNN-Klassifizierung mit einbezogen.

Einbeziehen der Höheninformation

Zuerst wird der Suchradius für die Höhe mit 100m begrenzt. Als Ergebnis erhält man einen Anstieg der Genauigkeit bei k=1 oder k=2. Bei größeren k fällt die Genauigkeit. Dieses Ergebnis zeigt, dass die Einschränkung des Suchradius für die geografische Höhe nicht zielführend ist da, wie man im vorhergehenden Kapitel zu verschiedenen k nächsten Nachbarn sieht, ein Optimum nicht im niedrigen Bereich von k liegt. Bei einem Suchradius von nur 50m sinkt die Genauigkeit bei allen k. Die Ursache für diese Verschlechterungen der Genauigkeit dürfte darin liegen, dass bei kleinen Suchradien zu wenig spektral ähnliche Referenzpixel pro Klasse vorliegen. Bei einer Einschränkung der Höhendifferenz auf 150m treten weder merkliche Verbesserungen noch Verschlechterungen ein. Man muss also annehmen, dass eine Einengung des Suchradius in der Höhe die Ergebnisse nicht positiv beeinflusst.

Einzig die Kategorien Latsche, Hochalpine Vegetation und Nadelwälder erfahren eine geringe Steigerung der Genauigkeit, während alle anderen Klasen nicht davon profitieren. Da es in dieser Arbeit aber hauptsächlich um die Untersuchung der großen Landbedeckungsklassen Wald, Ackerland und Grünland geht, ist die Verbesserung der weniger wichtigen Klassen auf Kosten der wichtigeren zu vernachlässigen. Die Idee der Verwendung der Höheninformation zur Steigerung der Genauigkeit wird deshalb verworfen.

Einbeziehen der Lageinformation

Der horizontale Suchradius des kNN-Programms wird für den ersten Versuch auf 10.000m beschränkt. Diese Maßnahme führt zu einer bedeutenden Verschlechterung der Genauigkeit. Auch die Begrenzung auf 50.000m führt zu einer Verschlechterung, welche allerdings weniger dramatisch ausfällt als beim engeren Suchradius. Es scheint, dass sich eine Beschränkung des Suchradius nicht positiv auf die Genauigkeit auswirkt, daher wird für die weitern Berechnungen auf eine horizontale Beschränkung des Suchrasters verzichtet.

3.3.4 Bandgewichtungen

Bei ersten Probeversuchen mit verschiedenen Gewichtungen der einzelnen Bänder wird festgestellt, dass meist eine Verbesserung bei einer Klasse zu einer gleichzeitigen Verschlechterung einer oder mehrerer anderer Klassen führt. Der Optimierungsvorgang ist ein überaus komplexer und die Wechselwirkungen zwischen den Gewichtungen und den Landbedeckungsklassen sind nicht oder kaum durchschaubar oder vorhersehbar. Man muss unweigerlich Kompromisse eingehen und entscheiden, ob manche Klassen „wichtiger“ für die gewünschten Ergebnisse sind als andere.

In den folgenden Arbeitsschritten wird versucht herauszufinden, wie die Ergebnisse auf Änderungen der Berechnungsparameter reagieren. Durch verschieden hohe Gewichtung der einzelnen Bänder und die Reaktionen der Genauigkeiten der Landbedeckungsklassen auf diese Veränderung wird mittels Iteration versucht, die Zusammenhänge zwischen Gewichtungen und Genauigkeiten zu ergründen und daraus eine optimale Gewichtung für die nachfolgenden Prozesse zu finden.

Jedes einzelne Band, die Bandkombination (Band 4 dividiert durch Band 3) sowie Slope, Hillshade und Aspect werden einzeln untersucht. Die Ergebnisse werden als Diagramm des Genauigkeitsverlaufes dargestellt. Zur Erzeugung dieser Diagramme wird so vorgegangen, dass bei der Gewichtung des jeweiligen Bandes beim Wert 0 begonnen und in 10er Schritten gesteigert wird, bis keine Verbesserung der Genauigkeit mehr eintritt. Die Gewichtungen der übrigen Bänder werden auf einen konstant bleibenden Referenzwert gesetzt.

Hillshade

Der Verlauf der Genauigkeit mit der Veränderung der Gewichtung von Hillshade ist sehr gleichmäßig. Mit k=7 wurde bei einer Gewichtung von 70 der höchste Wert von Kappa = 0.652 erreicht, welcher sich allerdings kaum von den anderen unterscheidet. Die Gewichtung von 70 wird in weiterer Folge verwendet.

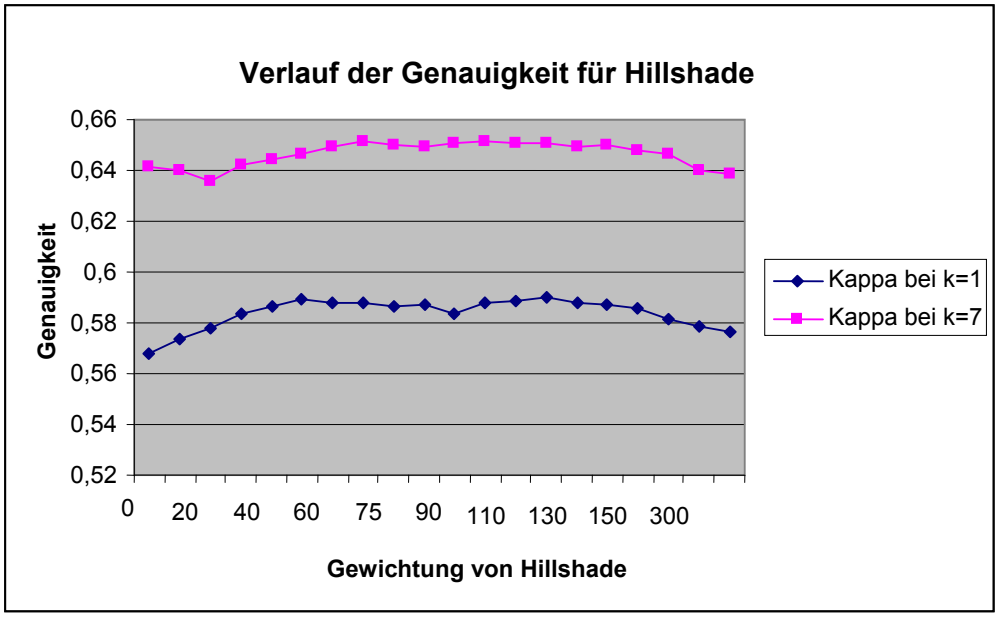


Abb. 10: Verlauf von Kappa bei k=1 und k=7 nach der Gewichtung von HS

Slope

Eine Änderung der Gewichtung von 0 auf einen anderen Wert führt nur zur Verschlechterung der Genauigkeit und ist absolut nicht zielführend. Die Gewichtung von 0 wird beibehalten.

Aspect

Da der Verlauf der Genauigkeit sich mit Veränderung der Gewichtung von Aspect nicht verändert wird eine zufällige Gewichtung von 25 verwendet.

Band1

Das Maximum der Genauigkeit liegt, mit k=7 und einer Gewichtung von 80, bei einem Wert von Kappa = 0.653

Band2

Das Maximum der Genauigkeit liegt, mit k=7 und einer Gewichtung von 230, bei einem Wert von Kappa = 0.654

Band3

Das Maximum der Genauigkeit liegt, mit k=7 und einer Gewichtung von 470, bei einem Wert von Kappa = 0.655

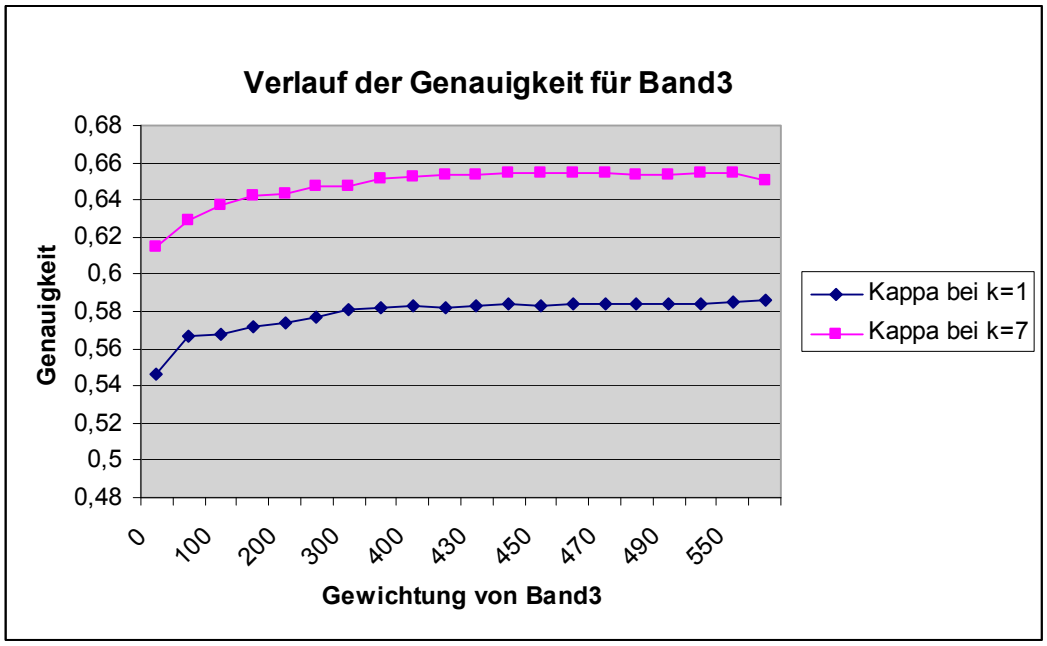


Abb. 11: Verlauf von Kappa bei k=1 und k=7 nach der Gewichtung von Band 3

Band4

Das Maximum der Genauigkeit liegt, mit k=7 und einer Gewichtung von 170, bei einem Wert von Kappa = 0.655

Band5

Das Maximum der Genauigkeit liegt, mit k=7 und einer Gewichtung von 25, bei einem Wert von Kappa = 0.655

Band6

Das Maximum der Genauigkeit liegt, mit k=7 und einer Gewichtung von 175, bei einem Wert von Kappa = 0.655

Band7

Das Maximum der Genauigkeit liegt, mit k=7 und einer Gewichtung von 20, bei einem Wert von Kappa = 0.655

Bandkombination

Das Maximum der Genauigkeit liegt, mit k=7 und einer Gewichtung von 20, bei einem Wert von Kappa = 0.655. Der Verl auf der Genauigkeit der Bandkom bination wird in Abbildung 12 dargestellt. Als Band kombination wird der „Ratio Vegetation Index“ (Band 4 dividiert durch Band 3) gewählt (RVI = L_{nir} / L_{rot})

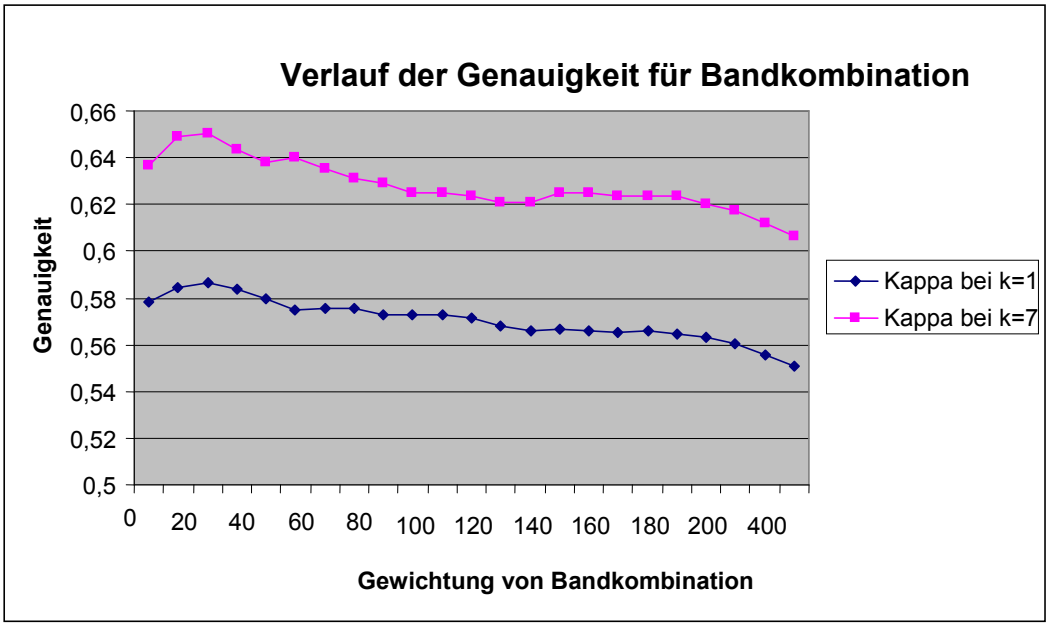


Abb. 12: Verlauf von Kappa bei K=1 und k=7 nach der Gewichtung der Bandkombination B4/B3

3.3.5 Zusammenfassende Betrachtung der Optimierungsschritte

Durch das oben durchgeführte Verfahren zur Anpassung der Gewichtung der Bänder ist ein opt imierter Satz von Koeffizient en entstanden. Dieser sieht nach einer Neuskalierung mit dem Faktor 0,01wie folgt aus:

B1	B2	B3	B4	B5	B6	B7	BC					HS	SL	ASP
0.8	2.3	4.7	1.7	0.25				1.75	0.2	0.2	0.7	0		0.25

Mit diesen Gewichtungen werden bei k=7 folgende Genauigkeiten erreicht:

- Kappa = 0.655
- Overall Accuracy = 0.692

Einflüsse verschiedener Werte von k

In Abbildung 13 ist deutlich zu s ehen, dass die höchste Genauigkeit bei k=7 erreicht wird. Das ist nicht verwunderlic h, da anh and der Abbildungen 8 und 9 im Kapitel „Werte von k“ ein k z wischen 4 und 8 als passender Wert festgel egt wird und alle weiteren Optimierungsschritte darauf abgestimmt sind.

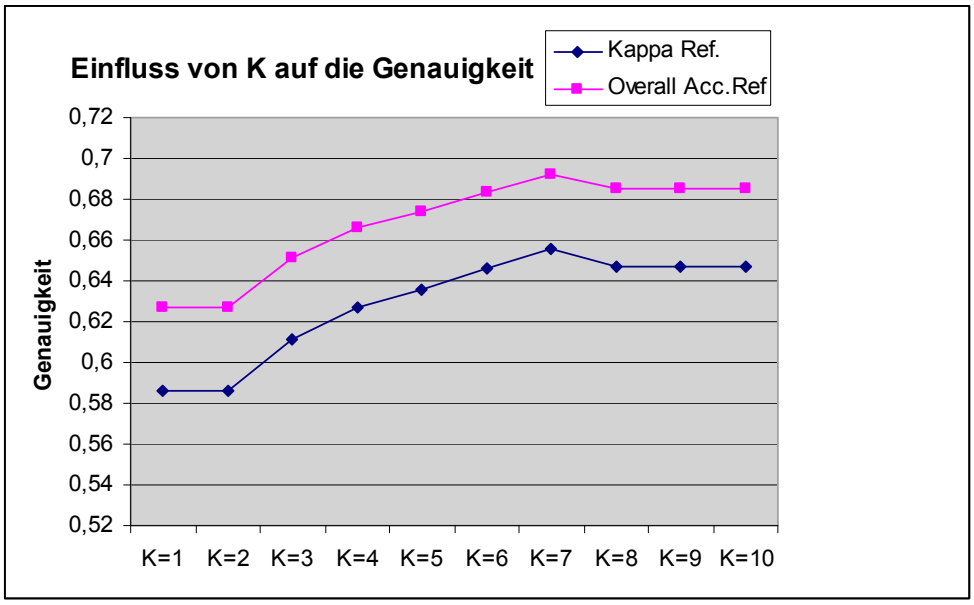


Abb. 13: Einfluss von k auf die Genauigkeit bei optimaler Gewichtung der einzelnen Bänder

Einflüsse der Outlier-Entfernung und der Bandgewichtungen

Der Fortschritt in der Verbesserung der Gen auigkeit der verschiedenen Schritte lässt sich in den Abbi ldungen 14 und 15 erkennen. Beim Vergleich der beiden

Abbildungen fällt der positive Einfluss eines höheren k auf die Genauigkeit auf. Die verworfenen Versuche der Beschränkung des horizontalen und des vertikalen Suchradius sind nicht in den Abbildungen dargestellt.

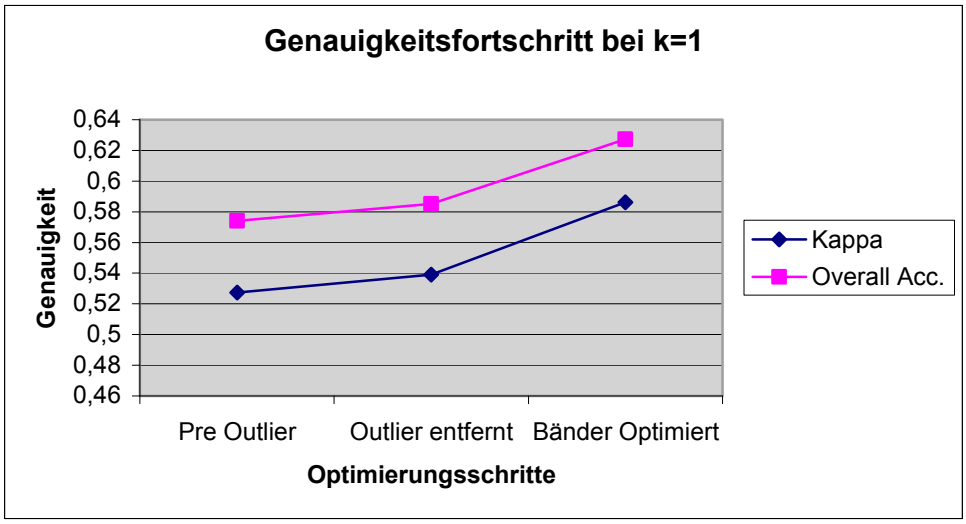


Abb. 14: Genauigkeitsverlauf von Kappa und Overall Acc. bei k=1

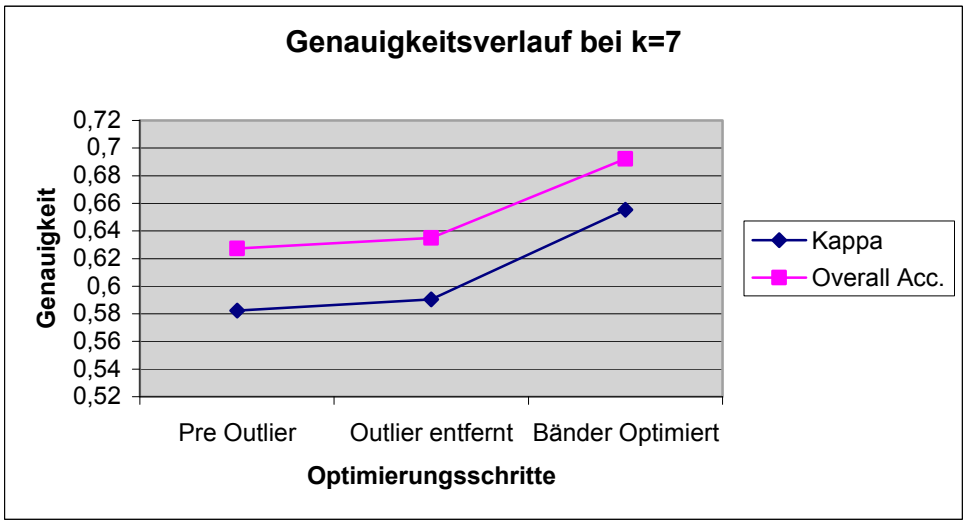


Abb. 15: Genauigkeitsverlauf von Kappa und Overall Acc. bei k=7

Der in Abbildung 14 und 15 dargestellte Optimierungsschritt „Pre Outlier“ bezeichnet die Genauigkeit vor dem Beginn der Trainingsdatenoptimierung, also vor dem Entfernen der Ausreißer. Im zweiten dargestellten Schritt sind die Ausreißer entfernt und die dadurch erzielte Steigerung der Genauigkeit wird ersichtlich. Den erkennbar größten Fortschritt in der Klassifizierungswahrscheinlichkeit erreicht man durch die Optimierung der einzelnen Landsatbänder, der Bandkombination sowie von Slope, Hillshade und Aspect, welche im Schritt „Bänder Optimiert“ zusammengefasst sind.

Wie in den Abbildungen 14 und 15 ersichtlich ist, trägt die Gewichtung der Bänder allgemein sehr zur Genauigk eit bei. Einzelne Bänder wir klich optimal zu gewichten bringt aber, wie in Kapitel 3.3.4 " Bandg ewichtungen“ erkennbar, nur eine eher geringe Steigerung der Genauigkeit. Man sollte aber wegen der doch deutlich erkennbaren Vorteile nicht darauf verzichten, die Bänder gut zu gewichten.

Einflüsse der einzelnen Landsat Bänder

Um die Einflüsse der einz elnen Bänder auf die Genauigkeit festzustellen, wird ausgehend von den Gewichtungen mit den optimalen Koeffizienten jedes einzelne Band nacheinander mit einer Gewichtung von 0 versehen. Die optimalen Gewichtungen der nicht betrachteten Bänder bleiben dabei unverändert. Die errechneten Veränderungen der Genauigkeit werden dann miteinander verglichen. Es wird versucht, dadurch Rücksc hlüsse auf möglic herweise für die Gesamtgenauigkeit besonders bedeutende Bänder zu bekommen.

In den Grafiken erkennt man, durch ve rschiedene Abweichungen vom Referenzwert der optimalen Gewichtung aller Bänder nach unten, den unterschiedlichen Einfluss der einzelnen Bänder. Je deutlic her die Genaui gkeit nachlässt, desto größer ist der Einfluss des jeweiligen Bandes.

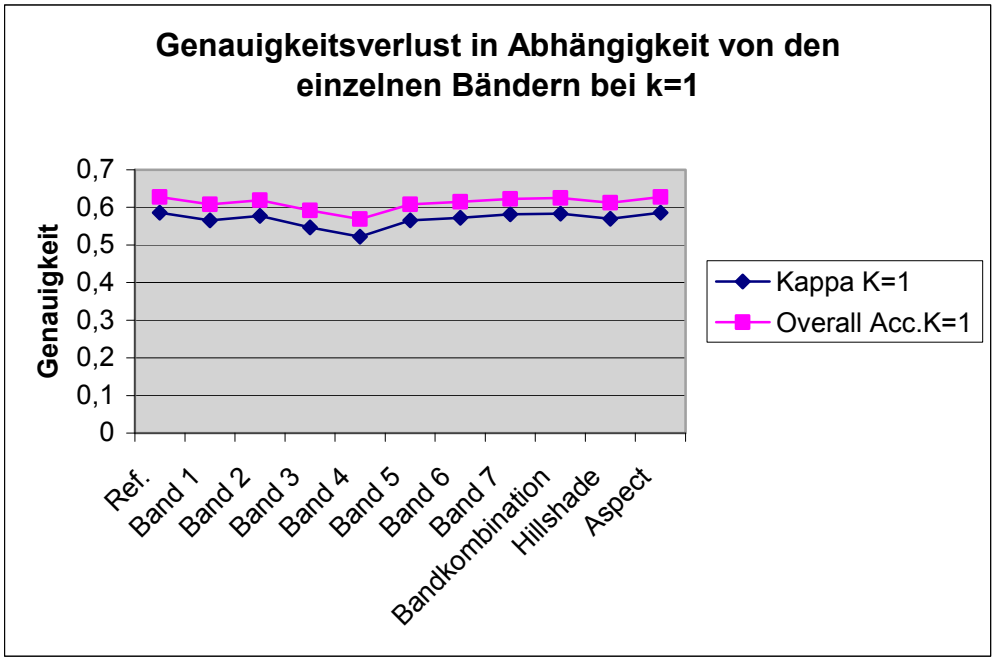


Abb. 16: Genauigkeitsverlust beim Weglassen der einzelnen Bänder bei k=1

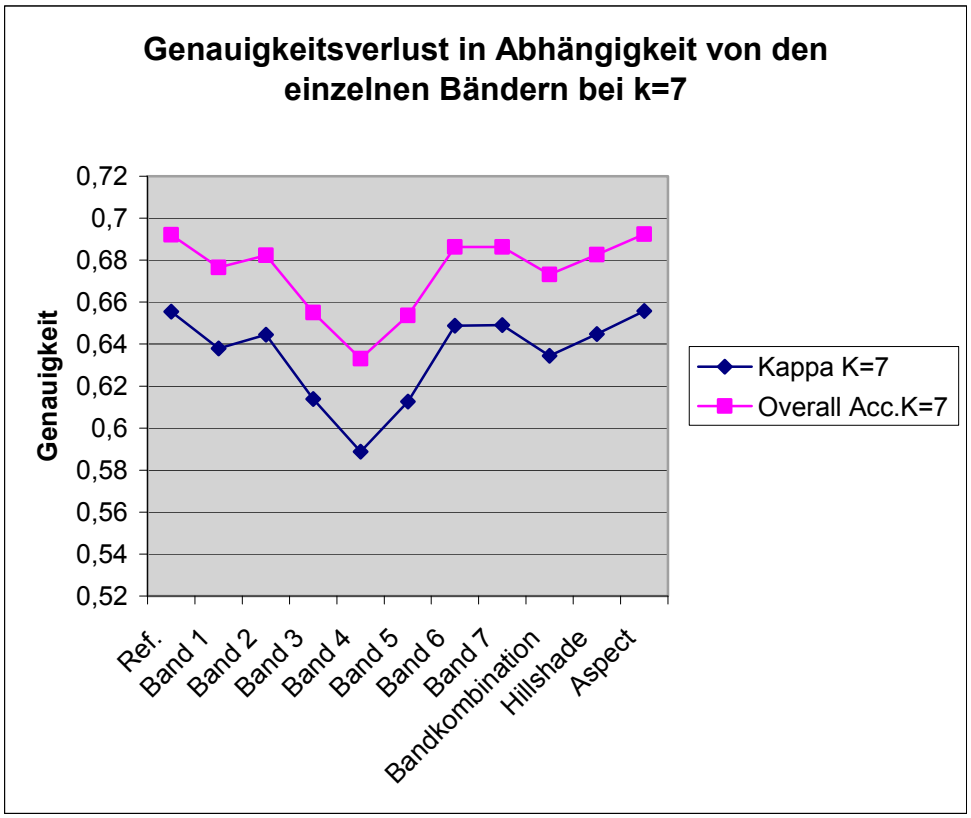


Abb. 17: Genauigkeitsverlust beim Weglassen der einzelnen Bänder bei k=7

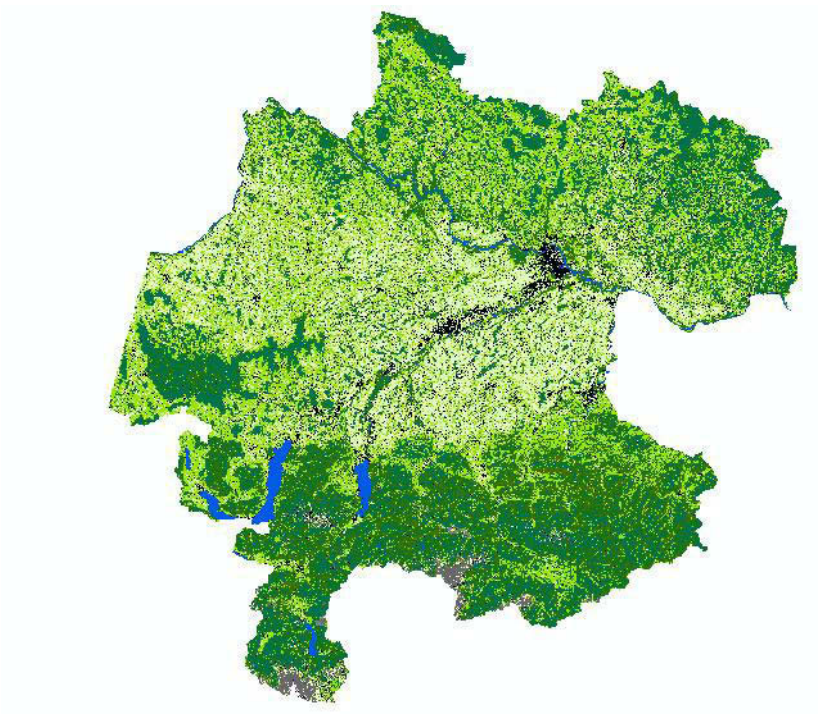
In den Abbildungen 16 und 17 kann man für zwei verschiedene Werte von k den Einfluss der einzelnen Bänder auf die Gesamtgenauigkeit erkennen. Es zeigen sich deutliche Unterschiede zwischen den Einflüssen der einzelnen Bänder. So hat das Weglassen der Bänder 3, 4 und 5 eine deutlichere Veränderung der Genauigkeit zufolge als das Weglassen der restlichen Bänder.

Es lässt sich daraus ableiten, dass bei der Optimierung der Gewichte der Bänder 3, 4 und 5 mit besonderer Sorgfalt vorgegangen werden sollte und die restlichen Bänder nicht so streng betrachtet werden müssen.

Generell ist zu den hier beschriebenen Versuchen der Parameteroptimierung anzumerken, dass die komplexen Wechselwirkungen zwischen den Parametern nicht berücksichtigt sind, da dies den Rahmen der vorliegenden Diplomarbeit gesprengt hätte. Eine mögliche Methode zur globalen Optimierung der Parameter einer kNN-Klassifikation wäre die Verwendung eines genetischen Algorithmus während des Optimierungsverlaufs (Tomppo 2004).

3.4 Klassifizierung des Satellitenbildes

Durch die Verwendung des Moduls MapMaker im kNN-Programm wird die Landbedeckung von Oberösterreich ermittelt. Das Modul liefert die Ergebnisse detailliert in Kartenform. In Abbildung 18 werden diese Ergebnisse, zusammengefasst zu „Landbedeckungsgroßklassen“ in Naturfarben dargestellt, um einen einfachen Eindruck der Landbedeckung Oberösterreichs zu gewinnen.



Legende:

- ***Schwarz*** *– Siedlung*
- ***Dunkelgrün*** *– Wald*
- ***Grün*** *– Grünland*
- ***Gelb*** *– Ackerland*
- ***Blau*** *– Wasser*
- ***Grau*** *– Fels*
- ***Weiß*** *- Schnee*

Abb. 18: vereinfachte Landbedeckung Oberösterreichs

3.4.1 Bezirksweise Flächenstatistik

Die Ergebnisse der Landbedeckungsklassifizierung aus dem Map-Maker-Modul des kNN-Programms werden unter Verwendung der in Kapitel 2.3.3 beschriebenen Maske auf Bezirksebene zusammengefasst. Die damit entstandene Tabelle enthält für jeden Bezirk die Flächen jeder einzelnen Landbedeckungsklasse. In weiterer Folge werden diese Daten mit jenen der Österreichischen Waldkarte sowie den Agrardaten aus der Invekos-Datenbank verglichen, um die Genauigkeit der automatischen Klassifikation zu überprüfen.

3.4.2 Fuzzy-Klassifizierung

Bei der Fuzzy-Klassifizierung wird einem Pixel nicht wie bei der harten Klassifizierung ein einziger Wert zugewiesen, sondern mehrere Werte, die nach ihrer Wahrscheinlichkeit gestaffelt sind. Bei der harten Klassifizierung bestimmt die „stärkste“ Wahrscheinlichkeit die Klasse des Pixels, welcher es mit 100% zugerechnet wird. Bei der Fuzzy-Klassifizierung überstimmt der größte Wert die kleineren nicht, sondern jeder im Pixel vorkommenden Klasse wird ihrer Wahrscheinlichkeit entsprechend ein Teil der Pixelfläche zugewiesen. Dadurch kann es im Gegensatz zur harten Klassifikation auch vorkommen, dass keine ganzzahligen Werte bei der Pixelsumme pro Klasse und Bezirk entstehen. Bei der hier verwendeten Klassifikation mit 7 nächsten Nachbarn könnte sich ein Pixel also nach folgenden Kriterien zusammensetzen: 5 der nächsten Nachbarn gehören der Klasse „Nadelwald dicht“ an, einer der Klasse „Mischwald dicht“ und einer der Klasse „Nadelwald schütter“. Damit wäre die Aufteilung der Pixelfläche folgende: 72% des Pixels wären „Nadelwald dicht“; 14% wären „Nadelwald schütter“ und ebenfalls 14% wären „Mischwald dicht“.

Um die Daten aus der Fuzzy-Klassifikation entsprechend der bezirksweisen Statistik auswerten zu können, wird das fuzzy-klassifizierte Bild, welches 25 Kanäle entsprechend der einzelnen Landbedeckungsklassen umfasst, in 25 Einzelbilder aufgeteilt. Es entstehen dadurch detaillierte Abbildungen der einzelnen Landbedeckungsklassen. Auf den unterschiedlichen Bildern der einzelnen Landbedeckungsklassen ist sehr gut der Unterschied zwischen den verschiedenen Landbedeckungen zu erkennen. Bei der in Abbildung 19 dargestellten Klasse „versiegelte Flächen“ sind Städte und Siedlungen beinahe weiß dargestellt, Wälder hingegen sehr dunkel bis fast schwarz.

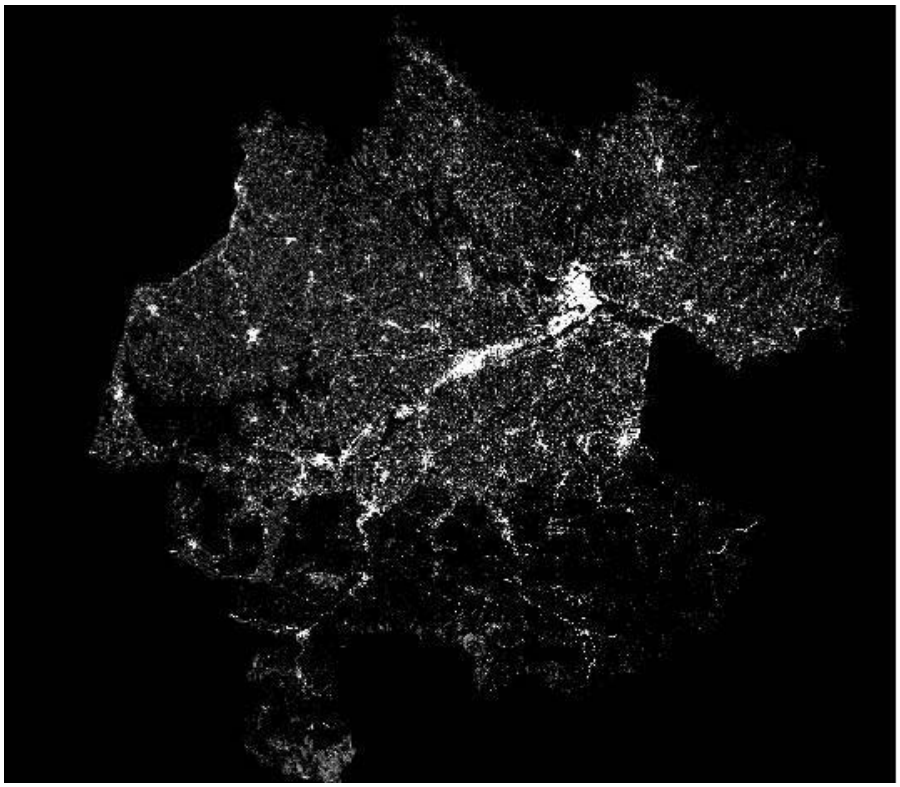


Abb. 19: Fuzzy klassifizierte Klasse 70: Versiegelte Fläche

Bei der in Abbildung 20 darges tellten Klasse „Nadelwald dicht“ werden Wälder sehr hell und Siedlungsgebiete sehr dunkel, bis fast schwarz dargestellt.

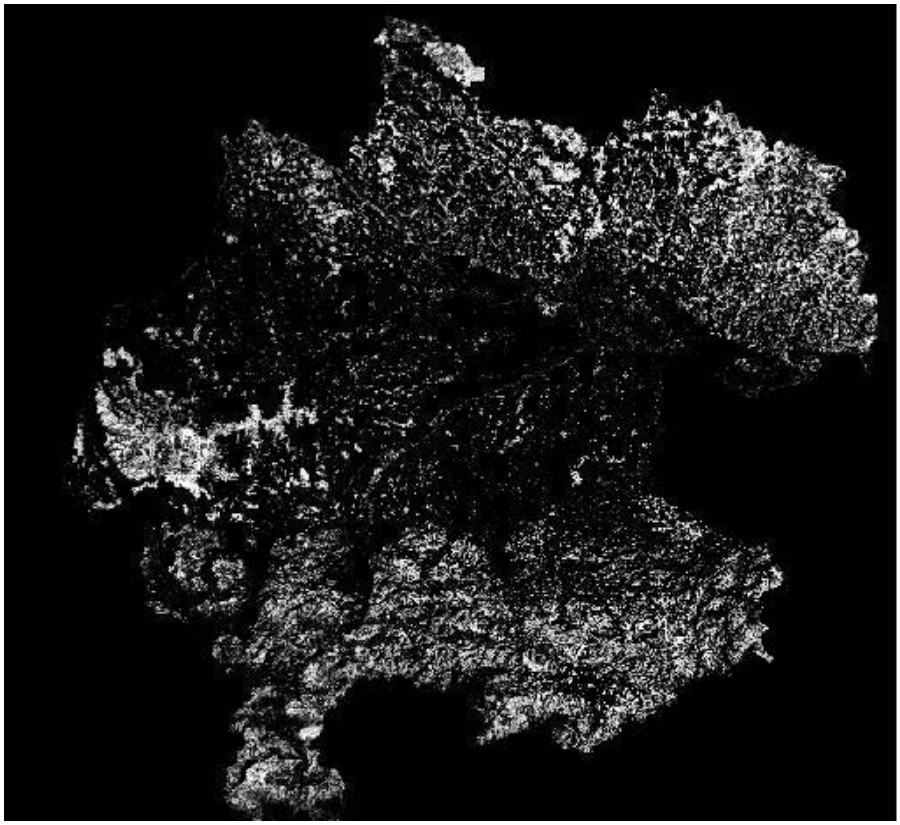


Abb. 20: Fuzzy klassifizierte Klasse 31: Nadelwald dicht

Die Aufteilung der Fuzzy-Klassifizierungser gebnisse auf Bezirke geschieht mittels ArcGis. Dadurch entstehen für jede einzelne Klasse aus dem dazu passenden Bild Tabellen, welche die Landbedeckung durch di e jeweilige Klasse auf Bezirksebene darstellen. Die Tabellen ze igen an, wie viele Pixel pro Bezirk die jeweilige Landbedeckungsklasse haben. Diese Pixelanzahl ist, um eine Fläche zu bekommen, noch mit der Pixelfläche von 30m x 30m zu multiplizieren.

Diese Einzeltabellen si nd danach z u einer alle Landbedeckungsklassen umfassenden Tabelle zusammen zu fassen, um einen Vergleic h mit den Daten aus der Waldkarte, aus Invekos oder aus der harten Klassifikation zu ermöglichen.

4. Ergebnisse der kNN Klassifikation

4.1 Ergebnisse für das gesamte Bundesland

Die einzelnen Landbedeckungs klassen werden im Wesentlichen nur mit spektral ähnlichen Klassen v erwechselt. Eine vor handene Verwechslung der einzelnen Waldklassen ist als nicht gravierend anzusehen und auch kaum vermeidbar. E s kommt hier häufig z u Überschneidungen einer Waldklasse mit unterschiedlic hem Kronenbedeckungsgrad. Diese Klassen glei chen Waldes mit verschiedenem Kronendeckungsgrad werden bei den späteren Unters uchungen ohnehin zusammengefasst. Es kommt auch zu le ichten Überschneidungen zwischen Mischwald mit Nadel- oder Laubwald. Dies e Überschneidungen sind abhängig vom Grad der Durchmischung des Waldes. Ü berschneidungen von Laub- und Nadelwald treten kaum auf.

In Tabelle 3 sind alle für die weiter en Aufgaben wichtigen Landbedeckungs klassen und ihre Genauigkeiten aufgelistet. Weniger wicht ige Klassen erzielen leider nicht so gute Genauigkeiten. Dies ist allerdings ni cht verwunderlich, da durch das „Tuning“ mit Kappa als Maß der Genauigkeit haupts ächlich die starken und für dies e Arbeit wichtigen Klassen weiter gestärkt werden. Es liegt auch im Ziel dieser Arbeit, diese besonders zu untersuchenden Klassen in ihrer Genauigkeit zu fördern.

	producer´s.Acc.	*user´s Acc.*	*Landbedeckung*
Cat11	0,756	0,775	Fels, Schotter
Cat12	0,333	0,556	Fels, Schotter mit Veg.
Cat13	0,850	1,000	Schnee
Cat20	0,961	0,980	Wasser
Cat31	0,872	0,742	Nadelwald
Cat34	0,653	0,470	Mischwald
Cat37	0,648	0,513	Laubwald
Cat61	0,695	0,758	Acker grün
Cat62	0,955	0,857	Acker vegetationslos
Cat65	0,780	0,716	Grünland
Cat66	0,835	0,668	Grünland gemäht
Cat70	0,674	0,851	Siedlung, Versiegelte Fl.

Tab.3: Wichtige Landbedeckungsklassen und ihre Genauigkeiten

Bei der Durchsicht von Tabelle 4 erkennt man, dass die fuzzy-klassifizierten Daten teilweise besser mit den als Referenzdaten verwendeten Flächen aus der Waldkarte und den Invekos-Daten übereinstimmen als die Daten aus der harten Klassifikation. Dieser Genauigkeitsgewinn ist jedoch nur bei Betrachtung Gesamt-Oberösterreichs zu erkennen.

Oberösterreich		„Referenzdaten“	harte Klassifikation	Fuzzy Klassifikation
Ges. Wald	ha	463.990	456.374	458.404
	%	42,4	41,7	41,9
Laubwald	ha	72.531	60.280	67.861
	%	6,6	5,5	6,2
Nadelwald	ha	298.207	249.355	2.54.388
	%	27,3	22,8	23,3
Mischwald	ha	74.927	142.358	131.613
	%	6,9	13,0	12,0
Ackerflächen	ha	251.181	242.645	259.457
	%	23,0	22,2	23,7

Tab.4: Gegenüberstellung der Ergebnisse aus den automatischen Klassifikationen mit den Referenzdaten für die Landbedeckungsklassen Wald und Ackerland

In Abbildung 21 werden auf Bezirksebene die Waldflächen aus den automatischen Klassifikationen den Referenzdaten gegenübergestellt.

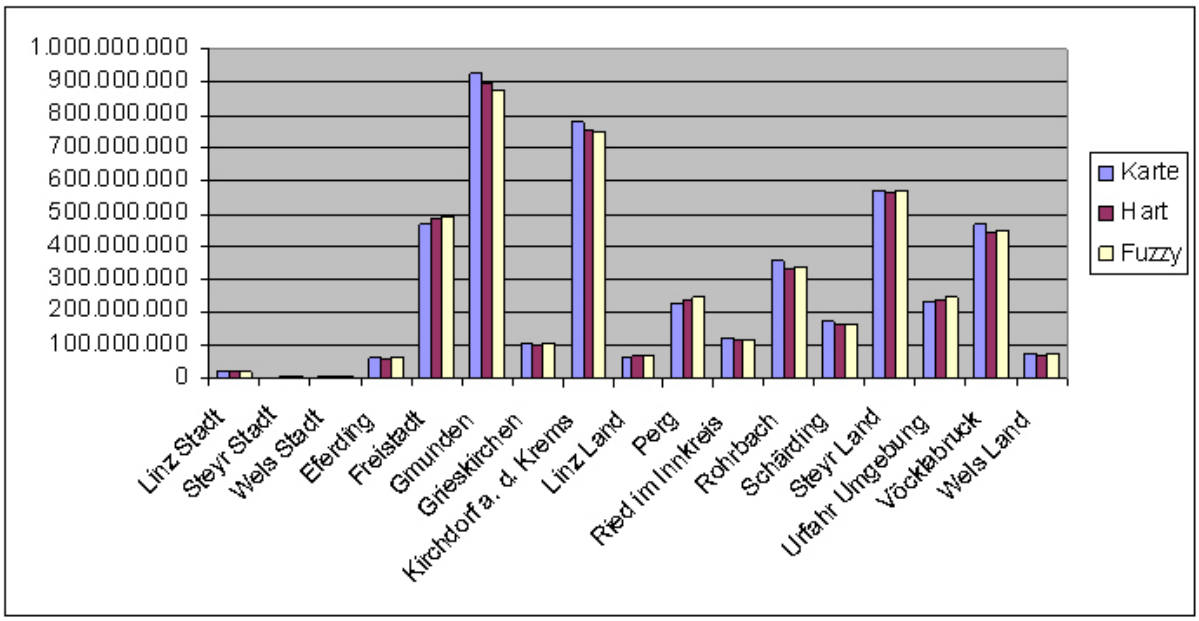


Abb. 21: Gegenüberstellung der Waldflächen aller Bezirke aus den automatischen Klassifikationen mit den Daten der Österreichischen Waldkarte

Die Fuzzy-Daten einzelner Bezirke stimmen nicht zwingend besser mit den Referenzdaten überein als die Daten aus der harten Klassifikation dies tun. So ist beispielsweise bei der Betrachtung der gesamten Waldfläche Oberösterreichs durch die Verwendung der Fuzzy-Daten eine Annäherung an die Referenzdaten (Waldkarte

42,4%) von 41,7% (harte Klassifikation) auf 41,9% Landbedeck ung festzustellen. Während im Bezirk Wels Stadt ebenfalls eine Annäherung an die Refer enzdaten (7,9%) von 7,2% auf 7,8% stattfindet, versch lechtert sich das Ergebnis im Bezirk Gmunden von 62,7% auf 61,2% bei einer Referenzfläche von 64,7%.

4.2 Ergebnisse der bezirksweisen Flächenstatistik für Wald und Ackerland

Die Ergebnisse der bezirksweisen statis tischen Berechnung aus Kapitel 3.4. 1 werden in Abbildung 22 als Kar te dargestellt. Die a bgebildeten Tortendiagramme stellen die Flächenv erteilung der Landbe deckungen auf Bez irksebene dar. Der besseren Übersichtlichkeit wegen we rden die Klassen z u wenigen, aussagekräftigeren Klassen zusammengefügt.

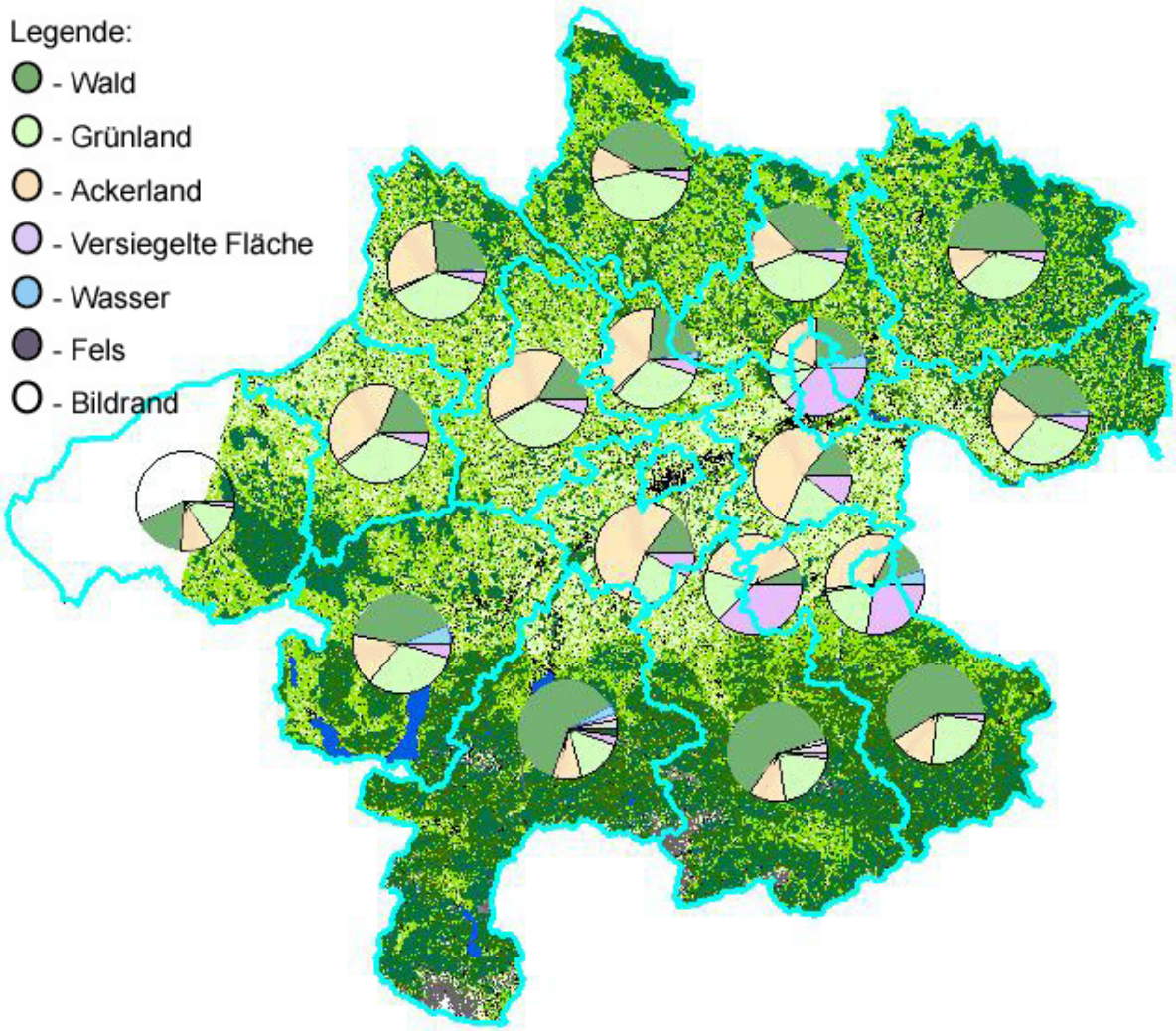


Abb. 22: Karte der bezirksweisen Flächenstatistik

4.2.1 Vergleich der Klassifizierungsergebnisse mit der Österreichischen Waldkarte

In diesem Kapitel sollen die Ergebnisse aus der harten Klassifikation und der Fuzzy Klassifikation mit den Referenzdaten der Österreichischen Waldkarte verglichen werden.

Gesamter Waldbestand Oberösterreichs

Der Vergleich des gesamten Waldbestandes in Oberösterreich zeigt eine Abweichung der Klassifikationsergebnisse von den Referenzdaten von maximal 0,7%. Die verschiedenen Datensätze der Waldbestände decken sich in der Gesamtfläche deutlich. Dies lässt sich auch beim Vergleich auf Bezirksebene erkennen. Einzig Braunau fällt in der Statistik ungewöhnlich auf, was daran liegt, dass bei der automatischen Klassifizierung für diesen Bezirk ein Großteil der Fläche fehlt. Dieser Bezirk darf also nicht so wie die anderen direkt verglichen werden, sondern benötigt später eine gesonderte Betrachtung. Aus diesem Grund wird der Bezirk Braunau in den Tabellen dieses Kapitels nicht aufgeführt.

	W Karte %	W Hart %	W Fuzzy %	Diff. H-F	Diff. K-H	Diff. K-F
Linz Stadt	17,44	21,38	21,79	-0,41	-3,94	-4,35
Steyr Stadt	10,82	14,11	14,97	-0,86	-3,29	-4,15
Wels Stadt	7,93	7,18	7,79	-0,61	0,75	0,15
Eferding	23,19	22,22	22,90	-0,68	0,97	0,30
Freistadt	46,98	48,80	49,64	-0,83	-1,83	-2,66
Gmunden	64,66	62,72	61,24	1,48	1,94	3,42
Grieskirchen	18,16	17,12	17,94	-0,82	1,04	0,21
Kirchdorf a.d Krems	62,63	60,99	60,39	0,60	1,64	2,24
Linz Land	12,90	14,42	15,02	-0,61	-1,52	-2,12
Perg	37,33	39,07	40,04	-0,97	-1,74	-2,71
Ried im Innkreis	20,63	19,71	20,18	-0,46	0,92	0,46
Rohrbach	43,31	40,23	40,82	-0,59	3,08	2,49
Schärding	27,49	26,37	26,96	-0,59	1,12	0,52
Steyr Land	58,60	58,34	58,70	-0,37	0,26	-0,11
Urfahr Umgebung	35,76	37,15	38,09	-0,94	-1,39	-2,33
Vöcklabruck	43,19	41,00	41,07	-0,07	2,19	2,12
Wels Land	16,35	15,35	15,97	-0,61	0,99	0,38
Oberösterreich	42,42	41,72	41,91	-0,19	0,70	0,51

Tab.5: Gegenüberstellung der Klassifizierungsergebnisse mit den Referenzdaten für den gesamten Waldbestand Oberösterreichs

Die Spalten von Tabelle 5 zeigen neben dem Bezirk die zu diesem gehörige prozentuelle Waldfläche aus der Österreichischen Waldkarte, die errechneten Flächen aus der harten Klassifikation und der Fuzzy Klassifikation. Darüber hinaus

sind die Differenzen von „Hart“ – „Fuzzy“, „K arte“ – „Hart“, „Karte“ – „Fuzzy“, jeweils in Prozent der Bezirksfläche, dargestellt. Dies ermöglicht einen einfachen Vergleich der verschiedenen Waldflächen. Die dargestellten Prozentwerte der Tabelle ergeben sich aus der Waldfläche pro Bezirk bzw. der Differenz der Waldfläche pro Bezirk und der gesamten Bezirksfläche.

Einzig Linz Stadt und Steyr Stadt haben mit mehr als 3% der Bezirksfläche relativ hohe Abweichungen. Diese Abweichungen dürften auf Gärten und Park s zurückzuführen sein, welche nicht in die Österreichische Waldkarte Eingang gefunden haben. Die 3% Abweichung im Bezirk Rohrbach ist wegen der sehr großen Waldbedeckung von mehr als 40% nicht so schwerwiegend zu sehen, ist aber möglicherweise teilweise auch auf das Fehlen eines sehr kleinen Teils der Bezirksfläche zurückzuführen. Dieses kl eine fehlende Stück ist haupt sächlich bewaldetes Gebiet und betrifft die restlichen Klassen kaum bis gar nicht.

Laubwald

Allgemein ist zu erkennen, dass der Laubw aldbestand in Oberösterreich mit ca. 6% Anteil an der Gesamtfläche des Bundes landes sehr gering is t. Einzig der Bezirk Steyr Land liegt mit ca. 12% Flächenant eil deutlich über dem Durchschnitt des Bundeslandes.

	LW Karte %	LW Hart %	LW Fuzzy %	Diff. H-F	Diff. K-H	Diff. K-F
Linz Stadt	9,42	6,65	6,58	0,07	2,77	2,84
Steyr Stadt	2,24	3,11	3,58	-0,47	-0,87	-1,34
Wels Stadt	2,36	1,48	1,84	-0,37	0,88	0,52
Eferding	8,57	5,87	6,06	-0,18	2,70	2,51
Freistadt	4,06	2,93	3,68	-0,75	1,12	0,37
Gmunden	7,58	7,33	8,55	-1,21	0,24	-0,97
Grieskirchen	4,90	3,13	3,52	-0,40	1,77	1,37
Kirchdorf a.d. Krems	7,98	8,19	9,28	-1,09	-0,21	-1,30
Linz Land	5,27	4,45	4,60	-0,15	0,81	0,66
Perg	6,68	5,06	5,79	-0,73	1,62	0,89
Ried im Innkreis	4,06	2,57	2,80	-0,23	1,49	1,26
Rohrbach	6,37	3,60	4,17	-0,57	2,77	2,20
Schärding	6,09	3,81	3,97	-0,15	2,27	2,12
Steyr Land	12,40	11,58	12,52	-0,94	0,82	-0,13
Urfahr Umgebung	6,29	4,49	5,17	-0,68	1,80	1,12
Vöcklabruck	5,15	4,62	5,29	-0,67	0,53	-0,14
Wels Land	4,33	3,08	3,41	-0,33	1,26	0,93
Oberösterreich	6,63	5,51	6,20	-0,69	1,12	0,43

Tab.6: Gegenüberstellung der Klassifizierungsergebnisse mit den Referenzdaten für die Landbedeckungsklasse Laubwald

Der Laubwaldbestand der verschiedenen Datensätze deckt sich wie schon die Ergebnisse des gesamten Waldbestandes auffallend gut. Er weist ledig lich eine Gesamtdifferenz von maximal 1,12% auf. Ei nzig Linz Stadt und Rohrbach fallen mit jeweils 2,77% der Bezirksfläche Differenz und Schärding mit 2,27% als ungenau auf. Die Abweichung in Linz ist wieder au f Waldbestände in Parks und Gärten zurückzuführen, welche in der Österreichi schen Waldkarte nic ht erfasst sind. In Schärding und Rohrbach sind Klass ifizierungsüberschneidungen mit der Klasse Mischwald für die Ungenauigk eiten verantwortlich. Dieses Phänomen wird bei der Betrachtung der Ergebnisse für Mischwald genauer untersucht.

Nadelwald

Der Nadelwaldbestand stellt mit ca . 27% Bedeckung des Bundes landes in Oberösterreich den eindeutig größten Teil des Waldbestandes.

	NW Karte %	NW Hart %	NW Fuzzy %	Diff. H-F	Diff. K-H	Diff. K-F
Linz Stadt	6,35	5,05	6,62	-1,57	1,29	-0,28
Steyr Stadt	5,69	5,74	6,52	-0,77	-0,05	-0,82
Wels Stadt	2,72	3,14	3,58	-0,44	-0,43	-0,87
Eferding	8,23	7,62	8,69	-1,07	0,61	-0,46
Freistadt	38,14	36,98	37,14	-0,16	1,16	1,00
Gmunden	40,33	32,17	31,54	0,63	8,16	8,80
Grieskirchen	9,11	8,75	9,42	-0,67	0,37	-0,31
Kirchdorf a. d. Krems	44,20	30,42	30,56	-0,14	13,78	13,64
Linz Land	6,35	4,86	5,57	-0,72	1,49	0,78
Perg	21,94	22,25	23,33	-1,07	-0,31	-1,39
Ried im Innkreis	13,40	12,51	12,78	-0,27	0,89	0,62
Rohrbach	28,74	26,39	26,93	-0,54	2,35	1,81
Schärding	16,13	14,89	15,62	-0,73	1,24	0,50
Steyr Land	32,09	23,95	25,10	-1,15	8,14	6,99
Urfahr Umgebung	25,71	23,63	24,29	-0,67	2,08	1,42
Vöcklabruck	27,69	22,72	23,58	-0,87	4,97	4,11
Wels Land	7,81	7,69	8,17	-0,48	0,11	-0,37
Oberösterreich	27,26	22,80	23,26	-0,46	4,47	4,01

Tab.7: Gegenüberstellung der Klassifizierungsergebnisse mit den Referenzdaten für die Landbedeckungsklasse Nadelwald

Der Nadelwaldbestand deckt sich mit einer maximalen Differenz von 4,47% (bei 22,8 bis 27,3% Gesamtbedeckung) ebenfalls relati v gut. Kirchdorf a. d. Krems weist aber mit 13,8% der Bezirksfläche eine sehr große Abweichung auf. Auch Gmunden und Steyr Land haben mit mehr als 8% eine rela tiv große Abweichung von den Daten der Österreichischen Waldkarte. Grund für di ese Abweichungen is t wieder eine Überschneidung mit der Landbedeckungsklasse Mischwald.

Mischwald

Als Referenzdatensatz werden die bei den Klassen „Laubwald-dominierter Mischwald“ und „Nadelwald-dominierter Mischwald“ aus der Österreichischen Waldkarte zu einer Klasse zusammengefasst. Dadurch ermöglicht sich ein direkter Vergleich mit den durch die automatisc he Klassifikation entstandenen Daten der Klasse Mischwald.

	MW Karte %	MW Hart %	MW Fuzzy %	Diff. H-F	Diff. K-H	Diff. K-F
Linz Stadt	1,59	9,68	8,59	1,09	-8,09	-6,99
Steyr Stadt	2,89	5,26	4,87	0,38	-2,37	-1,99
Wels Stadt	0,22	2,56	2,36	0,20	-2,34	-2,14
Eferding	2,35	8,70	8,10	0,60	-6,35	-5,75
Freistadt	2,95	8,89	8,78	0,11	-5,94	-5,84
Gmunden	16,48	21,28	19,37	1,91	-4,80	-2,89
Grieskirchen	1,00	5,24	5,00	0,25	-4,24	-3,99
Kirchdorf - Krems	10,37	21,43	19,45	1,98	-11,06	-9,08
Linz Land	1,12	5,10	4,85	0,26	-3,99	-3,73
Perg	6,18	11,76	10,91	0,85	-5,58	-4,73
Ried im Innkreis	1,13	4,63	4,60	0,03	-3,50	-3,46
Rohrbach	3,43	10,24	9,72	0,52	-6,81	-6,29
Schärding	1,61	7,67	7,37	0,29	-6,06	-5,76
Steyr Land	12,97	22,63	20,72	1,91	-9,66	-7,75
Urfahr Umgebung	2,09	9,03	8,61	0,42	-6,95	-6,53
Vöcklabruck	9,76	13,44	12,01	1,43	-3,67	-2,24
Wels Land	1,66	4,58	4,39	0,19	-2,92	-2,73
Oberösterreich	6,85	13,01	12,03	0,98	-6,16	-5,18

Tab.8: Gegenüberstellung der Klassifizierungsergebnisse mit den Referenzdaten für die Landbedeckungsklasse Mischwald

Die Flächenanteile von Misch wald laut Österreichischer Waldkar te und Mis chwald aus den automatischen Klassifikationen deck en sich auffallend schlecht. Es wir d beinahe doppelt so viel Misc hwald klassifiziert wie in der Karte als Mischwald ausgewiesen wird. Allerdings deckt die Fläche des zu viel klassifizierten Mischwaldes das leichte Defizit sowohl bei Laub- als auch bei Nadelwa ld ab. Ein gutes Beispiel dafür ist der Bezirk Kirchdorf a.d. Krems. Mit 1,64% (Tab.5) der Bezirksfläche ist die Gesamtabweichung sehr gerin g, allerdings mit sehr großen Abweichungen bei Nadel- und Mischwald. Bei Nadelwald wird um 13,78% (Tab.7) der Bezirksfläche zu wenig klassifiziert, wohingegen bei Mischwald 11,06% zu viel klassifiziert wird. Diese beiden Differenzen heben sich in der Gesamtstatistik beinahe auf.

Es ist anzunehmen, dass es einen Unterschied bei der Definition von Mischwald bei den Daten der Österreichischen Waldkarte und den Trainingsdaten dieser Arbeit gibt.

4.2.2 Vergleich der Klassifizierungsergebnisse mit den Invekos Daten

Ackerland stellt mit einem Flächenanteil von mehr als 22% eine bedeutende Landbedeckungsklasse Oberösterreichs dar.

	Invekos %	Acker Hart %	Acker Fuzzy %	Diff. H-F	Diff. I-H	Diff. I-F
Linz Stadt	9,80	19,33	22,09	-2,76	-9,53	-12,29
Steyr Stadt	27,01	33,79	35,22	-1,44	-6,78	-8,22
Wels Stadt	32,14	38,46	39,70	-1,25	-6,32	-7,57
Eferding	42,37	39,07	38,57	0,49	3,31	3,80
Freistadt	20,53	12,50	15,89	-3,40	8,03	4,63
Gmunden	5,42	9,62	10,78	-1,15	-4,21	-5,36
Grieskirchen	43,24	41,74	42,03	-0,29	1,50	1,20
Kirchdorf a.d Krems	10,05	12,12	13,43	-1,31	-2,06	-3,37
Linz Land	55,32	54,08	51,41	2,67	1,25	3,91
Perg	30,59	23,66	24,72	-1,06	6,93	5,87
Ried im Innkreis	39,98	40,87	41,15	-0,28	-0,89	-1,17
Rohrbach	16,66	12,01	16,63	-4,62	4,65	0,03
Schärding	32,41	30,81	32,61	-1,80	1,60	-0,20
Steyr Land	14,36	15,06	16,26	-1,20	-0,70	-1,90
Urfahr Umgebung	23,31	18,50	21,82	-3,33	4,81	1,49
Vöcklabruck	14,62	17,45	20,07	-2,61	-2,84	-5,45
Wels Land	54,77	54,12	52,10	2,02	0,65	2,66
Oberösterreich	22,96	22,18	23,72	-1,54	0,78	-0,76

Tab.9: Gegenüberstellung der Klassifizierungsergebnisse mit den Referenzdaten für die Landbedeckungsklasse Ackerland

Während das Gesamtergebnis mit einer maximalen Differenz von 0,78% eine sehr gute Übereinstimmung erzielt, weisen einige Bezirke doch deutliche Abweichungen von den Invekos-Daten auf. Bei der Auflistung der Ergebnisse wird, wie auch bei den Waldflächen, der Bezirk Braunau außer acht gelassen.

Bei Betrachtung von Tabelle 9 fallen zuerst die städtischen Bezirke auf, welchen im Vergleich zu viel Ackerfläche durch die automatische Klassifikation zugewiesen wird. Dies kann an einer Klassifikationsüberschneidung der beiden Klassen „Versiegelte Fläche“ und „Acker vegetationslos, anderer vegetationsloser Boden“ liegen, welche eine ähnliche spektrale Erscheinung haben können. Bezirke mit weniger dicht bebauten Siedlungsflächen erzielen eine bei weitem höhere Übereinstimmung. Einzig Freistadt und Perg weichen von dieser Übereinstimmung ab. Eine Erklärung für die Abweichungen dieser beiden Bezirke kann leider nicht gefunden werden.

4.2.3 Detaillierte Betrachtung einzelner wichtiger Bezirke

In diesem Kapitel sollen einzelne Bezirke stellvertretend für vergleichbare andere Bezirke genauer betrachtet und die Ergebnisse analysiert werden. Es wird versucht, hier repräsentative Bezirke auszuwählen. Es sollen je ein Bezirk mit viel Waldfläche, viel Ackerfläche und viel verbauter Fläche betrachtet und auf eventuelle Wechselbeziehungen zwischen einzelnen Klassen untersucht werden.

Gmunden

Der Bezirk Gmunden soll als Beispiel für Bezirke mit viel Waldbestand dienen. Dieser Bezirk weist eine Waldbedeckung von mehr als 60% auf, der Anteil an Ackerfläche beschränkt sich hingegen auf ca. 5-10% (anhängig vom betrachteten Datenmaterial). Die Bezirksgröße von Gmunden beträgt 14.329 ha und er ist damit einer der größten Bezirke Oberösterreichs. Davon sind ca. 92.000 ha mit Wald bedeckt. In Abbildung 23 werden die einzelnen Waldklassen und ihr Anteil an dieser beträchtlichen Fläche in m² dargestellt. Dabei werden die Flächen aus der Österreichischen Waldkarte denen aus der harten automatischen Klassifikation gegenübergestellt.

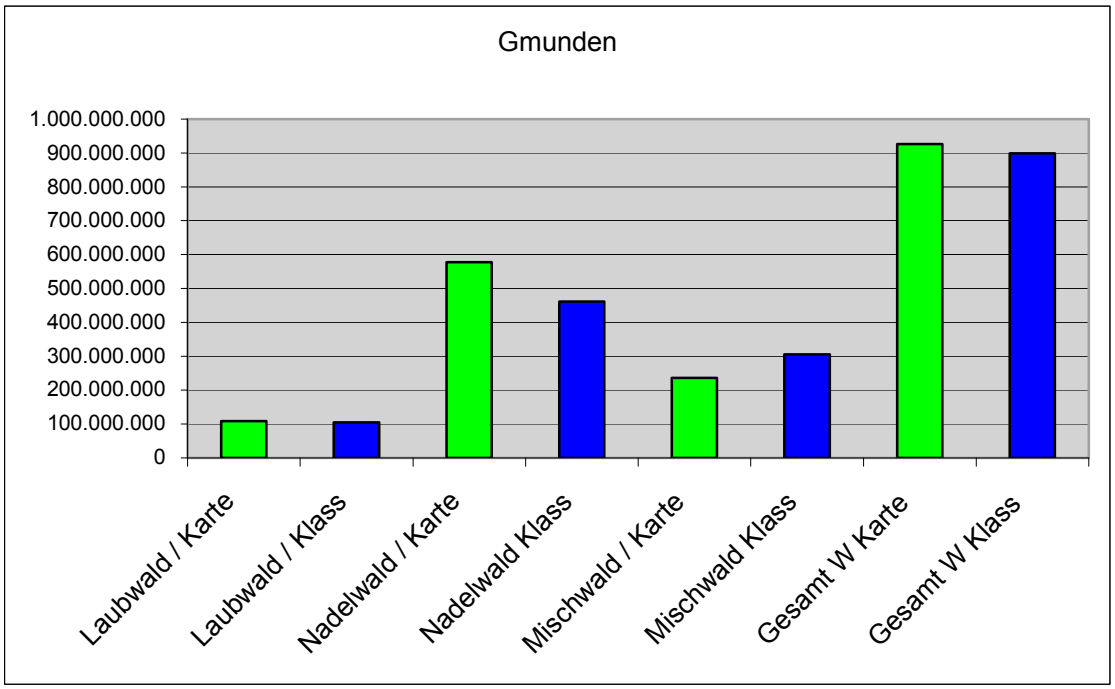


Abb. 23: Gegenüberstellung der Flächen der Waldlandbedeckungsklassen im Bezirk Gmunden

Die Landbedeckungsklasse Laubwald ist von allen Waldklassen am wenigsten vertreten. Sie macht nur ca. 7,5% der Bezirksfläche aus. Die Ergebnisse der

automatischen Klassifikation decken sich sehr gut mit den Referenzdaten aus der Österreichischen Waldkarte. Es kommt kaum zu Überschneidungen in der Klassifikation mit anderen Waldklassen.

Anders stellt sich das bei den Klassen Nadelwald und Mischwald dar. Nadelwald stellt mit zwischen 31,5% und 40,3% der Bezirksfläche, je nach betrachtetem Datenmaterial, den größten Anteil der Waldfläche. Dieser Unterschied der Flächenanteile ergibt sich aus einer in Abbildung 23 deutlich ersichtlichen Wechselbeziehung mit der Klasse Mischwald. Es wird bei der automatischen Klassifikation weniger Fläche der Klasse Nadelwald zugewiesen als in der Referenzdatenbank der Österreichischen Waldkarte. Im Gegenzug wird der Klasse Mischwald mehr Fläche zugeordnet. Dies kann sowohl auf eine Überschneidung der spektralen Merkmale der beiden Waldklassen als auch auf eine verschiedene Definition der beiden Waldklassen in den Trainingsdaten zur automatischen Klassifikation und der Österreichischen Waldkarte zurück zu führen sein. Die Gesamtfläche der Waldbedeckung deckt sich aus beiden Datensätzen jedoch wieder auffallend gut.

Gmunden	Ref. %	Hart %	Fuzzy %	Diff. Hart-Fuzzy	Diff. Ref-Hart	Diff. Ref-Fuzzy
Acker	5,42	9,62	10,78	-1,15	-4,21	-5,36
Wald Ges.	64,66	62,72	61,24	1,48	1,94	3,42
Laubwald	7,58	7,33	8,55	-1,21	0,24	-0,97
Nadelwald	40,33	32,17	31,54	0,63	8,16	8,80
Mischwald	16,48	21,28	19,37	1,91	-4,80	-2,89

Tab.10: Gegenüberstellung der Ergebnisse der wichtigsten Landbedeckungsklassen aus den Automatischen Klassifikationen mit den Referenzdaten für den Bezirk Gmunden*

In Tabelle 10 sind alle Waldklassen sowie die Ackerfläche des Bezirks als Prozentwerte der Bezirksfläche dargestellt. Auch die Differenzen zwischen den einzelnen Datensätzen werden in dieser Tabelle ersichtlich. „Ref %“ stellt den Wert aus der Österreichischen Waldkarte oder aus der Invekos Datenbank dar. „Hart %“ und „Fuzzy %“ geben jeweils das Ergebnis aus der harten Klassifikation oder aus der Fuzzy-Klassifikation wieder.

*Beim Lesen der Tabellen ist zu bedenken, dass der Wert für „Wald gesamt“ bei der Kategorie der Referenzdaten nicht bloß die Summe aus Laubwald, Nadelwald und Mischwald darstellt. Darin sind auch Schlagflächen aus der Österreichischen Waldkarte enthalten. In den Ergebnissen aus den automatischen Klassifikation ist auch die Kategorie „Wald Nordhang Schatten“ enthalten.

Bei Betrachtung der Ackerflächen in Tabelle 10 fällt auf, dass die beiden Ergebnisse der automatischen Klassifikationen sehr ähnlich sind. Beide Ergebnisse weichen jedoch deutlich von der Ackerfläche aus der Invekos Datenbank ab. Dies könnte von einer Überschneidung der spektralen Merkmale der Klassen „Ackerland vegetationslos und anderer vegetationsloser Boden“ und „Versiegelte Fläche“ kommen. Das heißt, dass „versiegelte Fläche“ irrtümlich als „vegetationsloser Boden“ klassifiziert wird. Eine andere Ursache ist, da es in diesem Bezirk viele Berge mit sehr spärlicher Vegetation gibt, dass dieser nicht agrarisch genützte Vegetationslose Boden einen großen Anteil an der Landbedeckungsklasse „Acker vegetationslos, anderer vegetationsloser Boden“ ausmacht.

Linz Stadt

Die Stadt Linz soll als Beispiel eines Bezirkes dienen, welcher sehr stark verbaut ist und der kaum Wald oder Ackergebiete aufweist. Von den 9.557 ha, die den Bezirk Linz Stadt ausmachen, sind je nach betrachteter Klassifikation 36,3% oder 33,7% versiegelte Fläche. Das entspricht einer Fläche von ca. 3.400 ha.

Bezirke:	Bezirksgröße [ha]	Versiegelte Flächen	Versiegelte Flächen	% Hart	% Fuzzy
		Harte Klass. [ha]	Fuzzy Klass. [ha]		
Linz Stadt	9.557	3.474	3.222	36,34	33,71

Tab.11: Gegenüberstellung der Landbedeckungsklasse „Versiegelte Fläche“ aus den automatischen Klassifikationen im Bezirk Linz

Diesem großen Anteil an versiegelter Fläche stehen lediglich ca. 20% Waldbedeckung gegenüber. Eine andere neben dem verbauten Gebiet sehr bedeutende Landbedeckungsklasse in Linz ist Grünland. Sie macht je nach Klassifizierungsart 18,2% oder 17,1% der Bezirksfläche aus.

	GL-Invekos %	GL Hart %	GL Fuzzy %
Linz Stadt	4,73	18,15	17,13

Tab.12: Gegenüberstellung der Ergebnisse aus den automatischen Klassifikationen mit den Referenzdaten für die Landbedeckungsklasse „Grünland“ im Bezirk Linz

Die große Differenz zu den Daten für Grünland aus der Invekos Datenbank ist darauf zurück zu führen, dass in dieser Datenbank nur agrarisch genütztes Grünland erfasst wird. Die automatische Klassifikation erkennt jedoch Parks, größere Gärten und sonstige Grünanlagen ebenfalls als Grünland. Deswegen wird mittels automatischer

Klassifikation viel mehr Grünland klassifiziert als es in der Invekos-Datenbank eingetragen ist.

Eine ähnliche Problematik ergibt sich auch bei den ohnehin geringen Waldbeständen im Bezirksgebiet von Linz. In der Österreichischen Waldkarte sind in Linz 17,4% der Bezirksfläche als Waldfläche identifiziert. Da aber bei der automatischen Klassifikation viele Park- und Grünflächen als Wald klassifiziert werden können, werden hier ~21% als Waldfläche ausgewiesen.

Linz Stadt	Ref. %	Hart %	Fuzzy %	Diff. H-F	Diff. R-H	Diff. R-F
Acker	9,80	19,33	22,09	-2,76	-9,53	-12,29
Wald Ges.	17,44	21,38	21,79	-0,41	-3,94	-4,35
Laubwald	9,42	6,65	6,58	0,07	2,77	2,84
Nadelwald	6,35	5,05	6,62	-1,57	1,29	-0,28
Mischwald	1,59	9,68	8,59	1,09	-8,09	-6,99

Tab.13: Gegenüberstellung der Ergebnisse der Landbedeckungsklassen Wald und Ackerland aus den Automatischen Klassifikationen mit den Referenzdaten für den Bezirk Linz

In Linz gibt es ein besonders starkes Ungleichgewicht zwischen den Referenzdaten der einzelnen Klassen und den Ergebnissen aus der automatischen Klassifikation. Die Ergebnisse aus den beiden automatischen Klassifikationsarten decken sich hingegen sehr gut. Es wird viel mehr Mischwald klassifiziert als in der Waldkarte eingetragen. Das kann neben der bereits beim Bezirk Gmunden besprochenen Möglichkeit einer unterschiedlichen Definition von Mischwald in den Trainingsdaten und der Österreichischen Waldkarte auch darauf zurückzuführen sein, dass Parks und Grünflächen sowohl mit Nadel- als auch mit Laubbäumen bepflanzt sind.

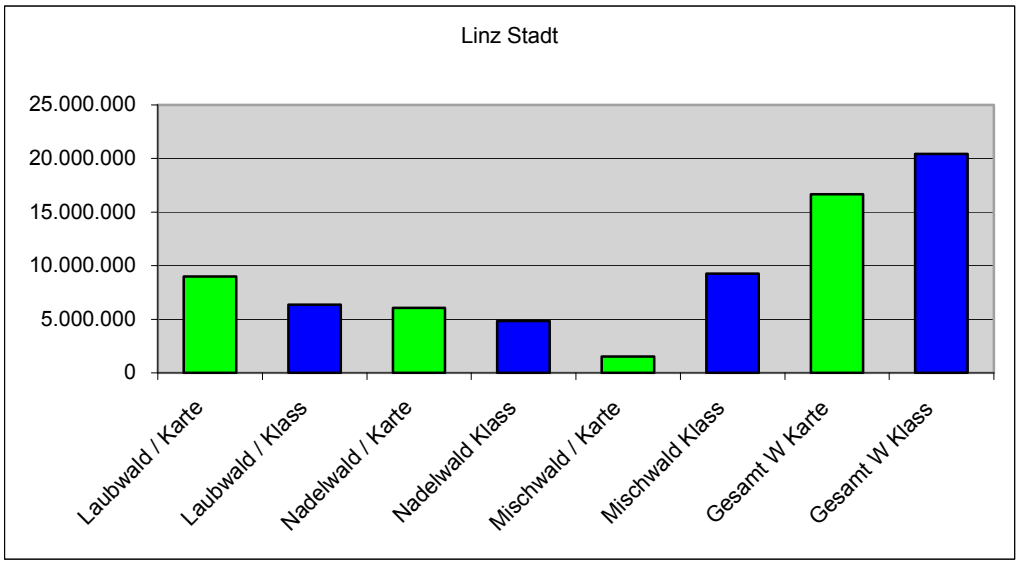


Abb. 24: Gegenüberstellung der Flächen der Waldlandbedeckungsklassen im Bezirk Linz

Bei Ackerland wird mittels der automatischen Klassifikationen bedeutend mehr Fläche zugewiesen als in der Invekos Datenbank eingetragen ist. Das liegt zum einen an der spektralen Überschneidung der Landbedeckungsklassen „Acker vegetationslos“ und „Versiegelte Fläche“, zum anderen daran, dass es in Linz zum Beispiel auf dem Gelände des Stahlwerks VOEST große brache, vegetationslose Schlackefelder und ähnliche Lagerplätze gibt. Diese Flächen werden ebenfalls der Kategorie „Acker vegetationslos, anderer vegetationsloser Boden“ zugewiesen und in weiterer Folge hier statistisch wie Ackerflächen behandelt.

Eine Kontrolle der automatischen Klassifizierung für Stadtgebiete ist aus den oben genannten Gründen mittels der in dieser Arbeit verwendeten Referenzdaten aus der Österreichischen Waldkarte und der Invekos Datenbank nicht durchführbar. Man müsste für Stadtgebiete andere Referenzdaten suchen oder die Trainingsdaten speziell für Stadtgebiete aus legen, wenn man den Schwerpunkt der automatischen Klassifikation auf verbaute Gebiete legen will.

Wels Land

Der Bezirk Wels Land soll hier als Beispiel für einen Bezirk mit sehr viel Ackerfläche dienen. Der Bezirk mit einer Fläche von 4.576 ha ist zu mehr als 50% mit Ackerland bedeckt. Das entspricht einer Fläche von über 2.300 ha und liegt deutlich über dem oberösterreichischen Durchschnittswert von ca. 22%.

	Invekos %	Acker Hart %	Acker Fuzzy %
Wels Land	54,77	54,12	52,10
Oberösterreich	22,96	22,18	23,72

Tab.14: Gegenüberstellung der Ergebnisse aus den automatischen Klassifikationen mit den Referenzdaten für die Landbedeckungsklasse „Ackerland“ im Bezirk Wels Land im Vergleich mit OÖ.

Die Ackerflächen der verschiedenen Datensätze decken sich auffallend gut, einzig die Daten aus der Fuzzy-Klassifikation weichen geringfügig ab.

Dieser sehr großen Bedeckung durch Ackerflächen steht eine sehr geringe Waldfläche von insgesamt nur ca.15% gegenüber.

Wels Land	Ref. %	Hart %	Fuzzy %	Diff. H-F	Diff. R-H	Diff. R-F
Acker	54,77	54,12	52,10	2,02	0,65	2,66
Wald Ges.	16,35	15,35	15,97	-0,61	0,99	0,38
Laubwald	4,33	3,08	3,41	-0,33	1,26	0,93
Nadelwald	7,81	7,69	8,17	-0,48	0,11	-0,37
Mischwald	1,66	4,58	4,39	0,19	-2,92	-2,73

Tab.15: Gegenüberstellung der Ergebnisse der Landbedeckungsklassen Wald und Ackerland aus den automatischen Klassifikationen mit den Referenzdaten für den Bezirk Wels Land

Auch in diesem Bezirk kommt es bei der automatischen Klassifikation wieder zu einem Überschuss von Mischwald. Der gesamte Waldbestand deckt sich jedoch in zufriedenstellendem Maß. Die Abweichungen der Waldflächen in den einzelnen Kategorien ergeben sich in diesem Bezirk auch durch die große Menge an „Schlagflächen“ aus der Waldkarte, welche keiner spezifischen Waldklasse zugeordnet, sondern nur der gesamten Waldfläche zugerechnet werden. Bei der automatischen Klassifizierung werden diese Schlagflächen einer anderen Wald- oder Nicht-Waldklasse zugeordnet.

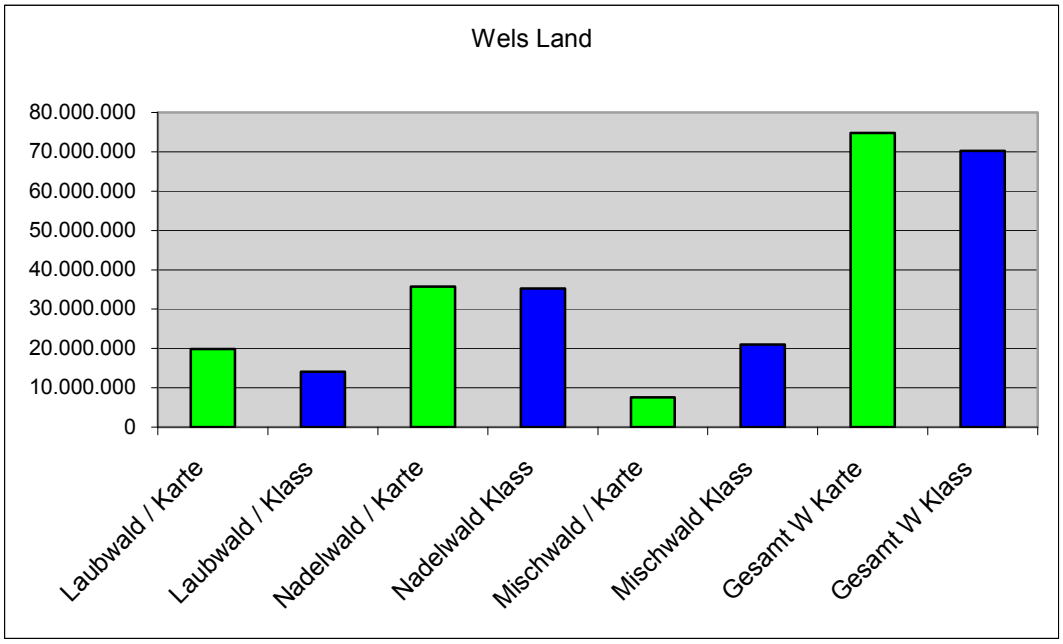


Abb. 25: Gegenüberstellung der Flächen der Waldlandbedeckungsklassen im Bezirk Wels Land

Braunau

Die bezirksweise Betrachtung soll mit dem Ausnahmefall Braunau abgeschlossen werden. Hier fehlen, da die Landsat-Szene nicht den ganzen Bezirk abdeckt, bei der automatischen Klassifikation große Teile der Flächen. Die prozentuellen Flächenbedeckungen aus den Invekos-Daten oder den Daten aus der Waldkarte werden mit der gesamten Bezirksgröße berechnet. Die aus der Klassifikation stammenden Flächen werden nur auf die zur Klassifikation zur Verfügung stehenden Fläche bezogen.

Braunau	Ref. %	Hart %	Fuzzy %	Diff. H-F	Diff. R-H	Diff. R-F
Acker	25,74	23,95	26,37	-2,42	1,80	-0,62
Wald Ges.	36,15	39,38	39,41	-0,03	-3,22	-3,25
Laubwald	4,42	2,47	2,87	-0,40	1,95	1,55
Nadelwald	27,15	28,18	28,15	0,03	-1,03	-1,00
Mischwald	3,25	8,73	8,39	0,34	-5,48	-5,15

Tab.16: Gegenüberstellung der Ergebnisse der Landbedeckungsklassen Wald und Ackerland aus den automatischen Klassifikationen mit den Referenzdaten für den Bezirk Braunau

Wie man in Tabelle 16 erkennen kann, ist Braunau ein Bezirk mit einer ausgewogenen Landbedeckungsaufteilung. Weder Ackerflächen noch Waldflächen dominieren die Landschaft. Es ist jedoch trotzdem deutlich die ländliche Form der Landbedeckung zu erkennen. Laut der Fuzzy- Klassifizierung bedecken Wald- und Ackerflächen zusammen 65,8% des Bezirksgebietes. Grünland bedeckt nach derselben Klassifikation lediglich 12,6%, versiegelte Flächen sogar nur 2,1%.

Das bei den anderen betrachteten Bezirken verwendete Balkendiagramm zur Darstellung der Waldflächen entfällt bei Braunau, da man die Flächen nicht direkt miteinander vergleichen kann.

5. Landbedeckungskarte Oberösterreichs

Die Landbedeckungskarte Oberösterreichs soll die Ergebnisse der in den vorangegangenen Kapiteln beschriebenen Berechnungen der Landbedeckung auf einen Blick darstellen.

Die Landbedeckungskarte ist aus den Daten der harten Klassifikation erstellt und berücksichtigt damit keine Anteile aus eventuellen weiteren Landbedeckungsklassen für einzelne Pixel. Aus der Fuzzy-Klassifikation ist jedoch bekannt, dass manche Pixel neben einer vorherrschenden oder dominierenden Klasse auch noch Anteile anderer Klassen enthalten können. Auch die genauen Werte dieser Anteile sind aus dieser Klassifikation bekannt. Diese sind jedoch nur tabellarisch erfasst und werden in der Karte nicht grafisch dargestellt.

Bei dieser Karte wird den einzelnen Landbedeckungsklassen jeweils eine Farbe zugeordnet welche diese Klasse auf der Karte optisch darstellt. Versiegelte Flächen werden zum Beispiel in Rot dargestellt. Dadurch lassen sich Siedlungen und ähnliche Strukturen auf der Karte leicht erkennen.

Wollte man Daten aus der Fuzzy-Klassifikation in die Karte mit einbeziehen, müsste man eine andere Art der Darstellung wählen. Einzelne Pixel können einfach in Form einer Grauskala oder Farbskala dargestellt werden, welche in ihrer Farbintensität dem Wert der Zugehörigkeit zu einer Landbedeckungsklasse entspricht. Jede Klasse müsste man in einer separaten Karte darstellen. Diese potentiell sehr große Anzahl an Landbedeckungskarten würde es dem Benutzer nicht mehr erlauben, die Landbedeckung eines Gebietes als Ganzes zu erfassen. (Wilson; Fotheringham 2008. S.265)

Beim Blick auf die Karte fällt sofort der beträchtliche Anteil an Waldflächen auf. Immerhin werden ca. 42% der Fläche von Oberösterreich von Wald bedeckt. Der Hauptteil des Waldbestandes konzentriert sich auf den südlichen Teil Oberösterreichs. Das Maximum an Waldbedeckung, mit einem Anteil von mehr als 60% der Bezirksfläche, befindet sich im Bezirk Gmunden. Ackerflächen machen

einen ebenfalls beträchtlichen Anteil an der Fläche Oberösterreichs aus, welcher bei ca. 23% liegt. Der Hauptteil der Ackerfläch en konzentriert sich auf den sehr flachen Teil etwa in der Mit te Oberösterreichs. Ebenfalls leicht zu erkennen sind die Ballungsräume Oberösterreichs. Am deutlichsten zeichnet sich Linz als größte Stadt mit vielen Industrieanlagen ab. Der Großteil Oberöst erreichs ist jedoch nicht dicht verbaut, sondern wird von Agrar-, Wiesen-, und Waldflächen dominiert.

Die Landbedeckungskarte Oberösterreichs ist beigelegt.

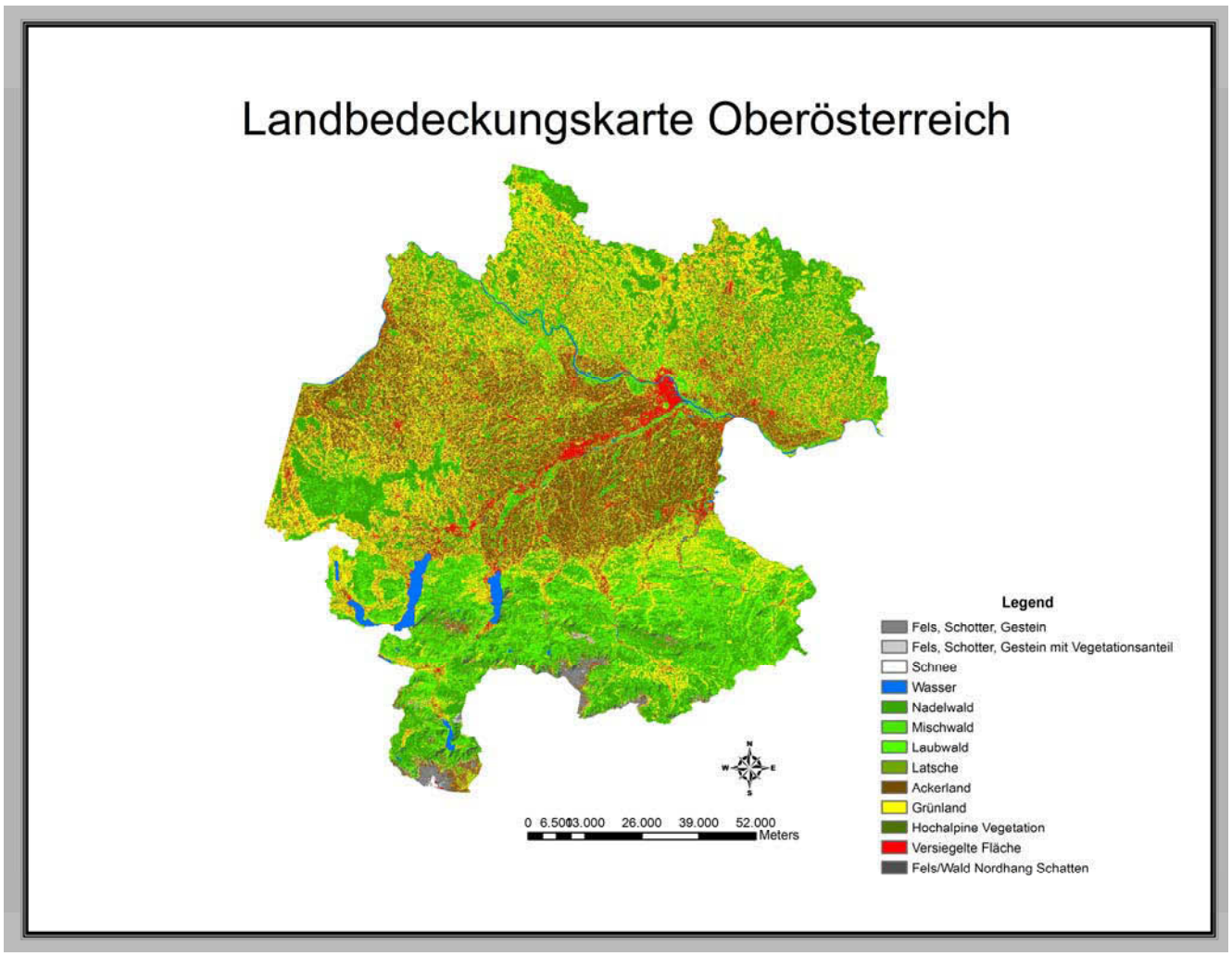


Abb. 26: Landbedeckungskarte Oberösterreich

6. Diskussion und Schlussfolgerung

kNN-Klassifikation

Wie schon in den meisten Arbeiten zum Thema Landbedeckungsklassifizierung aus Satellitenbildern mittels kNN-Klassifikation festgestellt, zeigt auch diese Arbeit, dass diese Methode für dieses Anwendungsgebiet sehr gut geeignet ist. Man sollte jedoch nicht zu viel Zeit für die Gewichtung der einzelnen Kanäle in der Berechnung verwenden, da der Genauigkeitsgewinn zwar sichtbar ist, aber auch vielen Wechselwirkungen unterliegt. Die Gewichtung der Kanäle sollte auf die konkrete, exakt definierte Aufgabenstellung abgestimmt sein: im Fall dieser Arbeit wird versucht, die großen Klassen der Wald- und Agrarflächen deutlich darzustellen, weil dafür passende Referenzdaten vorhanden sind, um die Ergebnisse zu überprüfen. Alle Versuche, die Genauigkeit dieser Klassen zu erhöhen, führen jedoch unweigerlich dazu, dass die Genauigkeit von kleineren Klassen geringer wird. Ein Trennen von Landbedeckungsklassen, welche sich spektral ähnlich sind, ist mittels der Kanalgewichtung nicht möglich.

Fuzzy-Klassifikation

Die Fuzzy-Klassifikation ermöglicht es, einzelnen Pixeln mehrere Landbedeckungsklassen zuzuweisen. Dadurch lässt sich eine Landschaft im Bezug auf ihre Landbedeckung, zumindest in tabellarischer Form, detaillierter darstellen. Ein tatsächlicher Genauigkeitsgewinn für ganze Regionen ist jedoch nicht sichtbar, da die harte Klassifikation für diese Anwendung bereits sehr gute Ergebnisse liefert. Der zusätzliche sehr große Arbeitsaufwand rechtfertigt nicht die kaum unterschiedlichen Ergebnisse, wenn man eine herkömmliche Landbedeckungskarte erstellen will. Man erhält zwar für die einzelnen Pixel mehr Detailinformation, welche sich jedoch in einer herkömmlichen Landbedeckungskarte nicht darstellen lassen.

Durch die Fuzzy-Klassifikation ließen sich eventuell Landbedeckungsklassen, welche im Gelände fraktal verteilt in kleinen Flächen vorkommen, aufspüren und darstellen. Diese nur in Pixelanteilen auftretenden Klassen werden bei der harten Klassifikation weitgehend unterdrückt.

Interessant wäre es, die Ergebnisse aus der Fuzzy-Klassifikation in anderer Form grafisch umzusetzen. Wie im Kapitel „Landbedeckungskarte Oberösterreichs“ beschrieben, könnte die Darstellung in Form einer Grauskalen-Karte für jede einzelne Landbedeckungsklasse dafür eine gute Möglichkeit sein. Man könnte, wie in dieser Arbeit geschehen, eine herkömmliche Karte der Landbedeckung eines Gebietes erstellen, um dem Benutzer eine übersichtliche Darstellung zu garantieren. Zusätzlich dazu würden die weiteren Karten der einzelnen Landbedeckungsklassen detailliert die Ergebnisse der Fuzzy Klassifikation darstellen. Eine weitere Darstellungsmöglichkeit wäre eine interaktive digitale Landbedeckungskarte, bei der man im Gesamtüberblick das Ergebnis der harten Klassifikation sieht, sich aber durch eine Detailansicht einzelner Pixel die Ergebnisse der Fuzzy-Klassifikation anzeigen lassen kann.

Verbesserung der Trainingsdaten

Die Qualität der Trainingsdaten ist essentiell für den Erfolg einer Landbedeckungsklassifizierung. Die Entfernung der Outlier aus der Crossvalidation zur Eliminierung von schlechten Daten aus dem Trainingsdatensatz ist eine gute Möglichkeit, die Qualität zu erhöhen. Es wird in dieser Arbeit jedoch gezeigt, dass es keinen Sinn macht, die bei der Crossvalidation besonders positiv auffallenden Trainingsdaten in besonderem Maß für die Klassifizierung heranzuziehen: Diese Daten werden nicht öfters für die eigentliche Interpretation der Landbedeckung herangezogen als andere. Es würde daher eher der Effekt eintreten, dass man die spektrale Vielfalt der Trainingsdaten vermindert, was in weiterer Folge zu schlechteren Ergebnissen bei der Berechnung der Landbedeckung führen würde.

Literaturverzeichnis / Quellenangabe

Bauer, Thomas, Skriptum Fernerkundung und Bilddatenverarbeitung (VU) Praktischer Teil LV 857.304 University of Natural Ressources and Applied Life Sciences (BOKU) Wien

Geoinformatik (29.August 2011): Geoinformatik Lexikon: http://www.geoinformatik.uni-rostock.de/einzel.asp?ID=529

Grillmayer, R. (2004): Vorwort zu Geodaten zur Landbedeckung in Österreich (R. Grillmayer / W. Schneider (Hrsg.)). Shaker Verlag Aachen

Koukal, Tatjana (2005): User Guide to the kNN-Programm. University of Natural Ressources and Applied Life Sciences (BOKU) Wien

Lillesand, Thomas M.; Kiefer, Ralph W. (2000): Remote Image Sensing and Image Interpretation (4th Edition) John Wiley & Sons Inc. New York

Mattiuzzi, Matteo (2007): Landbedeckungsklassifizierung aus Satellitenbildern mit der kNN-Methode am Beispiel Tirol. Diplomarbeit, University of Natural Ressources and Applied Life Sciences (BOKU) Wien

Mitchell, Tom (1997): Machine Learning. McGraw-Hill New York

Oberhauser, Otto (2005): Automatisches Klassifizieren. Peter Lang GmbH Europäischer Verlag der Wissenschaften. Frankfurt am Main

Popovici Paula (2008): Landbedeckungsklassifizierung aus Satellitenbildern am Beispiel Niederösterreich. Diplomarbeit, University of Natural Ressources and Applied Life Sciences (BOKU) Wien

Sachs, Lothar und Hedderich, Jürgen (2009): Angewandte Statistik (13.Auflage) Springer Verlag Berlin Heidelberg

Schneider, Werner (2002): Fotogrammetrie, Fernerkundung und Geoinformationswesen (Institut für Vermessung, Fernerkundung und Landinformation Universität für Bodenkultur, Wien)

Tomppo, Erkki (2004): Using coarse scale forest variables as ancillary information and weighting of variables in k-NN estimation: a genetic algorithm approach. Remote Sensing of Environment 92: 1-20.

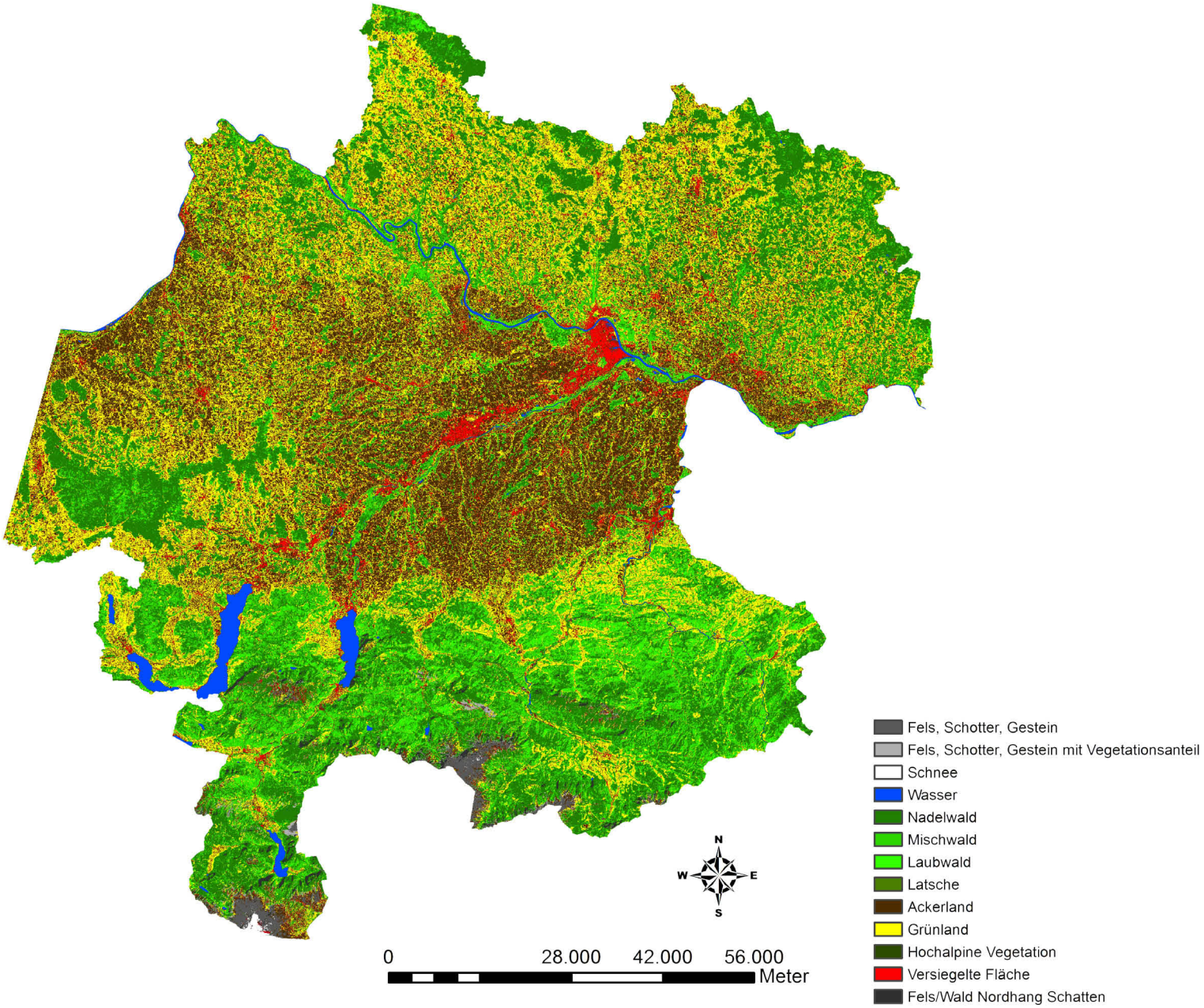
Wilson, John P. ; Fotheringham, Stewart (2008): The Handbook of Geographic Information Science. Blackwell Publishing Malden MA

Wikipedia (29.August 2011): Online Lexikon: http://en.wikipedia.org/wiki/Landsat_5

Abbildungsverzeichnis

Abb. 1: Landsat 5 Szene mit allen Passpunkten S.9
Abb. 2: Georeferenzierte Landsat 5 Szene mit ÖK 50 S.10
Abb. 3: Georeferenzierte Landsat 5 Szene mit ÖK 50 S.10
Abb. 4: Satellitenbild mit allen interpretierten Rasterpunkten S.14
Abb.5: Verteilung der Trainingsdaten - Anzahl der Trainingsdaten pro Landbedeckungsklasse S.17
Abb. 6: Waldkarte Oberösterreich S.18
Abb. 7: fertige OÖ Maske über dem Landsat Bild S.20
Abb. 8:Producers Accuracy abhängig von k S.28
Abb. 9:Users Accuracy abhängig von k S.28
Abb. 10: Verlauf von Kappa bei k=1 und k=7 nach der Gewichtung von HS S.31
Abb. 11: Verlauf von Kappa bei k=1 und k=7 nach der Gewichtung von Band 3 S.32
Abb. 12: Verlauf von Kappa bei K=1 und k=7 nach der Gewichtung der Bandkombination B4/B3S.33
Abb. 13: Einfluss von k auf die Genauigkeit bei optimaler Gewichtung der einzelnen Bänder S.34
Abb. 14: Genauigkeitsverlauf von Kappa und Overall Acc. bei k=1 S.35
Abb. 15: Genauigkeitsverlauf von Kappa und Overall Acc. bei k=7 S.35
Abb. 16: Genauigkeitsverlust beim Weglassen der einzelnen Bänder bei k=1 S.36
Abb. 17: Genauigkeitsverlust beim Weglassen der einzelnen Bänder bei k=7 S.37
Abb. 18: vereinfachte Landbedeckung Oberösterreichs S.38
Abb. 19: Fuzzy klassifizierte Klasse 70: Versiegelte Fläche S.40
Abb. 20: Fuzzy klassifizierte Klasse 31: Nadelwald dicht S.40
Abb. 21: Gegenüberstellung der Waldflächen aller Bezirke aus den automatischen Klassifikationen mit den Daten der Österreichischen Waldkarte S.43
Abb. 22: Karte der bezirksweisen Flächenstatistik S.44
Abb. 23: Gegenüberstellung der Flächen der Waldlandbedeckungsklassen im Bezirk Gmunden S.50
Abb. 24: Gegenüberstellung der Flächen der Waldlandbedeckungsklassen im Bezirk Linz S.53
Abb. 25: Gegenüberstellung der Flächen der Waldlandbedeckungsklassen im Bezirk Wels Land S.55
Abb. 26: Landbedeckungskarte Oberösterreich S.58

Tabellenverzeichnis

Tab.1: Landbedeckungsklassen S.13
Tab.2: Punkteverteilung der Trainingsdaten (Anzahl und Prozentanteil) in den Klassen S.16
Tab.3: Wichtige Landbedeckungsklassen und ihre Genauigkeiten S.42
Tab.4: Gegenüberstellung der Ergebnisse aus den automatischen Klassifikationen mit den Referenzdaten für die Landbedeckungsklassen Wald und Ackerland S.43
Tab.5: Gegenüberstellung der Klassifizierungsergebnisse mit den Referenzdaten für den gesamten Waldbestand Oberösterreichs S.45
Tab.6: Gegenüberstellung der Klassifizierungsergebnisse mit den Referenzdaten für die Landbedeckungsklasse Laubwald S.46
Tab.7: Gegenüberstellung der Klassifizierungsergebnisse mit den Referenzdaten für die Landbedeckungsklasse Nadelwald S.47
Tab.8: Gegenüberstellung der Klassifizierungsergebnisse mit den Referenzdaten für die Landbedeckungsklasse Mischwald S.48
Tab.9: Gegenüberstellung der Klassifizierungsergebnisse mit den Referenzdaten für die Landbedeckungsklasse Ackerland S.49
Tab.10: Gegenüberstellung der Ergebnisse der wichtigsten Landbedeckungsklassen aus den Automatischen Klassifikationen mit den Referenzdaten für den Bezirk Gmunden S.51
Tab.11: Gegenüberstellung der Landbedeckungsklasse „Versiegelte Fläche“ aus den automatischen Klassifikationen im Bezirk Linz S.52
Tab.12: Gegenüberstellung der Ergebnisse aus den automatischen Klassifikationen mit den Referenzdaten für die Landbedeckungsklasse „Grünland“ im Bezirk Linz S.52
Tab.13: Gegenüberstellung der Ergebnisse der Landbedeckungsklassen Wald und Ackerland aus den automatischen Klassifikationen mit den Referenzdaten für den Bezirk Linz S.53
Tab.14: Gegenüberstellung der Ergebnisse aus den automatischen Klassifikationen mit den Referenzdaten für die Landbedeckungsklasse „Ackerland“ im Bezirk Wels Land im Vergleich mit OÖ S.54
Tab.15: Gegenüberstellung der Ergebnisse der Landbedeckungsklassen Wald und Ackerland aus den automatischen Klassifikationen mit den Referenzdaten für den Bezirk Wels Land S.55
Tab.16: Gegenüberstellung der Ergebnisse der Landbedeckungsklassen Wald und Ackerland aus den Automatischen Klassifikationen mit den Referenzdaten für den Bezirk Braunau S.56

Landbedeckungskarte Oberösterreich