

AKTUALISIERUNG VON CORINE-  
LANDNUTZUNGSKARTEN MIT MULTISAISONALEN  
SATELLITENBILDERN

Diplomarbeit

an der Universität für Bodenkultur

Department für Raum, Landschaft und Infrastruktur (RaLI)

Institut für Vermessung, Fernerkundung und Landinformation (IVFL)

eingereicht von

Christoph Graf Adelman

betreut von

Univ. Prof. Dipl.-Ing. Dr. techn. Werner Schneider

unter Mitbetreuung von

Dipl.-Ing. Dr. nat. techn. Tatjana Koukal und

Mag. Franz Suppan

Hasperde, November 2010

## **Danksagung**

Ohne die Hilfe vieler Menschen wäre die Fertigstellung dieser Arbeit nicht möglich gewesen. Also möchte ich mich hiermit sehr bei all jenen bedanken, die mich in irgendeiner Form während des Studiums unterstützt haben.

Allen voran sind meine Eltern zu nennen, die mir das Studium überhaupt ermöglicht haben.

Bei der Umsetzung der Arbeit bin ich insbesondere meinen Betreuern Univ. Prof. Dipl.-Ing. Dr. techn. Werner Schneider und Dipl.-Ing. Dr. nat. techn. Tatjana Koukal zu Dank verpflichtet. Ihre Arbeit und Ideen sind Grundlage dieser Arbeit. Ihr Rat und ihre Beschäftigung mit den Einzelheiten der Thematik bilden den Rahmen des Erreichten. Auch die Unterstützung durch Dipl.-Ing. Paula Popovici und Mag. Franz Suppan war äußerst hilfreich. Insgesamt ist das Arbeiten am Institut für Vermessung, Fernerkundung und Landinformation wegen seiner zahlreichen hilfsbereiten Mitarbeiter und wegen der freundlichen Atmosphäre besonders angenehm.

Großer Dank gebührt weiter meinen Freunden, die mich während des Studiums begleitet und teils bei der Umsetzung der Arbeit unterstützt haben.

## **Kurzfassung**

Diese Arbeit beschreibt die Erstellung von Landbedeckungskarten anhand von Satellitenbilddaten. Für die Produktion der Karten wird die Klassifizierung durch die kNN-Methode verwendet. Die Klassifizierungen geschehen mit Referenzdaten zur Landbedeckung aus dem CORINE-Datensatz. Es werden auch Klassifizierungen untersucht, die sich auf Satellitenbilddaten von zwei Zeitpunkten beziehen. Mit der Verwendung des CORINE-Datensatzes ergibt sich das Problem von sogenannten Ausreißern, die bei mangelnder Aktualität und wegen der Generalisierung der CORINE-Daten vorkommen. Der Ausreißerproblematik konnte mit der Ausreißeridentifizierung und -eliminierung im Zuge von Klassifizierungen begegnet werden. Als Hauptergebnisse liegen Landbedeckungskarten sowie Genauigkeitswerte der Karten durch Verwendung von Kontrolldaten aus visueller Interpretation vor. Eine Aktualisierung und räumliche Verfeinerung der Karten gegenüber CORINE konnte erreicht werden. Die Karten werden insgesamt verbessert, wenn Ausreißer automatisch entfernt und Satellitenbilddaten von zwei Zeitpunkten (bzw. Jahreszeiten) verwendet werden.

Schlüsselwörter: Landbedeckungskarten, Satellitenbilder, CORINE, kNN-Klassifizierung, Ausreißer

## **Abstract**

This paper describes the production of land cover maps using satellite image data. The maps are generated by classifications with the kNN-method. These classifications use reference data on land cover from CORINE. Furthermore classifications of satellite image data from two dates are investigated. The use of CORINE-data implies the problem of outliers in the reference data, which may be outdated and which are geographically generalized. These problems were encountered by the use of outlier identification and elimination in the course of classifications. The main results of this work are land cover maps as well as accuracy data obtained by using reference data from visual interpretation. Updating and spatial refinements of maps in comparison to CORINE were achieved. The maps were altogether improved when outliers were eliminated automatically and reference data from satellite images of two dates (seasons) were used.

Keywords: land-cover maps, satellite images, CORINE, kNN-classification, outliers

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>7</b>
1.1	Ausgangssituation	7
1.2	Ziele der Arbeit	8
1.3	Vorgehensweise	8
<b>2</b>	<b>Datengrundlagen</b>	<b>9</b>
2.1	Eigenschaften des Untersuchungsgebietes	9
2.2	CORINE Land Cover	10
2.3	ASTER	12
2.4	Orthofotos	13
<b>3</b>	<b>Methoden</b>	<b>13</b>
3.1	Fernerkundungsdatensatz und Referenzdatensatz	13
3.2	Erzeugung der Datensätze	14
3.2.1	Erstellung des Rasters	14
3.2.2	Erstellung des bisaisonalen Fernerkundungsdatensatzes	14
3.2.3	„Homogenitätskriterium“ für die Rasterpunkte	15
3.2.4	„Auffüllpunkte“	15
3.2.5	Maskenerstellung	15
3.2.6	Klassenreduktion	16
3.3	kNN-Klassifizierung	18
3.4	Ausreißer	20
3.4.1	Ausreißeridentifizierung und -eliminierung	20
3.4.1.1	„Protokollmethode“	20
3.4.1.2	„Einschrittmethode“	21
3.4.2	Suche nach künstlichen Ausreißern	21
3.5	Erstellung des visuellen Kontrolldatensatzes	22
3.5.1	Erzeugung eines Referenzdatensatzes aus den visuell interpretierten Daten	25
3.6	Software	26
3.6.1	ArcMap	26
3.6.2	Erdas Imagine	26
3.6.3	kNN-Software	27
3.7	Variantenvergleiche	28
3.8	Methoden der Resultatsbewertung	30
3.8.1	Kreuzvalidierung der kNN-Referenzdatensätze	32
3.8.2	Vergleich der kNN-Klassifikation mit der visueller Interpretation	32
3.8.3	Visuelle Bewertung der produzierten Karten	33
3.8.4	Vergleich von Flächenanteilen	33
<b>4</b>	<b>Ergebnisse und Diskussion</b>	<b>34</b>
4.1	Identifikation und Eliminierung künstlicher Ausreißer	34
4.1.1	Ergebnisse der „Einschrittmethode“	34
4.1.2	Ergebnisse der „Protokollmethode“ nach 20 Runden	35
4.2	Kreuzvalidierung von kNN-Referenzdatensätzen	37

4.3	Genauigkeit der Klassifizierung im Vergleich mit der visuellen Interpretation _____	40
4.4	Karten und deren visuelle Bewertung _____	45
4.5	Vergleich der Flächenanteile aus dem CORINE-Datensatz und aus den produzierten Karten mit Flächenangaben aus anderen zuverlässigen Quellen _____	54
4.6	Diskussion der verschiedenen Bewertungsmethoden _____	59
<b>5</b>	<b>Zusammenfassung und Schlussfolgerungen _____</b>	<b>59</b>
<b>6</b>	<b>Literaturverzeichnis _____</b>	<b>63</b>

# 1 Einleitung

## 1.1 Ausgangssituation

Seitdem der Mensch Schrift und Zeichnung beherrscht, hält er Informationen über seine Umwelt auch in Darstellungen der Landoberfläche fest. Die Qualität von Karten entwickelt sich parallel zum technischen Stand und heute spielt die Beobachtung und Informationssammlung aus der Luft eine große Rolle.

Die Fernerkundung kann als „Gewinnung von Information über Art und Eigenschaften entfernter Objekte“ definiert werden (KRAUS UND SCHNEIDER 1988). Sie und ihre Anwendungen sind heute selbst aus dem Alltag nicht mehr wegzudenken und begegnen uns beispielsweise in Wettervorhersagen. Insbesondere durch die Verwendung von Satelliten hat sich die Fernerkundung in den letzten Jahrzehnten noch einmal stark weiterentwickelt. Dazu kommt der Fortschritt in den Bereichen der Aufnahme- sowie Datenverarbeitungstechnik (BAUER 2008).

In Übereinstimmung mit der Nachfrage nach flächigen Informationen über die Umwelt ist der Bedarf an Landbedeckungs- bzw. Landnutzungskarten heute von großer Bedeutung. Von Interesse sind auch Veränderungen der Landbedeckung und Hinweise auf die Dynamik in Landschaften. Dabei bestehen durch Fernerkundung z.B. Vorteile in Hinsicht auf die Größe der beobachtbaren Flächen und den reduzierten Aufwand im Vergleich zur Geländebegehung.

Bei der Interpretation und Beschreibung der Erdoberfläche und der Verwendung von Karten trifft man auf Klassen, die entweder die Landbedeckung oder die Landnutzung betreffen.

Unter Landbedeckung versteht man die Beschreibung der tatsächlichen Objekte, die die Oberfläche eines Gebietes trägt, sie ist also gegenständlich orientiert und kann beispielsweise Bezeichnungen wie „Gras“, „Laubwald“ oder „Häuser“ umfassen. Landnutzung dagegen bezieht sich auf die durch den Menschen bestimmte Verwendung der Fläche. So kann z.B. eine Fläche durch die Landbedeckung Gras gekennzeichnet sein, die Landnutzung derselben Fläche aber durch „sportliche Anlagen“ (wie Fußball- oder Golfplätze) beschrieben sein.

Fernerkundungsdaten sind durch räumliche, temporale, radiometrische und spektrale Auflösung gekennzeichnet. Die räumliche Auflösung beschreibt die kleinste identifizierbare Einheit bzw. die Pixel- („picture element“-) Größe in einem Rasterdatensatz. Die temporale Auflösung kann durch die zeitliche Verfügbarkeit der Daten beschrieben werden, also zum Beispiel in welchen Zeitabständen Informationen über ein bestimmtes Gebiet verfügbar sind. Die spektrale Auflösung wird durch die Anzahl der Kanäle des Aufnahmegerätes bestimmt, also wie viele verschiedene elektromagnetische Wellenlängenbereiche betrachtet werden. Die radiometrische Auflösung hingegen gibt an, in welchen Abstufungen die Intensität innerhalb des einzelnen Kanals gemessen werden kann.

Für die Arbeit motivierend ist die Vorstellung, die automatisierte Klassifizierung von Satellitenbildern so weit entwickelt zu sehen, dass (z.B. mit dem später beschriebenen kNN-Algorithmus (siehe Kapitel 3.3)) aufwandsarm praxisrelevante Karten hergestellt werden. Auf Grundlage von Satellitenbildern sollen also automatisch z.B. thematische Karten erstellt werden.

## 1.2 Ziele der Arbeit

Übergeordnetes Ziel der Arbeit ist die Untersuchung und Weiterentwicklung der Methodik zur Aktualisierung und räumlichen Verfeinerung bestehender Landbedeckungsdatensätze. Eines der Hauptziele ist zu untersuchen, wie es sich auf die Klassifizierungsqualität auswirkt, wenn man Daten von zwei Zeitpunkten im selben Untersuchungsgebiet kombiniert (siehe Kapitel 3.3).

Grund für diese Aufgabenstellung ist die Annahme, dass bei Kombination der Daten von zwei Zeitpunkten im Vergleich zur Verwendung nur eines Zeitpunktes zumindest bestimmte Klassen (vor allem Wiese und Acker) besser auseinandergehalten werden können. Dies liegt insbesondere an einer unterschiedlichen Erscheinung von Vegetation zu verschiedenen Zeitpunkten. So kann man visuell oder auch mittels eines Automatismus weniger genau bestimmen, ob ein Pixel zu Acker gehört, wenn es zu nur einem Zeitpunkt Spektralwerte von Vegetation aufweist. Wenn dagegen Aufnahmen von zwei Zeitpunkten vorliegen und zum Beispiel die Aprilaufnahme im Unterschied zur Augustaufnahme keine Vegetation aufweist, spricht dies für eine landwirtschaftliche Bodenbedeckung, denn Grünland würde jeweils Vegetation aufweisen.

Des Weiteren werden Effekte verschiedener Varianten der Herstellung von Referenzdatensätzen untersucht. Dabei soll auch die Identifizierung und Bereinigung von „Ausreißern“ (siehe Kapitel 3.3) analysiert und eventuell verfeinert werden.

## 1.3 Vorgehensweise

Wenn man die Methoden (siehe Kapitel 3) zusammenfasst, kann der gesamte Ablauf der Arbeit wie folgt beschrieben werden.

Für die Arbeit wurden ASTER-Bilddaten von Ober- und Niederösterreich verwendet (siehe Kapitel 2), welche für Klassifizierungen des Untersuchungsgebietes mit der kNN-Methode (siehe Kapitel 3.3) genutzt wurden. Referenzdaten für die Klassifizierungen stammten von CORINE. Hier ergab sich die Problematik der Generalisierung und fehlenden Aktualität des CORINE-Datensatzes, was zu Ausreißern (siehe Kapitel 3.3) im Referenzdatensatz führt. Zudem beinhaltet der CORINE-Datensatz das Problem, dass Landnutzungsdaten für eine Landbedeckungskarte verwendet werden, was ebenfalls zu Ausreißern führen kann (siehe Kapitel 2.1). Im Zuge der kNN-Kreuzvalidierung (Kapitel 3.6.3 und 3.8.1) wurde eine automatisierte Ausreißereliminierung (Kapitel 3.4.1) durchgeführt.

Die Qualität unterschiedlicher Klassifizierungsversionen wurde durch ständige Kreuzvalidierungen (siehe Kapitel 3.8.1) beurteilt. Dies hatte auch den Zweck, bestimmte Datenbearbeitungsvarianten auszuschließen und andere näher zu untersuchen. Weiter diente neben der visuellen Einschätzung der produzierten Karten (siehe Kapitel 3.8.3) auch die Erzeugung eines Referenzdatensatzes mit visuell interpretierten Kontrollpunkten (siehe Kapitel 3.5) zur Beurteilung der Klassifizierungen. Somit konnte die Genauigkeit der erhaltenen Klassifizierungen bzw. Karten mittels der visuell zugeordneten Landbedeckung abgeschätzt werden (siehe Kapitel 3.8.2).

## 2 Datengrundlagen

Die Datengrundlage besteht vor allem aus zwei Elementen. Die in der Arbeit verwendeten spektralen Informationen stammen aus ASTER-Bildern mit diversen Kanälen. Dabei standen 2 Szenen von je 2 Zeitpunkten zu Verfügung. Als Referenzdaten wurden CORINE-Daten verwendet.

Zusätzlich dienten Orthofotos als Referenzdaten für die visuelle Interpretation.

### 2.1 Eigenschaften des Untersuchungsgebietes

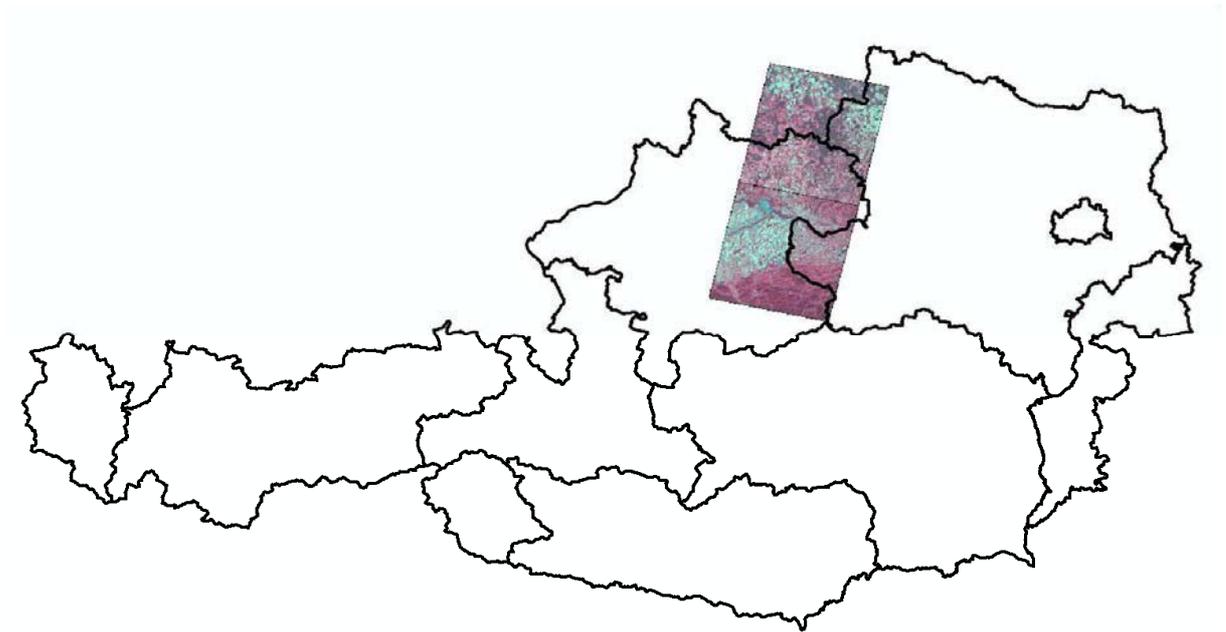


Abbildung 1: Lage der Augustszenen in Österreich

In der obigen Übersicht ist dargestellt, wo sich die zu untersuchenden Szenen innerhalb Österreichs und seiner Bundesländer befinden. Man sieht, dass ein Großteil der Szenen in Oberösterreich und weitere Teile in Niederösterreich sowie Tschechien liegen. Die auf tschechischem Gebiet liegenden Teile wurden nicht verwendet.

Bei der Betrachtung der Flächenanteile in Bezug auf die verwendeten ASTER-Bilder ergibt sich für die beiden Szenen des Augustzeitpunktes vom 31.08.2008 eine Gesamtfläche von 5.791 Quadratkilometern. Die Anteile der Flächen in Niederösterreich und Oberösterreich liegen bei 27% bzw. 73%.

Die Flächen, welche für die Herstellung der bisaisonalen Referenzdatensätze herangezogen wurden bzw. die Überschneidungsflächen der beiden Zeitpunkte, nehmen 5.256 Quadratkilometer ein, die sich zu 29% in Niederösterreich und zu 71% in Oberösterreich befinden.

## 2.2 CORINE Land Cover

CORINE steht für „Coordination of Information on the Environment“, also die „Koordinierung von Informationen über die Umwelt“ (ALBERTZ 2007). Es handelt sich um ein von der EU-Kommission in die Wege geleitetes Projekt. Darin ist vorgesehen, alle 10 Jahre europaweite Karten zu erstellen, welche die Landbedeckung bzw. Landnutzung darstellen. Der in dieser Arbeit verwendete Turnus wurde im Jahr 2000 durchgeführt. Ende 2009 ist der nächste Turnus auf Basis von Daten des Jahres 2006 abgeschlossen worden.

Zur Kartenerstellung werden Bilder des Erdbeobachtungssatelliten LANDSAT 7 visuell interpretiert. Durch diese Interpretation findet eine Landbedeckungskartierung statt. Die europäische Umweltagentur in Kopenhagen beurteilt in der Folge die dargestellte Flächennutzung.

CORINE ist eine Klassifizierung der Erdoberfläche, die gleichzeitig Landbedeckungs- sowie Landnutzungs-Klassen aufweist. Zudem beinhaltet der CORINE-Datensatz ein hohes Maß an Generalisierung der Landschaft. So gilt für ein Polygon bzw. eine interpretierte Fläche das Mindestmaß von 25 ha. Weiters gibt es Grundsätze für Generalisierungen im Rahmen der Definition einzelner Klassen, die insbesondere auf Mischklassen abzielen, also verschiedene Landbedeckungen enthalten.

Bei der Verwendung von CORINE-Daten für Referenzdatensätze stellt sich das Problem, dass CORINE eher Landnutzungs- als Landbedeckungsklassen aufweist, eine automatisierte, pixelbasierte Analyse aber nur auf die tatsächliche Bedeckung des Bodens abzielen kann.

## CORINE Landcover Nomenklatur (deutsch)

44 Klassen (Ö: 28 Klassen)

1. Bebaute Fläche	1.1. Städtisch geprägte Flächen	1.1.1. durchgängig städtische Prägung 1.1.2. nicht durchgängig städtische Prägung
	1.2. Industrie-, Gewerbe- und Verkehrsflächen	1.2.1. Industrie/Gewerbeflächen 1.2.2. Straßen/Eisenbahnnetze, funktionell zugeordnete Flächen 1.2.3. Hafengebiete 1.2.4. Flughäfen
	1.3. Abbauflächen, Deponien, Baustellen	1.3.1. Abbauflächen 1.3.2. Deponien, Abraummhalden 1.3.3. Baustellen
	1.4. Künstlich angelegte nicht landwirtschaftlich genutzte Flächen	1.4.1. Städtische Grünflächen 1.4.2. Sport/Freizeitanlagen
2. Landwirtschaft	2.1. Ackerflächen	2.1.1. Nicht bewässertes Ackerland 2.1.2. Regelmäßig bewässertes Ackerland 2.1.3. Reisfelder
	2.2. Dauerkulturen	2.2.1. Weinbauflächen 2.2.2. Obst/Beerenobstbestände 2.2.3. Olivenhaine
	2.3. Grünland	2.3.1. Wiesen und Weiden
	2.4. Heterogene landwirtschaftliche Flächen	2.4.1. Einjähr. Kulturen in Verbindung mit Dauerkulturen 2.4.2. Komplexe Parzellenstruktur 2.4.3. Landwirtschaftlich genutztes Land mit Flächen natürlicher Vegetation von signifikanter Größe 2.4.4. Land/Forstwirtschaftliche Flächen
3. Wälder und naturnahe Flächen	3.1. Wälder	3.1.1. Laubwälder 3.1.2. Nadelwälder 3.1.3. Mischwälder
	3.2. Kraut/Strauchvegetation	3.2.1. Natürliches Grünland 3.2.2. Heiden und Moorheiden 3.2.3. Hartlaubbewuchs 3.2.4. Wald/Strauch Übergangsstadien
	3.3. Offene Flächen ohne oder mit geringer Vegetation	3.3.1. Strände, Dünen, Sandflächen 3.3.2. Felsflächen ohne Vegetation 3.3.3. Flächen mit spärlicher Vegetation 3.3.4. Brandflächen 3.3.5. Gletscher/Dauerschneegebiet
4. Feuchtflächen	4.1. Feuchtflächen im Landesinneren	4.1.1. Sümpfe 4.1.2. Torfmoore
	4.2. Feuchtflächen an der Küste	4.2.1. Salzwiesen 4.2.2. Salinen 4.2.3. In der Gezeitenzone liegende Flächen
5. Wasserflächen	5.1. Wasserflächen im Landesinneren	5.1.1. Gewässerläufe 5.1.2. Wasserflächen
	5.2. Meeressgewässer	5.2.1. Lagunen 5.2.2. Mündungsgebiete 5.2.3. Meer und Ozean

Abbildung 2: CORINE-Klassen (ÖSTERREICHISCHES UMWELTBUNDESAMT 2010)

Abbildung 2 zeigt den Katalog an CORINE-Klassen. Die grau hinterlegten Klassen sind in Österreich nicht ausgewiesen. Von den 44 enthaltenen Klassen kommen somit 28 in Österreich vor. Das Untersuchungsgebiet weist 20 CORINE-Klassen auf.

## 2.3 ASTER

ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) ist ein Sensor auf dem NASA-Satelliten Terra. Er besitzt 14 Kanäle mit unterschiedlichen räumlichen Auflösungen. Die Kanäle 1 bis 3 haben eine 15m-Pixelgröße, die Kanäle 4 bis 9 eine 30m- und die Kanäle 10 bis 16 eine 90m-Pixelgröße. Die Flughöhe beträgt 705 km und der erfasste Geländestreifen hat eine Breite von 60 km, woraus dann 60 mal 60 km Szenen (d.h. 3600 km<sup>2</sup>) geschnitten werden. Die temporale Auflösung bzw. die Wiederholrate beträgt 16 Tage. (JENSEN 2007, BAUER 2008)

Die größte Bedeutung für die Arbeit hatten die Kanäle 1 bis 3 im Bereich des sichtbaren Lichtes und nahen Infrarots, auf Englisch mit VNIR abgekürzt („Visible and Near InfraRed“). Grün wird dabei in einem Wellenlängenbereich von 0.52–0.60 µm, rot von 0.63–0.69 und nahes Infrarot von 0.78–0.86 µm erfasst. Die Bänder des mittleren Infrarots (SWIR stehend für „ShortWave InfraRed“) waren aufgrund technischer Probleme auf dem Satelliten ausgefallen. Allerdings wären sie wegen der geringeren räumlichen Auflösung eher nur testweise verwendet worden, auch wenn sie zum Beispiel für die „Erkennung“ der Klasse Sumpf wegen ihrer Wassersensibilität nützlich sein könnten. Der Einfluss der thermalen Kanäle (TIR stehend für „Thermal Infrared“) auf die Qualität des Referenzdatensatzes und der Kartierung hingegen konnte untersucht werden.

Als Basis für die Arbeit dienten zwei Szenen, welche am 31. August 2008 aufgenommen wurden. Dazu kam ein weiterer Datensatz zweier im Vergleich zum ersten Datensatz leicht verschobener Szenen vom 9. April 2008. Zuerst fanden die August-Daten Verwendung, da sie als erstes zur Verfügung standen und auf diesen Bildern im Unterschied zu den Aprilbildern kein Schnee vorkommt. Die Aprilszenen liegen ca. 6,5 km weiter östlich, was eine Überschneidung der 60 km breiten Szenen auf 53,5 km bewirkt. Die Szenen befanden sich größtenteils in Oberösterreich und zu einem geringeren Anteil in Niederösterreich. Teile der nördlichen Szenen lagen in Tschechien, wurden aber nicht verwertet. Den Klassifizierungen, die sich auf Daten von 2 Zeitpunkten bezogen, lagen 2 Bilddatensätze zugrunde, welche aus den Überschneidungen (Kombinationen) der jeweils nördlichen und südlichen Szenen bestanden. Die nördliche Szene des Aprilzeitpunktes ergab also zusammen mit der nördlichen Szene des Augustzeitpunktes ein Bild mit entsprechend doppelter Anzahl an Kanälen. Gleiches gilt für die südlichen Szenen.

Zur Georeferenzierung der Daten durch den Anbieter ist festzuhalten, dass eine Kontrolle eine durchschnittliche Abweichung des August-ASTER-Bildes von knapp 15 Metern, also einem Pixel, von der ÖK 50 ergab. ÖK 50 bedeutet Österreichische Karte im Maßstab 1:50.000 und ist die vom Bundesamt für Eich- und Vermessungswesen herausgegebene amtliche topographische Karte Österreichs. Zur Feststellung der Abweichung wurden 15 im ASTER-Bild sowie in der ÖK 50 gut identifizierbare Stellen mit Punkten markiert und verglichen. Die meisten Punkte wiesen eine südliche Abweichung des ASTER-Bildes auf. Die Abweichung wird als tolerabel angesehen, wobei stets bedacht werden muss, dass fehlende geometrische Übereinstimmung der verschiedenen Daten als eine Hauptfehlerquelle für Klassifizierungen gilt (TOMPPÖ et. al 2008, a).

## 2.4 Orthofotos

Die im Internet aufgerufenen Orthofotos dienen vor allem der in Kapitel 3.5 beschriebenen visuellen Interpretation und Erstellung des Kontrolldatensatzes. Sie halfen bei der Bestimmung der Landbedeckungsklasse der Kontrollpunkte. Als Quelle und Datengrundlage dienen die im Internet öffentlich zugänglichen Geoinformationssysteme der Länder Niederösterreich und Oberösterreich. Namentlich sind dies der Niederösterreich-ATLAS und das Digitale Oberösterreichische Raum-Informationssystem (für die Internetadressen siehe Kapitel 6, Literaturverzeichnis).

# 3 Methoden

Übergeordnete Methodik ist die Klassifizierung von Landbedeckung mittels der kNN-Methode, wobei ASTER-Bilder mit CORINE-Referenzdaten klassifiziert werden. Als Einleitung des Methodenkapitels werden im Folgenden Eigenschaften der Datensätze beschrieben und Begriffsdefinition in Bezug auf die Daten aufgeführt. Die anschließenden Kapitel befassen sich mit der Erstellung der Datensätze sowie deren Verwendung und Bewertung.

## 3.1 Fernerkundungsdatensatz und Referenzdatensatz

Ein Fernerkundungsdatensatz, wie der ASTER-Bilddatensatz im vorliegenden Fall, dient als Grundlage für Berechnungen bestimmter Variablen wie hier der Landbedeckungsklasse. Ein Referenzdatensatz, wie im vorliegenden Fall der CORINE-Datensatz, erlaubt die Berechnungen und Vorhersagen zu den Fernerkundungsdaten. Er ist also eine Grundlage, auf der Aussagen über die größere Population des Fernerkundungsdatensatzes getroffen werden.

Bei einem Referenzdatensatz handelt es sich um Daten, auf welche andere Daten Bezug nehmen. Referenz leitet sich ab von (lat.) referre- wiederbringen, zurücktragen zu lat. ferre- tragen (KLUGE 2002). Referenzdaten in der Fernerkundung sind Daten, die für Maß- oder Vergleichszwecke für die Analyse der Fernerkundungsdaten dienen.

Eine Population und auch ein Datensatz bestehen aus einzelnen Beobachtungen bzw. Einheiten. Eine solche Beobachtung wird durch diverse Variablen charakterisiert. Manche dieser Variablen liegen für alle Einheiten des Datensatzes vor. Diese Variablen werden X-Variablen genannt und können durch Vergleich der Werte für Ähnlichkeitsmaße herangezogen werden. Variablen, die nur bei einem Teil der Beobachtungen bestimmt sind, werden Y-Variablen genannt. Beobachtungen, bei denen sowohl X- als auch Y-Variable vorliegen, werden Referenzbeobachtungen oder wie im Folgenden Referenzpunkte genannt. Die Gesamtheit der Referenzpunkte wird in der Folge als „kNN-Referenzdatensatz“ bezeichnet. Einheiten ohne bestimmte Y-Variable nennt man Ziel-Beobachtungen. Über die Ähnlichkeitsmaße der X-Variablen und anhand der bekannten Y-Variablen kann man die unbekanntes Y-Variablen der Ziel-Beobachtungen z.B. durch das kNN-Programm berechnen. Es gibt stetige und auch kategoriale Y-Variablen. Bei der Berechnung stetiger Y-Variablen durch die kNN-

Methode spricht man von Regression. Bei der Ermittlung von kategorialen Y-Variablen handelt es sich um Klassifizierung, womit sich diese Arbeit befasst.

Im Fall dieser Arbeit bezieht sich eine Einheit oder Beobachtung auf ein bestimmtes Pixel, das zusätzlich durch Koordinaten bestimmt ist. Die Lokalisierung der Pixelfläche wird bei den Referenzdaten durch die Angabe eines Punktes vorgenommen. Die X-Variablen sind hier spektrale Werte und die Y-Variablen beziehen sich auf die Landnutzung bzw. Landbedeckung. (KOUKAL 2009)

Die durch die CORINE-Daten repräsentierten Referenzdaten wurden nicht flächendeckend verwendet. Es erfolgte vielmehr eine „Ausdünnung“ durch Auswahl von rasterförmig verteilten „Referenzpunkten“. Für die Verwendung bei der kNN-Klassifizierung wurde ein „kNN-Referenzdatensatz“ hergestellt, der für die Referenzpunkte sowohl die Landbedeckungskategorie (Referenzinformation von CORINE) als auch die zugehörige Spektralinformation aus den ASTER-Bildern enthält.

## **3.2 Erzeugung der Datensätze**

Das Kapitel beschreibt das Zusammenfügen der verschiedenen Datenquellen zu Referenzdaten und deren Bearbeitung. Diese dienen als Grundlage für die später beschriebene Klassifizierung.

### **3.2.1 Erstellung des Rasters**

Für die Referenzpunkte sollte ein Raster mit der Rasterweite von (ungefähr) 500 Metern erstellt werden. Diese Rasterweite wurde auf exakt 495 m festgesetzt, um ein Vielfaches der Pixelgröße von 15 m zu haben und somit auch einen gleichbleibenden Abstand an Pixeln zwischen den Referenzpunkten zu bewirken. Also wurde das 33-fache der Pixelgröße als Rasterweite gewählt.

Für zwei Szenen, die insgesamt eine Fläche von 120 mal 60 km, also 7200 km<sup>2</sup> einnehmen, bedeutet das eine Anzahl von ca. 29400 Referenzpunkten. Diese ursprüngliche Anzahl reduzierte sich in einem ersten Schritt um die Punkte, welche in Tschechien liegen. Nach Auswahl der in Österreich liegenden Punkte verblieben 23881 Punkte im Datensatz.

Zur Erstellung des kNN-Referenzdatensatzes folgte das Auslesen der CORINE-Landbedeckungskategorie und der spektralen Informationen aus dem Fernerkundungsdatensatz, also das Verbinden der CORINE-Referenzinformation mit der Spektralinformation in Bezug auf eine bestimmte Koordinate.

### **3.2.2 Erstellung des bisaisonalen Fernerkundungsdatensatzes**

Um einen Datensatz zu erhalten, der aus den Kanälen beider Zeitpunkte zusammengesetzt ist, wurde eine räumliche Anpassung der beiden Zeitpunkte aneinander vorgenommen. Dabei fand die

Verschiebung der Aprilszenen in Bezug auf die nicht veränderten Augustszenen statt. Die Szenen wurden in Viertel aufgeteilt, worauf in diesen die Feststellung der räumlichen Abweichungen zwischen den Zeitpunkten durch Messung mittels Kontrollpunkten stattfand. Dabei ergab sich, dass bei der südlichen Szene die beiden unteren und die beiden oberen Viertel sehr ähnliche Abweichungen aufwiesen. In der nördlichen Szene wurde sogar eine übereinstimmende Abweichung aller Viertel festgestellt. Um einen „layerstack“, also eine Stapelung bzw. Kombination der Kanäle der beiden Zeitpunkte zu erstellen, fand hier entsprechend nur eine Verschiebung des Aprilzeitpunktbildes auf das Augustzeitpunktbild statt. Bei der südlichen Szene wurden vergleichbar zwei „layerstacks“ der Augustszenen mit den 2 verschiedenen verschobenen Teilen der Aprilszenen hergestellt.

### **3.2.3 „Homogenitätskriterium“ für die Rasterpunkte**

In einem weiteren Schritt erfolgte die Reduktion der Anzahl der Referenzpunkte um diejenigen, welche in der Nähe des Übergangs zweier Klassen im CORINE-Datensatz liegen. Dies dient der Reduzierung von Mischpixeln, also solchen, die nicht eindeutig einer Klasse zuzuordnen sind. Dazu wurde ein Feld von 3 mal 3 Pixeln, also 45 mal 45 Metern bzw. 0,2025 ha, um jeden Referenzpunkt gebildet. Wenn nicht alle Pixel des Feldes in derselben Klasse lagen, wurde der Referenzpunkt aus dem Datensatz entfernt. Dies ergab eine Reduzierung des Datensatzes von 23881 um 2205 Punkte oder 9,2% auf 21676 Punkte.

### **3.2.4 „Auffüllpunkte“**

Einige Klassen lieferten im regelmäßigen Raster von 495 Metern aufgrund ihrer geringen Fläche letztendlich sehr wenige Referenzpunkte, zumal sich ihre Anzahl durch die Ausreißerelimination (siehe Kapitel 3.4.1) weiter verringerte. Daher wurden zusätzliche Punkte gleichmäßig in Klassen mit weniger als 50 Referenzpunkten bzw. in die zugehörigen Flächen gesetzt, um dann aus ihnen Referenzpunkte zu erzeugen. Dadurch erhöhte sich die Anzahl des Referenzdatensatzes um 5600 auf 27276 Punkte.

### **3.2.5 Maskenerstellung**

Die Erstellung sogenannter Masken dient der Ausblendung von Informationen bestimmter Regionen. Wenn also an Orten der Bilder keine brauchbaren Daten vorhanden sind, was z.B. im Falle von Wolken, deren Schatten oder bei Vorhandensein von Schnee auftritt, werden an diesen Stellen keine Referenzpunkte platziert. Die Bilder beinhalten dort keine relevanten Informationen, und diese Stellen werden bei der Klassifizierung ausgespart. Entsprechende Flächen werden visuell abgegrenzt bzw. interpretiert. Dazu werden in der Software „ArcMap“ (siehe Kapitel 3.6.1) erstellte Polygone in eine

.bip-Datei umgewandelt, welche als „Maskfile“ fungiert und im kNN-Programm (siehe Kapitel 3.6.3) das Ignorieren dieser Regionen bewirkt.

### **3.2.6 Klassenreduktion**

Insbesondere aufgrund der Problematik der in CORINE gegebenen Vermischung von Landbedeckungs- und Landnutzungsaspekten wurde der Klassenkatalog für die weitere Verwendung innerhalb der Arbeit verändert. CORINE-Klassen wie „Sport und Freizeitanlagen“, die sich auf die Landnutzung beziehen, wurden entfernt. Der Klassenkatalog sollte nur Klassen beinhalten, welche allein die Landbedeckung beschreiben. Hauptgrund für diese Vorgehensweise ist der engere Zusammenhang zwischen den spektralen Werten eines Objektes mit der Landbedeckung als mit der Landnutzung.

Die CORINE-Klassen sind auch häufig durch Kombinationen (Mosaik) verschiedener Landbedeckungen gekennzeichnet. So besteht die Klasse „Landwirtschaft mit natürlicher Bodenbedeckung“ per CORINE-Definition vor allem aus einer Mischung von Landwirtschaft und Waldflächen (BOSSARD et al. 2000). Weiters wurden CORINE-Klassen wie „natürliches Grasland“ und „Wiesen und Weiden“, bei denen von gleichen spektralen Informationen ausgegangen werden kann, zusammengelegt. Der Klassenkatalog wurde also nach dem Kriterium eines Zusammenhangs der spektralen Information des einzelnen Pixels mit der jeweiligen Klasse abgeändert, welcher bei Landnutzungsklassen z.B. wegen der Mischung von Landbedeckungen innerhalb einer Klasse nicht gegeben ist. Die 20 im Gebiet vorkommenden CORINE-Klassen wurden auch in Abstimmung mit der visuellen Interpretation (siehe Kapitel 3.5 und 3.8.2) so letztendlich auf 9 reduziert. Es fanden auch andere Varianten der Klassenreduktion auf beispielsweise 12 Klassen statt, die hier jedoch nicht weiter beschrieben werden.

CORINE-Code	Bezeichnung der CORINE-Klasse	modifizierte Klasse
111	Flächen durchgängig städtischer Prägung	versiegelte Flächen
112	Flächen nicht-durchgängig städtischer Prägung	
121	Industrie- und Gewerbeflächen	
122	Straßen und Eisenbahn	
124	Flughäfen	weggelassen
131	Abbauflächen	übernommen
141	Städtische Grünflächen	weggelassen
142	Sport und Freizeitanlagen	weggelassen
211	Nicht bewässertes Ackerland	übernommen
231	Wiesen und Weiden	Grünland
242	Komplexe Parzellenstrukturen	weggelassen
243	Landwirtschaft mit natürlicher Bodenbedeckung	weggelassen
311	Laubwald	übernommen
312	Nadelwald	übernommen
313	Mischwald	weggelassen
321	Natürliches Grasland	Grünland
324	Wald-Strauch-Übergangsstadien	weggelassen
411	Sümpfe	übernommen
511	Gewässerläufe	übernommen
512	Wasserflächen	übernommen

Tabelle 1: Überarbeitung der vorkommenden CORINE-Klassen

Für die Modifizierung des Klassenkatalogs wurde jede einzelne Klasse in Bezug auf das Vorhandensein von Mischklassen, die spektrale Trennbarkeit von Klassen sowie die Landnutzungs-Landbedeckungsproblematik überprüft.

So fiel die Entscheidung, die Klassen „Flächen durchgängig städtischer Prägung“, „Flächen nicht-durchgängig städtischer Prägung“, „Industrie- und Gewerbeflächen“ und „Straßen und Eisenbahn“ zusammenzulegen. Die beiden erstgenannten unterscheiden sich auf Pixelebene nicht, sondern nur dadurch, dass zwischen die versiegelten Flächen der Klasse „Flächen nicht-durchgängig städtischer Prägung“ insbesondere Vegetation in Form von Gärten etc. eingestreut ist. Letztendlich kam es zu einer Zusammenfassung der 4 Klassen zur Klasse „versiegelte Flächen“. Die Klasse „Flughäfen“ wurde entfernt, da sie eine Mischung aus Grasflächen und versiegelten Flächen (Gebäude und Landebahn) ist. „Städtische Grünflächen“ setzen sich aus Bäumen und Grasflächen zusammen, weshalb die Klasse entfernt wurde. Gleiches gilt für „Sport- und Freizeitanlagen“, die vor allem als Golfplätze vorliegen. „Komplexe Parzellenstrukturen“ sind eine Zusammensetzung aus Wiesen und Weiden, Acker sowie Bäumen und wurden daher als Klasse entfernt. Die Klasse „Landwirtschaft mit natürlicher Bodenbedeckung“ ist als ein Wechsel von Landwirtschaft und Wald definiert (BOSSARD et al. 2000) und somit auf Pixelebene nicht definierbar, weshalb sie entfernt wurde. Auch die Klasse „Mischwald“ ist auf Pixelebene meist entweder Laub- oder Nadelwald. Daher fand sie keine Verwendung, was auch den Grund hat, dass sie im Laufe der visuellen Interpretation weder in den ASTER-Bildern noch auf den Orthofotos eindeutig zu identifizieren war. Das gleichzeitige Vorhandensein einer Klasse im visuell interpretierten Datensatz wie auch im Referenzdatensatz ist für die spätere Bewertung der Klassifizierung unabdingbar. Der Annahme folgend, dass „Natürliches

Grasland“ spektral nicht von „Wiesen und Weiden“ zu unterscheiden ist, wurden diese Klassen zu „Grünland“ zusammengefasst. Die Klasse „Wald-Strauch-Übergangsstadien“ wurde weglassen, da bei Betrachtung eines einzelnen Pixels dieses in eine Wald-Klasse oder in die Klasse „Grünland“ fallen würde und auch visuell kaum als „Wald-Strauch-Übergangsstadien“-Klasse identifizierbar ist.

### 3.3 kNN-Klassifizierung

Unter Klassifizierung versteht man das Zusammenfassen von Objekten oder Gegenständen. Der Begriff Klassifikation entwickelte sich aus den lateinische Worten „classis“ (Abteilung) und „ficare“ (machen) und beschreibt die systematische Einordnung in Klassen. Die Zuteilung der Objekte zu den verschiedenen Klassen erfolgt mittels der Objektmerkmale. Es werden somit Informationen geordnet und Muster erkannt (BROCKHAUS 2006, a).

In der Bildverarbeitung werden einzelne Pixel z.B. anhand spektraler Signaturen, also unterschiedlicher Reflexionseigenschaften der Oberflächen, Klassen wie „Wald“ oder „Gewässer“ zugeordnet (FERNERKUNDUNGSLEXIKON 2010).

In Bezug auf die Art der Klassifizierung unterscheidet man zwei Gruppen. Bei nicht überwachter, automatischer Klassifizierung sind die Klassen nicht vorgegeben. So werden in der Fernerkundung anhand von spektralen Ähnlichkeiten Gruppen mit ähnlichen Pixeln gebildet. Diese Gruppen werden durch Interpretation zu Klassen umfunktioniert.

Demgegenüber steht die überwachte Klassifizierung mit festgelegten Klassen. Beim Beispiel Fernerkundung werden hier Gebiete mit bekannten Klassen als sogenannte Trainingsgebiete genutzt. Für jede Klasse sucht man also typische Pixel. Anhand der statistischen Verteilung der Pixelwerte pro Klasse findet dann die Einteilung aller Pixel zu den entsprechenden Klassen statt.

In einer weiteren Ebene sind parametrische und nichtparametrische Verfahren zu unterscheiden. Parametrische Verfahren nehmen analytisch beschreibbare Wahrscheinlichkeitsverteilungen der Pixelwerte in jeder Klasse an. Die Verteilungsparameter dieser Wahrscheinlichkeitsverteilungen werden aus Trainingsdaten hergeleitet. (BAUER 2008). Die in der Arbeit verwendete und unter Kapitel 3.3 beschriebene kNN-Methode ist ein überwachtes parameterfreies Klassifizierungsverfahren.

Eine Variante ist die multitemporale Klassifizierung, die wegen häufig ähnlicher spektraler Signaturen der einzelnen Pflanzenarten oder auch Landnutzungsformen wie im Fall dieser Arbeit Anwendung findet. Sie nutzt durch verschiedene Wachstumsrhythmen verursachte spektrale Unterschiede.

Grundsätzlich ist die Unterscheidung verschiedener Pflanzenarten in der Fernerkundung wegen nicht absolut eindeutig trennbarer Merkmale nach wie vor schwer möglich. Auch das „Auftreten verschiedener Signaturen pro Pflanzenart in Abhängigkeit von Wachstumsphase, Feuchtigkeit des Erdreichs, atmosphärischen Bedingungen u.a.“ erschwert bislang Klassifizierungen und verhindert vollautomatisierte Verfahren (FERNERKUNDUNGSLEXIKON 2010). In der vorliegenden Arbeit werden bei der bisaisonalen Klassifizierung die sich mit der Wachstumsphase ändernden Signaturen für die Verbesserung des Klassifizierungsergebnisses ausgenutzt.

kNN bedeutet k-Nächste-Nachbarn, wobei k für eine wählbare Zahl steht. Die kNN-Methode ist ein Algorithmus zur Klassifizierung. Unter Algorithmus versteht man ein Lösungsverfahren, das vorgeschriebene Mittel bzw. Verfahren bei der Bearbeitung gleichartiger Aufgaben verwendet (BROCKHAUS 2006, b). In dieser Arbeit besteht der Algorithmus aus dem Ableiten von Klassen aus den Differenzen spektraler Werte.

Diesbezüglich werden im spektralen Merkmalsraum die euklidischen Distanzen eines zu klassifizierenden Elementes zu den Referenzelementen berechnet. Dazu wird die Summe der Quadrate aller Differenzen der einzelnen Kanäle zwischen dem zu klassifizierenden Element (Pixel) und jedem Referenzpixel berechnet und daraus jeweils die Wurzel gezogen (MITCHELL 1997). Sind zwei Pixel spektral identisch, beträgt ihre euklidische Distanz somit 0. Die k nächsten Nachbarn bestimmen, welcher Klasse das entsprechende Pixel zugeordnet wird. Weil die Distanz zu jedem Referenzpixel berechnet werden muss, besteht bei der kNN-Methode ein hoher Rechenaufwand, was aber angesichts der Kapazitäten heutiger Computer nicht mehr stark ins Gewicht fällt (STÜMER und KÖHL 2005).

Die Umsetzung geschah jeweils mit dem am IVFL entwickelten kNN-Programm (KOUKAL 2009) (siehe Kapitel 3.6.3) innerhalb der Module „crossvalidation“ und „mapmaker“. Das Modul „crossvalidation“ fand neben der Kreuzvalidierung auch in der Ausreißersuche und der Gegenüberstellung der Referenzdatensätze mit den visuell interpretierten Daten Anwendung.

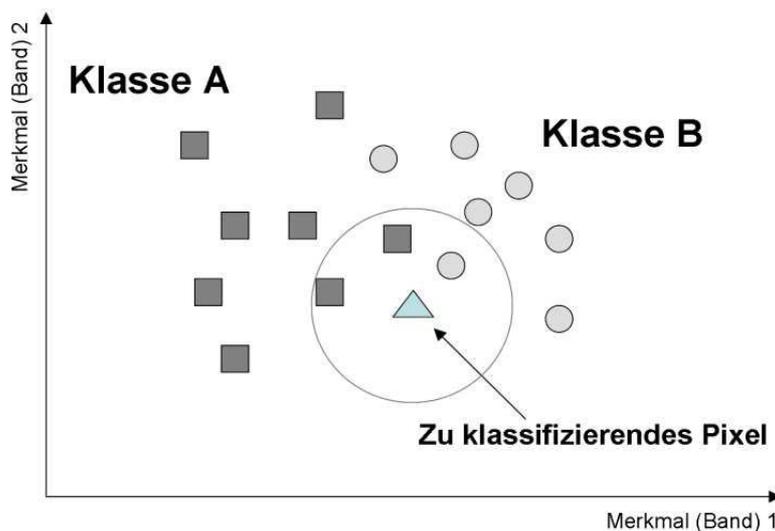


Abbildung 3: kNN-Methode für ein Beispiel mit  $k = 3$  (POPOVICI 2008)

Die obige Abbildung ist ein Beispiel für die Klassifizierung einer hier als Dreieck dargestellten Beobachtung in einem zweidimensionalen Merkmalsraum. Die Referenzdaten gehören im Beispiel entweder der mit Quadraten dargestellten Klasse A oder der durch Kreise symbolisierten Klasse B an. Bei der Klassifizierung mit  $k = 3$  sucht der Klassifizierungsalgorithmus die 3 nächsten Nachbarn als bestimmende Referenzen. Diese 3 Elemente mit den geringsten Distanzen befinden sich in einem Kreis mit dem zu klassifizierenden Pixel als Mittelpunkt, wobei 2 von ihnen der Klasse A und einer der Klasse B angehören. Deshalb wird der für die Klasse B sprechende Punkt von den 2 der Klasse A überstimmt, und die Klassifizierung zu Klasse A erfolgt.

### 3.4 Ausreißer

Als Ausreißer oder „Outlier“ bezeichnet man einen Messwert oder Befund, der nicht in die Messreihe oder Erwartung passt. Im Rahmen der vorliegenden Arbeit passt also bei einem Ausreißer im kNN-Referenzdatensatz die spektrale Information nicht zur CORINE-Landbedeckungsklasse. Das kann sich dadurch offenbaren, dass bei der Kreuzvalidierung (siehe Kapitel 3.8), bei der jeweils eines der Referenzpixel als unbekannt angenommen und durch Vergleich mit allen anderen Referenzpixeln klassifiziert wird, die gegebene nicht mit der berechneten Klasse übereinstimmt.

Dies kann daran liegen, dass ein Mischpixel (spektrale Uneindeutigkeit) vorliegt, das Pixel also unterschiedliche Landbedeckungen enthält. Dieses Problem wird allerdings durch das in Kapitel 3.2.3 beschriebene „Homogenitätskriterium“ verringert.

Auch die Generalisierung im Zuge der Erstellung des CORINE-Datensatzes führt zu Ausreißern. Dies ist beispielsweise der Fall, wenn ein Referenzpunkt in einer Fläche einer bestimmten Klasse liegt, die im CORINE-Datensatz nicht erfasst wurde, da sie die Mindestgröße von 25 ha nicht erreicht.

Des Weiteren kann eine Änderung der Landbedeckung zu einer Abweichung zwischen dem Klassifizierungsergebnis, welches sich auf Daten von 2008 stützt, und der sich auf das Jahr 2000 beziehenden CORINE- Landbedeckung führen und damit Ausreißer ergeben. (BROCKHAUS 2006, c; KOUKAL 2009; TOMPPU et. al 2008, b)

#### 3.4.1 Ausreißeridentifizierung und -eliminierung

Im Folgenden werden zwei Methoden der Ausreißereliminierung beschrieben, deren Untersuchung eines der zentralen Ziele der Arbeit war. Dabei ist die erste, hier „Protokollmethode“ benannte, Methode vor längerer Zeit entwickelt und bereits häufiger angewendet worden. Das zweite, hier „Einschrittmethode“ genannte, Verfahren wurde dagegen erst im Laufe dieser Arbeit entwickelt und erprobt.

##### 3.4.1.1 „Protokollmethode“

Die kNN-Software beinhaltet eine Ausreißer-Identifizierung. Die Kreuzvalidierung („crossvalidation“, siehe Kapitel 3.8.1) des Referenzdatensatzes wird bei  $k = 1$  durchgeführt. Dabei wird protokolliert, wie oft ein Referenzpunkt zur Klassifizierung anderer Punkte herangezogen und ob jeweils richtig klassifiziert wird. Am Ende erhält man für jedes Referenzpixel die Häufigkeit der Auswahl als nächster Nachbar sowie die Anzahl der richtigen Klassifizierungen und die entsprechende Quote.

Punkte, die häufig als nächster Nachbar ausgewählt werden und meist zu falschen Klassifizierungen führen, können als Ausreißer eliminiert werden. Punkte, die häufig ausgewählt

werden und (meist) zu richtigen Klassifizierungen führen, gehen in den endgültigen Referenzdatensatz ein (KOUKAL 2009).

Bei dieser Arbeit wurden Punkte als Ausreißer und „schlechte“ Referenzpunkte definiert und eliminiert, die mindestens zweimal ausgewählt wurden und deren „Trefferquote“ geringer als 50 Prozent war. Mit „Trefferquote“ ist der Anteil „korrekter“ Klassifizierungen eines Referenzpunktes an der Anzahl bzw. Häufigkeit seiner Auswahl zum nächsten Nachbarn gemeint. Eine „korrekte“ Klassifizierung bedeutet das Übereinstimmen der Klassifizierung mit der CORINE-Klasse. Referenzpunkte, die mindestens zweimal als nächste Nachbarn herangezogen wurden und dabei mindestens 66% Treffer erzielten, werden als „gute“ eingestuft. Als „indifferente“ sind Referenzpunkte bezeichnet, die weniger als zweimal ausgewählt wurden oder deren „Trefferquote“ zwischen 50 und 66 Prozent lag. Nach Entnahme der „guten“ sowie der „schlechten“ Referenzpunkte aus dem Gesamtdatensatz kann die Ausreißer-Identifizierung mit den verbleibenden „indifferenten“ Punkten abermals durchgeführt werden, um wieder „gute“ und „schlechte“ Referenzpunkte herauszufiltern. Dieser Prozess wird mehrfach wiederholt, bis keine Ausreißer mehr erkannt werden oder genügend „gute“ Referenzpunkte für sinnvolle Klassifizierungen vorhanden sind. Der endgültige Referenzdatensatz setzt sich aus den „guten“ Referenzpunkten der verschiedenen Filterungsstufen sowie wahlweise zusätzlich den „indifferenten“ Referenzpunkten der letzten Filterung zusammen.

#### **3.4.1.2 „Einschrittmethode“**

Bei diesem Verfahren wird der Ausgangsreferenzdatensatz mit einem k-Wert klassifiziert, von dem man annehmen kann, dass er gute Ergebnisse erzielt. Dann kann man Referenzpunkte, bei denen die gegebene (CORINE-)Klasse nicht mit der berechneten übereinstimmt, als Ausreißer identifizieren und entfernen.

Der so entstandene Datensatz kann zur Abschätzung seiner Qualität wiederum durch die „crossvalidation“ laufen. Macht man das mit verschiedenen „k-Werten“ zu Anfang des Prozesses, kann somit der „optimale“ k-Wert gefunden werden, der zum Beispiel die höchste Gesamtgenauigkeit des bereinigten Datensatzes bewirkt.

#### **3.4.2 Suche nach künstlichen Ausreißern**

Um die Effektivität der Ausreißersuche der verschiedenen Methoden objektiv beurteilen zu können, wurde ein Datensatz mit künstlich erzeugten Ausreißern erstellt. Dazu wurde die Klasse einer Auswahl von Referenzpunkten in einem bereits ausreißerbereinigten Datensatz bewusst verändert. Die Ausreißerbereinigung fand im Rahmen der „Einschrittmethode“ (durch eine Klassifizierung mit  $k = 22$ ) statt. Somit konnte man sicher sein, dass es sich bei den veränderten Daten auch wirklich um Ausreißer handelte. Um sicherzustellen, dass bei der anschließenden Klassifizierung die veränderten Referenzpunkte nicht die anderen dominieren (die Klasse „umkippt“), wurden maximal 10% der Referenzpunkte der Klasse im bereinigten Datensatz verändert. Die Erstellung künstlicher Ausreißer

fand bei zwei Klassen statt. In der Klasse „Nadelwald“, die 3876 Punkte im bereinigten Datensatz enthielt, wurden 97 Punkte zu „Grünland“ verändert. Die im bereinigten Datensatz 204 Punkte umfassende Klasse „versiegelte Flächen“ erfuhr eine Umwandlung von 20 Punkten zur Klasse „Acker“.

Die Suche nach den künstlichen Ausreißern wurde im bereinigten wie auch im unbereinigten Datensatz durchgeführt. Im letzteren Fall wurden die künstlichen Ausreißer in den unbereinigten Datensatz eingefügt, um zu untersuchen, wie effektiv die Suche nach Ausreißern in einem originalen Datensatz sein würde. Der Zweck der letzteren Untersuchung war festzustellen, wie die Ausreißersuche in einem Datensatz mit etwa „natürlicher“ Anzahl und Verteilung an Ausreißern funktionieren würde. Bei der „Protokollmethode“ fand die Ausreißersuche in 20 Durchgängen statt, um den Aufwand zu beschränken.

### **3.5 Erstellung des visuellen Kontrolldatensatzes**

Zur Qualitätsbewertung der Klassifizierungen fand die Erstellung eines Datensatzes auf Grundlage visueller Bewertung von Kontrollpunkten statt. Somit gab es eine unabhängige und entsprechend der Interpretation zuverlässige Quelle für die Validierung der Ergebnisse. Dieser Datensatz besteht demgemäß aus Kontrollpunkten, welche stichprobenartig eine Bewertung der Klassifizierung ermöglichen.

Das Ziel bestand darin, einen Kontrolldatensatz mit mindestens 50 Punkten pro Klasse zu erstellen. Die zu interpretierenden Punkte sollten aber nicht willkürlich, sondern in einem Raster ausgewählt werden.

Um beides zu gewährleisten, wurden im CORINE-Datensatz die Klassenflächen innerhalb des Untersuchungsgebietes festgestellt. Die Größen der Flächen geteilt durch 50 ergaben die pro Referenzpunkt repräsentierte Fläche und die Wurzel daraus bestimmte die klassenspezifische Rasterweite.

Zur folgenden Interpretation dienten die ASTER-Bilder der zwei Zeitpunkte sowie im Internet abrufbare Orthofotos (siehe Kapitel 2.3 und 2.4). Das Vorhandensein eines zweiten Zeitpunktes stellte sich dabei insbesondere für das Unterscheiden von Ackerflächen und Grünland sowie von Laub- und Nadelwald als hilfreich heraus.

Punkte, die nicht in homogenen bzw. zuordenbaren Pixeln lagen, wurden in die nächste nordöstlich gelegene „homogene“ Fläche mit erkennbarer Klasse verschoben.

Die folgenden Bilder dienen der Veranschaulichung des visuellen Interpretierens. Die Klasse eines Punktes wird unter Verwendung der ASTER-Daten zweier Zeitpunkte unter Zuhilfenahme eines Orthofotos bestimmt. Die ASTER-Bilder zeigen jeweils mit den Kanälen Grün, Rot und nahes Infrarot denselben Ausschnitt und haben dieselbe Ausdehnung.

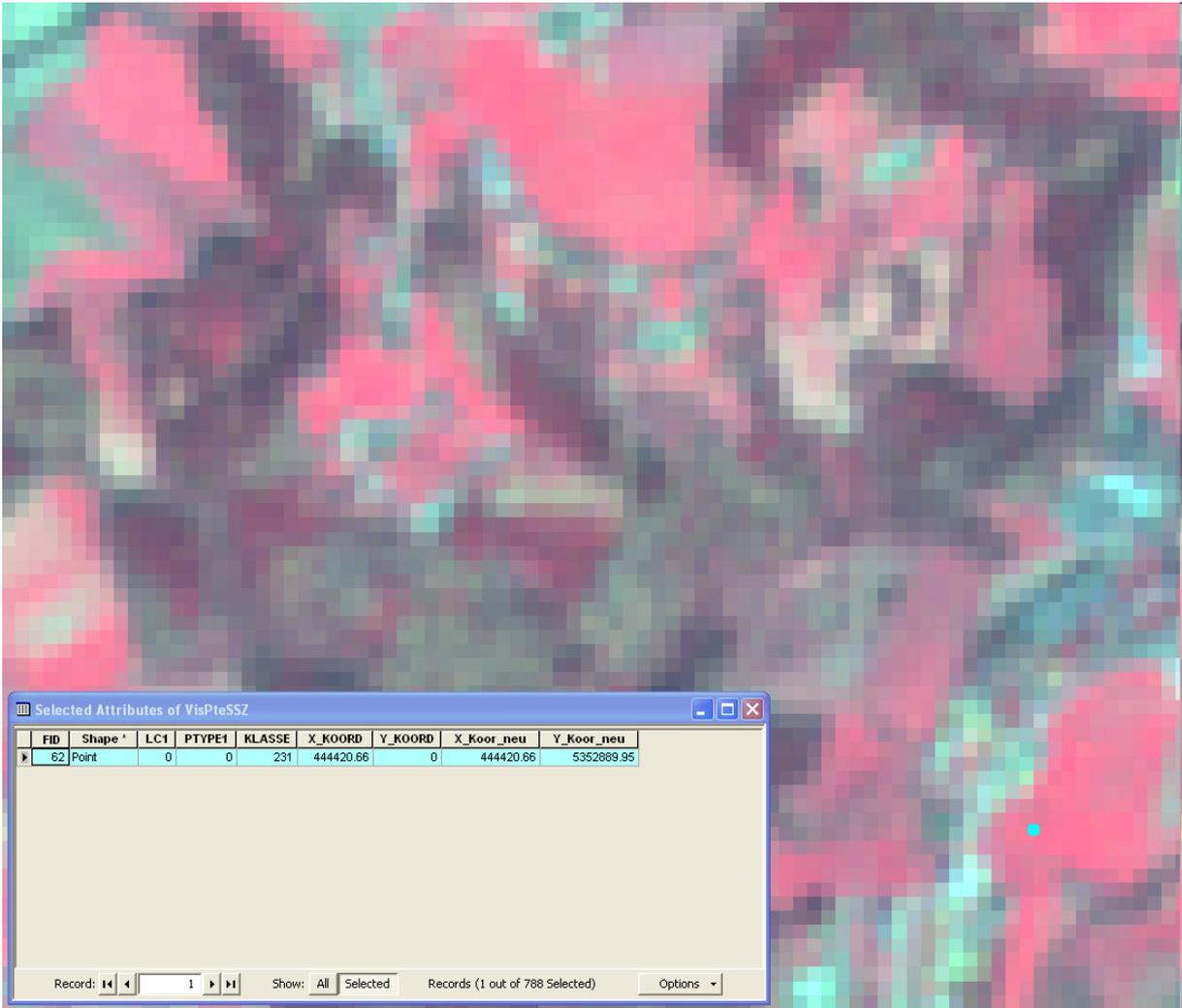


Abbildung 4: ASTER-Ausschnitt der Aprilaufnahme zur Veranschaulichung visuellen Interpretierens

Das obige ASTER-Bild stellt einen Ausschnitt des Aprilzeitpunktes dar. In der Attributbox unten im Bild sieht man die Tabelle mit nur einer Datenzeile, welche Eigenschaften des ausgewählten Punktes anzeigt. Letzteren findet man türkis gefärbt rechts auf gut halber Höhe der Attributbox in einer rotfarbigen Fläche. Die Farbe der Fläche spricht für photosynthetisch aktive Biomasse und die Struktur der Fläche für landwirtschaftliche Nutzung.

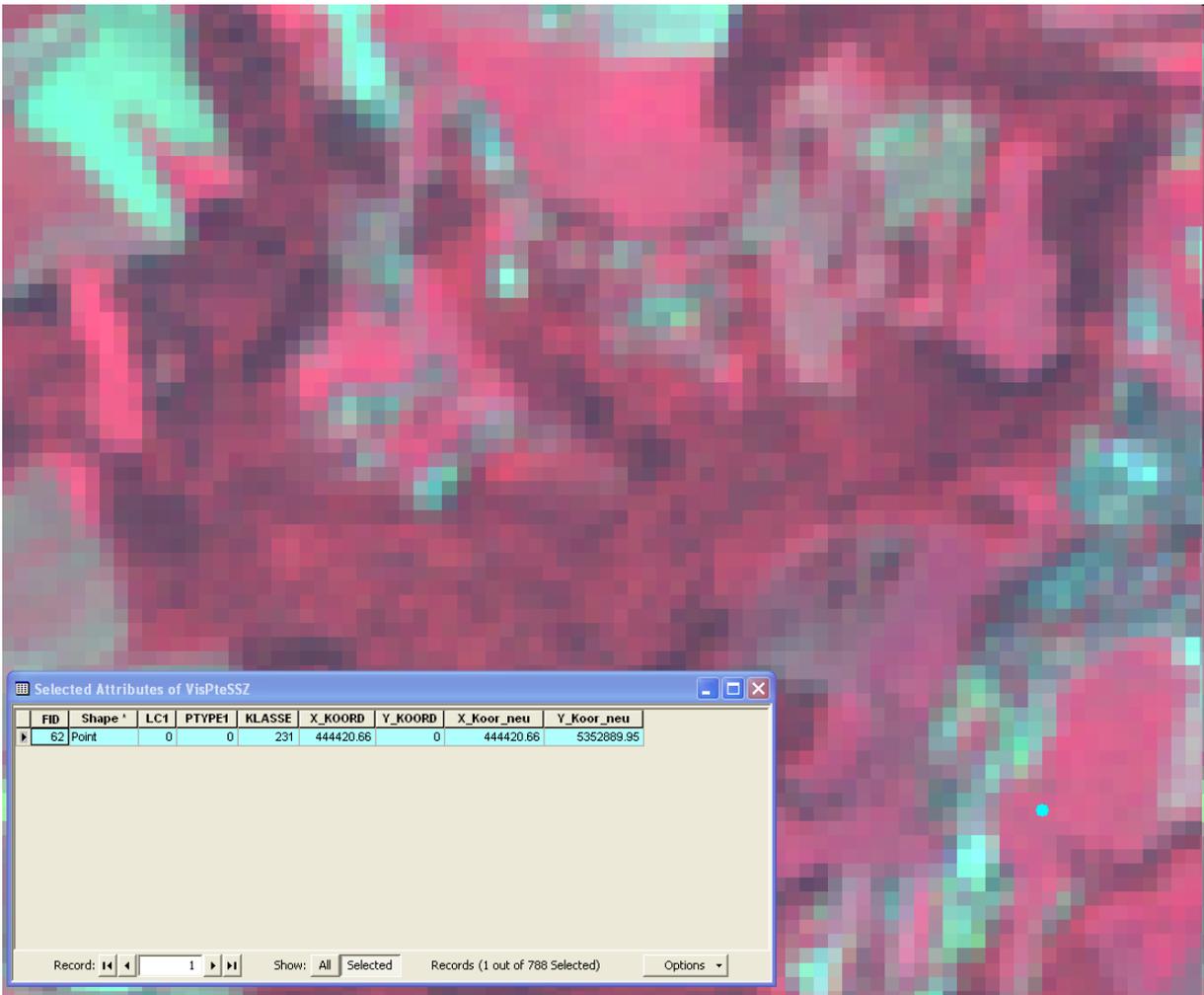


Abbildung 5: ASTER-Ausschnitt der Augustaufnahme zur Veranschaulichung visuellen Interpretierens

In Abbildung 5 ist derselbe Ausschnitt zum Augustzeitpunkt zu sehen. Der an gleicher Stelle befindliche Punkt liegt wieder in einer Fläche, welche die gleiche Form und rötliche Farbe wie der Aprilzeitpunkt aufweist und somit für photosynthetisch aktive Vegetation spricht. Weil zu beiden Zeitpunkten aktive Vegetation vorliegt und aufgrund der Struktur bzw. Homogenität der Fläche von einer landwirtschaftlich genutzten Fläche auszugehen ist, kann man annehmen, dass es sich bei der Fläche, in welcher der Punkt liegt, um Grünland handelt. Um eine Ackerfläche handelt es sich nicht, da sonst zu einem der beiden Zeitpunkte keine Vegetation vorhanden wäre. Zum Beispiel wäre bei Weizen-Anbau die Fläche Ende August vegetationsfrei, bei Mais-Anbau dagegen wäre sie im April vegetationsfrei.



Abbildung 6: Orthofoto zur Veranschaulichung visuellen Interpretierens

Auch die Betrachtung des obigen Orthofotos mit dem in der Mitte des Bildes rot gekennzeichneten Referenzpunkt spricht durch Form- und Farbgebung der Fläche dafür, den Punkt der Klasse Grünland zuzuordnen.

### 3.5.1 Erzeugung eines Referenzdatensatzes aus den visuell interpretierten Daten

Die visuell interpretierten Daten wurden in der Folge nicht nur als Kontrolldatensatz genutzt, sondern in umgewandelter Form auch als Referenzdatensatz verwendet. Die ursprünglich 731 visuell interpretierten Punkte wurden dazu in einem ersten Schritt nach Klassen sortiert. Jene Klassen, die einen geringeren Anteil an den 731 Punkten als den geschätzten Flächenanteil (siehe Kapitel 3.8.2) hatten, wurden vervielfältigt. Die Punkte der Klasse „Acker“ wurden verneunzehnfacht, die Punkte der Klasse „Grünland“ verzwölfacht, jene der Klasse „Laubwald“ vervierfacht und die „Nadelwald“-Punkte verelffacht. Somit enthält der erweiterte Datensatz 4931 Punkte und die Klassenanteile entsprechen in etwa den geschätzten Flächenanteilen. Dies soll gewährleisten, dass seltene Klassen bei der Klassifizierung nicht anhand ihres überproportionalen Punkteanteils zu häufig vorkommen und andere Klassen dominieren. In der Folge wurden den Punkten neue Koordinaten zugewiesen, um Abhängigkeiten von diesen bei der Klassifizierung zu minimieren. Der Datensatz kann allerdings keine aussagekräftigen Ergebnisse bei der Kreuzvalidierung liefern. Dies liegt daran, dass bei den vervielfältigten Klassen die nächsten Nachbarn der zu klassifizierenden Punkte ihre Duplikate sind. Der Datensatz wurde also nur zur Erstellung von Karten verwendet.

## **3.6 Software**

Im Folgenden werden die verwendeten Programme kurz beschrieben.

### **3.6.1 ArcMap**

ArcMap der Firma ESRI ist im Rahmen dieser Arbeit unter anderem für die Erstellung des Rasterdatensatzes aber auch teilweise für das Auslesen spektraler Daten aus den ASTER-Bildern genutzt worden. Die Verbindung der spektralen Daten der Referenzpunkte mit den CORINE-Landnutzungsklassen fand in diesem Programm statt. Zudem wurde es für das manuelle Hinzufügen weiterer Referenzpunkte, für die als „Homogenisierung“ bezeichnete Datenreduktion, für die Visualisierung der ASTER-Bilder und der produzierten Karten (sowie der Legendenerstellung), für die visuelle Interpretation und auch für die Erstellung der „maskfiles“ verwendet.

### **3.6.2 Erdas Imagine**

Dieses Programm diente zur Verbindung verschiedener Kanäle und Aufnahmezeitpunkte einer ASTER-Szene zu einem Bild („layerstack“). Außerdem fand hier die Umwandlung diverser Bildformate in das vom kNN-Programm ausschließlich zu verwendende Format .bip statt. Darüber hinaus wurden in Erdas Imagine diese Bilder visualisiert und spektrale Werte kontrolliert.

### 3.6.3 kNN-Software

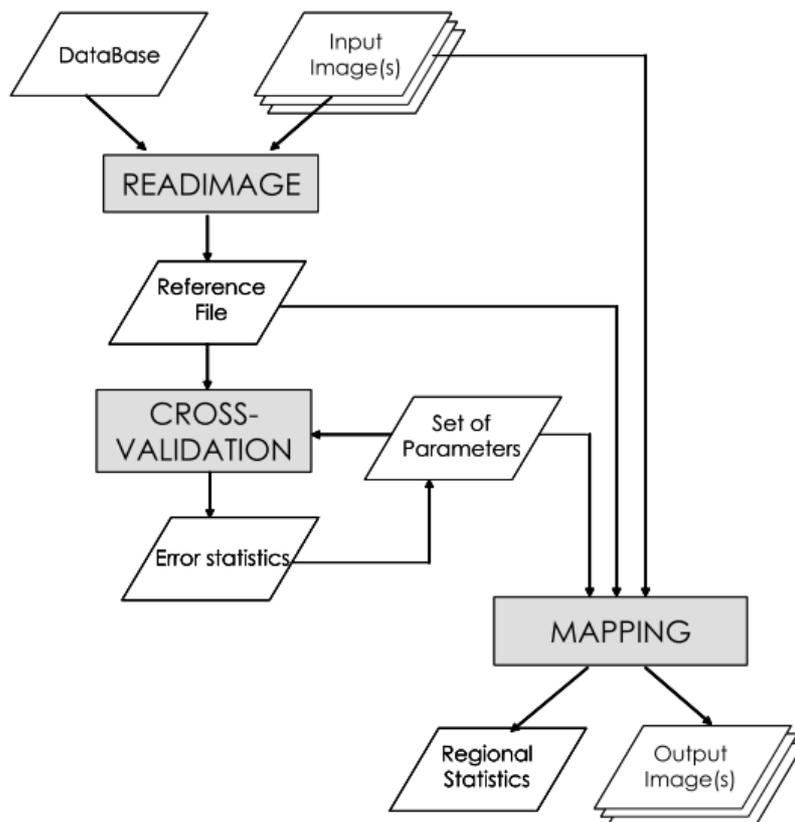


Abbildung 7: Ablaufschema des kNN-Programms (KOUKAL 2009)

Das Schema in Abbildung 7 zeigt die Module des kNN-Programms in Beziehung zu den Daten und ihrer Verarbeitung.

Die kNN-Software besteht aus 3 Modulen.

Das Modul „readimage“ sammelt bzw. vereinigt X- und Y-Variable (siehe Kapitel 3.1), indem es Pixelwerte eines Bildes ausliest und mit der CORINE-Klasse verbindet. In der „DataBase“ werden die Koordinaten der Referenzpunkte und die zugehörigen Klassenangaben, in den „Input Images“ die Fernerkundungsbilder bereitgestellt, die die Spektralwerte liefern. Als Ergebnis wird die kNN-Referenzdatei (Kapitel 3.1), welche in den anderen Modulen zur Anwendung kommt, erzeugt. Dabei können mittels „Masken-Dateien“ bestimmte Bereiche (wie Wolken) vom Auslesen ausgenommen werden.

Das zweite Modul „crossvalidation“ (Kreuzvalidierung) kann für Parameteroptimierungen oder zur Genauigkeitsabschätzung benutzt werden. Zudem beinhaltet es eine Funktion, durch welche zweifelhafte Referenzpunkte (Ausreißer, siehe Kapitel 3.4) erkannt werden können.

Mit dem Modul „mapping“ beinhaltet das Programm die Möglichkeit der Erzeugung von Karten. Dabei werden die Y-Variablen Pixel für Pixel unter Verwendung des Referenzdatensatzes geschätzt und als Bild (Karte) ausgegeben. (KOUKAL 2009)

### 3.7 Variantenvergleiche

Bei der Herstellung und Anwendung eines Referenzdatensatzes bestehen viele mögliche Varianten und Kombinationen an Methoden. Die Feststellung der entsprechenden Unterschiede ist ein Ziel der Arbeit. So gibt es bereits beim Ausgangsdatensatz und der Erstellung des Rasters für den Referenzdatensatz viele Möglichkeiten der Datenbearbeitung. Dazu kommt die mögliche Einbeziehung der „Auffüllpunkte“, also weiterer Punkte in Klassen, welche sonst nicht ausreichend im Referenzdatensatz vorhanden wären. Weiter gibt es die eventuelle Anwendung des „Homogenisierungskriteriums“. Auch der Einfluss verschiedener Methoden der Ausreißer-Eliminierung sowie unterschiedliche Anwendungen dieser Methoden sollten untersucht werden. Ein weiterer Untersuchungspunkt war die Einbeziehung bzw. Kombination verschiedener Kanäle oder Kanalgruppen. Ein besonderes Augenmerk lag auf dem Einfluss der Verwendung eines zweiten Zeitpunktes auf die Klassifizierung.

Insgesamt ergibt sich eine Vielzahl an Kombinationsmöglichkeiten, aus welchen im Rahmen der Arbeit eine begrenzte Anzahl ausgewählt wurde. Ein Schema der Kombinationsmöglichkeiten und möglichen Datenbearbeitungspfade zeigt die folgende Abbildung 8.

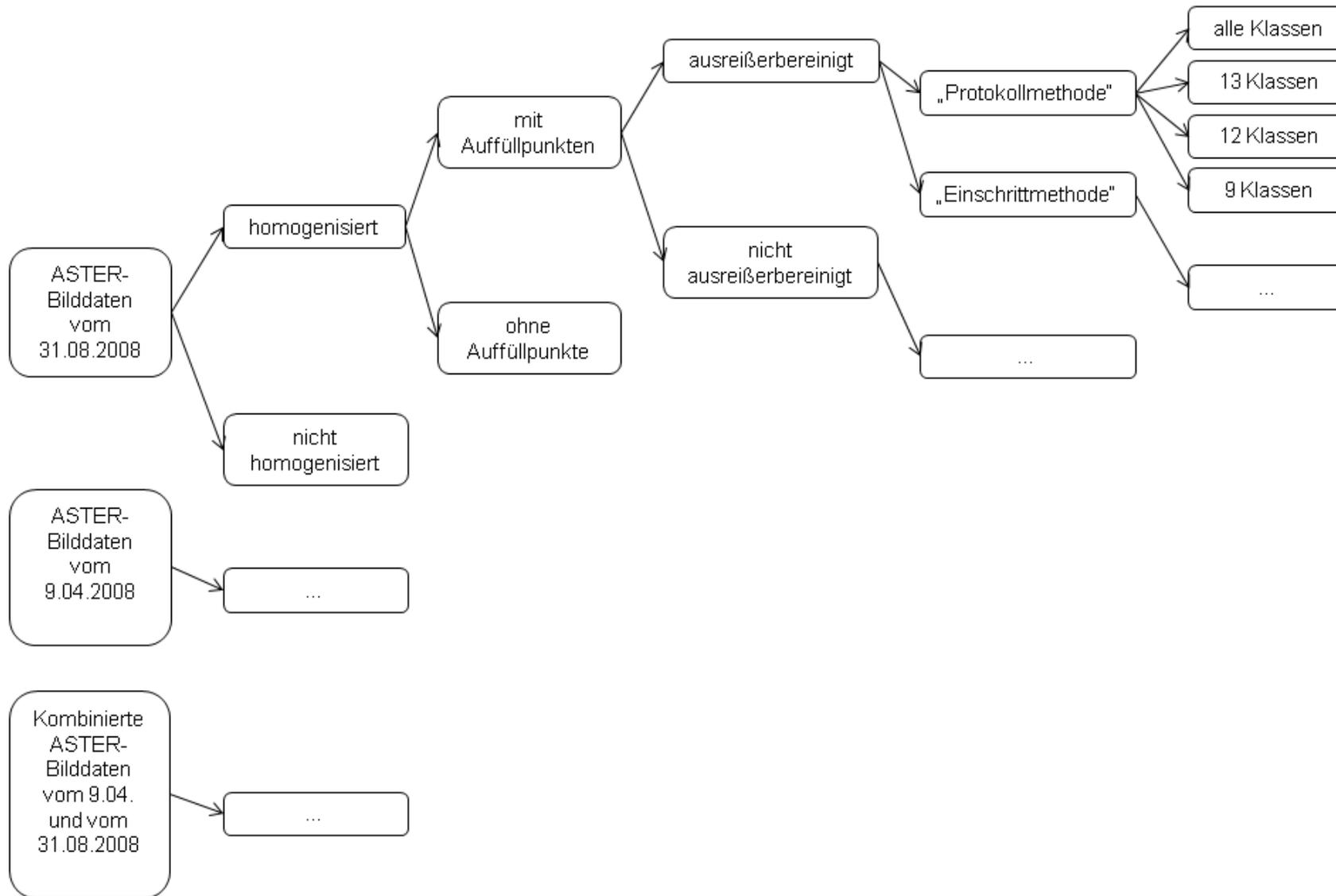


Abbildung 8: Schema der Auswertungsvarianten

Abbildung 8 zeigt die Pfade der Datenbehandlung und die verschiedenen Varianten der Erstellung von Referenzdatensätzen, entsprechend der im Methodikteil beschriebenen Bearbeitung der Daten. Kästchen mit „...“ bedeuten, dass hier weitere Auswertungen nach den entsprechenden weiter oben gezeigten Pfaden stattfinden.

Manche Äste des Schemas wurden bis in die letzte Spalte verfolgt. Andere hingegen wurden nicht weiter untersucht, weil eine andere Variante der Referenzdatenbehandlung deutlich sinnvoller erschien und insbesondere die Genauigkeitswerte für die ausschließliche Verfolgung eines Variantenzweiges sprachen. Beispielsweise ist nach der „Homogenisierung“ des Datensatzes in der Folge nur noch mit diesen Daten gerechnet worden. Auch die aus den Auffüllpunkten entstandenen Daten wurden hinterher stets verwendet, die Varianten ohne diese Vorgehensweise also nicht weiter verfolgt. Dies reduzierte die große Zahl an verschiedenen Behandlungsweisen der Ausgangsdaten, um zielgerichtet den Aufwand möglichst gering zu halten.

### **3.8 Methoden der Resultatsbewertung**

Zur Bewertung der Klassifikationen durch die Kreuzvalidierung sowie durch Vergleich mit der visuellen Interpretation (Kontrollpunkte) wurden Fehlermatrizen erstellt. Die visuelle Einschätzung der Karten als dritte Bewertungsmethode (siehe Kapitel 3.8.3) und der Flächenvergleich (siehe Kapitel 3.8.4) finden ohne sie statt.

Fehlermatrizen sind eine Sonderform von Kreuztabellen, die Soll-Werte und Ist-Werte gegenüberstellen. In dieser Arbeit wurden sie dazu verwendet, einen Überblick über die Genauigkeit der Klassifikationen im Rahmen von Kreuzvalidierungen und im Vergleich mit visuell erstellten Kontrolldaten zu erhalten. Eine Fehlermatrix enthält eine gleiche Anzahl von Spalten und Zeilen, entsprechend der Anzahl der Klassen. Zur Kennzeichnung der Klassen sind in Tabelle 2 die Klassencodes angegeben.

		CORINE-Klasse								Summe	Benutzer- genauigkeit	
		111	131	211	231	311	312	411	511			512
Klassifizierung	111	1381	83	212	96	13	38	5	5	24	1857	0,74
	131	65	73	42	10	6	18	3	1	17	235	0,31
	211	295	96	2549	393	39	70	4	2	7	3455	0,74
	231	182	31	346	1706	83	228	9	2	12	2599	0,66
	311	19	10	35	95	303	125	3		5	595	0,51
	312	64	7	56	310	211	3745	510	5	20	4928	0,76
	411						42	73			115	0,63
	511	3	2	2	3	1			104	2	117	0,89
	512	10	20	2	4	6	7		12	411	472	0,87
Summe	2019	322	3244	2617	662	4273	607	131	498	14373		
Produzenten- genauigkeit	0,68	0,23	0,79	0,65	0,46	0,88	0,12	0,79	0,83			

Kappa 0,64

Gesamtgenauigkeit 0,72

Tabelle 2: Beispiel einer Fehlermatrix

Die Spalten einer Fehlermatrix stellen die Soll-Werte der Klassifikation dar, also im Fall dieser Arbeit die übernommene Klasse aus CORINE oder der visuellen Interpretation. Die Zeilen enthalten die Ist-Werte bzw. hier die durch das kNN-Programm bestimmte Klasse. Ist ein Pixel richtig bzw. übereinstimmend klassifiziert worden, findet es sich in der Diagonalen der Tabelle wieder. Ist- und Soll-Wert stimmen überein.

Die Produzentengenauigkeit oder „producer’s accuracy“ wird folgendermaßen berechnet: In einer Klasse werden Punkte, die richtig klassifiziert sind, durch die Soll- (Referenzdaten-) Gesamtanzahl der Punkte geteilt. Dieser Wert beschreibt, wie gut die tatsächliche Landbedeckung von der Klassifizierung wiedergegeben wird. Wenn also beispielsweise 2019 Referenzpunkte in der Klasse mit dem Code „111“ liegen und 1381 davon richtigerweise durch die Klassifizierung dieser zugeordnet werden, ist die Produzentengenauigkeit dieser Klasse 0,68 bzw. 68% (siehe Tabelle 2).

Für den Wert der Benutzergenauigkeit oder „user’s accuracy“ werden die Punkte pro Klasse, die richtig klassifiziert sind, durch die Anzahl der Punkte, die durch die Klassifizierung dieser Klasse zugeordnet sind, geteilt. Hierdurch wird beschrieben, wie gut eine Karte bzw. Klassifizierung die Wirklichkeit wiedergibt, also wie zuverlässig die Karte ist. Wenn z.B. 1857 Referenzpunkte als Klasse mit dem Code „111“ klassifiziert werden und davon 1381 tatsächlich dieser angehören, ist die Benutzergenauigkeit 0,74 bzw. 74% (siehe Tabelle 2).

Die Gesamtgenauigkeit („overall accuracy“) gibt an, wie viele aller Elemente (Pixel) richtig klassifiziert wurden. Sie wird durch die Summe der Werte in der Tabellendiagonalen durch die Gesamtsumme aller Punkte bestimmt.

Ein weiterer Kennwert zur Charakterisierung der Genauigkeit einer Klassifizierung ist der Wert Kappa, der von der Gesamtgenauigkeit noch die Zufallsübereinstimmungen abzieht.

### **3.8.1 Kreuzvalidierung der kNN-Referenzdatensätze**

Die Kreuzvalidierung ist ein Verfahren zur Prüfung der Übertragbarkeit einer statistischen Analyse, die an Hand eines Datensatzes entwickelt wurde, auf einen unabhängigen Datensatz. Der Datensatz wird dazu in zwei Teilmengen eingeteilt. Die „Trainingsdaten“ werden zur Bestimmung der Parameter der „Testdaten“ verwendet und mit deren Parametern verglichen.

Bei der „crossvalidation“ in Bezug auf kNN werden innerhalb eines Referenzdatensatzes Punkte klassifiziert und mit ihrer bekannten Klasse verglichen. Dabei können ein einzelner Punkt oder auch eine Gruppe von Referenzpunkten durch den Rest der Referenzpunkte klassifiziert werden.

Im Fall dieser Arbeit ist nur die „leave-one-out“-Methode angewandt worden. Dabei wird immer ein einzelner Referenzpunkt als „Test“ verwendet und durch die restlichen klassifiziert. Dies wird für jeden der Punkte durchgeführt. (KOUKAL 2009)

### **3.8.2 Vergleich der kNN-Klassifikation mit der visueller Interpretation**

Der visuell erstellte Kontrolldatensatz beinhaltet die visuell zugeteilte Klasse, die Koordinaten sowie die spektralen Informationen. Um mit der visuellen Interpretation die kNN-Klassifikation zu bewerten, vergleicht man die mit Hilfe des kNN-Referenzdatensatzes ermittelte Klasse der Punkte des Kontrolldatensatzes mit der visuell interpretierten Klasse.

Demgemäß erfolgt eine Klassifizierung der Punkte des visuell interpretierten Datensatzes anhand ihrer spektralen Werte, wobei der kNN-Referenzdatensatz verwendet wird. Somit kann man die jeweils berechnete Klasse mit der visuell bestimmten Klasse vergleichen und Fehlermatrizen erstellen.

Bei der visuellen Interpretation sind die Punkte nicht gleichmäßig über die ganze relevante Fläche, sondern je zu 50 über die einzelnen Flächen der CORINE-Klassen verteilt. Somit ist die Auswertung nicht flächenrepräsentativ und Aussagen über Gesamtgenauigkeit und Benutzergenauigkeit können nicht getroffen werden.

Um dies zu beheben, wurden die Daten in den Fehlermatrizen in Anlehnung an die Flächenanteile der verschiedenen Klassen verändert. Die Anteile stammen aus einer Klassifizierung mit als sinnvoll beurteilten oder korrigierten Klassenflächen. Als Grundlage für die Flächenanteile diente eine Klassifikation mit  $k=13$  und 9 Klassen, deren Referenzdatensatz vorher mit der „Einschrittmethode“ (bei einer Kreuzvalidierung mit  $k=13$ ) ausreißerbereinigt wurde. Der Siedlungsanteil dieser Klassifizierung erschien visuell nicht sinnvoll. Vor allem Flächen, die tatsächlich Acker sind, wurden als „Siedlung“ klassifiziert. Daher erfolgte eine Übernahme des Siedlungsanteils aus einer Klassifikation mit  $k=13$  und allen 20 im Untersuchungsgebiet vorkommenden Klassen. Der zu Grunde liegende Referenzdatensatz war vorher durch die „Einschrittmethode“ (im Rahmen einer Kreuzvalidierung mit  $k=12$ ) ausreißerbereinigt worden. Die Siedlungsfläche erscheint in dieser Klassifizierung mit allen Klassen nämlich zuverlässiger als in der Klassifikation mit 9 Klassen. Die Differenz zum ursprünglichen Siedlungsanteil (aus der Klassifikation mit 9 Klassen) wurde der Klasse „Acker“ zugeteilt. Es fand auch eine Korrektur der

Anteile von Laub- und Nadelwald statt. Dazu wurden deren Anteile in der Klassifizierung mit 20 Klassen bestimmt, wobei dort der Mischwaldanteil zu je 50% Laub- und Nadelwald zugeordnet worden war. Die prozentualen Waldanteile von Laub- und Nadelwald fanden dann Anwendung auf die Gesamtwaldfläche der Klassifizierung, um wiederum die Flächenanteile von Laub- und Nadelwald zu berechnen.

Die Flächenanteile wurden dann auf die Fehlermatrix und besonders deren Spalten angewendet. Eine Spalte der Fehlermatrix bezieht sich auf eine bestimmte Klasse und trägt deren Namen. Sie stellt dar, wo sich die visuell interpretierten Punkte einer Spalte bzw. Klasse in der Klassifizierung wiederfinden.

Zur Anpassung der Tabelleninhalte an die Flächenproportionen wurden einzelne Felder der Kreuztabelle durch die Summe einer Spalte geteilt. Die Summe einer Spalte ist auch die Gesamtzahl der Punkte, die visuell einer Klasse zugeordnet worden sind. Durch das Teilen durch die Spaltensumme fand eine Normierung der Spaltensummen auf den Wert 1 statt. Weiter wurden die Felder der Spalte mit dem Flächenanteil der entsprechenden Klasse multipliziert, so dass die Spaltensummen den Flächenanteilen entsprachen. In der Folge wurden die Felder jeweils mit der ursprünglichen Gesamtanzahl visuell interpretierter Punkte multipliziert, um wieder auf die gleiche Gesamtanzahl zu kommen. Wenn man nun die Gesamtzahl der Punkte einer Spalte (also die Gesamtzahl der visuell einer Klasse zugeordneten und durch Flächenanteile korrigierten Punkte) durch die Gesamtanzahl aller Punkte teilt, erhält man den Flächenanteil der jeweiligen Klasse bzw. Spalte. Somit sind die Werte der Spalten flächenproportional und Genauigkeitswerte wie beispielsweise „Producer’s accuracy“ oder „Overall Accuracy“ können abgeschätzt werden.

### **3.8.3 Visuelle Bewertung der produzierten Karten**

Die Einschätzung der Karten- und Klassifizierungsqualität wird auch durch visuelle Beurteilung vorgenommen. Der „gesunde Menschenverstand“ taxiert, ob die produzierte Karte sinnvoll oder besser als eine andere erscheint. So erkennt das menschliche Gehirn Strukturen und Ausreißer aus diesen. Linien wie Straßen werden gut erkannt und auch logische Betrachtungen angestellt, die mit Hilfe des Computers nur schwer durchführbar sind. Beispielsweise erkennt die verstandesgeleitete menschliche Wahrnehmung ein Siedlungspixel innerhalb einer Waldfläche direkt als unwahrscheinlich. Wollte man für eine derartige Regel mit entsprechenden Ausnahmen Automatismen verwenden, müsste sehr großer Aufwand betrieben werden.

### **3.8.4 Vergleich von Flächenanteilen**

Zur weiteren Beurteilung der Qualität von CORINE und den produzierten Karten wurden deren Flächenanteile zuverlässigen Quellen gegenübergestellt. Diese stammten aus InVeKoS und der „Waldkarte“ in Oberösterreich und können als korrekt und zuverlässig angesehen werden. InVeKoS bedeutet Integriertes Verwaltungs- und Kontrollsystem. Es ist ein EU-Programm, das auch ein

„System zur Identifizierung landwirtschaftlich genutzter Parzellen“ beinhaltet (<http://www.invekos.at> - besucht am 15.04.2010). Die vom Bundesamt für Wald entwickelte „Waldkarte“ liefert österreichweit Informationen, die durch die Kombination von Fernerkundung und terrestrischen Waldinventurdaten gewonnen werden (BAUERHANSL 2009).

Die Flächen beim Vergleich der Klassifikationen mit CORINE beziehen sich auf jene Gemeinden, die vollständig in den in der Arbeit verwendeten Szenen liegen. Das Feststellen der Flächenanteile fand jeweils in ArcMap statt.

## **4 Ergebnisse und Diskussion**

Dieses Kapitel beinhaltet nur einen Teil der während der Arbeit produzierten Ergebnisse. Es sind im Folgenden nur brauchbare und erwähnenswerte Resultate aufgeführt, weshalb z.B. Klassifizierungen von Referenzdaten mit thermalen Kanälen, die insgesamt nicht verwendbar waren, nicht weiter beschrieben werden.

Kapitel 4.1 zeigt die Gegenüberstellung der beiden Ausreißerbereinigungsmethoden bei der Suche nach künstlichen Ausreißern. In den Kapiteln 4.2 bis 4.4 werden die Bewertungen der diversen Klassifizierungen dargestellt. Auch Kapitel 4.5 dient zur Beurteilung der Klassifizierungen, indem es deren Flächenbilanzen CORINE- sowie „Waldkarten“- und InVeKoS-Daten gegenüberstellt.

### **4.1 Identifikation und Eliminierung künstlicher Ausreißer**

Das Kapitel zeigt Unterschiede der Funktionsweisen der verschiedenen Ausreißerbereinigungsmethoden. Dies wird durch das Verfolgen der künstlich in die Datensätze eingebrachten Ausreißer ermöglicht (siehe Kapitel 3.4.2).

#### **4.1.1 Ergebnisse der „Einschrittmethode“**

Im ausreißerbereinigten Datensatz erzielte die „Einschrittmethode“ durch eine Klassifizierung mit  $k = 4$  die Identifizierung von 116 der 117, also von 99,15% der künstlichen Ausreißer. Dies bezieht sich auf die veränderten Punkte beider betroffenen Klassen. Zudem wurden mit der Methode im bereits bereinigten Datensatz 4,47% der Referenzpunkte „fälschlicherweise“ eliminiert. Im unbereinigten Datensatz wurden mit der Methode alle künstlichen Ausreißer gefunden.

#### 4.1.2 Ergebnisse der „Protokollmethode“ nach 20 Runden

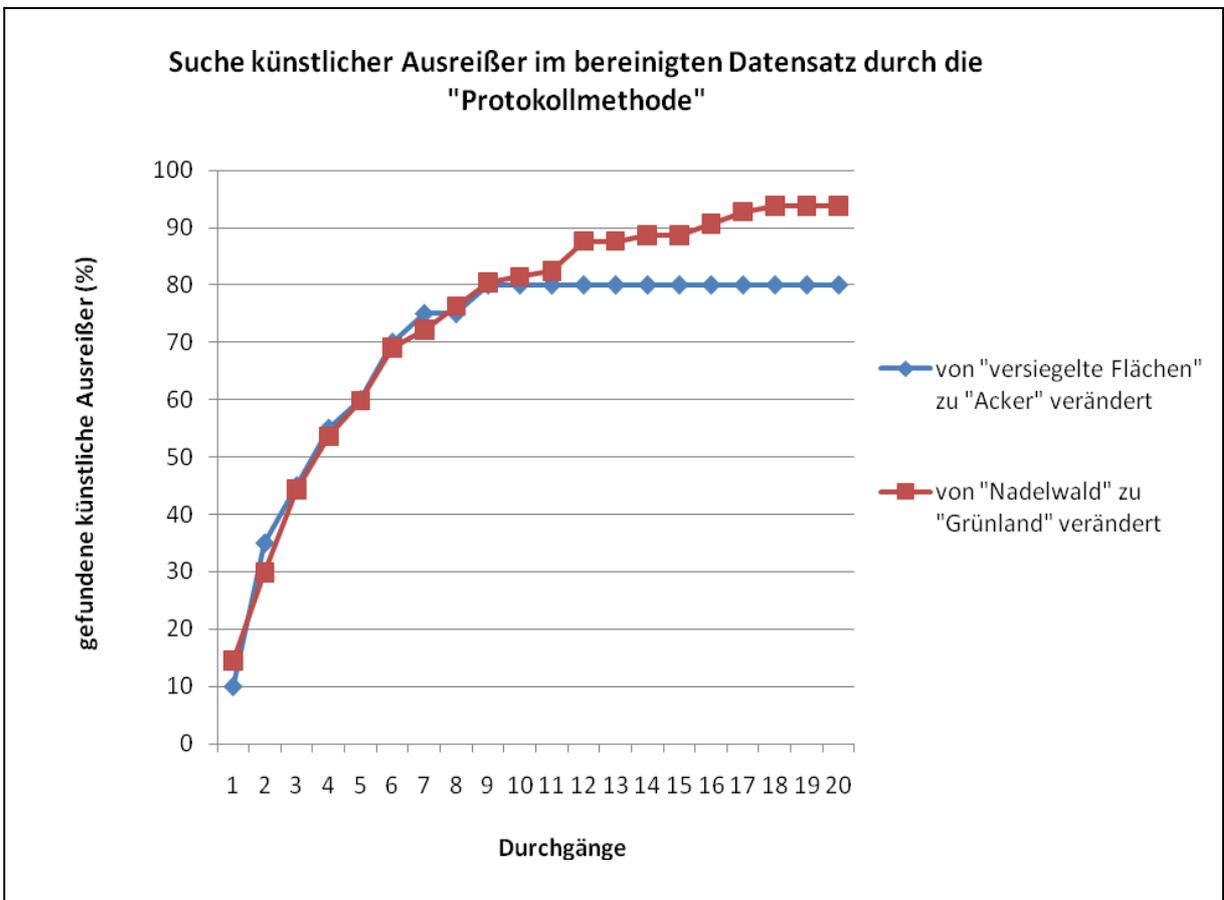


Abbildung 9: Suche künstlicher Ausreißer im bereinigten Datensatz mit der „Protokollmethode“

In der Abbildung ist dargestellt, wie viele künstliche Ausreißer im bereinigten Datensatz in Abhängigkeit von den Durchgängen der Ausreißersuche gefunden werden. Die Kurven zeigen die getrennte Betrachtung der in den zwei Klassen veränderten Punkte. Bis zum zehnten Suchdurchgang gleichen sich die Kurven, woraufhin die Kurve der gefundenen künstlichen Ausreißer, die von „versiegelte Flächen“ zu „Acker“ verändert wurden, stagniert und die entsprechende Kurve der von „Nadelwald“ zu „Grünland“ veränderten Punkte weiter leicht ansteigt. Der weniger stetige Kurvenverlauf der gefundenen künstlichen Ausreißer von „versiegelte Flächen zu Acker“ liegt daran, dass nur 20 Punkte von „versiegelte Flächen“ zu künstlichen Ausreißern der Klasse „Acker“ verändert wurden. Nach 20 Runden wurden 80% der von „versiegelte Flächen“ zu Acker und 93,8% der von „Nadelwald“ zu „Grünland“ veränderten Punkte gefunden. Es ist anzunehmen, dass durch weitere Durchgänge mehr künstliche Ausreißer gefunden würden.

Insgesamt wurden mit der „Protokollmethode“ im bereinigten Datensatz nach 20 Runden Ausreißersuche 91,5% der künstlichen Ausreißer beider betroffener Klassen gefunden. Im gleichen Zug werden 4,19% der Punkte in den Ausreißersuchdurchgängen 1 bis 20 „fälschlich“ eliminiert.

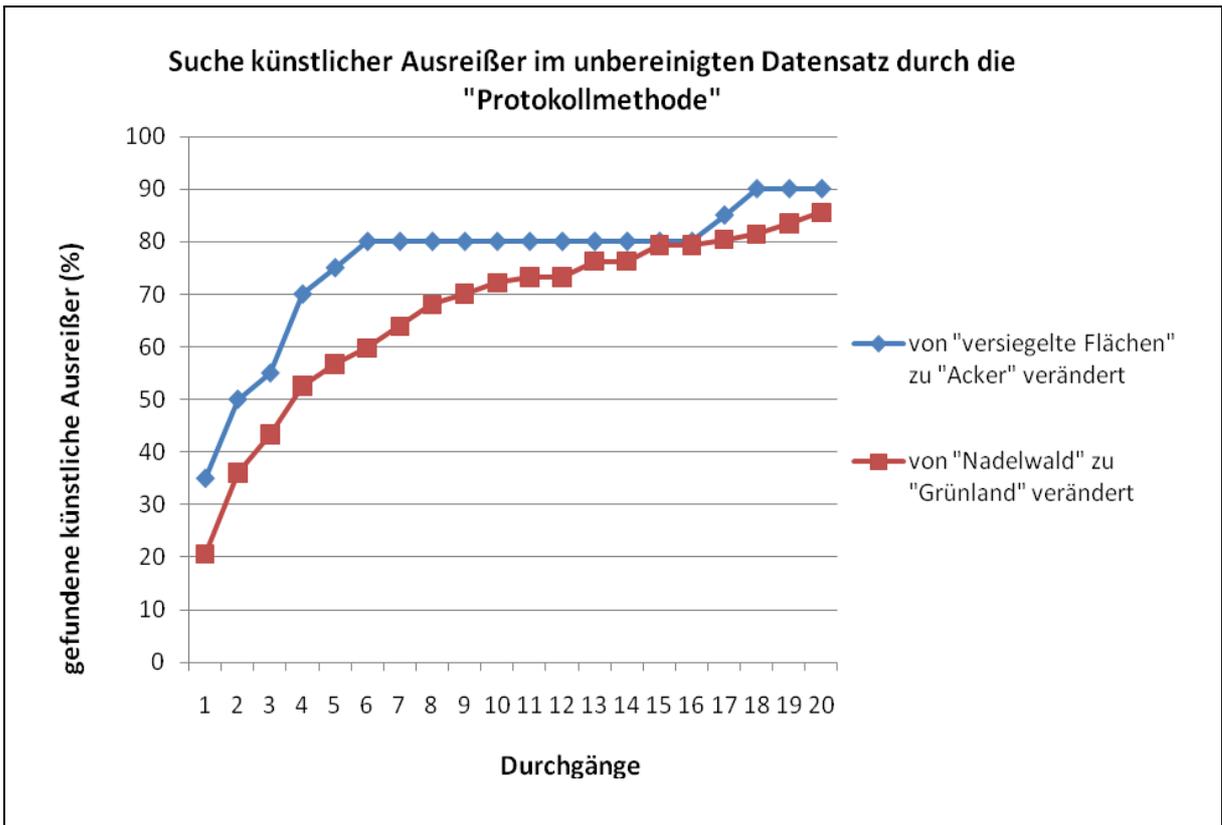


Abbildung 10: Suche künstlicher Ausreißer im unbereinigten Datensatz mit der „Protokollmethode“

Abbildung 10 zeigt den Anteil an gefundenen künstlichen Ausreißern in Abhängigkeit von Suchdurchgängen im unbereinigten Datensatz. Im Vergleich zum bereinigten Datensatz werden hier im ersten Durchgang mehr Ausreißer gefunden. Wie auch in Abbildung 9 ist die Kurve der gefundenen Ausreißer von „versiegelte Flächen zu Acker“ nicht so gleichmäßig geschwungen, was bedeutet, dass Ausreißer nicht stetig sondern eher sprunghaft eliminiert werden, was wieder an der geringen Zahl künstlicher Ausreißer von „Stadt zu Acker“ liegt. Während im bereinigten Datensatz mehr von „Nadelwald“ zu „Grünland“ veränderte Punkte gefunden wurden, sind es hier die von „versiegelte Flächen“ zu „Acker“ veränderten Punkte, die besser gefunden werden. Letztendlich konnten nach 20 Runden 90% der von „versiegelte Flächen“ zu „Acker“ und 85,6% der von „Nadelwald“ zu „Grünland“ veränderten Referenzpunkte gefunden werden. Im Hinblick auf beide Klassen sind im unbereinigten Datensatz 86,3% der künstlichen Ausreißer gefunden worden. Insgesamt ist zu sehen, dass bei der „Protokollmethode“ zu Anfang zwar deutlich mehr Ausreißer identifiziert werden, die Suche aber auch bei den späteren Suchdurchläufen noch erfolgreich verläuft.

Wenn man die beiden Methoden vergleicht, ist die „Einschrittmethode“ bezüglich der Anzahl gefundener künstlicher Ausreißer überlegen, was jedoch wahrscheinlich mittels weiterer Suchdurchgänge bei der „Protokollmethode“ ausgeglichen werden könnte. Beide Methoden eliminieren fälschlicherweise mit je gut 4% etwa gleich viele „gute“ Referenzpunkte.

## 4.2 Kreuzvalidierung von kNN-Referenzdatensätzen

Zunächst wurden Referenzdatensätze, die ohne und mit „Homogenisierung“ (siehe Kapitel 3.2.3) hergestellt wurden, vor der Ausreißereliminierung verglichen. Der Vergleich der Kreuzvalidierung mit und ohne Homogenisierung jeweils vor der Ausreißereliminierung ergab bessere Ergebnisse des homogenisierten Datensatzes. Dabei waren die homogenisierten Referenzpunkte um ca. 2% in Bezug auf Gesamtgenauigkeit sowie Kappa besser. Daher wurde in der Folge nur mit homogenisierten Daten weitergearbeitet, zumal mit den knapp 21700 Punkten immer noch eine große Anzahl an Referenzpunkten vorlag.

Die Kreuzvalidierungen von kNN-Referenzdatensätzen mit den ursprünglich 20 Corine-Klassen vor der Ausreißereliminierung erreichen mit unter 50% keine hohen Gesamtgenauigkeiten. Wenn jedoch die Anzahl der Klassen von 20 auf 12 Klassen reduziert wird, liegen die Gesamtgenauigkeiten etwa um 20% über dem Schnitt der anderen Datensätze. Dieses Ergebnis ist zu erwarten, da weniger Klassen vorliegen, in die ein zu klassifizierender Punkt fälschlicherweise eingeordnet werden kann. (Im Extrembeispiel eines Datensatzes, der nur aus einer Klasse besteht, läge die Gesamtgenauigkeit automatisch bei 100%).

Die Gesamtgenauigkeiten ausreißerbereinigter Datensätze liegen im Vergleich deutlich höher. Bei jenen Referenzdatensätzen mit allen Klassen liegt sie über 80% und bei jenen mit 12 Klassen bei bis zu 98%.

Die folgende Abbildung zeigt anhand von Gesamtgenauigkeiten aus Kreuzvalidierungen Unterschiede zwischen verschiedenen erstellten Referenzdatensätzen mit jeweils reduzierter Klassenanzahl in Wechselwirkung mit der Auswahl des k-Wertes.

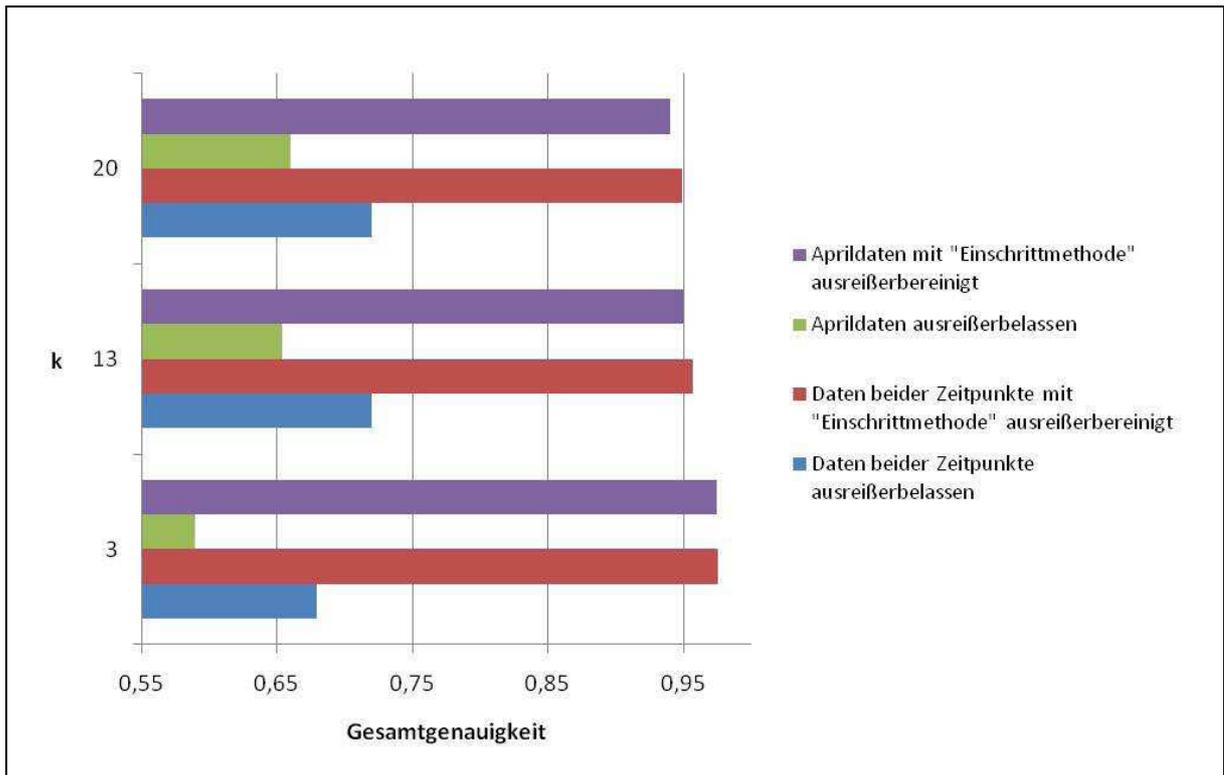


Abbildung 11: Vergleich von Klassifizierungen von mono- und bisaisonalen Datensätzen mit und ohne Ausreißerbereinigung durch Gesamtgenauigkeiten der Kreuzvalidierungen

Abbildung 11 stellt die Wirkung des k-Wertes, der Ausreißerelimination und der Wahl eines mono- oder bisaisonalen Fernerkundungsdatensatzes auf die Gesamtgenauigkeit dar. Die Ergebnisse stammen aus Kreuzvalidierungen von Referenzdatensätzen mit 13 Klassen, die nach der Klassifizierung in der Fehlermatrix zu 9 Klassen zusammengefasst wurden. Aus den Klassen 111, 112, 121, 122 ist die alle versiegelten Flächen umfassende Klasse 111 erstellt worden und aus den Klassen 321 und 231 die Klasse „Grünland“ entstanden (siehe Tabelle 1). Die Ausreißerbereinigungen erfolgten jeweils mit der „Einschrittmethode“ (durch eine Klassifikation mit  $k = 13$ ).

Generell ist festzustellen, dass ausreißerbereinigte Datensätze erheblich bessere Gesamtgenauigkeiten als ausreißerbelassene Datensätze erzielen. Im Schnitt sind dies um etwa 30% höhere Gesamtgenauigkeiten bei vergleichbaren Datensätzen. Auffällig ist auch eine steigende Gesamtgenauigkeit mit von 3 über 13 zu 20 steigendem k bei ausreißerbelassenen Referenzdatensätzen gegenüber einer fallenden Gesamtgenauigkeit bei entsprechend ausreißerbereinigten Daten. Die Wahl eines großen k-Wertes bei Daten mit hohem Ausreißeranteil wirkt den negativen Effekten der Ausreißer entgegen, bei hochwertigen Referenzdatensätzen überwiegt dagegen die bei großem k gegebene Verringerung der Genauigkeit durch Unterdrücken kleiner Klassen. Allerdings zeigen andere Referenzdaten-Bewertungsmethoden, dass dieses Ergebnis nicht verallgemeinerbar ist (siehe Diskussion der verschiedenen Bewertungsmethoden in Kapitel 4.6).

Darüber hinaus sind Unterschiede zwischen mono- und bisaisonalen Datensätzen markant. Bei den ausreißerbelassenen Datensätzen erzielt der bisaisonale Datensatz deutlich bessere Gesamtgenauigkeiten als der monosaisonale April-Datensatz. Beim Vergleich der ausreißerbereinigten

Daten wiederum gibt es kaum Unterschiede zwischen dem bisaisonalen Datensatz und dem Aprildatensatz.

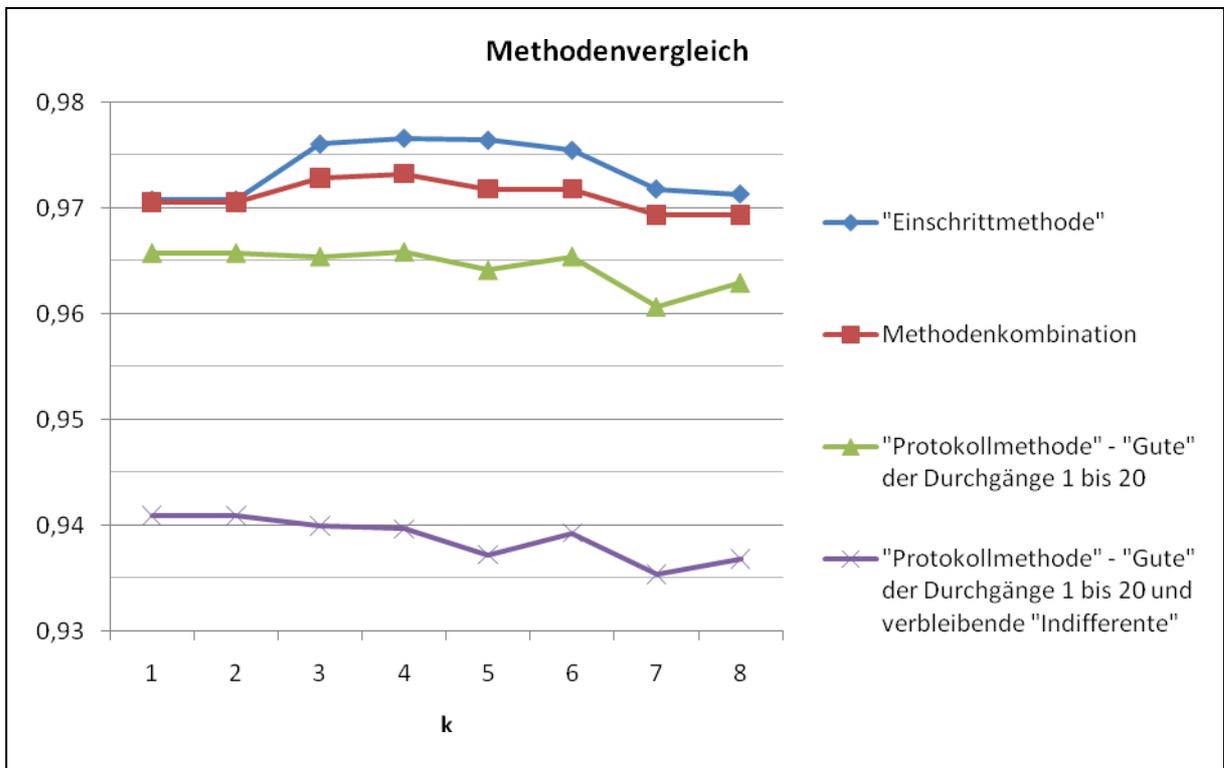


Abbildung 12: Methodenvergleich der Ausreißerbereinigung anhand von Gesamtgenauigkeiten aus Kreuzvalidierungen bisaisionaler Referenzdaten

Abbildung 12 zeigt vergleichend die Ergebnisse verschiedener Methoden der Ausreißerbereinigung. Die Datensätze enthalten jeweils neun Klassen. Am schlechtesten schneidet hier die „Protokollmethode“ mit 20 Runden der Ausreißersuche und in den Datensatz eingefügten „indifferenten“ Referenzpunkten der zwanzigsten Runde ab, die eine maximale Gesamtgenauigkeit von gut 94 Prozent erreicht. Am besten schneidet die „Einschrittmethode“ ab (bei der die Ausreißerbereinigung durch eine Klassifikation mit  $k = 11$  erfolgte). Ihre Gesamtgenauigkeit liegt bei maximal knapp 98%. Ein wenig schlechter schneidet die „Methodenkombination“ ab. Der Referenzdatensatz besteht hier aus den „Guten“ (siehe Kapitel 3.4.1.1) von einer Runde „Protokollmethode“ und den verbleibenden „Indifferenten“ abzüglich der Ausreißer, die mit der „Einschrittmethode“ (im Rahmen einer Klassifizierung durch  $k = 11$ ) entfernt wurden. Bezüglich der Gesamtgenauigkeit dazwischen liegt die „Protokollmethode“ mit einem Referenzdatensatz, der nur aus den „Guten“ von 20 Runden Ausreißersuche besteht und 96,5% Gesamtgenauigkeit erreicht. Es ist hier insgesamt keine eindeutige Überlegenheit einer Ausreißerbereinigungsmethode zu erkennen.

### 4.3 Genauigkeit der Klassifizierung im Vergleich mit der visuellen Interpretation

Aufnahmezeitpunkt	ausreißerbereinigt		ausreißerbelassen	
	Gesamtgenauigkeit	Kappa	Gesamtgenauigkeit	Kappa
April 2008	0,772	0,682	0,778	0,692
August 2008	0,799	0,723	0,796	0,720
April + August 2008	0,893	0,852	0,876	0,831

Tabelle 3: Genauigkeitswerte der Klassifizierung aus der Gegenüberstellung mit der visuellen Interpretation

Tabelle 3 zeigt die Gegenüberstellung diverser Klassifizierungen mit den visuell interpretierten Kontrollpunkten. Sie soll Unterschiede in Bezug auf Aufnahmezeitpunkt und Ausreißerbereinigung darstellen. Die Ausreißeridentifizierung und -eliminierung wurde jeweils im Zuge der „Einschrittmethode“ (durch eine Klassifizierung mit  $k = 13$ ) vorgenommen.

In Bezug auf die ausreißerbelassenen Daten zeigt sich eine Überlegenheit der Kombination aus April- und Augustzeitpunkt ( $Kappa = 0,852$ ), die eine deutlich bessere Gesamtgenauigkeit und einen höheren Kappa-Wert erreicht als der Augustzeitpunkt ( $Kappa = 0,723$ ), der wiederum dem Aprilzeitpunkt ( $Kappa = 0,682$ ) überlegen ist. Des Weiteren sind beim bisaisonalen Datensatz die ausreißerbereinigten den ausreißerbelassenen Daten überlegen, was beim Augustzeitpunkt nicht sehr deutlich ausfällt. Die Aprilwerte weisen sogar eine leichte Überlegenheit der ausreißerbelassenen Daten auf. Insgesamt zeigen die Daten einen positiven Einfluss der Ausreißerbereinigung und insbesondere der Nutzung eines zweiten Zeitpunktes. Der ausreißerbereinigte, bisaisonale Datensatz erzielt die besten Ergebnisse.

Im folgenden Abschnitt wird der Einfluss der Ausreißerbereinigung und der Wahl des  $k$ -Wertes auf die Klassifizierungsgenauigkeit diskutiert.

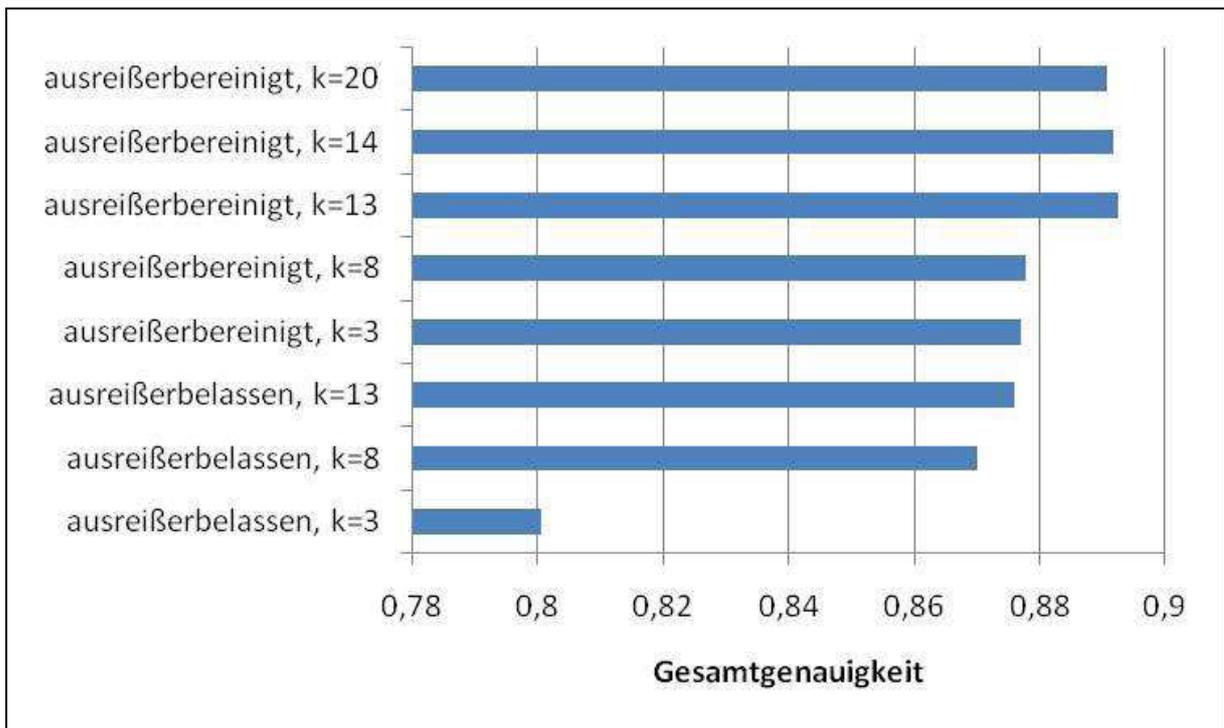


Abbildung 13: Gesamtgenauigkeit verschiedener Klassifikationen auf Grundlage bisaisoaler (teils ausreißerbelassener, teils mit der „Einschrittmethode“ ausreißerbereinigter) Fernerkundungsdaten im Vergleich mit der visuellen Interpretation

In Abbildung 13 ist ein Vergleich der Genauigkeiten von Klassifizierungen auf Grundlage bisaisoaler Referenzdatensätze zu sehen. Die Klassifizierungen geschahen mit auf 13 Klassen reduzierten Datensätzen. Im Anschluss fand eine Zusammenfassung verschiedener Klassen statt, so dass 9 Klassen mit der visuellen Interpretation verglichen werden konnten (siehe Tabelle 1). Vier Klassen wurden zur Klasse „versiegelte Flächen“ zusammengefasst und aus den Klassen „Wiesen und Weiden“ sowie „natürliches Grasland“ wurde „Grünland“. Die Ausreißerbereinigung fand jeweils durch die „Einschrittmethode“ (mittels einer Klassifikation mit  $k = 13$ ) statt.

Die ausreißerbereinigten Datensätze sind den ausreißerbelassenen grundsätzlich überlegen. Des Weiteren ist die Wahl höherer K-Werte von Vorteil für die Gesamtgenauigkeit. Allerdings ist mit dem durch  $k = 13$  klassifizierten Datensatz das Maximum bei guten 89 Prozent Gesamtgenauigkeit erreicht und die Klassifikationen mit  $k = 14$  und  $k = 20$  fallen im Vergleich ab.

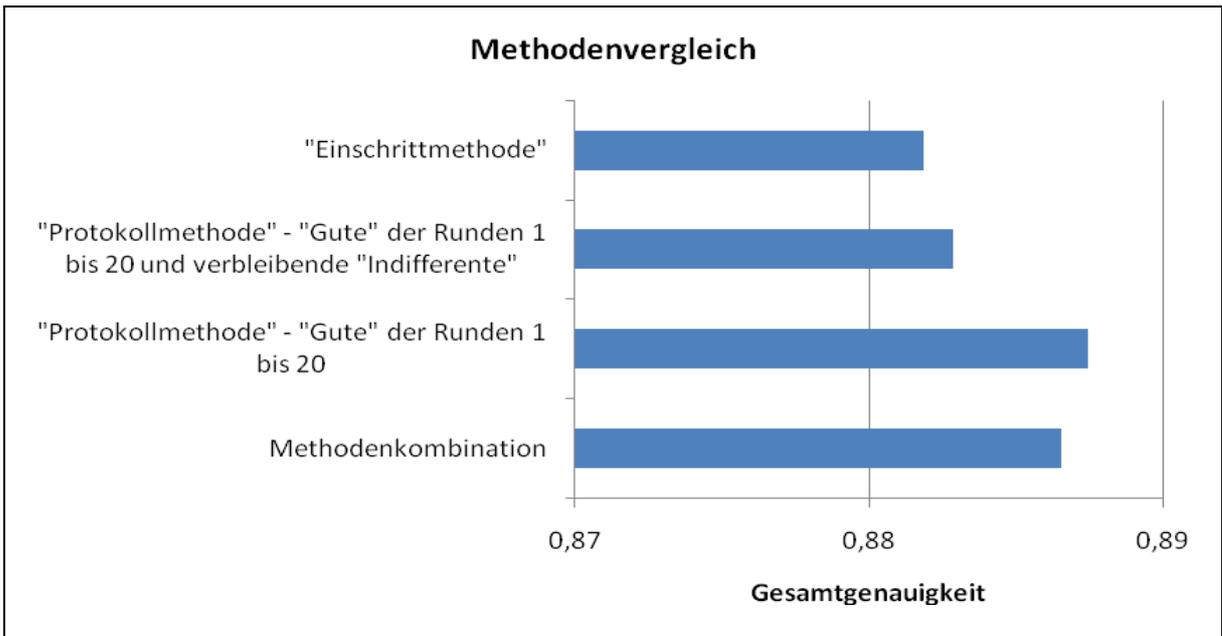


Abbildung 14: Vergleich von Ausreißerbereinigungsmethoden anhand der Gegenüberstellung von visueller Interpretation mit verschiedenen je mit  $k = 9$  klassifizierten, bisaisonalen und ausreißerbereinigten Datensätzen

Zur Bewertung der verschiedenen Ausreißerbereinigungsmethoden fand eine Gegenüberstellung von Klassifizierungen verschiedener bisaisonaler Datensätze mit der visuellen Interpretation, also dem Kontrolldatensatz, statt. Die mit „Einschrittmethode“ bezeichneten Daten wurden durch eine Klassifikation mit  $k = 10$  ausreißerbereinigt. Der mit „Methodenkombination“ bezeichnete Datensatz besteht aus den „Guten“ aus einer Runde „Protokollmethode“ und den verbleibenden Indifferenten, die durch die „Einschrittmethode“ (im Laufe einer Kreuzvalidierung mit  $k = 11$ ) ausreißerbereinigt wurden. Weiter sind Ergebnisse von Klassifizierungen auf Grundlage von Referenzdatensätzen, die mit der „Protokollmethode“ ausreißerbereinigt wurden, dargestellt. Der Datensatz bestehend aus „Guten“ von 20 Runden der „Protokollmethode“ erzielt insgesamt das beste Ergebnis des Vergleiches.

Zusammenfassend kann aber festgehalten werden, dass die Verwendung verschiedener Ausreißerbereinigungsmethoden oder deren Kombination wenig Einfluss auf die Genauigkeitsergebnisse aus dem Vergleich mit der visuellen Interpretation hat. Entsprechende Gesamtgenauigkeiten liegen alle zwischen 88 und 89 Prozent.

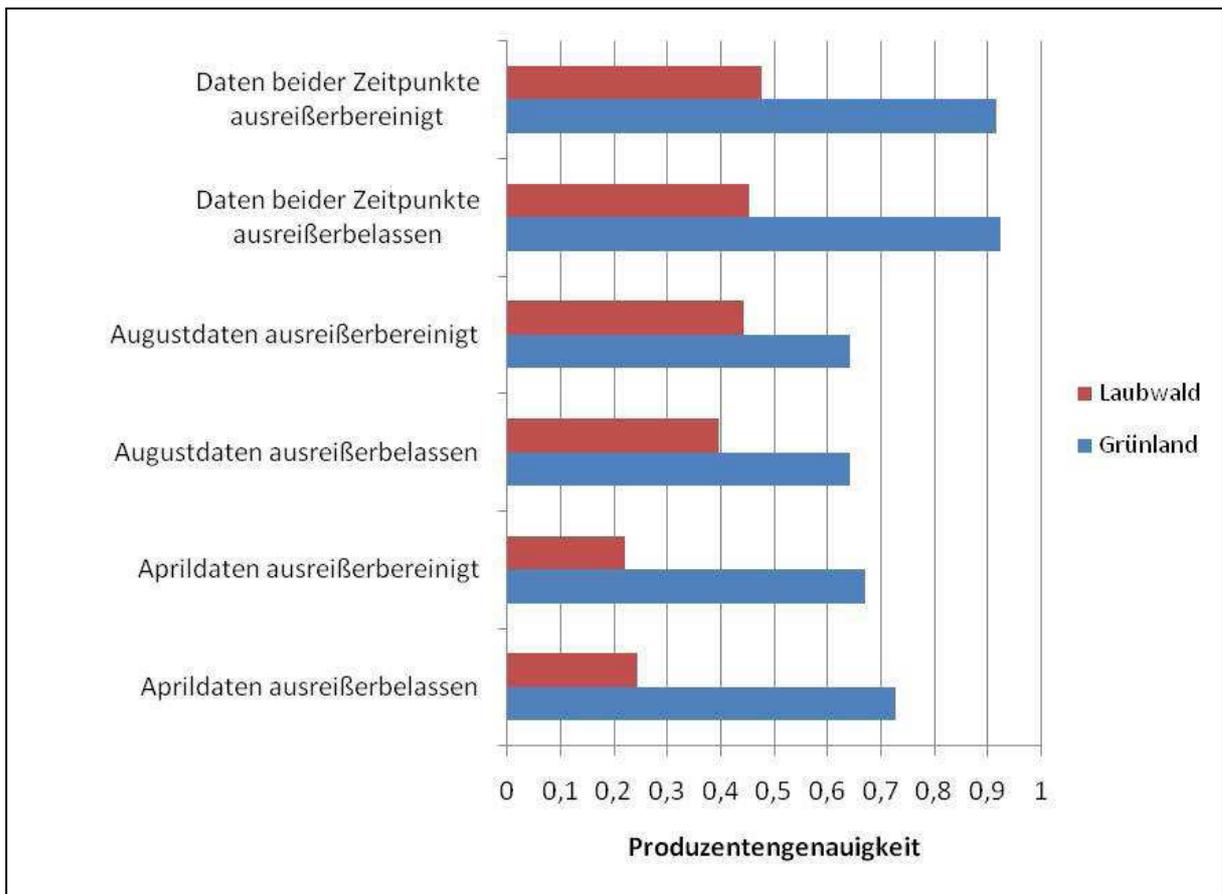


Abbildung 15: Produzentengenauigkeiten der Klassen "Laubwald" und "Grünland" aus Gegenüberstellungen von Klassifizierungen mit der visuellen Interpretation

In Abbildung 15 werden Produzentengenauigkeiten der Klassifizierungen mono- und bisaisonaler Daten gezeigt. Zudem ist der Einfluss der Ausreißerbereinigung dargestellt. Die Ausreißerbereinigung fand jeweils durch die „Einschrittmethode“ (im Zuge einer Klassifikation mit  $k = 13$ ) statt. Die Klassifizierungen erfolgten ebenfalls alle mit  $k = 13$ .

Insgesamt ist auffällig, dass die Klasse „Grünland“ eine jeweils deutlich höhere Produzentengenauigkeit aufweist als die Klasse „Laubwald“.

Wie im Vorfeld angenommen, zeigen sich Effekte in Bezug auf die Werte der Klasse „Grünland“, wenn Fernerkundungsdaten von zwei Jahreszeiten verwendet werden. Die Erwartung eines besseren Klassifizierungsergebnisses bei Verwendung von Bildern von 2 Aufnahmezeitpunkten bewahrheitet sich. Es wird eine Produzentengenauigkeit von etwa 92% gegenüber ca. 64% bei alleiniger Klassifizierung des Augustzeitpunktes erzielt. Auch der Aprilzeitpunkte erreicht etwa 20% geringere Produzentengenauigkeiten als der bisaisonale Datensatz. Bemerkenswert ist, dass die Klassifizierungen mit nicht ausreißerbereinigten Daten leicht bessere Ergebnisse erzielen, was insbesondere bei Referenzdaten des Aprilzeitpunktes der Fall ist und nicht den Erwartungen entspricht.

Auch beim Laubwald ist zu beobachten, dass die Verwendung von zwei Zeitpunkten bessere Produzentengenauigkeiten bewirkt. In Bezug auf die Ausreißerbereinigung ergibt sich ein abweichendes Ergebnis im Vergleich zur Klasse „Grünland“. Die Ausreißerbereinigung bewirkt bei

den bisaisonalen Daten und dem Augustdatensatz bessere Ergebnisse. Auch ist bei der Klasse „Laubwald“ der Augustzeitpunkt dem Aprilzeitpunkt überlegen, was beim Grünland umgekehrt ist.

Ob ein Aufnahme datum und ein Datenbearbeitungsverfahren als einem anderen überlegen einzustufen ist, hängt demnach im Besonderen von der Klasse ab.

Der nächste Abschnitt beschäftigt sich mit Benutzergenauigkeiten der Klassifizierungsergebnisse im Vergleich mit der visuellen Interpretation.

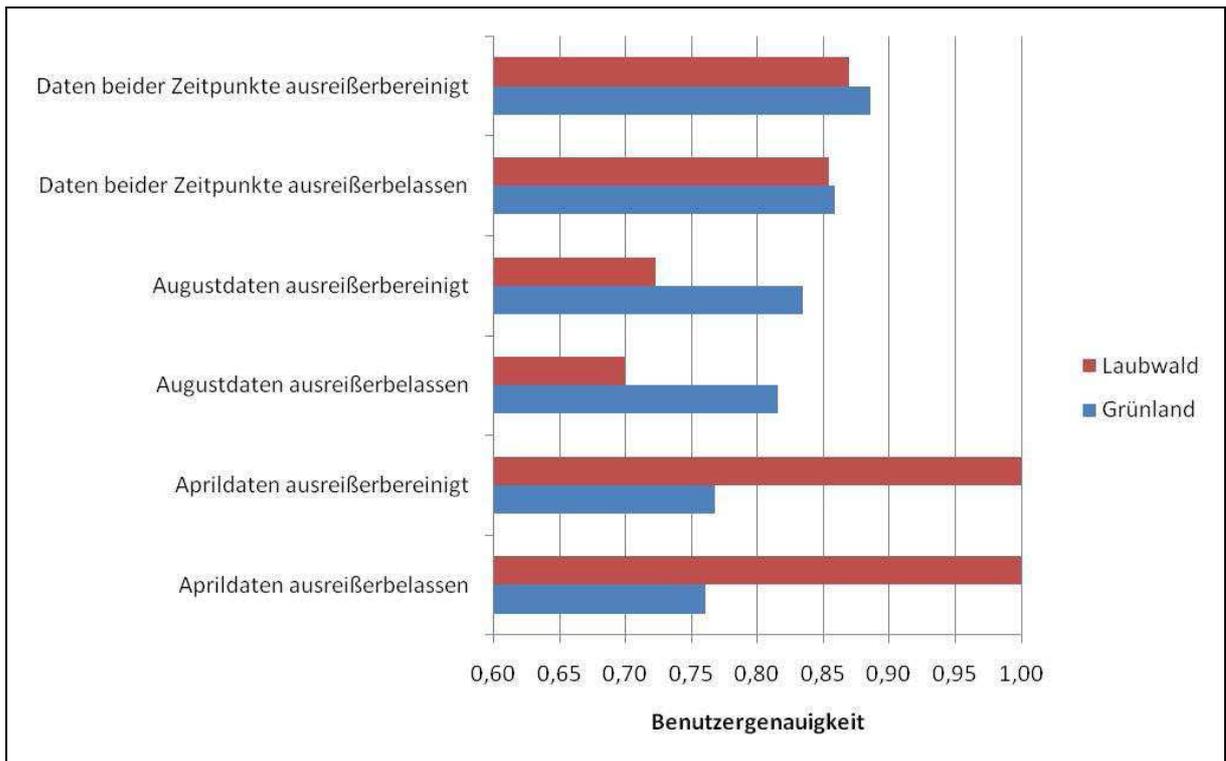


Abbildung 16: Benutzergenauigkeiten der Klassen "Laubwald" und "Grünland" aus Gegenüberstellungen von Klassifizierungen mit visueller Interpretation

Hier handelt es sich um einen Vergleich derselben Klassifizierungen wie in Abbildung 15, mit dem Unterschied, dass nicht Produzentengenauigkeit, sondern Benutzergenauigkeiten dargestellt sind. Es zeigen sich von der Produzentengenauigkeit völlig divergierende Ergebnisse. Insbesondere die Benutzergenauigkeiten der Klasse „Laubwald“ bei den Aprildatensätzen zeigen, dass einzelne Genauigkeitswerte nicht überbewertet und isoliert betrachtet werden dürfen. Die ausreißerbereinigten sowie die ausreißerbelassenen Aprildaten erreichen für Laubwald 100% Benutzergenauigkeit. Das heißt bildlich, dass ein Laubwald-Pixel einer klassifizierten Karte mit Sicherheit tatsächlich Laubwald ist. Diese Werte weichen stark von den Laubwald-Produzentengenauigkeiten der Aprildatensätze ab, die unter 25% erreichen (siehe Abbildung 15). Das letztere Ergebnis bedeutet, dass die Pixel, auf denen sich tatsächlich Laubwald befindet, nur zu einem Viertel richtigerweise zu diesem klassifiziert werden.

Die Benutzergenauigkeiten sind abgesehen von den Werten der Aprildaten für Laubwald bei bisaisonalen Datensätzen höher. Bezüglich der Ausreißerbereinigung zeigt sich eine Überlegenheit entsprechender Datensätze gegenüber ausreißerbelassenen Referenzdaten.

Als Essenz lässt sich somit festhalten, dass die Genauigkeiten der diversen Klassifizierungen mit den verschiedenen Klassen variieren. Eine bestimmte Kombination von Aufnahmezeitpunkt und Klassifizierungsmethode kann einer anderen in Bezug auf eine bestimmte Klasse klar überlegen, in Bezug auf eine andere dagegen klar unterlegen sein. Zudem zeigt sich, dass Werte wie die Benutzergenauigkeit nicht isoliert betrachtet werden können.

Der letzte Teil des Kapitels befasst sich eingehender mit Genauigkeitswerten bestimmter Klassen. Hierzu wurde die Unterscheidbarkeit der Klassen „Laubwald“ und „Nadelwald“ sowie „Acker“ und „versiegelte Flächen“ durch das kNN-Programm betrachtet. Dies erfolgte durch Vergleich einer Klassifizierung mit der visuellen Interpretation bzw. Kontrollpunkten. Die Klassifizierung basiert auf bisaisonalen, ausreißerbereinigten Daten mit ursprünglich 13 Klassen, die nach der Klassifizierung für den Vergleich mit dem Kontrolldatensatz zu 9 Klassen zusammengefasst worden sind. Die Ausreißerbereinigung erfolgte jeweils durch die „Einschrittmethod“ (im Zuge einer Klassifikation mit  $k = 13$ ). Die Werte zeigen, dass das kNN-Programm anhand der vorhandenen kNN-Referenzdatensätze einige Klassen schlecht auseinanderhalten kann.

So sind 9 Prozent der als Nadelwald klassifizierten Punkte laut visueller Interpretation Laubwald, was bedeutet, dass ebendieser Anteil des Nadelwalds einer Karte Laubwald sein sollte. Gleiches gilt für 12 Prozent der als Laubwald klassifizierten Punkte, die eigentlich Nadelwald sind. Weiter werden 44 Prozent der als Laubwald interpretierten Punkte zu Nadelwald klassifiziert. Dies bedeutet, dass 44 Prozent des in der Landschaft vorkommenden Laubwaldes in einer Karte als Nadelwald ausgewiesen wären.

Auch zwischen den Klassen „Acker“ und „versiegelte Flächen“ besteht diese Problematik. 18 Prozent der zu „versiegelte Flächen“ klassifizierten Punkte, also 18 Prozent einer Karte, sind laut visueller Interpretation Acker. Außerdem werden 2,4 Prozent der als Acker interpretierten Punkte zu „versiegelte Flächen“ klassifiziert.

Dieser Vergleich dient als Beispiel für die Schwierigkeit der Klassentrennung mittels der Referenzdaten, wobei anzumerken ist, dass jeder Referenzdatensatz Vor- und Nachteile bei der Bestimmung einzelner Klassen aufweist.

#### **4.4 Karten und deren visuelle Bewertung**

Das Kapitel zeigt auszugsweise einige Karten der Arbeit und beinhaltet überwiegend visuelle Bewertungen dieser Karten. Im Besonderen sind Effekte der Datenbearbeitungsmethoden auf die Karten dargestellt. Die Beurteilungen sind subjektiver und willkürlicher als die Genauigkeitsmaße aus Kreuzvalidierung und Vergleich mit Kontrolldaten, aber trotz dessen bei der Bewertung von Karten unverzichtbar. Zudem diente die Ansicht der Karten zur Orientierung und ließ Probleme erkennen, die anhand von Berechnungen genauer analysiert und dargestellt werden konnten.

Die folgende Abbildung ist eine ASTER-Aufnahme. Sie dient als Einführung der folgenden Karten von CORINE und einer Klassifizierung, die auch auf entsprechenden ASTER-Daten beruht. Zudem soll sie die Veranschaulichung der Klassifizierung mit dem kNN-Programm und den Vergleich von CORINE und Klassifizierung erleichtern.

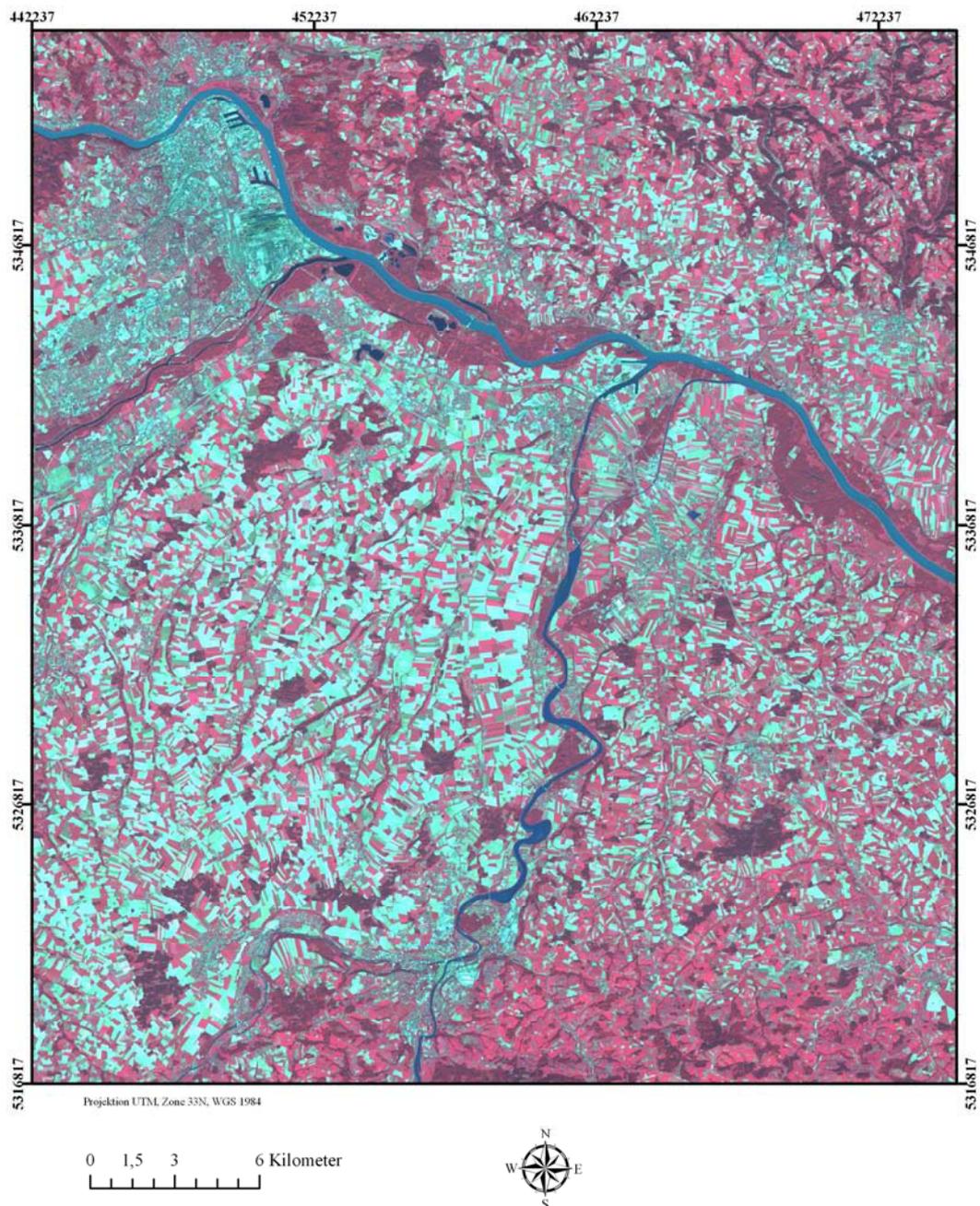
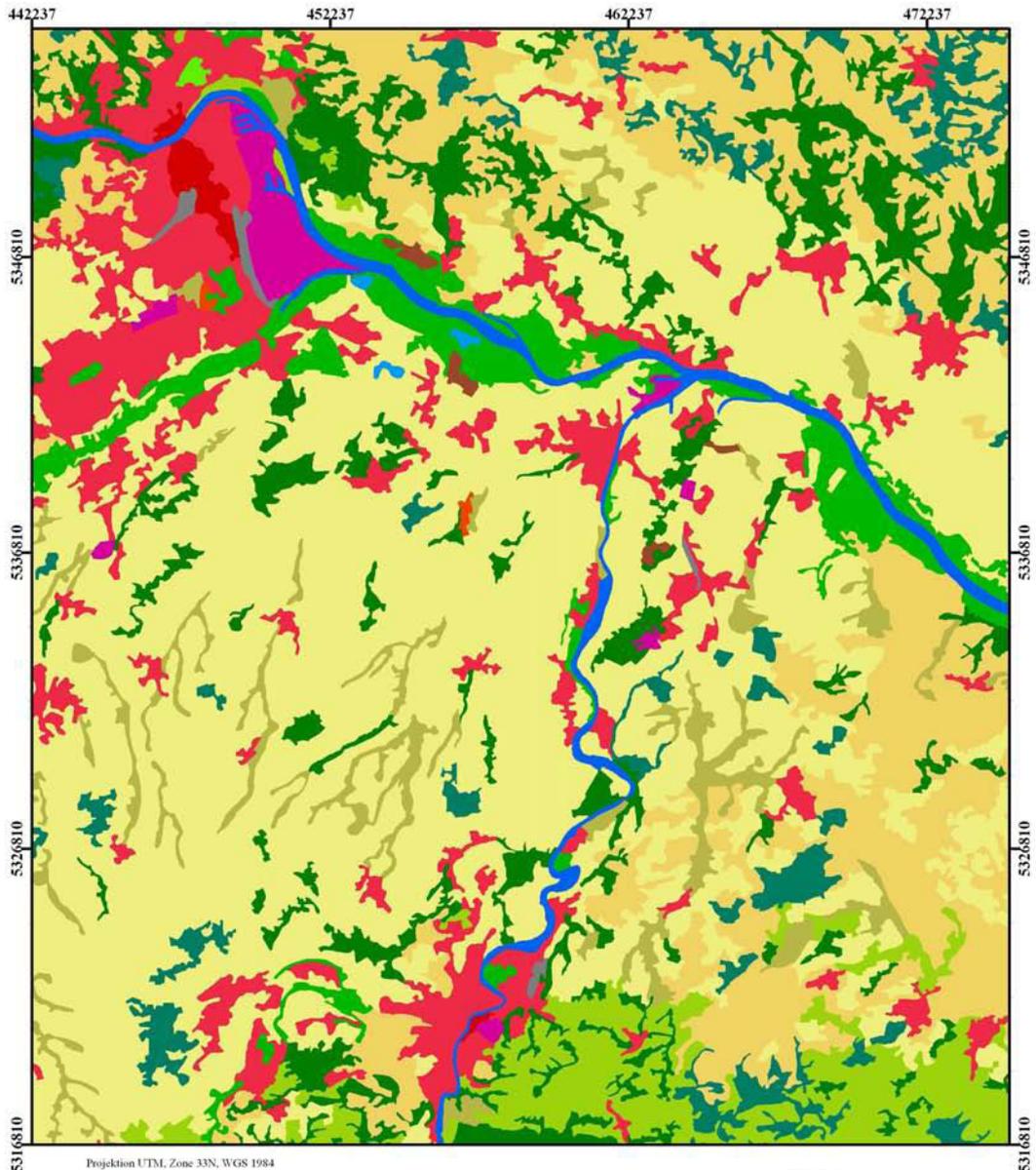


Abbildung 17: ASTER-Bildausschnitt der VNIR-Kanäle mit Linz im Nordwesten



Projektion UTM, Zone 33N, WGS 1984



**Legende**

- Klassenbezeichnung**
- Flächen durchgängig städtischer Prägung
  - Flächen nicht-durchgängig städtischer Prägung
  - Industrie- und Gewerbeflächen
  - Straßen und Eisenbahn
  - Flughäfen
  - Abbauflächen
  - Städtische Grünflächen
  - Sport und Freizeitanlagen
  - Nicht bewässertes Ackerland
  - Wiesen und Weiden
  - Komplexe Parzellenstrukturen
  - Landwirtschaft mit natürlicher Bodenbedeckung
  - Laubwald
  - Nadelwald
  - Mischwald
  - Natürliches Grasland
  - Heiden und Moorheiden
  - Wald-Strauch-Übergangsstadien
  - Flächen mit spärlicher Vegetation
  - Sümpfe
  - Gewässerläufe
  - Wasserflächen

Abbildung 18: CORINE-Kartenausschnitt mit Linz

Gemäß den CORINE-Vorgaben zur Generalisierung und gemäß den geringeren Auflösungen der für CORINE verwendeten LANDSAT-Daten zeigt Abbildung 18 keine feinen Strukturen. Es sind

eigentlich nur flächige Elemente zu erkennen und Strukturen wie Straßen oder einzelne Felder nicht identifizierbar.

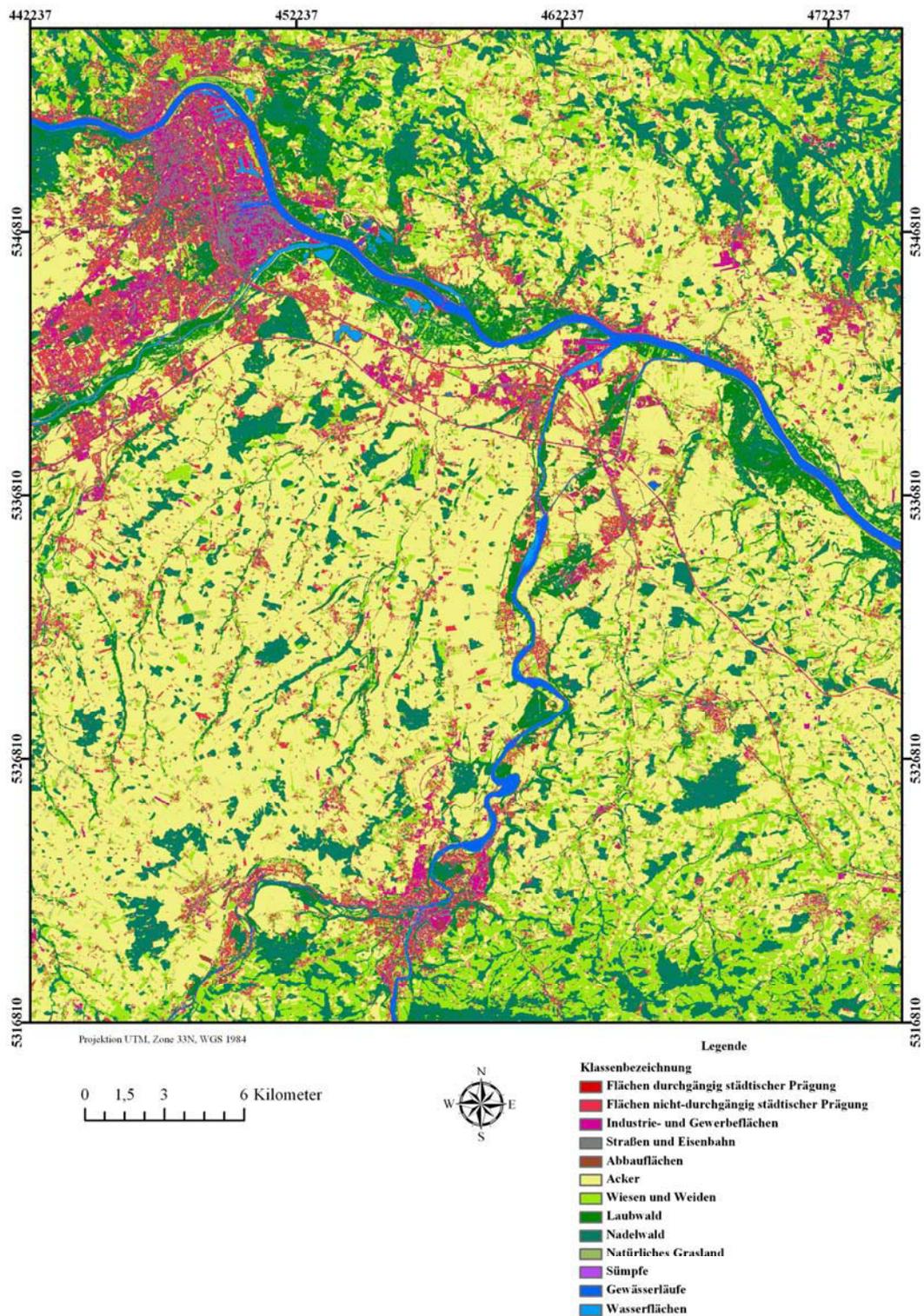


Abbildung 19: Linz enthaltender Ausschnitt einer Klassifizierung mit 13 Klassen auf Grundlage eines bisaisonalen, ausreißerbereinigten Datensatzes

Die mit kNN klassifizierte Karte in Abbildung 19 basiert auf einem bisaisonalen Referenzdatensatz mit 13 Klassen. Der Datensatz wurde im Rahmen der „Einschrittmethode“ (durch eine Klassifizierung mit  $k = 13$ ) ausreißerbereinigt. Die Bildproduktion im „mapmaker“ fand mit  $k = 13$  statt. Im Vergleich zur CORINE-Karte erkennt man deutlich den Gewinn an Details durch die 15m-Auflösung der ASTER-Bilder. Es sind insbesondere einzelne Felder, teils auch einzelne Gebäude (-gruppen) erkennbar, die wegen der Mindestflächenvorgaben im CORINE-Datensatz bei weitem nicht erfasst werden.

Es ist zu erkennen, dass das kNN-Programm im Unterschied zur generalisierten CORINE-Klassifizierung pixelweise analysiert bzw. berechnet. Wo z.B. in CORINE Mischwald als Klasse vorliegt, berechnet kNN eventuell abwechselnd Pixel (15x15m- Flächen) von Nadel- und Laubwald.

Die nächste Abbildung verdeutlicht den Unterschied der Wirkung verschieden erzeugter Referenzdatensätze auf die Kartenproduktion.

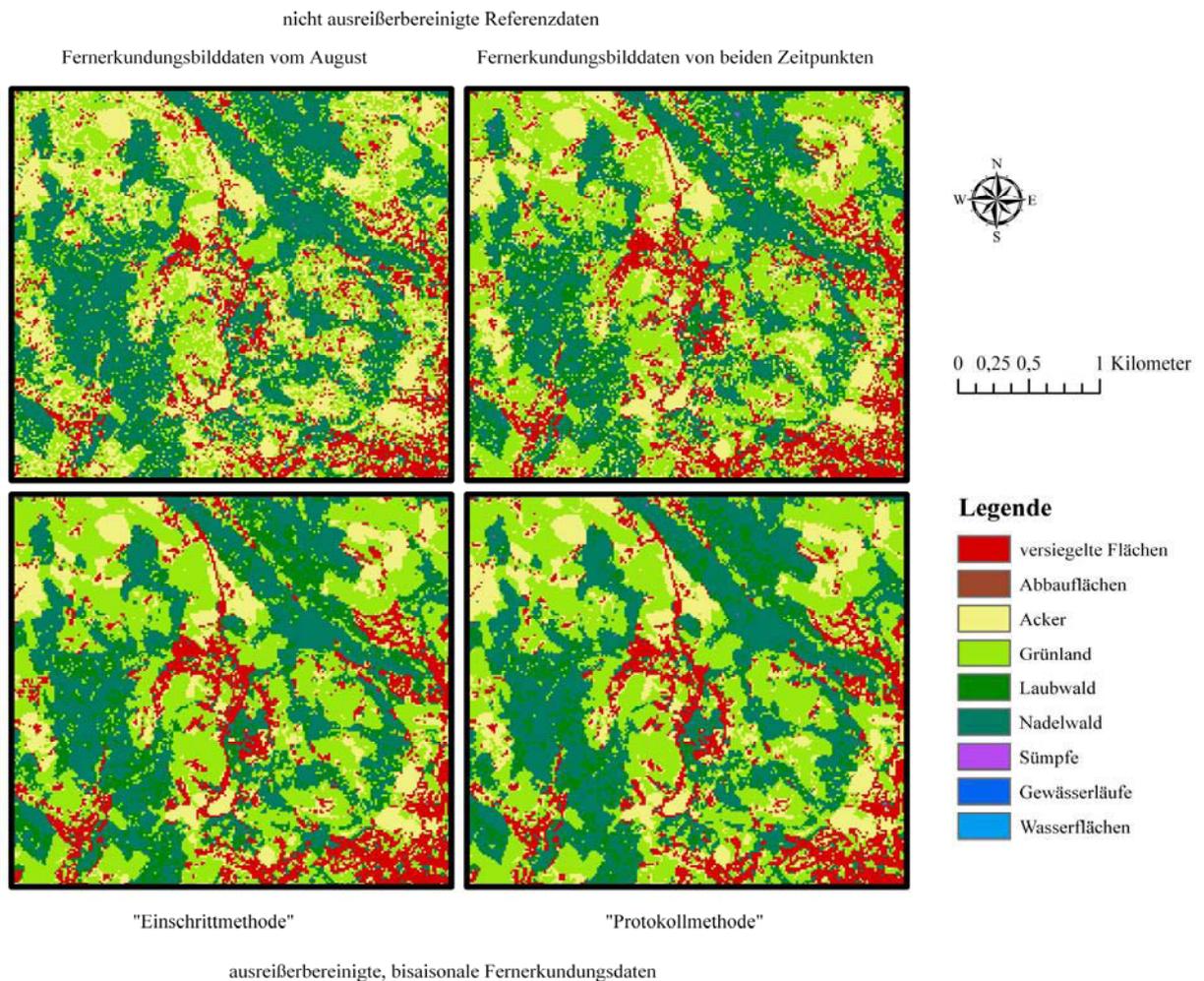


Abbildung 20: Vergleich der Karten aus verschiedenen Klassifizierungen

Den vier zu vergleichenden Kartenausschnitten in Abbildung 20 liegen vier Klassifizierungen mit je 9 Klassen zu Grunde. Die Klassifizierungen bzw. Kartenerstellungen erfolgten je mit  $k = 6$ . Jeweils ist derselbe Ausschnitt einer Fläche im Anschluss an den Nordwesten der Stadt Linz zu sehen.

Oben links ist ein Kartenausschnitt auf Grundlage der August-ASTER-Bilder und einem nicht ausreißerbereinigten Referenzdatensatz zu sehen. Man erkennt ein heterogenes, „gesprenkeltes“ Bild mit wenig flächigen Strukturen. Im Vergleich zu den restlichen Karten auf Grundlage bisaisonalen Datensätze ist die schlechtere Unterscheidung von Grünland und Acker zu nennen. Die in Karten aus bisaisonalen Datensätzen als flächige Strukturen zu erkennenden Wiesen sind hier nicht eindeutig von den Ackerflächen getrennt.

Rechts daneben befindet sich eine Karte aus bisaisonalen Fernerkundungsdaten, die einzelne Felder besser erkennen lässt. Allerdings sind in größeren Strukturen immer wieder fremde Klassen eingestreut, was an den belassenen Ausreißern liegt.

Die Karte links unten bezieht sich ebenfalls auf 2 Zeitpunkte und 9 Klassen. Ihr Referenzdatensatz wurde im Unterschied zur vorher genannten Karte durch die „Einschrittmethode“ (im Laufe einer Klassifizierung mit  $k = 13$ ) von Ausreißern bereinigt. Die Wirkung der Ausreißerbereinigung ist darin zu sehen, dass die Flächen deutlich homogener sind, also z.B. in Felder weniger Fremdklassen auftauchen. Dies ist als Qualitätssprung zu werten, da gerade im letztgenannten Beispiel in Bezug auf Acker oder bei Grünland homogene Flächen der Realität entsprechen. Z.B. enthält eine Wiese tatsächlich keine eingestreuten 15 mal 15 Meter großen Ackerflächen sondern ist einheitlich grasbewachsen.

Schließlich befindet sich rechts unten eine Karte ebenfalls auf Grundlage bisaisonaler Daten, die jedoch auf ausreißerbereinigte Referenzdaten aus der „Protokollmethode“ beruht. Diese wurde in 20 Runden durchgeführt und der Referenzdatensatz aus den „Guten“ der einzelnen Stufen zusammengesetzt. Im Vergleich zur Karte auf Grundlage der durch die „Einschrittmethode“ ausreißerbereinigten Referenzdaten hat diese Karte noch homogenere und klarer abgrenzbare Flächen mit weniger Fremdklassen in diesen. Das ist insbesondere in den Waldflächen und der Klasse „Grünland“ der Fall. Dieses Ergebnis deckt sich mit jenem der Gegenüberstellung von Kontrolldaten mit Klassifizierungen in Abbildung 14, wo die „Guten“ der Bereinigungsstufen 1-20 das beste Ergebnis erzielten.

Neben unterschiedlich erstellten Referenzdatensätzen hat auch die Wahl des  $k$ -Wertes einen deutlichen Einfluss auf die Qualität der Karten, was der nächste Abschnitt veranschaulicht.

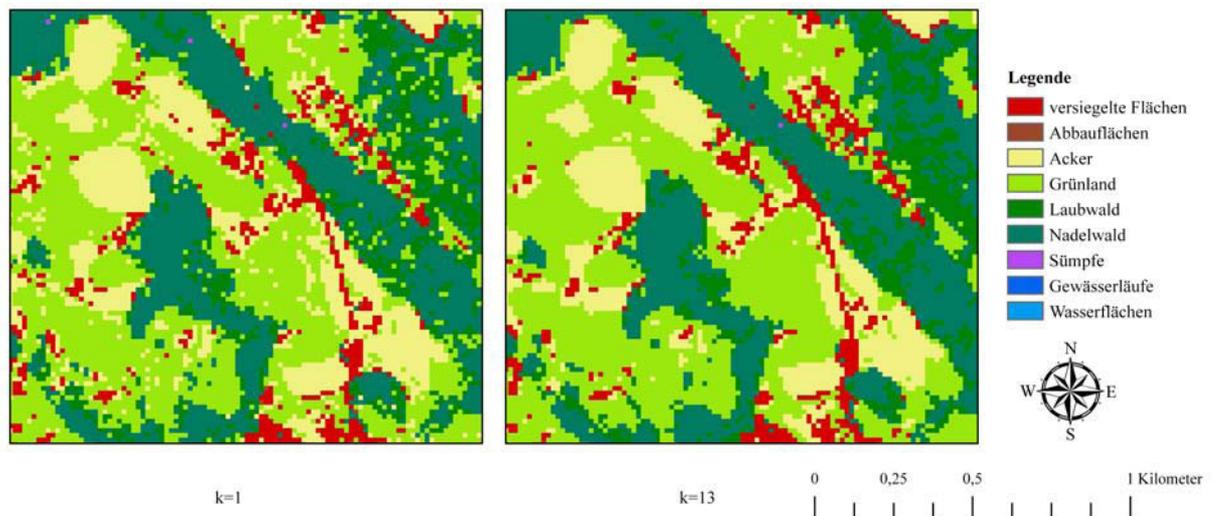


Abbildung 21: Wirkung des k-Wertes auf die Klassifizierung

Abbildung 21 zeigt zwei Karten desselben Ausschnittes nordwestlich von Linz. Es wurde jeweils ein bisaisonaler Datensatz klassifiziert, wobei die Referenzdaten durch die „Einschrittmethod“ (im Rahmen einer Klassifizierung mit  $k = 10$ ) ausreißerbereinigt worden waren. Die linke Karte wurde im „mapmaker“ mit  $k = 1$  erstellt und weist relativ viele Fremdklassen in den flächigen Strukturen auf. Im Vergleich dazu zeigt die mit  $k = 13$  klassifizierte Karte ein „ruhigeres“, weniger „gesprenkeltes“ Bild. Zum Beispiel befinden sich in den Grünlandflächen weniger Ackerpixel und im Wald weniger Grünlandpixel als in der linken Karte. Die Wahl des höheren k-Wertes hat insgesamt einen positiven Einfluss auf die Qualität der Karte.

Im Folgenden werden die bisher beschriebenen Klassifizierungen mit einer Klassifizierung verglichen, die nicht auf dem CORINE-Referenzdatensatz, sondern auf den visuell interpretierten Referenzdaten (dem Kontrolldatensatz) basiert. Abbildung 22 stellt dazu das Ergebnis der bislang besten Klassifizierung mit dem CORINE-Referenzdatensatz dem Ergebnis der Klassifizierung mit dem visuell erstellten Referenzdatensatz gegenüber. Zusätzlich sind die Ausgangs-ASTER-Bilder gezeigt.

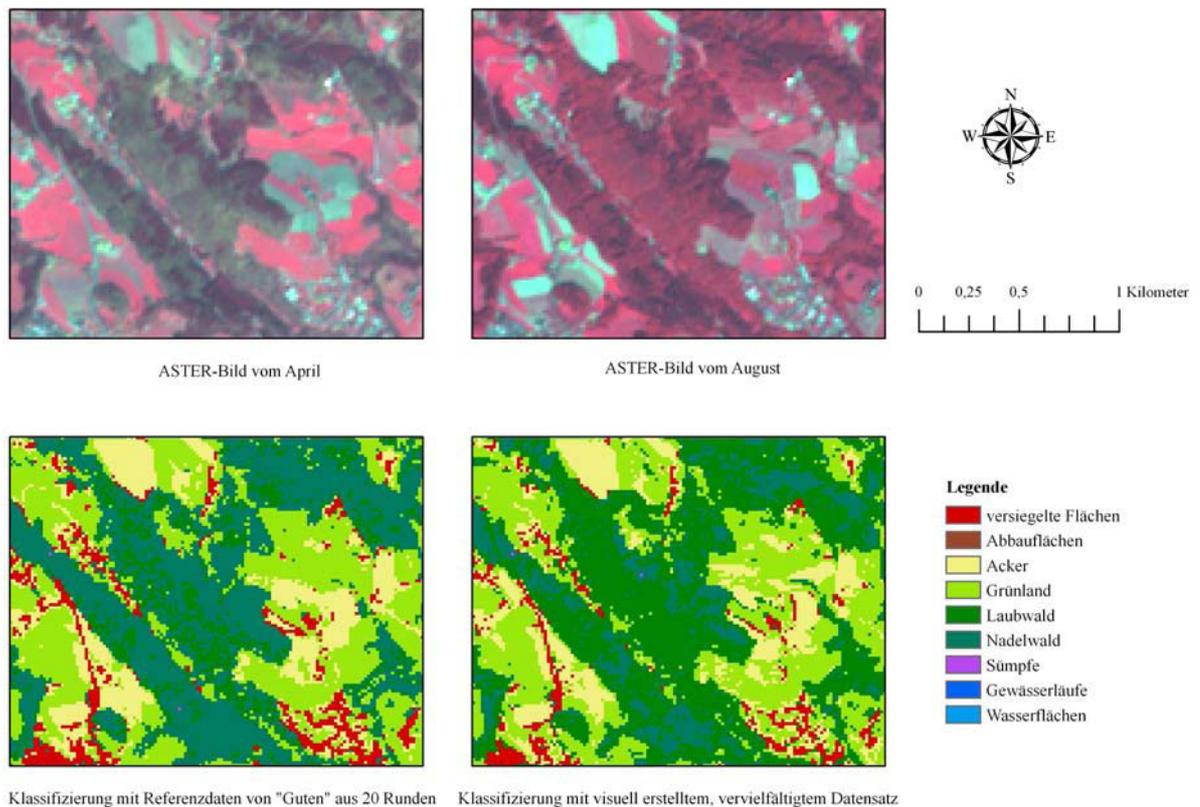


Abbildung 22: Vergleich einer Karte auf Grundlage der CORINE-Daten (durch die „Protokollmethode“ ausreißerbereinigt) mit einer Karte auf Grundlage eines visuell erstellten, „vervielfältigten“ Datensatzes

In den vier Bildern von Abbildung 22 ist jeweils ein Ausschnitt unmittelbar nordwestlich von Linz zu sehen. Die spektralen Werte der ASTER-Bilder sind in Summe Grundlage für die unten zu sehenden Klassifizierungen. Die Klassifizierungen sind jeweils mit  $k=6$  und auf Basis bisaisonaler Referenzdatensätze mit 9 Klassen durchgeführt worden. Vorrangig soll mit der Abbildung die Problematik der Klassifizierung von Laub- und Nadelwald veranschaulicht werden. Die Klassifizierung unter Verwendung des mit der „Protokollmethode ausreißerbereinigten CORINE-Referenzdatensatzes (bestehend aus „Guten“ von zwanzig Runden) weist deutlich mehr Nadelwald- als Laubwaldanteile auf, was bei der Klassifizierung unter Verwendung des visuell erstellten, vervielfältigten Datensatzes (siehe Kapitel 3.5.1) umgekehrt ist. Wenn man die ASTER-Bilder betrachtet, fällt auf, dass bei der letztgenannten Klassifizierung dort Laubwald vorhanden ist, wo im Aprilbild braungrüne und im Augustbild rote Pixel(-flächen) vorkommen. Nadelwald wird dort ausgewiesen, wo sich jeweils schwarzrote Pixel befinden. In Bezug auf die Unterscheidung von Laub- und Nadelwald ist diese Klassifizierung größtenteils richtig und entspricht der visuellen Einschätzung. Dieses Ergebnis legt nahe, dass die schlechte Unterscheidung von Laub- und Nadelwald (siehe auch Kapitel 4.5) ihren Ursprung im Referenzdatensatz und nicht in der schwierigen Unterscheidbarkeit der spektralen Werte hat. Diesbezüglich ist der Referenzdatensatz also ausreißerdominiert. Referenzpunkte, die spektral der Klasse Laubwald zuzuordnen sind, sind mit der Übernahme der CORINE-Klasse „Nadelwald“ zugeordnet worden. Die Anzahl dieser Ausreißer überwiegt offenbar die Anzahl von Referenzpunkten ähnlicher spektraler Werte, die richtigerweise die Klasse „Laubwald“ aufweisen.

Bei der Unterscheidungsproblematik anderer Klassen ist der visuell erstellte und vielfältige Referenzdatensatz nicht so überlegen, was darauf hindeutet, dass diese Unterscheidungen tatsächlich spektral schwierig sind, was sich bei der Betrachtung der ASTER-Bilder bestätigt. Dies ist zum Beispiel der Fall, wenn Ackerflächen zu beiden Zeitpunkten keine Vegetation tragen. Sie weisen dann jeweils helle Pixel auf, die visuell versiegelten Flächen gleichen.

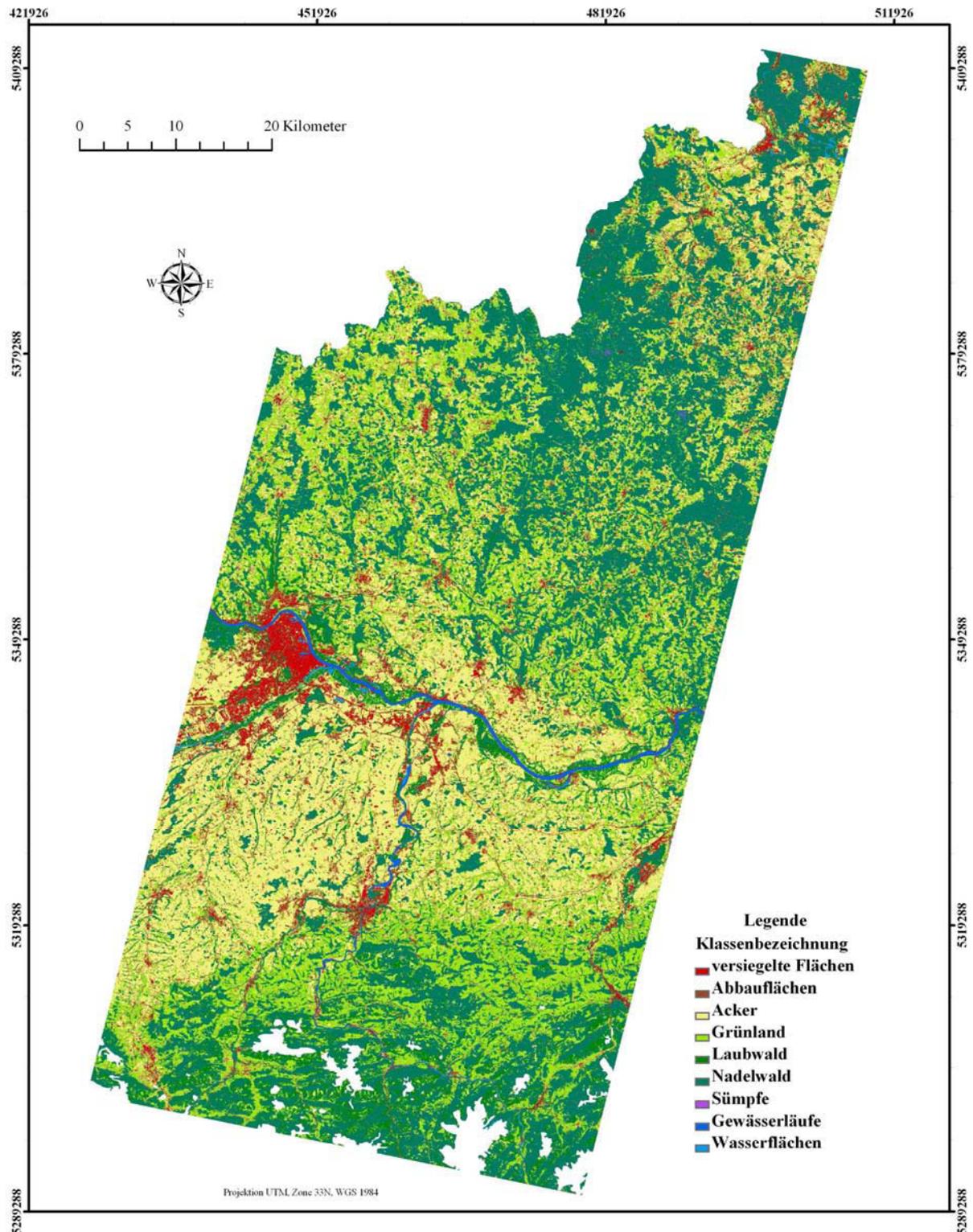


Abbildung 23: Klassifizierung des gesamten Untersuchungsgebietes

Die Karte in Abbildung 23 zeigt eine Klassifizierung mit  $k = 13$  auf Grundlage eines bisaisonalen Datensatzes mit 9 Klassen und Ausreißerbereinigung durch die „Einschrittmethode“ (im Zuge einer Klassifizierung mit  $k = 13$ ). Man sieht vor allem die unterschiedlichen Landschaften des Untersuchungsgebietes. Der nördliche Teil der Karte besteht hauptsächlich aus Wald- und Wiesenflächen, wobei im Nordwesten der Karte das Grünland überwiegt und sich im Nordosten auch eine Gegend mit hohem Ackeranteil befindet. Links in der Mitte der Karte an der Donau ist Linz als Ansammlung vieler rot dargestellter, versiegelter Flächen zu erkennen. In der Mitte der Karte und vor allem südlich der Donau erstrecken sich Niederungen mit großem Ackeranteil, bevor sich südlich höhere Landschaften mit hohem Grünland- und Nadelwaldanteilen anschließen. Die weißen Flächen sind der Maske geschuldet, die beim Aprilzeitpunkt die Schneeflächen ausblendet (siehe Kapitel 3.2.5). Im Ganzen lässt die Klassifizierung die Strukturen der Landschaften gut erkennen.

#### 4.5 Vergleich der Flächenanteile aus dem CORINE-Datensatz und aus den produzierten Karten mit Flächenangaben aus anderen zuverlässigen Quellen

CORINE			„Waldkarte“ bzw. InVeKoS	Differenz der Anteile
Code	Bezeichnung	Anteil [%]	Anteil [%]	[%]
112	Flächen nicht-durchgängig städtischer Prägung	4,1		
211	Nicht bewässertes Ackerland	14,9	23,0	8,0
231	Wiesen und Weiden	13,7		
242	Komplexe Parzellenstrukturen	20,8		
243	Landwirtschaft mit natürlicher Bodenbedeckung	1,1		
311	Laubwald	3,3	6,6	3,3
312	Nadelwald	19,9	27,3	7,4
313	Mischwald	16,6	6,8	-9,7
322	Heiden und Moorheiden	2,0		
332	Felsflächen ohne Vegetation	0,8		
333	Flächen mit spärlicher Vegetation	0,7		
511	Gewässerläufe	0,6		
512	Wasserflächen	0,9		

Tabelle 4: Flächenanteilsvergleich von CORINE mit InVeKoS und „Waldkarte“-Daten

Tabelle 4 stellt Flächenanteile von CORINE und Daten aus InVeKoS und der „Waldkarte“ (siehe Kapitel 3.8.4) in Oberösterreich gegenüber. Die Flächengrößen selbst können nicht verglichen werden, weil sich die Daten zur „Waldkarte“ und InVeKoS nicht auf ganz Oberösterreich beziehen.

Die Daten sind aufgrund der Landnutzungsproblematik bei CORINE generell schwer zu vergleichen, was sich besonders bei den landwirtschaftlichen Klassen äußert. Dies liegt daran, dass bei CORINE z.B. mehrere Klassen existieren, die zumindest in Anteilen landwirtschaftliche Flächen enthalten. Dabei handelt es sich um die Klasse „242“ „Komplexe Parzellenstrukturen“, die einen Anteil von knapp 21% einnimmt, sowie die Klasse „243“ „Landwirtschaft mit natürlicher Bodenbedeckung“, die bei CORINE 1,1% der Fläche Oberösterreichs ausmacht. Die Klasse „211“

„nicht bewässertes Ackerland“ bedeckt laut CORINE knapp 15% Oberösterreichs, die InVeKoS-Daten hingegen weisen einen Acker-Anteil von 23% also 8 Prozent mehr auf. Wenn man allerdings die Ackeranteile in den vorher genannten Klassen berücksichtigt, dürften die Anteile vergleichbar sein.

Beim Waldanteil stellt sich die Problematik nicht in ähnlicher Weise, da bei CORINE kaum waldfremde Klassen vorhanden sind, die Waldanteile enthalten. Die walddhaltende Klasse „Landwirtschaft mit natürlicher Bodenbedeckung“ nimmt nur 1,1% der Fläche ein und ist daher vernachlässigbar. Die drei jeweils aufgeführten Waldklassen „Laubwald“ („311“), „Nadelwald“ („312“) und „Mischwald“ („313“) sind somit vergleichbar. Die „Waldkarte“ enthält 6,6% Laubwald und damit doppelt so viel wie CORINE. Auch der Nadelwaldanteil der „Waldkarte“ liegt mit 27,3% höher als der CORINE-Nadelwaldanteil (19,9%). Dagegen hat der Mischwaldanteil bei CORINE mit 16,6% der Fläche Oberösterreichs einen um 9,7% höheren Anteil als die „Waldkarte“ mit 6,8%. Die Waldgesamtanteile der CORINE- und der „Waldkarten“-Daten liegen mit 39,8 bzw. 40,7 Prozent in einem vergleichbaren Bereich, weshalb sich folgern lässt, dass der höhere Mischwaldanteil im CORINE-Datensatz auf Generalisierung beruht und im „Waldkarten“-Datensatz eine feinere Unterscheidung bzw. höhere räumliche Auflösung für Laub-, Nadel- und Mischwald besteht.

Insgesamt zeigt der Vergleich, dass die CORINE-Daten wegen ihrer Landnutzungs-Landbedeckungs-Problematik sowie der Generalisierungen eine mit anderen Landbedeckungskarten schwer vergleichbare und verwendbare Basis sind.

Zur Einschätzung der Klassifizierungsqualität zeigt der nächste Abschnitt die Gegenüberstellung von Flächendaten einer klassifizierten Karte mit den als zuverlässig geltenden Quellen InVeKoS und „Waldkarte“.

Flächenanteile im Bezirk Freistadt [%]			
Klasse	Klassifizierung	„Waldkarte“- und InVeKoS-Daten	Differenz von „Waldkarte“ und InVeKoS-Daten zur Klassifizierung
Acker	14,3	20,5	6,2
Laubwald	2,0	5,5	3,5
Nadelwald	46,0	39,6	-6,4

Tabelle 5: Flächenanteilsvergleich einer Klassifizierung mit „Waldkarte“- und InVeKoS-Daten im Bezirk Freistadt

Die obige Tabelle vergleicht im ca. 99.400 ha großen Bezirk Freistadt „Waldkarte“- und InVeKoS-Daten mit einer Klassifizierung mit 9 Klassen und  $k = 13$ , die auf dem bisaisonalen Fernerkundungsdatensatz und Ausreißerbereinigung mit der „Einschrittmethode“ (bei einer Kreuzvalidierung mit  $k = 13$ ) beruht. In den „Waldkarte“-Daten findet sich bei Laub- und Nadelwald jeweils die Hälfte des Mischwaldanteils. Man erkennt, dass die Klassifizierung den Ackeranteil mit 14,3% im Vergleich zu InVeKoS mit 20,5% um 6,2% unterschätzt. Es liegt nahe, dass diese Fläche zu Grünland klassifiziert wurde. Die Klasse „Laubwald“ nimmt in den „Waldkarte“-Daten 5,5 gegenüber 2 Prozent bei der Klassifizierung ein. Der fehlende Anteil findet sich wahrscheinlich zu einem großen

Anteil in der Klasse „Nadelwald“ wieder, bei welchem die Klassifizierung einen Anteil von 46% aufweist. Die „Waldkarte“ hat hier 39,6% Flächenanteil. Der Vergleich zeigt insgesamt, dass die Klassifizierung Mängel gegenüber den zuverlässigen Quellen aus InVeKoS und „Waldkarte“ aufweist.

Code	Bezeichnung	Klassifizierungs-Anteil [%]	CORINE-Anteil [%]	Differenz zwischen Klassifizierung und CORINE [%]
112	Flächen nicht-durchgängig städtischer Prägung	2,30	5,22	-2,92
121	Industrie- und Gewerbeflächen	0,83	0,34	0,49
122	Straßen und Eisenbahn	0,70	0,07	0,63
142	Sport und Freizeitanlagen	0,89	0,07	0,82
211	Nicht bewässertes Ackerland	15,70	19,11	-3,40
231	Wiesen und Weiden	4,99	9,30	-4,32
242	Komplexe Parzellenstrukturen	36,03	28,44	7,59
243	Landwirtschaft mit natürlicher Bodenbedeckung	0,00	2,09	-2,09
311	Laubwald	1,44	2,27	-0,83
312	Nadelwald	23,08	22,52	0,55
313	Mischwald	9,80	9,41	0,39
321	Natürliches Grasland	0,82	0,02	0,80
324	Wald-Strauch-Übergangsstadien	1,61	0,07	1,54
511	Gewässerläufe	0,56	0,14	0,43

Tabelle 6: Flächenanteile von CORINE und einer Klassifizierung im Untersuchungsgebiet

In Tabelle 6 sind Flächenbilanzen von CORINE und einer auf CORINE basierenden Klassifizierung von bisaisonalen Fernerkundungsdaten aufgeführt. Die Klassifizierung mit  $k = 13$  basierte auf einem mittels der „Einschrittmethode“ (durch eine Klassifizierung mit  $k = 12$ ) ausreißerbereinigten Referenzdatensatz. Die Flächen beziehen sich auf jene Gemeinden, die vollständig in den in der Arbeit verwendeten Szenen liegen und ergeben in Summe ca. 3.886 Quadratkilometer. Von den 20 im Gebiet vorkommenden Klassen finden sich in der Tabelle nur jene, die entweder in der Klassifizierung oder in CORINE mehr als 0,5% Flächenanteil einnehmen.

Die Klasse „Flächen nicht-durchgängig städtischer Prägung“ nimmt in der Klassifizierung 2,30 und bei CORINE 5,22% ein, was eine Differenz von 2,92 Prozent am Gesamtflächenanteil bzw. einen mehr als doppelt so großen Anteil bei CORINE ergibt. Eine Erklärung liegt darin, dass diese Klasse als Landnutzungsklasse in CORINE auch Vegetationsflächen beispielsweise in Gärten und Parks im Siedlungsgebiet mit einschließt, die in der Klassifizierung vor allem mit den Landbedeckungen „Grünland“ und „Laubwald“ ausgewiesen sind.

„Nicht bewässertes Ackerland“ macht in der Klassifizierung 15,70 und bei CORINE 19,11 Prozent der Fläche aus. Die 3,40% Differenz finden sich unter anderem in dem bei der Klassifizierung deutlich größeren Anteil in der Klasse „Komplexe Parzellenstrukturen“ wieder. „Komplexe

Parzellenstrukturen“ nehmen bei der Klassifizierung 36,03% und damit 7,59% mehr als bei CORINE (28,44%) ein. Bei beiden Datensätzen ist sie die meistvertretene Klasse. Auch hier wird ein Teil der Problematik des CORINE-Datensatzes deutlich, da es sich um eine Mischklasse hauptsächlich aus Wiesen und Acker handelt, die sich eher auf Landnutzung als auf Landbedeckung bezieht. Bei pixelweise Analyse hat sie keinen Bestand, aber wegen ihres großen Anteils im Referenzdatensatz kann sie andere Klassen dominieren. Dies ist beispielsweise der Fall, wenn die Mischklasse viele Referenzpunkte enthält, die eigentlich typisch für eine andere Klasse sind. Man kann z.B. beobachten, dass Wiesenflächen in der Klassifizierung dieser Mischklasse zugeordnet werden, diese also ein zu hohes Gewicht besitzt. Daher weist CORINE bei „Wiesen und Weiden“ mit 9,30% einen fast doppelt so hohen Anteil wie die Klassifizierung mit 4,99% auf.

„Landwirtschaft mit natürlicher Bodenbedeckung“ findet sich in der Klassifizierung gar nicht wieder, in CORINE hingegen nimmt diese Klasse 2,09% der Fläche ein. Der „Nadelwald“-Anteil ist mit 23,08% bei der Klassifizierung und 22,52% bei CORINE ähnlich vertreten. Die Klasse „Natürliches Grasland“ weist dagegen wieder große Unterschiede auf und ist in CORINE (0,02%) im Gegensatz zur Klassifizierung (0,82%) kaum vertreten. Gleiches gilt für „Wald-Strauch-Übergangsstadien“ mit 1,61 Prozent Flächenanteil in der Klassifizierung und nur 0,07 in CORINE. Dies liegt wahrscheinlich daran, dass diese Klassen Referenzdaten enthalten, die anderen Klassen spektral gleichen. Somit dürften Flächen der Klasse „Wiesen und Weiden“ der Klasse „Natürliches Grasland“ zugeordnet werden. Teile der Referenzdaten zu „Wald-Strauch-Übergangsstadien“ gleichen wahrscheinlich spektral Laubwald, was zur Klassifizierung von Pixeln dieser Landbedeckung zu „Wald-Strauch-Übergangsstadien“ führt.

Der Vergleich zeigt insgesamt, dass die CORINE-Daten insbesondere wegen Ihres Charakters als landnutzungs-geprägter Datensatz nur eingeschränkt für die Verwendung als Referenzdatensatz geeignet sind. Häufig werden Flächen wie Wiesen oder auch Acker aufgrund ähnlicher Referenzpixel in Mischklassen diesen zugeordnet. Somit werden die Flächenanteile verzerrt.

Code	Bezeichnung	Klassifizierungs -Anteil [%]	CORINE -Anteil [%]	Differenz zwischen CORINE und Klassifizierung [%]
111	Flächen durchgängig städtischer Prägung	0,03	4,63	4,61
112	Flächen nicht-durchgängig städtischer Prägung	13,24	34,00	20,75
121	Industrie- und Gewerbeflächen	9,27	10,45	1,18
122	Straßen und Eisenbahn	11,42	2,05	-9,37
124	Flughäfen	0,44	0,00	-0,44
131	Abbauflächen	0,66	0,00	-0,66
141	Städtische Grünflächen	0,59	0,71	0,12
142	Sport und Freizeitanlagen	0,70	0,38	-0,32
211	Nicht bewässertes Ackerland	8,77	13,41	4,65
231	Wiesen und Weiden	2,39	1,78	-0,60
242	Komplexe Parzellenstrukturen	24,82	5,76	-19,06
243	Landwirtschaft mit natürlicher Bodenbedeckung	0,00	1,06	1,06
311	Laubwald	2,56	13,15	10,59
312	Nadelwald	7,33	0,11	-7,22
313	Mischwald	11,67	7,14	-4,53
321	Natürliches Grasland	0,20	0,00	-0,20
324	Wald-Strauch-Übergangsstadien	0,50	0,00	-0,50
411	Sümpfe	0,00	0,00	0,00
511	Gewässerläufe	3,13	4,73	1,59
512	Wasserflächen	2,29	0,64	-1,65

Tabelle 7: Flächenanteilsvergleich von CORINE und einer Klassifizierung im Bezirk Linz (Stadt)

Wenn man die gleiche Klassifizierung und die CORINE-Daten im etwa 9.560 ha großen Bezirk Linz (Stadt) betrachtet, wird die Problematik des CORINE-Datensatzes noch deutlicher. Die Flächenanteile der Klassifizierung weichen hier noch mehr von den CORINE-Daten ab. Es finden sich Abweichungen von über 10%. Die Klasse „Flächen nicht-durchgängig städtischer Prägung“ nimmt z.B. in der Klassifizierung gute 13% und damit fast 21 Prozent weniger ein als bei CORINE (34%). Eine ähnlich hohe Abweichung besteht bei der Klasse „Komplexe Parzellenstrukturen“, jedoch hat die Klassifizierung hier einen Flächenanteil von knapp 25% und damit gute 19% mehr als CORINE (knappe 6%). Auch die Klasse „Laubwald“ weist knapp 10,6% Abweichung zwischen CORINE (13,15%) und der Klassifizierung (2,56%) auf.

Grundsätzlich stellt sich die Frage, welcher der Datensätze fehlerhaft ist. Die Differenzen lassen sich durch Mängel in beiden Datensätzen erklären. Der CORINE-Datensatz beinhaltet unrealistische Flächenbilanzen durch Generalisierungs- und Mischklasseneffekte, die über den Referenzdatensatz auch auf die Klassifizierung übertragen werden.

Zusammenfassend kann festgehalten werden, dass die Datensätze CORINE, Waldkarte, InVeKoS und Klassifizierungen auf Grundlage der CORINE-Referenzdaten nur eingeschränkt miteinander verglichen werden können und unterschiedliche Flächenbilanzen ergeben. CORINE ist für Klassifizierungen mit pixelweiser Analyse ein Referenzdatensatz mit erheblichen Nachteilen. Die Gründe dafür liegen (neben allenfalls gegebenen „echten“ Fehlern) in der unterschiedlichen Generalisierung und unterschiedlichen Klassendefinitionen.

## 4.6 Diskussion der verschiedenen Bewertungsmethoden

Die vier Bewertungsmittel der Klassifizierungen (Kreuzvalidierung, Vergleich mit visueller Interpretation, visuelle Kartenbewertung, Flächenvergleiche) liefern unterschiedliche Ergebnisse und sind in Kombination zu betrachten. Die höchsten Genauigkeitswerte liefert die Kreuzvalidierung der Referenzdatensätze, welche die Qualität der Karten allerdings überschätzt. Der Vergleich der Klassifizierungen mit den visuell erstellten Kontrollpunkten liefert realistischere Werte. Weiters widersprechen sich die Ergebnisse der einzelnen Bewertungsmethoden in Einzelfällen. Insbesondere die Kreuzvalidierung des Referenzdatensatzes weicht in ihrer Aussage von anderen Bewertungsverfahren ab. Dies liegt wohl insbesondere an der Selbstreferenz dieser Daten. Mit Selbstreferenz ist hier gemeint, dass die später kreuzvalidierten Daten nicht zu ihrem Muster passende Daten entfernt haben und so ein zu gutes Ergebnis liefern. Es ist also nur unter Vorbehalt möglich, aus den Ergebnissen der Kreuzvalidierung Schlüsse auf die Qualität der Referenzdatensätze zu ziehen.

Einige Ergebnisse der Kreuzvalidierung scheinen nach Betrachtung der Genauigkeiten aus punktuellen Vergleichen mit visueller Interpretation und nach visueller Beurteilung der Karten mit „gesundem Menschenverstand“ wenig sinnvoll. So erzielen nicht ausreißerbereinigte Datensätze beispielsweise bei der Klasse Grünland bessere Ergebnisse, was den Ergebnissen der anderen Bewertungsverfahren widerspricht.

Ein recht übereinstimmendes Ergebnis ist eine gewisse Gegenläufigkeit der Datenbearbeitungsmethoden, was bedeutet, dass bei bereits erfolgter Bearbeitung des Datensatzes die Anwendung eines weiteren Verfahrens nicht mehr entsprechende Effekte hat (TOMPPPO et. al 2008, c). Ist z.B. ein Datensatz bereits ausreißerbereinigt tritt die Wirkung des kNN-Verfahrens in den Hintergrund und es werden häufig bessere Ergebnisse mit der Wahl niedrigerer k-Werte als in unbereinigten Datensätzen erreicht.

Insgesamt ist weiterhin eine Kombination von Bewertungsmöglichkeiten von Klassifizierungen angebracht. Die Kreuzvalidierung des Referenzdatensatzes allein hat wohl die geringste Aussagekraft für dessen Qualität. Mehr Gewicht ist auf einen unabhängigen Kontrolldatensatz in Kombination mit der visuellen Beurteilung der Karten zu legen. Auch der Vergleich mit anderen Quellen zu Flächenbilanzen liefert aussagekräftige Ergebnisse, die in Kombination mit den anderen Bewertungsverfahren nützlich sind. Dies ist zum Beispiel bei der Klasse „Nadelwald“ der Fall, wo der Flächenvergleich zeigt, dass deren Anteil durch die Klassifikationen mit CORINE-Referenzdaten überschätzt wird, was auch die Karte auf Grundlage der Klassifikation mit dem visuell erstellten Kontrolldatensatz bestätigt.

## 5 Zusammenfassung und Schlussfolgerungen

### Ergebnisse

Insgesamt hat sich gezeigt, dass die angewendete Methodik mit dem zentralen Element der kNN-Klassifizierung für die automatische Erzeugung von Landbedeckungskarten gut geeignet ist. Die Ziele der Arbeit wurden insofern erreicht, als eine Verfeinerung und Aktualisierung des CORINE-

Datensatzes durchgeführt werden konnte. Allerdings war dies nur unter der Voraussetzung der Reduzierung der CORINE-Klassen auf spektral sinnvolle Klassen möglich (siehe Kapitel 2.2). Zudem enthalten die erstellten Karten nach wie vor Mängel und geben die Landbedeckung auf Pixelebene nicht völlig korrekt wieder (siehe beispielsweise Kapitel 4.3). Ohnehin kann eine Klassifizierung z.B. mit dem kNN-Programm immer nur so gut sein, wie die Datengrundlagen und der Referenzdatensatz es ermöglichen.

Mit der „Einschrittmethode“ konnte eine neue Methode der Ausreißerbereinigung entwickelt werden (siehe Kapitel 3.4.1.2), die gute Resultate liefert. Zudem wurde die Anwendung der „Protokollmethode“ eingehender untersucht und unter anderem festgestellt, dass die Aufnahme verbleibender „indifferenten“ (siehe Kapitel 3.4.1.1) in den Referenzdatensatz keinen positiven Effekt auf die Klassifizierung hat.

Darüber hinaus konnte auf Grundlage visueller Interpretation noch eine weitere Methode der Referenzdatenerzeugung entwickelt werden (siehe Kapitel 3.5.1). Die visuell erstellten und flächenproportional vervielfältigten Referenzdaten lieferten gute Resultate bei der Kartenerstellung, welche auch zur Einschätzung der anderen Klassifizierungen nützlich waren.

Ein wichtiges Ergebnis ist, dass die Verwendung von Fernerkundungsdaten eines zweiten Zeitpunktes positive Effekte auf die Klassifizierungsergebnisse hat und so höhere Genauigkeiten und bessere Kartenqualitäten erreicht werden (siehe insbesondere Kapitel 4.3 und 4.4). Zu den Datengrundlagen ist außerdem zu sagen, dass sich die Verwendung von ASTER-Daten mit ihrer recht hohen räumlichen Auflösung von 15 Metern bewährt hat. Der CORINE-Datensatz dagegen hat mehrere Probleme bei der Verwendung als Referenzdatensatz offenbart. Die Übernahme von Klassen in den Referenzdatensatz, die eher Landnutzung als Landbedeckung zuzuordnen oder als Mischklassen einzuteilen sind, schlägt sich negativ auf die produzierten Karten nieder (siehe beispielsweise die Diskussion von Tabelle 6). Zudem wird der Referenzdatensatz durch die Generalisierung im CORINE-Datensatz beeinträchtigt (siehe die Diskussion von Abbildung 22). Des Weiteren zeigen sich Diskrepanzen, wenn Klassifizierungen unter Verwendung von CORINE-Referenzdaten mit visuell interpretierten Kontrolldaten verglichen werden, da letztere sich auf die Landbedeckung beziehen.

Im Laufe der Anwendung von Referenzdatensätzen ist eine mehrschichtige Einschätzung ihrer Qualität unabdingbar. In dieser Arbeit wurden vier Bewertungsmethoden für die Klassifizierungen bzw. die Referenzdaten verwendet. Es fanden Kreuzvalidierungen, die Gegenüberstellung von Klassifizierungen mit visueller Interpretation, visuelle Einschätzungen der produzierten Karten sowie der Vergleich von Flächenbilanzen mit zuverlässigen Quellen statt. Die Ergebnisse der Arbeit deuten darauf hin, dass Kreuzvalidierungen insbesondere ausreißerbereinigter Datensätze zu hohe Genauigkeitswerte liefern und ein visuell erstellter Kontrolldatensatz verlässlichere Aussagen zulässt, was insbesondere dann der Fall ist, wenn wie im Fall von CORINE die Quelle des Referenzdatensatzes ungenau ist (und einen hohen Ausreißeranteil aufweist). Indessen ist eine Betrachtung der produzierten Karten unabkömmlich, besonders weil so Strukturen besser bewertet werden können.

Der Vergleich der beiden Methoden der Ausreißerbereinigung zeigte keine großen Unterschiede. Ein Vorteil der „Einschrittmethode“ ist ihre aufwandärmere Anwendung. Ein Nachteil der Methode ist, dass Referenzpunkte von Klassen, die nicht häufig vorhanden sind, schnell als

Ausreißer eliminiert werden. Dies ist besonders der Fall, wenn die Klasse weniger als  $k$  Referenzpunkte beinhaltet. Ein Nachteil der „Protokollmethode“ könnte sein, dass bei späteren Durchgängen die Ausreißersuche unzuverlässiger wird. Die Anzahl der Indifferenten nimmt im Laufe der Durchgänge immer weiter ab, womit auch die Qualität der Klassifizierung dieses Restdatensatzes abnehmen dürfte. Zur Kombination der zwei Methoden der Ausreißeridentifizierung und -eliminierung ist zu bemerken, dass sie keine besonders guten Ergebnisse erzielte und wohl nicht weiter lohnend ist.

Als Fazit lässt sich festhalten, dass das menschliche Gehirn bei der Kartenerstellung nicht ersetzbar ist. Zum Beispiel werden linienhafte Strukturen wie Straßen, die man mit dem Auge erkennt, bei der Klassifizierung oft nicht durchgehend der visuell erkannten Klasse zugeteilt. Der Mensch verwendet hier regelbasierte Zuteilungen bei Strukturerkennung in der Landschaft, zu welchen das pixelweise arbeitende kNN-Programm nicht in der Lage ist. Das Gehirn mit seiner Fähigkeit zur Integration vielwertiger Informationen ist bislang nicht durch Computer zu ersetzen.

## **Ausblick**

Der abschließende Abschnitt der Arbeit beschäftigt sich mit gegebenenfalls zu empfehlenden weiterführenden Untersuchungen.

In Bezug auf die Auswahl der Daten ist es vielleicht möglich, besser geeignete Fernerkundungsdaten zu finden. Es erscheint sinnvoll, die Auswahl der Zeitpunkte der Fernerkundungsdaten sowie die Anzahl der Zeitpunkte genauer zu untersuchen. Z.B. ist es eventuell von Vorteil, einen Zeitpunkt in den Datensatz zu integrieren, der völlig außerhalb der Vegetationszeit liegt, um den spektralen Kontrast zwischen verschiedenen Klassen mit mindestens zeitweiligem Pflanzenbestand oder auch den Kontrast zwischen immergrünem Nadelwald und Laubwald zu erhöhen. Für die Unterscheidung von Acker und versiegelten Flächen wäre es nützlich, wenn zu mindestens einem Zeitpunkt Vegetation auf den Ackerflächen vorhanden ist, was im Fall dieser Arbeit nicht immer gegeben war. In den Ackerflächen wurden Pixel häufig zu „versiegelten Flächen“ klassifiziert, wenn zu beiden Zeitpunkten in der üblichen Farb-Infrarot-Darstellung graugrünliche Farbwerte, also spärliche Vegetation, vorhanden waren. Für die Unterscheidung von Acker- und Grünland ist wichtig, dass bei den Ackerflächen zu einem Zeitpunkt keine Vegetation vorhanden ist, was wohl z.B. wegen des Anbaus von Winterweizen und Raps durch den Zeitpunkt außerhalb der Vegetationszeit nicht allein gewährleistet wäre. Ein weiterer Referenzzeitpunkt im Hochsommer erscheint daher sinnvoll. Um diesen Anforderung zu genügen, ist eventuell die Verwendung eines dritten Zeitpunktes sinnvoll. Beispielsweise wäre der weitere Zeitpunkt zur Vermeidung der Acker-Siedlungs-Problematik hilfreich. Wenn der zusätzliche Zeitpunkt Vegetation aufweist, ist die Unterscheidung der Klassen durch das kNN-Programm wahrscheinlich gewährleistet, auch wenn zwei Zeitpunkte siedlungsähnliche Spektralwerte liefern. Ein Referenzzeitpunkt könnte also im Juni liegen, wo die Ackerflächen mit großer Wahrscheinlichkeit Vegetation tragen, ein weiterer Ende August und der dritte im November.

Bei der Ausreißersuche wäre es in Erwägung zu ziehen, den „Filterprozess“ in der „Protokollmethode“ zu automatisieren, also die verschiedenen Stufen der „Filterung“ zu verbinden. Es wäre hilfreich, wenn das Programm selbstständig Ausreißer identifizieren und eliminieren könnte bzw. je nach Vorgaben Zusammenstellungen der „guten“, „schlechten“ und „indifferenten“ Referenzpunkte

durchführte. Dazu müssten die „Indifferenten“ einer Filterungsstufe automatisch in die Ausreißeridentifizierung der nächsten Stufe einfließen. Der Prozess wäre insbesondere über die Anzahl der Filterungsstufen sowie über die Kriterien der Ausreißerdefinition, also Häufigkeit der Auswahl als nächster Nachbar sowie „Trefferquote“, zu steuern („Trefferquote“ beschreibt, wie oft berechnete und hier durch CORINE gegebene Klasse übereinstimmen).

Die Arbeit bietet eine Perspektive für die automatische Klassifizierung durch kNN in bislang nicht kartierten Gegenden. Durch visuelle Interpretation besteht die Möglichkeit der Erstellung eines geeigneten Referenzdatensatzes für die Anwendung im kNN-Programm.

In einem Gebiet, zu welchem Satellitenbilder, aber keine Informationen zur Landbedeckung vorliegen, könnte man einen Datensatz erstellen, der pro Klasse eine ausreichende Anzahl an Referenzdaten enthält. Die Klassen wären durch visuelle Interpretation des Gebietes festzulegen. Die Referenzdaten könnten dann entsprechend den Flächenanteilen der Klassen, die beispielsweise in einem Raster abgeschätzt werden, gewichtet werden. Zudem könnte man die Referenzpunkte der Klassen entsprechend den Flächenanteilen der Klassen vervielfältigen, so dass im somit künstlich vergrößerten Datensatz die Anzahlen der Referenzpunkte der verschiedenen Klassen proportional zu den Flächenanteilen der Klassen sind, was der möglichen Dominanz selten vorkommender Klassen entgegenwirken soll (siehe Kapitel 3.5.1). Zur Bewertung eines visuell erstellten, vervielfältigten Datensatzes müsste ein weiterer visuell erstellter Datensatz als Kontrollmittel angelegt werden.

Weiters könnte man zur Aktualisierung von Landbedeckungskarten und zur Beobachtung der Landschaftsdynamik einen festen Raster an Referenzpunkten mit bekannter Landbedeckungsklasse erstellen, wobei diese aktuell gehalten werden muss. Die zu klassifizierenden Bilder müssten räumlich an diesen Raster angepasst werden, um dann ihre spektralen Werte auszulesen und mit dem so gewonnen Referenzdatensatz die Klassifizierung durchzuführen.

Im Zusammenhang mit Klassifizierungen unkartierter Gegenden besteht auch die Überlegung eines universellen Gesamt-kNN-Referenzdatensatzes, der für entsprechende Landbedeckungen jeweilige Standardwerte in den verschiedenen Kanälen beinhaltet, so dass nicht für jedes neue Projekt bzw. auszuwertende Bild ein neuer Referenzdatensatz angelegt werden muss. Wegen der atmosphärischen Einflüsse ist ein solcher Datensatz bislang nicht generell erstellbar (TOMPPPO et. al 2008, d).

## 6 Literaturverzeichnis

- ALBERTZ, J.: Einführung in die Fernerkundung, 3. Aktualisierte und erweiterte Auflage, Wissenschaftliche Buchgesellschaft Darmstadt, 2007, S. 180 ff.
- BAUER, T.: Fernerkundung und Bilddatenverarbeitung, Vorlesung am Institut für Vermessung, Fernerkundung und Landinformation (IVFL) der Universität für Bodenkultur Wien, 2008
- BAUERHANSL, C.: Waldkarten aus Satellitenbildern, Bundesforschungs- und Ausbildungszentrum für Wald, Naturgefahren und Landschaft (BFW) (Hg.), Jahresbericht 2008, Wien, 2009, S. 10
- BOSSARD, M.; Feranec J. und Otahel, J.: Technical report No 40-CORINE land cover technical guide-Addendum 2000, European Environment Agency, Copenhagen, 2000
- BROCKHAUS, Enzyklopädie in 30 Bänden, 21., völlig neu bearbeitete Auflage, Redaktionelle Leitung: Zwahr, A., F.A. Brockhaus GmbH, Leipzig, 2006, a) Band 15: Kind bis Krus, Stichwort Klassifikation, S. 112, b) Band 1: A bis Anat, Stichwort Algorithmus, S. 526, c) Band 2: Anau bis Ausf, Stichwort Ausreißer, S. 756
- DIGITALES OBERÖSTERREICHISCHES RAUM-INFORMATION-SYSTEM, DORIS interMAP: <http://doris.ooe.gv.at/viewer/%28S%28sblzh545wx0ww145efjty2r%29%29/init.aspx?ks=alk&karte=orthofotos>, 2010
- FERNERKUNDUNGSLEXIKON: Stichwort Klassifizierung, <http://www.fe-lexikon.info/lexikon-k.htm#klassifizierung>, 2010
- <http://www.invekos.at> - besucht am 15.04.2010
- Jensen, J.R.: Remote sensing of the environment- an earth resource perspective, 2nd edition, Prentice-Hall Inc., Upper Saddle River, NJ., 2007, p. 24
- KLUGE ETYMOLOGISCHEN WÖRTERBUCHS DER DEUTSCHEN SPRACHE, Walter de Gruyter GmbH & Co. KG, Berlin, 2002
- KOUKAL, T.: User Guide to the kNN-Program, Institut für Vermessung, Fernerkundung und Landinformation (IVFL) der Universität für Bodenkultur Wien, 2009
- KRAUS, K. und Schneider, W.: Fernerkundung, Bd.1, Physikalische Grundlagen und Aufnahmetechniken, Ferd. Dümmlers Verlag Bonn, 1988
- MITCHELL, T.M.: Machine Learning, McGraw Hill, 1997, p. 231 ff.
- NIEDERÖSTERREICH-ATLAS, NÖatlas:: <http://www.intermap1.noel.gv.at/webgisatlas/%28S%28iwcclk45dqfusk55o3kdr055%29%29/init.aspx>, 2010
- ÖSTERREICHISCHES BUNDESUMWELTAMT: CORINE-Nomenklatur, [http://www.umweltbundesamt.at/fileadmin/site/umweltthemen/raumplanung/1\\_flaechennutzung/CORINE/CORINE\\_Nomenklatur.pdf](http://www.umweltbundesamt.at/fileadmin/site/umweltthemen/raumplanung/1_flaechennutzung/CORINE/CORINE_Nomenklatur.pdf), 2010
- POPOVICI, P.: Landbedeckungsklassifizierung aus Satellitenbildern am Beispiel Niederösterreich, Diplomarbeit am Institut für Vermessung, Fernerkundung und Landinformation (IVFL) der Universität für Bodenkultur Wien, 2008, S. 24 f.
- Stümer, W.; Köhl, M.: Kombination von terrestrischen Aufnahmen und Fernerkundungsdaten mit Hilfe der k-Nächste-Nachbar-Methode zur Klassifizierung und Kartierung von Wäldern, Photogrammetrie; Fernerkundung;, Geoinformation, Heft 1, 2005, Seiten 23-36
- TOMPO, E.; OLSSON, H.; STÄHL, G.; NILSSON, M.; HAGNER, O.; and KATILA, M.: Combining national forest inventory field plots and remote sensing data for forest databases, Remote Sensing of Environment, Volume 112, Issue 5, 15 May 2008, Pages 1982-1999, a) p. 1987, b) p. 1992, c) p. 1989, d) p. 1996

## Tabellenverzeichnis

Tabelle 1: Überarbeitung der vorkommenden CORINE-Klassen.....	17
Tabelle 2: Beispiel einer Fehlermatrix .....	31
Tabelle 3: Genauigkeitswerte der Klassifizierung aus der Gegenüberstellung mit der visuellen Interpretation .....	40
Tabelle 4: Flächenanteilsvergleich von CORINE mit InVeKoS und „Waldkarte“-Daten.....	54
Tabelle 5: Flächenanteilsvergleich einer Klassifizierung mit „Waldkarte“- und InVeKoS-Daten im Bezirk Freistadt .....	55
Tabelle 6: Flächenanteile von CORINE und einer Klassifizierung im Untersuchungsgebiet .....	56
Tabelle 7: Flächenanteilsvergleich von CORINE und einer Klassifizierung im Bezirk Linz (Stadt)...	58

## Abbildungsverzeichnis

Abbildung 1: Lage der Augustszenen in Österreich .....	9
Abbildung 2: CORINE-Klassen (ÖSTERREICHISCHES UMWELTBUNDESAMT 2010) .....	11
Abbildung 3: kNN-Methode für ein Beispiel mit $k = 3$ (POPOVICI 2008) .....	19
Abbildung 4: ASTER-Ausschnitt der Aprilaufnahme zur Veranschaulichung visuellen Interpretierens .....	23
Abbildung 5: ASTER-Ausschnitt der Augustaufnahme zur Veranschaulichung visuellen Interpretierens .....	24
Abbildung 6: Orthofoto zur Veranschaulichung visuellen Interpretierens .....	25
Abbildung 7: Ablaufschema des kNN-Programms (KOUKAL 2009) .....	27
Abbildung 8: Schema der Auswertungsvarianten .....	29
Abbildung 9: Suche künstlicher Ausreißer im bereinigten Datensatz mit der „Protokollmethode“ .....	35
Abbildung 10: Suche künstlicher Ausreißer im unbereinigten Datensatz mit der „Protokollmethode“ .....	36
Abbildung 11: Vergleich von Klassifizierungen von mono- und bisaisonalen Datensätzen mit und ohne Ausreißerbereinigung durch Gesamtgenauigkeiten der Kreuzvalidierungen .....	38
Abbildung 12: Methodenvergleich der Ausreißerbereinigung anhand von Gesamtgenauigkeiten aus Kreuzvalidierungen bisaisonaler Referenzdaten .....	39
Abbildung 13: Gesamtgenauigkeit verschiedener Klassifikationen auf Grundlage bisaisonaler (teils ausreißerbelassener, teils mit der „Einschrittmethode“ ausreißerbereinigter) Fernerkundungsdaten im Vergleich mit der visuellen Interpretation .....	41
Abbildung 14: Vergleich von Ausreißerbereinigungsmethoden anhand der Gegenüberstellung von visueller Interpretation mit verschiedenen je mit $k = 9$ klassifizierten, bisaisonalen und ausreißerbereinigten Datensätzen .....	42
Abbildung 15: Produzentengenauigkeiten der Klassen "Laubwald" und "Grünland" aus Gegenüberstellungen von Klassifizierungen mit der visuellen Interpretation .....	43
Abbildung 16: Benutzergenauigkeiten der Klassen "Laubwald" und "Grünland" aus Gegenüberstellungen von Klassifizierungen mit visueller Interpretation .....	44
Abbildung 17: ASTER-Bildausschnitt der VNIR-Kanäle mit Linz im Nordwesten .....	46
Abbildung 18: CORINE-Kartenausschnitt mit Linz .....	47
Abbildung 19: Linz enthaltender Ausschnitt einer Klassifizierung mit 13 Klassen auf Grundlage eines bisaisonalen, ausreißerbereinigten Datensatzes .....	48
Abbildung 20: Vergleich der Karten aus verschiedenen Klassifizierungen .....	49
Abbildung 21: Wirkung des $k$ -Wertes auf die Klassifizierung .....	51
Abbildung 22: Vergleich einer Karte auf Grundlage der CORINE-Daten (durch die „Protokollmethode“ ausreißerbereinigt) mit einer Karte auf Grundlage eines visuell erstellten, „vervielfältigten“ Datensatzes .....	52
Abbildung 23: Klassifizierung des gesamten Untersuchungsgebietes .....	53