

Universität für Bodenkultur Wien
Studienrichtung Landschaftsplanung und Landschaftspflege

Diplomarbeit

Landbedeckungsklassifizierung aus Satellitenbildern am Beispiel Niederösterreich

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

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

unter Mitbetreuung von
Dipl.-Ing. Dr.nat.techn. Tatjana Koukal
Mag. Franz Suppan

eingereicht von
Paula Popovici

Wien, November 2008

Danksagung

An dieser Stelle möchte ich mich bei meiner Familie bedanken, die mich darin bekräftigt hat mein Studium in Österreich zu machen. Der Dank gilt meiner Schwester Felicia, meiner Mutter und meinem Vater.

Herzlicher Dank gebührt ganz besonders meinen Betreuern Univ.Prof. Dipl.-Ing. Dr.techn. Werner Schneider, Dipl.-Ing. Dr.nat.techn. Tatjana Koukal und Mag. Franz Suppan, die mich sehr beim Schreiben und Durchführen meiner Diplomarbeit unterstützt haben.

Nicht zuletzt bedanke ich mich bei allen Mitarbeitern und Diplomanden des Instituts für Vermessung, Fernerkundung und Landinformation für ein sehr freundliches Arbeitsklima. Auch standen sie mir stets mit Rat und Tat zur Seite. Weiters bedanke ich mich bei all meine Studienkollegen und Freunden, die mich während meiner Studienzeit begleitet haben.

Kurzfassung

Das Ziel dieser Diplomarbeit ist die Entwicklung eines Verfahrens zur Aktualisierung von bestehenden Landbedeckungsdatensätzen unter Verwendung aktueller Satellitenbilder. Es wird ein Klassifizierungsverfahren nach der Methode der k nächsten Nachbarn (kNN) eingesetzt.

Die Satellitenbilder werden unter Verwendung von Referenzdaten aus einer veralteten Landbedeckungskarte klassifiziert. Das Klassifizierungsverfahren erkennt Inkonsistenzen zwischen den spektralen Charakteristiken der Satellitenbilder und den Referenzdaten und kann damit jene Anteile der Referenzdaten eliminieren, die fehlerhaft oder überholt sind.

Das Verfahren wurde auf die Landbedeckungsdatensätze SINUS und CORINE angewendet und zeigt gute Ergebnisse. Als Anwendungsbeispiel wird eine Landbedeckungskarte von Niederösterreich mit 13 Landbedeckungskategorien erzeugt.

Abstract

The goal of this diploma thesis is the development of a method for the actualization of existing land cover datasets using current satellite images. For this, the kNN (k nearest neighbours) method was used as a classification method.

The satellite images are classified using reference data from outdated land cover maps. The classification procedure is able to detect inconsistencies between the spectral characteristic of the satellite images and the reference data and can thereby eliminate those portions of the reference data, which are incorrect or out of date.

This method was applied to the land cover datasets SINUS and CORINE and shows good results. As a practical example, a land cover map of Lower Austria with 13 land cover categories was generated.

Inhaltsverzeichnis

1	Einleitung	3
1.1	Motivation	3
1.2	Ziel der Arbeit	4
1.3	Definition: Landbedeckung und Landnutzung	5
2	Klassifikationsverfahren	7
2.1	Klassifikationen im Allgemeinen	7
2.2	Klassifikationen in der Fernerkundung	8
2.3	kNN-Methode	10
3	Untersuchungsgebiet Niederösterreich	12
4	Datengrundlage	14
4.1	LANDSAT	14
4.2	Digitales Höhenmodel-SRTM	17
4.3	Referenzdatensätze	18
4.3.1	SINUS-Datensatz	18
4.3.2	CORINE Landcover	20
4.3.3	Visuelle Interpretation	23
5	Vorarbeiten	24
5.1	Bildvorverarbeitung allgemein	24
5.2	Georeferenzierung der Szenen	25
5.3	Radiometrische Korrektur	27
5.3.1	Luftlichtkorrektur	28
5.3.2	Topographische Korrektur	28
5.4	Erstellen einer Bildmaske	30
6	Durchführung der Klassifikation	32
6.1	Erstellen eines Rasters von Referenzpunkten	32
6.2	Visuelle Interpretation	33
6.3	Übernahme der Landbedeckungsklasse aus SINUS und CORINE	34
6.3.1	SINUS	34
6.3.2	CORINE	35
6.4	kNN-Methode	35
6.4.1	Modul Readimage	35

6.4.2	Modul Crossvalidation	36
6.4.3	Modul Mapmaker	37
6.5	Fehlerquellen	37
6.6	Methoden zur Reduzierung von Klassifikationsfehlern	38
6.6.1	Optimierung der kNN-Parameter	38
6.6.2	Bereinigen der Referenzdatensätze	40
	Verschieben von Referenzpunkten	40
	Eliminierung von Ausreißern	41
7	Ergebnisse	44
7.1	Karten	44
7.1.1	Klassifikation mit Referenzpunkten aus CORINE	44
7.1.2	Klassifikation mit Referenzpunkten aus SINUS	49
7.1.3	Bereinigen der Referenzdatensätze	53
	Verschieben von Referenzpunkten	53
	Eliminierung von Ausreißern	53
7.2	Fehlerabschätzung	54
7.2.1	Kontrollpunkte und Kreuzvalidierung	54
7.2.2	Rasterweiten und Punkteanzahl	55
7.2.3	Eliminierung von Ausreißern	57
8	Diskussion und Schlussfolgerungen	59

1 Einleitung

1.1 Motivation

Die Fernerkundung hat sich in den letzten Jahren rasant entwickelt und an Bedeutung zugenommen. Das Angebot an Fernerkundungsdaten ist stark gestiegen. In vielen Bereichen ist heutzutage die Gewinnung von Informationen über Erdoberfläche, Atmosphäre oder Lithosphäre ohne Fernerkundung schwer vorstellbar. So finden die Daten, meist in Form von Bildern mit einer großen Fülle von Informationen über das abgebildete Gelände, verschiedenste Anwendungen in Geographie, Geologie, Kartographie, Katastrophenschutz, Naturschutz, Archeologie, etc. Sie liefern unter anderem folgende Informationen:

- Höhenrelief und Wasserwege
- Gesteinstypen, Lagerstätten
- Landbedeckung und Landnutzung
- Urbanisierung (Ausbreitung der Städte)
- Forstinventurdaten, Holzpotentialabschätzungen, Waldschadenskartierung, Forstwegenetze, etc.
- landwirtschaftliche Erntevorhersagen, Anbauflächen, Grundlagen für die Überprüfung von subventionierten Brachflächen, precision farming, etc.
- Vegetationsphänologie (Aspektfolge)
- Bestandsaufnahme von Natur und Landschaft (Ist-Zustandserhebung), Flächenbewertungen und Visualisierungen für die Landschaftsplanung

Während vor Jahren die Kartierung der Bodenoberfläche nur durch Geländebegehungen möglich war, ist heutzutage die Kartierung von Landbedeckung und Landnutzung mit Hilfe von satellitengestützten, optischen Fernerkundungssensoren möglich. Diese Sensoren liefern in regelmäßigen Zeitabständen Informationen über spektrale Eigenschaften der Erdoberfläche, die mittels entsprechender Fachkenntnisse in Informationen zur Landbedeckung beziehungsweise Landnutzung übersetzt werden können. Auch die Methoden für die Auswertung der Fernerkundungsdaten haben sich weiterentwickelt. Während diese traditionell durch visuelle Bildinterpretation erfolgt ist, wurden seit einiger

Zeit zahlreiche rechnergestützte Methoden entwickelt, mit deren Hilfe Landnutzungsinformationen aus digitalen Fernerkundungsdaten weitaus effizienter abgeleitet werden können. Inzwischen stehen eine große Anzahl von Methoden der Bildbearbeitung und Bildklassifizierung zur Verfügung.

Zu den Stärken der Fernerkundung zählen unter anderem: Erfassung großer Flächen und somit eine zeitliche Einsparung und eine verbesserte Kosteneffizienz, höhere Reproduzierbarkeit von Erhebungen und langfristige Dokumentationen von Zeitschnitten, die jederzeit neu ausgewertet werden können.

Generell liefern Landbedeckungsdatensätze elementare Informationen zur wissenschaftlichen Beurteilung der Interaktionen zwischen Mensch und Umwelt.

Das Wissen um die Beschaffenheit und die Veränderungen der Landbedeckung ist für viele umweltrelevante Fragestellungen von essenzieller Bedeutung. Aktuelle Landbedeckungskarten stellen somit eine wichtige Datengrundlage dar. Trotz des hohen Bedarfs an zuverlässigen, aktuellen Informationen zur Landbedeckung existieren in Europa bisher keine Landbedeckungsdatensätze, die regelmäßig mit verhältnismäßig geringem Arbeitsaufwand aktualisiert werden können. (Muhar, 2003)

Derzeit sind vor allem waldbezogene Daten verfügbar, zukünftig aber wäre zu prüfen, ob eine effiziente Landschaftsinventur eingerichtet werden kann, die die tatsächliche Landnutzung, die Nutzungsintensität, Nutzungsüberlagerungen, die Zugänglichkeit, den Versiegelungsgrad usw. erhebt. (Muhar, 2003)

1.2 Ziel der Arbeit

Die vorliegende Diplomarbeit geht vom Bedarf an Landnutzungskarten und deren laufender Aktualisierung aus.

Als Methode für die weitgehende automatisierte Ausstellung von Landbedeckungskarten aus multispektralen Satellitenbildern hat sich in vorangehenden Arbeiten am Institut für Vermessung, Fernerkundung und Landinformation der Universität für Bodenkultur die kNN-Methode bewährt. Die vorliegende Arbeit hat zum Ziel, diese Methode weiterzuentwickeln und insbesondere ihre Brauchbarkeit für die automatisierte Aktualisierung bestehender Landbedeckungskarten zu analysieren. Dazu soll die Robustheit der kNN-Methode gegenüber Fehlern und Abweichungen (Ausreißern) im Referenzdatensatz ausgenutzt werden.

Die kNN-Methode hat das Potenzial, Inkonsistenzen zwischen den spektralen Charakteristiken der Satellitenbilder und den Referenzdaten zu erkennen und damit jene Anteile der Referenzdaten zu eliminieren, die fehlerhaft oder überholt sind.

Dieses Verfahren soll für die Aktualisierung der bestehenden Datensätze SINUS (aus dem Programm “Kuturlandschaftsforschung“ (Hollaus und Suppan, 2003) und CORINE (Hözl, 2003) adaptiert und am Beispiel Niederösterreich angewendet werden.

1.3 Definition: Landbedeckung und Landnutzung

Bevor auf das Thema Klassifikation näher eingegangen wird, ist es notwendig, die Begriffe Landbedeckung und Landnutzung zu diskutieren, da es nicht selten zu einer Verwechslung dieser zwei Termini kommt.

Als **Landbedeckung** (land cover) wird die Erdoberflächenbedeckung bezeichnet. Wesentliche Elemente sind die Vegetation, Gewässer, Eis, Felsen, Sand sowie die künstliche Bodenbedeckung wie Gebäude und Strassen. Landbedeckung ist eines der wichtigsten Elemente zur Beschreibung und Untersuchung der Umwelt.

Landnutzung (land use) wird als eine Verbindung zwischen der Landbedeckung und den Aktivitäten der Menschen in ihrer Umwelt verstanden. Bei der Landnutzung erfolgt die Beschreibung in Bezug auf die Funktion der Landbedeckung, der sozioökonomischen Nutzung der Fläche, wie zum Beispiel landwirtschaftliche Flächen, Naturschutzflächen oder Wohnbauflächen und Gewerbeflächen. Die Landnutzung kann nur teilweise, oft gar nicht unmittelbar beobachtet werden. Beispielsweise ist Wiese eine Landbedeckung, Weideland oder Tennisplatz bezeichnen wiederum die Landnutzung auf der Fläche. Erholungsgebiet ist eine Landnutzungsform; diese kann mehreren Landbedeckungstypen zugeordnet werden, wie beispielsweise Sand (Strand), Wälder, Wassergebiete (See, Fluss), etc. (Radberger, 2001)

Die vorhandenen Landbedeckungsdatensätze enthalten sowohl Landbedeckungskategorien als auch Landnutzungskategorien, weil in der Praxis oft eine eindeutige Trennung zwischen den zwei Kategorien nicht möglich ist.

In Österreich stehen mittlerweile eine Anzahl an Landbedeckungskarten zur Verfügung. Dazu zählen SINUS (Landbedeckungsdatensatz aus der Kulturlandschaftsforschung) (Hollaus und Suppan, 2003), Landinformationssystem Österreich (Unterscheidung zwischen Siedlungsraum, Agrarraum, Waldflächen und alpines Höhenstockwerk) (Segger, 2003), Datensätze zur Naturraumkartierungen in Oberösterreich (beauftragt von der Naturschutzabteilung der oberösterreichische Landesregierung, bilden wichtige Zusatzinformationen zur naturräumlichen Gesamtsituation des Landes) (Dorninger, 2003) und auch die amtlichen Karten, der Grundstückskataster, Flächenwidmung- und Bebaungspläne, Daten über Gefahrenzonen, Deponien, Schutzgebiete oder die Umweltsituation. (Grillmayer und Schneider, 2003)

Leider gibt es europaweit nicht allzu viele allgemeine Landbedeckungskarten. In den meisten Fällen sind sie themenspezifisch. Als einziger ganzflächiger Landbedeckungsdatensatz, der europaweit erstellt wurde, steht CORINE Land Cover zur Verfügung, der nicht nur die Mitgliedstaaten der europäischen Union, sondern auch die osteuropäischen und nordafrikanischen Staaten abdeckt. An themenspezifischen Karten gibt es Waldbedeckungskarten, Brandgefahrenkarten usw.. (Hölzl, 2003)

Weltweit steht GLC2000 (Global Land Cover, 22 detaillierte Bodenbedeckungstypen) zur Verfügung. (<http://www-tem.jrc.it/glc2000/>)

2 Klassifikationsverfahren

2.1 Klassifikationen im Allgemeinen

Klassifizierung (von lat. classis, „Klasse“, und facere, „machen“) nennt man das Zusammenfassen von Objekten oder Begriffen, die ähnliche Merkmale besitzen, zu Klassen (Kategorien, Gruppen, Mengen). Die Gesamtheit aller Klassen bildet eine Klassifikation.

Die Klassifikation dient als konzeptionelle Grundlage für anschließende Analysen.

Bei der Aufstellung einer Klassifikation müssen einige Grundprinzipien beachtet werden: die Bildung der Klassen muss anhand derselben Klassifikationsmerkmale erfolgen, die Klassen müssen einander ausschließen, das heißt, jedes Objekt darf nur einer einzigen Klasse angehören, und der Umfang der Klassen insgesamt muss gleich dem Umfang der zu klassifizierenden Gesamtheit sein (jedes Objekt muss einer Klasse zugeordnet werden). (Brendler, 2004)

Ein erster Schritt bei der Erstellung einer Klassifikation ist die Klassenbildung. Dann folgt die Einteilung der Objekte in die Klassen. Dieser Prozess kann nie fehlerfrei erfolgen. Gründe dafür sind, unter anderen:

- **Undeutliche Kriterien und unscharfe Grenzen:** Oft ist es unklar, wann ein Objekt zu einer Klasse gehört und wann nicht. Daher wird es schwierig, dieses Objekt richtig zu klassifizieren. Beispiele sind: ab wie viel Prozent Vegetationsanteil gehört ein Pixel zu einer „Vegetationsklasse“ (Grünland, Wald, Acker) oder wie viel Anteil an „versiegelter Fläche“ muss ein Pixel haben, um es einer Siedlungskategorie zuzuordnen?
- **Falsche Merkmale:** Objekte können nur dann in Klassen eingeordnet werden, wenn die betrachteten Merkmale tatsächlich eine Unterscheidung der Klassen ermöglichen. So ist es schwierig zu beurteilen, ob anhand der Blattfarbe einer Pflanze diese der Klasse „Baum“, „Strauch“ oder „einjährige Pflanze“ zugeordnet wird: die Blattfarbe hat im Allgemeinen keinerlei Aussagekraft über die Zugehörigkeit einer Pflanze zu diesen Klassen.
- **Nichttrennbarkeit:** Wenn zu wenige oder nichtssagende Merkmale betrachtet werden, erscheinen die Objekte durcheinander gemischt und es wird keine klare Trennung möglich. Möchte man etwa anhand der Form, der Größe und der Vegetationsbedeckung einen Spielplatz von einem Ackerfeld unterscheiden, so könnten

sich Spielplätze und Ackerfelder in diesen Merkmalen ähnlich sein und eine eindeutige Trennung nahezu unmöglich machen.

- **Ausreißer:** Messfehler oder ungewöhnlich ausgeprägte Einzelexemplare können dazu führen, dass ein Objekt falsch klassifiziert wird.
- **Restobjekte:** Bei der Klasseneinteilung kann eine Gruppe von Restobjekten übrig bleiben, die sich in keine der bestehenden Klassen einordnen lässt, und für die sich auch keine neue Klasse schaffen lässt. Oft muss eine Kategorie, in der alle diese Objekte erfasst werden, wie zum Beispiel „nicht klassifizierbar“, festgelegt werden.
- **Vereinfachung:** Beim Klassifizieren handelt es sich um eine Zusammenfassung und Vereinfachung der komplexen Sachverhalte der Natur. Daher gibt es keine 100%ig richtige Klassifizierung.

Selbst wenn alle Merkmale eines Objektes bekannt sind, kann es unter Umständen falsch klassifiziert werden. Wenn Fehler in der Vorgehensweise anzunehmen sind und man sich nicht sicher über die richtige Zuordnung ist, empfiehlt es sich, jeder Entscheidung eine Angabe über ihre Zuverlässigkeit beizufügen. (Wikipedia, am 12.08.08)

2.2 Klassifikationen in der Fernerkundung

Für viele Anwendungen von Satellitenbildern wird Information in Form einer Klassifikation aus den Bildern gewonnen. Vielfach wird pixelweise klassifiziert. Dabei wird jedem Pixel eines Bildes eine bestimmte Klasse zugeordnet. Die Kategorien stellen thematische Informationen dar. Das Ergebnis der Klassifizierung ist eine thematische Karte.

Satellitenbilder können mit verschiedenen Methoden klassifiziert werden, von der visuellen Interpretation bis zu automatisierten Verfahren.

Die automatisierte Klassifizierung kann geometrisch als eine Aufteilung des Merkmalsraums gesehen werden. „Ein Merkmalsraum ist der n-dimensionale Raum, der aufgespannt werden kann, wenn zu einer Klassifizierung n Merkmale herangezogen werden“ (Geoinformatik, am 09.10.08). Der Merkmalsraum in der Bilddatenklassifizierung ist von der Anzahl der verwendeten Spektralbänder bestimmt. Jedes Bildpixel kann mithilfe seiner Pixelwerte, die einen Merkmalsvektor bilden, in diesem Raum eindeutig eingeordnet werden. (Glossardatenbank, am 09.10.08) Die automatisierte Klassifizierung wird in unüberwachte und in überwachte Klassifizierung unterteilt.

Bei der unüberwachten Klassifizierung („**Clusteranalyse**“) wird die Gesamtheit der Bildelemente in eine Anzahl von Klassen unterteilt, die jeweils Bildelemente ähnlicher

spektraler Eigenschaften enthalten. Dabei macht man sich keine Gedanken über die Definition der Klassen. Die Unterteilung erfolgt automatisch. Die Bedeutung der einzelnen Klassen wird nachträglich durch Interpretation bestimmt.

Für die Durchführung der überwachten Klassifizierung werden Referenzdaten (in Form von „Trainingsgebieten“) gebraucht. Zunächst werden die gewünschten Klassen vorgegeben (definiert). Als nächstes werden in den Satellitendaten Pixel ausgewählt, die als Stichprobe zur Beschreibung der Klasse dienen. Die Auswahl dieser „Trainingsgebiete“ erfordert meist eine gewisse Erfahrung des Anwenders und Informationen über das Gelände. Danach wird jedes Pixel des gesamten Satellitenbildes mit den spektralen Charakteristiken der Trainingsgebiete verglichen und einer Klasse zugeordnet. (Albertz, 2007)

Für die überwachte Klassifizierung stehen verschiedene Algorithmen zur Verfügung, unter anderen: Minimum-Distance-Verfahren, Maximum-Likelihood-Verfahren, künstliche neuronale Netze und die k-Nearest-Neighbour-Methode. Die Auswahl einer geeigneten Methode obliegt dem Anwender. Hier spielen mehrere Faktoren eine Rolle, wie z.B.: Art der gewünschten Klassen und aussagekräftige Bildmerkmale zur Klassifizierung (spektrale Merkmale sind eher für automatisierte Verfahren geeignet als Textur-, Struktur- oder Kontext-Merkmale), vorhandene Erfahrungen, finanzielle Mittel, Verfügbarkeit von Daten, Geländebeschaffenheit des Gebietes (großflächige, reliefarme Landwirtschafts- oder Waldgebiete sind für automatisierte Verfahren geeignet, während inhomogene, komplexe, stark variierende Gebiete wie Städte oder Gebirgsformationen größere Probleme bereiten).

Das **Minimum-Distance-Verfahren** ist einfach und erfordert nur einen geringen Rechenaufwand. Zunächst werden für die Trainingsgebiete jeder Objektklasse die Mittelwerte in den einzelnen Spektralkanälen berechnet. Ein Bildelement wird derjenigen Klasse zugewiesen, deren Mittelpunkt im Merkmalsraum am nächsten liegt, also den minimalen Abstand aufweist. Das Verfahren hat den Nachteil, dass die unterschiedlichen Streubereiche der Merkmalswerte nicht berücksichtigt werden. So kann ein Pixel fälschlicherweise einer Klasse zugeordnet werden, der es mit großer Wahrscheinlichkeit nicht angehört. (Albertz, 2007)

Das **Maximum-Likelihood-Verfahren** (Verfahren der größten Wahrscheinlichkeit) setzt voraus, dass die Pixelwerte jeder Klasse im Merkmalsraum normalverteilt sind. Diese Normalverteilungen können mit Flächen gleicher Wahrscheinlichkeit, sogenannten Mahalanobis-Ellipsoiden, veranschaulicht werden. Die Verteilungsparameter werden für jede Objektklasse berechnet. Für jedes zu klassifizierende Pixel wird die Zugehörigkeitswahrscheinlichkeit zu jeder Klasse berechnet. Das Pixel wird der Klasse mit der größten Zugehörigkeitswahrscheinlichkeit zugewiesen. Diese Methode ist die genaueste unter den parametrischen Verfahren (Albertz, 2007)

Mit **künstlichen neuronalen Netzen** wird versucht, das Verhalten natürlicher Neuronen im menschlichen Gehirn mit Computern zu simulieren. Klassifikationsergebnisse

zeigen, dass neuronale Netze in der Lage sind, Fernerkundungsbilder effizient zu klassifizieren. Vorteilhaft wirkt sich dabei aus, dass die Einstufung der Merkmale parameterfrei erfolgt und keine Annahme über die Datenverteilung notwendig ist. Aufgrund dieser Tatsache ist auch die Einbeziehung von Umgebungsinformation benachbarter Bildpunkte in die Bewertung leicht zu implementieren (Bischof u. a., 1992)

Zu den überwachten parameterfreien Klassifizierungsverfahren gehört auch die **kNN-Methode**, die im folgenden Abschnitt genauer beschrieben wird.

2.3 kNN-Methode

Das k-Nächste-Nachbarn-Verfahren ist, so wie die Methode der künstlichen neuronalen Netze, ein nicht parametrischer Algorithmus. Gleich wie bei den künstlichen neuronalen Netzen werden hier keine Annahmen über die Verteilungen der Merkmale in den einzelnen Klassen getroffen. Die Klassifikation erfolgt mittels einer Abstandsfunktion (z.B. Euklidische Distanz) im Merkmalsraum, wobei die k nächsten Referenzpixel gesucht werden. k stellt die Anzahl der Nachbarn dar, die betrachtet werden und wird vom Anwender vorgegeben. Das zu klassifizierende Bildpixel wird jener Klasse zugewiesen, in der sich die meisten Nachbarn befinden. (Mitchell, 1997)

In Abbildung 2.1 ist zu sehen, dass es beim k-Nächste-Nachbarn-Verfahren durchaus vorkommen kann, dass das Ergebnis je nach gewähltem k-Wert unterschiedlich sein kann: Im Beispiel von Abbildung 2.1 würde k=1 die Klasse B ergeben, während k=3 die Klasse A ergibt.

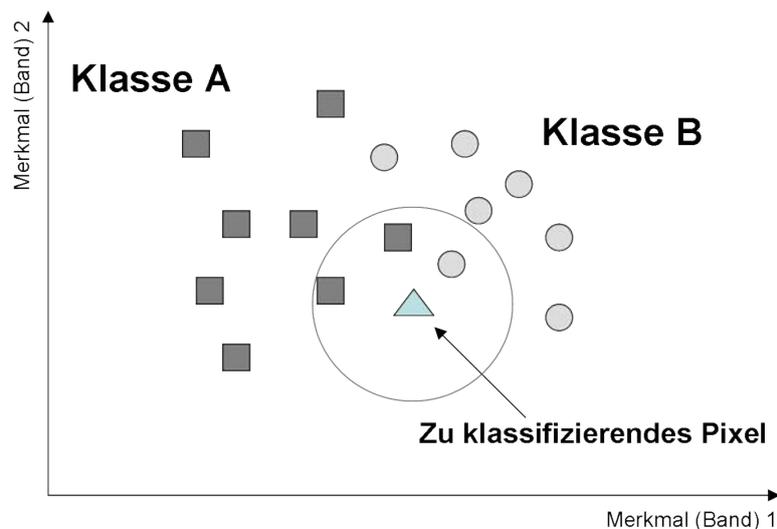


Abbildung 2.1: kNN-Methode für Beispiel k=3

Die Ergebnisse, die mit einer kNN-Klassifizierung erzielt werden, sind stark von der Optimierung der kNN-Parameter abhängig. So kann man z.B. diejenigen Spektralkanäle stärker gewichten, in denen der Unterschied der spektralen Signatur zwischen den einzelnen Objektklassen größer ist. Wenn man beispielsweise eine Differenzierung von Vegetationstypen anstrebt, ist es sinnvoll, die Infrarot-Spektralkanäle stärker zu gewichten, weil in diesen Kanälen die einzelnen Vegetationsklassen besser trennbar sind. Auch eine Gewichtung der nächsten Nachbarn kann empfehlenswert sein: Den näherliegenden Nachbar kann man eine größere Aussagekraft für das zu klassifizierende Pixel geben als den weiter entfernt liegenden.

In Vergleich zu parametrischen Verfahren wird bei dieser Methode eine größere Menge an Referenzpunkten benötigt, damit auch Klassen mit einem großen Streubereich gut repräsentiert sind.

Eine Klasse kann auch auf mehrere Punktwolken über den ganzen Merkmalsraum verteilt sein. Wenn eine Klasse aus Subklassen besteht (wie z.B. die Klasse „Acker“ aus „grüner Ackervegetation“, „offener Boden“ usw.), so kann sie bei Verwendung der kNN-Methode trotzdem von Anfang an als eine einzige Klasse gesehen werden.

Beim Einsatz der kNN-Methode wird zunächst eine Kreuzvalidierung durchgeführt (siehe Kap. 6.4.2). Mit Hilfe der Kreuzvalidierung kann die Eignung der gegebenen Merkmale für die vorgesehene Klassifizierung an Hand des Referenzdatensatzes beurteilt werden. Es wird durch die Kreuzvalidierung sichtbar, welche Klassen sich überschneiden, bei welchen die Wahrscheinlichkeit einer Fehlklassifizierung höher ist. Ebenso können fehlerhafte Referenzdaten (Ausreißer) aus dem Referenzdatensatz eliminiert werden. Bei den anderen Verfahren wäre das nicht auf diese Weise möglich, da dort die Klassen in ihrer Gesamtheit betrachtet werden, das heißt z.B. durch ihre Verteilungsparameter und nicht durch die einzelnen Referenzpunkte beschrieben werden.

Es besteht damit die Möglichkeit, den Referenzdatensatz nachträglich zu verbessern. Im Rahmen dieser Arbeit wird auch untersucht, inwieweit Referenzdatensätze von nicht aktuellen Landbedeckungskarten für eine weitere Klassifikation verwendet werden können. Die Erdoberfläche ist ständigen Veränderungen unterworfen: Waldflächen werden kahlgeschlagen, Siedlungsgebiete wachsen, Brachen oder neue Anbauflächen entstehen. Das bedeutet, dass Klassenzuordnungen, die aus Referenzdaten veralteter Landbedeckungskarten stammen, bei manchen Referenzpunkten falsch sind. Die Referenzpunkte können bei der Kreuzvalidierung identifiziert und anschließend entfernt werden. Damit ergibt sich die Möglichkeit zur Aktualisierung des veralteten Landbedeckungsdatensatzes.

3 Untersuchungsgebiet Niederösterreich

Lage

Niederösterreich ist mit 19 177km² das größte Bundesland Österreichs. Es umgibt Wien und liegt zu beiden Seiten der Donau zwischen der Mündung der Enns im Westen und der March im Osten; grenzt im Westen an Oberösterreich, im Süden an die Steiermark, im Südosten an das Burgenland, im Nordosten an die Slowakei und im Norden an Tschechien. Geographisch ist Niederösterreich in 4 Regionen aufgeteilt: Weinviertel (Nordosten), Waldviertel (Nordwesten), Industrieviertel (Südosten), Mostviertel (Südwesten). (Österreich Lexikon AEIOU, am 24.07.08)

Landschaftsbild

Niederösterreich erstreckt sich vom Böhmischem Massiv¹ im Norden, bis zu den Alpen im Süden. Dazwischen liegt das Alpenvorland. Im Osten breiten sich das Wiener Becken (Industrieviertel) und das Marchfeld aus. Nördlich der Donau erstrecken sich wellige Hügel von 500-800 m Höhe mit Wäldern, Mooren und Teichen (Waldviertel), während im Mostviertel die Ausläufer der Kalkalpen mit Berghöhen bis 2.000 m üNN das Landschaftsbild prägen. Nach Osten ist das Böhmisches Massiv durch die Randhöhe des Manhartsbergs begrenzt, der in Nord-Südrichtung verläuft und das nördliche Niederösterreich in Waldviertel und Weinviertel trennt. Die Linie wird südlich weitergeführt von dem Wienerwald mit der Thermenlinie², die wiederum die Grenze zwischen Mostviertel und Industrieviertel aufzeichnet. (Österreich Lexikon AEIOU, am 24.07.08)

Klima

Das Klima der einzelnen Regionen Niederösterreichs weist erhebliche Unterschiede auf: Von kontinentalen Klimafaktoren mit geringen Niederschlägen, die das Wiener Becken und das Karpatenvorland prägen, zur atlantisch bestimmten Klimazone, die jährliche Regenmengen zwischen 800 und 1000 mm im Alpenvorland aufweist. Das Hügelgebiet des Weinviertels und die Beckenlandschaften liegen im Bereich des pannonischen Klimas. Den Süden von Niederösterreich prägt alpines Klima.

1 geologische Bezeichnung für den Teil des alten mitteleuropäischen Grundgebirges aus Granit (Weinsberger, Mauthausener, Eisgarner Granit), Gneis und kristallinen Schiefen, der Südböhmen, Westmähren und den nördlichen Teil Österreichs umfasst. Das Mühl- und das Waldviertel gehören zur Gänze zum Böhmischem Massiv

2 Thermenlinie geologisches Bruchsystem am Abbruch der Nördlichen Kalkalpen zum Wiener Becken mit warmen Quellen. Sie verläuft entlang der Linie Bad Fischau- Bad Vöslau- Baden- Mödling in das Wiener Stadtgebiet (Oberlaa)

In den südlichen gebirgigen Landesteilen und im Waldviertel herrschen Mischwälder und Nadelwälder (Fichte) vor. In den heute vom Ackerbau dominierten tiefer gelegenen Gebieten des Alpenvorlands, des Weinviertels und des Wiener Beckens sind Eichen-Hainbuchen-Wälder heimisch. Im Wienerwald sind ausgedehnte Buchenbestände charakteristisch. Ein besonderes Charakteristikum sind die im Raum Mödling-Vöslau und im Steinfeld zu findenden Schwarzföhrenwälder. (Österreich Lexikon AEIOU, am 24.07.08)

Landwirtschaft

Niederösterreich ist in erster Linie ein Agrarland. Ca. 42% der gesamten Landesfläche sind Ackerland. Ein weiterer wichtiger landwirtschaftlicher Aspekt ist der Weinbau, der vor allem - wie schon der Name sagt - im Weinviertel, aber auch in der Wachau und in der Thermenregion zu finden ist. Auch die Forstwirtschaft ist ein bedeutender Wirtschaftsfaktor; fast 40% der Fläche in Niederösterreich sind als Waldfläche ausgewiesen. (Österreich Lexikon AEIOU, am 24.07.08)

Flächennutzung

Niederösterreich trägt in vielfacher Hinsicht zur Wirtschaft Österreichs bei. So haben wirtschaftlich wichtige Flächenkategorien in Niederösterreich einen höheren prozentuellen Anteil als in Österreich insgesamt. Das gilt insbesondere für Ackerbauflächen, Weingärten und Infrastrukturflächen.

letzte Aktualisierung: 23.8.2007	Niederösterreich		Österreich		NÖ in % von Ö
	<i>absolut</i>	<i>relativ</i>	<i>absolut</i>	<i>relativ</i>	
Katasterfläche in km² Stand: 1.1.2006	19,177.78	100.0	83,870.95	100.0	22.9
Dauersiedlungsraum	11,276.53	58.8	31,451.61	37.5	35.9
Flächenwidmung					
Baufläche	210.95	1.1	670.97	0.8	31.4
landw. Nutzung	9,588.89	50.0	26,335.48	31.4	36.4
Weingärten	345.20	1.8	503.23	0.6	68.6
Wald	7,536.87	39.3	3,6232.25	43.2	20.8
Strassen	556.16	2.9	1,761.29	2.1	31.6

Quelle: Statistikkreferat WKO, NÖ, <http://wko.at/wknoe/stat/>

Tabelle 3.1: Flächenwidmung in Niederösterreich

4 Datengrundlage

4.1 LANDSAT

Aufgrund der Großräumigkeit, Kosteneffizienz, genügenden spektralen und räumlichen Auflösung und allgemeinen Verfügbarkeit wurden für diese Arbeit LANDSAT-Daten gewählt. Auch für die Erstellung der Referenzdatensätze wurden LANDSAT-Szenen verwendet.

LANDSAT-Satelliten sind Umweltsatelliten (primär), welche zur Beobachtung von Vegetation, Schnee/Eis, Geologie, Siedlungen etc. Anwendung finden. Durch die vorhandenen Spektralkanäle sind LANDSAT-Daten besonders für die Erhebung der Landbedeckung geeignet und stellen eine vielseitige Informationsquelle für Analysen der Umweltsituation und ihrer Veränderung dar.

Die gegenwärtig in einer Höhe von 705km im Umlauf befindlichen LANDSAT-Satelliten überfliegen alle 16 Tage den gleichen Ort, so dass ca. 20 Aufnahmen eines Ortes pro Jahr aufgezeichnet werden. Leider können nicht alle Bilder verwendet werden, da oft die atmosphärischen Bedingungen, wie Wolken, die Sicht auf die Erdoberfläche behindern.

Heute stehen die Daten von LANDSAT 5 TM (Thematik Mapper) und LANDSAT 7 ETM+ (Enhanced Thematic Mapper) zur Verfügung. Die Bahnen der zwei Satelliten sind gegeneinander versetzt, so dass alle 8 Tage ein Überflug durch einen der beiden Satelliten erfolgt. Dies ist für die Beobachtung von schnellen Veränderungen der Bodenbedeckung eine wichtige Datengrundlage. LANDSAT 5 TM hat in den Kanälen 1 bis 5 und 7 eine Pixelgröße von 30m und in Thermalkanal 6 eine Pixelgröße von 120m. LANDSAT 7 ETM+ verfügt zusätzlich über einen weiteren Kanal (panchromatisch) mit der Pixelgröße 15 m. Dies entspricht einer 4-fachen Verbesserung der Detailerkennbarkeit ($30^2/15^2$). Jedoch werden die Daten, die LANDSAT 7 ETM+ liefert, weniger eingesetzt, weil der Ausfall des Scan Line Corrector am 31. Mai 2003 zu einer deutlichen Verschlechterung der Datenqualität geführt hat. (http://LANDSAT.usgs.gov/products_slccoeffbackground.php)

Eine LANDSAT-Szene ist 182×175 km² groß. Für die Abdeckung Niederösterreichs werden 2 LANDSAT 5 TM Szenen benötigt und zwar die Szenen path 190/row 26 und path 190/row 27. Für die vorliegende Arbeit werden Bilder verwendet, die am 20.7.2007 aufgenommen wurden.

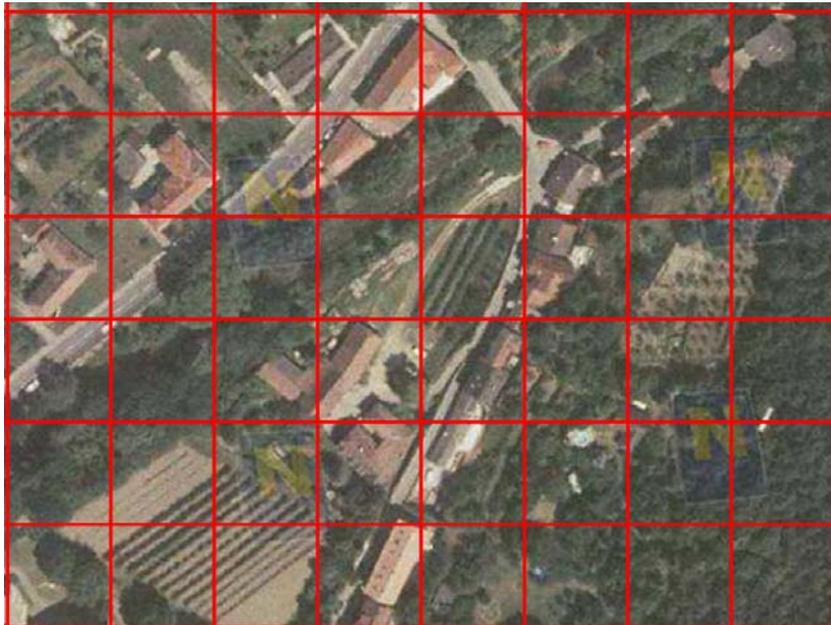


Abbildung 4.1: Visualisierung der Pixelgröße von LANDSAT durch Vergleich mit einem Luftbild. Die Seitenlänge der Quadrate beträgt 30 m und entspricht der Pixelgröße von LANDSAT. In dem Bild sind rechts unten relativ homogene Flächen. Bei den anderen Pixeln setzt sich die spektrale Signatur aus den Spektralsignaturen der beinhalteten Objekte zusammen.

Bei der digitalen Aufnahme von Satellitendaten entstehen Rasterbilder. Jedes Bild besteht aus einer Matrix von gleichartigen Bildelementen, den Pixeln (von engl. „picture element“). Die Größe dieser Pixel kennzeichnet die räumliche Auflösung. Das räumliche Auflösungsvermögen ist einer der wichtigsten Faktoren bei der Interpretation von Fernerkundungsdaten, da die Erkennbarkeit von Details auf den Aufnahmen direkt von der räumlichen Auflösung abhängt. (Stech, am 20.09.08)

- Bei den LANDSAT-5 TM-Aufnahmen hat jedes Bildelement 30m Kantenlänge, das heißt eine Fläche von 900m^2 . Wenn innerhalb eines Pixels zwei oder mehrere Objekte mit unterschiedlichen Spektralwerten liegen, dann zeigt das Pixel einen durch Gewichtung mit den Flächenanteilen entstandenen mittleren Wert. Die Zuordnung zu bestimmten Objekttypen wird dadurch erschwert. Es fällt daher schwer, Details im Satellitenbild wiederzufinden wie beispielsweise Straßenzüge oder einzelne Häuser.
- Neben dem Verhältnis von Objektgröße zu Pixelgröße hängt die Erkennbarkeit der Objekte auch von der Objektform ab. Flächenhafte Objekte von ausreichender Größe werden in mehreren Pixeln erfasst. In der Mitte der Fläche entsprechen die Spektralwerte der Pixel genau den Spektralwerten des Objekts (keine Mischpixel). Linienhafte Objekte bestimmen den spektralen Wert mehrerer nebeneinander liegender Pixel. Wenn die Breite dieser Objekte nicht die erwünschte Größe hat, dann „vermischen“ sich die Spektralwerte der Objekte mit der Umgebung.

- Auch der spektrale Kontrast der Objekte zueinander bestimmt die Erkennbarkeit: eine schmale Straße ist normalerweise nicht zu erkennen. Wenn die TM-Aufnahme jedoch im Winter gemacht wird (Straße geräumt, Umgebung schneebedeckt), könnte man die Straße (dunkel) von der Umgebung (weiß) unterscheiden. Die Objekte müssen einen gewissen Helligkeits- oder Farbkontrast zu ihrer Umgebung aufweisen, um überhaupt sichtbar abgebildet zu werden.

Kanal	Spektralbereich	Anwendungen und Merkmale
1	0,45 – 0,52 μ m (Blau)	Für Studien in Küstenbereichen geeignet, da tiefes Eindringen in Wasser; wird wegen der starken Absorption des Chlorophylls genutzt zur Unterscheidung von Boden und Vegetation.
2	0,52 – 0,60 μ m (Grün)	Misst die in diesem Spektralbereich vergleichsweise hohe Reflexion von (gesunder) Vegetation an Land und im Wasser; Unterscheidung verschiedener Vegetationstypen möglich.
3	0,63 – 0,69 μ m (Rot)	Misst die Chlorophyllabsorption durch gesunde Vegetation; Differenzierung der Pflanzenarten; gut geeignet zur Darstellung von Straßen und unbedecktem Boden.
4	0,76 – 0,90 μ m (Nahes Infrarot)	(Minimum der Chlorophyllreflexion) misst die in diesem Spektralbereich sehr hohe Reflexion gesunder Vegetation → wird genutzt zur Abschätzung der Biomasse; Erkennung der Küstenlinien wegen geringer Eindringtiefe in Wasser
5	1,55 – 1,75 μ m (Mittleres Infrarot)	Zeigt den Feuchtezustand der Vegetation, aber auch des Bodens; ermöglicht die Unterscheidung zwischen Wolken und Schnee; zeigt Straßen, unbedeckten Boden und Wasser; guter Kontrast bei unterschiedlichen Vegetationstypen, keine Beeinflussung durch Dunst.
6	10,4 – 12,5 μ m (Thermales Infrarot), 120m	Messung von Wärmestrahlung (auch nachts); Ermittlung von Stress in der Vegetation; zeigt unterschiedliche Bodenfeuchte an; ermöglicht Tiefenmessungen in Seen; Schwachpunkt: geringere Auflösung.
7	2,08 – 2,35 μ m (Mittleres Infrarot), 30m	Zeigt unbedeckten Boden; geeignet zur Vegetationsdifferenzierung; Unterscheidung von Mineralien und Gestein sowie hydrothermalen Veränderungen.

Quelle: GeoAg, Landsat-Multispektral-Aufnahme, <http://geo-ag-wesel.de/multispektral.htm>

Tabelle 4.1: Empfindlichkeitsbereich der Spektralkanäle von LANDSAT 5

4.2 Digitales Höhenmodell-SRTM

Für diese Arbeit wurde das SRTM-Höhenmodell verwendet.

SRTM-Daten (Shuttle Radar Topography Mission) wurden in 11 Tagen im Februar 2000 aus dem Weltraum aufgezeichnet. Das Shuttle wurde in eine Umlaufbahn mit einer Neigung von 57 Grad geschickt. Dies erlaubte ca. 80% der Landfläche aufzunehmen, zwischen 60 Grad nördlicher und 56 Grad südlicher Breite (damit wurden 95% der bewohnte Fläche der Erde abgedeckt). Das Space Shuttle umkreiste die Erde 16-mal pro Tag. Während der 11-tägigen Mission wurden 176 Umlaufbahnen der Erde abgeschlossen. Die Mission war für die Topographie der Erdoberfläche ausgerichtet, so dass die Ozeane nicht berücksichtigt wurden.

Die Geländehöhen wurden durch Radar-Interferometrie gemessen. Vorteile von Radarsystemen gegenüber optischen Systemen bestehen darin, dass Radarsysteme Tag und Nacht und unabhängig von der Bewölkung eingesetzt werden können.

(<http://www2.jpl.nasa.gov/srtm/>)

Die zunächst veröffentlichten SRTM-Daten weisen Messfehler auf (aus technischen Gründen und wegen Reduzierung der Datenmengen). Seit 2005 stehen die Daten mit verbesserter Qualität zur Verfügung (Version 2). Viele Fehler konnten jedoch nicht vollständig beseitigt werden:

- Die Höhendaten geben die Oberflächenstruktur der Erde wieder. Dies muss nicht der Höhe des Erdbodens entsprechen, sondern schließt Bewuchs und Bebauung ein, was zu Abweichungen bei topografischen Kartenwerken führen kann.
- An einigen Stellen weisen die Daten „Löcher“ auf (Fehlpixel). Für diese Orte konnte die Höhe nicht zuverlässig ermittelt werden, weil z.B. die Radarstrahlen von gewissen Oberflächen (Wasserflächen, Eisflächen, wie Gletscher) zu stark vom Sensor weg reflektiert wurden oder Teile von Bergen im Schattenbereich des Sensors lagen (Problem des Seitensichtradars). Um die Stellen zu füllen, für die keine Höheninformation bestimmt werden konnte, sind verschiedene Verfahren üblich: Interpolation aus den vorhandenen Werten (dies führt insbesondere bei größeren Lücken in Gebirgsregionen zu Ungenauigkeiten) und die Korrektur mittels Daten aus anderen Quellen.

(<http://de.wikipedia.org/wiki/SRTM-Daten>)

Der **USGS** hat für die freie Nutzung zwei unterschiedliche Datensätze zur Verfügung gestellt:

- Für das Gebiet der USA sind die Daten mit einer Auflösung von einer Bogensekunde (30m) in Länge und Breite verfügbar (SRTM-1). Diese Daten werden in Stücken von 1×1 Grad mit einer Größe von 3601×3601 Pixel angeboten.

- Für den Rest der Erdoberfläche werden nur Daten mit einer reduzierten Auflösung von 3 Bogensekunden (90m) angeboten (SRTM-3). Die Größe der 1×1 Grad-Stücke beträgt hierbei 1201×1201 Pixel.

Die vom USGS veröffentlichten Daten sind als Public Domain zum Herunterladen kostenfrei verfügbar und dürfen uneingeschränkt verwendet werden. Die nicht veröffentlichten Daten mit einer Bogensekunde (30m) Auflösung von Gebieten außerhalb der USA unterliegen der Kontrolle des US-Verteidigungsministeriums. Das Benutzerinterface des USGS erlaubt zudem eine frei definierbare rechteckige Auswahl beliebiger Gebiete, für welche die Datensätze individuell zusammengestellt und an den Benutzer übertragen werden. (<http://edc.usgs.gov/srtm/>)

Bei der **GLCF** (Global Land Cover Facility) ist das Herunterladen von SRTM-Daten nach LANDSATszenen möglich. (<http://glcf.umiacs.umd.edu/data/srtm/>)

Das SRTM-Modell wurde am Institut für Vermessung, Fernerkundung und Landinformation, BOKU, Wien, weiter bearbeitet. Dies war mit gewissen Schwierigkeiten verbunden. Die Abweichung der Höhenwerte in den SRTM-Daten durch Bewuchs und Bebauung von konventionell ermittelten Höhendaten ist zu beachten. National erfasste Höhenmodelle haben im Gegensatz zu SRTM-Daten unterschiedliche Bezugssysteme, was deren Weiterverarbeitung erschwert.

Jedoch ist für den Zweck der vorliegenden Arbeit kein hochgenaues, digitales Höhenmodell notwendig. Das Modell wird nur für geometrische und radiometrische Korrekturen herangezogen, für die eine grobe Schätzung der Höhendaten aus (zumal die LANDSAT-Szenen eine Pixelauflösung von 30m haben).

4.3 Referenzdatensätze

4.3.1 SINUS-Datensatz

SINUS wurde vom Institut für Vermessung, Fernerkundung und Landinformation (IFVL) der Universität für Bodenkultur, Wien erstellt und war Teil des Forschungsprojekts „Landschaftsökologische Strukturelemente als Indikator der Nachhaltigkeit“ (Spatial Indicators for Land Use Sustainability - SINUS). (Hollaus und Suppan, 2003)

Ziel dieses Projekt war es, Indikatoren für die Bewertung und langfristige Beobachtung der österreichischen Kulturlandschaften zu entwickeln, um die Zusammenhänge zwischen den räumlichen Mustern der Landschaftselemente und der ökologischen Situation zu erkennen und zu erklären.

Als Datengrundlage für die Erstellung dieses Landbedeckungsdatensatzes dienten LANDSAT 5 TM Satellitenbilder, aufgenommen im Zeitraum von 1993 bis 1996. Das Bild, das

Niederösterreich abdeckt, ist am 1.8.1994 aufgenommen. Zusätzlich stand eine grobmaßstäbliche Gliederung der Kulturlandschaften Österreichs zur Verfügung. (Hollaus und Suppan, 2003)

Für die Erstellung des SINUS-Datensatzes wurde eine automatische objektbasierte (segmentweise) Klassifizierung durchgeführt. Segmente sind räumlich zusammenhängende Mengen von Pixeln, die eine spektrale und texturale Homogenität aufweisen und die sich von anderen Pixeln in ihrer unmittelbaren Umgebung unterscheiden.

Für die Klassifikation wurde ein Algorithmus entwickelt, der statistische sowie wissensbasierte Ansätze verwendet. Unter Einbeziehung der Spektral- und Texturattribute, der Attribute betreffend der Form und Lage der Segmente, der aus dem digitalen Höhenmodell abgeleiteten Informationen (Seehöhe, Exposition, Neigung) sowie der Informationen über die Naturraumzugehörigkeit wurde die segmentweise Klassifizierung durchgeführt.

Code	Landbedeckungsklassen – Land Cover
1.1.	Vegetationsfreie Standorte, die nicht versiegelt sind und keinen Humushorizont besitzen, Vegetationsbedeckung < 30 % (z.B. Schottergruben, Steinbrüche, Deponien, Fels, Schutt, Schotter, Moränen, Baustellen)
1.2.	Eis- und Schneeflächen
2.0.	Wasserflächen (offene Wasserflächen wie z.B. Seen, nicht aber Mischbereiche von Wasserflächen mit angrenzender Vegetation, beispielsweise schmale Auwaldgewässer)
3.1.	Nadelwald – Bestandesschluss > 40 % (Laubwaldanteil < 20 %)
3.2.	Mischwald – Bestandesschluss > 40 %
3.3.	Laubwald – Bestandesschluss > 40 % (Nadelwald < 20 %; Diese Klassen beinhalten auch Kleingehölzkomplexe)
4.1.	Grünland mit viel Biomasse (meist mit hohem Feuchtegrad korreliert)
4.2.	Grünland mit wenig Biomasse (meist mit geringerem Feuchtegrad korreliert)
4.3.	Schilf, Moore
4.9.	Grünland gemischt (Grünland mit wechselndem Biomasseanteil, Vorkommen von Sträuchern, Gebüsch, Einzelbäumen)
5.0.	Weingärten, Obstplantagen
6.0.	Offener Boden (geackertes Feld)
6.1.	Reife Ackervegetation (gelb, meist Halmfruchtäcker)
6.2.	Grüne Ackervegetation (z.B.: Mais, Zuckerrüben, aber auch Halmfruchtäcker in einem frühen Vegetationsstadium)
6.3.	Unspezifizierbare Ackervegetation (meist Mischungen von reifer und grüner Ackervegetation)
7.1.	Geringer Versiegelungsgrad (stark durchgrünte Siedlungsbereiche)
7.2.	Mittlerer Versiegelungsgrad (Siedlungsbereiche mit geringem Grünanteil)
7.3.	Hoher Versiegelungsgrad (stark verbaute Siedlungsbereiche z.B. Stadtkerne, Industrieflächen)
9.0.	Zurückweisungsklassen, Schatten, Wolken

Tabelle 4.2: SINUS Landbedeckungsklassen

Anschließend wurde die Genauigkeit der Klassifikation mit 200 Testquadraten (1km×1km) über ganz Österreich verteilt visuell überprüft und gegebenenfalls adaptiert.

Der Landbedeckungsdatensatz ist für ganz Österreich verfügbar und wird für wissenschaftliche Zwecke in aufaggrierter Rasterform abgegeben, wobei die Rasterweite variabel ist ($\geq 90\text{m}$). Die Aktualität wird durch die Aufnahmezeitpunkte der Satellitenbilder bestimmt und kann somit mit der Zeitspanne 1993 bis 1996 spezifiziert werden. (Hollaus und Suppan, 2003)

Da für die Kulturlandschaftsforschung landwirtschaftsökologische Fragestellungen im Hinblick auf Struktur und Nutzung (Landwirtschaft, Siedlung...) im Vordergrund standen und nicht die tatsächliche Bodenbedeckung im Hinblick auf die Pflanzenart, sind hier Kategorien anzutreffen, die teilweise die Landnutzung beschreiben.

4.3.2 CORINE Landcover

CORINE, Synonym für „koordinierte Erfassung von Informationen über die Umwelt“ (Coordination of Information on the Environment) ist ein von der Kommission der Europäischen Union im Jahr 1985 gegründetes Programm.

Im Rahmen des CORINE Programmes werden unterschiedlichste Projekte hinsichtlich einer einheitlichen Erfassung von Umweltdaten durchgeführt (CORINE Air, CORINE Biotopes, CORINE Landcover). Das Projekt wurde nicht nur in den Mitgliedstaaten gestartet, sondern auch auf osteuropäische und nordafrikanische Staaten ausgedehnt.

Das Ziel des CORINE Landcover Projektes war eine einheitliche europaweite Erfassung der Bodenbedeckung bzw. Landnutzung auf der Grundlage von Satellitendaten.

In Österreich führt das Umweltbundesamt Wien das CORINE-Landcover-Projekt seit Mitte 1992 durch.

Für die Koordination und fachliche Beratung bezüglich des CORINE Landcover Projektes wurde 1995 ein „European Topic Centre on Landcover“ unter der Leitung von MDC (Centre for Environmental Satellite Data) in Schweden eingerichtet. Die „Gemeinsame Forschungsstelle der Europäischen Union“ (JRC) in Ispra, Italien, entwickelte eine Methodik zur Erfassung der Bodenbedeckung bzw. Landnutzung, welche von allen an diesem Projekt teilnehmenden Staaten anzuwenden ist.

Die Erhebung erfolgte mittels einer computerunterstützten visuellen Interpretation. Als digitale Basisdaten dienten Satellitendaten (LANDSAT 5 TM). Unter Verwendung der CORINE Landcover Nomenklatur (siehe Tabelle 4.3) und weiterer Hilfsmittel wie topographischer Karten, thematischer Karten, Luftbilder, Statistiken etc. wurde die Bodenbedeckung hochgezeichnet (kleinste Erhebungseinheit 25 Hektar, Mindestbreite bei länglichen Elementen 100 m).

Die Landbedeckungskarte wird alle 10 Jahre aktualisiert. Bis jetzt ist der Datensatz ein einziges Mal modifiziert worden, von CLC90 zu CLC2000. Hölzl (2003)

Die CORINE-Landcover-Nomenklatur ist in 3 Ebenen (Levels) aufgebaut. (siehe Tabelle 4.3)

Der Datensatz wird in eine Vektorform mit einer kleinsten Kartiereinheit von 25 ha abgegeben und ist für nicht kommerzielle Zwecke kostenlos.

Link: <http://gis.umweltbundesamt.at/download/CORINE/Register1.jsp>

Die CORINE-Nomenklatur beschreibt sowohl Landbedeckungskategorien als auch Landnutzungskategorien. Wie schon erwähnt, kann die Landnutzung nicht allein aus den Bilddaten gewonnen werden. Da für die Erstellung des CORINE-Datensatzes eine visuelle Interpretation unter Zuhilfenahme anderer Datenquellen durchgeführt wurde, war es möglich, auch Landnutzungskategorien auszuweisen.

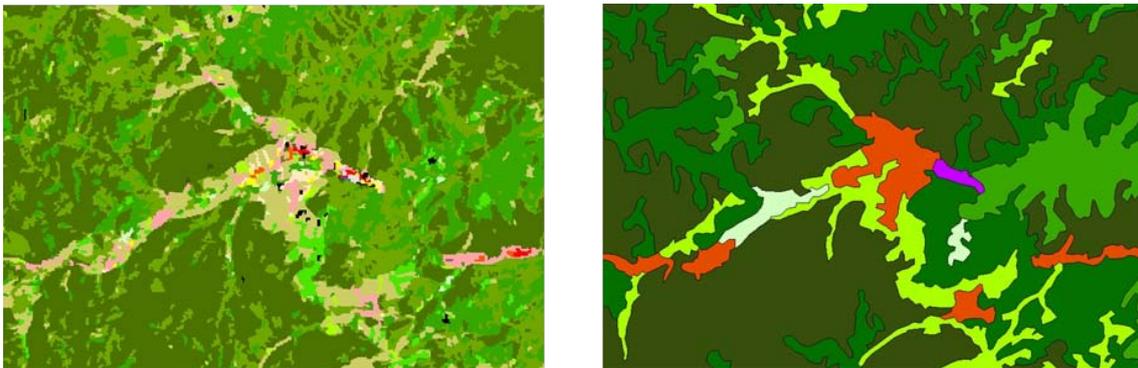


Abbildung 4.2: SINUS und CORINE: Vergleich der räumlichen Auflösung (Generalisierung)

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
		<i>1.3.2. Deponien, Abraumhalden</i>
		<i>1.3.3. Baustellen</i>
	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
		<i>2.1.2. Regelmäßig bewässertes Ackerland</i>
		2.1.3. Reisfelder
	2.2. Dauerkulturen	2.2.1. Weinbauflächen
		<i>2.2.2. Obst/Beerenobstbestände</i>
		<i>2.2.3. Olivenhaine</i>
	2.3. Grünland	2.3.1. Wiesen und Weiden
	2.4. Heterogene landwirtschaftliche Flächen	<i>2.4.1. Einjähr. Kulturen in Verbindung mit Dauerkulturen</i>
		2.4.2. Komplexe Parzellenstruktur
		2.4.3. Landwirtschaftlich genutztes Land mit Flächen natürlicher Vegetation von signifikanter Größe
<i>2.4.4. Land/Forstwirtschaftliche Flächen</i>		
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
		<i>3.2.3. Hartlaubbewuchs</i>
		3.2.4. Wald/Strauch Übergangsstadien
	3.3. Offene Flächen ohne oder mit geringer Vegetation	<i>3.3.1. Strände, Dünen, Sandflächen</i>
		3.3.2. Felsflächen ohne Vegetation
		3.3.3. Flächen mit spärlicher Vegetation
<i>3.3.4. Brandflächen</i>		
3.3.5. Gletscher/Dauerschneegebiet		
4. Feuchtflächen	4.1. Feuchtflächen im Landesinneren	4.1.1. Sümpfe
		4.1.2. Torfmoore
	<i>4.2. Feuchtflächen an der Küste</i>	<i>4.2.1. Salzwiesen</i>
		<i>4.2.2. Salinen</i>
<i>4.2.3. In der Gezeitenzone liegende Flächen</i>		
5. Wasserflächen	5.1. Wasserflächen im Landesinneren	5.1.1. Gewässerläufe
		5.1.2. Wasserflächen
	<i>5.2. Meeresgewässer</i>	<i>5.2.1. Lagunen</i>
		<i>5.2.2. Mündungsgebiete</i>
		<i>5.2.3. Meer und Ozean</i>

Quelle: Umweltbundesamt-Raumplanung-CORINE Nomenklatur

Tabelle 4.3: CORINE-Landcover: grau/kursive Klassen kommen in Österreich nicht vor. Die 1. Spalte beschreibt Level 1, die 2. Spalte beschreibt Level 2, die 3. Spalte beschreibt Level 3

4.3.3 Visuelle Interpretation

Bei den Referenzdatensätzen aus SINUS und CORINE handelt es sich um ältere Datensätze. Zwischen ihrer Herstellung bzw. der Aufnahme der zu ihrer Herstellung verwendeten Bilddaten und der Aufnahme der in der vorliegenden Arbeit verwendeten Satellitenbilder können Änderungen der Landbedeckung eingetreten sein. Um auch einen aktuellen Referenzdatensatz zu haben, wurde ein solcher durch visuelle Interpretation der aktuellen, für die Klassifikation verwendeten LANDSAT-Bilder an Punkten eines regelmäßigen Stichprobenrasters (siehe Kap. 6.1) hergestellt.

Zusätzlich zu den LANDSAT-Bildern wurden fallweise auch Orthophotos verwendet.

Bis auf kleine Änderungen wurde das Klassifikationsschema von SINUS übernommen. (siehe Tabelle 4.2)

Klasse	Beschreibung
11	vegetationsfreie Standorte, wie Schottergruben, Fels, Baustellen
20	Wasserflächen
31	Nadelwald
32	Mischwald
33	Laubwald
41	Grünland, mit viel Biomasse
42	Grünland, mit wenig Biomasse
44	unbestimmbare landwirtschaftliche Flächen mit viel Biomasse
45	unbestimmbare landwirtschaftliche Flächen mit wenig Biomasse
49	Grünland gemischt, mit wechselndem Biomasseanteil, Vorkommen von Sträuchern, Gebüsch, Einzelbäumen
50	Weingärten
60	offener Boden (geackertes Feld)
61	reife Ackervegetation (gelb, meist Halmfrüchte)
62	grüne Ackervegetation (Mais, Zuckerrüben, auch Halmfrüchte in einem früheren Stadium)
63	unspezifische Ackervegetation (meist Mischungen von reifer und grüner Ackervegetation)
71	geringer Versiegelungsgrad
72	mittlerer Versiegelungsgrad
73	hoher Versiegelungsgrad
90	Wolken
91	Schatten

Tabelle 4.4: Klassen der visuellen Interpretation

5 Vorarbeiten

5.1 Bildvorverarbeitung allgemein

Ziel der Bilddatenvorverarbeitung ist es, die Bilder in die richtige geometrische Lage zu bringen, d.h. den Bezug zu einem Geländekoordinatensystem herzustellen, sowie diejenigen Fehler zu beseitigen, die durch Variation der (u.a. topographisch bedingten) Sonneneinstrahlung und der atmosphärischen Zustände entstehen. (Georeferenzierung, radiometrische Korrektur)

Die erworbenen Originaldaten sind nicht ohne jegliche Vorbearbeitung abgegeben worden. Die geometrischen Fehler wurden schon teilweise korrigiert. Unter diesen sind die Korrektur von Verzerrungen durch Erdkrümmung, Erdrotation¹ und Vorwärtsbewegung des Satelliten während der Abtastung einer Zeile.

Jede Korrektur erfolgt durch eine Transformation. „Durch eine Transformation T wird dann eine diskrete zweidimensionale Grauwertfunktion $g(x,y)$ des Eingabebildes in die ebenfalls diskrete zweidimensionale Grauwertfunktion $g'(x',y')$ des Ausgabebildes transformiert“. (Albertz, 2007)

Durch geometrische Transformationen sollten die Bilder in ihrer Form verändert werden, die Grauwerte sollten dabei erhalten bleiben. Auf Grund der Transformationsgleichung ergeben sich aber keine ganzzahligen Werte für die Bildkoordinaten des Ausgabebildes, x',y' . Deshalb muss eine Regel eingeführt werden, nach der die Grauwertzuweisung erfolgen soll (Resamplingverfahren). Dazu gibt es allgemein 3 Methoden:

- Verfahren der nächsten Nachbarschaft („Nearest Neighbour“): Dabei ändern sich die Grauwerte nicht. Jedem Pixel im Ausgabebild wird der Grauwert jenes Pixels im Eingabebild zugewiesen, welches den berechneten Koordinaten x',y' am nächsten liegt. Das kann aber an Grenzen zu einer störenden Treppenstruktur führen und manche Werte können verschwinden, andere verdoppelt werden.
- Bilineare Interpolation: Diese berechnet den Grauwert durch lineare Interpolation zwischen den vier direkt benachbarten Grauwerten. Die bilineare Interpolation wird oft eingesetzt.

¹ Zeilenabtaster wie LANDSAT benötigen eine bestimmte Zeit zur Erfassung einer Szene. Während der Aufnahmedauer dreht sich die Erde in östliche Richtung, so dass der Startpunkt der letzten Bildzeile weiter westlich liegt als derjenige der ersten Bildzeile.

- Kubische Faltung („Cubic Convolution“): Diese Interpolation verwendet die Werte von 4×4 umliegenden Pixeln. Sie liefert die visuell ansprechendsten Ergebnisse.

Es ist ein Nachteil der beiden letzten Methoden, dass sie zu einem Kontrastverlust führen. Für eine kNN-Klassifikation, die auf den Spektraleigenschaften jedes Pixels beruhen, ist dies nicht sehr vorteilhaft, weil hier erwünscht ist, dass die Spektralwerte unterschiedlicher Landbedeckungskategorien stark von einander abweichen.

Bei der radiometrischen Korrektur werden die Grauwerte verändert, während die geometrische Eigenschaften erhalten bleiben. Radiometrische Transformationen werden sowohl zur Korrektur von verschiedenartigen Störeinflüssen (z.B. Atmosphäreneinfluss) eingesetzt als auch zur Bildverbesserung im Hinblick auf eine spätere Auswertung. (Alberty, 2007)

5.2 Georeferenzierung der Szenen

Um die Satellitenbilder mit den Referenzdatensätzen verschneiden zu können, müssen sie mit diesen geometrisch exakt übereinstimmen.

Durch die Georeferenzierung werden die Satellitenbilder auf die Projektion der Referenz-Datensätze gebracht (geokodiert). Lageverschiebungen innerhalb des Satellitenbildes, die sich auf Grund von Geländeunterschieden ergeben haben, werden beseitigt. Die geometrische Korrektur impliziert die Übertragung von Pixeln aus der ursprünglichen Bildmatrix in eine neue Position im Ergebnisbild. Die Bilder werden differenziell (stückweise) verschoben, skaliert und gedreht.

Die Entzerrung erfolgt mit einer Transformationsgleichung, die mit Hilfe von Passpunkten berechnet wird, und einem Resamplingverfahren. Die Passpunkte müssen im Bild eindeutig identifizierbar sein und ihre Koordinaten im Geländekoordinatensystem müssen entweder bekannt sein oder einem Referenzdatensatz entnommen werden. (Alberty, 2007)

Mit Hilfe dieser Passpunkte (GCPs, ground control points) können die geometrischen Beziehungen zwischen Bild und Gelände hergestellt und die Transformationsparameter berechnet werden. Geländehöhenunterschiede können berücksichtigt werden, wenn ein digitales Geländemodell vorhanden ist. Für die Auswahl von Passpunkten (GCPs) muss Folgendes beachtet werden:

- ausreichende Anzahl (genügend Information für die Regressionsschätzung des zu erstellenden Transformationsgleichungssystems)
- gute räumliche Verteilung (vor allem in Bildecken und entlang der Bildseiten, dazu im Bildbereich gut verteilte Punkte)
- eindeutige Erkennbarkeit (gute Kontrastverhältnisse)

- möglichst unbeeinflusst von zeitlichen Veränderungen (Strassenkreuzungen, Brückenmitten, etc.)

Es wurde das Koordinatensystem UTM (WGS-1984-Zone 33N) verwendet. In jedem Satellitenbild wurden bis zu 35 Passpunkte gewählt. Die Genauigkeit der Transformation kann durch die Residualwerte an den Passpunkten und durch die mittlere quadratische Abweichung (root mean square error, RMSE) angegeben werden. Der maximale RMSE (root mean square error) lag bei 1,52 Pixel, was bei einer Pixelgröße von 30m einer Abweichung von etwa 45m entspricht. Als Passpunkte wurden häufig Strassenkanten, Strassenkreuzungen, Brückenmitten, einzelne, gut sichtbare Gebäude, Baumgruppen usw. verwendet.

In der Abbildung 5.1 ist die Verteilung der Passpunkte ersichtlich. Die Transformationsgleichung wurde für jedes Bild separat berechnet. Damit aber die Bilder auch gut an einander angepasst sind, wurden im Überlappungsbereich die gleichen Punkte für beide Bilder genommen.

Die Koordinaten der Passpunkte wurden entnommen aus:

1. **Orthophotos** für die Passpunkte in Österreich (von den Homepages der Länder Niederösterreich, Kärnten und Steiermark). Gründe: sehr gute Objekterkennbarkeit, hohe Genauigkeit, Verfügbarkeit in mehreren Referenzsystemen.
2. **Austria Map** für die Grenzgebiete (Die Austria Map reicht bis zu 60km in die Nachbarländer) und für das Burgenland, wo keine Orthophotos zu Verfügung standen.
3. **Google Earth** für die Passpunkte in Tschechien, Ungarn, Slowakei und Slowenien. Nachteil: Die Auflösung ist teilweise schlechter es gibt möglicherweise Lageungenauigkeiten.
4. georeferenzierte **LANDSAT Szene** 190/26 aus dem Jahre 2001. Hier sind die Passpunkte wegen der gleichen Pixelauflösung leicht zu orten.

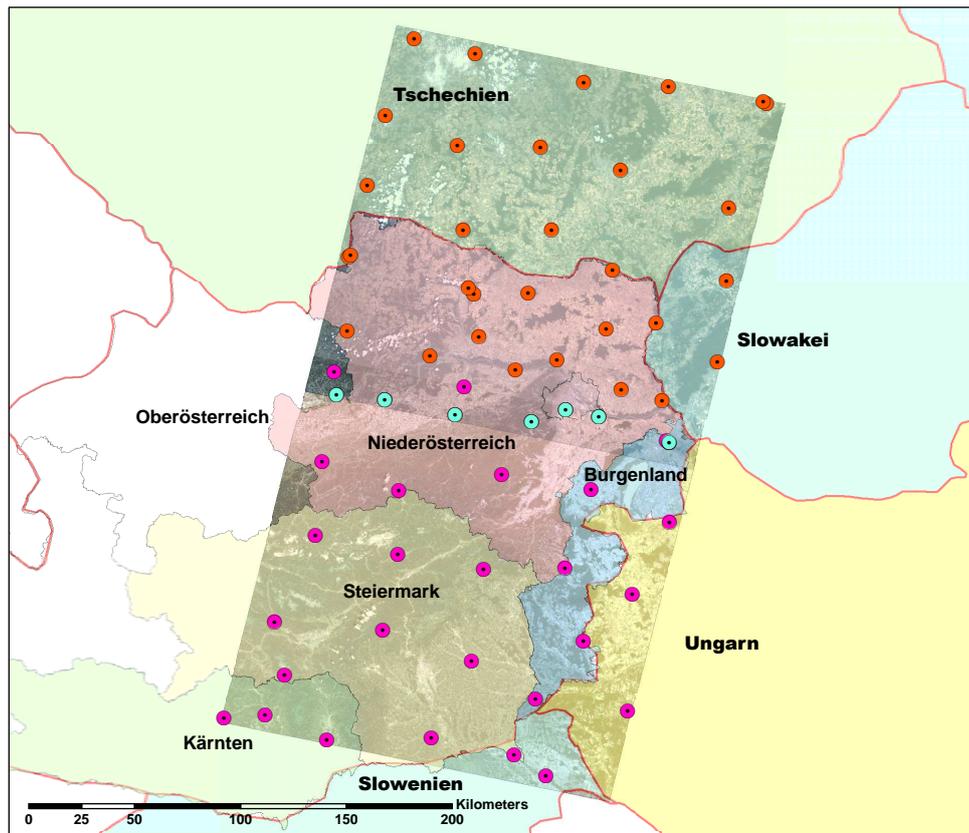


Abbildung 5.1: LANDSAT-Szenen nach dem Georeferenzieren, Passpunktverteilung

5.3 Radiometrische Korrektur

Ziel der radiometrischen Korrektur ist es, Faktoren zu eliminieren, die die Bildinformation beeinflussen, die nicht objektspezifisch sind und die sich den eigentlichen Spektralwerten der Objekte in störender Weise systematisch überlagern, wie Atmosphäreinfluss und durch die Geländeform (Topographie) auftretende Effekte. Nach einer erfolgreichen Korrektur sollte die spektrale Signatur von Störungen bereinigt sein und allein von Reflexionseigenschaften der Bodenbedeckung abhängen. Die geometrischen Eigenschaften der Bilddaten bleiben unverändert.

Beide LANDSATbilder, die am selben Tag unmittelbar nacheinander aufgenommen sind, wurden mit den selben Parametern korrigiert, da angenommen werden kann, dass sich der Zustand der Atmosphäre nicht so schnell verändert hat. Anschließend wurde überprüft, ob die Korrektur mit diesen Werten, für beide Bilder gute Ergebnisse liefert.

5.3.1 Luftlichtkorrektur

Unter Luftlicht versteht man diejenige Strahlung, die von der Sonne ausgeht und von der Atmosphäre direkt zum Sensor zurückgestreut wird. Das Luftlicht überlagert also die von der Geländeoberfläche reflektierte Strahlung.

Das Luftlicht führt stets zu einer Aufhellung und Kontrastminderung. Es hängt stark vom Spektralbereich, vom momentanen Zustand der Atmosphäre und von anderen Faktoren wie Sonnenstand, Beobachtungsrichtung, sowie vom Geländere relief ab.

Das Luftlicht kann direkt aus den Bildern abgeschätzt werden. Diese Methode beruht auf der Annahme, dass dunkle Objekte (wie z.B. klares Wasser oder Schatten) keine Reflexion aufweisen und daher theoretisch einen Pixelwert von Null haben müssten. Jede Abweichung von Null ist auf atmosphärische Einflüsse zurückzuführen und kann durch Subtraktion dieses Wertes vom Originalbildwert eliminiert werden. Das Luftlicht wird als Konstante über das gesamte Bild angenommen. Es ist für jeden einzelnen Kanal zu ermitteln, da seine Intensität mit abnehmender Wellenlänge zunimmt.

5.3.2 Topographische Korrektur

Die topographische Korrektur wurde durchgeführt, weil die Geländemorphologie ausgeprägt ist. Im südwestlichen Teil von Niederösterreich reichen die Geländehöhen über 1500m, und die Beleuchtungsunterschiede sind auffällig.

Die topographische Korrektur soll die Beleuchtungsunterschiede, die durch den variierenden Sonnenstand und die Topographie hervorgerufen werden, kompensieren. Es gibt verschiedene Verfahren für die topographische Korrektur: Cosinus-Korrektur, C-Korrektur, Minnaert-Korrektur, SCS-Korrektur. Eine Voraussetzung für die genannten Korrekturmethode n ist die Verfügbarkeit eines digitalen Höhenmodells, da die Beleuchtungswinkel und die Hangneigung benötigt werden.

Bei der **Cosinus-Korrektur** wird mit Einbeziehung der Geländeneigung und des Einstrahlungswinkels, unter Voraussetzung Lambertscher Objektreflexion, der Reflexionswert auf den Wert einer ebenen Fläche korrigiert.

C-Korrektur hat die Cosinus-Korrektur als Basis. Durch einen zusätzlichen empirischen Faktor C soll die diffuse Streustrahlung der Atmosphäre nachgebildet und damit die Überkorrektur bei flachen Sonneneinstrahlungswinkel reduziert werden.

Die **Minnaert-Korrektur** erweitert die Cosinus-Korrektur um einen oberflächenspezifischen Exponenten, der die Abweichung von lambertschen Reflexionseigenschaften beschreibt.

Die **SCS-Korrektur** wurde speziell für bewaldete Gebiete entwickelt. Die Wechselwirkungen zwischen Strahlung und bewaldeter Oberfläche werden auf zwei Ebenen betrachtet, und zwar für das Kronendach als Ganzes und für die einzelne Baumkrone. Die Beleuchtung des Kronendachs ist stark von der Topographie abhängig, während die geometrischen Verhältnisse von Sonne und Einzelkrone von der Topographie unabhängig sind, da die Ausrichtung des Einzelbaumes unabhängig von der Geländeneigung ist. (Gu und Gillespie, 1997) Das der SCS-Methode zugrunde liegende mathematische Beleuchtungsmodell wird durch folgende Gleichung beschrieben:

$$L_N(\lambda) = L_O(\lambda) \cdot \frac{\cos s \cdot \cos sz}{\cos i} \quad (5.1)$$

$L_O(\lambda)$ ist die Strahldichte der von einer geneigten bewaldeten Fläche reflektierten Strahlung, $L_N(\lambda)$ ist die (fiktive, normalisierte) Strahldichte vom selben Waldbestand auf einer ebenen Fläche. s ist die Geländeneigung, sz ist der Zenitwinkel der Sonne, und i ist der Winkel zwischen Sonneneinstrahlung und Geländennormaler. (Gu und Gillespie, 1997)

Dieses Modell berücksichtigt aber nur den direkten Anteil der einfallenden Strahlung. Tatsächlich setzt sich die Einstrahlung aus einer direkten und einer diffusen Komponente zusammen. Durch Vernachlässigung des diffusen Strahlungsanteils kann es zu einer Überkorrektur in Bereichen geringer Beleuchtung kommen. Dieser Effekt ist im korrigierten Bild in Form von starken Aufhellungen sichtbar und führt zu Fehlklassifikationen.

Daher wurde eine Methode entwickelt, die es ermöglicht, die Anteile der direkten und der diffusen Einstrahlung abzuschätzen und bei der Korrektur zu berücksichtigen. Das für die Berücksichtigung der diffusen Einstrahlung modifizierte Modell wird durch die folgende Gleichung beschrieben:

$$L_N(\lambda) = L_O(\lambda) \cdot \frac{\cos s \cdot \cos sz}{c_s(\lambda) \cdot \cos i + [1 - c_s(\lambda)] \cdot \cos s \cdot \cos sz} \quad (5.2)$$

$C_s(\lambda)$ ist hier der Anteil der direkten Sonneneinstrahlung an der Globalstrahlung. Die Werte von $c_s(\lambda)$ werden folgendermaßen ermittelt: In Bereichen ausgeprägten Reliefs, wo hohe Beleuchtungsunterschiede vorkommen, werden Pixel von Testflächen möglichst einheitlicher Landbedeckung (Wald) ausgelesen. Jeder bei einer bestimmten Geländeneigung ermittelte Pixelwert wird mit Hilfe von Gleichung 5.3 unter Annahme unterschiedlicher Werte von $c_s(\lambda)$ in normalisierte Werte umgerechnet. Für jeden angenommenen Wert von $c_s(\lambda)$ wird mit Hilfe eines linearen Regressionsansatzes (Gleichung 5.3) die verbliebene Abhängigkeit von den Beleuchtungsverhältnissen (Winkel i) in Form des Faktors k ermittelt. Es sind die $c_s(\lambda)$ -Werte von Interesse, bei denen die Regressionsgerade eine Steigung von $k = 0$ aufweist, d.h. keine Abhängigkeit vom Beleuchtungswinkel

i gegeben ist. (Koukal u. a., 2005)

$$L_N(\lambda) = k \cdot \cos i + d \quad (5.3)$$

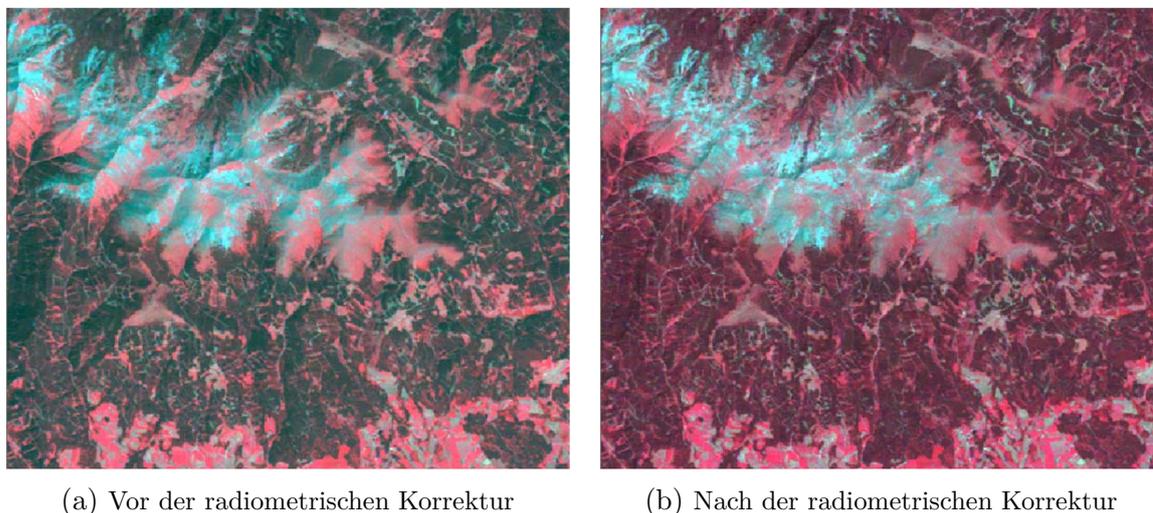


Abbildung 5.2: Radiometrische Korrektur

5.4 Erstellen einer Bildmaske

Es wurde eine Maske erstellt, die sicherstellt dass die kNN-Klassifizierung nur auf das Gebiet von Niederösterreich außerhalb von Wolken- und Wolkenschattenflächen angewendet wird.

Die Gründe dafür sind die folgenden: Die Referenzpunkte wurden nur aus dem Gebiet von Niederösterreich entnommen. In den angrenzenden Gebieten kann es daher zu Fehlklassifikationen kommen: In Niederösterreich kommt beispielsweise die Klasse 43 (Schilf) nicht vor, wohl aber im Burgenland. Aufgrund fehlender Referenzpixel würde Schilf im Burgenland einer anderen, unzutreffenden Klasse zugeordnet werden.

Ebenso kann eine Analyse der Genauigkeit der Klassifizierung nur dann durchgeführt werden, wenn die Flächen außerhalb von Niederösterreich sowie Wolkenflächen nicht klassifiziert sind.

Bei der Erstellung der Maske wurde jedoch der Raum Wien absichtlich nicht ausgeschlossen, um visuell zu überprüfen, ob die Klassifikation mit Referenzpunkten von anderen Gebieten auch gute Ergebnisse liefern kann. Für eine weitere Bearbeitung der Karte wäre es möglich, die Bereiche nachträglich zu entfernen.

Wolken und Wolkenschatten machen eine Fläche von etwa 260km² aus. Ein kleiner Teil im Westen von Niederösterreich ist von der LANDSAT-Szene nicht abgedeckt (478km²). Insgesamt konnten damit 3,85% von der Gesamtfläche Niederösterreichs nicht klassifiziert werden.

6 Durchführung der Klassifikation

6.1 Erstellen eines Rasters von Referenzpunkten

Mit Hilfe von ArcGis wurden mehrere Raster mit unterschiedlicher Rasterweite (5 km, 2,5km, 1km, 0,5km) erstellt. Die Rasterpunkte dienen als Referenzpunkte für die kNN-Klassifikation. Die Landbedeckung für jeden Rasterpunkt wurde einerseits durch visuelle Interpretation (siehe Kapitel 6.2) bestimmt und andererseits bestehenden Landbedeckungsdatensätzen, nämlich SINUS und CORINE (siehe Kapitel 6.3), entnommen (Tabelle 6.1).

Die unterschiedlichen Rasterweiten wurden aus folgenden Gründen gewählt:

- Die Güte einer Klassifikation mit kNN hängt unter anderem von der Repräsentativität des Referenzdatensatzes ab, wobei jede Klasse in ausreichender Anzahl vertreten sein sollte. Die Rasterweite muss also so gewählt werden, dass diese Bedingung auch für die seltenste Klasse gegeben ist. Die Mindestanzahl an Referenzpunkten, die für ein gutes Klassifikationsergebnis benötigt wird, ist schwierig im Vorhinein zu bestimmen. Es wurden daher unterschiedliche Rasterweiten gewählt, um den Einfluss der Punktzahl auf das Klassifikationsergebnis beobachten zu können.
- Aus Zeitgründen wurden nur die Rasterpunkte des 5km-Rasters visuell interpretiert. Der Auswerter kann bei der visuellen Interpretation die Anzahl der Punkte in einzelnen Klassen beeinflussen. Fällt nämlich ein Referenzpunkt auf ein Mischpixel, bei dem es nicht eindeutig ist, welcher Klasse es angehört, so wird der Referenzpunkt in die „geringer vertretene“ Klasse verschoben. Befindet sich ein Referenzpunkt im regelmäßigen Raster beispielsweise zwischen den Kategorien „Nadelwald“ und „Vegetationsfrei“, so wird er fast immer in die Klasse „Vegetationsfrei“ verschoben, weil die Häufigkeit der Klasse „Nadelwald“ um ein Vielfaches

Rasterweite (km)	Anzahl der Rasterpunkte	Quelle Referenzinformation
5,0	ca. 750	Visuelle Interpretation
2,5	ca. 3 000	SINUS bzw. CORINE
1,0	ca. 18 800	
0,5	ca. 70 900	

Tabelle 6.1: Anzahl der Rasterpunkte bei unterschiedlichen Rasterweiten

größer ist als jene der Klasse „Vegetationsfrei“. Dadurch sind bei der visuellen Interpretation auch seltene Klassen im Referenzdatensatz gut vertreten und kleinere Rasterweiten nicht erforderlich.

- Für die Übernahme der Landbedeckungsklasse aus den Datensätzen SINUS und CORINE ist die Anzahl der Referenzpunkte in den Klassen proportional zum Anteil der Klassen an der Gesamtfläche. Damit auch die kleinflächigen Landbedeckungskategorien mit ausreichender Referenzpunktezahl repräsentiert sind, muss daher mit einem engeren Raster gearbeitet werden als bei der Gewinnung von Referenzinformation durch visuelle Interpretation.
- Ein weiterer Grund, warum bei der Übernahme der Landbedeckungsklasse aus den Datensätzen SINUS und CORINE eine größere Anzahl an Referenzpunkten benötigt wird, besteht darin, dass ein Teil der Referenzpunkte fehlerhaft ist (Mischpixel, zeitliche Veränderungen, Klassifikationsfehler, Fehler durch Generalisierung) und die Referenzdatensätze nachträglich bereinigt werden müssen.

Die Rasterweiten wurden so gewählt, dass die Rasterpunkte jeweils in Pixelmitte liegen, damit jedem Rasterpunkt eindeutig und leicht erkennbar ein Pixel zugeordnet werden kann. Die realisierten Rasterweiten betragen somit ein Vielfaches der Pixelgröße (30 m) und weichen geringfügig von den ursprünglich angestrebten Rasterweiten ab (z.B.: 1 km → 990 m). In weiterer Folge werden zur Vereinfachung die ursprünglich angestrebten Rasterweiten genannt.

6.2 Visuelle Interpretation

An jedem Rasterpunkt des 5km-Rasters wurde die aktuelle Landbedeckung durch visuelle Interpretation (Grundlage: LANDSAT-Bild bzw. Orthophotos¹) ermittelt. Mit Hilfe dieses Referenzdatensatzes wurde eine aktuelle Landbedeckungskarte erstellt. Die Landbedeckungskarten, die unter Verwendung der Referenzdatensätze SINUS und CORINE erstellt wurden, wurden unter anderem anhand dieser aktuellen Landbedeckungskarte beurteilt.

Für die visuelle Interpretation wurden die Klassen ähnlich jenen des SINUS-Datensatzes ausgewählt (siehe Tabelle 4.4). Insgesamt werden 3 Klassen hinzugefügt. Zunächst Klasse 44 („unbestimmbare landwirtschaftliche Flächen mit viel Biomasse“) und Klasse 45 („unbestimmbare landwirtschaftliche Flächen mit wenig Biomasse“), da die Zuordnung der Flächen zu Acker oder Grünland nicht eindeutig war. Diese Flächen besitzen teilweise die Eigenschaften von Ackerflächen (zu unterschiedlichen Zeitpunkten, unterschiedliche Biomasseanteile), ihre Form weist jedoch auf Grünland hin (nicht rechteckig).

¹ Die Orthophotos befinden sich auf der Homepage des Landes Niederösterreich (NÖ Karten): |<http://www.noel.gv.at/index.wai.html>|

Weiters wurde die SINUS-Klasse 90 („Wolken“, „Schatten“, „nicht klassifizierbar“) in die Klassen 90 („Wolken“) und 91 („Schatten“¹) aufgeteilt.

Die Acker-Kategorien (Klassen: 60, 61, 62 und 63) wurden unverändert von SINUS übernommen, da es bei der visuellen Interpretation des zu klassifizierenden Bildes möglich ist, zwischen „reifer Ackervegetation“, „grüner Ackervegetation“ und „offenem Boden“ zu unterscheiden. Für einen späteren Vergleich mit den Landbedeckungskarten, die unter Verwendung der Referenzdatensätze SINUS und CORINE erstellt wurden, können diese Klassen zu einer zusammengefasst werden (Vorteil von kNN-Klassifikation), welche dann der in Kap. 6.3.1 erwähnten Klasse 65 entspricht.

6.3 Übernahme der Landbedeckungsklasse aus SINUS und CORINE

An jedem Rasterpunkt des 2,5km-, 1km- und 0,5km-Rasters wird die Landbedeckungsklasse aus den Datensätzen SINUS und CORINE ausgelesen.

6.3.1 SINUS

Der auf SINUS beruhende Referenzdatensatz muss überarbeitet werden, da SINUS nicht aktuell ist und einige absehbare Probleme auftreten:

- Punkte, die im aktuellen Satellitenbild auf Wolken oder Wolkenschatten fallen, besitzen durch das Übernehmen der Landbedeckungskategorie aus SINUS eine Klasse, die nicht zutreffend ist (siehe Abbildung 6.1). Diese Punkte werden entfernt.
- Referenzpunkte, für welche aus SINUS die Klasse 90 („nicht klassifizierbar“) übernommen wird, werden entfernt, weil sich Schatten und Wolken nicht mehr an derselben Stelle befinden.
- Die Ackerflächen würden nur durch Zufall dieselbe Ackerkategorie aufweisen; wo es vor Jahren „offener Boden“ war, kann heute eine Bedeckung durch eine Feldfrucht bestehen (z.B. „Grüne Ackervegetation“). Daher werden alle Ackerklassen (60, 61, 62, 63) in eine einzige Klasse (65) zusammengefasst.

In Niederösterreich kommt die Klasse 43 („Schilf“) nicht vor. Zum gegebenen Aufnahmezeitpunkt ist außerdem die Klasse 12 („Schnee“, „Eis“) nicht vertreten.

1 Schatten, der von Wolken verursacht wird; der topographisch bedingter Schatten wurde radiometrisch korrigiert

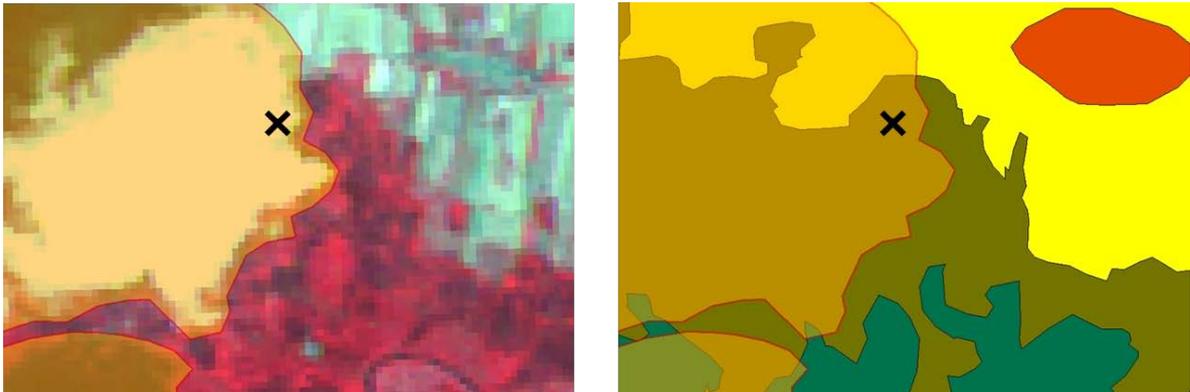


Abbildung 6.1: Fehlerhafter Punkt aufgrund von Wolken. Für den gezeigten Referenzpunkt wird aus SINUS bzw. CORINE die Landbedeckungskategorie „Wald“ übernommen, aus dem aktuellen Satellitenbild aber die Spektralwerte einer Wolke.

6.3.2 CORINE

Die CORINE-Kategorien sind im Gegensatz zu den SINUS-Kategorien von zeitlichen Veränderungen nicht so stark beeinflusst.

Bei CORINE handelt es sich teilweise um Landnutzungskategorien. So sind beispielsweise die Ackerflächen nicht durch die Feldfruchtart charakterisiert, sondern nach der Form und Bewirtschaftungsart (siehe CORINE-Kategorien, Tabelle 4.3). Daher müssen zunächst nur die Referenzpunkte entfernt werden, die im aktuellen Satellitenbild auf Wolken oder Wolkenschatten fallen.

6.4 kNN-Methode

Das kNN-Programm beinhaltet 3 Module, mit deren Hilfe die Referenzdaten aufbereitet und analysiert werden und die Klassifikation durchgeführt wird. (Koukal, 2008)

6.4.1 Modul Readimage

Mit dem Modul Readimage werden die Werte der Referenzpixel aller Kanäle aus dem Bild selektiert. Zu diesen Kanälen zählen die 7 Spektralkanäle des LANDSAT-Bildes und die Werte für Geländeneigung und Beleuchtung. Die Geländeneigung kann beispielsweise als ein Unterscheidungsmerkmal zwischen „Grünland“ und „Acker“ verwendet werden („Grünland“ kommt häufig in geneigtem Gelände vor, „Acker“ eher nicht).

6.4.2 Modul Crossvalidation

Mit dem Modul Crossvalidation werden Statistiken berechnet, um die Genauigkeit der Klassifikation im Vorhinein zu schätzen. Dadurch kann eine optimale Einstellung der kNN-Parameter und die Eliminierung von Ausreißern („outlier“) erfolgen.

Aus dem Referenzpunktedatensatz wird für einen Punkt nach dem anderen die Landbedeckungsklasse als eine Unbekannte behandelt, und mit Hilfe der restlichen Punkte mit der k-Nächste-Nachbarn Methode klassifiziert. Die Klasse, die das Pixel auf diese Weise erhält, wird mit der tatsächlichen (Soll-)Klasse verglichen.

Als Ergebnis der Crossvalidation (Kreuzvalidierung) wird eine Fehlermatrix erstellt, das ist eine Kreuztabelle, deren Spalten- und Zeilenanzahl der Anzahl an vorkommenden Klassen entspricht. Die Spalten enthalten Daten für die Soll-Klassen, die Zeilen enthalten Daten für die klassifizierten Klassen. In jeder Zeile steht die Anzahl der Pixel, die zu der betreffenden Soll-Klasse gehören und bei der Klassifikation der betreffenden klassifizierten Klasse zugewiesen wurden. Die Zahlen in der Diagonale zeigen die Anzahl der Pixel, die richtig klassifiziert wurden.

Aus dieser Fehlermatrix können verschiedene Fehlermaße abgeleitet werden: (Congalton und Green, 1999)

Die **Hersteller-Genauigkeit** („producer’s accuracy“) erhält man durch Division der richtig klassifizierten Pixel durch die Gesamtzahl der Punkte die tatsächlich dieser Klasse angehören. In Tabelle 6.2 beträgt die Hersteller-Genauigkeit für die Klasse „Vegetationsfrei“ 82%, da insgesamt 9 von 11 Referenzpixel dieser Klasse richtig klassifiziert wurden.

Die **Nutzer-Genauigkeit** („user’s accuracy“) wird berechnet, indem man die Anzahl der korrekt klassifizierten Punkte in einer Klasse durch die Summe der Punkte, die insgesamt dieser Klasse zugeordnet wurden, dividiert. In dem angeführten Beispiel (siehe

		Soll-Klassen						Summe	User.Acc.
		Vegfrei	Wasser	Siedlung	Grünland	Wein	Siedlung		
Klassif.ergebnis	Vegfrei	9	0	1	0	0	0	10	0.90
	Wasser	0	8	0	0	0	0	8	1.00
	Wald	0	0	41	0	0	0	41	1.00
	Grünland	0	0	0	4	1	0	5	0.80
	Wein	0	0	1	1	19	3	24	0.79
	Siedlung	2	0	0	0	3	27	32	0.84
	Summe:	11	8	43	5	23	30	120	
Prod.Acc.:	0.82	1.00	0.95	0.80	0.83	0.90			

Kappa: 0.868

Overall Acc.: 0.900

Tabelle 6.2: Beispiel einer Fehlermatrix

Tabelle 6.2) beträgt dieser Wert für die Klasse „Vegetationsfrei“ 90%, da von den insgesamt 10 Referenzpixel, 9 richtig klassifiziert wurden. Die Nutzergenauigkeit ist ein Maß für die Verlässlichkeit der erzeugten Karte und gibt an, wie gut die Karte die Realität wiedergeben würde.

Die **Gesamtgenauigkeit** („Overall Accuracy“) wird folgendermaßen berechnet: die Werte, die in der Diagonalen der Matrix angeführt sind, werden aufsummiert und die Summe durch die Gesamtzahl an Fällen (=Gesamtzahl an Referenzpixel) dividiert.

Der **Kappa**-Wert (\hat{K}) ist ein über alle Klassen berechnetes Maß für die Klassifikationsgenauigkeit. Der Unterschied zu der Gesamtgenauigkeit besteht darin, dass Kappa die zufällige Übereinstimmung zwischen Referenz und Klassifikation eliminiert. Kappa besagt, wie viel besser eine Klassifizierung ist als eine reine Zufallsübereinstimmung. Ein Wert von 1 beschreibt eine perfekte Klassifizierung, ein Wert von 0 hingegen ein Zufallsergebnis.

Der Kappa Wert hat außerdem eine Bedeutung für den Vergleich von Fehlermatrizen. Anhand des Kappawertes kann man feststellen, ob eine Klassifikation signifikant besser ist als eine andere.

Die hier beschriebenen Größen sind unter anderem von der Wahl des k-Wertes abhängig. Grundsätzlich wird mit einem größeren k die Genauigkeit verbessert.

6.4.3 Modul Mapmaker

Mit dem Modul Mapmaker erfolgt die eigentliche Klassifikation. Jedes Bildpixel wird mit Hilfe der Referenzpunkte unter Verwendung der nächsten Nachbarn einer Klasse zugewiesen.

Im Modul Mapmaker können maximale Distanzen im geographischen Raum (lagemäßig, höhenmäßig) angegeben werden. Dadurch können z.B. Klassen in bestimmten Höhen ausgeschlossen werden. Felsen kommen beispielsweise eher nicht in 200 m Höhe vor, Acker wiederum nicht in einer Höhe von 1500m.

6.5 Fehlerquellen

Abhängig von der Klassifizierungsmethode oder dem Referenzdatensatz können folgende Probleme bzw. Fehler auftreten:

- Offensichtliche Änderung der Landbedeckung zwischen den Zeitpunkten der Erhebung des Referenzdatensatzes und der Bildaufnahme
- Nichttrennbarkeit zweier Klassen anhand ihrer spektralen Eigenschaften (Überlagerung der Spektralwerte)

- Mischpixel
- Geringe Zahl an Referenzpunkten bei manchen Klassen
- Schlechte Pixelzuordnung aufgrund von Fehlern in den als Referenz verwendeten Landbedeckungsdatensätzen (bei CORINE durch Generalisierung, bei SINUS durch die automatische Auswertung)
- Bei Referenzdaten aus visueller Interpretation: subjektive Interpretationsfehler; da es dazu jedoch keine verlässliche Information gibt, werden diese Fehler nicht weiter beachtet.

6.6 Methoden zur Reduzierung von Klassifikationsfehlern

6.6.1 Optimierung der kNN-Parameter

Gewichtung der Spektralkanäle

Bei der kNN-Klassifikation wird die Euklidische Distanz zur Feststellung der nächsten Nachbarn mit Hilfe der Pixelwerte in den unterschiedlichen Kanälen berechnet. Die Kanäle sollen so gewichtet werden, dass ein bedeutendes Unterscheidungsmerkmal herbeigeführt wird. Wenn man beispielsweise einen Unterschied zwischen Vegetationsklassen erreichen möchte, so ist es sinnvoll, die Infrarot-Kanäle stärker zu gewichten als die Kanäle des sichtbaren Lichts.

In dieser Diplomarbeit wurde nicht näher auf die Feststellung einer optimalen Gewichtung der Spektralkanäle eingegangen. Es wurden für jede Klasse die Mittelwerte in den einzelnen Kanälen berechnet, um die Unterschiede zu verdeutlichen. Diejenigen Kanäle, in denen die Unterschiede am größten waren, wurden ausgewählt.

Die Suche nach optimalen Gewichten der Spektralkanäle ist zeitaufwendig. Ein Optimum für alle Klassen konnte nicht gefunden werden. Die Verbesserung einer Klasse bedeutet manchmal eine Verschlechterung einer anderen.

Die Spektralkanäle 4,5,6 und die Neigung wurden bei der Klassifizierung 5- bis 10fach stärker gewichtet als die Kanäle 1,2,3 und 7. Die Beleuchtung wurde nicht für die Klassifikation verwendet.

Wahl des k-Wertes

Die Wahl des k-Wertes wirkt sich auf die Klassifikationsgenauigkeit aus. Wie in Abbildung 6.2 ersichtlich, steigt die Genauigkeit mit dem k-Wert, erreicht bei einem gewissen k-Wert den Höhepunkt und fällt dann wieder. Diese mit steigendem k-Wert eintretende Verbesserung ist durch das Hinzuziehen weiterer Informationen aus dem spektralen

Umfeld begründbar. Bei kleinen k -Werten ist der k NN-Algorithmus hingegen anfällig gegenüber fehlerhaften Referenzinformationen. Eine Erklärung dafür, dass ab einem gewissen k -Wert Kappa wieder fällt, ist, dass die spektrale Eigenschaft von einem „weit entfernten“ Referenzpixel nicht mehr der spektralen Eigenschaft des zu klassifizierenden Pixels entspricht. Ein weiterer Grund ist, dass manche Klassen wenig Referenzpunkte enthalten und bei einem großen k -Wert die Anzahl der betrachteten k nächsten Nachbarn die Anzahl der Referenzpunkte in der Klasse überschreiten kann.

Die Informationen der k nächsten Referenzpixel können in Abhängigkeit von der spektralen Distanz unterschiedlich gewichtet werden.

In diese Arbeit wurden für die Klassifikation k -Werte von 1 bis 10 verwendet (in der Abbildung 6.2 findet Kappa seinen Höchstwert bei $k=10$).

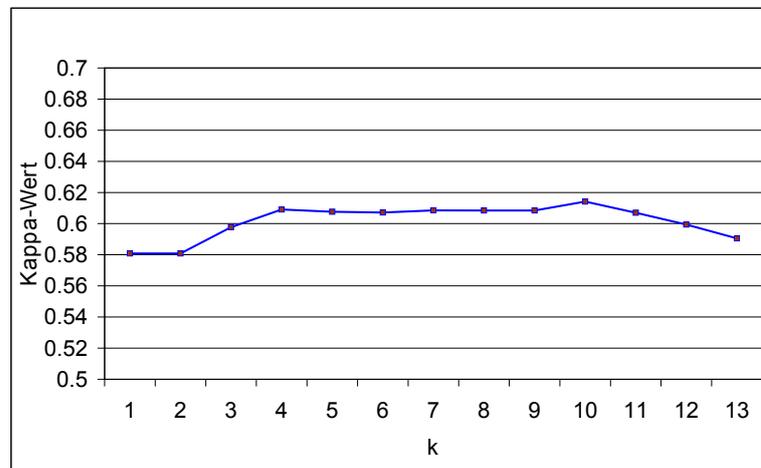


Abbildung 6.2: Kappa-Wert in Abhängigkeit von k

Suchradius

Die Suche nach Referenzpixeln kann geographisch sowohl in horizontaler Richtung als auch in vertikaler Richtung eingeschränkt werden. Diese Einschränkung erfolgt durch die Angabe einer maximalen geographischen Entfernung vom zu klassifizierenden Pixel. Das kann zu guten Ergebnissen führen, wenn die Landbedeckungskategorien regional aufgeteilt sind, kann sich aber auch negativ auswirken, wenn für Klassen geringer Häufigkeit keine ausreichende Anzahl an gleichmäßig über das Bild verteilten Referenzpixel vorhanden ist.

In Niederösterreich ist diese Einschränkung nach horizontaler oder vertikaler Entfernung nicht zielführend, weil zum einen die Höhenunterschiede allgemein nicht so gravierend sind und zum anderen, weil keine hinreichend signifikante regionale Aufteilung der Landbedeckungsklassen vorhanden ist.

6.6.2 Bereinigen der Referenzdatensätze

Wie in Kap. 6.3.1 beschrieben, wurden gleich zu Beginn jene Referenzpixel entfernt, welche zu vorhersehbaren Problemen bei der Klassifikation führen würden (z.B. aufgrund von Wolken oder Wolkenschatten). Um auch andere, zwar vermutete aber ad-hoc nicht lokalisierbare Fehler im Referenzdatensatz zu beseitigen, wurden folgende Methoden getestet: das Verschieben von Referenzpunkten und die Suche und anschließende Eliminierung von Ausreißern.

Verschieben von Referenzpunkten

Um zu vermeiden, dass Referenzpunkte auf Mischpixel liegen, wurden die Referenzpunkte teilweise verschoben. Dies erfolgte in einem automatisierten Prozess: Wenn die umgebende Fläche (9 Pixel) eines Referenzpunktes bezogen auf die als Referenz verwendete Landbedeckungskarte (z.B. SINUS) inhomogen ist, dann wird der Referenzpunkt solange verschoben, bis er sich in einer homogenen Umgebung befindet. Die Verschiebung erfolgt, wie in Abbildung 6.3 gezeigt, jeweils um 2 Pixel von der ursprünglichen Position in diagonaler Richtung. Wird für einen Rasterpunkt keine geeignete Position gefunden, so wird er aus dem Referenzdatensatz eliminiert.

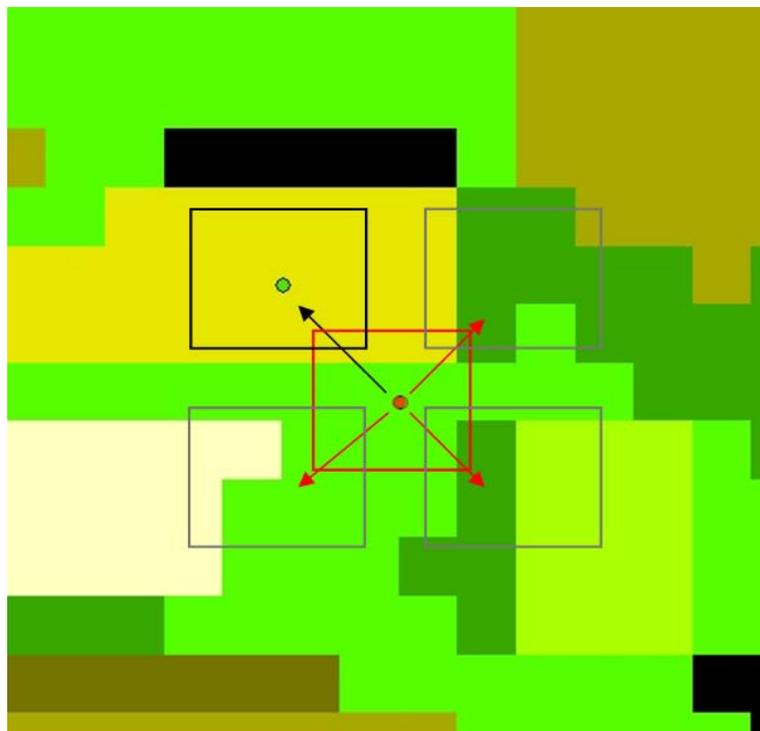


Abbildung 6.3: Verschieben von Referenzpunkten

Diese Methode wurde nur beim SINUS-Referenzdatensatz angewendet, da bei CORINE durch die großflächige Abgrenzung die Umgebung der meisten Referenzpunkte ohnehin homogen ist und Mischpixel aufgrund der Generalisierung auch innerhalb einer abgegrenzten Fläche vorkommen.

Angewendet beispielsweise auf die Referenzpunkte des 2,5km-Rasters wurden von den insgesamt 3000 Punkten auf diese Weise ca. 1200 Punkte verschoben und ca. 550 Referenzpunkte eliminiert. Die restlichen 1250 Referenzpunkte blieben unverändert.

Eliminierung von Ausreißern

Ein wichtiger Teil dieser Diplomarbeit ist die Suche und Eliminierung von Ausreißern („outlier“). Diese können durch Veränderung der Bodenbedeckung, aufgrund einer fehlerhaften Klassifikation des Referenzdatensatzes, aufgrund der Generalisierung oder infolge von Mischpixeln auftreten.

Prinzip der Outlier-Suche

Die Outlier-Suche erfolgt mit Hilfe der Crossvalidation. Für jeden Referenzpunkt wird protokolliert, wie oft er während eines gesamten Crossvalidation-Durchlaufs als nächster Nachbar ausgewählt wird und wie viele Referenzpunkte dadurch richtig klassifiziert werden. Aufgrund dieser Informationen können die Referenzpunkte in 3 Gruppen eingeteilt werden:

1. „Gute“: Referenzpunkte, die häufig als nächster Nachbar ausgewählt wurden und meistens zu einer richtigen Klassifikation geführt haben.
2. „Schlechte“: Referenzpunkte, die häufig als nächster Nachbar ausgewählt wurden, meistens aber zu einer falschen Klassifikation geführt haben.
3. „Indifferente“: Referenzpixel, über die keine Aussage getroffen werden kann, weil sie entweder zu selten als nächster Nachbar ausgewählt wurden oder weil sie weder zu einem überwiegenden Teil zu einer richtigen noch zu einer falschen Klassifikation geführt haben.

Die „Guten“ werden zur Seite gelegt, die „Schlechten“, bei denen es sich mit großer Wahrscheinlichkeit um die gesuchten Outlier handelt, werden entfernt. Mit den „Indifferenten“ kann der Vorgang wiederholt werden. Dieser Ablauf wird solange wiederholt, bis keine Outlier mehr gefunden werden oder die Klassifikationsgenauigkeit nicht mehr signifikant steigt (Abbildung 6.4). Der bereinigte Referenzdatensatz setzt sich schließlich aus den „Guten“ und den nach dem letzten Durchgang noch verbleibenden „Indifferenten“ zusammen.

Würde man nur die „Guten“ als Referenzpunkte für die Klassifikation verwenden, so wäre zwar die mit Hilfe der Crossvalidation geschätzte Klassifikationsgenauigkeit sehr hoch, jedoch würden die Referenzpunkte die tatsächliche spektrale Verteilung

der Klassen möglicherweise nicht mehr ausreichend repräsentieren. Dies würde bei der Klassifikation „neuer“ Pixel voraussichtlich zu schlechten Ergebnissen führen.

Die Outlier-Suche wird mit $k=1$ durchgeführt, da der kNN-Algorithmus bei $k=1$ besonders anfällig gegenüber Ausreißern ist und somit Ausreißer leichter auffindbar sind.

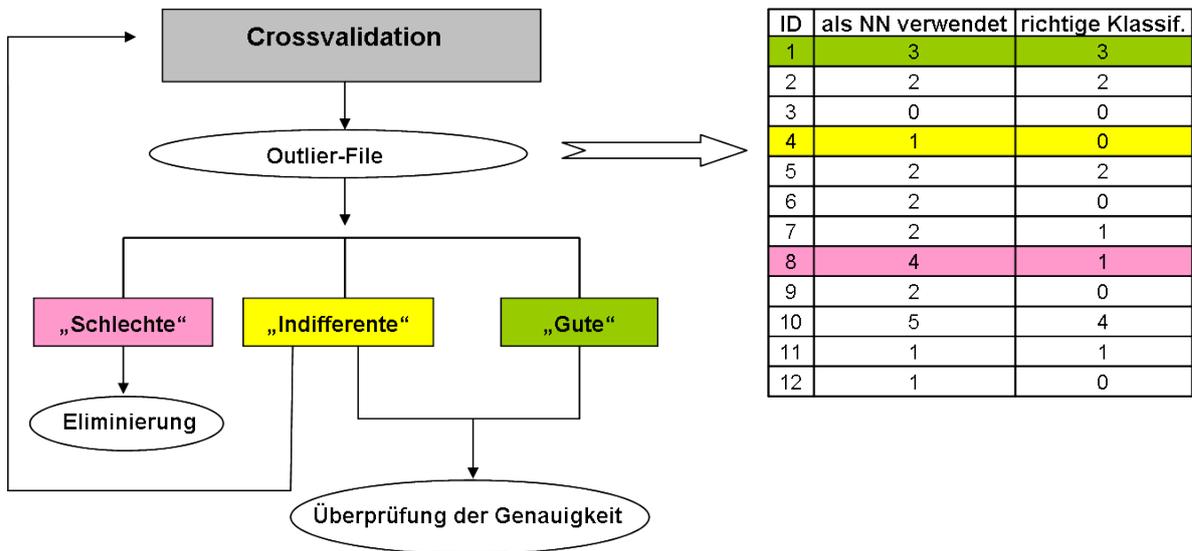


Abbildung 6.4: „Outlier“-Suche

Um zu testen, inwieweit Outlier mit der beschriebenen Methode gefunden werden, wurden Änderungen im Referenzdatensatz vorgenommen, die eine Veränderung der Landbedeckung simulieren sollen. Bei 10% der Referenzpunkte der Klasse „Wald“ (Klassen 31, 32 und 33) wurde die Landbedeckung von „Wald“ auf „Grünland“ (Klassen 41 und 42) und bei 10% der Referenzpunkte der Klasse „Siedlung“ (Klassen 71 und 72) wurde die Landbedeckung auf „Acker“ (Klassen 60, 61 und 62) geändert.

Durch die oben beschriebene Outlier-Suche konnten nach 5 Iterationen 53% (47 von 88) der von Wald auf Grünland und 41% (5 von 12) der von Siedlung auf Acker geänderten Referenzpunkte als Outlier identifiziert werden (siehe Tab. 6.3). Eine Erklärung, warum die Acker-Outlier zu einem geringeren Teil gefunden wurden als die Wald-Outlier, könnte sein, dass sich die beiden Klassen „Siedlung“ und „Acker“ (vor allem reife Ackervegetation, offener Boden) spektral ähnlicher sind als die Klassen „Wald“ und „Grünland“.

Versuch: 10% Wald → Grünland (88 Pkt.)

10% Siedlung → Acker (12 Pkt.)

Punkte gesamt	Outlier gefunden	Davon Grünland	Grünland anfangs	Davon Acker	Acker anfangs
2274	176	18	88	2	12
1862	95	13	70	2	10
1612	58	8	57	1	8
1451	43	4	49	0	7
1332	28	4	45	0	7
		47		5	

Nach 5 Runden → Wald wird zu 53 % gefunden, Siedlung zu 41 %

Tabelle 6.3: Auffinden absichtlich in den Datensatz eingefügter Ausreißer (Test)

7 Ergebnisse

Die Ergebnisse werden in Form von Karten für die visuelle und in Form von Fehlerstatistiken für die quantitative Beurteilung präsentiert.

Da es nicht möglich ist, alle Ergebnisse in dieser Arbeit ausführlich zu dokumentieren, werden hier nur die wichtigsten Ergebnisse aufgezählt anhand von Beispielen gezeigt:

- Gegenüberstellung des SINUS-Datensatzes mit der Klassifikation mit Referenzpunkten aus SINUS (Ausschnitt)
- Gegenüberstellung des CORINE-Datensatzes mit der Klassifikation mit Referenzpunkten aus CORINE (Ausschnitt)
- Fehlerstatistiken (mit Kontrollpunkten) für Klassifikationen mit verschiedenen Referenzdatensätzen
- Auswirkungen der Eliminierung von Ausreißern

7.1 Karten

7.1.1 Klassifikation mit Referenzpunkten aus CORINE

In Abbildung 7.1 bis Abbildung 7.3 wird für einen Ausschnitt Niederösterreichs das Ergebnis der Klassifikation mit Referenzpunkten aus CORINE dem CORINE-Datensatz gegenübergestellt.

Es wird ersichtlich, dass mit Hilfe der kNN-Klassifikation nicht nur eine Aktualisierung des CORINE-Datensatzes erfolgt, sondern auch eine Verfeinerung in der Darstellung der Landbedeckung. So werden z.B. Straßen dargestellt, die bei CORINE nicht erfasst worden sind. Auch beispielsweise Wald- und Grünlandflächen werden räumlich feiner abgegrenzt, wodurch die Landschaftsstruktur besser erkennbar wird.

Kategorien, die sich auf die Landnutzung (und nicht auf die Landbedeckung) beziehen, werden in der Klassifikation teilweise nicht berücksichtigt. Dies geschieht dann, wenn für eine Kategorie nur wenige Referenzpunkte vorhanden sind und es eine spektral ähnliche Klasse mit einer großen Anzahl an Referenzpunkten gibt, welche sich auf die Landbedeckung (und nicht auf die Landnutzung) bezieht. So werden beispielsweise Flächen, die laut CORINE zur Klasse „Sport- und Freizeitanlagen“ (Landnutzung)

gehören, als „Grünland“ (Landbedeckung) klassifiziert. Gibt es hingegen keine spektral ähnlichen, dominierenden Landbedeckungsklassen, so werden die betreffenden Flächen richtig klassifiziert. Ein Beispiel dafür ist die Klasse „nicht durchgängige städtische Prägung“.

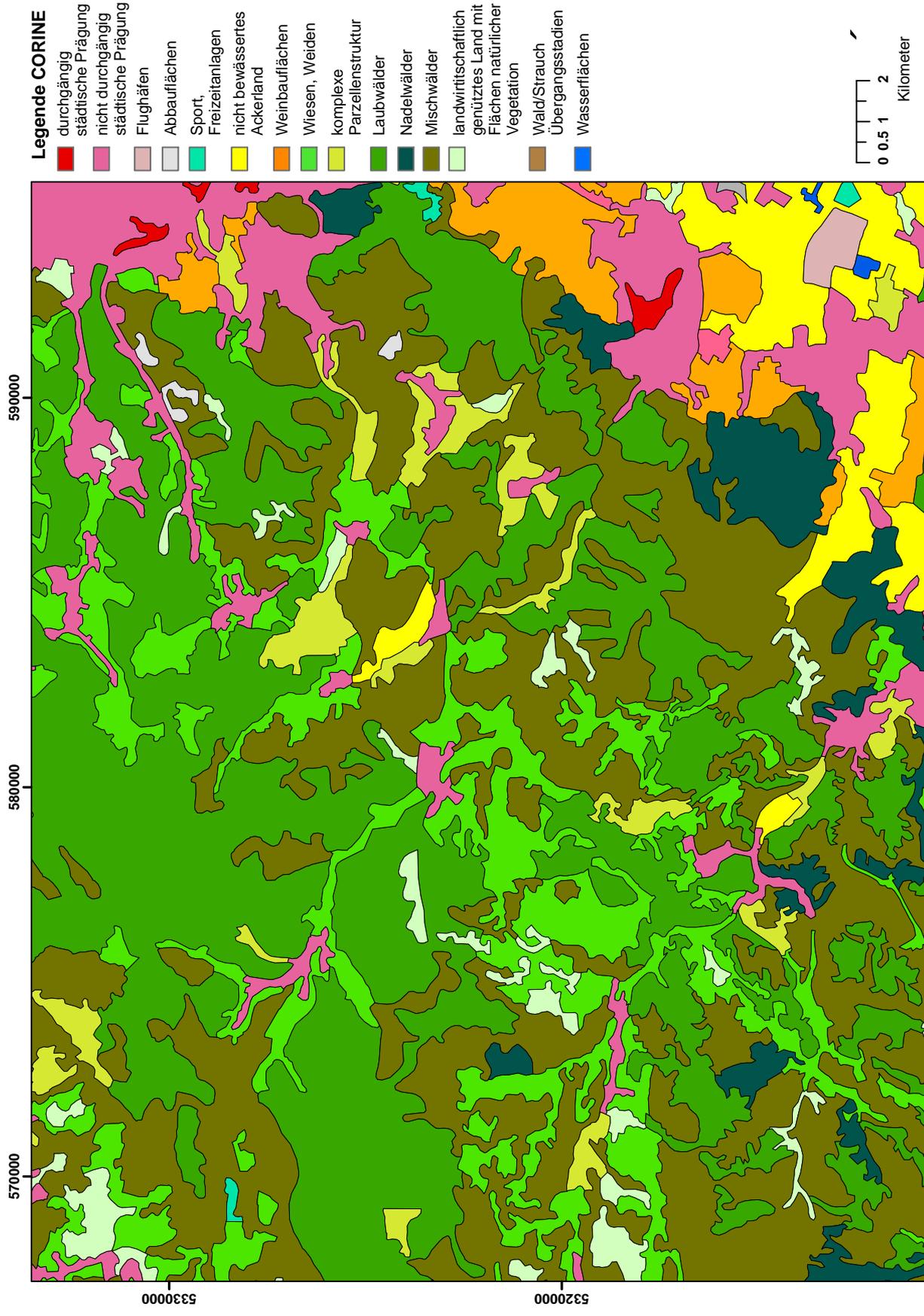
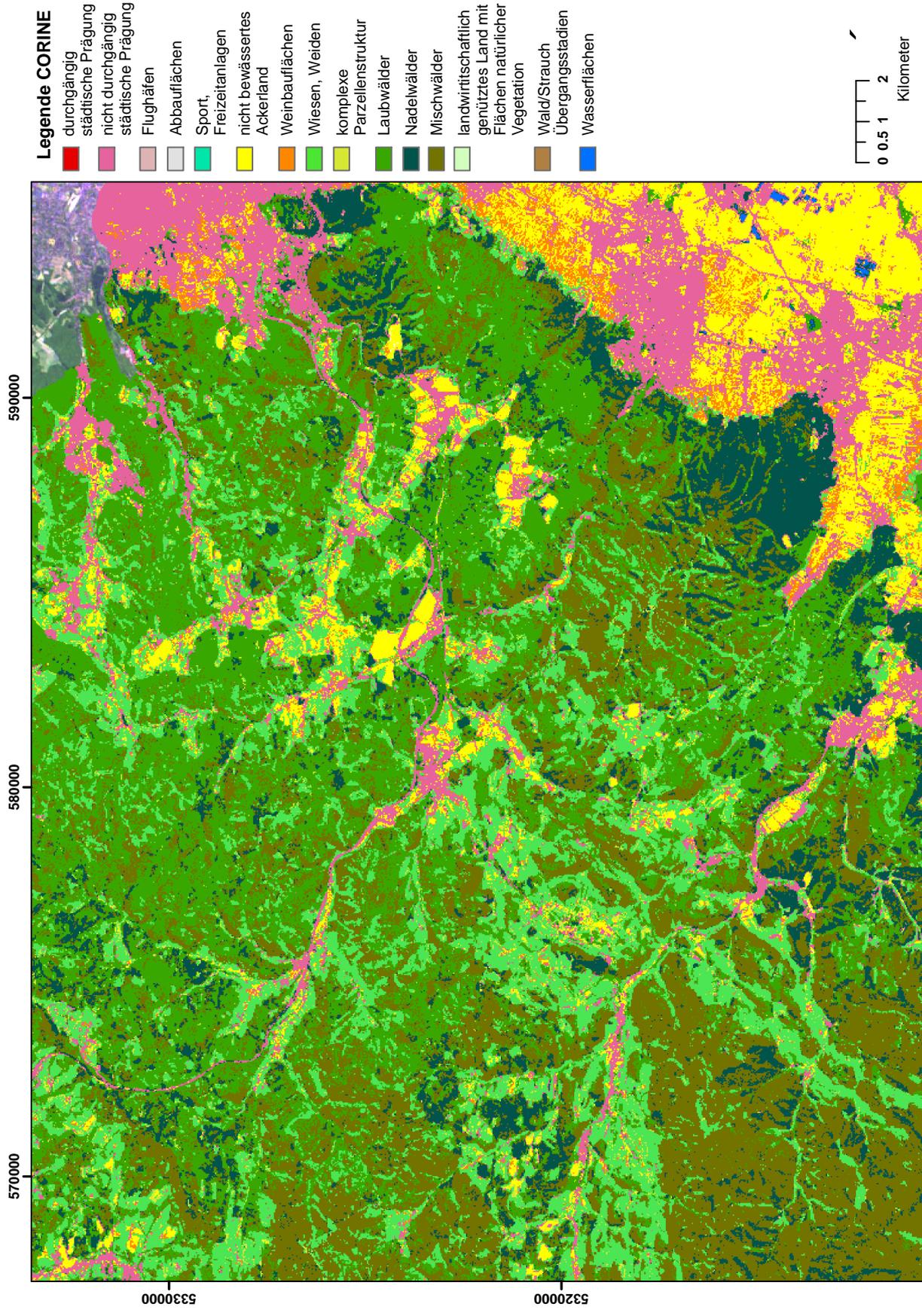


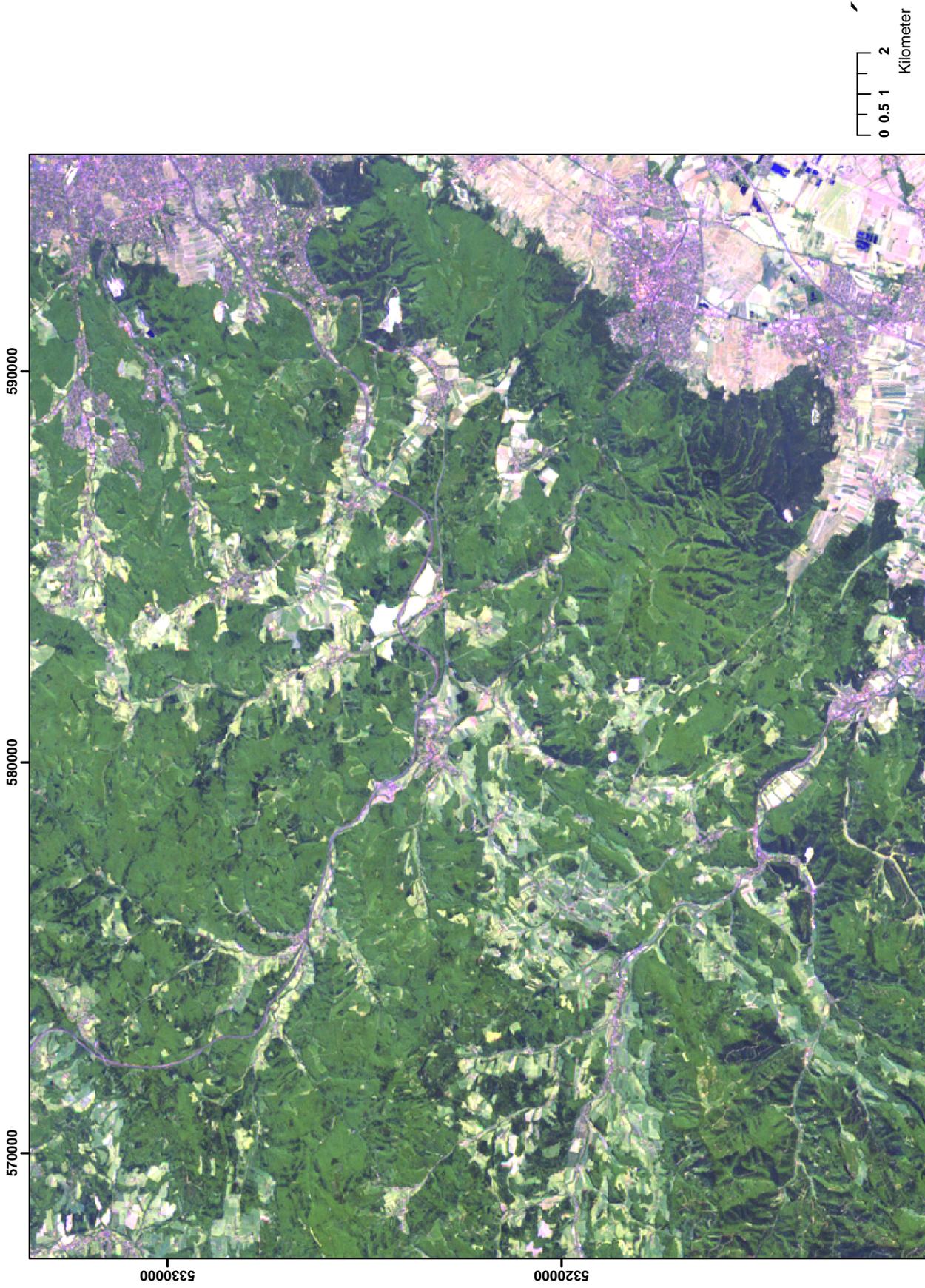
Abbildung 7.1: CORINE-Datensatz, Ausschnitt: südlicher Wienerwald

Projektion UTM, Zone 33N, WGS84



Projektion UTM, Zone 33N, WGS84

Abbildung 7.2: Klassifikation mit Referenzpunkten aus CORINE, 1km-Raster, Ausreißer-bereinigter Referenzdatensatz, k=5, Ausschnitt: südlicher Wienerwald



Projektion UTM, Zone 33N, WGS84

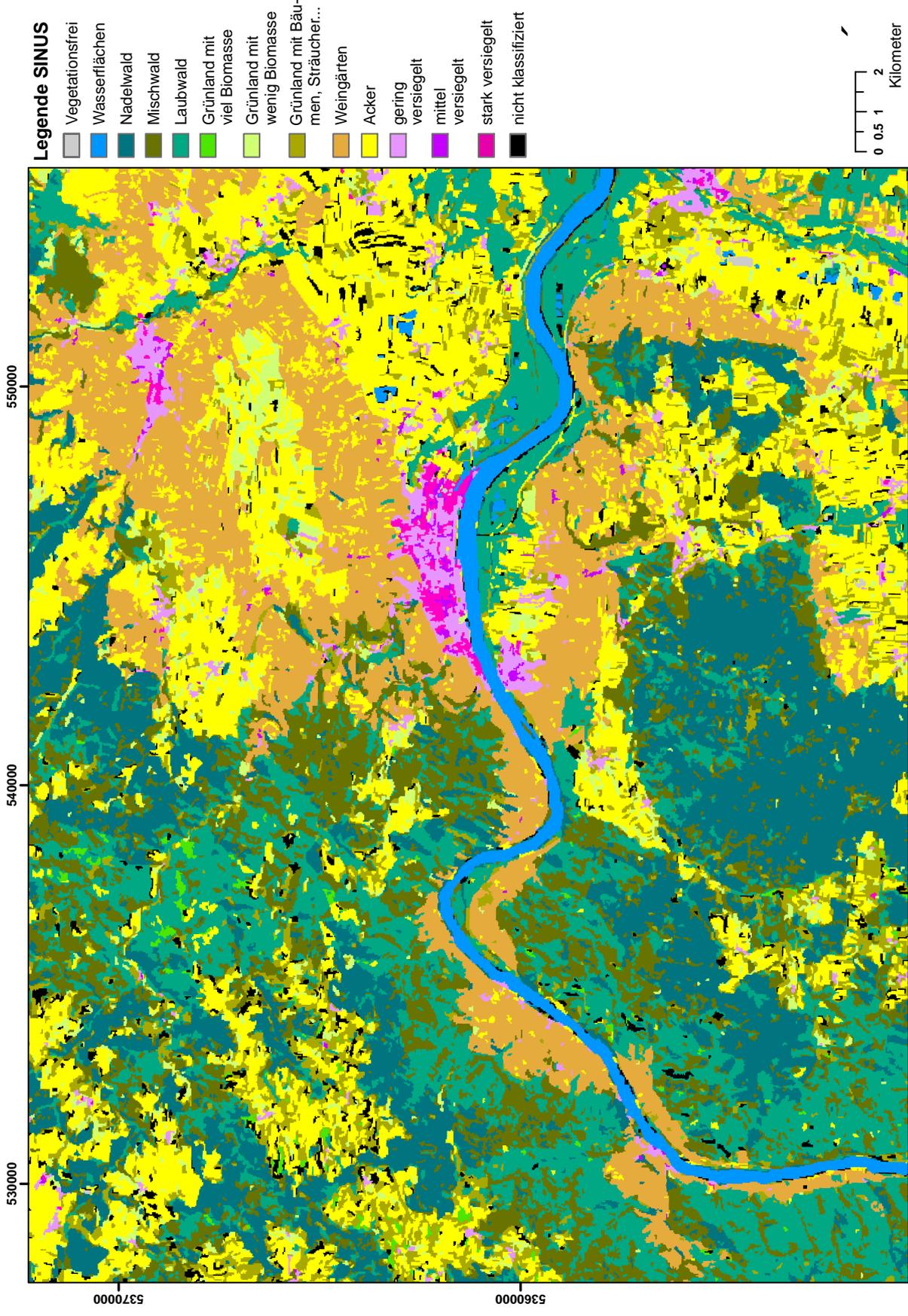
Abbildung 7.3: LANDSAT 5 TM, Bandkombination: 7,5,2, Ausschnitt: südlicher Wienerwald

7.1.2 Klassifikation mit Referenzpunkten aus SINUS

In Abbildung 7.4 bis Abbildung 7.6 wird für einen Ausschnitt Niederösterreichs das Ergebnis der Klassifikation mit Referenzpunkten aus SINUS dem SINUS-Datensatz gegenübergestellt. Beim Vergleich der beiden Datensätze muss berücksichtigt werden, dass die SINUS-Klassifikation objektbasiert erfolgt ist, die kNN-Klassifikation hingegen pixelbasiert. Ebenso, wenn auch nicht so stark wie bei CORINE, ist eine Verfeinerung der Landschaftsstruktur erkennbar. Die Unterschiede hinsichtlich Generalisierung sind jedoch nicht so groß, dass sie das Erkennen von Veränderungen der Landbedeckung wesentlich beeinträchtigen würden.

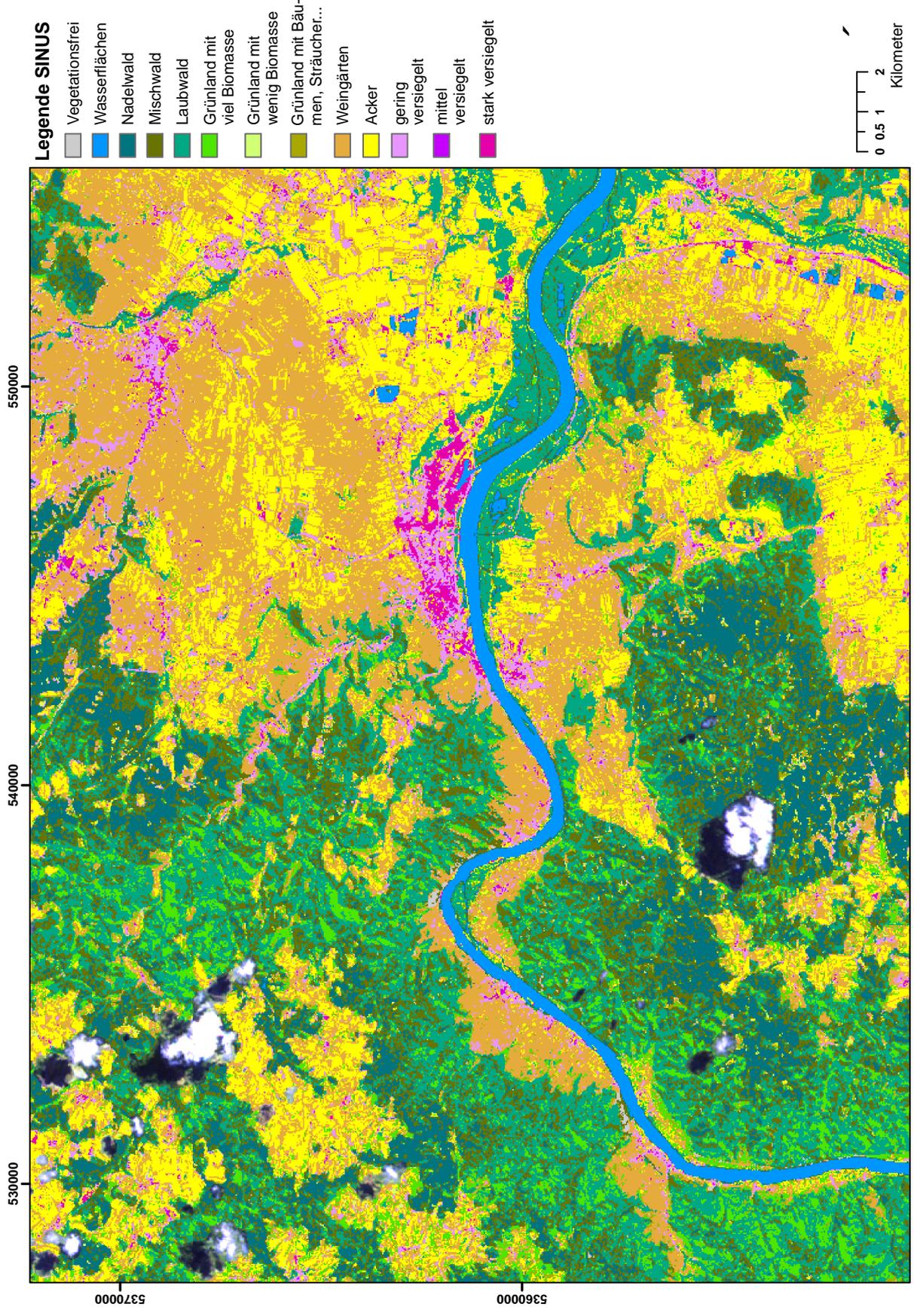
Bei der Klassifikation von Ackerflächen spielt der zeitliche Aspekt eine besonders große Rolle. Hier kommt es sowohl im Laufe eines Jahres als auch von Jahr zu Jahr zu wesentlichen Veränderungen der Landbedeckung (Beispiel „offener Boden“ - „grüne Ackervegetation“ - „reife Ackervegetation“ oder Wechsel der Feldfruchtart von Jahr zu Jahr). Es mussten für die Klassifikation daher alle Unterklassen der Klasse „Acker“ zusammengefasst werden. Pixel können auf diese Weise eindeutig als Ackerfläche klassifiziert werden, eine Differenzierung nach Entwicklungsstadium oder Feldfruchtart ist jedoch nicht möglich. Dadurch sind in der kNN-Klassifikation die einzelnen Parzellen der Ackerflächen nicht mehr erkennbar.

Die in der SINUS-Karte nicht klassifizierten Flächen (in Abb. 7.4 schwarz dargestellt) werden mit Hilfe von kNN klassifiziert.



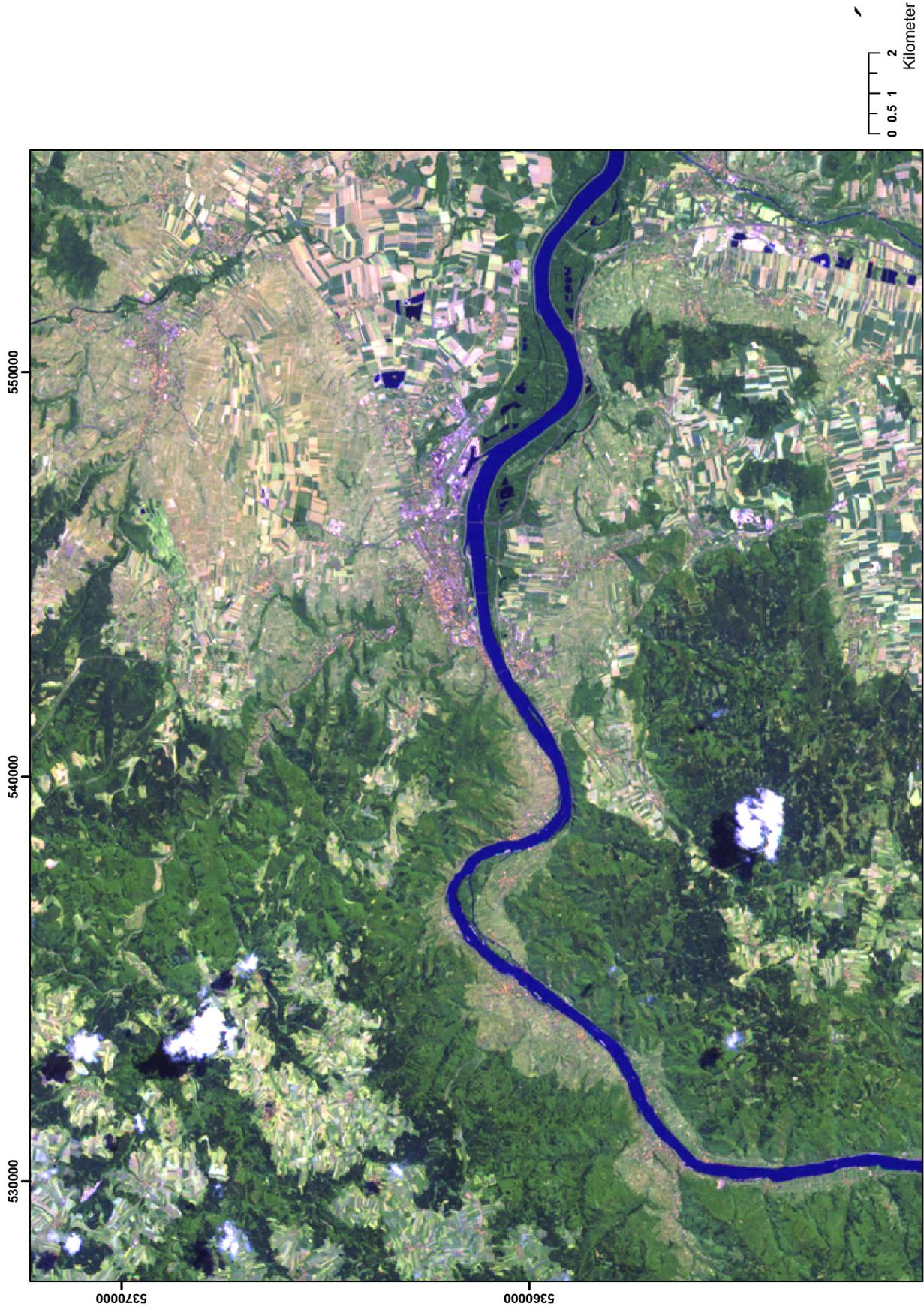
Projektion UTM, Zone 33N, WGS84

Abbildung 7.4: SINUS-Landbedeckungskarte, Ausschnitt: Krems und Umgebung



Projektion UTM, Zone 33N, WGS84

Abbildung 7.5: Klassifikation mit Referenzpunkten aus SINUS, 0,5km-Raster, reduzierter und Ausreißer-bereinigter Referenzdatensatz, k=5, Ausschnitt: Kreams und Umgebung



Projektion UTM, Zone 33N, WGS84

Abbildung 7.6: LANDSAT 5 TM, Bandkombination: 7,5,2, Ausschnitt: Krems und Umgebung

7.1.3 Bereinigen der Referenzdatensätze

Verschieben von Referenzpunkten

Ziel der Verschiebung von Referenzpunkten in homogene Flächen (siehe Kap. 6.6.2) ist es zu vermeiden, dass Referenzpunkte auf Mischpixel fallen. Durch das Verschieben von Referenzpunkten in homogene Flächen konnte allerdings keine eindeutige Verbesserung der Klassifikation erzielt werden. Ein Nachteil der Methode besteht darin, dass durch das Verschieben von Referenzpunkten in homogene Flächen kleinflächige oder linienhafte Landbedeckungsobjekte (z.B. Straßen) nicht mehr mit einer ausreichenden Anzahl an Referenzpunkten vertreten sind.

Eliminierung von Ausreißern

In Abbildung 7.7, kann man den Unterschied zwischen einer Klassifikation mit einem nicht Ausreißer-bereinigten Referenzdatensatz und einer Klassifikation mit einem Ausreißer-bereinigten Referenzdatensatz erkennen. Die Klassifikation wurde mit Referenzpunkten aus CORINE (Rasterweite 1km) durchgeführt. Für beide Klassifikationen wurde ein k -Wert von 5 verwendet. Durch das Eliminieren von Ausreißern kommt es zu einer Glättung des Klassifikationsergebnisses.

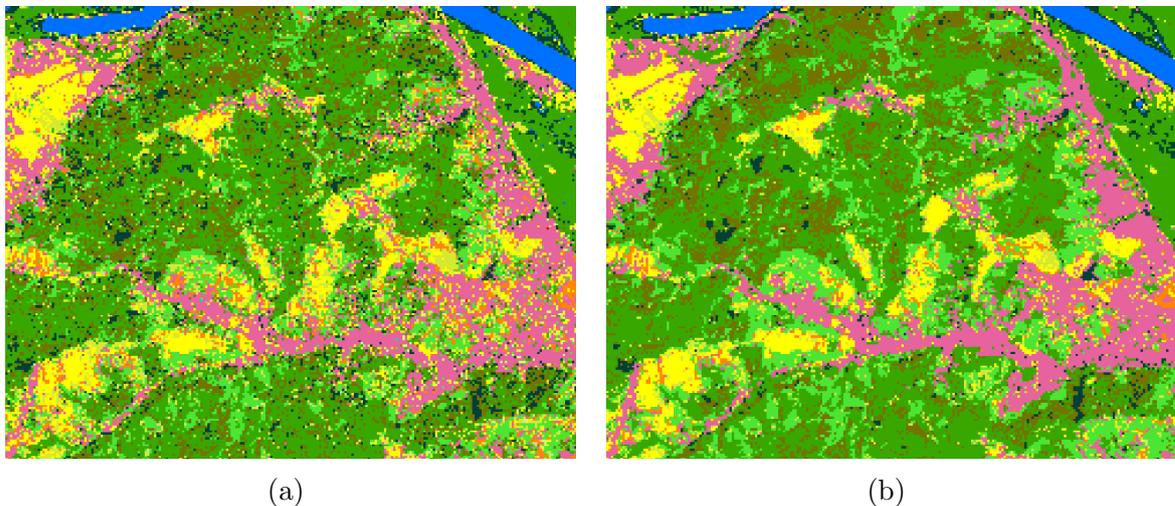


Abbildung 7.7: Klassifikation mit Referenzpunkten aus CORINE, 1km Raster, $k=5$. (a) Referenzdatensatz original, (b) Referenzdatensatz Ausreißer-bereinigt.

7.2 Fehlerabschätzung

7.2.1 Kontrollpunkte und Kreuzvalidierung

Die Schätzung der Klassifikationsgenauigkeit erfolgte teilweise durch Kreuzvalidierung und teilweise mit Hilfe von Kontrollpunkten. Kontrollpunkte sind Punkte, die ausschließlich für die Beurteilung des Klassifikationsergebnisses, nicht aber als Referenzpunkte bei der kNN-Klassifikation verwendet werden.

Um die Genauigkeit einer Klassifikation zu schätzen, die mit Referenzpunkten aus SINUS oder CORINE erstellt wurde, weist die Kreuzvalidierung Nachteile auf. So kann es vorkommen, dass ein Referenzpunkt aufgrund seiner spektralen Eigenschaften richtig klassifiziert wird, die Soll-Klasse aufgrund von Veränderungen der Landbedeckung jedoch nicht mehr richtig ist. In diesem Fall wird die Nicht-Übereinstimmung von Soll- und Ist-Klasse als Klassifikationsfehler gewertet, obwohl das Klassifikationsergebnis der aktuellen Landbedeckung entspricht.

Bei der Schätzung der Klassifikationsgenauigkeit mit Hilfe von Kontrollpunkten kommt es zu diesen Problemen nicht, da die Soll-Klasse anhand des aktuellen Satellitenbildes ermittelt wurde. Für die Beurteilung der Klassifikationen mit Referenzdatensätzen aus SINUS werden die Punkte, an denen die Landbedeckung durch visuelle Interpretation ermittelt wurde (Kap. 6.2; Tab. 7.1), als Kontrollpunkte verwendet. Dazu wurden Klassen wie folgt zusammengefasst: die Klassen 60, 61, 62 und 63 werden zu Klasse 65 („Acker“) zusammengefasst, die Klasse 44 wird in die Klasse 41 integriert und die Klasse 45 in die Klasse 42.

Klasse	Beschreibung der Klasse	Referenzpunkte
11	Vegetationsfreie Flächen	22
20	Wasserflächen	9
31	Nadelwald	45
32	Mischwald	66
33	Laubwald	68
41	Grünland mit viel Biomasse (BM)	23
42	Grünland mit wenig BM	26
44	unbestimmbare landwirt.Flächen mit viel BM	28
45	unbestimmbare landwirt.Flächen mit wenig BM	20
49	Grünland mit Bäumen, Sträucher etc.	42
50	Weingärten	35
60	offener Boden (Acker)	50
61	reife Ackervegetation	102
62	grüne Ackervegetation	72
63	unspezifizierbare Ackervegetation	66
71	gering versiegelt	43
72	mittel versiegelt	12
73	stark versiegelt	11
Summe		740

Tabelle 7.1: Anzahl an visuell interpretierter Referenzpunkte pro Klasse

7.2.2 Rasterweiten und Punkteanzahl

In der Tabelle 7.2 sind für jede SINUS-Klasse die Anzahl der Referenzpunkte und deren Anteil an der Gesamtpunkteanzahl bei den verschiedenen Rasterweiten (2,5km, 1km,

Klasse	Fläche in % lt.SINUS	Raster 2,5km		Raster 1km			Raster 0,5km	
		Anz.Pkt.	in %	Anz.Pkt.	in %	reduziert	Anz.Pkt.	in %
11	0.16	2	0.07	28	0.15	28	114	0.16
20	0.51	13	0.45	85	0.46	85	360	0.52
31	14.34	401	13.75	2612	14.25	312	9745	14.14
32	9.19	265	9.09	1762	9.62	312	6588	9.56
33	17.11	519	17.80	3217	17.55	317	12219	17.74
41	1.84	67	2.30	320	1.75	320	1226	1.78
42	2.78	97	3.33	528	2.88	328	2020	2.93
49	12.45	368	12.62	2324	12.68	324	8632	12.53
50	2.07	68	2.33	436	2.38	336	1510	2.19
65	33.27	1015	34.82	6361	34.70	1111	23945	34.75
71	2.56	79	2.70	468	2.55	268	1878	2.72
72	0.41	7	0.24	92	0.50	92	314	0.45
73	0.42	14	0.48	91	0.50	91	344	0.49
Pkt.gesamt		2915		18324		3924	68895	

Tabelle 7.2: SINUS-Fläche in Prozent und Punkteanzahl pro Klasse bei unterschiedlichen Rasterweiten

0,5km) aufgelistet.

Die Prozentanteile der Punkte pro Klasse stimmen bei allen 3 Rasterweiten mit den Flächenanteilen laut SINUS recht gut überein. Beim 2,5km Raster sind einzelne Klassen (11, 72) nur mit sehr geringen Punktzahlen vertreten, wodurch es bei der kNN-Klassifikation zu schlechten Ergebnissen für diese Klassen kommen würde. Beim 1km Raster sind kleine Klassen mit einer ausreichenden Zahl an Referenzpunkten vertreten, bei großen Klassen gibt es allerdings einen Überschuss an Referenzpunkten (z.B.: Klasse 33: 3217 Punkte). Diese überschüssigen Punkte führen zu langen Rechenzeiten bei der kNN-Klassifikation, ohne aber einen zusätzlichen Nutzen für die spektrale Beschreibung der Klassen zu haben. Ähnliches gilt in verstärktem Maße für den 0,5km Raster. Der 1km Raster scheint demzufolge der am besten geeignete zu sein. Fehlerabschätzungen werden daher für die mit Hilfe des 1km Rasters erstellten Referenzdatensätze gezeigt. Außerdem werden Ergebnisse gezeigt, die mit Hilfe eines reduzierten Referenzdatensatzes erzeugt wurden (siehe Tabelle 6.1, Raster 1km reduziert). Die Auswahl überschüssiger Punkte erfolgte zufällig.

Wie in Tabelle 7.1 ersichtlich, ist im Referenzdatensatz, der durch visuelle Interpretation erzeugt wurde, der Unterschied der Referenzpunkte pro Klasse kleiner als bei den Referenzdatensätzen aus SINUS. Wie in Kap. 6.1 beschrieben, wurden Referenzpunkte zugunsten kleiner Klassen verschoben, wenn sie im regelmäßigen Raster auf Mischpixel lagen. Kleine Klassen sind daher gemessen an ihrem tatsächlichen Flächenanteil im Referenzdatensatz überrepräsentiert.

Sowohl beim reduzierten Referenzdatensatz als auch beim Referenzdatensatz, der durch visuelle Interpretation erzeugt wurde, stellt sich die Frage, ob es sich negativ auf das Klassifikationsergebnis auswirkt, wenn die Klassen abweichend von ihren tatsächlichen Häufigkeiten vertreten sind (vgl. Davies, 1988).

In Tabelle 7.3 sind die Kappa-Werte für verschiedene Referenzdatensätze zusammenge-

Referenz		SINUS, 1km Raster		Visuelle Interpretation
		original (ca. 18.000 Punkte)	reduziert (ca. 3.900 Punkte)	
				740 Punkte
VOR Outlier- Eliminierung	k=1	0.394	0.376	0.581
	k=5	0.480	0.483	0.608
NACH Outlier- Eliminierung	k=1	0.428	0.399	–
	k=5	0.480	0.499	–

Standardabweichung von Kappa = +/- 0.02

Tabelle 7.3: Kappa-Werte für verschiedene Referenzdatensätze, berechnet anhand der Kontrollpunkte

stellt und der Klassifikationsgenauigkeit gegenübergestellt, die erreicht wird, wenn der durch visuelle Interpretation erstellte Referenzdatensatz verwendet wird.

7.2.3 Eliminierung von Ausreißern

In den Abbildungen 7.8 und 7.9 ist dargestellt, wie sich Kappa nach jeder Runde der Outlier-Suche und -Eliminierung verändert. Die Kurve steigt nach den ersten Runden stärker bis sie schließlich nach 5 Runden abflacht. Die Kappa-Werte wurden durch Kreuzvalidierung bestimmt.

Tabelle 7.3 kann entnommen werden, welchen Einfluss die Eliminierung von Ausreißern auf die Klassifikationsgenauigkeit hat. Die angegebenen Kappa-Werte wurden mit Hilfe der Kontrollpunkte berechnet. Im Vergleich zu den Abbildungen 7.8 und 7.9 sind die mit Hilfe der Kontrollpunkte berechneten Kappa-Werte niedriger. Die Ursache dürfte darin liegen, dass durch die Eliminierung der Ausreißer die Variationsbreite im Referenzdatensatz eingeschränkt und dadurch die Repräsentativität des Referenzdatensatzes herabgesetzt wird. Es werden teilweise also Referenzpunkte als Ausreißer deklariert, bei denen es sich zwar um eine seltene aber wichtige Ausprägung einer Klasse handelt.

Zusammenfassend kann festgestellt werden, dass die Eliminierung von Outliern bei $k=1$ eine deutlichere Verbesserung der Klassifikationsgenauigkeit bewirkt als bei größeren k -Werten. Dies entspricht den Erwartungen, da bei einem größeren k die Anfälligkeit gegenüber Ausreißern prinzipiell geringer ist.

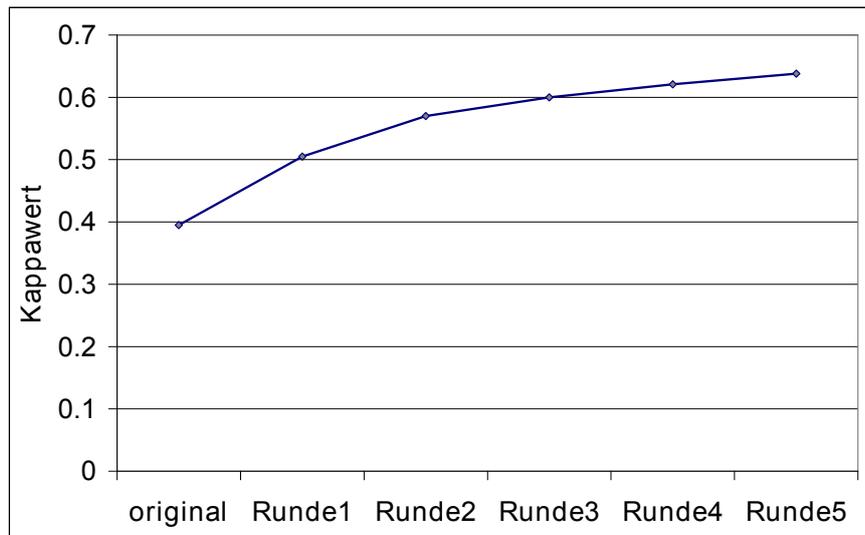


Abbildung 7.8: Kappa-Werte nach der Outlier-Suche, original Datensatz (ca. 18 000 Punkte) für $k=1$

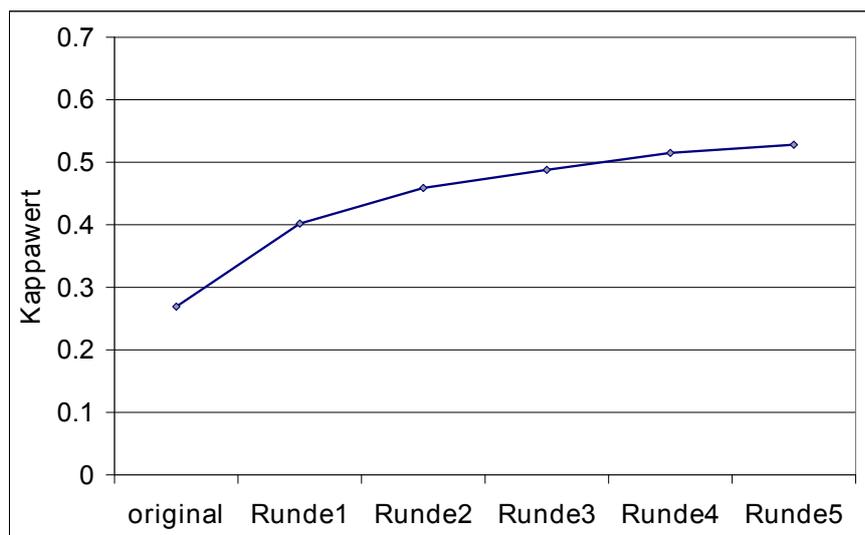


Abbildung 7.9: Kappa-Werte nach der Outlier-Suche, reduzierter Datensatz (ca. 3 900 Punkte) für $k=1$

8 Diskussion und Schlussfolgerungen

In der Arbeit hat sich insgesamt bestätigt, dass die k-Nearest-Neighbour-Klassifizierungsmethode ein zuverlässiges und relativ einfaches Verfahren ist, das sich für den Einsatz bei der Landbedeckungsklassifizierung eignet. Es wurden verschiedene Aspekte der kNN-Methode untersucht. Die Ergebnisse können wie folgt bewertet werden, und es können die folgenden Schlüsse gezogen werden:

Optimierung der kNN-Parameter

Die Gewichtung der Spektralkanäle hat keinen großen Einfluss auf das Klassifizierungsergebnis. Die Erklärung dafür dürfte sein, dass viele Klassen große spektrale Ähnlichkeiten aufweisen, im spektralen Merkmalsraum nicht gut getrennt sind und daher auch mit beliebiger Gewichtung der Spektralkanäle nicht besser trennbar werden. In dieser Hinsicht stellen insbesondere die „vegetationsfreien“ Klassen ein Problem dar, wie zum Beispiel die Klasse „mittlerer Versiegelungsgrad“ und offener Boden in der „Ackerklasse“ oder die Klasse „hoher Versiegelungsgrad“ und Schotterflächen beziehungsweise Felsen in der Klasse „Vegetationsfrei“.

Verbesserungen der Trennbarkeit mancher Klassen durch Kanalgewichtungen ziehen fast immer Verschlechterungen für andere Klassen nach sich.

Ein höheres k bringt eine Genauigkeitsverbesserung mit sich. Voraussetzung ist eine ausreichende Anzahl an Referenzpunkten, sonst werden die schwach vertretenen Klassen unterdrückt. Mit einem größeren k werden auch die Ausreißer in den Referenzdaten unterdrückt, auch wenn sie nicht im Zuge einer speziellen „Outlier“-Suche aus dem Datensatz entfernt werden.

Probleme mit den Referenzdaten und Möglichkeiten zur Verbesserung der Referenzdaten

Wegen der Pixelgröße von 30m ist das Mischpixelproblem groß.

Bei Referenzpunkten, die in der Nähe von Landbedeckungsgrenzen liegen, kann es - verschärft durch Fehler in der Georeferenzierung - zu Mischpixeln kommen. Eine Verschiebung dieser Referenzpunkte in eine homogene Fläche kann eine Verbesserung bringen. Verschiebungen können sich aber nachteilig auf kleinflächige Landbedeckungskategorien

auswirken, die dadurch einen Verlust wichtiger Referenzpunkte erfahren können. Besonders tritt dieses Problem bei der Klasse „Versiegelte Fläche“ (im speziellen Straßen) auf.

Weiters gibt es Klassen, die eine Mischung aus mehreren Landbedeckungskategorien darstellen (wie „geringer Versiegelungsgrad“, typisch für eine Einfamilienhaussiedlung, bestehend aus versiegelter Fläche, Vegetation und anderen Oberflächen. Hier sind Mischpixel für die Klasse charakteristisch. Das Mischungsverhältnis variiert aber, so dass keine zufriedenstellende Klassifikation möglich ist. Es handelt sich hier um ein bei der kNN-Klassifikation ungelöstes Problem.

Ein weiteres Problem sind Outlier in den Referenzdaten. Die Outlier-Suche funktioniert grundsätzlich gut. Man kann eine Verbesserung der Gesamtgenauigkeit (ersichtlich aus der Fehlermatrix) erreichen und diese Verbesserung auch visuell in der Ergebniskarte erkennen. Bei der Outlier-Suche ist darauf zu achten, dass man eine ausreichende Anzahl an Referenzpunkten hat. Dies kann mit kleinen Rasterweiten für die Referenzpunkte erreicht werden. Eine Verbesserungsmöglichkeit könnte sich durch eine Aufteilung des Referenzdatensatzes nach den Klassen ergeben: Wenn die Outlier-Suche nur innerhalb einer bestimmten Gruppe von Klassen durchgeführt wird, die untereinander gut trennbar sind, könnte die Outlier-Suche erfolgreicher werden. Es war in der vorliegenden Arbeit nicht möglich, diese Idee weiter zu verfolgen.

Die Outlier-Suche kann auch zu einem Verlust wichtiger Referenzpunkte führen. So gibt es Klassen, die über große Bereiche des Merkmalsraums mehrgipfelig (mit unterschiedlichen Anteilen) verteilt sind. Dann kann es vorkommen, dass auch zur Klasse gehörige Referenzpunkte als Outlier angesehen und eliminiert werden. Für dieses Problem gibt es noch keine Lösung.

Die Outlier-Suche scheint bei Referenzdatensätzen mit ungefähr gleicher Besetzung der Klassen besser zu funktionieren. Eine „Ausdünnung“ von zu stark mit Referenzdaten besetzten Klassen dürfte insbesondere auch bei Klassifikation mit großem k positive Effekte haben, da ansonsten bei großem k meist die stark besetzten Klassen dominieren.

Visualisierung von Veränderungsflächen und Darstellung der Klassifikationsergebnisse

Die Outlier-Suche funktioniert für die Darstellung von Veränderungen (Aktualisierung) gut. Bei der Aktualisierung von SINUS-Datensätzen, vor allem aber von CORINE-Datensätzen sind insbesondere die Auswirkungen der unterschiedlichen räumlichen Auflösung und Generalisierung einerseits der veralteten Datensätze und andererseits der mit LANDSAT pixelweise aktualisierten Datensätze von den tatsächlichen Veränderungen der Landbedeckung nicht klar zu trennen. Flächenbilanzen von Veränderungen können daher nicht erstellt werden.

Es ist ein Vorteil der räumlich höher auflösenden automatisierten Klassifikation im Vergleich zur generalisierten Kartierung in CORINE, dass die Landschaftselemente in ihrer Größe, Form und Anordnung anschaulich gemacht werden und somit eine bessere landschaftsökologische Analyse ermöglichen.

Bei der Klassifizierung mit Referenzpixeln aus CORINE wurde der Raum von Wien nicht sehr gut klassifiziert. Da die Referenzpunkte nur aus dem Raum Niederösterreich ausgewählt wurden, kommt die Klasse „durchgängig städtische Prägung“ (Stadtkerngebiet) nicht vor. So ist das Stadtbild eher monoton, mit nur einer Kategorie „nicht durchgängig städtische Prägung“ dargestellt.

Im Gegensatz dazu liefert die Klassifikation mit Referenzpunkten aus SINUS sehr gute Ergebnisse für den Raum Wien (gute Referenzen in Niederösterreich).

Klassifikationsmethodik

Die kNN-Methode hat sich allgemein gut bewährt. Da keine parametrische Verteilung der Stichproben im Merkmalsraum vorausgesetzt wird, können vielfältige Erscheinungen in der Landschaft gut kategorisiert werden.

Anzustreben wäre insbesondere noch eine bessere Auswahl der Referenzpunkte. Wenn diese nicht über das ganze Gebiet verteilt wären, sondern nur über Teilbereiche (Gebiete, die eine große Anzahl an Landbedeckungskategorien aufweisen), so könnte erreicht werden, dass die Anzahl der Punkte gleichmäßiger auf die Klassen verteilt sind. Dies könnte anfangs zu einem Zeitverlust führen, jedoch helfen, Probleme in den nachfolgenden Schritten zu vermeiden.

Kombinationen der kNN-Klassifizierung mit Segmentierungsmethoden, regel- und wissensbasierter Klassifizierung sowie radiometrisch-topographischer Korrektur könnten verbesserte Ergebnisse liefern.

Eine automatische Auswertung bringt eine sehr große Zeiteinsparung, sie bleibt aber noch immer fehlerbehaftet. Der Mensch, mit seiner Fähigkeit, eine Fülle von visuellen Informationen und Kontextinformationen heranzuziehen und auszuwerten, bleibt dem Computer überlegen.

Tabellenverzeichnis

3.1	Flächenwidmung in Niederösterreich	13
4.1	Empfindlichkeitsbereich der Spektralkanäle von LANDSAT 5	16
4.2	SINUS Landbedeckungsklassen	19
4.3	CORINE-Landcover	22
4.4	Klassen der visuellen Interpretation	23
6.1	Anzahl der Rasterpunkte bei unterschiedlichen Rasterweiten	32
6.2	Beispiel einer Fehlermatrix	36
6.3	Auffinden absichtlich in den Datensatz eingefügter Ausreißer (Test) . .	43
7.1	Anzahl an visuell interpretierter Referenzpunkte pro Klasse	55
7.2	Punkteanzahl bei unterschiedlichen Rasterweiten	55
7.3	Kappa-Werte für verschiedene Referenzdatensätze	56

Abbildungsverzeichnis

2.1	kNN-Methode	10
4.1	Pixelauflösung von LANDSAT 5	15
4.2	SINUS und CORINE: Räumliche Auflösung	21
5.1	LANDSAT-Szenen nach dem Georeferenzieren, Passpunktverteilung . .	27
5.2	Radiometrische Korrektur	30
6.1	Fehlerhafte Punkte auf Grund der Wolken	35
6.2	Kappa-Wert in Abhängigkeit von k	39
6.3	Verschieben von Referenzpunkten	40
6.4	„Outlier“-Suche	42
7.1	CORINE-Datensatz, Ausschnitt: südlicher Wienerwald	46
7.2	Klassifikation mit Referenzpunkten aus CORINE	47
7.3	LANDSAT 5 TM, Ausschnitt: südlicher Wienerwald	48
7.4	SINUS-Landbedeckungskarte, Ausschnitt: Krems und Umgebung . . .	50
7.5	Klassifikation mit Referenzpunkten aus SINUS	51
7.6	LANDSAT 5 TM, Ausschnitt: Krems und Umgebung	52
7.7	Vergleich Referenzdatensatz original / Ausreißer-bereinigt	53
7.8	Kappa-Werte nach der Outlier-Suche (original)	58
7.9	Kappa-Werte nach der Outlier-Suche (reduziert)	58

Literaturverzeichnis

- [Albertz 2007] ALBERTZ, J.: *Einführung in die Fernerkundung, 3.Auflage*. Bd. 3. Auflage. WBG, Darmstadt, 2007
- [Bischof u. a. 1992] BISCHOF, H. ; SCHNEIDER, W. ; PINZ, A.: Multi-Spectral Classification of Landsat Images using Neural Networks. In: *IEEE Transactions on Geoscience and Remote Sensing, Volume 30, Issue 3* (1992), S. 482–490
- [Brendler 2004] BRENDLER, S.: Klassifikation der Namen. In: BRENDLER, A. (Hrsg.): *Namenarten und ihre Erforschung*, Baar, 2004, S. 70, 71. – Url: <http://www.isis.unibas.ch/download.php?id=13046>, am 30.10.08
- [Congalton und Green 1999] CONGALTON, R. ; GREEN, K.: *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Lewis Publishers, 1999
- [Davies 1988] DAVIES, E. R.: Training sets and a priori probabilities with the nearest neighbour method of pattern recognition. In: *Pattern Recognition Letters, Volume 8* (1988), S. 11–13
- [Dorninger 2003] DORNINGER, G.: Datensätze zur Naturraumkartierung in Oberösterreich. In: GRILLMAYER, R. (Hrsg.) ; SCHNEIDER, W. (Hrsg.): *Geodaten zur Landbedeckung in Österreich*, Shaker Verlag, 2003, S. 67–73
- [Geoinformatik am 09.10.08] GEOINFORMATIK, Lexikon: *Merkmalsraum*. am 09.10.08. – url: <http://www.geoinformatik.uni-rostock.de/einzel.asp?ID=-685625833>
- [Glossardatenbank am 09.10.08] GLOSSARDATENBANK: *Merkmalsraum*. am 09.10.08. – url: <http://mars.geographie.uni-halle.de/mlucampus/geoglossar/>
- [Grillmayer und Schneider 2003] GRILLMAYER, R. ; SCHNEIDER, W.: *Geodaten zur Landbedeckung in Österreich*. Shaker Verlag, 2003
- [Gu und Gillespie 1997] GU, D. ; GILLESPIE, A.: Topographic Normalization of LANDSAT TM, Images of Forest based on Subpixel Sun-Canopy-Sensor Geometry. In: *Remote Sensing of Environment, Volume 64* (1997), S. 166–175
- [Hölzl 2003] HÖLZL, M.: CORINE Landcover, Erfahrungen bei der Aktualisierung und Abgabe von Landbedeckungsdaten. In: GRILLMAYER, R. (Hrsg.) ; SCHNEIDER, W. (Hrsg.): *Geodaten zur Landbedeckung in Österreich*, Shaker Verlag, 2003, S. 21–31

- [Hollaus und Suppan 2003] HOLLAUS, M. ; SUPPAN, F.: Landbedeckungsdatensatz aus der Kulturlandschaftsforschung (SINUS). In: GRILLMAYER, R. (Hrsg.) ; SCHNEIDER, W. (Hrsg.): *Geodaten zur Landbedeckung in Österreich*, Shaker Verlag, 2003, S. 33–43
- [Koukal 2008] KOUKAL, T.: *User Guide to the kNN-Program*. Institut für Vermessung, Fernerkundung und Landinformation, Universität für Bodenkultur Wien: , February 2008
- [Koukal u. a. 2005] KOUKAL, T. ; SCHNEIDER, W. ; SUPPAN, F.: Radiometric-topographic Normalization in Mountainous Terrain for LANDSAT-TM-based Forest Parameter Assessment by the kNN Method. In: OLUIĆ, Marinko (Hrsg.): *New Strategies for European Remote Sensing*, Proceedings of the 24th Symposium of the European Association of Remote Sensing Laboratories, Dubrovnik, 25-27 May 2004, Millpress, 2005, S. 239–246
- [Österreich Lexikon AEIOU am 24.07.08] LEXIKON AEIOU Österreich: *Niederösterreich*. am 24.07.08. – url: <http://aeiou.iicm.tugraz.at/aeiou.encyclp.n/n670068.htm>
- [Mitchell 1997] MITCHELL, T.: *Machine Learning*. McGraw Hill, 1997
- [Muhar 2003] MUHAR, A.: Bedarf an Landbedeckungskarten in der Landschaftsplanung. In: GRILLMAYER, R. (Hrsg.) ; SCHNEIDER, W. (Hrsg.): *Geodaten zur Landbedeckung in Österreich*, Shaker Verlag, 2003, S. 97–103
- [Radberger 2001] RADBERGER, R.: *Monitoring der Verstädterung im Großraum Istanbul mit den Methoden der Fernerkundung und der Versuch einer räumlich-statistischen Modellierung*. Universität Göttingen, Dissertation. 2001
- [Seger 2003] SEGER, M.: Landinformationssystem Österreich. In: GRILLMAYER, R. (Hrsg.) ; SCHNEIDER, W. (Hrsg.): *Geodaten zur Landbedeckung in Österreich*, Shaker Verlag, 2003, S. 45–55
- [Stech am 20.09.08] STECH, C.: *Einführung in die Fernerkundung*. Erstellt im Rahmen des Projektes ENGL/EMIR. am 20.09.08. – URL http://www.uni-kiel.de/forum-erdkunde/unterric/material/einf_fe/kapitel_5.htm
- [Wikipedia am 12.08.08] WIKIPEDIA: *Klassifizierung*. am 12.08.08. – url: <http://de.wikipedia.org/wiki/Klassifizierung>