

# Hochschule Neubrandenburg Studiengang Geoinformatik

# Fernerkundliche Detektion von Heilpflanzen in Deutschland und Burkina Faso

Masterarbeit vorgelegt von: Thomas Marzinke

zum Erlangen des akademischen Grades Master of Engineering (M.Eng)

Erstprüfer: Prof. Dr. Ralf Löwner	
Zweitprüfer: Prof. Dr. Erik Borg	
eingereicht am:	

urn:nbn:de:gbv:519-thesis2020-0115-2

# Inhaltsverzeichnis

1	Motivation	5
2	Ziele	5
3	Untersuchungsgebiete 3.1 Untersuchungsgebiet bei Kraase (Deutschland)	6 6 7
4	Problemstellung 4.1 Fernerkundung von Pflanzen	9 10 11
5	Methodik  5.1 Datengrundlage 5.2 Preprocessing 5.3 Merkmalsraumerweiterung 5.4 Merkmalsreduktion 5.4.1 Berechnung der Korrelationskoeffizienten bei großem Ausgangsdatenvolumen 5.5 Jeffries-Matusita Abstand 5.6 Clusteralgorithmen 5.6.1 k-Means Clustering 5.6.2 EM Clustering 5.7 Klassifikationsalgorithmen 5.7.1 Euklidische (Spektrale) Distanz Klassifikation 5.7.2 Maximum Likelihood Klassifikation 5.7.3 Mahalanobis Abstand Klassifikation 5.7.4 Korrelationskoeffizienten-basierte Klassifikation (Spectral Correlation Mapper - SCM) 5.7.5 Klassifikation mithilfe der kanonischen Korrelationsanalyse 5.7.6 Random Forest Klassifikation	13 13 15 16 22 24 25 25 27 30 31 31 33 36 38
6	6.1 Auswahl der Trainingsgebiete	39 42 44 51
7	Fernerkundung von Röhricht	53
8	Fernerkundung von Weiden	60
9	Fernerkundung von Cassia sieberiana	67
10	Fernerkundung von Vitellaria paradoxa	69
11	Zusammenfassung	<b>7</b> 6
12	Ausblick	<b>7</b> 6

Glossar																		
Quellenverzeichnis																		
Datenquellen																		
Abbildungsverzeichn	is .	 								·								
Formelverzeichnis .		 																
Tabellenverzeichnis																		

#### Abstract

Supporting the project DigiMeP at the University of Applied Sciences Neubrandenburg, methods of remote sensing of medicinal plants are investigated and provided on basis of Sentinel data (as technological spear-head among the remote sensing missions freely available to the public), which are freely available within the Copernicus programme of the European Commission. For this purpose, spectrum-based, probabilistic and non probabilistic, as well as machine learning methodologies are applied to classification problems. Due to the seasonal variability in the characteristics of vegetation, a special focus is set on multitemporal methods. The selected instruments are firstly tested in the vicinity of the village of Kraase in Mecklenburg - Western Pomerania as a known and easily accessible area for control purposes and finally transferred to problems in Burkina Faso.

# Kurzfassung

Zur Unterstützung des Projektvorhabens DigiMeP an der Hochschule Neubrandenburg werden Methoden der Fernerkundung von Heilpflanzen auf der Grundlage der im Rahmen des Copernicus Programms der Europäischen Kommission frei verfügbaren Daten aus den Sentinel-Missionen (als technologische Speerspitze unter den der Öffentlichkeit frei verfügbaren Fernerkundungsmissionen) untersucht und bereitgestellt. Zu diesem Zweck werden sowohl Spektrums-basierte, probabilistische, als auch nicht probabilistische sowie Methodiken des maschinellen Lernens auf Klassifikationsprobleme angewandt. Wegen der jahreszeitlichen Variabilität in den Eigenschaften der Vegetation liegt ein besonderer Fokus auf multitemporale Methoden. Das ausgewählte Instrumentarium wird zunächst in der näheren Umgebung des Dorfes Kraase in Mecklenburg-Vorpommern als bekanntes und zur Kontrolle leicht begehbares Gebiet erprobt und schließlich auf Problemstellungen in Burkina Faso übertragen.

# Danksagung

In erster Linie möchte ich meinen Betreuern Herrn Prof. Dr. Ralf Löwner und Herrn Prof. Dr. Erik Borg für Ihre Unterstützung danken. Großer Dank gebührt meiner Familie: meiner Mutter Gundula, meinem kleinen Bruder Christian und meinem Vater Klaus, die mir sehr halfen die schwere Last der großen Anstrengungen mit zu tragen, mich stets unterstützten und mir in vielen Situationen des Alltags helfend beistanden. Ich danke allen Dozentinnen und Dozenten unseres Fachbereiches an der Hochschule Neubrandenburg, die mir das nötige Wissen vermittelten, aus dem es mir möglich war diese Arbeit zu vollenden und stets ansprechende, mitreißende, interessante, aber zugleich manchmal auch humorvolle Vorlesungen hielten. Dabei bin ich insbesondere unserem Rektor Herrn Prof. Dr. Gerd Teschke, unserem Dekan Herrn Prof. Dr. Sven Brämer, seinem Vorgänger Herrn Prof. Dr. Wolfgang Kresse und unserem Studienleiter Herrn Prof. Dr. Andreas Wehrenpfennig für die gute Leitung und Gestaltung unseres Studiengangs dankbar. Ich möchte auch unserer Studiensekretärin Frau Ariane Haß, ihrer Vorgängerin Frau Petra Schuldt und unserer Sekretärin Frau Anja Eutin sehr für ihre Hilfe bei der Organisation des Studiums und der diese Arbeit betreffenden Angelegenheiten danken. Außerdem bin ich auch meinen ehemaligen Lehrerinnen und Lehrern aus der Schulzeit am Richard-Wossidlo-Gymnasium in Waren sehr dankbar. Im Speziellen danke ich Herrn Andreas Liedtke, der meine Aufmerksamkeit als erster auf das Thema der Fernerkundung von Pflanzenspezies lenkte und mein Interesse für die Physik immer wieder aufs neue weckte, den Herren Bernd Dethloff, Christian Schneeweiß und Friedwald Wagenknecht, die mich für die Mathematik und Informatik faszinierten sowie unserem Direktor Herrn Manfred Glas. Nicht zuletzt möchte ich mich bei unserer Hochschuldruckerei für den Druck dieser Arbeit bedanken.

#### 1 Motivation

Für den Großteil ihrer Geschichte standen der Menschheit ausschließlich Medikamente aus natürlichen Quellen zur Verfügung und die medizinische Praxis war eng mit religiösen Vorstellungen und Traditionen verknüpft. Der Aufstieg synthetischer Medikamente begann vor etwa 200 Jahren mit den ersten Anwendungen von Chloroform und Äther zur Anästhesie. Dank bedeutender Fortschritte in Wissenschaft und Technik Anfang des 20. Jahrhunderts nimmt die Anzahl synthetische Präparate seitdem stetig zu und industrielle Herstellungsverfahren sichern ihre allgemeine Verfügbarkeit (Sutcliffe, 1992). Viele Medikamente basieren dabei auf pflanzlichen Grundlagen und den Erkenntnissen aus der jahrtausendealten traditionellen Medizin (Farag, 2015; Yuan, 2016).

In Burkina Faso leben viele unterschiedlicher Volksgruppen nebeneinander (Trautmann, 2017; Gre-BE, 2017). Etwa 62% der Bevölkerung sind Muslime, 23% katholische und 7% evangelische Christen. Die Anhänger traditioneller afrikanischer Religionen machen rund 7% aus (CIA, 2018a). Rund 80% der Bevölkerung macht Gebrauch von traditionellen Heilmethoden (WHO, 2001). Westafrika ist das Ursprungsland der Voodoo-Religion (Maranise, 2012). Im Nachbarland Benin machen die Voodoosi einen Bevölkerungsanteil von etwa 12% aus (CIA, 2018b). Die traditionelle afrikanische Heilkunde ist eng mit der Spiritualität der Heiler verknüpft (Mokgobi, 2014). Das Konzept von Krankheit in der traditionellen afrikanischen Medizin unterscheidet sich von dem der westlichen Medizin. Das Auftreten von Krankheiten wird in der traditionellen afrikanischen Medizin mit natürlichen, kulturellen und sozialen Ursachen in Verbindung gebracht. Dabei können kulturelle und soziale Ursachen auch übernatürlichen Ursprungs sein und beispielsweise Verzauberungen oder böse Geister umfassen. Die traditionelle afrikanische Medizin soll vom westlichen Standpunkt aus nicht als "alternative" oder "komplementäre" Medizin betrachtet werden. Es handelt sich vielmehr um ein über Jahrtausende etabliertes Gesundheitssystem, welches auch Lösungsvorschläge zur Heilung von Krankheiten bietet, mit denen die westliche Medizin bislang Schwierigkeiten hat (OZIOMA, 2019). Durch die strikte Ablehnung moderner westlicher Medizin und Maßnahmen zur Gesundheitsvorsorge können die Vorstellungen der traditionellen Religion und Medizin deren Akzeptanz in der Bevölkerung verringern und so einen negativen Einfluss auf die Gesundheit der Allgemeinheit haben (STOOP, 2019). Daher spielen die traditionellen Vorstellungen eine wichtige Rolle beim Aufbau eines nachhaltigen und integrativen Gesundheitssystems; einem Prozess von dem langfristig sowohl die traditionelle afrikanische, als auch die westliche Medizin profitieren können. Insbesondere die Vielzahl der in der traditionellen afrikanischen Medizin eingesetzten Heilpflanzen kommt als vielversprechende Grundlage für die Entwicklung neuartiger Medikamente infrage (Kahumba, 2015).

Die traditionelle Medizin erfährt in Burkina Faso staatliche Förderung (WHO, 2001; Kasilo, 2010) und wird von der Regierung in den Aufbau einer nachhaltigen Gesundheitsinfrastruktur integriert (Busia, 2010). Einige Verhaltenseisen der traditionellen Heiler stellen jedoch auch eine Bedrohung für die Artenvielfalt dar und werden in Teilen des Landes für das Verschwinden einzelner Spezies verantwortlich gemacht (Thiombiano, 2010a).

#### 2 Ziele

Zur Unterstützung der traditionellen Heiler vor Ort, der Förderung eins nachhaltigen Ressourcenmanagements und zum Erhalt der Artenvielfalt in den Zielländern sollen Verfahren zur Fernerkundung (wildwachsender) Pflanzen auf Grundlage der im Rahmen des Copernicus-Programms (Berger, 2009; ESA, 2018) der Europäischen Kommission der Allgemeinheit frei zur Verfügung gestellten Daten (Open Data) der Sentinel 2 - Mission (ESA, 2015; GMES PROGRAMME, 2012) untersucht, beurteilt und bereitgestellt werden. Dabei liegt der Fokus auf der Identifikation einzelner Spezies bzw. von Pflanzengemeinschaften. Um die Unabhängigkeit der Untersuchungen zu gewährleisten und die finanziellen Aufwendungen zu minimieren, werden ausschließlich der Allgemeinheit frei zugängliche Daten (Open Data) und freie und quelloffene Software (FOSS - free and open source software) genutzt. Auch die Resultate der Untersuchungen sollen jedem Interessierten und für weitere Nachforschungen frei zur Verfügung stehen.

Die Untersuchungen werden auch im Hinblick auf den zunehmenden Verlust traditionellen Heilwissens

durchgeführt (RAMIREZ, 2007; CÁMARA-LERET, 2019) und sollen durch die Unterstützung der traditionellen Heiler einen Beitrag zur Konservierung dieses wertvollen Kulturgutes leisten. Im Hinblick auf die schwache Wirtschaftslage Burkina Fasos (WORLD BANK GROUP, 2018) wird darüber hinaus angestrebt, mit der Arbeit einen Beitrag zur nachhaltigen Entwicklungshilfe im Einklang mit den Zielen der Agenda 2030 für nachhaltige Entwicklung der Vereinten Nationen zu leisten (HOLLOWAY, 2018; GENERALVERSAMMLUNG DER VEREINTEN NATIONEN, 2015).

# 3 Untersuchungsgebiete

Die Untersuchungsgebiete liegen in Burkina Faso in der Umgebung der Stadt Bobo-Dioulasso im Gouvernement Hauts-bassins und in der näheren Umgebung des Dorfes Kraase in Mecklenburg-Vorpommern (Deutschland). Dabei dient das Untersuchungsgebiet bei Kraase der Erprobung und Verifizierung der vorgestellten Methodiken in einem bekannten und leicht erreichbaren Gebiet sowie zur genaueren Beurteilung ihrer Effektivität und Anwendbarkeit auf die Problemstellung.

#### 3.1 Untersuchungsgebiet bei Kraase (Deutschland)

Das Untersuchungsgebiet bei Kraase in Mecklenburg-Vorpommern (Landkreis Mecklenburgische Seenplatte) erstreckt sich über ein rechteckiges Gebiet innerhalb der Zone 33 des UTM-Koordinatensystems, welches durch die Koordinaten 350540m und 368630m (Rechtswert) sowie 5925650m und 5942300m (Hochwert) begrenzt wird (etwa zwischen 12,74° östlicher Länge und 53,45° nördlicher Breite sowie 13,02° östlicher Länge und 53,62° nördlicher Breite; siehe Abbildung 1). Die räumliche Abgrenzung des Untersuchungsgebietes ist dabei bis auf die Berücksichtigung der Anforderungen an leichte Erreichbarkeit und guter bereits vorhandener Ortskenntnis willkürlich.

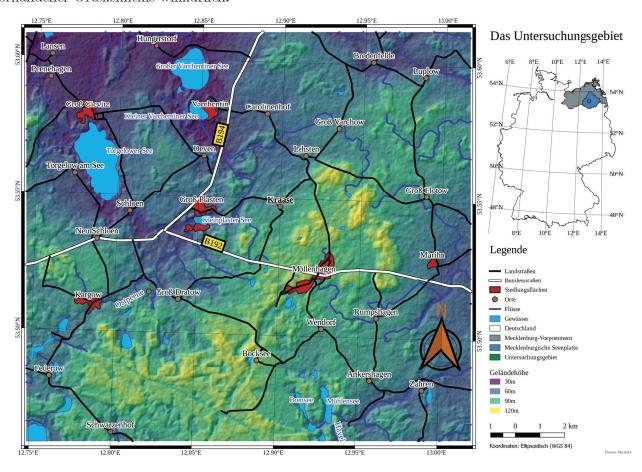


Abbildung 1: Übersichtskarte für das Untersuchungsgebiet bei Kraase

Die größte Siedlung innerhalb des Untersuchungsgebietes ist mit rund 1500 Einwohnern Möllenhagen mit den Ortsteilen Bauernberg, Freidorf, Groß Varchow, Hoppenbarg, Kraase, Lehsten, Rethwisch, Rockow und Wendorf im Zentrum des Gebiets (Statistisches Amt Mecklenburg-Vorpommern, 2019). Zu den größten Seen innerhalb des Untersuchungsgebietes zählen der Torgelower See im Nordwesten sowie der große und kleine Varchentiner See im Norden. Im Südosten entspringt der Fluss Havel an den Quellseen Bornsee und Mühlensee. Als lokal bedeutendster Fluss entspringt die Ostpeene im Westen des Untersuchungsgebietes unweit des Dorfes Kargow, von wo aus er zunächst in östlicher Richtung und weiter flussabwärts Richtung Norden durch den Torgelower See verläuft. Es handelt sich um ein Flachland mit dem Tannenberg (125,2m über NN) im Südosten von Kraase als höchster Erhebungen (Amt für Geo-Information, Vermessung und Katasterwesen, 2016).

Die Landschaft wurde maßgeblich durch vergangene Eiszeiten geprägt (Moränenlandschaft) und ist von zahlreichen Toteis- und Rillenseen übersät (FRAEDRICH, 2016). Das Untersuchungsgebiet wird großlandschaftlich vom Rückland der Mecklenburgischen Seenplatte mit dem oberen Peenegebiet im Nordwesten und dem oberen Tollensegebiet im Südosten bis in den Nordosten dominiert. Im Süden und Südwesten schließt sich der Hochrücken der Mecklenburgischen Seenplatte mit dem Neustrelitzer Kleinseenland an. Es herrscht ein humides Kontinentalklima mit jährlichen Niederschlagssummen zwischen rund 500mm und 650mm vor (UMWELTMINISTERIUM MECKLENBURG-VORPOMMERN, 2003).

Die Mecklenburgische Seenplatte beheimatet zahlreiche ökologisch wertvolle zum Teil naturbelassene Vegetationsformen, von denen weitläufige Flächen auf Schutzgebiete und Nationalparks entfallen. Innerhalb des Untersuchungsgebietes erstreckt sich der Müritz-Nationalpark über den Südwesten. Mit rund 262'400 Einwohnern und einer mittleren Bevölkerungsdichte von 48 Einwohnern pro Quadratkilometern stellt der Landkreis eine sehr bevölkerungsarme Region dar. Die bedeutendsten Wirtschaftszweige sind traditionell die Land- und Forstwirtschaft sowie der Tourismussektor. Erneuerbare Energieträger haben sich als neuer Wirtschaftszweig mit großem Zukunftspotential etabliert (REGIONALER PLANUNGSVERBAND MECKLENBURGISCHE SEENPLATTE, 2017).

#### 3.2 Untersuchungsgebiet bei Bobo Dioulasso (Burkina Faso)

Das Untersuchungsgebiet im Südwesten Burkina Fasos bei Bobo-Dioulasso im Gouvernement Hauts-Bassins erstreckt sich über ein rund 100 Quadratkilometer großes rechteckiges Gebiet innerhalb der Zone 30 des UTM-Koordinatensystems, begrenzt durch die Koordinaten 300000m und 409800m (Rechtswert), sowie 1190220m und 1300020m (Hochwert). Dies entspricht etwa dem Gebiet zwischen 4,83° westlicher Länge und 10,76° nördlicher Breite sowie 3.83° westlicher Länge und 11,76° nördlicher Breite in ellipsoidischen Koordinaten, bezogen auf das WGS 84 Ellipsoid (siehe Abbildung 2).

Die räumliche Abgrenzung des Untersuchungsgebietes basiert auf der Aufteilung des Level 2A Produktes der Sentinel 2 Satellitenfernerkundungsmission auf 100km×100km große Kacheln (ESA, 2015). Die Ausmaße des Untersuchungsgebietes entsprechen denen einer Kachel des Level 2A Produktes der Sentinel 2 Mission. Wegen der Lage der Kachel am Rand eines Aufnahmestreifens des Sentinel 2A Satelliten fehlen in den verwendeten Szenen jeweils die Daten innerhalb eins dreieckigen Gebietes im Osten und Südosten des Untersuchungsgebietes (purpur hervorgehobener Bereich in Abbildung 2).

Der Großteil des Untersuchungsgebietes liegt innerhalb des Gouvernements Hauts Bassins (Obere Becken). Im Norden befinden sich kleinere Gebiete innerhalb von la Boucle du Mouhoun ("Schwarzer Volta-Bogen") und im Süden von Cascades ("Wasserfälle"). Das Gouvernement Hauts Bassins macht mit einer Fläche von 25'606km² etwa 9.4% der Gesamtfläche Burkina Fasos aus (MINISTÈRE DE L'AGRICULTURE ET DE L'Hydraulique, 2011). Nach der Volkszählung aus dem Jahr 2006 lebten zu dieser Zeit mit etwa 1'469'600 Einwohnern rund 10,5% der Bevölkerung Burkina Fasos in Hauts Bassins. Damit ist es das Gouvernement mit der landesweit zweitgrößten Bevölkerungszahl. Die Bevölkerungsdichte ist mit rund 57 Einwohnern je Quadratkilometer höher, als in der Mecklenburgischen Seenplatte. Die größte Siedlung innerhalb des Untersuchungsgebietes ist mit über 244'000 Einwohnern Bobo-Dioulasso, die zweitgrößte Stadt Burkina Fasos, gefolgt von Bama mit rund 69'700 Einwohnern (MINISTÈRE DE L'ECONOMIE ET DES FINANCES, 2008). Die größten ethnischen Gruppen in Hauts Bassins sind die Bobo, die Dioula, die Mossi, die Bwaba,

die Sénoufo und die Peulh. Der Großteil der Bevölkerung lebt auf dem Land. Der mit Abstand bedeutendste Wirtschaftssektor ist die Landwirtschaft. Darüber hinaus spielen auch der Handel und der (artisanale) Bergbau eine Rolle. Die Arbeitswelt zeichnet sich durch eine hohe Kinderbeschäftigung aus (MINISTÈRE DE L'AGRICULTURE ET DE L'HYDRAULIQUE, 2011).

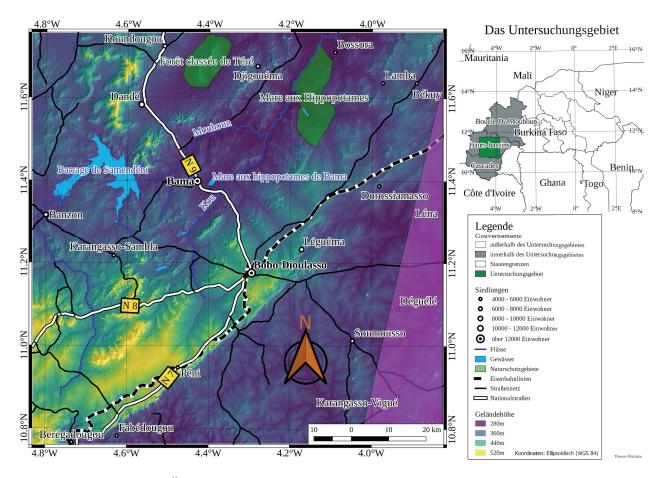


Abbildung 2: Übersichtskarte für das Untersuchungsgebiet bei Bobo Dioulasso

In Abhängigkeit von den Niederschlagsmengen und dem Ausmaß des Wasserstaus ist der Samendeni Stausee (International Monetary Fund, 2008) der größte See innerhalb des Untersuchungsgebietes. Der Stausee wird vom Fluss Mouhoun (bzw. schwarzer Volta) gespeist, dem größten Fluss der Region. Im Norden des Untersuchungsgebietes erstreckt sich das Ramsar Biosphärenreservat des "Mare aux Hippopotames" Feuchtgebietes am Ostufer des Mouhoun (Poda, 2014). Etwas weiter westlich liegt der "Téré Wald" als weiteres bedeutendes Naturschutzgebiet (MINISTÈRE DE L'ENVIRONNEMENT ET DU CADRE DE VIE, 2007).

Die Geomorphologie des Untersuchungsgebietes wird im Nordwesten durch den südlichen Rand des Taoudéni Beckens dominiert (HEIDT, 2017). Das Plateau des Leo-Man Schild erstreckt sich über den Südosten des Gebiets (DIRKS, 2003). Die Grenze zwischen beiden Landschaftsformen verläuft vom Südosten nach Nordwesten und ist insbesondere im Süden durch markante Höhenunterschiede gekennzeichnet. Geologisch überwiegen präkambrische Granite und Gneise auf dem Leo-Man Schild und Sandsteinsedimente im Gebiet des Taoudéni Beckens (DIPAMA, 2010a). Die Böden Burkina Fasos sind generell als sehr nährstoffarm einzustufen. Innerhalb des Untersuchungsgebietes überwiegen ausgewaschene eisenhaltige Böden. In der näheren Umgebung der Flussläufe, vor allem des Mouhouns, kommen hydromorphe Böden vor. Insbesondere innerhalb des Taoudéni Beckens sind vereinzelt auch schwach entwickelte Böden anzutreffen. Böden ohne oder mit stark eingeschränkter Pedogenese bilden die Ausnahme und kommen nur auf kleineren Ge-

bieten im Norden und Osten vor (Traore, 2010; Ruess, 2017).

In Burkina Faso herrscht ein tropisches Klima mit einer kurzen Regenzeit von Juni bis Oktober und einer langen Trockenzeit von November bis Mai. Dabei variieren die klimatischen Bedingungen und die Ausprägungen der Trocken- und Regenzeiten von Nord nach Süd. Das Untersuchungsgebiet liegt in der sudanesischen Klimazone des Südens mit einer mittleren Dauer der Regenzeit von 5 bis 6 Monaten, Temperaturen zwischen 20°C und 25°C und mittleren jährlichen Niederschlagssummen um 1100mm (DIPAMA, 2010b).

# 4 Problemstellung

Die Anwendung von Methodiken der Fernerkundung bei der Suche nach Heilpflanzen und ihren Habitaten kommt als kostengünstige und effiziente Alternative zu in-situ Datenerfassungen infrage. Insbesondere frei verfügbare Daten aus staatlichen Satellitenfernerkundungsmissionen erfüllen die notwendigen Anforderungen hinsichtlich Kostenminimierung, Unabhängigkeit und Abrufbarkeit. Ihre Nutzbarkeit für Anwendungen der Fernerkundung von Pflanzenspezies wird jedoch aufgrund der jeweiligen Sensorspezifika eingeschränkt (beispielsweise durch die räumliche und spektrale Auflösung). Bei geringeren räumlichen Auflösungen werden unterschiedliche Vegetationsformen in erster Linie anhand ihrer spektralen Charakteristika mithilfe unterschiedlicher Klassifikationsalgorithmen gegeneinander abgegrenzt.

#### 4.1 Fernerkundung von Pflanzen

Fernerkundung kann bei der Erforschung von Ökosystemen eine kostengünstige und effektive Alternative zur in-situ Datenerfassung darstellen. Dabei sind die Anwendungsmöglichkeiten mannigfaltig und umfassen neben der Kartographie der Vegetation und ihrer Dichte auch die Bestimmung bzw. Abschätzung vegetativer Merkmale wie z.B. Artenreichtum, spezifische botanische Zusammensetzung und Zustand der Pflanzen oder die Analyse von Landbedeckungsänderungen (Kerr, 2003; Reddy, 2015).

Die grundlegenden Lösungsansätze zur Fernerkundung von Vegetation liegen in der Anwendung spektraler Indizes und von Klassifikationsmethoden (Albertz, 2009). Darüber hinaus werden die Verfahren zur Fernerkundung einzelner botanischer Spezies in indirekte und direkte Methodiken unterteilt (Tan, 2008). Dabei werden Spezies mithilfe direkter Methoden anhand ihrer spezifischen Eigenschaften identifiziert, während bei indirekten Methoden auch aus den vorherrschenden Umweltbedingungen (wie z.B. Temperatur, Feuchtigkeitsgehalt oder Geländehöhe) und anderen Indikatoren mit nachgewiesenen oder vermuteten Einflüssen auf die Biodiversität (wie beispielsweise der Struktur des Habitats, dem Chlorophyllgehalt und der biologischen Produktivität der Vegetation) auf das mögliche Vorhandensein unterschiedlicher Spezies geschlossen wird.

Spektrale Indizes stellen ein ebenso einfaches wie auch effektives Verfahren zur Fernerkundung von Vegetation dar. Sie werden im Allgemeinen als arithmetische Ausdrücke über den spektralen Kanälen eines Fernerkundungssensors definiert. Insbesondere zur Fernerkundung von Vegetationsmerkmalen steht eine Vielzahl unterschiedlicher spektraler Indizes zur Verfügung. Dabei seien als bekannte Beispiele der normalisierte Differenz-Vegetationsindex (NDVI) als Indikator für die Vegetationsdichte und der normalisierte Differenz-Wasserindex (NDWI) zur Feststellung des Wassergehalts genannt (BANNARI, 1995; XUE, 2017). Zur unkomplizierten Veranschaulichung verschiedener Eigenschaften der untersuchten Vegetation können mehrere spektrale Indizes in einem RGB-Komposit zusammengestellt werden (PUJIONO, 2013; RAMDANI, 2015).

Traditionell werden bei der Klassifikation von Fernerkundungsdaten überwachte und unüberwachte bzw. Cluster - Methoden unterschieden. Mithilfe unüberwachter Klassifikationsmethoden lassen sich aus Fernerkundungsdaten direkt und ohne Vorwissen auf der Grundlage spektraler Pixel(/Element)-basierter Statistiken unter Anderem Informationen über unterschiedliche (vegetative) Landbedeckungsarten sowie die Zusammensetzung und Beschaffenheit der beobachteten Gebiete gewinnen. Dabei zielen unüberwachte Klassifikationsalgorithmen in erster Linie auf die Identifikation und Abgrenzung von (lokalen) Ballungen sogenannter "Cluster" - in der untersuchten Datengrundlage ab. Darüber hinaus können mithilfe derartiger

Verfahren auch mögliche unbekannte Strukturen in einem Datensatz aufgedeckt werden. Im Gegensatz dazu werden bei überwachten Klassifikationsverfahren Gebiete mit bekannten Eigenschaften als sogenannte "Trainingsdaten" vordefinierten Klassen zugeordnet und anschließend jedes Element (bei Bilddaten Pixel) aus der Datengrundlage über einen Algorithmus anhand von, aus den Trainingsdaten berechneten, Eigenschaften eindeutig auf die Klassen verteilt. Hierfür ist im Allgemeinen umfassendes Wissen über die Bedingungen vor Ort (im Bezug zur Problemstellung: Anteile und die Verteilung unterschiedlicher Pflanzenarten in der Vegetation) erforderlich. Insbesondere für die Identifikation einzelner botanischer Spezies spielt dabei die räumliche und spektrale Auflösung der Fernerkundungssensoren eine entscheidende Rolle, da sich verschiedene Spezies in erster Linie aufgrund ihrer Spektren unterscheiden lassen, das Spektrum eines Gebietes bei höherer spektraler Auflösung genauer bestimmt werden kann und bei höherer räumlicher Auflösung kleinere Gebiete mit einer geringeren Anzahl unterschiedlicher vorkommender Spezies betrachtet werden können (im Optimalfall lassen sich einzelne Pflanzenteile wie Blätter oder Blüten gegeneinander abgrenzen). Mit höherer räumlicher Auflösung und einer größeren Anzahl spektraler Kanäle lassen sich unterschiedliche Arten und ihre räumlichen Verteilungen genauer bestimmen. Daher werden zur Fernerkundung einzelner Spezies vornehmlich Daten von hyperspektralen Sensoren mit hoher räumlicher Auflösung eingesetzt. Bei geringeren Auflösungen ist in der Regel eher die Identifikation dominanter Spezies möglich. Da es sich bei Vegetation um ein dynamisches System handelt, ist auch das temporale Auflösungsvermögen eines Fernerkundungssensors ist für die Fernerkundung von Pflanzen von Bedeutung. Mehrere Aufnahmen zu unterschiedlichen vegetativen Phasen können für die Abgrenzung unterschiedlicher Spezies vorteilhaft sein. Zur Validierung und Verifikation der Resultate werden die Ergebnisse mit den vor Ort vorgefundenen Bedingungen (in einzelnen ausgewählten Gebieten) verglichen. (XIE, 2008)

Die Spektren einzelner Pflanzen sind von einer Vielzahl unterschiedlicher Parameter (wie z.B. Wassergehalt, Tageszeit, Vitalität, evtl. Befall durch Krankheiten, Parasiten, Pilze, Alter (auch nach dem Abschneiden, wenn Spektren nicht direkt an der Pflanze gemessen werden) etc. abhängig und werden in der Regel durch in-situ Messungen bestimmt. Die Unterscheidung verschiedener Pflanzenarten ist anhand der an ihren Blättern gemessenen Spektren möglich, wobei spektrale Indizes zur Unterstützung bei der statistischen Analyse genutzt werden können (Prospere, 2014).

Mit Hilfe unbemannter Drohnen (UAVs) lassen sich die zur Unterscheidung einzelner Spezies erforderlichen Aufnahmen mit hoher räumlicher Auflösung vergleichsweise kostengünstig (im Gegensatz zur Satellitenfernerkundung) anfertigen. Auf Grundlage derartiger Aufnahmen lassen sich mithilfe multispektraler Datengrundlagen unterschiedliche Baumarten über die Abgrenzung von Baumgruppen und einzelner Baumkronen (unter Anwendung von Segmentierungsalgorithmen) und anschließender Klassifizierung anhand spektraler Parameter und Indizes, zuverlässig unterscheiden (BAENA, 2017).

Im Nahbereich können neben den spektralen Charakteristika (GAI, 2011) auch die Blattgeometrien bei der Identifikation einzelner Spezies berücksichtigt werden (NOBLE, 2009).

Zu den möglichen Anwendungsgebieten der Fernerkundung in der Landwirtschaft zählen beispielsweise die Überwachung der Pflanzenvitalität für ein nachhaltiges und zielgerichtetes Dünge- und Pflegemanagement (CLEVERS, 2017) sowie die Feststellung und räumliche Eingrenzung von Schädlings- (AL-KINDI, 2017) oder Krankheitsbefall (JIN, 2018).

#### 4.2 Fernerkundung im Rahmen von DigiMeP

Das Projekt "Digital Ressource Management for Traditional Medicinal Plants" (Digitales Ressourcenmanagement für traditionelle Heilpflanzen, DigiMeP) wurde im Jahr 2018 an der Hochschule Neubrandenburg initiiert. Die Ziele des Vorhabens bestehen in der Erforschung traditioneller Heilpflanzen in Marokko und Burkina Faso, dem Aufbau einer digitalen Infrastruktur für ein nachhaltiges Ressourcenmanagement sowie dem landwirtschaftlichen Anbau ausgewählter Arten und der Vermarktung der Erzeugnisse.

Im Rahmen des Projektes bestehen Perspektiven für die Anwendung von Methodiken der Fernerkundung bei der Suche nach ausgewählten Spezies und im weiteren Verlauf auch bei der Überwachung der Anbaugebiete von Heilpflanzen. Fernerkundung kann eine kostengünstige und effektive Alternative zu in-situ Analysen darstellen und einen Beitrag zu deren effektiver Planung leisten.

Die Anwendungsmöglichkeiten der Fernerkundung von Pflanzenspezies werden im wesentlichen durch die

technischen Vorgaben der verfügbaren Datengrundlagen beschränkt. Zu den entscheidenden Anforderungen für die unterschiedlichen Verfahrensweisen zählen das räumliche und spektrale Auflösungsvermögen der genutzten Sensoren. Unterschiedliche Spezies lassen sich dabei insbesondere auf der Grundlage hyperspektraler Daten (hohe Anzahl verfügbarer Spektralkanäle) mit hoher räumlicher Auflösung zuverlässiger gegeneinander Abgrenzen (XIE, 2008). Im Rahmen des DigiMeP-Projekts müssen in der aktuellen Phase darüber hinaus auch minimale Kosten und gute Verfügbarkeit als weitere Kriterien berücksichtigt werden. Unter den derzeit verfügbaren Fernerkundungsmissionen wird die Mehrzahl der Anforderungen durch die Sentinel-2 Satellitenmission des Copernicus Programms der europäischen Kommission bestmöglich erfüllt (siehe Abschnitt 5.1 "Datengrundlage"). Dabei erlauben jedoch die räumlichen Auflösungen der spektralen Kanäle des Multispectral Instrument (MSI) an Bord der Sentinel-2 Satelliten von 10, 20 und 60 Metern keine Identifikation einzelner Spezies auf der Grundlage geometrischer Merkmale (z.B. von Kronen und Blättern) und die Analysen beschränken sich hier auf multispektrale Klassifikationen mit Erweiterung des Merkmalsraumes um spektrale Indizes.

#### 4.3 Physikalische Grundlagen

Das wesentliche physikalische Prinzip, auf welchem die gesamte Fernerkundung aufbaut, ist die Photon-Elektron-Wechselwirkung (Albertz, 2009; Jutzi, 2017; Toth, 2017). Diese Wechselwirkung wird mit der Quantenelektrodynamik (QED) beschrieben. Die Theorie der QED dient darüber hinaus auch der Beschreibung sämtlicher Fragestellungen aus der Chemie, da die chemischen Eigenschaften aller Substanzen durch die Elektronen in den Orbitalen der Atome und Moleküle bestimmt werden (FEYNMAN, 2007). Demnach werden Absorption und Emission von Licht durch den Übergang von Elektronen auf unterschiedliche diskrete Energieniveaus der Atome bzw. Moleküle verursacht, wobei die Energie der Photonen gerade der Differenz zwischen den jeweiligen Zuständen entspricht. Dabei findet Absorption statt, wenn Photonen einer bestimmten Frequenz ihre Energie an die Elektronen eines Atoms bzw. Moleküls abgeben und es so von einem Zustand niedrigerer Energie in einen Zustand höherer Energie versetzen und Emission, wenn die Elektronen beim Übergang von einem energetisch höheren in einen energetisch niedrigeren Zustand die Energiedifferenz als Photonen abgeben. Die diskreten Energiezustände, die ein Molekül bzw. ein Atom annehmen kann hängen von dessen Art - also der jeweiligen Substanz bzw. dem Element - ab. Somit ergibt sich aus der Summe der Wirkungen von absorbierten und emittierten Licht für jeden Stoff ein charakteristisches elektromagnetisches Spektrum (Feynman, 2009). Dieser Sachverhalt wird in der Spektroskopie ausgenutzt, um über das Spektrum des von einer Probe ausgestrahlten Lichts auf das Vorhandensein bestimmter Substanzen zu schließen (Sharma, 2017; Merkt, 2011). Zu den Anwendungsbereichen der Spektroskopie gehören unter anderem die Mülltrennung (KILIAN, 2020), die Untersuchung moderner Halbleitermaterialien (Olthof, 2019), die Analyse der Zusammensetzung von Sternen (Rosswog, 2020) sowie die Suche nach organischen Substanzen in der interstellaren Materie (Fuchs, 2010) in der Astrophysik. Die Spektroskopie bildet das Fundament eines umfangreichen Instrumentariums der Fernerkundung mit multi- und hyperspektralen Sensorsystemen (Schaepman, 2009). Neben unterschiedlichen Landbedeckungsarten können auch Vegetationsformen, Pflanzengemeinschaften und einzelne Spezies anhand ihrer spektralen Signaturen unterschieden werden, sodass letztendlich auch die Fernerkundung von Pflanzen mithilfe von multi- oder hyperspektralen Sensoren ein spektroskopisches Verfahren darstellt (Goetz, 1985). Bei der Fernerkundung von Pflanzenspezies anhand ihrer spektralen Signatur werden unterschiedliche Klassifikationsmethoden genutzt, um die Primitive eines Datensatzes gegen ein Referenzspektrum abzugleichen (XIE, 2008).

Im Unterschied zu den spektroskopischen Verfahren der chemischen Analytik werden bei Anwendungen aus der Fernerkundung in der Regel die Spektren größerer Gebiete betrachtet, sodass ein Bildpunkt aus einer Aufnahme häufig mit mehreren Realweltobjekten korrespondiert, welche sich darüber hinaus auch in ihrer Art unterscheiden können. Im Zusammenhang mit Klassifikationsproblemen kann diese Tatsache die gleichzeitige Anwesenheit von Objekten unterschiedlicher Klassenzugehörigkeit innerhalb eines Bildpunktes bedeuten (Mischpixel; Choodarathnakara, 2012). In der Fernerkundung werden die Messungen auch durch die Zusammensetzung und den Zustand der Atmosphäre beeinflusst, wodurch die Auswertung erschwert werden kann. Darüber hinaus ist auch die Anzahl der verfügbaren Spektralkanäle und damit die

spektrale Auflösung der in der Fernerkundung eingesetzten Sensorsysteme in der Regel geringer, als bei spektroskopischen Verfahren in der chemischen Analytik. In der Fernerkundung werden hinsichtlich der spektralen Auflösung multi- und hyperspektrale Sensoren unterschieden. Dabei ist die Anzahl der verfügbaren Spektralkanäle bei hyperspektralen Sensorsystem größer, als bei multispektralen. Die Abgrenzung von multi- gegen hyperspektrale Sensorsysteme anhand der Anzahl der verfügbaren Spektralkanäle ist in der Literatur nicht eindeutig. Hyperspektrale Sensorsysteme können über mehrere Hundert spektrale Kanäle verfügen (Toth, 2017). Mit seinen 12 Spektralkanälen (im Level-2A Produkt verfügbar) zählt das MSI der Sentinel 2 Satellitenfernerkundungsmission zu den multispektralen Sensorsystemen (ESA, 2015). Wegen der zugleich hohen räumlichen Auflösung und der freien Verfügbarkeit der Daten für die Allgemeinheit ist die Sentinel 2 Mission zum aktuellen Zeitpunkt die für die hier vorgesehenen Anwendungen am besten geeignetste Fernerkundungsmission. Abbildung 3 zeigt einen Ausschnitt aus einer Aufnahme des MSI der Sentinel 2 Mission vom 26.07.2019 (ESA, 2019) für die Varchentiner Seen innerhalb des Untersuchungsgebietes bei Kraase mit dem aus der Aufnahme abgeleiteten Spektrum für ein Gebiet (bzw. Bildpunkt/Pixel) auf dem kleinen Varchentiner See und für Wasser typischen Spektren aus hochauflösenden Messungen. Die Namen der einzelnen Spektralkanäle des MSI sind in der Darstellung an den entsprechenden Positionen blau hervorgehoben. Auch bei der geringeren spektralen Auflösung des MSI lassen sich Charakteristika aus den höher aufgelösten Spektren für Wasser wiedererkennen (z.B. die Maxima nahe der Wellenlängen von 600 bzw. 700 nm und das Minimum bei etwa 680nm).

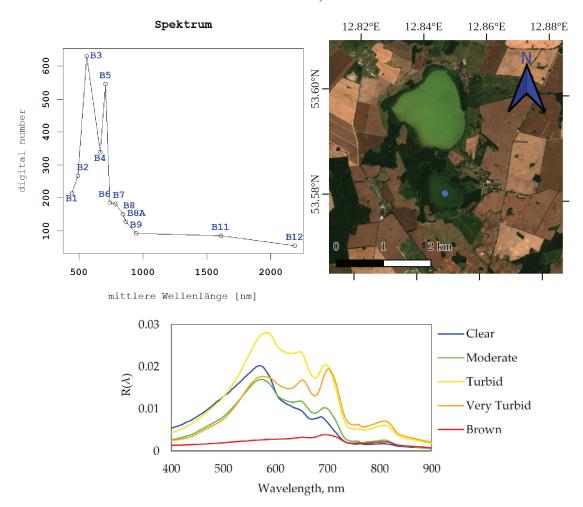


Abbildung 3: Spektrum für ein Gebiet auf dem kleinen Varchentiner See, aufgenommen am 26.07.2019 mit dem MSI von Sentinel 2 (oben links), mit der Position des Gebietes (blauer Punkt) innerhalb des Untersuchungsgebietes bei Kraase (RGB-Echtfarbenkomposit oben rechts) und typischen Spektren für Wasser bei unterschiedlichen Bedingungen (unten; UUDEBERG, 2019)

#### 5 Methodik

Mit den Daten aus den Sentinel Missionen stellt die Europäische Kommission eine Quelle mit großem Anwendungspotential - auch für die Fernerkundung von Heilpflanzen - für jedermann zur freien Verfügung. Insbesondere die Aufnahmen des Multispectral Instruments (MSI) der Sentinel 2 Satelliten kommen für die Fernerkundung von Pflanzen infrage. Da die räumliche Auflösung dieses Sensors für morphologische Analysen (beispielsweise der Kronen- oder Blattgeometrien) zu gering ist, beschränken sich die anwendbaren Methoden auf spektrale Klassifikationen (mit erweiterten Merkmalsräumen).

Um die effektive Verarbeitung der Datengrundlage zu ermöglichen, die Interpretierbarkeit der Ausgangsdaten und damit die Qualität der Endergebnisse zu erhöhen, müssen die Daten im Preprocessing korrigiert und für die folgenden Analysen aufbereitet werden.

Bei Klassifikationen ist eine Erweiterung des Merkmalsraumes um zusätzliche Merkmale möglich, um die Resultate zu verbessern. Andererseits kann die Leistungsfähigkeit klassischer Klassifikatoren in hochdimensionalen Merkmalsräumen abnehmen, sodass die Beschränkung auf eine Merkmalsauswahl zweckmäßig ist - auch um den Zeitaufwand einer Klassifikation zu verringern.

Für die Klassifikation der multispektralen Fernerkundungsdaten werden unterschiedliche überwachte Klassifikationsalgorithmen implementiert. Implementierungen unüberwachter Klassifikations- bzw. Clusterverfahren werden insbesondere zur Abgrenzung der Trainingsgebiete und Extraktion von Zielstrukturen aus den Ergebnissen von Einklassenfall Klassifikationen bereitgestellt.

Sämtliche Algorithmen und Analysen werden in der freien und quelloffenen statistischen Rechen- und Programmierumgebung R implementiert (R CORE TEAM, 2019). Dabei kommen die R-Bibliotheken "sp" in der Version 1.3-1 (PEBESMA, 2018) für grundlegende Funktionalitäten zur Verarbeitung von Geodaten, "rgdal" in der Version 1.4-3 (BIVAND, 2019) für den Import und Export von Geodaten und "raster" in der Version 2.8-19 (HIJMANS, 2019) zur Verarbeitung von Rasterdaten zum Einsatz. Für weitergehende räumliche Analysen sowie zur Aufbereitung und Darstellung der Endergebnisse wird die FOS Software QGIS (QGIS DEVELOPMENT TEAM, 2018) als GIS genutzt.

#### 5.1 Datengrundlage

Für die Fernerkundung von Heilpflanzen sollen hier die Daten aus den Sentinel Satellitenfernerkundungsmissionen des Copernicus Programms der europäischen Kommission zum Einsatz kommen. Die Daten aus allen Sentinel-Satellitenmissionen stehen der Allgemeinheit kostenfrei zur freien Verfügung (COPERNICUS SPACE COMPONENT MISSION MANAGEMENT TEAM, 2019).

Es sind insgesamt 6 Sentinel-Missionen (Sentinel 1-6) geplant, von denen bisher 4 den operationellen Betrieb aufgenommen haben (Sentinel 1-3 und Sentinel 5 Precursor). Bei Sentinel 1 handelt es sich um eine Radarmission (SAR). Das Multispectral Instrument (MSI) von Sentinel 2 dient zur Bereitstellung hochauflösender multispektraler Aufnahmen. An Bord von Sentinel 3 befinden sich Instrumente zur Überwachung des Klimas und der Atmosphäre: das Ocean and Land Cover Instrument mit 21 spektralen Kanälen und einer räumlichen Auflösung (ground sample distance: GSD) von 300m, das Surface Temperature Instrument zur Messung der Oberflächentemperatur (500m GSD) und schließlich das SAR radar altimeter (SRAL) zur Messung der Topographie. Derzeit befinden sich für die Missionen Sentinel 1-3 jeweils zwei Satelliten (A und B) im Orbit, wobei für die nächsten Jahre eine Aufstockung auf jeweils 4 Satelliten (um C und D; auch zur Ablösung der ersteren nach ihrer voraussichtlichen aktiven Phase) geplant ist. Mit der vollständigen Einsatzbereitschaft der Sentinel 1-6 Missionen wird zurzeit Mitte der 2020er Jahre gerechnet. Eine Erweiterung der Flotte um zusätzliche Sentinel Missionen ist möglich und für die Zeit nach 2030 sind weitere Nachfolgemissionen geplant. Dabei werden die Sentinel 4 und 5 Missionen der Untersuchung der (chemischen) atmosphärischen Zusammensetzung und der Luftqualität dienen. Für die Übergangsphase bis zur Bereitschaft der Sentinel 5 Satelliten befindet sich aktuell der Sentinel 5 Precursor Satellit im Orbit. Der Beginn der Radar Altimeter Mission Sentinel 6 mit dem Start des Sentinel 6 A Satelliten ist schließlich für das Jahr 2020 vorgesehen. (COPERNICUS SPACE COMPONENT MISSION MANAGEMENT TEAM, 2019; EUROPEAN COMMISSION, 2019; EUROPEAN COMMISSION, 2017; TOTH, 2017)

Für die Fernerkundung von Heilpflanzen bieten sich insbesondere die hochauflösenden multispektralen Auf-

nahmen des MSI der Sentinel 2 Mission an. Die Sentinel 2 Mission ist als Fortführung und Verbesserung der multispektralen Satellitenfernerkundungsmissionen SPOT und Landsat konzipiert. Die Aufnahmen des MSI bieten 13 spektrale Bänder unterschiedlicher Bandbreiten und räumlicher Auflösungen (GSD von 10m bis 60m, siehe Abbildung 4; ESA, 2015; GMES PROGRAMME, 2012). Verschiedene Pflanzengemeinschaften und dominante Spezies lassen sich bereits mithilfe der TM- und und ETM-Instrumente der Landsat-Missionen identifizieren (XIE, 2008). Wegen der Verbesserungen (in erster Linie der höheren räumlichen Auflösung) des MSI von Sentinel-2 im Vergleich zu den Instrumenten von Landsat, kann von einer besseren Eignung der Sentinel-Daten zur Fernerkundung von Vegetationsmerkmalen ausgegangen werden. Darüber hinaus ist für derartige Anwendungen auch die höhere temporale Auflösung der Sentinel-2 Daten (zwischen 2 Überflügen liegen, dank der Nutzung zweier baugleicher Satelliten, für Gebiete in Äquatornähe 5 Tage) von Vorteil (ZHANG, 2017).

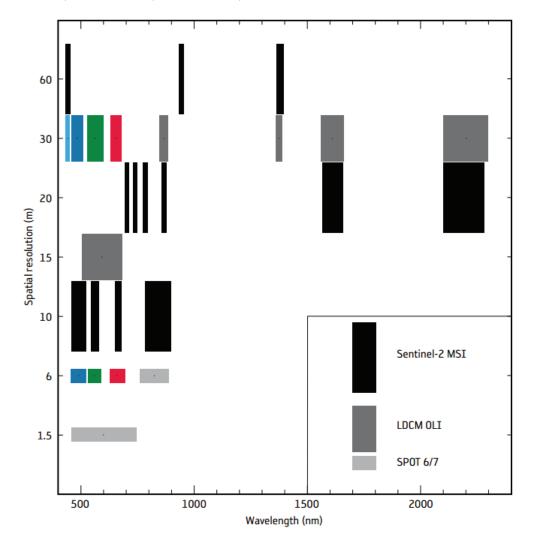


Abbildung 4: Spektralbänder des MSI von Sentinel 2 mit Bandbreite und räumlicher Auflösung (GSD) im Vergleich mit dem SPOT 6/7 Instrument und dem OLI von Landsat (GMES PROGRAMME, 2012)

Zur Erhöhung der Interpretierbarkeit der Satellitenaufnahmen müssen die Daten vor der weiteren Verwendung bei der Fernerkundung von Pflanzenspezies geometrisch, radiometrisch und atmosphärisch korrigiert werden (XIE, 2008). Für die Daten aus der Sentinel 2 Mission werden 3 unterschiedliche Produkttypen definiert (siehe Abbildung 5). Das für die Problemstellung erforderliche atmosphärisch korrigierte Level-2A Produkt (BOA: bottom of atmosphere) wird seit dem Anfang des Jahres 2019 systematisch generiert und

steht den Anwendern zur freien Verfügung (KNOWELDEN, 2019). Dadurch entfällt beim Nutzer die bis dahin erforderliche manuelle Atmosphärenkorrektur mit "Sen2Cor" als zusätzliche Fehlerquelle und sehr zeitaufwändiger Teilschritt beim Preprocessing (CLERK, 2020). Da das Band 10 des MSI der Detektion von Zirruswolken dient und keine Oberflächeninformationen enthält, ist es nicht im Level 2A Produkt enthalten, sodass für diesen Produktlevel nur 12 Bänder (1,2,3,4,5,6,7,8,8A,9,11 und 12) verfügbar sind (ESA, 2015).

Name	High-level Description	Production & Distribution	Data Volume
Level-1B	Top-of-atmophere radiances in sensor geometry	Systematic generation and on- line distribution	27 MB (each 25x23km²)
Level-1C	Top-of-atmosphere reflectances in cartographic geometry	Systematic generation and on- line distribution	500 MB (each 100x100km²)
Level-2A	Bottom-of-atmosphere reflectances in cartographic geometry (prototype product)	Generation on user side (using Sentinel-2 Toolbox)	600 MB (each 100x100km²)

Abbildung 5: Sentinel 2 Produkttypen (ESA, 2015)

#### 5.2 Preprocessing

Aufgrund des Wegfalls der Atmosphärenkorrektur als sehr aufwändiger Teilschritt bei der Vorbereitung der Daten durch die systematische Generierung und freie Verfügbarkeit des Level 2A Produktes, beschränkt sich das Preprocessing auf das räumliche Zuschneiden der Aufnahmen auf das Untersuchungsgebiet und das Hochskalieren (Resampling) der einzelnen Bänder auf identische räumliche Auflösungen (siehe Abbildung 6). Bei der Auswahl der Level 2A Produkte wird darauf geachtet, dass für die weiteren Analysen ausschließlich wolkenfreie Aufnahmen genutzt werden. Da für alle weiteren Berechnungen die freie Rechenund Programmierumgebung R (R CORE TEAM, 2019) zum Einsatz kommt, wird auch das Preprocessing mit R durchgeführt und die Ausgangsdatensätze für die folgenden Analysen in ein kompatibles Format exportiert.

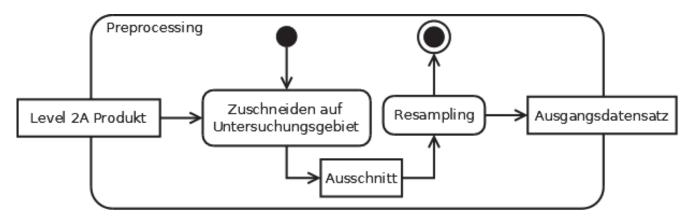


Abbildung 6: Preprocessing Prozesskette

Wegen der räumlichen Abgrenzung des Untersuchungsgebietes bei Bobo Dioulasso auf die Ausmaße einer Kachel des Level 2A Produktes entfällt für dieses Untersuchungsgebiet das Zuschneiden als Teilschritt. Um Informationsverlust zu vermeiden und die Vorteile der räumlich höher aufgelösten Bänder des MSI optimal auszunutzen, wird die Auflösung aller Bänder beim Resampling auf die höchstmöglich verfügbare Auflösung (GSD 10m bei den Bändern 2,3,4 und 8) angeglichen. Dabei wird bei den 8 Bändern von geringerer räumlicher Auflösung ein bilineares Interpolationsverfahren für das Hochskalieren genutzt. Die resultierenden Ausgangsdaten werden je Aufnahme zu "RasterBricks" (Objekt der R-Bibliothek "raster") zusammengefasst und als GeoTiff exportiert.

#### 5.3 Merkmalsraumerweiterung

Bei der Klassifikation multispektraler Daten können neben den verfügbaren Spektralkanälen auch über diesen definierte spektrale Indizes als zusätzliche Merkmale genutzt werden (ROTTENSTEINER, 2017). Derartige Merkmalsraumerweiterungen können bei der Fernerkundung von Pflanzen vorteilhaft sein (PROSPERE, 2014; KOBAYASHI, 2020). Daher werden hier im Vorfeld von Klassifikationen zusätzlich 4 spektrale Indizes mit Relevanz für die Problemstellung zur Erweiterung des Merkmalsraumes berechnet: der Leaf Chlorophyll Index (LCI: Blatt Chlorophyll Index), der Enhanced Vegetation Index 2 (EVI2: verbesserter Vegetationsindex 2), der Modified Chlorophyll Absorption in Reflectance Index (MCARI: modifizierter Chlorophyll Absorption im Reflexionsgrad Index) und der Normalized Difference Moisture Index (NDMI: normalisierter Differenz Feuchtigkeitsindex; siehe Tabelle 1).

Tabelle 1: spektrale Indizes zur Erweiterung des Merkmalsraumes mit Formeln (bezogen auf die Spektralbänder des MSI) und Quellen

Index	Formel	Quelle
LCI	$\frac{B8 - B5}{B8 + B4}$	(Datt, 1999)
EVI2	$2,4*\frac{B8-B4}{B8+B4+1}$	(JIANG, 2008)
MCARI	$[(B5 - B4) - 0, 2 \times (B5 - B3)] \times \frac{B5}{B4}$	(Daughtry, 2000)
NDMI	$\frac{B8 - B11}{B8 + B11}$	(Hardisky, 1983)

#### 5.4 Merkmalsreduktion

Die Leistungsfähigkeit klassischer Klassifikatoren - wie z.B. nach dem kleinsten Abstand oder der Maximum Likelihood Klassifikation - kann bei höherdimensionalen Problemstellungen durch den sogenannten "Fluch der Dimensionalität" (Bellman, 1957) bzw. das "Hughes-Phänomen" (Hughes, 1968) negativ beeinflusst werden. Bei der Maximum Likelihood Klassifikation kann die Inverse der Kovarianzmatrix instabil werden, wenn Merkmale stark korreliert sind (Albertz, 2009). Im Extremfall ist die Kovarianzmatrix singulär und damit nicht invertierbar. Bei höherdimensionalen Problemstellungen liefern stark korrelierte Merkmale weniger zusätzliche Informationen. Eine weitere Ursache für die Abnahme der Leistungsfähigkeit besteht im exponentiellen Wachstum des multidimensionalen Volumens mit zusätzlichen Dimensionen, sodass hochdimensionale Räume in der Regel dünn besetzt sind. Moderne Verfahren des maschinellen Lernens sind gegenüber dieses Problems robuster. Dabei kann Hochdimensionalität sogar von Vorteil sein (Gorban, 2018). Die Auswirkungen des "Fluchs der Dimensionalität" sind insbesondere in der hyperspektralen Fernerkundung von Bedeutung. Zur Lösung des Problems stehen unterschiedliche Verfahren zur Merkmalsauswahl und Informationsextraktion zur Verfügung (Thilagavathi, 2019).

Da stark korrelierte Merkmale auch redundant sind und damit weniger zusätzliche Informationen liefern

(Hall, 1999), wird hier die Merkmalsauswahl auf Grundlage der Korrelation zwischen Merkmalen als entscheidenden Parameter vorgenommen. Als Maß für die Korrelation zwischen zwei Merkmalen wird dabei der Pearson'sche Korrelationskoeffizient  $r_{xy}$ :

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$
(1)

mit der Anzahl n der Elemente der Merkmale x und y, dem i-ten Element  $x_i$  des Merkmals x, bzw.  $y_i$  des Merkmals y sowie den Mittelwerten  $\bar{x}$  und  $\bar{y}$  der Merkmale x und y genutzt (Bronstein, 2012). Die Korrelationen zwischen den einzelnen Merkmalen sind unterschiedlich stark ausgeprägt (siehe Tabelle 2 und Abbildung 7 für ein Beispiel). Häufig bestehen starke Korrelationen zwischen Spektralbändern deren Wellenlängenbereiche nahe beieinander liegen (wie z.B. die Bänder 6,7,8 und 8A beim MSI) oder sich überschneiden (wie z.B. die Bänder 8 und 8A beim MSI).

Tabelle 2: Pearson'sche Korrelationskoeffizienten für die Merkmale aus einer Aufnahme des MSI von Sentinel 2 vom 25.08.2019 (ESA, 2019) für das Untersuchungsgebiet bei Kraase

	B1	B2	В3	B4	В5	В6	В7	В8	B8A	В9	B11	B12
B1	1											
B2	0.93	1										
В3	0.92	0.98	1									
B4	0.93	0.98	0.97	1								
B5	0.93	0.95	0.97	0.96	1							
B6	-0.02	-0.06	0.07	-0.06	0.13	1						
B7	-0.24	-0.24	-0.17	-0.29	-0.12	0.96	1					
В8	-0.24	-0.24	-0.16	-0.28	-0.11	0.93	0.97	1				
B8A	-0.26	-0.26	-0.19	-0.3	-0.13	0.95	0.99	0.97	1			
В9	-0.26	-0.25	-0.19	-0.28	-0.13	0.91	0.95	0.91	0.96	1		
B11	0.84	0.84	0.84	0.86	0.91	0.28	0.05	0.06	0.07	0.08	1	
B12	0.93	0.93	0.91	0.96	0.94	0.04	-0.19	-0.18	-0.19	-0.17	0.95	1
LCI	-0.74	-0.77	-0.77	-0.78	-0.73	0.47	0.64	0.65	0.67	0.66	-0.51	-0.68
EVI2	-0.81	-0.85	-0.84	-0.87	-0.79	0.42	0.61	0.61	0.63	0.61	-0.61	-0.78
MCARI	-0.56	-0.64	-0.54	-0.68	-0.47	0.57	0.65	0.62	0.64	0.59	-0.39	-0.58
NDMI	-0.83	-0.81	-0.77	-0.86	-0.83	0.10	0.31	0.33	0.29	0.23	-0.89	-0.93
	LCI	EVI2	MCARI	NDMI								
LCI	1											
EVI2	0.97	1										
MCARI	0.61	0.72	1									
NDMI	0.65	0.75	0.61	1								

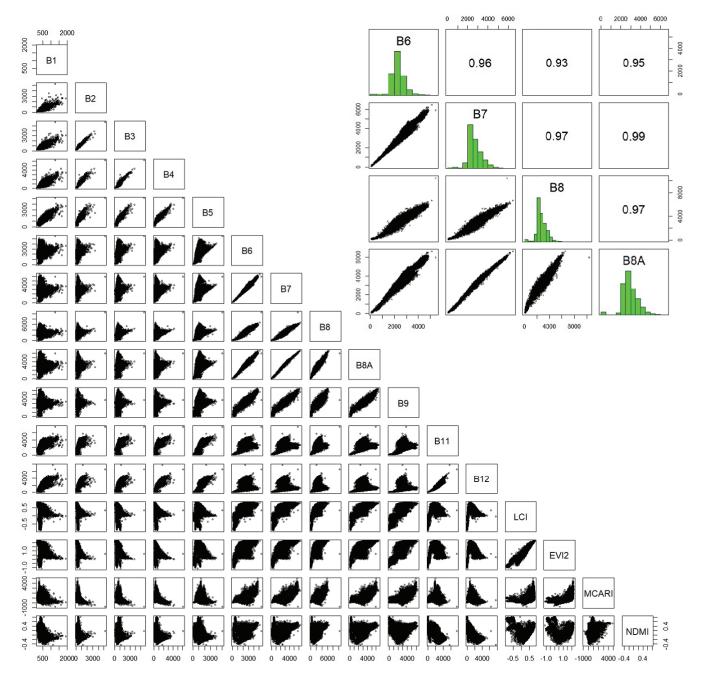


Abbildung 7: Streudiagramme für die Merkmale aus der Aufnahme vom 25.08.2019 (ESA, 2019) für das Untersuchungsgebiet bei Kraase (siehe Tabelle 2 für Korrelationskoeffizienten); Detail oben rechts für eine Auswahl stark korrelierter Merkmale mit Histogrammen und Korrelationskoeffizienten; der Absolutbetrag des Korrelationskoeffizienten zwischen zwei Merkmalen liegt umso näher bei 1, je besser sich die Punktwolke aus dem Streudiagramm durch eine Gerade approximieren lässt

Für die Merkmalsauswahl soll hier aus allen verfügbaren Merkmalen einer Aufnahme ein Tupel von Merkmalen mit minimaler Korrelation unter einem Schwellwert bestimmt werden. Dazu werden zunächst alle Korrelationen zwischen den verfügbaren Merkmalen einer Aufnahme nach Gleichung 1 berechnet und in einer Matrix zusammengestellt (analog zu Tabelle 2). Für die Auswahl eines Tupels von Merkmalen mit minimaler Korrelation unter einem zuvor festgelegten Schwellwert werden zunächst alle Elemente auf der Hauptdiagonalen der Matrix der Korrelationskoeffizienten entfernt. Da es sich bei diesen Elementen jeweils um die Korrelationen der Merkmale mit sich selbst handelt, werden sie stets 1 und damit für die

Merkmalsauswahl unbedeutend sein. Anschließend wird aus der Matrix der Korrelationskoeffizient mit dem kleinsten Absolutbetrag ausgewählt. Die zugehörigen Merkmale bilden die ersten Elemente des Tupels der Merkmalsauswahl. Ist der Absolutbetrag des ausgewählten Korrelationskoeffizienten größer oder gleich dem im Vorfeld festgelegten Schwellwert, so existiert kein Tupel von Merkmalen minimaler Korrelation kleiner als dieser Schwellwert und der Auswahlprozess wird ohne Ergebnis abgebrochen. Im nächsten Schritt werden alle Zeilen und Spalten der Korrelationskoeffizientenmatrix gelöscht, die mit Merkmalen korrespondieren, welche mit den zuvor ausgewählten Merkmalen stärker als der Schwellwert korreliert sind. Die Merkmalsauswahl ist vollständig, wenn im Anschluss die Anzahl der Zeilen bzw. Spalten der Korrelationskoeffizientenmatrix (entspricht den für die Auswahl noch infrage kommenden Merkmalen) gerade der Anzahl der bereits ausgewählten Merkmale entspricht. In diesem Fall existieren außer den bereits gewählten Merkmalen keine weiteren, die für die Auswahl infrage kommen. So lange die Anzahl der Spalten bzw. Zeilen der Korrelationskoeffizientenmatrix an dieser Stelle noch größer als die Anzahl der bereits ausgewählten Merkmale ist, wird der Prozess ab der Auswahl des Korrelationskoeffizienten mit dem kleinsten Absolutbetrag aus der Matrix wiederholt. In jedem Durchlauf wird überprüft, ob der Betrag des kleinsten Korrelationskoeffizienten größer oder gleich dem im Vorfeld festgelegten Schwellwert ist und die Merkmalsauswahl mit den bereits gewählten Merkmalen als Resultat abgebrochen werden kann. Der Algorithmus für dieses Verfahren der Merkmalsauswahl ist in Abbildung 8 zusammengefasst und wird als Funktion in R implementiert. Darüber hinaus wird für die Berechnung der Matrix der Korrelationskoeffizienten für die Merkmale einer Aufnahme eine separate Funktion implementiert.

Als Beispiel ergibt sich für die Aufnahme des MSI vom 25.08.2019 (ESA, 2019) für das Untersuchungsgebiet bei Kraase das Tupel aus den Spektralbändern B1,B2 und B11 sowie den spektralen Indizes NDMI, MCA-RI und EVI2 als Merkmalsauswahl nach minimalen Korrelationskoeffizienten unter einem Schwellwert von 0.9 (vergleiche Tabelle 2).

Um die Auswirkungen dieses Verfahrens zur Merkmalsauswahl auf Resultate von Klassifikationen mit klassifikatoren anhand eines Beispiels zu verdeutlichen, wird jeweils eine Maximum Likelihood Klassifikation als Einklassenfall für Nadelwälder für das Untersuchungsgebiet bei Kraase mit und ohne Merkmalsauswahl (über alle verfügbaren Merkmale) durchgeführt (siehe Abschnitt 5.7.2 "Maximum Likelihood Klassifikation" und 6 "Fernerkundung von Nadel- und Mischwald"). Da sich die immergrünen Nadelwälder in Aufnahmen aus dem Winter deutlicher mit dem bloßen Auge von Laubwäldern abgrenzen lassen, wird eine Aufnahme des MSI von Sentinel 2 vom 16.02.2019 (ESA, 2019) als Ausgangsdatensatz genutzt. So sind die Gebiete innerhalb des Waldes im Osten von Möllenhagen in denen Nadelgehölz dominiert gut in dem unteren Kartenausschnitt in Abbildung 9 für den Vergleich auszumachen. Der vorgestellte Algorithmus zur Merkmalsauswahl liefert für diese Aufnahme für Schwellwerte von 0.7 bzw. 0.8 jeweils identische Tupel mit den Spektralkanälen B1, B8 und B11 sowie den spektralen Indizes MCARI, EVI2 und NDMI als Auswahl. Im Vergleich zu dem Ergebnis der Einklassen-ML-Klassifikation ohne Merkmalsauswahl sind die resultierenden Werte der Likelihood-Funktion bei der Klassifikation mit Merkmalsauswahl im Durchschnitt größer. Die Ursache für diese Auswirkung liegt in dem dichter besetzten Merkmalsraum bei Beschränkung auf die Merkmalsauswahl: Wegen des exponentiellen Zuwachs des Volumens des Merkmalsraumes mit zusätzlichen Dimensionen wird dieser bei gleichem Umfang der Datengrundlage (auch in Bezug auf die Trainingsdaten) umso dichter besetzt sein, je weniger Merkmale berücksichtigt werden. Innerhalb des in Abbildung 9 berücksichtigten Ausschnitts aus den Wertebereichen der Likelihood-Funktionen zwischen -200 und 0 finden sich in dem Resultat der Klassifikation mit Merkmalsauswahl mehr zusätzliche Treffer in Waldgebieten mit geringerer Dichte von Nadelgehölz. Dabei gilt jedoch zu berücksichtigen, dass die Spannweite des Wertebereichs der Likelihood-Funktion bei der Klassifikation mit Merkmalsauswahl geringer ist (etwa von -808'000 bis -10), als bei der Klassifikation ohne Merkmalsauswahl (etwa von -21'000'000 bis -33). Letztendlich lässt sich durch die Beschränkung auf die Merkmalsauswahl auch eine erhebliche Zeitersparnis bei der Klassifikation der Daten erzielen.

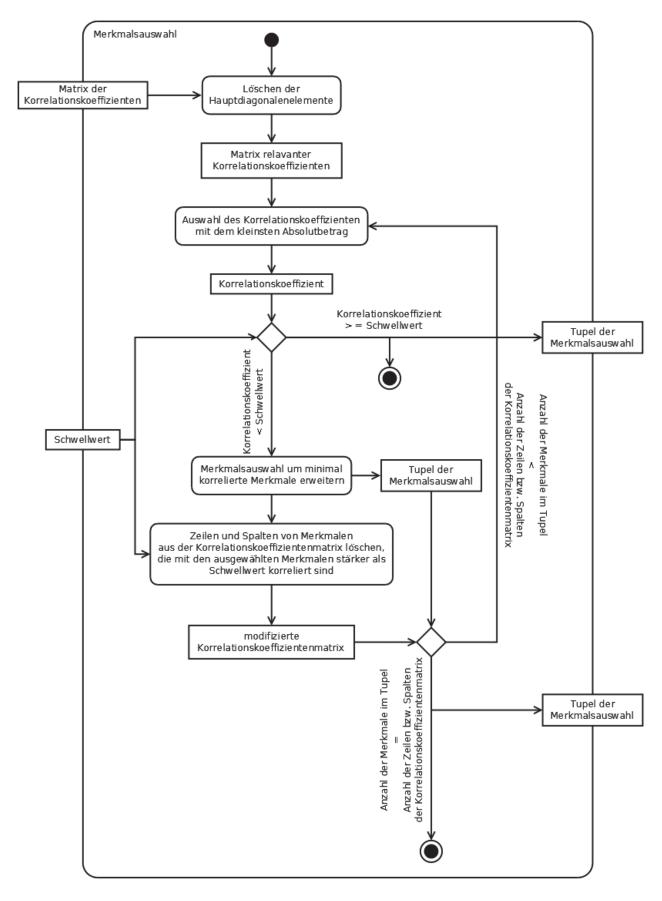


Abbildung 8: Workflow Merkmalsauswahl

# **Einklassen MLC**

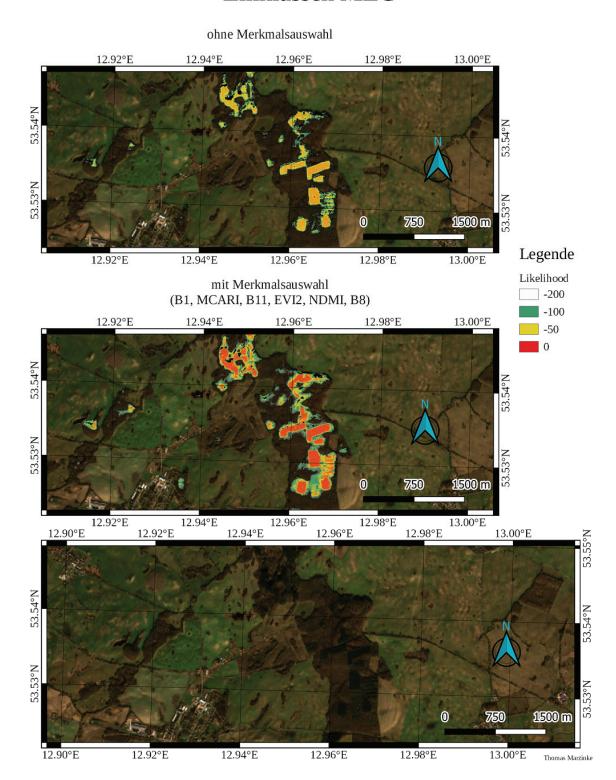


Abbildung 9: Ausschnitte aus Resultaten von Maximum Likelihood Klassifikationen als Einklassenfall einer Aufnahme des MSI vom 16.02.2019 für ein Waldgebiet im Osten von Möllenhagen innerhalb des Untersuchungsgebietes bei Kraase für Nadelwälder ohne (oben) und mit Merkmalsauswahl (Mitte) nach minimaler Korrelation unter einem Schwellwert; unten: RGB-Echtfarbenkomposit der Originaldaten zum Vergleich

#### 5.4.1 Berechnung der Korrelationskoeffizienten bei großem Ausgangsdatenvolumen

Die Berechnung der Pearson'schen Korrelationskoeffizienten für alle Merkmale einer Aufnahme ist bei der Berücksichtigung aller Primitive (bzw. Bildpunkte/Pixel) eine sehr aufwändige Aufgabe. Der Zeitaufwand hält sich für Aufnahmen aus dem vergleichsweise kleinen Untersuchungsgebiet bei Kraase noch in Grenzen, ist für Aufnahmen aus dem Untersuchungsgebiet bei Bobo Dioulasso wegen des exponentiellen Zuwachs der Anzahl zu berücksichtigender Primitive mit der Fläche jedoch unpraktikabel groß. Daher bietet sich die Bereitstellung einer alternativen Methode zur Berechnung der Korrelationskoeffizienten auf der Grundlage einer zufälligen Stichprobe an. Dabei ist die Festlegung des Stichprobenumfangs von entscheidender Bedeutung, um die Berechnung der Korrelationskoeffizienten mit ausreichender Genauigkeit zu gewährleisten. Für bivariate normalverteilte Zufallsvariablen folgt aus dem Konfidenzintervall für Korrelationskoeffizienten für die erste Näherung des Stichprobenumfangs  $n_0$  zur Bestimmung eines Korrelationskoeffizienten  $\theta$  der geschätzten Größe  $\tilde{\theta}$  innerhalb eines Konfidenzintervalls der Weite w mit der Irrtumswahrscheinlichkeit  $\alpha$ 

 $n_0 = 4c^2 (1 - \tilde{\theta}^2)^2 \left(\frac{z_{\alpha/2}}{w}\right)^2 + b \tag{2}$ 

mit dem  $\alpha/2$ -Quantil  $z_{\alpha/2}$  der standardisierten Normalverteilung und den Konstanten c=1 und b=3 für den Pearson'schen Korrelationskoeffizienten (BONETT, 2000). Für die hier vorgesehenen Anwendungen ist eine Weite w des Konfidenzintervalls von 0.01 (Bestimmung des Korrelationskoeffizienten auf den Prozent genau) mit einer Irrtumswahrscheinlichkeit von 5% ( $\alpha=0.05$  - signifikant) ausreichend. Problematisch ist die Abhängigkeit des Stichprobenumfangs  $n_0$  von der Größe des zu bestimmenden Korrelationskoeffizienten  $\tilde{\theta}$  (Schätzwert). Die Korrelationen zwischen den Merkmalen einer Aufnahme können unterschiedlich stark ausgeprägt sein (siehe Abbildung 7 und Tabelle 2). Die Lösung besteht in der Analyse des Verhaltens von  $n_0$  für unterschiedliche Werte von  $\tilde{\theta}$  bei  $\alpha=0.05$  und w=0.01 (siehe Abbildung 10).

#### Entwicklung Stichprobenumfang

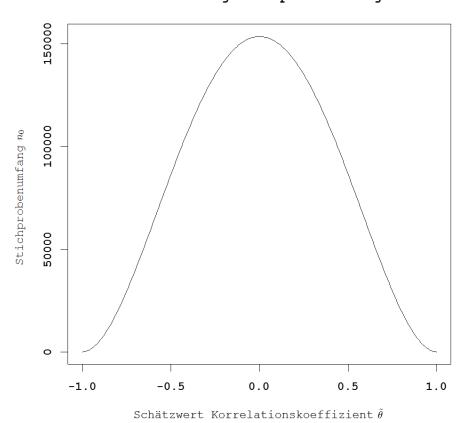


Abbildung 10: Entwicklung des erforderlichen Stichprobenumfangs  $n_0$  für unterschiedliche Schätzwerte  $\tilde{\theta}$  des Korrelationskoeffizienten bei w=0.01 und  $\alpha=0.05$ 

Der Stichprobenumfang erreicht sein Maximum von etwa 150'000 bei einem Schätzwert des Korrelationskoeffizienten von 0 und wird für Schätzwerte von -1 bzw. +1 minimal. Der Pearson'sche Korrelationskoeffizient kann als Wahrscheinlichkeit interpretiert werden, dass die gegeneinander abgetragenen Messwerte zweier Merkmale auf einer Geraden (Regressionsgerade) liegen (BRONSTEIN, 2012). Dementsprechend stellt sich die Bestimmung eines Korrelationskoeffizienten von 0 (vollständig unkorrelierte Merkmale) als besonders aufwändig dar, da zur Approximation der Messwerte in diesem Fall unendlich viele Geraden mit einer Wahrscheinlichkeit von 0 infrage kämen. Hingegen ist der erforderliche Stichprobenumfang zur Bestimmung des Korrelationskoeffizienten bei maximaler Korrelation (+/-1; alle Messungen liegen mit 100%iger Wahrscheinlichkeit auf einer Geraden) minimal. Für den Grenzfall eines Korrelationskoeffizienten von +/-1 genügen 2 Messungen - die in jeden Fall zu Korrelationskoeffizienten von +/-1 führen. Letztendlich ist es möglich, einen maximal erforderlichen Stichprobenumfang  $n_0$  bei konstanter Weite w des Konfidenzintervalls und Irrtumswahrscheinlichkeit  $\alpha$  zu bestimmen, bei dem eine Bestimmung von beliebig großen Korrelationskoeffizienten mindestens mit den festgesetzten Anforderungen an die Genauigkeit gewährleistet ist.

Zur genauen Bestimmung des höchstmöglichen Stichprobenumfangs bei festgelegtem w und  $\alpha$  wird die erste und zweite partielle Ableitung von Gleichung 2 nach dem Schätzwert  $\tilde{\theta}$  des Korrelationskoeffizienten bestimmt:

$$\frac{\partial n_0}{\partial \tilde{\theta}} = -16c^2 \tilde{\theta} (1 - \tilde{\theta}^2) \left(\frac{z_{\alpha/2}}{w}\right)^2 \tag{3}$$

$$\frac{\partial^2 n_0}{\partial \tilde{\theta}^2} = -16c^2 \left(\frac{z_{\alpha/2}}{w}\right)^2 (1 - 3\tilde{\theta}^2) \tag{4}$$

Für  $\tilde{\theta}=0$  mit w=0.01 und  $\alpha=0.05$  wird Gleichung 3 gleich null (notwendige Bedingung für eine Extremstelle; Bronstein, 2012) und Gleichung 4 kleiner null (hinreichende Bedingung für lokale Maxima), folglich hat Gleichung 2 an der Stelle  $\tilde{\theta}=0$  ein lokales Maximum. Daraus ergibt sich ein höchstens erforderlicher Stichprobenumfang von  $n_0=153'667$ . Da der Unterschied des in zweiter Näherung berechneten Stichprobenumfangs im Vergleich zu  $n_0$  nur gering ist (Bonett, 2000), die Genauigkeit des bestimmten Korrelationskoeffizienten mit größerem Stichprobenumfang zunimmt und der Mehraufwand bei der Berechnung vertretbar ist, wird der Umfang der Stichprobe auf 200'000 aufgerundet und eine alternative Funktion zur Berechnung des Korrelationskoeffizienten auf Grundlage einer solchen zufälligen Stichprobe bei großem Ausgangsdatenvolumen in R implementiert.

Bei der Bestimmung von Korrelationskoeffizienten zwischen Merkmalen aus Beispielaufnahmen auf der Grundlage zufälliger Stichproben variablen Umfangs lässt sich eine schnelle Konvergenz der ermittelten Korrelationskoeffizienten gegen den tatsächlichen Wert mit zunehmenden Stichprobenumfang beobachten (siehe Abbildung 11). Häufig geht der relative Fehler bereits für Stichprobenumfänge deutlich kleiner als 200'000 in Bereiche über, die für die hier vorgesehenen Anwendungen ausreichend sind, sodass die Bestimmung der Korrelationskoeffizienten mit der alternativen Methode über eine Stichprobe mit ausreichender Genauigkeit sichergestellt ist. Generell konvergiert der absolute Fehler für Korrelationskoeffizienten mit größerem Absolutbetrag schneller gegen null. In Teilbereichen können die Näherungswerte für stärker korrelierte Merkmale unter Umständen auch schlechter sein, als für schwächer korrelierte.

#### Absolute Fehler

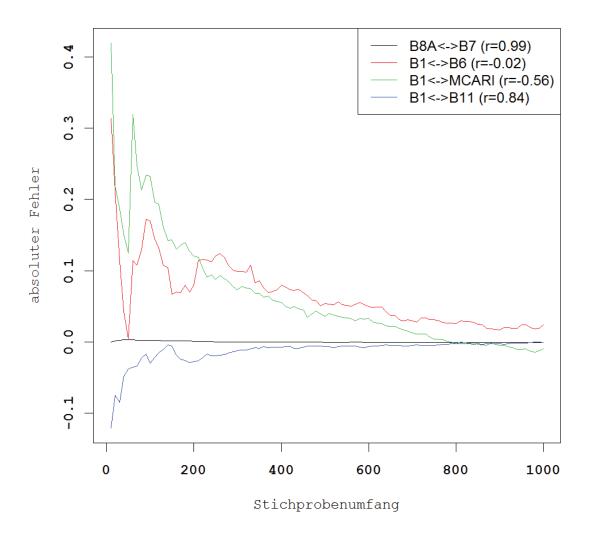


Abbildung 11: Absolute Fehler (Korrelationskoeffizient aus Stichprobe — wahrer Wert) von, aus zufälligen Stichproben unterschiedlichen Umfangs ermittelten, Korrelationskoeffizienten für unterschiedlich stark korrelierte Merkmale aus der Beispielaufnahme vom 25.8.2019 (ESA, 2019) für das Untersuchungsgebiet bei Kraase (vergleiche Tabelle 2)

#### 5.5 Jeffries-Matusita Abstand

Bei Mehrklassenfall-Klassifikationen kann der Jeffries-Matusita Abstand zur Beurteilung der Separabilität der durch die jeweiligen Trainingsdaten definierten Klassenverteilungen genutzt werden. Der Jeffries-Matusita Abstand  $J_{ij}$  mit:

$$J_{ij} = \int_{\mathbf{x}} \{ \sqrt{p(\mathbf{x}|\omega_i)} - \sqrt{p(\mathbf{x}|\omega_j)} \}^2 d\mathbf{x}$$
 (5)

kann als mittlerer Abstand zwischen zwei Wahrscheinlichkeitsdichtefunktionen  $p(\boldsymbol{x}|\omega_i)$  und  $p(\boldsymbol{x}|\omega_j)$  über alle Positionen  $\boldsymbol{x}$  interpretiert werden. Im Zusammenhang mit Klassifikationsproblemen beziehen sich diese Wahrscheinlichkeitsdichtefunktionen auf die i-te und j-te Klasse und werden mithilfe der zugehörigen Trainingsdaten  $\omega_i$  bzw.  $\omega_j$  definiert. Dabei entsprechen die Positionen  $\boldsymbol{x}$  Vektoren im Merkmalsraum (Merkmalsvektorern - jedem Primitiv wird ein Merkmalsvektor zugeordnet). Folglich sind  $p(\boldsymbol{x}|\omega_i)$  und

 $p(\boldsymbol{x}|\omega_j)$  die Wahrscheinlichkeitsdichtefunktionen des Merkmalsvektors  $\boldsymbol{x}$  unter der Bedingung er gehört zur *i*-ten Klasse, definiert durch die Trainingsdaten  $\omega_i$ :  $p(\boldsymbol{x}|\omega_i)$ , bzw. zur *j*-ten Klasse, definiert durch die Trainingsdaten  $\omega_i$ :  $p(\boldsymbol{x}|\omega_i)$ . Für multivariate Normalverteilungen folgt für den Jeffries-Matusita Abstand:

$$J_{ij} = 2(1 - e^{-B}) (6)$$

Mit dem Bhattacharyya Abstand B:

$$B = \frac{1}{8} (\boldsymbol{m}_i - \boldsymbol{m}_j)^{\mathrm{T}} \left\{ \frac{\Sigma_i + \Sigma_j}{2} \right\}^{-1} (\boldsymbol{m}_i - \boldsymbol{m}_j) + \frac{1}{2} \ln \left\{ \frac{|(\Sigma_i + \Sigma_j)/2|}{\sqrt{|\Sigma_i||\Sigma_j|}} \right\}$$
(7)

Dabei sind  $m_i$  bzw.  $m_j$  die Mittelwertsvektoren der Klassen i bzw. j und  $\Sigma_i$  bzw.  $\Sigma_j$  jeweils deren Varianz-Kovarianzmatrizen (berechnet aus den Trainingsdaten). (RICHARDS, 2006)

Zur Berechnung des paarweisen Jeffries-Matusita Abstands aus Trainingsdatensätzen der zugehörigen Klassen wird eine Funktion in R implementiert. Dabei werden die Resultate für alle Klassen aus den Trainingsdaten in einer Matrix zusammengefasst. Der Jeffries-Matusita Abstand kann Werte zwischen 0 und 2 annehmen. Dabei indiziert ein Abstand über 1.9 gute, zwischen 1.0 und 1.9 moderate und kleiner als 1.0 schlechte Klassenseparabilität (Ahmad, 2012). Bei schlechter Klassenseparabilität ist die Abgrenzung der Klassen gegeneinander bzw. die Auswahl der Trainingsdaten (nach Möglichkeit) zu überarbeiten.

#### 5.6 Clusteralgorithmen

Die Clusteranalyse (manchmal auch unüberwachte Klassifikation bzw. unsupervised Classification) dient dem Aufdecken von Ballungen (Clustern) in einer Datengrundlage. Entsprechende Algorithmen kommen hier zur Verfeinerung der Abgrenzung der Trainingsgebiete in den Ausgangsdaten und zur Extraktion von Zielstrukturen aus den Resultaten von Einklassenfall Klassifikationen zur Anwendung. In diesem Zusammenhang werden der k-Means als Schwerpunkt-basierter und der EM-Cluster als statistischer Algorithmus genutzt und anwendungsbezogen als Funktionen in R implementiert. Da diese Cluster-Algorithmen in der Regel nur lokale Maxima aufdecken (HASTIE, 2009), können mehrere Durchläufe mit verschieden Ausgangswerten erforderlich sein, um Tendenzen zur Clusterbildung in den Daten zuverlässig aufzudecken.

#### 5.6.1 k-Means Clustering

Der k-Means Cluster Algorithmus basiert auf der iterativen Zuordnung aller verfügbaren Datenpunkte (in Bezug auf Klassifikationsprobleme als Merkmalsvektoren interpretierbar) zu Clustern nach dem kleinsten Abstand zu deren initial zufällig gewählten Schwerpunkten. Die Koordinaten der Clusterzentren werden dabei in jedem Iterationsschritt aus den jeweils zugewiesenen Datenpunkten neu berechnet (MACQUEEN, 1967). Die Anwendung des k-Means Algorithmus resultiert in einer Partitionierung des Merkmalsraumes in Voronoi-Polygone (in zwei Dimensionen, siehe Abbildung 12; in 3 Dimensionen sind es Polyeder und in mehrdimensionalen Räumen entsprechende mehrdimensionale Erweiterungen).

Der k-Means Cluster Algorithmus wird hier für Rasterdatensätze mit beliebiger Anzahl an Merkmalen (Layer) mit jeweils identischer räumlicher Auflösung als Funktion in R implementiert. Dabei werden alle Primitive (im Fall von Rasterdaten: Bildpunkte/Pixel; bezogen auf die Problemstellung als räumliche Gebiete interpretierbar) des Ausgangsdatensatzes in beliebig vielen Iterationen auf eine frei wählbare Anzahl von Clustern mit initial zufällig gewählten Schwerpunkten verteilt. Das Resultat wird wiederum in einem Rasterdatensatz mit zu den Ausgangsdaten identischer räumlicher Auflösung bereitgestellt. Die Pixelwerte des Resultats entsprechen der Clusterzugehörigkeit der Bildpunkte (ganze Zahlen von 1 bis zur Anzahl der Cluster). Dabei wird zunächst der Zufallsgenerator mit einem zuvor festgesetzten Startwert (zur Gewährleistung der Reproduzierbarkeit der Ergebnisse und um den Durchlauf des Algorithmus mit unterschiedlichen Ausgangswerten zu ermöglichen) initialisiert und eine zufällige Stichprobe aus dem Ausgangsdatensatz genommen. Die Stichprobe dient der Bestimmung der initialen Clusterzentren und ist daher gerade vom Umfang gleich der Anzahl der Cluster.

Anschließend wird für alle Primitive nach Gleichung 8 mit der Dimension des Merkmalsraumes n, den

Koordinaten  $\mu_i$  des Clusterzentrums und den Werten  $x_i$  eines Primitivs für das *i*-te Merkmal jeweils der euklidische Abstand d zu den Clusterzentren im Merkmalsraum berechnet.

$$d = \sqrt{\sum_{i=1}^{n} (\mu_i - x_i)^2}$$
 (8)

Die Abstände aller Primitive zu den Clusterzentren im Merkmalsraum werden für jedes Cluster jeweils in einem Rasterdatensatz (Abstandsraster) zusammengefasst. Darauf folgt die Zuordnung der Primitive zu den Clustern nach minimalem Abstand vom Clusterzentrum. Im resultierenden Raster der (räumlichen) Clusterverteilung ist jedem Primitiv nach der Clusterzugehörigkeit eine ganze Zahl zugeordnet. Da diese Zuordnung der Primitive zu Clustern nach dem minimalen Abstand vom Clusterzentrum gerade der (überwachten) euklidischen/spektralen Distanz Klassifikation (Albertz, 2009) entspricht, wird der Teilprozess für eventuelle spätere Anwendungen und zur Modularisierung in einer separaten Funktion implementiert.

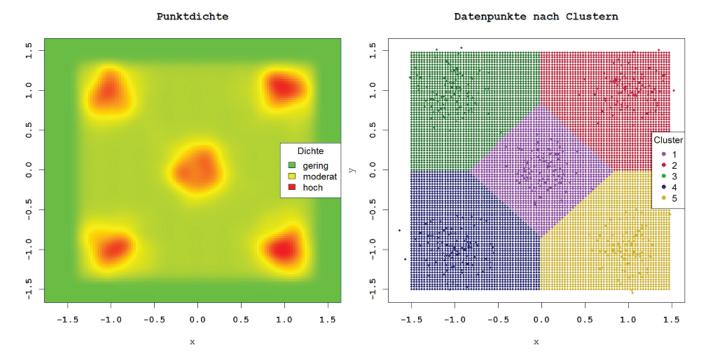


Abbildung 12: Beisiel für k-Means Clustering zur Verdeutlichung der Partitionierung des Merkmalsraumes in Voronoi-Polygone: Eine Punktwolke bestehend aus einem regelmäßigen quadratischen Punktraster über  $-1.5 \le x \le 1.5$  und  $-1.5 \le y \le 1.5$ , überlagert mit 5 bivariat normalverteilten Punktwolken von jeweils 100 Punkten mit unterschiedlichen Mittelwerten und einer Standardabweichung von  $\sigma = 0.2$  jeweils für beide Variablen (Dichteverteilung links) wird mit dem k-Means Algorithmus (Implementierung aus dem R-Paket "stats" (R CORE TEAM AND CONTRIBUTORS, 2019)) geclustert (rechts: Punkte eingefärbt nach Clusterzugehörigkeit)

Ist im Anschluss die vorgesehene Anzahl an Iterationen erreicht, erfolgt die Ausgabe der Clusterverteilung als Resultat des k-Means Clustering. Andernfalls werden die Koordinaten der Clusterzentren jeweils als Schwerpunkte der Cluster im Merkmalsraum auf Grundlage der resultierenden Clusterverteilung und den Ausgangsdaten nach Gleichung 9 mit dem Schwerpunktvektor  $\mu$ , der Dimension n des Merkmalsraumes, dem Einheitsvektor  $e_i$  für das i-te Merkmal, der Anzahl m der Primitive eines Clusters und dem Wert  $x_{ij}$  des j-ten Primitivs für das i-te Merkmal neu berechnet und die Zuteilung der Primitive auf die Cluster

mit den neuen Zentren wiederholt. Der Algorithmus ist in Abbildung 13 zusammengefasst.

$$\mu = \sum_{i=1}^{n} e_i \frac{1}{m} \sum_{j=1}^{m} x_{ij}$$
 (9)

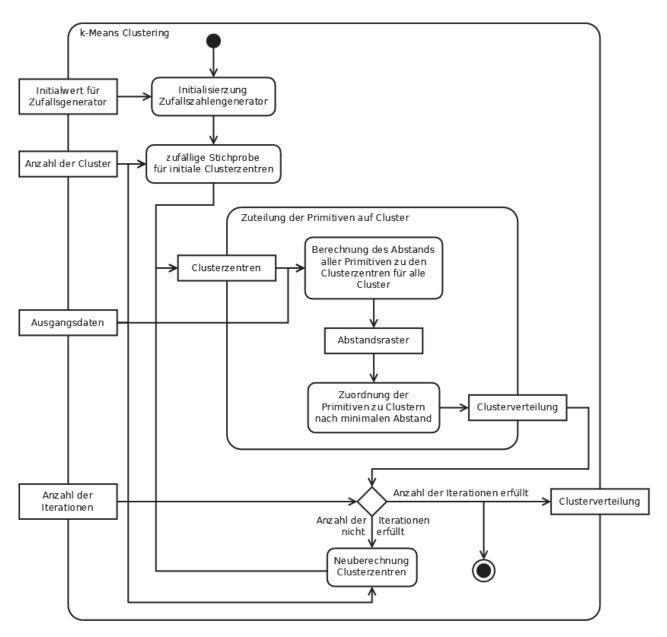


Abbildung 13: Workflow k-Means Clustering

#### 5.6.2 EM Clustering

Der Expectation-Maximization (EM; etwa: "Erwartungs-Maximierung") Algorithmus dient der iterativen Bestimmung von Maximum Likelihood (etwa: "höchstwahrscheinlichen") Schätzwerten für statistische Modellparameter und latente (unbeobachtete) Variablen, die nur indirekt durch die erfassten Daten beobachtet werden. Dabei werden in jeder Iteration jeweils in einem Expectation-Schritt (Erwartungsschritt) "erwartete" bedingte Wahrscheinlichkeitsdichteverteilungen für die unbeobachteten Variablen gegeben der Messwerte und aktuellen Schätzwerte der Parameter berechnet, die im anschließenden Maximization-Schritt (Maximierungsschritt) genutzt werden, um neue Schätzwerte für die Modellparameter zu bestimmen (DEMPSTER, 1977).

Der EM-Algorithmus kann als Cluster Algorithmus genutzt werden, wenn die latenten Variablen als Indikator für die Zugehörigkeit zu Clustern mit entsprechenden statistischen Modellen interpretiert werden. Als statistisches Modell für die Cluster kommen häufig Mischmodelle aus Gauß'schen Normalverteilungen (Gaussian Mixture Models) zur Anwendung (HASTIE, 2009). Bezogen auf die mehrdimensionale Problemstellung lassen sich die M Cluster statistisch mithilfe multivariater Gauß'scher Wahrscheinlichkeitsdichteverteilungen  $\phi_{\theta_j}(\boldsymbol{x})$  mit dem Mittelwertsvektor  $\boldsymbol{\mu}_j$  und der Varianz-Kovarianz-Matrix  $\boldsymbol{\Sigma}_j$  als Parameter  $\boldsymbol{\theta}_j$  für das j-te Cluster und dem Merkmalsvektor  $\boldsymbol{x}$  eines Primitivs modellieren. Daraus folgt für die vollständige Likelihood-Funktion (ZEIDLER, 2013)  $p(\boldsymbol{x}, z|\boldsymbol{\theta})$  aller N Primitive:

$$p(\boldsymbol{x}, z | \boldsymbol{\theta}) = \prod_{i=1}^{N} \prod_{j=1}^{M} \delta_{z_i j} \pi_j \phi_{\boldsymbol{\theta}_j}(\boldsymbol{x}_i)$$
(10)

mit der latenten Variablen z für die Clusterzugehörigkeit (ganze Zahl  $z_i$  von 1 bis M (Anzahl der Cluster) zusätzlich für jedes Primitiv), den Parametern  $\boldsymbol{\theta}$  der multivariaten Gauß'schen Normalverteilungen, dem Kroneckersymbol  $\delta_{z_ij}$  (1 für  $z_i=j$  (Zugehörigkeit des i-ten Primitivs mit dem Merkmalsvektor  $\boldsymbol{x}_i$  zum j-ten Cluster), ansonsten 0) und der Priori-Wahrscheinlichkeit  $\pi_j$  für das j-te Cluster mit der multivariaten Gauß'schen Wahrscheinlichkeitsdichteverteilung  $\phi_{\boldsymbol{\theta}_j}$ . Entsprechend folgt für die logarithmierte vollständige Likelihood-Funktion  $\ell(\boldsymbol{\theta}; \boldsymbol{x}, z)$  (log likelihood):

$$\ell(\boldsymbol{\theta}; \boldsymbol{x}, z) = \sum_{i=1}^{N} \sum_{j=1}^{M} \delta_{z_i j} \left[ \ln \pi_j - \frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln|\Sigma_j| - \frac{1}{2} (\boldsymbol{x}_i - \boldsymbol{\mu}_j)^{\mathrm{T}} \Sigma_j^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_j) \right]$$
(11)

mit der Dimension des Merkmalsraumes n. Das EM-Clustering zielt auf die Bestimmung der latenten Variablen z und Modellparameter  $\boldsymbol{\theta}$  unter der Bedingung, dass Gleichung 11 maximal wird. Die Maximierung der Likelihood-Funktion wäre wegen der Summe über alle Primitive ohne die Einführung latenter Variablen sehr aufwändig. Durch die Erweiterung um latente Variablen lassen sich die Modellparameter der einzelnen multivariaten Gauß'schen Normalverteilungen iterativ als gewichtete Mittelwertsvektoren und Varianz-Kovarianz-Matrizen auf der Grundlage (zufälliger) initialer Schätzwerte für die latenten Variablen der Primitive bestimmen. Dabei werden in jeder Iteration zunächst die M Gewichte  $\gamma_{ij}$  für jedes Primitiv aus der bedingten Wahrscheinlichkeitsdichteverteilung für die latente Variable z gegeben dem zugehörigen Merkmalsvektor  $\boldsymbol{x}_i$  und den geschätzten Parametern  $\boldsymbol{\theta}$  über den Satz von Bayes berechnet (Expectation-Schritt):

$$\gamma_{ij} = p(z_i = j | \boldsymbol{x}_i, \boldsymbol{\theta}) = \frac{\pi_j \phi_{\boldsymbol{\theta}_j}(\boldsymbol{x}_i)}{\sum_{j=1}^{M} \pi_j \phi_{\boldsymbol{\theta}_j}(\boldsymbol{x}_i)}$$
(12)

und anschließend die gewichteten Mittelwertsvektoren  $\mu_j$  und Varianz-Kovarianz-Matrizen  $\Sigma_j$  als neue Schätzwerte für die Parameter  $\theta$  des statistischen Modells sowie die neuen Priori-Wahrscheinlichkeiten  $\pi_j$  für die Cluster als Mittelwerte der Gewichte  $\gamma_{ij}$  (Maximization-Schritt):

$$\boldsymbol{\mu}_{j} = \frac{\sum_{i=1}^{N} \gamma_{ij} \boldsymbol{x}_{i}}{\sum_{i=1}^{N} \gamma_{ij}}$$

$$(13)$$

$$\Sigma_{j} = \frac{\sum_{i=1}^{N} \gamma_{ij} (\boldsymbol{x}_{i} - \boldsymbol{\mu}_{j}) (\boldsymbol{x}_{i} - \boldsymbol{\mu}_{j})^{\mathrm{T}}}{\sum_{i=1}^{N} \gamma_{ij}}$$
(14)

$$\pi_j = \frac{1}{N} \sum_{i=1}^N \gamma_{ij} \tag{15}$$

Dieser EM-Clustering Algorithmus mit multivariaten Gauß'schen Normalverteilungen zur statistischen Modellierung der Cluster wird hier analog zum k-Means Cluster Algorithmus (siehe Abschnitt 5.6.1 "k-Means Clustering") für Rasterdatensätze mit beliebiger Anzahl an Merkmalen mit jeweils identischer räumlicher Auflösung, zur Aufteilung aller Primitive (bzw. Pixel) auf eine frei wählbare Anzahl an Clustern in beliebig vielen Iterationen als Funktion in R implementiert. Eine Zusammenfassung des Algorithmus ist in Abbildung 14 dargestellt.

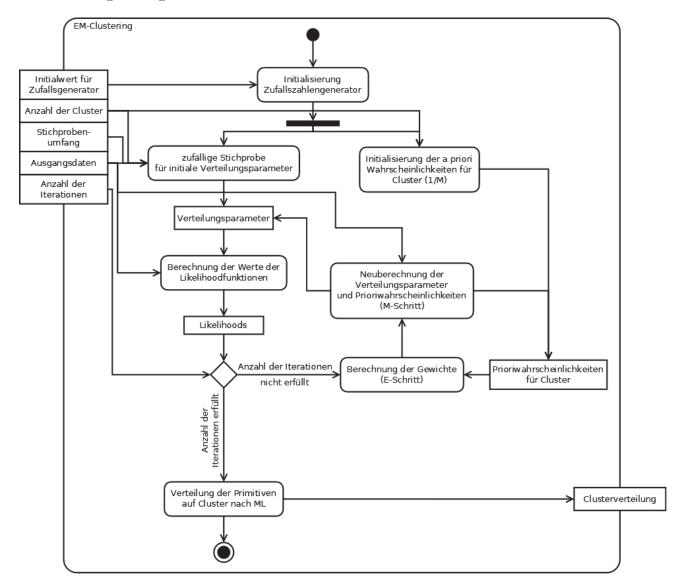


Abbildung 14: Workflow EM Clustering

Dabei werden zunächst die initialen Verteilungsparameter in Analogie zur Implementierung des k-Means Cluster Algorithmus anhand einer zufälligen Stichprobe aus den Ausgangsdaten bestimmt. Der Stichprobenumfang bezieht sich jeweils auf die einzelnen Stichproben zur Ermittlung der jeweiligen Verteilungsparameter, ist beliebig variierbar und findet sich daher unter den Eingangsparametern. Die Priori-Wahrscheinlichkeiten für die Cluster werden als Reziproke der Anzahl der Cluster (1/M - die Zugehörigkeit eines beliebigen Primitivs zu einem der Cluster ist für alle Cluster gleich wahrscheinlich) initialisiert. Anhand der Verteilungsparameter werden die Werte (Likelihoods) der Likelihood-Funktionen (Wahrscheinlichkeitsdichteverteilungen für die einzelnen Cluster) für die Ausgangsdaten berechnet. Da sich diese Berechnung analog auch als Teilschritt bei der Maximum Likelihood Klassifikation (Albertz,

2009) wiederfindet, wird sie für eventuelle spätere Anwendungen und zur Modularisierung in einer separaten Funktion implementiert. Ist im Anschluss die vorgesehene Anzahl an Iterationen erreicht, so werden die Primitive der Ausgangsdaten jeweils nach dem höchsten Wert der Likelihood-Funktionen (Maximum Likelihood (ML) Methode) auf die zugehörigen Cluster verteilt und die resultierende Clusterverteilung als Raster mit zu den Ausgangsdaten identischer räumlicher Auflösung bereitgestellt. Andernfalls wird der EM Algorithmus mit der Neuberechnung der Gewichte aus den Werten der Likelihood-Funktionen und den Priori-Wahrscheinlichkeiten der Cluster (E-Schritt) gefolgt von der Anpassung der Verteilungsparameter und Priori-Wahrscheinlichkeiten (M-Schritt) fortgesetzt.

#### 5.7 Klassifikationsalgorithmen

Klassifizierungsalgorithmen dienen der (überwachten) Zuteilung von Primitiven anhand ihrer Merkmale auf vordefinierte Klassen. Die Klassifikation kann auf der Grundlage unterschiedlicher Kriterien - den Klassifikatoren - erfolgen. Mit dem Klassifikator wird ein Modell definiert, nach dem die Primitive auf die Klassen verteilt werden. Die jeweiligen Modellparameter werden vor der Klassifikation aller Primitive in einer Trainingsphase aus einer Teilmenge der Datengrundlage von Primitiven mit bekannter Klassenzugehörigkeit - den Trainingsdaten - bestimmt. Dieser grundlegende Ablauf lässt sich auf die Mehrzahl der unterschiedlichen Klassifizierungsalgorithmen übertragen und ist in Abbildung 15 zusammengefasst. Die im Folgenden beschriebenen Klassifizierungsalgorithmen werden nach diesem Schema zur Klassifikation von Rasterdatensätzen mit einer beliebigen Anzahl an Merkmalen (Layer) identischer räumlicher Auflösung als Funktionen in R implementiert. Dabei werden die Trainingsdaten als Polygone mit einem Attribut für die Klassenzugehörigkeit in einem Vektordatensatz zusammengefasst. Klassifikatoren werden für den Mehrklassenfall so implementiert, dass jedem Primitiv (Pixel) in den Ergebnissen genau eine Klasse aus den Trainingsdaten zugeordnet wird und keine unklassifizierten Primitive vorkommen.

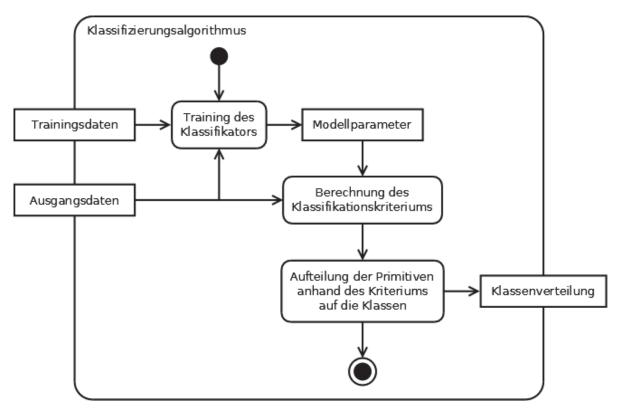


Abbildung 15: grundlegender Workflow für Klassifizierungsalgorithmen

Bei der Fernerkundung von Pflanzen ist häufig nur eine bestimmte Spezies von Interesse, sodass die Über-

tagung von Klassifikationsalgorithmen auf den Einklassenfall angebracht ist (Khan, 2014). Die Anwendung von Mehrklassenfall- oder binären Klassifikationsalgorithmen bei der Identifikation einzelner Spezies kann zu deren zuverlässiger Abgrenzung gegen andere Vegetationsformen umfangreiche Trainingsdatensätze erfordern und eine zusätzliche Fehlerquelle darstellen. Daher werden für die Euklidische (Spektrale) Distanz -, die Maximum Likelihood - und die korrelationskoeffizienten-basierte Klassifikation zusätzliche Funktionen für den Einklassenfall in R implementiert. Da keine Abwägung gegen andere Klassen möglich ist (Wegfall des letzten Bearbeitungsschritts in Bezug zum allgemeinen Algorithmus aus Abbildung 15), entsprechen die Resultate dieser Einklassenfall Klassifikationen gerade den räumlichen Verteilungen der jeweiligen Klassifikationskriterien (der euklidischen (spektralen) Distanz, der Werte der Likelihood-Funktion und der Korrelationskoeffizienten). Dabei lassen sich "Treffer" aus diesen Verteilungen mithilfe von Clusteranalysen als Anhäufungen von Primitiven mit optimalen Werten für das entsprechende Kriterium extrahieren, um für jedes Primitiv zu einer binären Entscheidung (Anwesenheit/Abwesenheit einer bestimmten Spezies) zu gelangen.

#### 5.7.1 Euklidische (Spektrale) Distanz Klassifikation

Bei der Euklidischen (Spektralen) Distanz Klassifikation wird die Klassifikation der Primitive anhand ihres euklidischen Abstands zu den, aus den Trainingsdaten bestimmten, Schwerpunkten der Klassen im Merkmalsraum als Klassifikationskriterium vorgenommen. Die Primitive werden der Klasse zugeordnet, für die ihr euklidischer Abstand zum jeweiligen Schwerpunkt im Merkmalsraum minimal wird (Albertz, 2009). Damit lässt sich die Euklidische (Spektrale) Distanz Klassifikation als Analogie zum k-Means Clustering (siehe Abschnitt 5.6.1 "k-Means Clustering") betrachten, wo die Verteilung der Primitive auf die Cluster ebenfalls nach dem minimalen euklidischen Abstand zu den Clusterschwerpunkten im Merkmalsraum erfolgt. Entsprechend kann der euklidische Abstand d der Primitive mit den Werten  $x_i$  für das i-te Merkmal zu den Klassenschwerpunkten mit den Koordinaten  $\mu_i$  in einem n-dimensionalen Merkmalsraum analog nach Gleichung 8 bestimmt werden. Die Schwerpunktvektoren  $\mu$  der Klassen werden in der Trainingsphase jeweils nach Gleichung 9 mit dem Einheitsvektor  $e_i$  für das i-te Merkmal und den zugehörigen Werten  $x_{ij}$  von Primitiven mit bekannter Klassenzugehörigkeit aus einem Trainingsdatensatz vom Umfang m berechnet. Im Gegensatz zum k-Means Clustering erfolgt keine iterative Neubestimmung der Klassenschwerpunkte und -Zugehörigkeit der Primitive. Wie das k-Means Clustering führt die Euklidische (Spektrale) Distanz Klassifikation zu einer Partitionierung des Merkmalsraumes in Voronoi-Polygone (bzw. deren mehrdimensionalen Erweiterungen; siehe Abbildung 12).

#### 5.7.2 Maximum Likelihood Klassifikation (MLC)

Die Maximum Likelihood Klassifikation ist eine weit verbreitete statistische Klassifikationsmethode (Albertz, 2009). Ähnlich wie das EM-Clustering basiert das Verfahren auf der Optimierung einer Likelihood-Funktion. Die Klassen werden mithilfe von statistischen Verteilungsfunktionen modelliert, deren Parameter in der Trainingsphase aus einem Trainingsdatensatz geschätzt werden. Dabei wird jede Klasse in der Regel jeweils mit einer Verteilungsfunktion beschrieben. Zur Herleitung der Likelihood-Funktion wird vom Satz von Bayes ausgegangen:

$$p(i|\mathbf{x}) = \frac{p(\mathbf{x}|i)p(i)}{p(\mathbf{x})}$$
(16)

Dabei ist  $p(i|\mathbf{x})$  die bedingte Wahrscheinlichkeitsdichteverteilung (Posteriori Verteilung), dass ein Primitiv gegeben dem Merkmalsvektor  $\mathbf{x}$  zur i-ten Klasse gehört;  $p(\mathbf{x}|i)$  die bedingte Wahrscheinlichkeitsdichteverteilung für die Beobachtung des Merkmalsvektors  $\mathbf{x}$  unter der Bedingung, dass das Primitiv zur i-ten Klasse gehört; p(i) die Priori Wahrscheinlichkeit für die i-te Klasse und  $p(\mathbf{x})$  die Wahrscheinlichkeit für die Beobachtung des Merkmalsvektors  $\mathbf{x}$ . Im Rahmen der Maximum Likelihood Klassifikation wird die Wahrscheinlichkeit  $p(\mathbf{x})$  für gewöhnlich vernachlässigt oder als Normierungskonstante genutzt (Ahmad, 2012). Andernfalls kann  $p(\mathbf{x})$  nach Gleichung 17 über den Satz von der totalen Wahrscheinlichkeit aus den Wahrscheinlichkeitsdichteverteilungen und Priori Wahrscheinlichkeiten der M Klassen bestimmt werden

und das Verfahren geht in eine Maximum A Posteriori (MAP) Klassifikation über (BASSETT, 2019).

$$p(\mathbf{x}) = \sum_{i=1}^{M} p(\mathbf{x}|i)p(i)$$
(17)

Zur Modellierung der Klassen kommen häufig multivariate Gauß'sche Normalverteilungen zur Anwendung, sodass für die logarithmierten Likelihood-Funktionen  $\ell(i; \boldsymbol{x})$  nach Gleichung 16

$$\ell(i; \boldsymbol{x}) = \ln p(i) - \frac{n}{2} \ln (2\pi) - \frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu}_i)^{\mathrm{T}} \Sigma_i^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_i)$$
(18)

mit der Dimension n des Merkmalsraumes folgt. Die Parameter  $\mu_i$  (Mittelwertsvektor) und  $\Sigma_i$  (Varianz-Kovarianz-Matrix) der multivariaten Gauß'schen Normalverteilungen werden in der Trainingsphase bestimmt. Für jede Klasse wird jeweils eine Likelihood-Funktion mit den zugehörigen Parametern  $\mu_i$  und  $\Sigma_i$ aufgestellt, deren Werte für jedes Primitiv als Klassifikationskriterium berechnet werden. Jedes Primitiv wird der Klasse zugeordnet, für welche die Likelihood-Funktion  $\ell(i;x)$  maximal wird. Bei der Maximum Likelihood Klassifikation werden die Klassen in einem zweidimensionalen Merkmalsraum (bzw. zweidimensionalen Projektionen mehrdimensionaler Merkmalsräume) durch Kegelschnitte (Punkte, Geraden, Hyperbeln, Parabeln, Kreise und Ellipsen; siehe Abbildung 16) begrenzt (Ahmad, 2012). In Abbildung 17 ist eine Teilmenge der Datenpunkte der Primitive (Pixel bzw. Gebiete) aus der Aufnahme des MSI von Sentinel 2 vom 16.2.2019 für das Untersuchungsgebiet bei Kraase (ESA, 2019; siehe Abbildung 9 für einen Ausschnitt) in zweidimensionalen Projektionen des durch die ausgewählten Merkmale (vergleiche Beispiel aus Abschnitt 5.4 "Merkmalsreduktion" und Abbildung 9) aufgespannten Merkmalsraumes dargestellt. Dabei sind in jedem Diagramm jeweils zwei Merkmale paarweise gegeneinander abgetragen, sodass die Projektionsebenen im mehrdimensionalen Merkmalsraum senkrecht aufeinander stehen. Die Datenpunkte sind nach ihrer Klassenzugehörigkeit aus einer beispielhaften Maximum Likelihood Klassifikation eingefärbt.

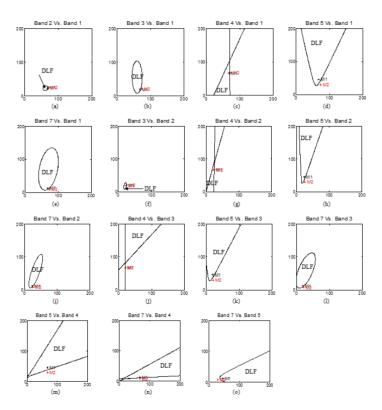


Abbildung 16: Begrenzungen zwischen jeweils zwei Klassen in zweidimensionalen Projektionen eines mehrdimensionalen Merkmalsraumes bei der Maximum Likelihood Klassifikation (Ahmad, 2012)

#### Datenpunkte nach Klassen

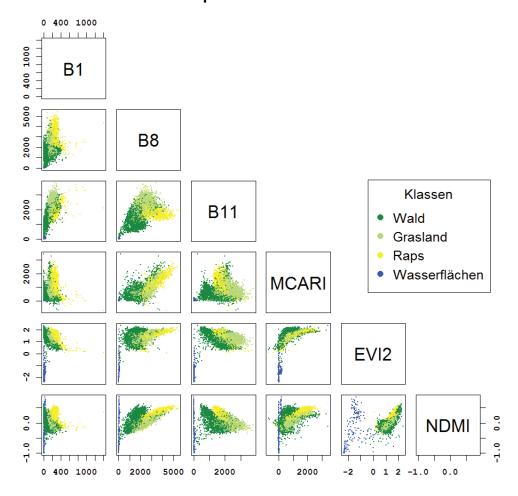


Abbildung 17: Beispiel für das Resultat einer Maximum Likelihood Klassifikation im Merkmalsraum

#### 5.7.3 Mahalanobis Abstand Klassifikation

Der Mahalanobis Abstand ist ein statistisches Distanzmaß zur Bestimmung des Abstands zwischen einem Punkt und einer Gauß'schen Normalverteilung bzw. zwischen zwei Gauß'schen Normalverteilungen mit gemeinsamer Varianz-Kovarianz-Matrix und unterschiedlichen Mittelwerten in einem multidimensionalen Raum (Mahalanobis, 1936). Bei der Klassifikation von Fernerkundungsdaten kann der quadrierte Mahalanobis Abstand  $d_M^2$  des Datenpunktes eines Primitivs mit dem Merkmalsvektor  $\boldsymbol{x}$  zu einer multivariaten Gauß'schen Normalverteilung mit den Parametern  $\boldsymbol{\mu}_i$  (Mittelwertsvektor) und  $\boldsymbol{\Sigma}_i$  (Varianz-Kovarianz-Matrix) für die i-te Klasse im Merkmalsraum als Klassifikationskriterium genutzt werden (Albertz, 2009):

$$d_M^2 = (\boldsymbol{x} - \boldsymbol{\mu}_i)^{\mathrm{T}} \Sigma_i^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_i)$$
(19)

Jedes Primitiv wird der Klasse zugeordnet, für die Gleichung 19 minimal wird. In Analogie zur Maximum Likelihood Klassifikation werden in der Trainingsphase die Parameter  $\mu_i$  und  $\Sigma_i$  der multivariaten Gauß'schen Normalverteilungen zur Modellierung der Klassen aus den Trainingsdaten bestimmt. Der Mahalanobis Abstand als Klassifikationskriterium unterscheidet sich von der logarithmierten Likelihood Funktion bei der MLC nur in der zusätzlichen Berücksichtigung der Priori Verteilung p(i), den Summanden  $-\frac{n}{2}\ln(2\pi)$  und  $-\frac{1}{2}\ln|\Sigma_i|$  sowie dem Faktor  $-\frac{1}{2}$  (vergleiche Gleichung 18 und 19); wobei die Priori Verteilung häufig vernachlässigt wird und der Term  $-\frac{n}{2}ln(2\pi)$  bei gleichbleibender Dimensionalität des

Merkmalsraumes für alle Klassen konstant ist und daher keine Auswirkungen auf das Klassifikationsergebnis hat.

Zur Herleitung der Formen der Klassengrenzen kann davon ausgegangen werden, dass sie im Merkmalsraum genau dort auftreten werden, wo das Klassifikationskriterium für (mindestens) zwei Klassen identisch ist. Daraus folgt für die Klassengrenzen zwischen einer *i*-ten und *j*-ten Klasse:

$$(\boldsymbol{x} - \boldsymbol{\mu}_i)^{\mathrm{T}} \Sigma_i^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_i) = (\boldsymbol{x} - \boldsymbol{\mu}_i)^{\mathrm{T}} \Sigma_i^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_i)$$
(20)

beziehungsweise

$$(x - \mu_i)^{\mathrm{T}} \Sigma_i^{-1} (x - \mu_i) - (x - \mu_j)^{\mathrm{T}} \Sigma_j^{-1} (x - \mu_j) = 0$$
(21)

Da es sich bei Gleichung 21 um eine mehrdimensionale quadratische Gleichung handelt (BRONSTEIN, 2012), werden die Klassen wie bei der MLC in 2 Dimensionen durch Kurven 2. Ordnung (Kegelschnitte), in 3 Dimensionen durch Flächen 2. Ordnung und in mehr Dimensionen durch entsprechende mehrdimensionale Erweiterungen begrenzt (AHMAD, 2012).

# 5.7.4 Korrelationskoeffizienten-basierte Klassifikation (Spectral Correlation Mapper - SCM)

Durch seine Eigenschaft die Wahrscheinlichkeit wiederzugeben, mit der die Beobachtungen zweier Merkmale auf einer Geraden liegen (Bronstein, 2012), lässt sich der Pearson'sche Korrelationskoeffizient (siehe Gleichung 1) als Kriterium für einen Klassifikationsalgorithmus instrumentalisieren. Dabei werden statt allen Beobachtungen aus zwei Merkmalen die Merkmalsvektoren zweier Primitive gegeneinander abgetragen (bzw. Merkmalsvektoren als Merkmale bei einer Korrelationsanalyse interpretiert). Für zwei identische Primitive würden die Beobachtungen auf einer Geraden liegen. In Abbildung 18 sind die Merkmalsvektoren zweier Bildpunkte (jeweils einer auf dem kleinen und einer auf dem großen Varchentiner See) aus einer Beispielaufnahme des MSI von Sentinel 2 vom 26.7.2019 (ESA, 2019; vergleiche Abbildung 3) für das Untersuchungsgebiet bei Kraase gegeneinander abgetragen. Darüber hinaus ist in dem Diagramm auch die resultierende Regressionsgerade dargestellt. Beide Positionen befinden sich auf Wasserflächen. Die Merkmalsvektoren lassen sich gegeneinander abgetragen gut durch eine Gerade approximieren. Für die Beobachtungen ergibt sich ein Pearson'scher Korrelationskoeffizient r von 0.98.

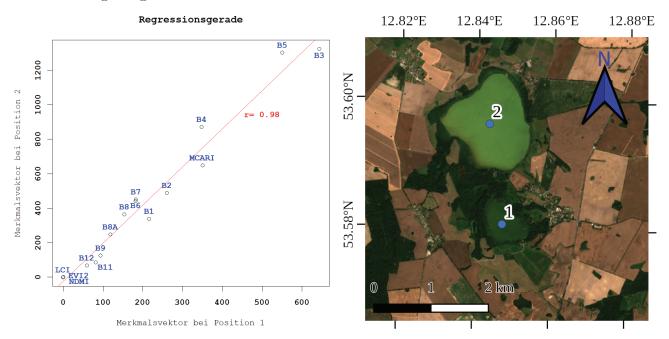


Abbildung 18: Regressionsgerade für die Merkmalsvektoren zweier Bildpunkte aus einer Beispielaufnahme des MSI vom 26.7.2019 für das Untersuchungsgebiet bei Kraase (links) mit Beschriftungen für die einzelnen Merkmale und den Positionen der Bildpunkte in der Aufnahme (rechts)

Bei der Klassifikation mit dem Pearson'schen Korrelationskoeffizienten als Kriterium wird für jede Klasse in der Trainingsphase ein Referenz-Merkmalsvektor (in der Regel der Schwerpunktvektor) aus den Trainingsdaten bestimmt und die Korrelationen der Merkmalsvektoren der Primitive mit den Referenzvektoren berechnet. Jedes Primitiv wird der Klasse zugeordnet deren Referenzvektor am stärksten mit dem zugehörigen Merkmalsvektor korreliert ist. Der resultierende Klassifikator ähnelt dem Spectral Angle Mapper (DE CARVALHO JÚNIOR, 2000).

Zur Herleitung der Klassengrenzen wird Gleichung 1 auf Klassifikationsprobleme übertragen und vektoriell interpretiert:

$$r_{\boldsymbol{x}\boldsymbol{\mu}_i} = \frac{(\boldsymbol{x} - \overline{\boldsymbol{x}})^{\mathrm{T}} (\boldsymbol{\mu}_i - \overline{\boldsymbol{\mu}_i})}{|\boldsymbol{x} - \overline{\boldsymbol{x}}| |\boldsymbol{\mu}_i - \overline{\boldsymbol{\mu}_i}|}$$
(22)

Dabei ist  $r_{\boldsymbol{x}\boldsymbol{\mu}_i}$  der Pearson'sche Korrelationskoeffizient zwischen dem Merkmalsvektor  $\boldsymbol{x}$  eines Primitivs und dem Schwerpunktvektor  $\mu_i$  der i-ten Klasse und die Elemente der Vektoren  $\overline{\boldsymbol{x}}$  bzw.  $\overline{\boldsymbol{\mu}_i}$  sind die Mittelwerte der Koordinaten von  $\boldsymbol{x}$  bzw.  $\boldsymbol{\mu}_i$ . An den Klassengrenzen sind die Korrelationskoeffizienten der Merkmalsvektoren mit den Schwerpunktvektoren für (mindestens) zwei Klassen identisch. Daraus folgt für die Grenze zwischen einer i-ten und j-ten Klasse:

$$\frac{(\boldsymbol{x} - \overline{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{\mu}_{i} - \overline{\boldsymbol{\mu}_{i}})}{|\boldsymbol{x} - \overline{\boldsymbol{x}}||\boldsymbol{\mu}_{i} - \overline{\boldsymbol{\mu}_{i}}|} = \frac{(\boldsymbol{x} - \overline{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{\mu}_{j} - \overline{\boldsymbol{\mu}_{j}})}{|\boldsymbol{x} - \overline{\boldsymbol{x}}||\boldsymbol{\mu}_{j} - \overline{\boldsymbol{\mu}_{j}}|}$$
(23)

bzw.

$$(\boldsymbol{x} - \overline{\boldsymbol{x}})^{\mathrm{T}} \left( \frac{1}{|\boldsymbol{\mu}_i - \overline{\boldsymbol{\mu}_i}|} (\boldsymbol{\mu}_i - \overline{\boldsymbol{\mu}_i}) - \frac{1}{|\boldsymbol{\mu}_j - \overline{\boldsymbol{\mu}_j}|} (\boldsymbol{\mu}_j - \overline{\boldsymbol{\mu}_j}) \right) = 0$$
(24)

Da es sich bei Gleichung 24 um eine Gleichung ersten Grades handelt (BRONSTEIN, 2012), werden die Klassen durch Ebenen begrenzt. Bei zweidimensionalen Projektionen höherdimensionaler Merkmalsräume können die Grenzen zwischen den Klassen die Form von Geraden annehmen. Da der Pearson'sche Korrelationskoeffizient für Paare beliebiger Merkmalsvektoren in zweidimensionalen Merkmalsräumen stets 1 ergibt (zwei Punkte liegen immer mit einer Wahrscheinlichkeit von 1 auf einer Geraden), lässt sich die SCM Klassifikation nicht in zweidimensionalen Merkmalsräumen anwenden. In Abbildung 19 ist in Analogie zu dem Beispiel aus Abbildung 17 eine Teilmenge der Datenpunkte der Primitive aus der Aufnahme des MSI von Sentinel 2 vom 16.2.2019 für das Untersuchungsgebiet bei Kraase (ESA, 2019; siehe Abbildung 9 für einen Ausschnitt) in zweidimensionalen Projektionen des durch die Merkmalsauswahl (vergleiche Beispiel aus Abschnitt 5.4 "Merkmalsreduktion" und Abbildung 9) aufgespannten Merkmalsraumes dargestellt. Die Datenpunkte sind hier nach ihrer Klassenzugehörigkeit aus einer beispielhaften SCM Klassifikation eingefärbt. Für die Klassifikation werden dabei die selben Trainingsdaten und die selbe Merkmalsauswahl verwendet, wie in dem Beispiel aus Abbildung 17.

#### Datenpunkte nach Klassen

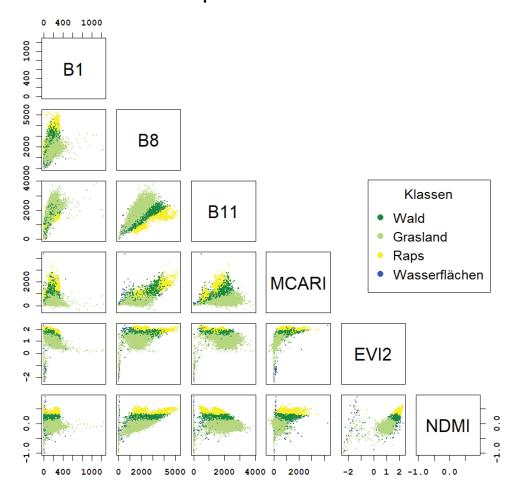


Abbildung 19: Beispiel für das Resultat einer SCM Klassifikation im Merkmalsraum

#### 5.7.5 Klassifikation mithilfe der kanonischen Korrelationsanalyse

Die kanonische Korrelationsanalyse ist ein Verfahren der multivariaten Statistik zur Untersuchung von Zusammenhängen zwischen zwei Mengen von Zufallsvektoren in einem Datensatz. Es beruht auf der Maximierung der Korrelation zwischen Linearkombinationen der beiden Sätze von Zufallsvektoren mit Gewichtsvektoren - den kanonischen Gewichten (Hoteling, 1936). Die kanonische Korrelationsanalyse lässt sich auch auf die Klassifikation von Fernerkundungsdaten übertragen (Lee, 1999). Die beiden Gruppen von Zufallsvektoren sind dabei zum einen die Merkmalsvektoren der zu klassifizierenden Primitiven und zum anderen Referenz-Merkmalsvektoren für die Klassen. Da die Analyse Pixelweise durchgeführt wird, umfasst die erste Menge lediglich einen Merkmalsvektor. Als Referenz-Merkmalsvektoren können in der Trainingsphase die jeweiligen Schwerpunktvektoren der einzelnen Klassen aus den Trainingsdaten bestimmt werden, sodass der Umfang der zweiten Menge gerade der Anzahl der Klassen M entspricht. Beide Gruppen werden in Matrizen zusammengefasst. Folglich ist die Matrix  $X^{(1)}$  für die Merkmalsvektoren der Primitive vom Typ  $d \times 1$  und  $X^{(2)}$  für die Referenz-Merkmalsvektoren der Klassen vom Typ  $d \times M$  mit der Dimension d des Merkmalsraumes. Das Ziel der kanonischen Korrelationsanalyse ist die Bestimmung von Gewichtsvektoren  $w_{(1)}$  und  $w_{(2)}$  - den kanonischen Gewichten, sodass die Korrelation zwischen den Linearkombinationen

$$z_{(1)} = X^{(1)} \boldsymbol{w}_{(1)} \tag{25}$$

und

$$\mathbf{z}_{(2)} = X^{(2)} \mathbf{w}_{(2)} \tag{26}$$

maximal wird. Diese Aufgabe entspricht der Minimierung des Winkels zwischen den Linearkombinationen  $\boldsymbol{z}_{(1)}$  und  $\boldsymbol{z}_{(2)}$  und die kanonischen Gewichte können als Lösungen von Eigenwertproblemen bestimmt werden. Insbesondere ergeben sich bei Normierung (Mittelwert subtrahieren) und Standardisierung (durch Standardabweichung dividieren) der Zufallsvektoren (Spalten von  $X^{(1)}$  und  $X^{(2)}$ ) die Gewichtsvektoren  $\boldsymbol{w}_{(2)}$  als Eigenvektoren der Matrix  $\boldsymbol{\Sigma}_{(2)(2)}^{-1}\boldsymbol{\Sigma}_{(2)(1)}^{-1}\boldsymbol{\Sigma}_{(1)(1)}^{-1}\boldsymbol{\Sigma}_{(1)(2)}^{-1}$  mit den Varianz-Kovarianz-Matrizen

$$\Sigma_{(1)(1)} = \frac{1}{d-1} X_{(1)}^{\mathrm{T}} X_{(1)}, \tag{27}$$

$$\Sigma_{(2)(2)} = \frac{1}{d-1} X_{(2)}^{\mathrm{T}} X_{(2)}$$
 (28)

und

$$\Sigma_{(1)(2)} = \Sigma_{(2)(1)}^{\mathrm{T}} = \frac{1}{d-1} X_{(1)}^{\mathrm{T}} X_{(2)}$$
(29)

(Uurtio, 2017). Das Primitiv mit dem Merkmalsvektor  $X^{(1)}$  wird der Klasse zugeordnet, auf deren Referenz-Merkmalsvektor in  $X^{(2)}$  das größte Gewicht aus dem Gewichtsvektor  $\boldsymbol{w}_{(2)}$  als Eigenvektor zum größten Eigenwert von  $\Sigma_{(2)(2)}^{-1}\Sigma_{(2)(1)}^{-1}\Sigma_{(1)(1)}^{-1}\Sigma_{(1)(2)}^{-1}$  entfällt (Lee, 1999). Zur Verdeutlichung der Auswirkungen einer Klassifikation mithilfe der kanonischen Korrelationsanalyse

Zur Verdeutlichung der Auswirkungen einer Klassifikation mithilfe der kanonischen Korrelationsanalyse im Merkmalsraum ist in Abbildung 20 in Analogie zu den Beispielen aus den Abbildungen 17 und 19 eine Teilmenge der Primitive aus der Beispielaufnahme des MSI von Sentinel 2 vom 16.2.2019 für das Untersuchungsgebiet bei Kraase (ESA, 2019) in zweidimensionalen Projektionen des durch die Merkmalsauswahl aufgespannten Merkmalsraumes dargestellt. Die Datenpunkte sind nach ihrer Klassenzugehörigkeit aus einer beispielhaften Klassifikation mithilfe der kanonischen Korrelationsanalyse eingefärbt.

#### Datenpunkte nach Klassen

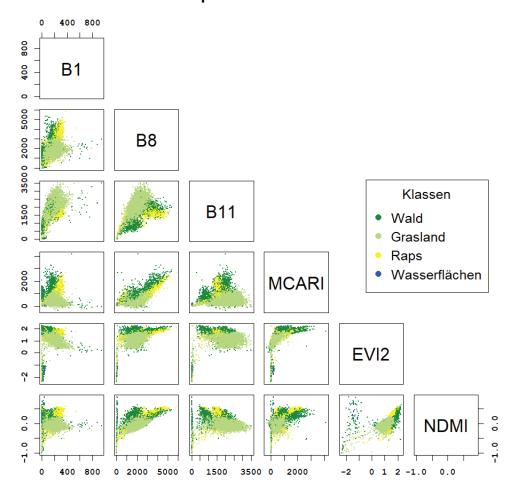


Abbildung 20: Beispiel für das Resultat einer Klassifikation mithilfe der kanonischen Korrelationsanalyse im Merkmalsraum

#### 5.7.6 Random Forest Klassifikation

Random Forests (etwa: zufällige Entscheidungswälder) sollen als moderne Methode des maschinellen Lernens zur Lösung von Klassifikationsproblemen auf die Problemstellung angewendet werden. Bei Random Forests handelt es sich um ein Verfahren des Ensemble-Lernens. Dabei soll die Leistungsfähigkeit schwacher Klassifikatoren durch die Kombination von Ergebnissen mehrerer Instanzen des Klassifikators gesteigert werden. Bei Random Forests werden binäre Entscheidungsbäume als grundlegender Klassifikator genutzt. Ein Merkmalsvektor durchläuft einen Baum vom Wurzel- bis in die Blattknoten. In jedem Knoten wird ein binärer Test über den Daten durchgeführt, welche im Anschluss - in Abhängigkeit von dem Ergebnis - entweder an den rechten oder linken Kindknoten weitergereicht werden. Das Ergebnis der Klassifikation ergibt sich aus der Verteilung auf die Blattknoten. Bei Random Forests durchläuft ein Merkmalsvektor eine Vielzahl verschiedener Entscheidungsbäume, deren Einzelergebnisse zu einem Endergebnis (beispielsweise als gewichtetes Mittel) zusammengefasst werden. Die resultierenden Klassengrenzen im Merkmalsraum können bei der Random Forest Klassifikation komplexe Formen annehmen. Die Generierung der einzelnen Entscheidungsbäume erfolgt bis zu einem gewissen Grad (abhängig von der Implementierung) zufallsbestimmt. Beispielsweise können die Tests in den Knoten eines Baumes zufällig aus einer vordefinierten Menge gezogen, oder die Schwellwerte für die Tests zufällig gewählt werden. Random Forests ermöglichen

auch bei großen Datenmengen eine effektive Klassifikation und werden darüber hinaus kaum vom "Fluch der Dimensionalität" (siehe Abschnitt 5.4 "Merkmalsreduktion") beeinflusst. (HÄNSCH, 2017)

Im Gegensatz zu den bis hierher beschriebenen Klassifikatoren wird für Random Forests keine eigene Implementierung bereitgestellt, stattdessen wird für entsprechende Klassifikationen die R Bibliothek "randomForest" in der Version 4.6-14 genutzt (Liaw, 2018), eine Implementierung der Breiman'schen Random Forests (Breiman, 2001).

In Analogie zu den vorausgegangenen Beispielen (siehe Abbildung 17, 19 und 20) ist in Abbildung 21 eine Teilmenge der Primitive aus der Beispielaufnahme des MSI von Sentinel 2 vom 16.2.2019 für das Untersuchungsgebiet bei Kraase (ESA, 2019) in zweidimensionalen Projektionen des Merkmalsraumes dargestellt. Die Datenpunkte der Primitive sind nach ihrer Klassenzugehörigkeit aus einer Random Forest Klassifikation mit 100 Entscheidungsbäumen eingefärbt.

#### Datenpunkte nach Klassen

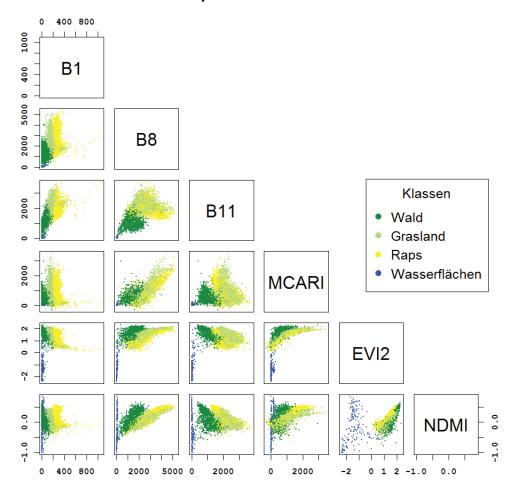


Abbildung 21: Beispiel für das Resultat einer Random Forest Klassifikation im Merkmalsraum

## 6 Fernerkundung von immergrünen Nadel- und Mischwald

Viele Ökosysteme zeichnen sich durch eine hohe Dynamik aus. Verschiedene Pflanzenarten werden in der Regel anhand von Merkmalen unterschieden, die mit der Zeit variieren oder auch nur zu bestimmten Zeitpunkten zu beobachten sind (z.B. Knospen, Blättern, Blüten oder Früchten; Pahlow, 2013). Folglich kann es sich bei der Bestimmung von Pflanzen um eine multitemporale Problemstellung handeln. Auch

bei der Fernerkundung von Pflanzenspezies kann die Anwendung multitemporaler Verfahren von Vorteil sein (Ghoussein, 2019; Heupel, 2018; Tarantino, 2019).

Ein eindrucksvolles Beispiel für die Unterscheidung von Pflanzen auf der Grundlage zeitlich variabler Merkmale ist die Differenzierung zwischen immergrünen Nadel- und Laubwäldern: Während Laubbäume ihre Blätter im Herbst verlieren, tragen die meisten Nadelbäume ihre grünen Nadeln ganzjährig. Viele Nadelund Laubbäume werden auch als Heilpflanzen genutzt und einige Spezies bevorzugen Standorte in Nadelbzw. Mischwäldern. Beispielsweise kommen Bärlapp (Lycopodium clavatum) und Wintergrün (Chimaphila umbellata) in Nadelwäldern vor, während Waldmeister (Galium odoratum) und Lungenkraut (Pulmonaria officinalis) in Laubwäldern anzutreffen sind (Pahlow, 2013). Folglich kann das Wissen um die räumliche Verteilung von immergrünen Nadel- und Mischwäldern eine wichtige Hilfestellung bei der gezielten Suche nach Heilpflanzen sein.

Daher soll hier ein mögliches multitemporales Verfahren für die Fernerkundung von Pflanzenspezies zunächst anhand eines Beispiels zur Fernerkundung von immergrünem Nadel- und Mischwald in dem Untersuchungsgebiet bei Kraase vorgestellt werden. Das Ziel dieses multitemporalen Ansatz zur Klassifikation ist die Ausnutzung von spezifischen Variationen in den Spektren der jeweiligen Pflanzenarten zu unterschiedlichen Zeitpunkten, um die Abgrenzung verschiedener Spezies zu verfeinern und so das Resultat der Klassifikation im Vergleich zu unitemporalen Verfahren zu verbessern. Dabei sind für immergrüne Nadel- und Laubbäume die charakteristischen Abweichungen in den zugehörigen Spektren beim Vergleich zwischen Sommer und Winter bereits mit bloßem Auge wahrnehmbar. Aus diesem Grund werden für die Analyse zwei Aufnahmen des MSI von Sentinel 2 verwendet: eine Aufnahme aus dem Sommer (vom 26.7.2019) und eine aus dem Winter (vom 16.2.2019; ESA, 2019).

Der Ablauf der Analyse ist in Abbildung 22 skizziert. Dabei wird zunächst für beide Aufnahmen eine Merkmalsauswahl nach minimaler Korrelation unter einem Schwellwert (siehe Abschnitt 5.4 "Merkmalsreduktion") bestimmt. Für die Aufnahme vom 16.2.2019 ergeben sich für Schwellwerte von 0.7 und 0.8 jeweils identische Tupel von untereinander minimal korrelierten Merkmalen mit den Spektralbändern B1, B8 und B11 sowie den spektralen Indizes EVI2, MCARI und NDMI. Bei der Aufnahme vom 26.7.2019 umfasst das Tupel minimal korrelierter Merkmale für einen Schwellwert von 0.8 die Spektralbänder B4 und B6 sowie die spektralen Indizes NDMI, MCARI und LCI. Im weiteren Verlauf der Analyse werden für jede Aufnahme nur die Merkmale aus der jeweiligen Merkmalsauswahl berücksichtigt.

Die Abgrenzung der Trainingsgebiete erfolgt unter Berücksichtigung mehrerer Clusterverteilungen zu unterschiedlichen Startwerten für die Ausgangsdaten. Die Trainingsdaten werden in Gebieten definiert, in denen die vorkommenden Spezies und ihre Verteilung bekannt sind. Die Klassifikationen werden für immergrünen Nadel- und Mischwald jeweils als Einklassenfall MLC vorgenommen. Der Ansatz zur multitemporalen Klassifikation wird hierbei vorerst nur bei der Fernerkundung von immergrünen Nadelwald genutzt und basiert auf der Kombination der unitemporalen Einzelergebnisse. Die Fernerkundung der Mischwälder erfolgt ausschließlich anhand der Aufnahme aus dem Sommer. Die Verteilungen der beiden Waldtypen werden mithilfe von Clusteralgorithmen aus den Klassifikationsergebnissen extrahiert. Für das Endergebnis wird die Verteilung von Mischwald (als allgemeinere Klasse) mit der Verteilung von immergrünen Nadelwald (als spezifischere Klasse) überlagert. Dabei werden Gebiete, in denen beide Klassen anzutreffen sind, immergrünem Nadelwald zugeordnet. Das Resultat wird im Anschluss genutzt, um einen Random Forest für eine multitemporale Klassifikation über das gesamte Ausgangsdatenvolumen zu trainieren.

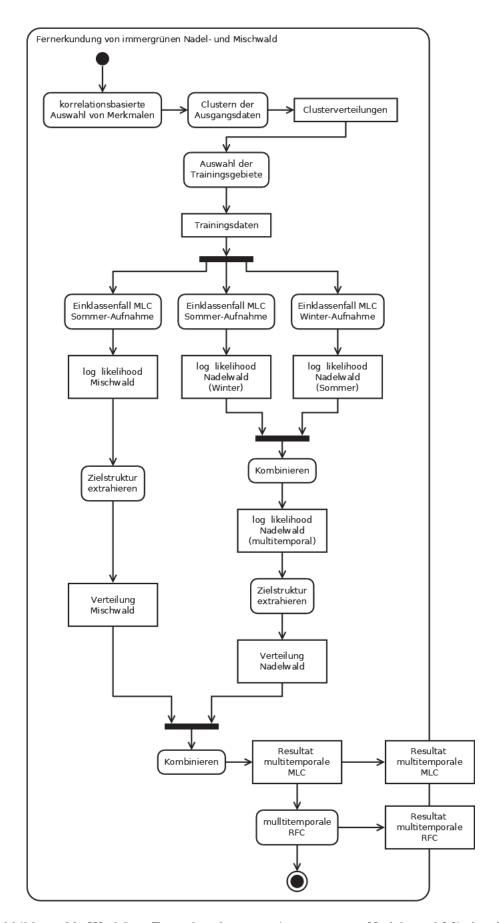


Abbildung 22: Workflow Fernerkundung von immergrünen Nadel- und Mischwald

#### 6.1 Auswahl der Trainingsgebiete

Die Abgrenzung der Trainingsgebiete soll grundsätzlich in einem möglichst kleinen räumlichen Gebiet erfolgen. Diese Beschränkung auf ein kleinstmögliches Gebiet beruht insbesondere auf dem Prinzip aus der Geostatistik, nach dem die räumliche Varianz in kleineren Gebieten geringer ist, als in größeren (KRIGE, 1951). Übertragen auf die Problemstellung ergibt sich aus diesem Prinzip insbesondere die Schlussfolgerung, dass in einem kleineren Gebiet in der Regel weniger unterschiedliche Pflanzenspezies vorkommen, als in einem größeren. Folglich sind in kleineren Untersuchungsgebieten für die Gruppierung von Gebieten (als Primitive) mit Spezies ähnlicher spektraler Eigenschaften mithilfe von Clusteralgorithmen weniger Cluster erforderlich und die Gefahr von Verwechselungen der Zielspezies mit Spezies ähnlicher spektraler Eigenschaften wird minimiert. Darüber hinaus resultiert aus der Beschränkung auf ein kleineres Gebiet auch eine erhebliche Zeitersparnis bei der Abarbeitung der Clusteralgorithmen.

Die Trainingsgebiete werden in einem Wald ausgewählt, in dem die Verteilung unterschiedlicher Spezies in weiten Teilen aus der Ortskenntnis und Begehungen bekannt ist. Dabei handelt es sich um das Waldgebiet östlich von Möllenhagen (siehe Abbildung 23). Der Wald erstreckt sich von Nord nach Süd über eine Länge von etwa 4 Kilometern bei einer durchschnittlichen Breite von annähernd einem Kilometer. Der nördliche Teil liegt weiter im Westen, als der südliche. Die botanische Zusammensetzung ist sehr heterogen und umfasst neben unterschiedlichen Laub- (Bilder 1, 2, 3, 8 und 9 in Abbildung 23) und Nadelgehölzen (Bilder 6, 7 und 10 in Abbildung 23) teilweise auch "Exoten" wie beispielsweise Zypressen (Bild 4 in Abbildung 23). An einigen Stellen treffen reiner Laub- und Nadelwald direkt aufeinander (Bild 5 in Abbildung 23). Auch die Dichte des Bewuchs (vergleiche Bild 8 und 9 für Laubwald bzw. 6 und 7 für Nadelwald in Abbildung 23) und das Alter der Bäume (vergleiche z.B. Bild 3 und 9 in Abbildung 23 für Laubbäume) unterliegt innerhalb des Waldes starken Schwankungen. Im gesamten Wald sind Spuren von unterschiedlich intensiver Abholzung (siehe Bild 10 in Abbildung 23) vorzufinden. Insbesondere im nördlichen Teil des Waldes finden sich zur Mitte des Jahres 2020 teilweise haushohe Ansammlungen von Baumstämmen und weite Landstriche in denen kein Baum mehr steht. In einigen Gebieten sind bereits Anpflanzungen mit neuen Bäumen anzutreffen.

Zur Definition der Trainingsdaten werden die lokalen Verhältnisse in Begehungen festgestellt und mit Fotografien dokumentiert. Die Trainingsdaten sollen anhand von Gebieten definiert werden, in denen die beiden unterschiedlichen Waldtypen in möglichst reiner Form vertreten sind (hohe Konzentration an immergrünem Nadelgehölz für immergrünen Nadelwald und an Laubgehölz für Mischwald). Für die zuverlässige Abgrenzung der Trainingsgebiete werden die Ausgangsdaten mehrfach sowohl mit dem k-Means-, als auch mit dem EM-Cluster Algorithmus für unterschiedliche Startwerte geclustert. Das Clustern erfolgt bei jeder Aufnahme nur über die Merkmale der zugehörigen Merkmalsauswahl. Dabei sollen nach Möglichkeit unterschiedliche Cluster für die Zielstrukturen (immergrüner Nadel- und Mischwald) in den Ausgangsdaten aufgedeckt werden. Die Grenzen der Trainingsgebiete werden folglich so festgelegt, dass innerhalb der einzelnen Gebiete möglichst wenige unterschiedliche Cluster (im Optimalfall 1) in den Clusterverteilungen vorkommen. Für die Aufnahmen aus dem Winter und aus dem Sommer werden mit dieser Methode jeweils unterschiedliche Trainingsgebiete definiert. Bei der Aufnahme aus dem Winter sind lediglich Trainingsdaten für immergrünen Nadelwald erforderlich. Die Trainingsdaten für die Aufnahme vom 26.7.2019 sind mit jeweils 3 Clusterverteilungen für unterschiedliche Startwerte aus dem k-Means- und EM-Cluster Algorithmus in Abbildung 24 dargestellt. Bei der Aufnahme vom 26.7.2019 beträgt der Jeffries-Matusita Abstand zwischen den Verteilungen für immergrünen Nadel- und Mischwald aus den Trainingsdaten rund 1.997, was eine gute Separabilität indiziert. In einigen Clusterverteilungen aus Abbildung 24 sind nahe der Position 6 aus Abbildung 23 Unregelmäßigkeiten im Vergleich mit dem RGB-Echtfarbenkomposit der Aufnahme vom 16.2.2019 festzustellen: Das Gebiet westlich dieser Position wird in vielen Clusterverteilungen dem selben Cluster zugeordnet, wie immergrüner Nadelwald, obwohl es in der Aufnahme aus dem Winter nicht grün erscheint und folglich die Bäume dort im Winter keine Blätter oder Nadeln tragen. Der Grund für diese Abweichung ist eine Eigenart der Spezies, deren dominantes Vorkommen in diesem Gebiet aus den Begehungen bekannt ist: die Lärche (Larix). Lärchen verlieren ihre Nadeln im Herbst. Daher werden Wälder mit Lärchen hier nicht zu den immergrünen Nadelwäldern gezählt und das Gebiet den Trainingsgebieten für Mischwald zugeordnet. Innerhalb des Gebietes zur Abgrenzung der Trainingsdaten kommen Lärchen außerdem südlich von Position 8 in Abbildung 23 (rechte Seite in Bild 8) und östlich von Position 9 (gegenüber dem in Bild 9 dargestellten Bereich) vermehrt vor.

## Gebiet zur Auswahl der Trainingsgebiete im Wald östlich von Möllenhagen



Abbildung 23: Das Waldgebiet östlich von Möllenhagen in der Aufnahme vom 16.2.2019 (RGB-Echtfarbenkomposit oben links; ESA, 2019) mit dem Gebiet für die Auswahl der Trainingsdaten (purpurnes Polygon), Bildern von der Situation vor Ort vom 27.3.2020 (1-9) und 6.10.2019 (10) und den Positionen der Beobachtungen (nummerierte purpurne Punkte)

## Abgrenzung der Trainingsgebiete

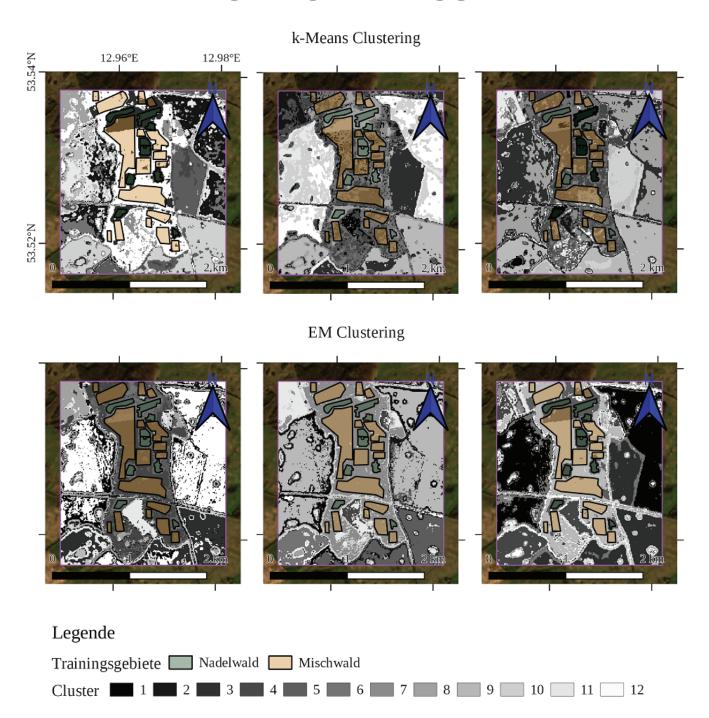


Abbildung 24: Clusterverteilungen für unterschiedliche Startwerte beim k-Means- und EM-Cluster Algorithmus mit den Trainingsgebieten für die Aufnahme vom 26.7.2019 (ESA, 2019)

#### 6.2 Multitemporale Klassifikation

Multitemporale Klassifikationsverfahren lassen sich als Anwendungen von Datenfusion interpretieren. Grundsätzlich können dabei zwei unterschiedliche Vorgehensweisen unterscheiden werden: I.) die Erweiterung des Merkmalsraumes um Merkmale aus Aufnahmen zu mehreren unterschiedlichen Zeitpunkten (Stacked Vector) und II.) die Kombination von Resultaten unitemporaler Klassifikationen zu einem multitemporalen Endergebnis (Jeon, 1999; Ghamisi, 2019).

Der hier vorgestellte Ansatz zur multitemporalen ML Klassifikation basiert ebenfalls auf der Kombination von unitemporalen Klassifikationsergebnissen. Dazu wird von der logarithmierten Likelihood-Funktion nach Gleichung 18 ausgegangen und Verteilungen der Likelihood-Funktionen für Aufnahmen zu unterschiedlichen Zeitpunkten als Priori Verteilungen p(i) genutzt. Ausgegangen von einer Zeitreihe von Aufnahmen als Beispiel, werden zunächst die Verteilungen der Likelihood-Funktionen der einzelnen Klassen für die erste Aufnahme wie bei einer klassischen unitemporalen ML Klassifikation ohne Priori Verteilung p(i) bestimmt. Die resultierenden Verteilungen werden bei der Berechnung der Verteilungen der Likelihood-Funktionen für die folgende Aufnahme als Priori Verteilungen p(i) genutzt und analog bei den Berechnungen für jede weitere Aufnahme die Resultate aus der vorhergehenden als Priori Verteilungen p(i) verwendet. Da so nach Gleichung 18 die einzelnen Verteilungen der Likelihood-Funktionen summiert werden, ist die zeitliche Reihenfolge nicht mehr von Bedeutung und die Berechnungen können auch separat erfolgen, um zuletzt zu einem multitemporalen Endergebnis zusammengefasst zu werden. Folglich lässt sich - unter Berücksichtigung des Logarithmus - die resultierende multitemporale Verteilung der Klassen-Likelihood als Produkt der unitemporalen Verteilungen darstellen (Benediktson, 1990).

Bei Mehrklassenproblemen erfolgt die Zuordnung der Primitive auf die einzelnen Klassen auf Grundlage der resultierenden multitemporalen Verteilungen der Likelihood-Funktionen wie üblich nach der Maximum Likelihood Methode. Da hier für die Problemstellung der Fernerkundung von Heilpflanzen die Anwendung von Einklassenfall Klassifikationen von Vorteil ist (siehe Abschnitt 5.7 "Klassifikationsalgorithmen"), entfällt die Zuteilung der Primitive auf unterschiedliche Klassen und das Resultat der multitemporalen Klassifikation ist die Verteilung der Likelihood-Funktion für die Zielklasse. Um zu einer binären Verteilung für die An- bzw. Abwesenheit einer Spezies für das Untersuchungsgebiet zu gelangen, wird hier die resultierende Verteilung der Likelihood-Funktion mit dem k-Means Algorithmus geclustert und das Cluster mit den im Durchschnitt höchsten Werten der Likelihood-Funktion als Zielstruktur extrahiert. Das Resultat wird mit den Erfahrungen aus der Ortskenntnis über die Verteilung der Spezies abgeglichen und überprüft. Insbesondere bei einer niedrigen Anzahl an Clustern für den k-Means Algorithmus und bei Spezies, die innerhalb des Untersuchungsgebietes nur vereinzelt vorkommen, können im Endergebnis viele "falsch positive" Treffer vorkommen, die durch wiederholtes Clustern eines Ausschnitts aus der Likelihood Funktion entlang des ersten Ergebnis ausgeschlossen werden können.

Für die Fernerkundung von immergrünem Nadel- und Mischwald wird dieser multitemporale Ansatz für die ML Klassifikation nur bei der Fernerkundung immergrünen Nadelwaldes genutzt. Die Verteilung von Mischwald wird aus einer unitemporalen Einklassenfall ML Klassifikation der Aufnahme aus dem Sommer (vom 26.7.2019) bestimmt.

Für die multitemporale Einklassenfall ML Klassifikation zur Fernerkundung von immergrünem Nadelwald sind in Abbildung 25 Ausschnitte aus den unitemporalen Verteilungen der logarithmierten Likelihood-Funktion für die Aufnahmen vom 16.2.2019 und 26.7.2019 sowie der resultierenden multitemporalen Verteilung für das Gebiet zur Definition der Trainingsdaten dargestellt. Die Darstellung ist auf die absoluten Maxima fokussiert. Um die Vergleichbarkeit zu gewährleisten, sind für jede Verteilung jeweils die 10% aller Gebiete mit den höchsten Werten der Likelihood-Funktion abgebildet. Dabei handelt es sich hauptsächlich um die Gebiete mit reinem immergrünen Nadelwald. Da der flächenmäßige Anteil an zusammenhängenden Flächen reinen immergrünen Nadelwaldes in dem betrachteten Gebiet leicht unter 10% liegt, finden sich auch Treffer in Gebieten von Mischwald mit geringerer Konzentration von immergrünem Nadelgehölz. Die immergrünen Nadelwälder lassen sich in dem Resultat für die Aufnahme aus dem Winter zuverlässiger abgrenzen. Bei der Klassifikation der Aufnahme aus dem Sommer werden auch dem mit Lärchen (Larix) bewachsenen Gebiet im Norden hohe Werte der Likelihood-Funktion zugeordnet und einige längliche Gebiete immergrünen Nadelwaldes im Osten sind schwerer abzugrenzen. Beim Vergleich der beiden unitemporalen Ergebnisse gilt außerdem zu berücksichtigen, dass die Klassifikationen über unterschiedlichen Merkmalsräumen durchgeführt werden und die Dimensionalität des Merkmalsraumes für die Aufnahme vom 26.7.2019 um eins kleiner ist, als für die Aufnahme aus dem Winter.

# Resultate immergrüner Nadelwald (Maxima)

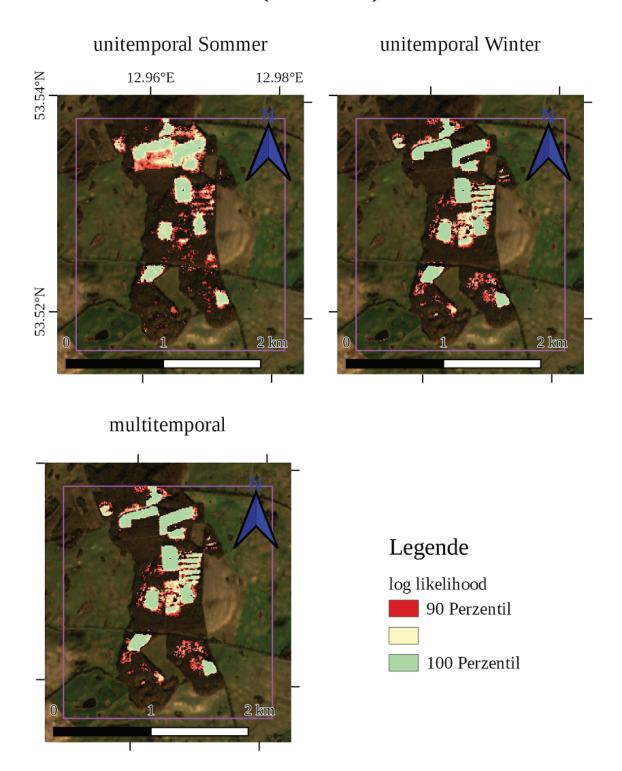


Abbildung 25: unitemporale Einzelergebnisse und multitemporales Gesamtergebnis der multitemporalen Einklassenfall ML Klassifikation für immergrünen Nadelwald innerhalb des Gebietes zur Definition der Trainingsdaten (purpurnes Polygon) mit RGB Echtfarbenkomposit der Aufnahme des MSI von Sentinel 2 vom 16.2.2019 (ESA, 2019) im Hintergrund

Zur weiteren Analyse der Ergebnisse der multitemporalen Einklassenfall ML Klassifikation für immergrünen Nadelwald werden die Resultate für Gebiete betrachtet, in denen die Verteilung von immergrünem Nadel- und Mischwald aus den Trainingsdaten bekannt ist. Da für die Aufnahme aus dem Winter keine Trainingsdaten für Mischwald erforderlich sind, werden hierfür die Trainingsgebiete für die Aufnahme vom 26.7.2019 genutzt. In Abbildung 26 sind Histogramme für die resultierenden Verteilungen der logarithmierten Likelihood-Funktionen innerhalb der Trainingsgebiete dargestellt. Da die Dimensionalität des Merkmalsraumes für die Klassifikation der Aufnahme aus dem Winter um eins größer ist, als bei der Aufnahme aus dem Sommer, sind die resultierenden Werte der Likelihood-Funktion durch den exponentiellen Zuwachs des hyperdimensionalen Volumens für die Aufnahme aus dem Winter generell deutlich kleiner, als für die Aufnahme aus dem Sommer. Die Werte unterscheiden sich in etwa um eine Größenordnung (Faktor 10). Daher ist der direkte Vergleich beider Klassifikationsergebnisse schwierig.

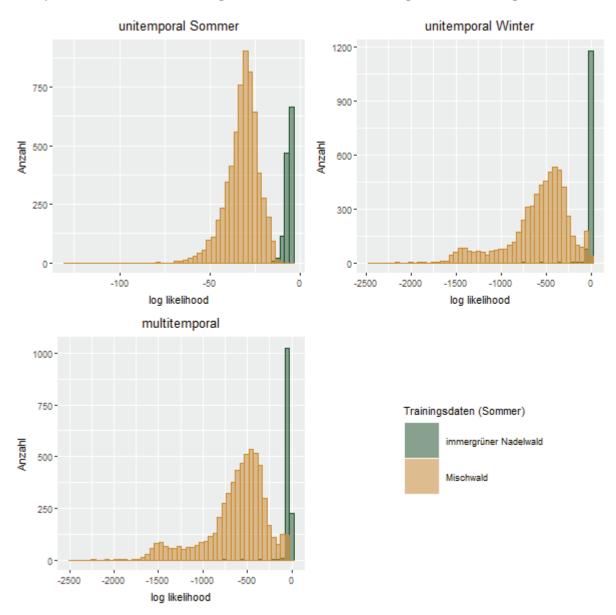


Abbildung 26: Histogramme für Ausschnitte aus den unitemporalen Einzelergebnissen und dem multitemporalen Gesamtergebnis der multitemporalen Einklassenfall ML Klassifikation für immergrünen Nadelwald innerhalb der Trainingsgebiete für die Aufnahme aus dem Sommer

Die Mittelwerte, Spannweiten, Minima und Maxima der resultierenden Verteilungen der logarithmierten Likelihood-Funktionen für immergrünen Nadelwald innerhalb der Trainingsgebiete sind in Tabelle 3 zusammengefasst.

Tabelle 3: einige statistische Kenngrößen für die resultierenden Verteilungen der Likelihood-Funktionen innerhalb der Trainingsgebiete

Verteilung	Minimum	Mittelwert	Maximum	Spannweite
Sommer: immergrüner Nadelwald	-17.79	-6.77	-4.50	13.29
Sommer: Mischwald	-128.87	-32.24	-12.35	116.51
Winter: immergrüner Nadelwald	-1'121.87	-19.18	-8.45	1'113.42
Winter: Mischwald	-2'461.28	-583.71	-11.12	2'450.16
multitemporal: immergrüner Nadelwald	-1'147.29	-36.06	-23.43	1'123.85
multitemporal: Mischwald	-2'508.36	-626.06	-34.53	2'473.83

Die Absolutwerte dieser Parameter unterscheiden sich zwischen den Resultaten für die Aufnahme aus dem Sommer und Winter deutlich. Dieser Unterschied ist zwischen dem Resultat aus der Aufnahme aus dem Winter und dem multitemporalen Endergebnis geringer, da das multitemporale Endergebnis aus der Addition der im Durchschnitt betragsmäßig vergleichsweise geringen Werte der logarithmierten Likelihood-Funktion aus dem Sommer zu den deutlich höheren aus dem Winter hervorgegangen ist. In allen Verteilungen befinden sich die absoluten Maxima in Trainingsgebieten für immergrünen Nadelwald und die absoluten Minima in Trainingsgebieten für Mischwald. Diese Beobachtung ist konsistent mit der Tatsache, dass es sich um Verteilungen von Likelihood-Funktionen für immergrünen Nadelwald handelt. Die Spannweite der Ausschnitte aus den Verteilungen über alle Trainingsgebiete ist bei dem Resultat für die Aufnahme aus dem Sommer am kleinsten und bei dem multitemporalen Resultat am größten. Die Spannweite der resultierenden Verteilung wird hier also durch den multitemporalen Ansatz erweitert. Diese Entwicklung lässt sich auch über die Trainingsgebiete hinaus beobachten: Innerhalb des gesamten Untersuchungsgebietes beträgt die Spannweite der resultierenden Verteilung für die Aufnahme aus dem Sommer 41'406.54, für die Aufnahme aus dem Winter 22'095.81 und für das multitemporale Endergebnis 47'449.76. Da sich die Werte der Likelihood-Funktion bei einer größeren Spannweite über einen weiteren Bereich verteilen, kann dadurch auch die Abgrenzung einer charakteristischen Zielstruktur (hier immergrüner Nadelwald) als Ballung bei maximalen Werten der Likelihood-Funktion mithilfe von Clusteralgorithmen verbessert werden. Im Gegensatz zu den Verteilungen innerhalb der Trainingsgebiete ist im gesamten Untersuchungsgebiet die Spannweite der Verteilung für die Aufnahme aus dem Winter am kleinsten. Um die Ursachen für diese Unregelmäßigkeit aufzudecken, sind in Abbildung 27 Histogramme der unitemporalen logarithmierten Verteilungen der Likelihood-Funktionen für das gesamte Untersuchungsgebiet mit Ausschnitten in der näheren Umgebung des absoluten Maximums dargestellt. Es zeigt sich, dass die große Spannweite für die resultierende Verteilung bei der Aufnahme aus dem Sommer durch Ausreißer verursacht wird. Nur etwa 0.01% der Werte liegen unter -26'000. Bei der Verteilung für die Aufnahme aus dem Winter liegen etwa 0.02% der Werte unter -20'000. Folglich liegt bei beiden Verteilungen der Großteil der Werte in etwa auf dem Intervall zwischen -20'000 und 0. Da es sich um Verteilungen von logarithmierten Likelihood-Funktionen für immergrünen Nadelwald handelt, ist zu erwarten, dass es sich - wie in Abbildung 26 - bei den Ansammlungen nahe des absoluten Maximums um die Werte für Merkmalsvektoren von Flächen mit immergrünem Nadelwald handelt. Für die nähere Umgebung der absoluten Maxima weisen die Histogramme starke Ähnlichkeiten mit denen aus Abbildung 26 auf.

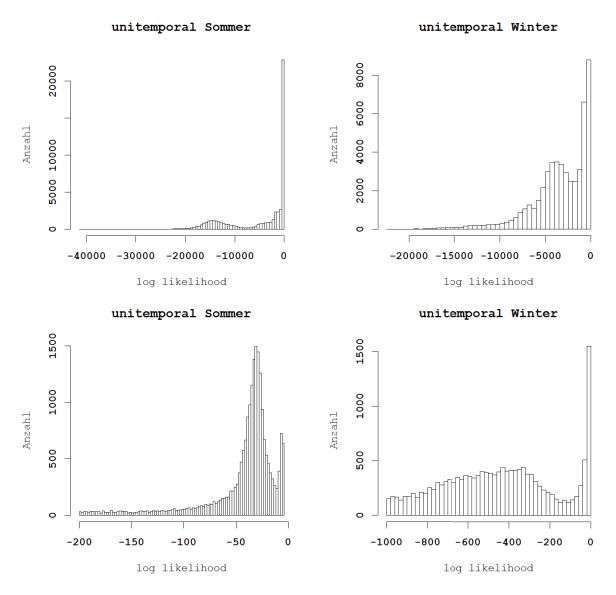


Abbildung 27: Histogramme für die unitemporalen Einzelergebnisse der multitemporalen Einklassenfall ML Klassifikation für immergrünen Nadelwald (gesamtes Untersuchungsgebiet)

Im Gegensatz zu Tabelle 3 sind die statistischen Kenngrößen Minimum, Mittelwert, Maximum und Spannweite für die resultierenden Verteilungen innerhalb der Trainingsgebiete in Tabelle 4 als relative Werte in Bezug auf das absolute Maximum und die Spannweite des gesamten Datenumfangs zusammengetragen. Für die Aufnahme aus dem Sommer nimmt die Spannweite der Werte der Likelihood-Funktion innerhalb der Trainingsgebiete für immergrünen Nadelwald 10.69% des gesamten Wertebereichs innerhalb aller Trainingsgebiete ein, für die Aufnahme aus dem Winter sind es 45.39% und für das multitemporale Endergebnis 45.23%. Im Vergleich zu der Verteilung für die Aufnahme aus dem Winter hat sich dieser Bereich im multitemporalen Endergebnis verkleinert. Ähnlich verhält es sich mit den Spannweiten für die Werte innerhalb der Trainingsgebiete für Mischwald: Auch hier ist der Bereich im Resultat für die Aufnahme aus dem Sommer mit 93.68% am kleinsten und in der Verteilung für die Aufnahme aus dem Winter mit 99.89% am größten. Im Vergleich zu dem Ergebnis für die Aufnahme aus dem Winter hat sich der Bereich im multitemporalen Endergebnis um 0.33% auf 99.56% verkleinert.

Tabelle 4: einige statistische Kenngrößen für die resultierenden Verteilungen der Likelihood-Funktionen innerhalb der Trainingsgebiete als relative Werte (Angaben in %)

Verteilung	Minimum	Mittelwert	Maximum	Spannweite
Sommer: immergrüner Nadelwald	-10.69	-1.83	0.00	10.69
Sommer: Mischwald	-100.00	-22.30	-6.32	93.68
Winter: immergrüner Nadelwald	-45.39	-0.44	0.00	45.39
Winter: Mischwald	-100.00	-23.45	-0.11	99.89
multitemporal: immergrüner Nadelwald	-45.23	-0.51	0.00	45.23
multitemporal: Mischwald	-100.00	-24.25	-0.45	99.56

Die Überschneidungen zwischen den Wertebereichen von immergrünem Nadel- und Mischwald sind bei der Abgrenzung der Werte der Likelihood-Funktionen für immergrünen Nadelwald als Ballungen nahe dem absoluten Maximum mithilfe von Clusteralgorithmen aus den Verteilungen problematisch: In diesem Bereichen nimmt die Likelihood-Funktion sowohl für Primitive, die der Zielklasse (hier: immergrüner Nadelwald) zuzuordnen sind, als auch für solche, die nicht zur Zielklasse gehören identische Werte an. Es besteht die Gefahr der fälschlichen Zuordnung von Gebieten, die tatsächlich mit immergrünem Nadelwald bewachsen sind, zu Mischwäldern (Fehler erster Art) und von Gebieten, die tatsächlich mit Mischwald bewachsen sind, zu immergrünen Nadelwäldern (Fehler zweiter Art). Daher ist eine minimale Weite dieses Bereichs bzw. eine möglichst geringe Anzahl von Primitiven, die in diesen Bereich fallen für die Abgrenzung der Zielstruktur von Vorteil. Der Bereich, in dem die Werte der Likelihood-Funktion für die Verteilungen in Trainingsgebieten mit immergrünem Nadel- und Mischwald identisch sind, ist mit rund 45.28% der gesamten Spannweite bei dem Resultat für die Aufnahme aus dem Winter am größten und mit rund 4.37% bei dem Resultat für die Aufnahme aus Sommer am kleinsten. Im Vergleich zu dem unitemporalen Resultat aus dem Winter hat sich der Bereich bei dem multitemporalen Endergebnis um etwa 0.5% auf rund 44.78% verkleinert. Obwohl sich der immergrüne Nadelwald in der Aufnahme aus dem Winter besser abgrenzen lässt, als in der Aufnahme aus dem Sommer, ist der kritische Bereich in dem Resultat für die Aufnahme aus dem Winter mehr als 10 mal so groß, wie in dem Resultat für die Aufnahme aus dem Sommer. Das Histogramm für den Ausschnitt aus dem Resultat für die Aufnahme aus dem Winter in Abbildung 26 weist darauf hin, dass die Werte der logarithmierten Likelihood-Funktion nur für sehr wenige Gebiete in Trainingsgebieten für immergrünen Nadelwald kleiner als -100 sind. Tatsächlich sind in der resultierenden Verteilung für die Aufnahme aus dem Winter innerhalb der Trainingsgebiete für immergrünen Nadelwald nur etwa 1% der Werte kleiner als -158, sodass sich der kritische Bereich auf etwa 6% der gesamten Spannweite verkleinert, wenn nur die 99% der Primitive mit den größten Werten der logarithmierten Likelihood-Funktion berücksichtigt werden. Es liegt nahe, dass es sich bei den Gebieten mit sehr kleinen Werten der logarithmierten Likelihood-Funktion innerhalb der Trainingsgebiete für immergrünen Nadelwald um Ausreißer handelt, die auf die fälschliche Zuteilung von Gebieten mit geringem Bewuchs von immergrünem Nadelgehölz oder vermehrtem Aufkommen von Laubgehölz auf den Trainingsdatensatz für immergrünen Nadelwald zurückzuführen sind. Ein weiterer Grund für die zum Teil großen Überschneidungen ist die Definition des Mischwaldes als die allgemeinere Klasse: Innerhalb von Flächen mit Mischwald können auch Teile mit immergrünen Nadelbäumen bewachsen sein. Darüber hinaus basiert die Abgrenzung der Trainingsgebiete für Mischwald ausschließlich auf der Aufnahme aus dem Sommer. Zu dieser Jahreszeit sind die Reinformen der beiden Waldtypen schwerer gegeneinander abzugrenzen (siehe Abbildung 24 und 25) und jüngere immergrüne Nadelbäume können in den Aufnahmen durch das Blattwerk älterer Laubbäume verdeckt werden. Beim Vergleich der Trainingsgebiete für Mischwald in Abbildung 24 und dem unitemporalen Resultat für die Aufnahme aus dem Winter in Abbildung 25 fällt auf, dass Teile der länglichen Gebiete immergrünen Nadelwaldes im Osten des Waldes auf Trainingsgebiete für Mischwald entfallen. Damit lassen sich Gebiete mit hohen Werten der logarithmierten Likelihood-Funktion innerhalb der Trainingsgebiete für Mischwald erklären.

Für die Abgrenzung einer Zielstruktur aus den Werten einer Likelihood-Funktion als Ballung nahe dem absoluten Maximum mithilfe von Clusteralgorithmen kann auch ein möglichst großer Abstand zwischen den Zentren unterschiedlicher Cluster vorteilhaft sein. Bei den Ausschnitten aus den Verteilungen der Likelihood-Funktionen betragen die relativen Abstände zwischen den Mittelwerten der Verteilungen innerhalb der Trainingsgebiete für immergrünen Nadel- und Mischwald für die Aufnahme aus dem Sommer rund 20.47%, für die Aufnahme aus dem Winter rund 23.01% und im multitemporalen Endergebnis 23.74%. Der relative Abstand zwischen den Mittelwerten der Verteilungen der beiden Waldtypen ist in dem Resultat für die Aufnahme aus dem Sommer. Diese Tatsache kann eine Erklärung für die bessere Abgrenzbarkeit immergrünen Nadelwaldes aus dem Resultat für die Aufnahme aus dem Winter sein. Der relative Abstand zwischen den Mittelwerten beider Verteilungen ist in dem multitemporalen Endergebnis am größten.

Die Anwendung des multitemporalen Ansatz zur Einklassenfall ML Klassifikation wirkt sich hier im Endergebnis günstig auf die Abgrenzbarkeit der Zielstruktur als Ballung nahe dem absoluten Maximum aus: Im multitemporalen Endergebnis hat sich die Spannweite der Verteilung im Vergleich zu den unitemporalen Einzelergebnissen erweitert. In den resultierenden multitemporalen Verteilungen innerhalb der Trainingsgebiete hat sich der Überschneidungsbereich der Verteilungen für immergrünen-Nadel und Mischwald verkleinert und der relative Abstand zwischen den Mittelwerten der Verteilungen vergrößert.

#### 6.3 Resultat

Die Verteilung von immergrünem Nadel- und Mischwald innerhalb des Untersuchungsgebietes wird als Endergebnis aus den resultierenden Verteilungen der Likelihood-Funktionen für die beiden Waldtypen als Ballung nahe dem absoluten Maximum mithilfe des k-Means Clusteralgorithmus extrahiert. Die mit Mischwald bewachsenen Gebiete werden als allgemeinere Klasse mit der Verteilung von immergrünem Nadelwald überlagert. Die resultierende Verteilung ist in Abbildung 28 dargestellt.

Zur Kontrolle wird das Resultat mit den Erfahrungen aus der Ortskenntnis und dem RGB-Echtfarbenkomposit der Aufnahme des MSI von Sentinel 2 vom 16.2.2019 (ESA, 2019) verglichen. In RGB-Echtfarbenkompositen von Aufnahmen aus dem Winter lassen sich immergrüne Nadelwälder gut mit mit dem bloßen Auge gegen Laub- und Mischwald abgrenzen und die resultierende Verteilung entspricht den vorgefundenen Verhältnissen. Beispielsweise handelt es sich bei dem Teil des Müritz-Nationalparks als größtes zusammenhängendes Waldgebiet im Süden des Untersuchungsgebiets hauptsächlich um immergrünen Nadelwald. Auch kleinere Gebiete von Mischwald - in diesem Fall Laubwald - in der näheren Umgebung der Quellseen der Havel finden sich korrekt im Endergebnis wieder. Größere Gebiete von Laubwald in den Wäldern nördund südwestlich von Kraase und in der Umgebung der Varchentiner Seen sind im Resultat korrekt in der Verteilung für Mischwald vorzufinden. Die Klassifikation ist sensitiv: Waldgebiete werden bereits bei geringer Konzentration immergrünen Nadelgehölzes im Bewuchs immergrünen Nadelwäldern zugeordnet (beispielsweise im Wald östlich von Möllenhagen).

Das Endergebnis lässt sich durch den multitemporalen Ansatz für die ML-Klassifikation im Vergleich zu unitemporalen Verfahren verbessern. Wegen der hohen Sensitivität beim Auffinden von immergrünem Nadelgehölz ist auch bei der Fernerkundung von Spezies in Gebieten mit vergleichsweise geringer Dichte des Bewuchs von einer guten Eignung des Verfahrens auszugehen. Eine wichtige Voraussetzung für dieses multitemporale Verfahren ist die Anwesenheit der Zielspezies in den Trainingsgebieten für jede Aufnahme. Da Random Forests weniger vom "Fluch der Dimensionalität" beeinflusst werden, wird ein Ausschnitt aus dem Resultat der multitemporalen Einklassenfall ML Klassifikation für immergrünen Nadel- und Mischwald genutzt, um einen Random Forest aus 100 Bäumen zur multitemporalen Klassifikation über alle 32 verfügbaren Merkmale der Aufnahmen vom 16.2.2019 und 27.7.2019 (Stacked Vector Approach) zu trainieren. Als Trainingsgebiet wird dazu die nähere Umgebung von Möllenhagen als ein Gebiet mit sehr heterogener Landbedeckung genutzt. Das Resultat der multitemporalen Random Forest Klassifikation ist mit dem zugehörigen Trainingsgebiet in Abbildung 29 dargestellt. Die Verteilung stimmt bis auf kleinere Unterschiede mit dem Ergebnis der multitemporalen Einklassenfall ML Klassifikation überein. Die multitemporale RF Klassifikation wird analog mit 100 Bäumen für einen abermals um die insgesamt 32 Merkmale der Aufnahmen vom 28.11.2018 und 25.8.2019 des MSI von Sentinel 2 (ESA, 2019) erweiter-

ten 64-dimensionalen Merkmalsraum wiederholt. Da sich das Resultat kaum von dem aus Abbildung 29 unterscheidet, wird hier auf dessen Darstellung verzichtet.

# Resultat multitemporale MLC



Abbildung 28: Karte der resultierenden Verteilung von immergrünem Nadel- und Mischwald innerhalb des Untersuchungsgebietes mit einem RGB-Echtfarbenkomposit der Aufnahme des MSI von Sentinel 2 vom 16.2.2019 im Hintergrund (ESA, 2019)

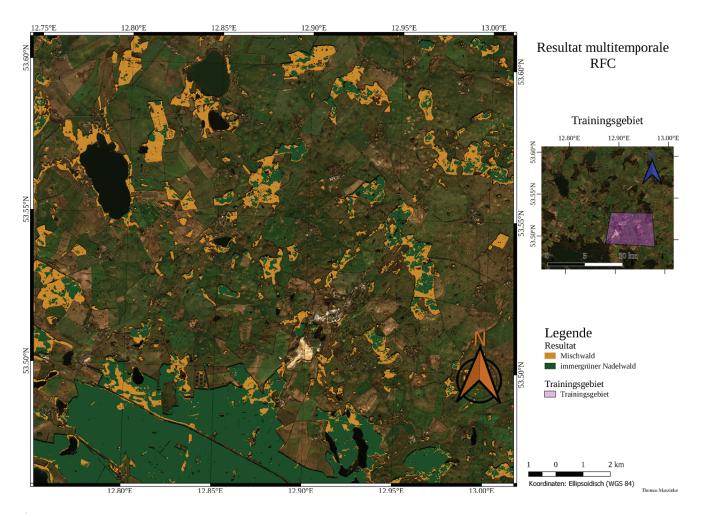


Abbildung 29: Karte der aus der multitemporalen RF Klassifikation resultierenden Verteilung von immergrünem Nadel- und Mischwald innerhalb des Untersuchungsgebietes mit einem RGB-Echtfarbenkomposit der Aufnahme des MSI von Sentinel 2 vom 16.2.2019 im Hintergrund (ESA, 2019) und einer Karte des Trainingsgebiets

## 7 Fernerkundung von Röhricht

Röhricht ist häufig an feuchten Stellen anzutreffen und kann daher als Indikator für entsprechende Habitate dienen. In solchen Lebensräumen kommen auch Heilpflanzen wie z.B. der Bitterklee (Menyanthes trifoliata), Kalmus (Acorus Calamus), Sonnentau (Drosera rotundifolia) oder die Sumpfdotterblume (Caltha palustris) vor (Pahlow, 2013). Folglich kann das Wissen um die räumliche Verteilung von Röhrichtbeständen, wie bei immergrünen Nadel- und Mischwäldern, bei der gezielten Suche nach Heilpflanzen vorteilhaft sein. Somit soll das Verfahren zur multitemporalen Einklassenfall ML Klassifikation aus Abschnitt 6 "Fernerkundung von immergrünen Nadel- und Mischwald" als nächstes Anwendungsbeispiel an der Fernerkundung von Röhrichtbeständen innerhalb des Untersuchungsgebietes bei Kraase erprobt werden.

Auch im Spektrum von Röhricht lassen sich - ähnlich wie bei immergrünen Nadel- und Laubgehölzen - charakteristische Abweichungen bereits mit dem bloßen Auge wahrnehmen: Die Pflanzen sind während der Vegetationsphase im Sommer saftig grün, während das abgestorbene Stroh zwischen Herbst und Frühjahr des folgenden Jahres beige bis gelblich-braun erscheint. Zwischen den frischen Pflanzen kann zum Teil auch noch das Stroh aus den Vorjahren stehen.

Die Analyse für Röhricht wird analog zu der vorangegangenen, in Abbildung 22 skizzierten Analyse für immergrünen Nadel- und Mischwald durchgeführt. Da die Einklassenfall ML Klassifikation hier nur für

eine Spezies und als multitemporale Klassifikation durchgeführt wird, entfällt in Abbildung 22 der Zweig für die unitemporale Klassifikation für Mischwald und die multitemporale Einklassenfall ML Klassifikation wird hier nicht für immergrünen Nadelwald, sondern für Röhricht durchgeführt. Darüber hinaus werden für die multitemporale Einklassenfall ML Klassifikation für Röhricht neben den Aufnahmen des MSI von Sentinel 2 vom 16.2.2019 und 26.7.2019 auch die Aufnahmen vom 28.11.2018 und 25.8.2019 genutzt (ESA, 2019). Für jede Aufnahme sind die Spektralbänder und spektralen Indizes der jeweiligen Merkmalsauswahl nach minimaler Korrelation unter einem Schwellwert in Tabelle 5 zusammengefasst.

Tabelle 5: Spektralbänder und spektrale Indizes der Merkmalsauswahl für die Aufnahmen des MSI von Sentinel 2 vom 28.11.2018, 16.2.2019, 26.7.2019 und 25.8.2019 für das Untersuchungsgebiet bei Kraase

Aufnahme vom	Spektralbänder	spektrale Indizes
28.11.2018	B1, B8, B11	LCI, MCARI, NDMI
16.2.2019	B1, B8, B11	EVI2, MCARI, NDMI
26.7.2019	B4, B6	LCI, MCARI, NDMI
25.8.2019	B2, B6, B11	EVI2, MCARI, NDMI

Die Abgrenzung der Trainingsgebiete erfolgt nach der gleichen Vorgehensweise, wie bei der Analyse für immergrünen Nadel- und Mischwald. Darüber hinaus wird hier noch zusätzlich darauf geachtet, dass die Trainingsgebiete in allen Aufnahmen vollständig mit Röhricht bewachsen sind. Dadurch soll vermieden werden, dass für jede Aufnahme ein separater Satz von Trainingsdaten erforderlich ist. Die Trainingsdaten werden in Gebieten definiert, die aus der Ortskenntnis für einen besonders großflächigen und dichten Röhrichtbewuchs bekannt sind. Dabei handelt es sich um den Klein Plastener See, den Borksee nordöstlich von Groß Dratow und das Nordwestufer des Großen Varchentiner Sees (siehe Abbildung 30). Bei der Abgrenzung der Trainingsgebiete ist zu beachten, dass das Röhricht in den Uferbereichen häufig durch Büsche und Bäume unterbrochen werden kann (siehe Bild 1 und 2 in Abbildung 30). Da es sich dabei nicht um die Zielspezies handelt, müssen entsprechende Gebiete aus den Trainingsdaten ausgeschlossen werden. Die aus der multitemporalen Einklassenfall ML Klassifikation resultierende Verteilung von Röhricht innerhalb des Untersuchungsgebietes ist in Abbildung 31 dargestellt. Da nur ein flächenmäßig kleiner Teil des Untersuchungsgebietes mit Röhricht bewachsen ist, gestaltet sich die Abgrenzung der Werte der Likelihood-Funktion für Gebiete mit Röhricht als Ballung nahe dem absoluten Maximum schwieriger, als bei dem Beispiel für immergrünen Nadel- und Mischwald. Insbesondere bei einer geringen Anzahl an Clustern finden sich in dem Cluster nahe dem absoluten Maximum auch viele Werte für Gebiete, die tatsächlich nicht mit Röhricht bewachsen sind ("falsch-positive" Treffer). Daher wird die resultierende Verteilung für das Cluster nahe dem absoluten Maximum der Werte der Likelihood-Funktion nochmals mithilfe des k-Means Algorithmus geclustert, um die Werte für mit Röhricht bewachsene Gebiete gegen die "falsch-positiven" Treffer abzugrenzen. Die resultierende Verteilung von Röhricht innerhalb des Untersuchungsgebietes spiegelt sehr gut die Erfahrungen aus der Ortskenntnis wieder. Auch zahlreiche kleinere und teilweise sehr kleine mit Röhricht bewachsene Wasserlöcher finden sich korrekt im Endergebnis wieder. An einigen Gewässern kann das Röhricht in den Uferbereichen durch darüber wachsende Bäume verdeckt werden. Ein gutes Beispiel dafür ist der kleine Varchentiner See (siehe Abbildung 32). Insbesondere der schmale Schilfgürtel entlang des Südostufers des Sees wird durch Bäume verdeckt, die sehr nahe beieinander unmittelbar am Uferrand stehen. Der Schatten der Bäume ist auch in dem RGB-Echtfarbenkomposit in Abbildung 32 wiederzuerkennen. In der Fotografie in Abbildung 32 ist ein Bereich entlang des langgestreckten Nordostufer des Sees abgebildet, in dem das Röhricht darüber hinaus durch dichtes Gebüsch entlang des Uferverlaufs zurückgedrängt wird. Im Vordergrund ist ein Teil des Röhrichts vom Südostufer des Sees wiederzuerkennen, welches durch die darüber stehenden Bäume verdeckt wird.

## Gebiete zur Auswahl der Trainingsgebiete

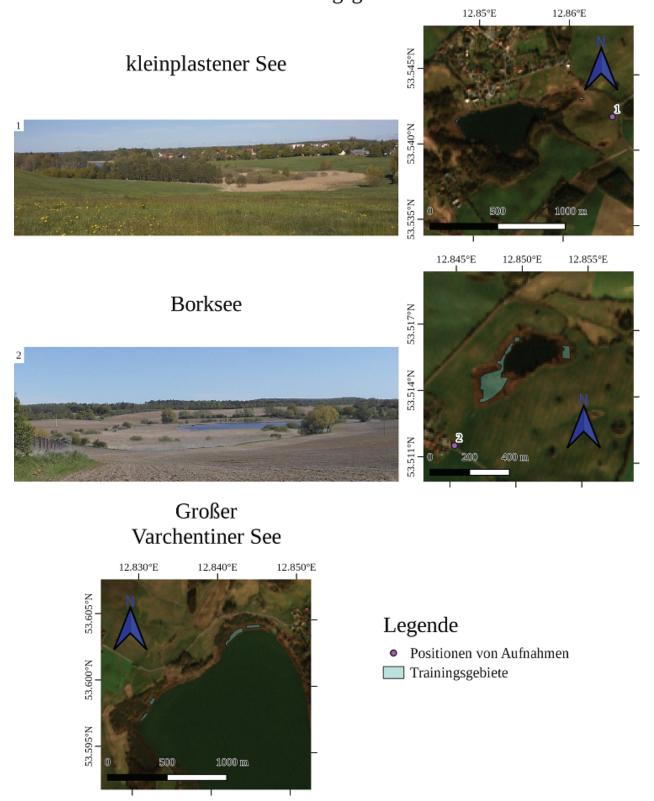


Abbildung 30: Trainingsgebiete für die Röhricht-Analyse in der Aufnahme vom 16.2.2019 (RGB-Echtfarbenkomposite im Hintergrund; ESA, 2019) mit Bildern von der Situation vor Ort vom 6.5.2020 und deren jeweiligen Aufnahmepositionen (purpurne Punkte)

## Resultat multitemporale $\operatorname{MLC}$



Abbildung 31: Karte der resultierenden Verteilung von Röhricht innerhalb des Untersuchungsgebietes mit einem RGB-Echtfarbenkomposit der Aufnahme des MSI von Sentinel 2 vom 16.2.2019 im Hintergrund (ESA, 2019)

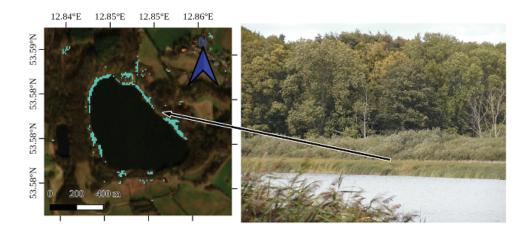


Abbildung 32: Detail aus der Karte in Abbildung 31 für den Kleinen Varchentiner See mit Aufnahme von der Lage vor Ort vom 20.8.2019

Unter der Annahme, dass die resultierende Verteilung von Röhricht dessen Vorkommen innerhalb des Untersuchungsgebietes korrekt wiedergibt, lassen sich - in Analogie zu dem vorausgegangenen Beispiel für immergrünen Nadel- und Mischwald - deren Eigenschaften in den Zwischenergebnissen und ihre Entwicklung im Rahmen des multitemporalen Verfahrens zur Einklassenfall ML Klassifikation genauer untersuchen. Dazu sind in Tabelle 6 die Minima, Maxima, Mittelwerte und Spannweiten der unitemporalen Verteilungen der Likelihood-Funktionen innerhalb und außerhalb der nach dem Endergebnis mit Röhricht bewachsenen Gebiete für die einzelnen Aufnahmen dargestellt. Darüber hinaus sind in Tabelle 7 in Analogie zu Tabelle 6 die gleichen Parameter für multitemporale Verteilungen innerhalb und außerhalb der nach dem Endergebnis mit Röhricht bewachsenen Gebiete zusammengetragen. Bei den multitemporalen Verteilungen wird zeitlich von der Aufnahme vom 28.11.2018 ausgegangen und für jedes weitere multitemporale Resultat zusätzlich die jeweils zeitlich nachfolgende Aufnahme berücksichtigt.

Tabelle 6: einige statistische Kenngrößen für die resultierenden Verteilungen der logarithmierten Likelihood-Funktionen innerhalb der nach dem Endergebnis mit bzw. nicht mit Röhricht bewachsenen Gebiete in den unitemporalen Zwischenergebnissen als absolute und relative Werte (in %, bezogen auf das absolute Maximum und den gesamten Wertebereich)

Verteilung	Min	imum	Mit	telwert	Maximum		Span	nweite
(Röhricht)	${\it absolut}$	relativ	absolut	relativ	absolut	relativ	absolut	relativ
28.11.2018	-38.50	$-4.84 \times 10^{-6}$	-18.13	$-1.37 \times 10^{-6}$	-10.07	0.00	28.43	$4.84 \times 10^{-6}$
16.2.2019	-41.67	$-1.08 \times 10^{-4}$	-18.45	$-2.37 \times 10^{-5}$	-11.99	$-1.08 \times 10^{-7}$	29.68	$1.08 \times 10^{-4}$
26.7.2019	-33.68	-0.04	-14.03	$-7.62 \times 10^{-3}$	-9.86	$-1.29 \times 10^{-5}$	23.82	0.04
25.8.2019	-40.48	-0.07	-19.75	-0.02	-13.81	$-1.93 \times 10^{-4}$	26.67	0.07
(kein								
Röhricht)								
28.11.2018	$-5.88 \times 10^{8}$	-100.00	-862	$-1.40 \times 10^{-4}$	-10	0.00	$5.88 \times 10^{8}$	100.00
16.2.2019	$-2.74 \times 10^7$	-100.00	-381	$-1.34 \times 10^{-3}$	-12	0.00	$2.74 \times 10^7$	100.00
26.7.2019	$-5.48 \times 10^4$	-100.00	-137.96	-0.23	-9.85	0.00	$5.48 \times 10^4$	100.00
25.8.2019	$-3.72 \times 10^4$	-100.00	-122.75	-0.29	-13.74	0.00	$3.71 \times 10^4$	100.00

Tabelle 7: einige statistische Kenngrößen für die resultierenden Verteilungen der logarithmierten Likelihood-Funktionen innerhalb der nach dem Endergebnis mit bzw. nicht mit Röhricht bewachsenen Gebiete in multitemporalen Ergebnissen als absolute und relative Werte (in %, bezogen auf das absolute Maximum und den gesamten Wertebereich)

Verteilung	Min	imum	Mittelwert		Maximum		Spannweite	
(Röhricht)	absolut	relativ	absolut	relativ	absolut	relativ	absolut	relativ
28.11.2018	siehe Tabelle 6							
+16.2.2019	-64.89	$-5.31 \times 10^{-6}$	-47.61	$-2.37 \times 10^{-6}$	-33.68	0.00	31.21	$5.31 \times 10^{-6}$
+26.7.2019	-80.67	$-5.42 \times 10^{-6}$	-66.23	$-2.96 \times 10^{-6}$	-48.82	0.00	31.85	$5.42 \times 10^{-6}$
+25.8.2019	-100.67	$-5.41 \times 10^{-6}$	-91.49	$-3.85 \times 10^{-6}$	-68.85	0.00	31.82	$5.41 \times 10^{-6}$
(kein								
Röhricht)								
28.11.2018	siehe Tabelle 6							
+16.2.2019	$-5.88 \times 10^{8}$	-100.00	-1254	$-2.10 \times 10^{-4}$	-34	0.00	$5.88 \times 10^{8}$	100.00
+26.7.2019	$-5.88 \times 10^{8}$	-100.00	-1396	$-2.30 \times 10^{-4}$	-51	0.00	$5.88 \times 10^{8}$	100.00
+25.8.2019	$-5.88 \times 10^{8}$	-100.00	-1524	$-2.50\times10^{-4}$	-101	$-1.00 \times 10^{-5}$	$5.88 \times 10^{8}$	99.99

Die Mittelwerte der Verteilungen innerhalb der nach dem Endergebnis mit Röhricht bewachsenen Gebiete liegen in allen unitemporalen Zwischenergebnissen näher am absoluten Maximum, als die Mittelwerte der Verteilungen außerhalb der mit Röhricht bewachsenen Gebiete (siehe Tabelle 6). Die Wertebereiche der Verteilungen innerhalb der mit Röhricht bewachsenen Gebiete machen in allen unitemporalen Einzelergebnisse nur sehr kleine Bruchteile der gesamten Spannweite aus. Die Anteile sind in den Resultaten für die Aufnahmen vom 28.11.2018 und 16.2.2019 kleiner, als für die Aufnahmen vom 26.7.2019 und 25.8.2019. In allen unitemporalen Zwischenergebnissen liegen die Gebiete mit den höchsten Werten der Likelihood-Funktion außerhalb der Gebiete, die nach dem Endergebnis mit Röhricht bewachsen sind. Daher nehmen die Spannweiten der Verteilungen innerhalb der nicht mit Röhricht bewachsenen Gebiete jeweils den gesamten Wertebereich der Likelihood-Funktion ein und der Überschneidungsbereich zwischen den Verteilungen innerhalb der mit bzw. nicht mit Röhricht bewachsenen Gebiete wird maximal. Die Anzahl der Bildpunkte, für welche die Likelihood-Funktion maximal wird, ohne dass sie in der Verteilung der mit Röhricht bewachsenen Gebiete aus dem Endergebnis vorkommen, liegt für alle unitemporalen Einzelergebnisse unter 5 und ist damit sehr gering. Eine mögliche Ursache für diese Ausreißer kann ein Bewuchs mit Röhricht sein, der in anderen Aufnahmen durch darüber wachsende Vegetation verdeckt wird. Es kann sich auch um die Auswirkungen von Zerstörung oder Neuentstehung von Beständen handeln. Da jedoch nur sehr wenige Primitive betroffen sind und kein Muster in deren Auftreten erkennbar ist, wird es sich in diesem Fall um statistische Ausreißer in den unitemporalen Resultaten handeln, die mithilfe des multitemporalen Ansatzes ausgeschlossen werden können.

Da die Überschneidungen zwischen den Verteilungen innerhalb der im Endergebnis mit bzw. nicht mit Röhricht bewachsenen Gebiete in allen unitemporalen Einzelergebnissen die gesamte Spannweite der Verteilung innerhalb der mit Röhricht bewachsenen Gebiete umfassen und die Spannweiten der Verteilungen innerhalb der mit Röhricht bewachsenen Gebiete für die Aufnahmen vom 28.11.2018 und 16.2.2019 kleinere Anteile des gesamten Wertebereichs ausmachen, sind für diese Aufnahmen auch die Bereiche, in denen sich die beiden Verteilungen überschneiden, im Verhältnis zur gesamten Spannweite geringer. In dem unitemporalen Einzelergebnis für die Aufnahme vom 28.11.2018 fallen 468'322 Primitive, die nicht Teil der mit Röhricht bewachsenen Gebiete im Endergebnis sind, in den Überschneidungsbereich zwischen beiden Verteilungen. In dem Resultat für die Aufnahme vom 16.2.2019 sind es 476'634, vom 26.7.2019 1'211'300 und vom 25.8.2018 1'293'011. Dieser hohe Anteil "falsch positiver" Treffer weist darauf hin, dass in den unitemporalen Einzelergebnissen einigen mit Röhricht bewachsenen Gebieten im Vergleich mit dem multitemporalen Endergebnis verhältnismäßig geringere Werte der Likelihood-Funktion zugeordnet werden. Es zeigt sich, dass ein Großteil der "falsch positiven" Treffer in den Resultaten für die Aufnahmen vom 26.7.2019 und vom 25.8.2018 auf andere Vegetationsformen (insbesondere Wälder) entfällt. Zu dieser Zeit zeigt sich das Schilf in saftigem grün und ähnelt damit vielen anderen Pflanzen (siehe Fotografie in Abbildung 32), was eine erschwerte Abgrenzbarkeit erklären kann. Darüber hinaus ist der relative Abstand der Minima der Verteilungen innerhalb der mit Röhricht bewachsenen Gebiete vom absoluten Maximum in den Resultaten für die Aufnahmen vom 26.7.2019 und 25.8.2019 größer, als für die Aufnahmen vom 28.11.2018 und 16.2.2019. Das gleiche gilt auch für die relativen Abstände der Mittelwerte der Verteilungen vom absoluten Maximum. Folglich sind die Werte der Verteilungen innerhalb der mit Röhricht bewachsenen Gebiete in den unitemporalen Resultaten für die Aufnahmen vom 26.7.2019 und 25.8.2019 verhältnismäßig kleiner, als für die Aufnahmen vom 28.11.2018 und 16.2.2019. Die absoluten Abstände zwischen den Mittelwerten der Verteilungen innerhalb der mit bzw. nicht mit Röhricht bewachsenen Gebiete sind in den Resultaten für die Aufnahmen vom 28.11.2018 und 16.2.2019 größer, als für die Aufnahmen vom 26.7.2019 und 25.8.2019. Wobei es sich bei den relativen Abständen der Mittelwerte gerade umgekehrt verhält. Bei den multitemporalen Resultaten wächst die Spannweite der Verteilung der Werte der Likelihood-Funktion mit jeder zusätzlich berücksichtigten Aufnahme weiter an. In diesem Beispiel ist das Wachstum

Bei den multitemporalen Resultaten wächst die Spannweite der Verteilung der Werte der Likelihood-Funktion mit jeder zusätzlich berücksichtigten Aufnahme weiter an. In diesem Beispiel ist das Wachstum verhältnismäßig sehr gering (in Tabelle 7 nicht erkennbar) und kleiner, je mehr Aufnahmen berücksichtigt werden. Die absoluten und relativen Spannweiten der Verteilungen innerhalb der im Endergebnis mit Röhricht bewachsenen Gebiete nehmen in den multitemporalen Resultaten bis zur dritten berücksichtigten Aufnahme vom 26.7.2019 zu. Dabei wird die Zunahme der Spannweite geringer, je mehr Aufnahmen berücksichtigt werden. Schließlich nimmt die Spannweite (sowohl absolut, als auch relativ) mit der vierten

berücksichtigten Aufnahme vom 25.8.2019 im multitemporalen Endergebnis wieder leicht ab. In den multitemporalen Resultaten existieren keinerlei Ausreißer außerhalb der mit Röhricht bewachsenen Gebiete näher am absoluten Maximum, als das Maximum der Verteilung innerhalb der mit Röhricht bewachsenen Gebiete. Mit jeder weiteren berücksichtigten Aufnahme wird der Überschneidungsbereich zwischen den Verteilungen innerhalb der mit bzw. nicht mit Röhricht bewachsenen Gebiete kleiner. Da die Verteilung der mit Röhricht bewachsenen Gebiete aus dem Endergebnis durch Clustern mit dem k-Means Algorithmus hervorgegangen ist, existiert im multitemporalen Endergebnis zwischen den beiden Verteilungen kein Überschneidungsbereich. Sowohl die absoluten, als auch die relativen Mittelwerte beider Verteilungen entfernen sich in den multitemporalen Resultaten mit jeder zusätzlich berücksichtigten Aufnahme weiter voneinander.

Schließlich wird auch die Verteilung von Röhricht innerhalb des Untersuchungsgebietes als Resultat der multitemporalen Einklassenfall ML Klassifikation genutzt, um einen Random Forest aus 100 Bäumen auf die Klassifikation des gesamten Ausgangsdatenvolumens (Stacked Vector Approach) zu trainieren. Die Trainingsdaten werden in möglichst homogenen Gebieten definiert, in denen die resultierende Verteilung von Röhricht gut mit den Erfahrungen aus den Begehungen und der Ortskenntnis übereinstimmt. Das Resultat der multitemporalen RF Klassifikation ist mit den zugrunde gelegten Trainingsdaten in Abbildung 33 dargestellt.

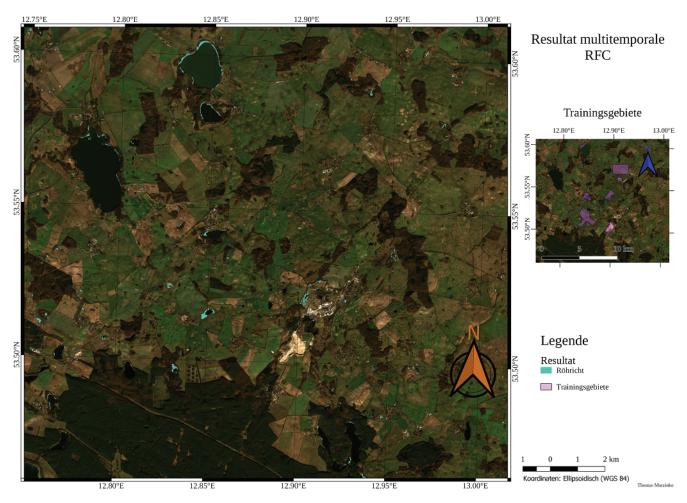


Abbildung 33: Karte der aus der multitemporalen RF Klassifikation resultierenden Verteilung von Röhricht innerhalb des Untersuchungsgebietes mit einem RGB-Echtfarbenkomposit der Aufnahme des MSI von Sentinel 2 vom 16.2.2019 im Hintergrund (ESA, 2019) und einer Karte der Trainingsgebiete

### 8 Fernerkundung von Weiden

Das Vorkommen zahlreicher Pflanzenarten ist in vielen Fällen (bis zu einem bestimmten Grad) an charakteristische Umweltbedingungen gebunden. Auf Grundlage dieser Tatsache kann die Fernerkundung von Heilpflanzen auch als eine Aufgabe der Standortfindung interpretiert und das zusätzliche Priori Wissen um das Vorkommen bestimmter Spezies unter typischen Umweltbedingungen genutzt werden, um insbesondere die Resultate leistungsschwacher Klassifikatoren zu verbessern.

Die Weide (Gattung Salix) findet in der Volksmedizin als fiebersenkende, schmerzstillende und harntreibende Heilpflanze Anwendung. Die fiebersenkende und schmerzstillende Wirkung der Weidenrinde wird auf die enthaltene Salicylsäure zurückgeführt. Weiden kommen verstärkt in feuchten Gebieten vor und stehen oft in der Nähe von Wasser (Pahlow, 2013). Schon in einem alten keltischen Gedicht heißt es "Wähle die Weide am Bach" und "Salix", der lateinische Name der Gattung in der Botanik, leitet sich aus dem keltischen "sallis" für "nah am Wasser" ab (Mania, 2018). Für das Anwendungsbeispiel zur Fernerkundung von Weiden soll dieses Priori Wissen über einen Ansatz mit unscharfen Mengen in Analogie zur Standortwahl (Eastman, 1999) berücksichtigt werden, um die Resultate multitemporaler Euklidischer (Spektraler) Distanz (siehe Abschnitt 5.7.1 "Euklidische (Spektrale) Distanz Klassifikation") und korrelationskoeffizienten-basierter Klassifikationen (siehe Abschnitt 5.7.4 "korrelationskoeffizientenbasierte Klassifikation") zu verbessern. Dabei wird hier der Ansatz zur multitemporalen Klassifikation ebenfalls über das Konzept unscharfer Mengen realisiert.

Der grundlegende Ablauf der Analyse ist den beiden vorausgegangenen Beispielen identisch (siehe Abbildung 22). Wie bei dem Beispiel zur Fernerkundung von Röhricht (siehe Abschnitt 7 "Fernerkundung von Röhricht") konzentriert sich die Analyse auf die Fernerkundung einer Zielspezies innerhalb des Untersuchungsgebietes bei Kraase. Bei den multitemporalen Klassifikationen werden die Aufnahmen des MSI von Sentinel 2 vom 28.11.2018, 16.2.2019, 26.7.2019 und 25.8.2019 berücksichtigt (siehe Tabelle 5 für die Merkmalsauswahl). Die Abgrenzung der Trainingsgebiete erfolgt in Analogie zu den vorausgegangenen Beispielen. Die Trainingsdaten für Weiden werden in zwei unterschiedlichen Wiesen nördlich (bei Hoppenbarg) und südlich von Kraase definiert. Die Trainingsgebiete sind in Abbildung 34 mit Fotografien von der Situation vor Ort dargestellt.

Für die multitemporale Klassifikation werden die unitemporale Euklidische (Spektrale) Distanz- und die korrelationskoeffizienten-basierte Klassifikation über einen Ansatz mit unscharfen Mengen zu multitemporalen Verfahren erweitert. Dabei werden über den Ansatz mit unscharfen Mengen in den Endergebnissen zusätzliche Priori Informationen berücksichtigt.

Zu Vergleichszwecken und weiteren Untersuchung der einzelnen Verfahren wird zunächst eine multitemporale Maximum Likelihood Klassifikation für Weiden in Analogie zu den vorausgegangenen Beispielen durchgeführt und eine Auswahl aus dem Resultat als Trainingsdatensatz für eine Random Forest Klassifikation mit einem Entscheidungswald aus 100 Bäumen genutzt. Ausschnitte aus beiden Resultaten mit Bildern von der Situation vor Ort sind in Abbildung 35 dargestellt. Die Resultate verdeutlichen, dass sich mithilfe der multitemporalen Maximum Likelihood und Random Forest Klassifikation auch Ansammlungen von Pflanzen aufspüren lassen, die nur wenige Individuen umfassen und sich in den Aufnahmen über eine geringe Anzahl an Bildpunkten erstrecken. In Abbildung 35 sind als Beispiele verifizierte Treffer für zwei Gebiete dargestellt, in denen Weiden nur vereinzelt vorkommen. Links sind Treffer am Rande eines kleines Sumpfes auf einem Acker westlich von Kraase dargestellt. Bei diesem Sumpf kommen Weiden ausschließlich an dieser Stelle und in sehr geringer Anzahl vor (ein älterer Baum und etwa 5 junge). Rechts im Bild finden sich Treffer bei einem kleinen Wasserloch auf einem Acker nordöstlich von Kraase. Derartige Wasserlöcher kommen in der Umgebung häufig vor und sind in der Regel mit Weiden gesäumt. Beide Treffer finden sich sowohl in der aus der multitemporalen MLC, als auch aus der multitemporalen RFC resultierenden Verteilung für Weiden innerhalb des Untersuchungsgebietes.

## Gebiete zur Auswahl der Trainingsgebiete

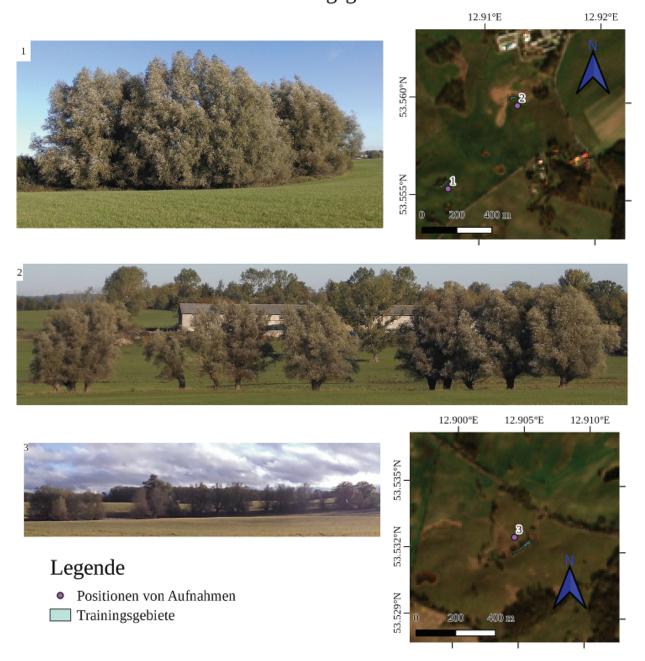


Abbildung 34: Trainingsgebiete für die Weiden-Analyse in der Aufnahme vom 16.2.2019 (RGB-Echtfarbenkomposite im Hintergrund; ESA, 2019) mit Bildern von der Situation vor Ort vom 6.10.2019 (1 und 2) und 3.11.2019 (3) und deren jeweiligen Aufnahmepositionen (purpurne Punkte)

Für die multitemporale Fernerkundung von Weiden über einen Ansatz basierend auf unscharfen Mengen wird hier von der Euklidischen (Spektralen) Distanz Klassifikation (vergleiche Abschnitt 5.7.1 "Euklidische (Spektrale) Distanz Klassifikation") im Einklassenfall ausgegangen. Für alle Primitive wird nach Gleichung 8 der euklidische (spektrale) Abstand  $d_{Weiden}(x)$  des zugehörigen Merkmalsvektors x zum Schwerpunktvektor der Trainingsdaten für Weiden im Merkmalsraum berechnet. In Analogie zu den vorausgegangenen Beispielen sollen die Zugehörigkeitswerte für Gebiete mit Weiden maximal (nahe 1) werden. Daher wird

das reziproke Abstandsquadrat  $(d_{Weiden}(\boldsymbol{x}))^{-2}$  (näher an 1, wenn der Euklidische (Spektrale) Abstand zum Schwerpunktvektor der Trainingsdaten für Weiden gering ist) als Zugehörigkeitsfunktion  $\mu_{Weiden}(\boldsymbol{x})$  einer unscharfen Menge für Weiden definiert. Die Zugehörigkeitswerte werden jeweils für jede Aufnahme berechnet und die resultierenden unitemporalen räumlichen Verteilungen von  $\mu_{Weiden}(\boldsymbol{x})$  mit der T-Norm (unscharfer Mengenschnitt) des T-T-Konorm-Paares aus algebraischem Produkt und algebraischer Summe zu einer multitemporalen Verteilung überlagert. Dies entspricht, in Analogie zu den vorangegangenen Beispielen zur multitemporalen Maximum Likelihood Klassifikation, der Multiplikation der unitemporalen Einzelergebnisse und legt eine Interpretation der multitemporalen Maximum Likelihood Klassifikation mit unscharfen Mengen nahe, wobei die Likelihood-Funktion die Rolle einer Zugehörigkeitsfunktion  $\mu_i(\omega)$  übernimmt (Bronstein, 2012).

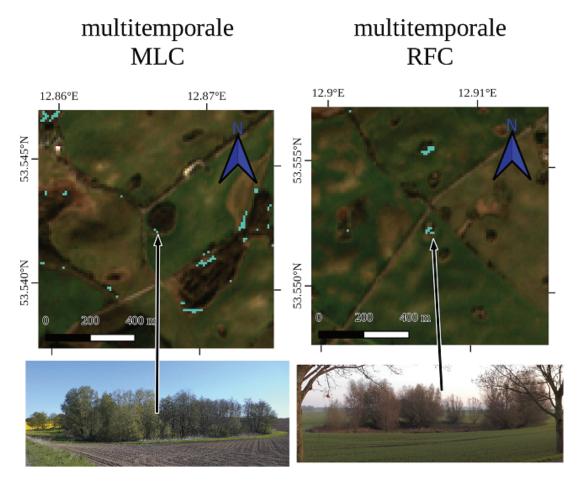


Abbildung 35: Ausschnitte aus den Resultaten für vereinzelte Weiden am Rand eines Sumpfes westlich (links) und Weiden am Rand eines Wasserloches nordöstlich (rechts) von Kraase: links: Resultat der Maximum Likelihood Klassifikation, rechts: Random Forest Klassifikation; blau-grüne Flächen: Weiden; Hintergrund: RGB-Echtfarbenkomposit der Sentinel 2 Aufnahme vom 16.2.2019 (ESA, 2019); mit Bildern von der Situation vor Ort (unten links: vom 6.5.2020 und rechts: vom 10.11.2019)

Zur Integration des zusätzlichen Prioriwissens wird das multitemporale Resultat über die algebraische T-Norm mit unscharfen Mengen für Waldgebiete mit einer Größe von mehr als 0.5 Hektar (binär (constraint): keine Weiden in Wäldern; aus der Nadel- und Mischwaldanalyse, siehe Abschnitt 6 "Fernerkundung von immergrünen Nadel- und Mischwald"), Wasserflächen (binär (constraint): keine Weiden auf Wasserflächen; über die multitemporale Anwendung von "water in wetlands" (WIW); LEFEBVRE, 2019), die Nähe zu Gewässern und feuchten Gebieten (reziproke Distanz (+1): Weiden kommen eher in der Nähe zum Wasser und in feuchten Gebieten vor; über die multitemporale Anwendung von WIW) und die Feuchtigkeit in

einem Gebiet (mittlerer positiver NDMI aus allen Aufnahmen: Weiden kommen eher in feuchten Gebieten vor) überlagert.

Ausschnitte aus dem Resultat der multitemporalen Einklassenfall Klassifikation und dem Endergebnis nach Berücksichtigung des Prioriwissens sind in Abbildung 36 dargestellt. Da die absoluten Maxima in beiden Verteilungen stark durch Ausreißer beeinflusst werden, sind jeweils die 9 Prozent der Zugehörigkeitswerte zwischen dem 90 und 99 Perzentil nahe dem absoluten Maximum dargestellt. Das Resultat lässt sich durch die Berücksichtigung von zusätzlichem Prioriwissen deutlich verbessern. So zählen z.B. die zahlreichen Treffer in Wäldern zu den offensichtlichsten Fehlern im Resultat der multitemporalen Klassifikation (Weiden kommen nur sehr selten in Wäldern vor), welche sich über den Ansatz mit unscharfen Mengen nahezu vollständig ausschließen lassen. Auf der anderen Seite wird so das Endergebnis in Richtung des Priori Wissens verschoben, wodurch zum Teil neue Fehlerquellen entstehen können. Beispielsweise werden hier Fehltreffer auf feuchten Wiesen verstärkt, hohe Zugehörigkeitswerte für viele Vegetationstypen in Gewässernähe erreicht und Weiden an trockenen Standorten vernachlässigt. Diese Verschiebung erschwert auch das Auffinden von Strukturen mit guter Korrespondenz mit Weiden mithilfe von Cluster-Algorithmen (siehe Abbildung 37). Gebiete mit Weiden verteilen sich oft über viele verschiedene Cluster bzw. Subcluster innerhalb größerer Strukturen.

Als Alternative zur multitemporalen Euklidischen (Spektralen) Distanz Klassifikation wird das Verfahren analog mit dem "Spectral Correlation Mapper" (SCM; siehe Abschnitt 5.7.4 "korrelationskoeffizientenbasierte Klassifikation (Spectral Correlation Mapper - SCM)"; DE CARVALHO JÚNIOR, 2000) als grundlegendem unitemporalen Klassifikator umgesetzt. Da der Korrelationskoeffizient zwischen 0 und 1 liegt und die Werte bei der Anwendung als Klassifikator umso näher bei 1 liegen, je stärker die Ähnlichkeit mit der Zielklasse ausgeprägt ist, lässt er sich hier direkt als Zugehörigkeitsfunktion definieren. Ein Ausschnitt aus einer Verteilung von Clustern mit guter Korrespondenz mit Weiden in der unter Berücksichtigung des Priori Wissens resultierenden multitemporalen Verteilung der Zugehörigkeitswerte ist in Abbildung 38 dargestellt. Dabei werden für das Clustern der resultierenden Verteilung der Zugehörigkeitswerte die selben Parameter (Startwert, Anzahl der Klassen und Iterationen) verwendet, wie bei dem Beispiel für die multitemporale Euklidische (Spektrale) Distanz Klassifikation mit Integration von Priori Wissen (Ausschnitt aus dem Resultat in Abbildung 37). Auf diese Weise wird die Vergleichbarkeit der beiden Endergebnisse sichergestellt. Es zeigt sich, dass die Cluster mit guter Korrespondenz mit Weiden im Endergebnis der multitemporalen korrelationskoeffizienten-basierten Klassifikation deutlich umfangreicher (inklusive Fehltreffern) ausfallen, als bei der multitemporalen Euklidischen (Spektralen) Distanz Klassifikation. So machen die Cluster 1, 2, 3, 6, 11 und 10 aus Abbildung 38 im Bezug auf das gesamte Untersuchungsgebiet einen flächenmäßigen Anteil von etwa 27.11% aus, während es bei den Clustern 2, 5, 6, 8, 10 und 12 aus Abbildung 37 nur 13.71% sind. Die Abgrenzung von Gebieten mit Weiden aus dem Endergebnis lässt sich durch die weitere Aufteilung der Cluster in Subcluster verbessern.

Unter der Annahme, dass die aus der multitemporalen MLC resultierende Verteilung von Weiden innerhalb des Untersuchungsgebietes (siehe Abbildung 35 links für einen Ausschnitt) die vorherrschenden Verhältnisse korrekt wiederspiegelt, lassen sich die aus den multitemporalen Euklidischen (Spektralen) Distanz und korrelationskoeffizienten-basierten Klassifikationen mit Berücksichtigung von Priori Wissen resultierenden Verteilungen der Zugehörigkeitswerte auf dieser Grundlage weiter untersuchen. Die Spannweite der Verteilung der Zugehörigkeitswerte innerhalb der nach der multitemporalen MLC mit Weiden bewachsenen Gebiete reicht bei dem Endergebnis der multitemporalen Euklidischen (Spektralen) Distanz Klassifikation von 0 bis etwa  $1.37 \times 10^{-21}$  und bei der multitemporalen SCM Klassifikation von 0 bis etwa 0.048. Da die Spannweite in beiden Fällen bei 0 beginnt, was nach dem Ansatz mit unscharfen Mengen für Gebiete der Fall ist, in denen keine Weiden vorkommen, wird das Vorkommen von Weiden in einem Teil der Gebiete, die nach der multitemporalen MLC mit Weiden bewachsen sind, in beiden Resultaten ausgeschlossen. Der flächenmäßige Anteil dieser Gebiete macht bei der multitemporalen Euklidischen (Spektralen) Distanz Klassifikation etwa 10.76% und bei der multitemporalen SCM Klassifikation etwa 13.79% der nach der multitemporalen MLC mit Weiden bewachsenen Gebiete aus. Etwa 78% der bei beiden Klassifikationen zusammen ausgeschlossenen Gebiete sind in den jeweiligen Endergebnissen identisch.

Auch in der aus der multitemporalen RF Klassifikation resultierenden Verteilung von Weiden innerhalb des

Untersuchungsgebietes (siehe Abbildung 35 rechts für einen Ausschnitt) wird das Vorkommen von Weiden in einigen Gebieten ausgeschlossen, die nach dem Resultat der multitemporalen MLC mit Weiden bewachsen sind. Im Vergleich zu den Resultaten der multitemporalen Klassifikationen mit unscharfen Mengen ist deren Anteil mit etwa 94.92% bei der multitemporalen RF Klassifikation am größten. Beim Vergleich der durch die multitemporale RF Klassifikation ausgeschlossenen Gebiete mit denen der Klassifikationen mit unscharfen Mengen stellt sich heraus, dass 100% der durch die multitemporale Euklidische (Spektrale) Distanz Klassifikation ausgeschlossenen Gebiete auch durch die multitemporale RF Klassifikation ausgeschlossen werden. Bei der multitemporalen SCM Klassifikation sind es etwa 98.57%.

Die Maxima beider aus den multitemporalen Klassifikationen mit unscharfen Mengen resultierenden Verteilungen der Zugehörigkeitswerte innerhalb der (nach der multitemporalen MLC) mit Weiden bewachsenen Gebiete entsprechen nicht den absoluten Maxima der jeweiligen Verteilungen der Zugehörigkeitswerte im gesamten Untersuchungsgebiet. Bei der multitemporalen Euklidischen (Spektralen) Distanz Klassifikation haben lediglich 5 Gebiete (etwa  $1.66 \times 10^{-4}\%$  der Gesamtfläche des Untersuchungsgebietes) außerhalb der mit Weiden bewachsenen Flächen größere Zugehörigkeitswerte, als die innerhalb. Bei der multitemporalen SCM Klassifikation sind es mit 6'074 immerhin etwa 0.2% des gesamten Untersuchungsgebietes. In der aus der multitemporalen Euklidischen (Spektralen) Distanz Klassifikation resultierenden Verteilung der Zugehörigkeitswerte erreichen etwa 44.88% aller Gebiete innerhalb des gesamten Untersuchungsgebietes Zugehörigkeitswerte größer als das Minimum der Verteilung innerhalb der mit Weiden bewachsenen Gebiete (ausgeschlossen der Gebiete mit dem Zugehörigkeitswert 0). In der Verteilung aus der multitemporalen SCM Klassifikation sind es mit rund 40.52% etwas weniger. Werden dabei, um Ausreißer mit geringen Zugehörigkeitswerten auszuschließen, nur die 99% der mit Weiden bewachsenen Gebiete mit den größten Zugehörigkeitswerten betrachtet, so verkleinern sich die Anteile bei der aus der multitemporalen Euklidischen (Spektralen) Distanz Klassifikation resultierenden Verteilung auf etwa 21.52% und bei der Verteilung aus der multitemporalen SCM Klassifikation auf etwa 28.82% des gesamten Untersuchungsgebietes. Folglich erreichen einige der mit Weiden bewachsenen Gebiete in beiden resultierenden Verteilungen Zugehörigkeitswerte nahe dem absoluten Maximum und die Zugehörigkeitswerte aller mit Weiden bewachsenen Gebiete liegen eher nahe dem absoluten Maximum. Nach dem Endergebnis der multitemporalen MLC sind jedoch lediglich rund 0.46% der Gesamtfläche des Untersuchungsgebietes mit Weiden bewachsen. Jedoch sind die Zugehörigkeitswerte von mit Weiden bewachsenen Gebieten in den aus den multitemporalen Klassifikationen mit unscharfen Mengen resultierenden Verteilungen über ein Intervall gestreut, auf dem sich auch die Zugehörigkeitserte von etwa zwischen 20% und 45% aller anderen Gebiete innerhalb des gesamten Untersuchungsgebietes wiederfinden. Daher ist von einer starken Durchmischung der Zugehörigkeitswerte von mit Weiden bewachsenen Gebieten mit denen anderer Gebiete auszugehen, wodurch sich die Abgrenzung der Zielstrukturen erschwert.

## multitemporale Euklidische (Spektrale) Distanz Klassifikation

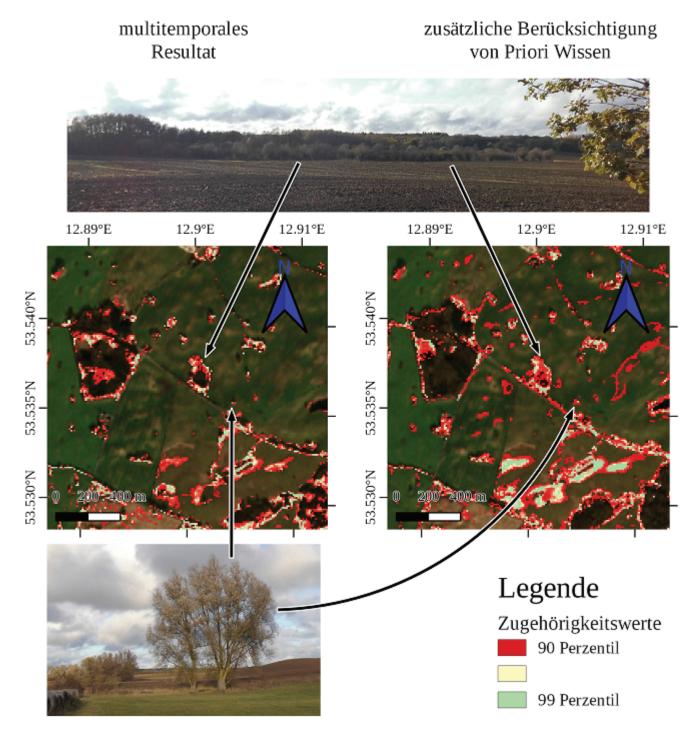


Abbildung 36: Ausschnitte aus den Resultaten: links: räumliche Verteilung der aus der multitemporalen Einklassenfall Klassifikation resultierenden Zugehörigkeitswerte, rechts: Verteilung der Zugehörigkeitswerte nach der zusätzlichen Berücksichtigung des Priori Wissens, im Hintergrund: RGB-Echtfarbenkomposit der Aufnahme vom 16.2.2019 (ESA, 2019) für den Südosten von Kraase, mit Bildern von der Situation vor Ort (vom 3.11.2019)

## Clusterverteilung für Zugehörigkeitswerte

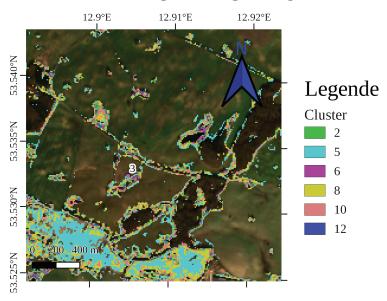


Abbildung 37: Ausschnitt aus einer Verteilung von Clustern mit guter Korrespondenz mit Weiden (bestimmt mit dem k-Means Algorithmus) in den Zugehörigkeitswerten des Endergebnisses der multitemporalen Euklidischen (Spektralen) Distanz Klassifikation unter Berücksichtigung von Priori Informationen für das Gebiet aus Abbildung 36, unterschiedliche Farben indizieren unterschiedliche Cluster; Hintergrund: RGB-Echtfarbenkomposit der Aufnahme vom 16.2.2019 (ESA, 2019)

## Clusterverteilung für Zugehörigkeitswerte

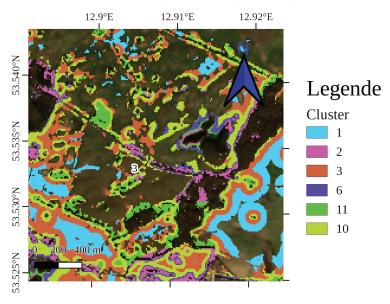


Abbildung 38: Ausschnitt aus einer Verteilung von Clustern mit guter Korrespondenz mit Weiden (bestimmt mit dem k-Means Algorithmus) in den Zugehörigkeitswerten des Endergebnisses der multitemporalen SCM Klassifikation unter Berücksichtigung von Priori Informationen für das Gebiet aus Abbildung 36, unterschiedliche Farben indizieren unterschiedliche Cluster; Hintergrund: RGB-Echtfarbenkomposit der Aufnahme vom 16.2.2019 (ESA, 2019)

### 9 Fernerkundung von Cassia sieberiana

Als erstes Anwendungsbeispiel für das Untersuchungsgebiet bei Bobo Dioulasso und Banfora soll die multitemporale MLC zur Fernerkundung von Cassia sieberiana genutzt werden. Die Spezies Cassia sieberiana (ein kleiner bis mittelgroßer Strauch bzw. Baum) wird in Burkina Faso als traditionelles Heilmittel z.B. bei Malaria und Gelbsucht eingesetzt (Thiombiano, 2010a). Für diese Pflanze sind antimikrobielle (Olutayo, 2012), antiplasmodiale (Aliyu, 2013) und antivirale (Leteane, 2012) Wirkungen dokumentiert.

Die multitemporale Fernerkundung von Cassia sieberiana wird in Analogie zu dem Beispiel für die Fernerkundung von Röhricht durchgeführt (siehe Abschnitt 7 "Fernerkundung von Röhricht"). Der Ablauf entspricht im Wesentlichen dem in Abbildung 22 skizzierten Workflow für die Fernerkundung von immergrünen Nadel- und Mischwald. Dabei wird die MLC hier wie bei der Fernerkundung von Röhricht nur für eine Spezies - Cassia sieberiana - und multitemporal durchgeführt. Die Validierung des Resultats gestaltet sich für Cassia sieberiana schwierig. Daher lassen sich keine Gebiete, in denen die resultierende Verteilung gut die tatsächlichen Verhältnisse widerspiegelt, zur Definition der Trainingsdaten für einen Random Forest finden. Deshalb wird im Anschluss an die multitemporale MLC keine zusätzliche Random Forest Klassifikation vorgenommen. Die multitemporale MLC wird wie bei der Fernerkundung von Röhricht über 4 Aufnahmen des MSI von Sentinel 2 durchgeführt (ESA, 2019). Dabei handelt es sich um die Aufnahmen vom 13.1.2019, 28.1.2019, 22.2.2019 und 14.3.2019 für das Untersuchungsgebiet bei Bobo Dioulasso. In Tabelle 8 sind für jede Aufnahme die Merkmale aus der Merkmalsauswahl nach minimaler Korrelation unter einem Schwellwert (siehe Abschnitt 5.4 "Merkmalsreduktion") zusammengetragen. Dabei wird für jede Merkmalsauswahl ein einheitlicher Schwellwert von 0.7 für den Korrelationskoeffizienten genutzt.

Tabelle 8: Spektralbänder und spektrale Indizes der Merkmalsauswahl für die Aufnahmen des MSI von Sentinel 2 vom 13.1.2019, 28.1.2019, 22.2.2019 und 14.3.2019 für das Untersuchungsgebiet bei Bobo Dioulasso

Aufnahme vom	Spektralbänder	spektrale Indizes
13.1.2019	B2, B8	LCI, MCARI, NDMI
28.1.2019	B5	LCI, MCARI, NDMI
22.2.2019	B5	LCI, MCARI, NDMI
14.3.2019	B1, B7	LCI, MCARI, NDMI

Im Gegensatz zu den vorausgegangenen Anwendungsbeispielen für das Untersuchungsgebiet bei Kraase liegt für Burkina Faso keine Informationsgrundlage für die Abgrenzung der Trainingsgebiete vor. Es sind nur wenige Informationen über die Zielspezies und deren Verteilungen innerhalb des Untersuchungsgebietes oder Teilen davon bekannt. Auch die Möglichkeit, die Gegebenheiten vor Ort in Begehungen festzustellen, ist für das Untersuchungsgebiet bei Bobo Dioulasso nicht gegeben. Daher muss zur Definition der Trainingsdaten auf alternative Informationsquellen zurückgegriffen werden.

Die Trainingsdaten für Cassia sieberiana werden auf der Grundlage von georeferenzierten Daten über das Auftreten der Spezies von der Global Biodiversity Information Facility (GBIF) definiert (GBIF, 2020). Die Daten stammen aus vielen unterschiedlichen Quellen und sind nicht vollständig georeferenziert. Zur Definition der Trainingsdaten werden nur die georeferenzierten Punktdaten genutzt. Die Genauigkeit der Koordinaten ist für die Beobachtungen innerhalb des Untersuchungsgebietes nicht dokumentiert. Innerhalb des Untersuchungsgebietes wird Cassia sieberiana bisher an 41 unterschiedlichen Orten beobachtet. Das Beobachtungsdatum ist für die überwiegende Mehrzahl der Beobachtungen dokumentiert: sie stammen aus einem Zeitraum zwischen 1977 und 2002. Für den Großteil der Daten ist 2002 als Erfassungsdatum angegeben. Da es sich bei der Vegetation um ein dynamisches System handelt und die jüngsten Beobachtungen 17 Jahre in der Vergangenheit liegen, kann die Zielspezies inzwischen von einigen Beobachtungsorten verschwunden sein.

Wegen der großen Unsicherheiten bei der Definition der Trainingsdaten und ihres äußerst kleinen Umfangs ist von einem sehr ungenauen Endergebnis auszugehen. In Abbildung 39 sind die 10% der Gebiete innerhalb des Untersuchungsgebietes mit den höchsten Werten der resultierenden multitemporalen Likelihood-Funktion zusammen mit den Beobachtungspositionen aus den GBIF-Daten dargestellt.

#### Resultat multitemporale MLC: Cassia sieberiana

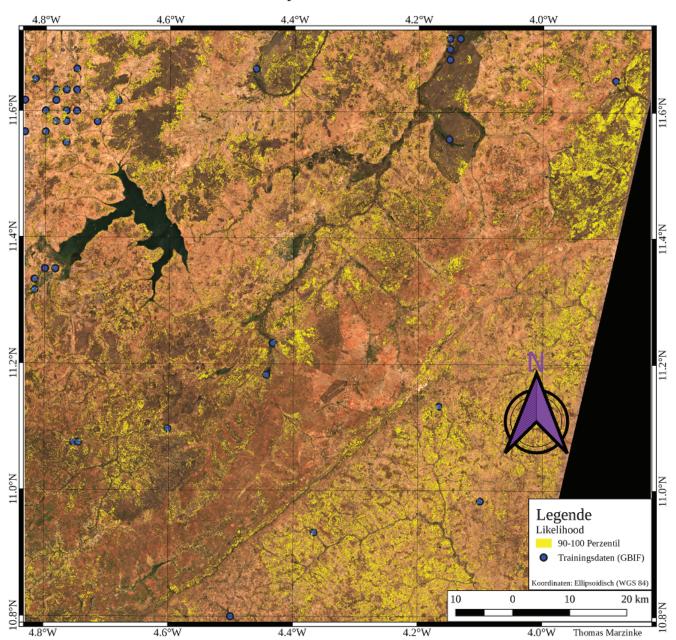


Abbildung 39: Resultat der multitemporalen MLC für Cassia sieberiana (10% der Gebiete mit den höchsten Werten der Likelihood-Funktion innerhalb des Untersuchungsgebietes) mit Beobachtungspositionen aus den GBIF-Daten (GBIF, 2020) und RGB-Echtfarbenkomposit der Aufnahme vom 22.2.2019 (ESA, 2019) im Hintergrund

Im Endergebnis finden sich in dicht bewachsenen Gebieten wie z.B. Teilen der Mare aux Hippopotames oder des Téré-Waldes weniger Gebiete mit hohen Werten der Likelihood-Funktion. Die Validierung des Resultats gestaltet sich wegen des Informationsmangels schwierig. Auf einigen Flächen kann die Anwesen-

heit von Cassia sieberiana mit hoher Sicherheit ausgeschlossen werden. Dazu zählen Wasserflächen, das dicht bebaute Stadtgebiet von Bobo Dioulasso und Flächen ohne Bewuchs. Innerhalb von Bobo Dioulasso und auch auf dem Flughafen der Stadt fallen nur sehr wenige Gebiete unter die 10% mit den größten Werten der Likelihood-Funktion. Auf offenen Wasserflächen sind keine "Treffer" und nur sehr wenige auf dünn oder nicht bewachsenen Flächen zu finden.

## 10 Fernerkundung von Vitellaria paradoxa

Im Gegensatz zu Cassia sieberiana liegen für Vitellaria paradoxa zusätzlich Daten über die Auftrittswahrscheinlichkeit der Spezies innerhalb des Untersuchungsgebietes bei Bobo Dioulasso vor. Daher soll sich das abschließende Anwendungsbeispiel mit der Fragestellung befassen, wie derartige Zusatzinformationen in der Analyse genutzt werden können.

Der Shea- oder Karitébaum (Vitellaria paradoxa) wird in der traditionellen Medizin zur Behandlung von Bluthochdruck, schlecht heilenden Wunden, Gelbsucht, Hämorrhoiden und Fieber genutzt (Thiombiano, 2010a). Für Vitellaria paradoxa ist in der Region des Gouvernements Hauts-Bassins auch die Anwendung bei neurologischen Störungen, Hexerei, Halluzinationen und Bewusstseinsverlust belegt (Kinda, 2017). Darüber hinaus dienen die Frucht und das aus den Kernen gewonnene Fett als Nahrungsmittel (Thiombiano, 2010b).

Im Gegensatz zu den vorausgegangenen Anwendungsbeispielen erlauben die Zusatzinformationen über die räumliche Verteilung der Auftrittswahrscheinlichkeit von Vitellaria paradoxa weitergehende Analysen. So wird hier zusätzlich zur multitemporalen Einklassenfall Klassifikation untersucht, welches der unter Abschnitt 5.7 "Klassifikationsalgorithmen" vorgestellten Klassifikationsverfahren für den Mehrklassenfall am besten geeignet ist, um die Verteilung der Auftrittswahrscheinlichkeit in einer unitemporalen Klassifikation auf der Grundlage einer zufälligen Stichprobe zu rekonstruieren. Dabei wird für die Random Forest Klassifikation neben dem unitemporalen auch ein multitemporales Verfahren zur Klassifikation genutzt, um die Resultate beider Ansätze miteinander zu vergleichen. Schließlich wird für Vitellaria paradoxa eine multitemporale Einklassenfall MLC in Analogie zu dem Beispiel für die Fernerkundung von Cassia sieberiana (siehe Abschnitt 9 "Fernerkundung von Cassia sieberiana") durchgeführt und das Resultat mit den Daten zur Verteilung der Auftrittswahrscheinlichkeit der Spezies innerhalb des Untersuchungsgebietes verglichen.

Für Vitellaria pradoxa (und viele weitere der in Burkina Faso heimischen Spezies) ist im "Biodiversitätsatlas" des BIOTA (Biodiversity Monitoring Transect Analysis in Africa) Programms für Burkina Faso auch eine räumliche Verteilung der Auftrittswahrscheinlichkeit der Pflanze in Burkina Faso zu finden (Thiom-BIANO, 2010b). Über die genaue Herkunft und Entstehungsgeschichte der Karten zu den Verteilungen der Auftrittswahrscheinlichkeiten lassen sich jedoch keinerlei Angaben finden. Es lässt sich auch keine Kopie des grundlegenden Datenmaterials auftreiben, sodass die Daten zur Verteilung der Auftrittswahrscheinlichkeit für die weitere Verwendung aus den Karten rekonstruiert werden müssen. Dazu wird zunächst eine digitale Kopie der Karte als Rasterdatensatz angefertigt und georeferenziert. Da der Maßstab der Karten zur Verteilung der Auftrittswahrscheinlichkeit aus dem Biodiversitätsatlas sehr groß ist, lässt sich bei der Georeferenzierung eine ausreichende Genauigkeit erreichen, wenn nur die Koordinaten der Schnittpunkte der Gitterlinien genutzt werden. Um die Verteilung der Auftrittswahrscheinlichkeit aus der georeferenzierten Karte zu extrahieren, wird sie mithilfe des k-Means Cluster Algorithmus geclustert. Da die unter Abschnitt 5.6.1 "k-Means Clustering" beschriebene Implementierung noch nicht optimiert und für größere Datenmengen sehr zeitaufwändig ist, wird an dieser Stelle für das k-Means Clustering die Implementierung aus dem R Paket "stats" genutzt (R CORE TEAM AND CONTRIBUTORS, 2019). Dabei wird die Anzahl der Cluster aus der Menge der vorhandenen Kartenelemente (in diesem Fall sind das die unterschiedlichen diskreten Auftrittswahrscheinlichkeiten, Markierungen von Probeentnahmestellen, Schriftzüge, Gewässer und Gitterlinien) bestimmt. Es entstehen Lücken in der rekonstruierten Verteilung der Auftrittswahrscheinlichkeit an den Orten, wo die Verteilung durch andere Kartenelemente verdeckt wird. Da die Klasse für die höchste Auftrittswahrscheinlichkeit von Vitellaria paradoxa innerhalb des Untersuchungsgebietes nur sehr vereinzelt vorkommt, werden entsprechende Gebiete ebenfalls ausgeschlossen. Die rekonstruierte Verteilung der Auftrittswahrscheinlichkeit von Vitellaria paradoxa innerhalb des Untersuchungsgebietes ist mit der zugrunde gelegten Karte aus dem Biodiversitätsatlas in Abbildung 40 dargestellt.

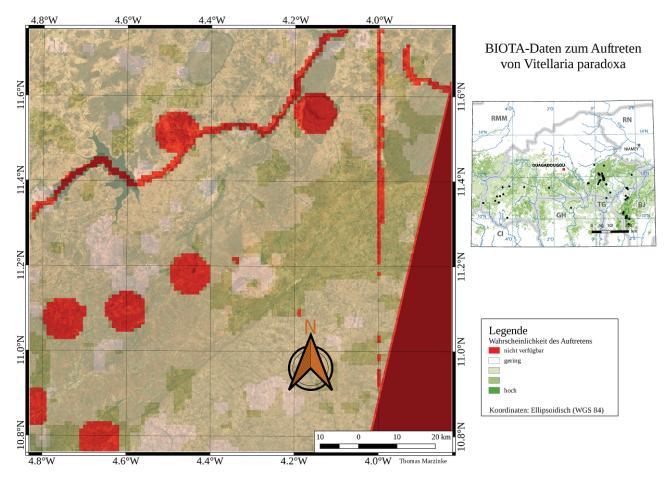


Abbildung 40: Rekonstruktion der räumlichen Verteilung der Auftrittswahrscheinlichkeit von Vitellaria paradoxa innerhalb des Untersuchungsgebietes aus einer Karte aus dem Biodiversitätsatlas für Burkina Faso (rechts; Thiombiano, 2010b) mit RGB-Echtfarbenkomposit der Aufnahme vom 22.2.2019 (ESA, 2019) im Hintergrund

Bei der Verteilung der Auftrittswahrscheinlichkeit gilt zu berücksichtigen, dass das BIOTA Programm in den Zeitraum zwischen 2000 und 2010 fällt. Das grundlegende Datenmaterial ist also bereits mindestens 10 Jahre alt und kann folglich von den aktuellen Verhältnissen abweichen. Beispielsweise fallen unterschiedliche Teile des bebauten Gebietes von Bobo Dioulasso oder der Wasserfläche des Samendéni Stausees in Alle 3 diskreten Niveaus der Auftrittswahrscheinlichkeit innerhalb des Untersuchungsgebietes. Dabei ist die Unregelmäßigkeit bei Bobo Dioulasso am wahrscheinlichsten auf einen Zuwachs der Siedlungsfläche zurückzuführen und der Samendéni Stausee hat sich erst nach der Beobachtungsperiode des BIOTA Programms entwickelt. Darüber hinaus ist die räumliche Auflösung der Verteilung der Auftrittswahrscheinlichkeit (GSD etwa 1km) deutlich geringer, als bei den Aufnahmen des MSI von Sentinel 2 (GSD zwischen 10m und 60m). Daher wird die räumliche Verteilung der Auftrittswahrscheinlichkeit - wie bei den Daten der geringer aufgelösten Kanäle des MSI (vergleiche Abschnitt 5.2 "Preprocessing") - auf eine GSD von 10m hochskaliert (mit nächster Nachbar Interpolation, um die 3 diskreten Klassen zu erhalten). Auf der Grundlage der Verteilung der Auftrittswahrscheinlichkeit wird zunächst untersucht, welcher Klassen

sifizierungsalgorithmus am besten dazu geeignet ist, die Verteilung aus einer zufälligen Stichprobe in einer unitemporalen Mehrklassenfall Klassifikation zu rekonstruieren. Dabei werden die Maximum Likelihood, Mahalanobis Abstand, korrelationskoeffizienten-basierte, Euklidische (Spektrale) Distanz, Random Forest

und die Klassifikation mithilfe der Kanonischen Korrelationsanalyse miteinander verglichen. Die Aufnahme vom 13.1.2019 (siehe Tabelle 8 für die Merkmalsauswahl) wird jeweils mit jedem einzelnen Klassifikationsalgorithmus klassifiziert, wobei alle Algorithmen mit dem selben Trainingsdatensatz trainiert werden. Für den Trainingsdatensatz wird aus jeder der drei unterschiedlichen Klassen für die Auftrittswahrscheinlichkeit innerhalb des Untersuchungsgebietes jeweils eine zufällige Stichprobe vom Umfang 1'000 gezogen, sodass der Trainingsdatensatz insgesamt vom Umfang 3'000 ist. Die Verteilungen der drei unterschiedlichen Klassen für die Auftrittswahrscheinlichkeit werden in den Trainingsdaten mithilfe des Jeffries-Matusita Abstands auf ihre Separabilität untersucht (siehe Tabelle 9).

Tabelle 9: Jeffries-Matusita Abstände (gerundet) für die drei unterschiedlichen Klassen der Auftrittswahrscheinlichkeit in den Trainingsdaten

Auftrittswahrscheinlichkeit	gering	mäßig	hoch
gering	0	0.036	0.115
mäßig	-	0	0.061
hoch	-	-	0

Die Separabilität der Klassen ist sehr schlecht. Dieses Problem ist im Wesentlichen darauf zurückzuführen, dass alle Gebiete des gesamten Untersuchungsgebietes exklusiv auf lediglich drei Klassen verteilt werden. Daraus ergeben sich zwangsläufig große Überschneidungen (beispielsweise aus den unterschiedlichen Klassen für die bebauten Gebiete von Bobo Dioulasso oder die Wasserfläche des Samendéni Stausees). Auch die geringe räumliche Auflösung des Ausgangsdatensatzes wirkt sich nachteilig auf die Separabilität aus. Wegen des Hochskalierens sind mehrere Primitive unterschiedlicher Klassenzugehörigkeit in einem Bildpunkt möglich. Trotz der insgesamt sehr schlechten Separabilität lassen sich in den Jeffries-Matusita Abständen der Verteilungen der Klassen für die verschiedenen Auftrittswahrscheinlichkeiten voneinander Tendenzen erkennen, die im Zusammenhang mit der grundlegenden Problemstellung stehen: Der Abstand zwischen den Verteilungen der Klassen für geringe und hohe Auftrittswahrscheinlichkeit ist größer, als der zwischen den Verteilungen für geringe und mäßige bzw. mäßige und hohe Auftrittswahrscheinlichkeit.

Mit den Trainingsdaten werden unterschiedliche Klassifikationsalgorithmen zur Klassifikation der Aufnahme vom 13.1.2019 in die drei Klassen für die Auftrittswahrscheinlichkeit trainiert. Zum Vergleich der Algorithmen wird jedes Resultat mit der Verteilung der Auftrittswahrscheinlichkeit aus dem Biodiversitätsatlas verglichen, jeweils eine Wahrheitsmatrix aufgestellt sowie der Kappa-Koeffizient, die Produzenten-, Nutzer- und Gesamtgenauigkeit berechnet (Banko, 1998; siehe Tabelle 10 bis 15). Alle rekonstruierten Verteilungen der Auftrittswahrscheinlichkeit stimmen nur sehr schlecht mit der zugrunde gelegten Verteilung aus dem Bidiversitätsatlas überein. Die Gesamtgenauigkeiten der Maximum Likelihood und der Euklidischen (Spektralen) Distanz Klassifikation liegen unter  $33\frac{1}{3}\%$  und sind damit in diesem Fall geringer, als bei der zufälligen Verteilung der Gebiete auf die 3 Klassen. Die höchste Gesamtgenauigkeit wird bei der Klassifikation mithilfe der kanonischen Korrelationsanalyse erzielt. Bei dieser Klassifikationsmethode ist jedoch auch der Kappa-Koeffizient am geringsten: Ein Indiz dafür, dass die "gute" Trefferquote zufällig bedingt ist. Die zweitgrößte Gesamtgenauigkeit (mit 36.56% immer noch nur etwas besser als der Zufall) wird mit der Random Forest Klassifikation erzielt, bei welcher der Kappa-Koeffizient mit 0.041 am größten ist. Folglich stellt das Resultat der Random Forest Klassifikation die beste Rekonstruktion der Verteilung aus dem Biodiversitätsatlas dar. Bei der Anwendung eines multitemporalen Ansatz für die Random Forest Klassifikation (Stacked Vector) lassen sich die Gesamtgenauigkeit und der Kappa-Koeffizient weiter verbessern (siehe Tabelle 16). Dabei wird die multitemporale Random Forest Klassifikation über alle verfügbaren Merkmale der Aufnahmen vom 13.1.2019, 28.1.2019, 22.2.2019 und 14.3.2019 durchgeführt, wobei wie für die unitemporale Random Forest Klassifikation ein Entscheidungswald mit 100 Bäumen genutzt wird. Das Resultat der multitemporalen Random Forest Klassifikation ist die beste Rekonstruktion der Verteilung aus dem Biodiversitätsatlas und ist in Abbildung 41 dargestellt.

Eine wesentliche Ursache für die schlechte Übereinstimmung der rekonstruierten Verteilungen mit derjenigen aus dem Biodiversitätsatlas ist die schlechte Qualität der Trainingsdaten (siehe Tabelle 9), welche auch im Zusammenhang mit den unterschiedlichen räumlichen Auflösungen der Verteilung aus dem Biodiversitätsatlas und der für die Klassifikation genutzten Aufnahmen bzw. der resultierenden Verteilung steht. Da die räumliche Auflösung der resultierenden Verteilung größer ist, als die der zugrunde gelegten, wird die Verteilung auch verdichtet. Das führt dazu, dass innerhalb eines Bildpunktes der zugrunde gelegten Verteilung mehrere Klassen der verdichteten Verteilung vorkommen können und so das Resultat weniger mit der zugrunde gelegten Verteilung übereinstimmt - auch, wenn die verdichtete Verteilung die tatsächlichen Verhältnisse besser wiedergibt. Die Qualität der verdichteten Verteilung wird hier zusätzlich dadurch eingeschränkt, dass die Trainingsdaten (da keine Informationen für eine bessere Abgrenzung vorliegen) zufällig aus den zugrunde gelegten Verteilungen der Klassen für unterschiedliche Auftrittswahrscheinlichkeiten gezogen sind. Die Verdichtung lässt sich nur mit räumlich höher aufgelösten Verteilungen der Auftrittswahrscheinlichkeit zuverlässig validieren.

Tabelle 10: Wahrheitsmatrix mit Genauigkeitsmaßen für das Resultat der Maxlimum Likelihood Klassifikation

Auftrittswahrscheinlichkeit BIOTA-Daten→	gering	mäßig	hoch	Nutzergenauigkeit
Resultat \( \psi \)	8011118	<u> </u>		
gering	5'089'113	41'948'330	13'310'596	8.43%
mäßig	1'663'778	14'589'563	4′521′096	70.23%
hoch	1'393'655	12'693'086	5'438'991	27.86%
Produzentengenauigkeit	62.47%	21.07%	23.37%	Gesamtgenauigkeit: 24.96%
Kappa-Koeffizient			0.019	

Tabelle 11: Wahrheitsmatrix mit Genauigkeitsmaßen für das Resultat der Mahalanobis Abstand Klassifikation

	gering	mäßig	hoch	Nutzergenauigkeit
Resultat↓				
gering	1'170'685	6'855'473	2'042'682	11.63%
mäßig	1'698'660	15'122'251	3'796'535	73.35%
hoch	5'277'211	47'253'347	17'431'505	24.92%
Produzentengenauigkeit	14.37%	21.84%	74.91%	Gesamtgenauigkeit: 33.51%
Kappa-Koeffizient			0.037	

Tabelle 12: Wahrheitsmatrix mit Genauigkeitsmaßen für das Resultat der Euklidischen (Spektralen) Distanz Klassifikation

Auftrittswahrscheinlichkeit				Nutzergenauigkeit
$BIOTA-Daten \rightarrow$	gering	mäßig	hoch	
Resultat↓				
gering	4'118'672	30'662'270	8'326'328	9.55%
mäßig	616′913	4'697'899	1'434'266	69.61%
hoch	3'410'970	33'870'901	13′510′130	26.60%
Produzentengenauigkeit	50.56%	6.79%	58.06%	Gesamtgenauigkeit: 22.18%
Kappa-Koeffizient			0.030	

Tabelle 13: Wahrheitsmatrix mit Genauigkeitsmaßen für das Resultat der korrelationskoeffizientenbasierten Klassifikation

Auftrittswahrscheinlichkeit				Nutzergenauigkeit
$\parallel$ BIOTA-Daten $\rightarrow$	gering	mäßig	hoch	
Resultat↓				
gering	3'104'382	25'232'317	8'533'451	8.42%
mäßig	3'183'672	25'898'437	7'738'568	70.34%
hoch	1'857'299	18'090'248	6'995'299	25.96%
Produzentengenauigkeit	38.11%	37.41%	30.06%	Gesamtgenauigkeit: 35.77%
Kappa-Koeffizient			0.022	

Tabelle 14: Wahrheitsmatrix mit Genauigkeitsmaßen für das Resultat der Klassifikation mithilfe der Kanonischen Korrelationsanalyse

Auftrittswahrscheinlichkeit				Nutzergenauigkeit
$\parallel$ BIOTA-Daten $\rightarrow$	gering	mäßig	hoch	
Resultat↓				
gering	1'274'057	11'034'728	4'177'988	7.73%
mäßig	6'180'101	51'870'118	16'781'046	69.32%
hoch	692′398	6'326'230	2'311'691	24.78%
Produzentengenauigkeit	15.64%	74.92%	9.93%	Gesamtgenauigkeit: 55.10%
Kappa-Koeffizient			0.011	

Tabelle 15: Wahrheitsmatrix mit Genauigkeitsmaßen für das Resultat der Random Forest Klassifikation

Auftrittswahrscheinlichkeit				Nutzergenauigkeit
$BIOTA-Daten \rightarrow$	gering	mäßig	hoch	
Resultat↓				
gering	3'517'565	21'364'235	6'334'783	11.27%
mäßig	2'430'261	23'885'661	7'538'217	70.55%
hoch	2'198'730	23'981'180	9'397'725	26.41%
Produzentengenauigkeit	43.18%	34.50%	40.38%	Gesamtgenauigkeit: 36.56%
Kappa-Koeffizient			0.041	

Tabelle 16: Wahrheitsmatrix mit Genauigkeitsmaßen für das Resultat der multitemporalen Random Forest Klassifikation

Auftrittswahrscheinlichkeit				Nutzergenauigkeit
$BIOTA-Daten \rightarrow$	gering	mäßig	hoch	
Resultat↓				
gering	4′506′946	19'661'184	3'987'773	16.01%
mäßig	2'055'723	25'073'135	6'593'647	74.35%
hoch	1′583′887	24'496'757	12'689'305	32.73%
Produzentengenauigkeit	55.32%	36.22%	54.53%	Gesamtgenauigkeit: 42.00%
Kappa-Koeffizient	0.118			

# Rekonstruktion der Auftrittswahrscheinlichkeit von Vitellaria Paradoxa mit multitemporaler RFC

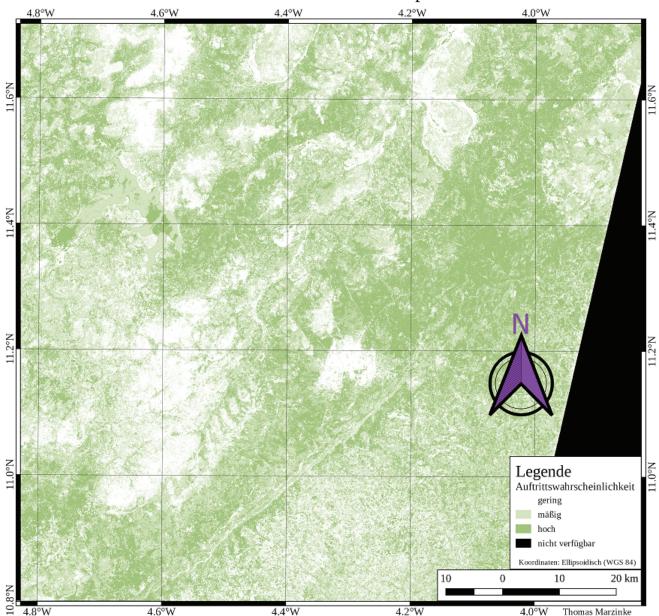


Abbildung 41: Rekonstruktion der räumlichen Verteilung der Auftrittswahrscheinlichkeit von Vitellaria paradoxa innerhalb des Untersuchungsgebietes mithilfe einer multitemporalen RFC aus der Verteilung der Auftrittswahrscheinlichkeit aus dem Biodiversitätsatlas für Burkina Faso (Thiombiano, 2010b)

Auch für Vitellaria paradoxa sind bei der Global Biodiversity Information Facility Daten mit Beobachtungspositionen der Spezies verfügbar (GBIF, 2020), sodass eine multitemporale Einklassenfall MLC in Analogie zu dem Beispiel für die Fernerkundung von Cassia sieberiana (siehe Abschnitt 9 "Fernerkundung von Cassia sieberiana") durchgeführt werden kann. Innerhalb des Untersuchungsgebietes wird Vitellaria paradoxa zwischen 1963 und 2003 an 109 unterschiedlichen Orten beobachtet. Wie bei den Daten zu Cassia sieberiana ist die Genauigkeit der Koordinatenangaben auch bei den Beobachtungspositionen von Vitellaria Paradoxa nicht dokumentiert. Für die multitemporale Einklassenfall MLC werden die selben Aufnahmen und Merkmale genutzt, wie bei dem Beispiel zur Fernerkundung von Cassia sieberiana (siehe Tabelle 8).

In Abbildung 42 sind die 10% der Gebiete innerhalb des Untersuchungsgebietes mit den höchsten Werten der resultierenden multitemporalen Likelihood-Funktion zusammen mit den Beobachtungspositionen aus den GBIF-Daten dargestellt. Beim Vergleich mit der Verteilung der Auftrittswahrscheinlichkeit von Vitellaria paradoxa aus dem Biodiversitätsatlas fällt auf, dass etwa 10.75% der Gebiete mit hoher Auftrittswahrscheinlichkeit der Spezies zu den 10% der Gebiete mit den höchsten Werten der resultierenden multitemporalen Likelihood-Funktion aus dem Resultat gehören, bei den Gebieten mit mäßiger Auftrittswahrscheinlichkeit sind es etwa 10.37% und bei denen mit geringer 8.80%. Das Resultat ist also mit der Verteilung der Auftrittswahrscheinlichkeit aus dem Biodiversitätsatlas konsistent.

### Resultat multitemporale MLC: Vitellaria paradoxa

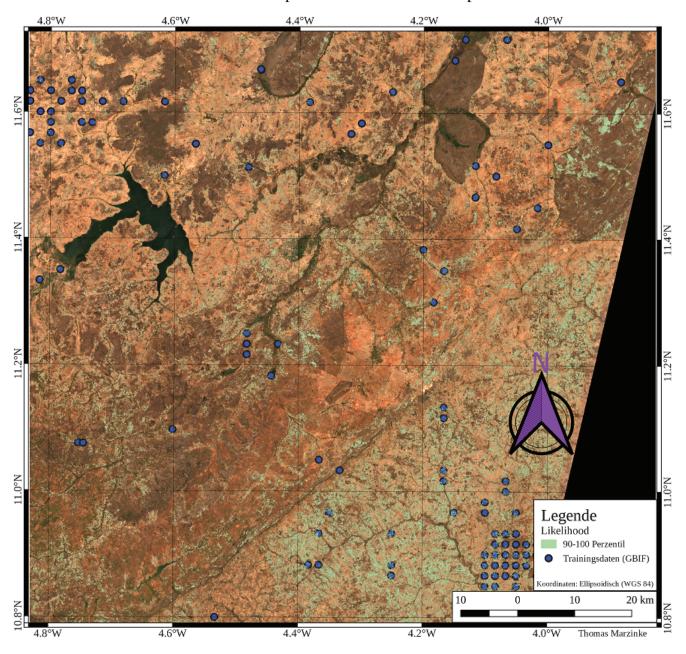


Abbildung 42: Resultat der multitemporalen MLC für Vitellaria paradoxa (10% der Gebiete mit den höchsten Werten der multitemporalen Likelihood-Funktion innerhalb des Untersuchungsgebietes) mit Beobachtungspositionen aus den GBIF-Daten (GBIF, 2020) und RGB-Echtfarbenkomposit der Aufnahme vom 22.2.2019 (ESA, 2019) im Hintergrund

## 11 Zusammenfassung

Unterschiedliche Klassifizierungsalgorithmen werden in R implementiert und auf die Problemstellung der Fernerkundung von Heilpflanzen in Deutschland (Umgebung des Dorfes Kraase) und Burkina Faso (Umgebung von Bobo Dioulasso und Banfora) angewandt. Das Instrumentarium der eingesetzten Klassifizierungsalgorithmen umfasst neben probabilistischen auch nicht probabilistische, Spektrum-basierte und Methoden des maschinellen Lernens. Neben Mehrklassenfall Klassifikationen wird ein besonderes Augenmerk auf Klassifikationen im Einklassenfall gelegt, deren Einsatz insbesondere bei der Fernerkundung einzelner Spezies von Vorteil ist. Die Klassifikationsergebnisse lassen sich darüber hinaus auch durch die Anwendung unterschiedlicher multitemporaler Verfahren für die Maximum Likelihood, Euklidische (Spektrale) Distanz, korrelationskoeffizienten-basierte und Random Forest Klassifikation weiter verbessern.

Verschiedene Verfahren zur Fernerkundung von Heilpflanzen werden anhand mehrerer Beispielanwendungen für unterschiedliche Spezies innerhalb des Untersuchungsgebietes bei Kraase vorgestellt und erfolgreich erprobt. Die Resultate spiegeln die tatsächlich vor Ort vorgefundenen Verhältnisse sehr gut wieder und belegen die Zuverlässigkeit der genutzten Verfahren.

Für das Untersuchungsgebiet in Burkina Faso liegen nur sehr wenige Informationen über die vorkommenden Spezies und deren räumliche Verteilung vor. Daher wird das auf dem Untersuchungsgebiet bei Kraase erprobte Instrumentarium hier auf die Fernerkundung zweier Spezies (Cassia sieberiana und Vitellaria paradoxa) angewandt, für die auch Informationen über ihre räumliche Verteilung aus zusätzlichen Quellen vorliegen. Wegen des Informationsmangels gestaltet sich die Validierung der Resultate für das Untersuchungsgebiet bei Bobo Dioulasso schwierig.

Die Unabhängigkeit der Untersuchungen wird durch die ausschließliche Nutzung freier Daten und Software (FOSS) sichergestellt, wodurch auch die Kosten auf ein Minimum reduziert werden können. Als Ausgangsdaten für die Fernerkundung von Heilpflanzen werden die Daten der Satellitenfernerkundungsmission Sentinel 2 der Europäischen Kommission genutzt. Diese stehen der Allgemeinheit jederzeit zur freien Verfügung und stellen zum aktuellen Zeitpunkt eine technologische Speerspitze unter den frei verfügbaren Satellitenfernerkundungsmissionen dar.

#### 12 Ausblick

Das vorgestellte Instrumentarium zur Fernerkundung von Heilpflanzen liefert für das Untersuchungsgebiet bei Kraase sehr zuverlässige Endergebnisse und kann daher bei zukünftigen Untersuchungen eine effiziente Ergänzung und geeignete Unterstützung bei der in-situ Datenerfassung sein. Bei Problemstellungen, in denen die Datenerfassung mit sehr großem zeitlichen oder finanziellen Aufwand verbunden ist und die Ansprüche an die Genauigkeit erfüllt werden, kann die Fernerkundung eine günstige Alternative darstellen. Die Genauigkeit der Verfahren lässt sich dabei durch die Nutzung von Sensorsystemen mit höherem räumlichen und spektralen Auflösungsvermögen sowie die Erweiterung des Instrumentariums um zusätzliche Methodiken zur Berücksichtigung weiterer Parameter (wie beispielsweise der Kronen- oder Blattgeometrien) weiter verbessern. Darüber hinaus ist die Fernerkundung von Spezies als nicht-destruktive Methodik in stark gefährdeten Lebensräumen vorteilhaft, wenn die durch eine flächendeckende in-situ Datenerfassung verursachten Schäden nicht vertretbar sind. Gleiches gilt in schwer zugänglichen Gebieten oder solchen, in denen die Gefährdung für die in-situ Datenerfassung zu groß ist. In jedem Fall sind zur Verifikation der Resultate auch in-situ Daten erforderlich und unter Umständen lässt sich nicht jede Spezies mithilfe der Fernerkundung gleich gut aufspüren.

Wegen des Mangels an Informationen für das Untersuchungsgebiet in Burkina Faso steht hier eine detailliertere Beurteilung der Endergebnisse noch aus. Es ist zu erwarten, dass sich die Resultate auf der Grundlage zusätzlichen Datenmaterials verbessern lassen.

Das vorgestellte Instrumentarium lässt sich als Ausgangspunkt für mögliche Anwendungen der Fernerkundung im Rahmen des geplanten DigiMeP Projektes nutzen. Die Untersuchungen können als Unterstützung bei der Beurteilung der Anwendbarkeit von Methodiken der Fernerkundung im Rahmen des Projektes dienen.

## Anhang

Glossar

BIOTA - Biodiversity Monitoring Transect Analysis in Africa

**BOA** - Bottom Of Atmosphere

**bzw.** - beziehungsweise

CIA - Central Intelligence Agency

**DAAC** - Distributed Active Archive Center

**DC** - District of Columbia

**DEM** - Digital Elevation Model (Digitales Geländemodell)

DigiMeP - DIGItal ressource management for traditional MEdicinal Plants

DLR - Deutsches Zentrum für Luft- und Raumfahrt

Ed. - Editor Edn. - Editoren

EM - Expectation-Maximization

**EOSDIS** - Earth Observing System Data and Information System

**ESA** - European Space Agency

et al. - et alii

ETM - Enhanced Thematic Mapper EVI2 - Enhanced vegetation Index 2

FAO - Food and Agriculture Organization of the United Nations

**FOS(S)** - Free ans Open Source (Software)

GBIF - Global Biodiversity Information Facility

GIS - Geographic Information System

GmbH - Gemeinschaft mit beschränkter Haftung

**GMES** - Global Monitoring for Environment and Security

**GSD** - Ground Sample Distance

IEEE - Institute of Electronical and Electronics Engineers

JPL - Jet Propulsion Laboratory

LAI - Leaf Area Index

LCI - Leaf Chlorophyll Index

m - Meter

MAP - Maximum A Posteriori

MCARI - Modified Chlorophyll Absorption in Reflectance Index

MDPI - Multidisciplinary Digital Publishing Institute

ML(C) - Maximum Likelihood (Classification)

MSI - Multispectral Instrument NASA - North American Space Agency

NDMI - Normalized Difference Moisture Index

NN - Normal Null

OLI - Operational Land ImagerOMG - Object Management Group

PNAS - Proceedings of the National Academy of Sciences of the United States of America

**pp.** - pages (Seiten)

**QED** - Quantenelektrodynamik

QGIS - Quantum GIS (FOSS-GIS-Software)
RF(C) - Random Forest (Classification)

RGB - Rot Grün Blau

SAR - Synthetic Aperture RadarSCM - Spectral Correlation Mapper

TM - Thematic Mapper

USA - United States of AmericaUML - Unified Modeling Language

UNESCO - United Nations Educational, Scientific and Cultural Organization

UTM - Universale Transversale Mercator-Projektion

WGS - World Geodetic SystemWHO - World Health Organization

WIW - Water In Wetlandsz.B. - zum Beispiel

### Quellenverzeichnis

#### Quellenverzeichnis

AHMAD, A. & QUEGAN, S. (2012): Analysis of Maximum Likelihood Classification Technique on Landsat 5 TM Satellite Data of Tropical Land Covers. - 2012 IEEE International Conference on Control System, Computing and Engineering: 280-285.

ALBERTZ, J. & WIGGENHAGEN, M. (2009): Taschenbuch zur Photogrammetrie und Fernerkundung, 5., völlig neu bearbeitete und erweiterte Auflage. - Herbert Wichmann Verlag, Heidelberg, Deutschland.

ALIYU, Z. Y., YUSHA'U, M. J. & ALIYU, B. S. (2013): Anti Malarial Activity of Cassia Sieberiana Leaf Extracts. - Open Conf. Proc. J. 4, Special Supplement of 5th International Drug Discovery & Therapy 2013: 72-76.

AL-KINDI, K. M., KWAN, P., ANDREW, N. R. & WELCH, M. (2017): Remote sensing and spatial statistical techniques for modelling Ommatissus lybicus (Hemiptera: Tropiduchidae) habitat and population densities. - PeerJ 5(3752).

AMT FÜR GEOINFORMATION, VERMESSUNG UND KATASTERWESEN (2016): Digitale Topographische Karte 1:25000 Mecklenburg-Vorpommern, Amt für Geoinformation, Vermessung und Katasterwesen, Landesamt für innere Verwaltung Mecklenburg-Vorpommern, Schwerin, Deutschland.

BAENA, S., MOAT, J., WHALEY, O. & BOYD, D. S. (2017): Identifying species from the air: UAVs and the very high resolution challenge for plant conservation. - PLOS ONE 12(11).

Banko, G. (1998): A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data and of Methods Including Remote Sensing Data in Forest Inventory. - International Institute for Applied Systems Analysis, Laxenburg, Österreich.

BANNARI, A., MORIN, D., BONN, F. & HUETE, A. R. (1995): A Review of Vegetation Indices. - Remote Sens. Rev. 13(1-2): 95-120.

BASSETT, R. & DREIDE, J. (2019): Maximum a posteriori estimators as a limit of Bayes estimators.
- Math. Program. 174: 129-144.

Bellman, R. (1957): Dynamic Programming. - Princeton University Press, Princeton, New Jersey, USA.

BENEDIKTSSON, J. A., SWAIN, P. H. & ERSOY, O. K. (1990): Neural Network Approaches Versus Statistical Methods in Classification of Multisource Remote Sensing Data. - IEEE Trans. Geosci. Remote Sens. 28(4): 540-552.

BERGER, M., HERLAND, E.-A., ASCHBACHER, J., BRIGGS, S., JOHANNESSEN, J., HANSSEN, R., MORENO, J. & HAUGLUSTAINE, D. (2009): New Possibilities For Science ESA's operational Sentinel missions. - ESA Bulletin 140.

BIVAND, R., KEITT, T., ROWLINGSON, B., PEBESMA, E., SUMNER, M., HIJMANS, R., ROU-AULT, E., WARMERDAM, F., OOMS, J. & RUNDEL, C. (2019): Bindings for the 'Geospatial' Data

- Abstraction Library (Version: 1.4-3), R Foundation for Statistical Computing, Wien, Österreich.
- BONETT, D. G. & WRIGHT, T. A. (2000): Sample Size Requirements For Estimating Pearson, Kendall and Spearman Correlations. Psychometrika 65(1): 23-28.
- Breiman, L. (2001): Random Forests. Mach. Learn. 45(1): 5-32.
- BRONSTEIN, I.N., SEMENDJAJEW, K.A., MUSIOL, G. & MÜHLIG, H. (2012): Taschenbuch der Mathematik 8., vollständig überarbeitete Auflage. Verlag Harri Deutsch, Frankfurt am Main, Deutschland.
- Busia, K. & Kasilo, O. M. J. (2010): An overview of traditional medicine in ECOWAS Member States. In: Lusamba-Dikassa, P. et al. (Edn.): The African Health Monitor, African Traditional Medicine Day, 31 August, Special Issue. WHO Regional Office for Africa, Brazzaville, Republik Kongo, 16-24.
- CÁMARA-LERET, R., FORTUNA, M. A. & BASCOMPTE, J. (2019): Indigenous knowledge networks in the face of global change. Proc. Natl. Acad. Sci. U.S.A. 116(20): 9913-9918.
- CHOODARATHNAKARA, A. L., KUMAR, T. A., KOLIWAD, S. & PATIL, C. G. (2012): Mixed Pixels: A Challenge in Remote Sensing Data Classification for Improving Performance. Int. J. Adv. Res. Comput. Eng. Technol. 1(9): 261-271.
- CIA (2018a): Africa: Burkina Faso. In: The World Factbook. Central Intelligence Agency (CIA), Washington, District of Colombia, USA.
- CIA (2018b): Africa: Benin. In: The World Factbook. Central Intelligence Agency (CIA), Washington, District of Colombia, USA.
- CLERK, S. & MPC TEAM (2020): S2 MPC Level 2A Data Quality Report Issue 22. European Space Agency (ESA), Paris, Frankreich.
- CLEVERS, J. G. P. W., KOOISTRA, L. & VAN DEN BRANDE, M. M. M. (2017): Using Sentinel-2 Data for Retrieving LAI and Leaf and Canopy Chlorophyll Content of a Potato Crop. Remote Sens. 9(5,405).
- COPERNICUS SPACE COMPONENT MISSION MANAGEMENT TEAM (2019): Sentinel High Level Operations Plan Issue 3 Revision 0. European Space Agency (ESA), Paris, Frankreich.
- **Datt**, **B.** (1999): A New Reflectance Index for Remote Sensing of Chlorophyll Content in Higher Plants: Tests using Eucalyptus Leaves. J. Plant Physiol. 154(1): 30-36.
- DAUGHTRY, C. S. T., WALTHALL, C. L., KIM, M. S., BROWN DE COLSTOUN, E. & MCMURTREY, J. E. (2000): Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance. Remote Sens. Environ. 74(2): 229-239.
- **DE CARVALHO JÚNIOR, O. A. & MENESES, P. R.** (2000): Spectral Correlation Mapper (SCM): An Improvement on the Spectral Angle Mapper (SAM). Proceedings of the Ninth Annual JPL Airborne Earth Science Workshop: 65-74.
- **DEMPSTER**, A. P., LAIARD, N. M. & RUBIN, D. B. (1977): Maximum Likelihood from Incomplete Data via the EM Algorithm. J. R. Stat. Soc. Series B Stat. Methodol. 39(1): 1-38.
- DIPAMA, J. M. & ANNE, C. A. T. (2010a): Principal environmental factors of Burkina Faso 4.2 Geology and geomorphology. In: Thiombiano, A. & Kampmann, D. (Edn.): Biodiversity Atlas of West Africa Volume II Burkina Faso. Project management agency of the German Aerospace Center (PT-DLR), Environment, Culture, Sustainability, Bonn, Deutschland, 126-128.
- **DIPAMA**, J. M. (2010b): Principal environmental factors of Burkina Faso 4.1 Climate. In: Thiombiano, A. & Kampmann, D. (Edn.): Biodiversity Atlas of West Africa Volume II Burkina Faso. Project management agency of the German Aerospace Center (PT-DLR), Environment, Culture, Sustainability, Bonn, Deutschland, 122-124.

- DIRKS, P. H. G. M., BLENKINSOP, T. G. & JELSMA, H. A. (2003): The Geological Evolution Of Africa. In: De Vivo, B., Grasemann, B. & Stüwe, K. (Edn.): Geology Volume 4. UNESCO Encyclopedia Of Life Support Systems Publishers, Oxford, England, 230-265.
- EASTMAN, J. R. (1999): Multi-criteria evaluation and GIS. In: Longley, P. A., Goodchild, M. F., Maguire, D. J. & Rhind, D. W. (Edn.): Geographical Information Systems Volume 1 Principles and Technical Issues. John Wiley & Sons, Hoboken, New Jersey, USA, 493-502.
- **ESA** (2015): Sentinel-2 User Handbook, Issue 1, Revision 2. European Space Agency (ESA), Paris, Frankreich.
- ESA (2018): sentinels Space For Copernicus. European Space Agency (ESA), Paris, Frankreich.
- **EUROPEAN COMMISSION** (2017): Guidance Document for Horizon 2020 Work Programme 2018-2020. European Commission, Brüssel, Belgien.
- **EUROPEAN COMMISSION** (2019): Annex to the Commission Implementing Decision on the financing of the Copernicus Programme and on the adoption of the work programme for 2020. European Commission, Brüssel, Belgien.
- FARAG, M., MOHAMMED, M. S., FOUD, I., AHMED, W. J. & MOHAMED, M. S. (2015): The Role Of Natural Products In Drug Discovery And Development. World J. Pharm. Res. 4(3): 22-33.
- FEYNMAN, R. P., LEIGHTON, R. B. & SANDS, M. (2007): Feynman-Vor-lesungen über Physik Band 1 Mechanik, Strahlung, Wärme, Definitive Edition. Oldenbourg Wissenschaftsverlag GmbH, München, Deutschland.
- FEYNMAN, R. P., LEIGHTON, R. B. & SANDS, M. (2009): Feynman-Vor-lesungen über Physik Band 3 Quantenmechanik, Definitive Edition. Oldenbourg Wissenschaftsverlag GmbH, München, Deutschland.
- FRAEDRICH, W. (2016): Spuren der Eiszeit, Landschaftsformen in Mitteleuropa. Springer Spektrum, Springer-Verlag Berlin Heidelberg, Berlin, Deutschland.
- FUCHS, G. W. (2010): Kreislauf eines kosmischen Überlebenskünstlers. In: Bührke, T. & Wengenmayr, R. (Edn.): Geheimnisvoller Kosmos. Wiley-VCH Verlag, Weinheim, Deutschland, 90-95.
- GAI, Y., FAN, W., Xu, X. & ZHANG, Y. (2011): Flower species identification and coverage estimation based on hyperspectral remote sensing data. 2011 IEEE International Geoscience and Remote Sensing Symposium: 1243-1246.
- GENERALVERSAMMLUNG DER VEREINTEN NATIONEN (2015): Transformation unserer Welt: die Agenda 2030 für nachhaltige Entwicklung, Resolution der Generalversammlung, verabschiedet am 25. September 2015. Vereinte Nationen, New York, USA.
- GHAMISI, P., RASTI, B., YOKOYA, N., WANG, Q., HÖFLE, B., BRUZZONE, L., BOVOLO, F., CHI, M., ANDERS, K., GLOAGUEN, R., ATKINSON, P. M. & BENEDIKTSSON, J. A. (2019): Multisource and Multitemporal Data Fusion in Remote Sensing. IEEE Trans. Geosci. Remote Sens. 7(1): 6-39.
- GHOUSSEIN, Y., NICOLAS, H., HAURY, J., FADEL, A., PICHELIN, P., ABOU HAMDAN, H. & FAOUR, G. (2019): Multitemporal Remote Sensing Based on an FVC Reference Period Using Sentinel-2 for Monitoring Eichhornia crassipes on a Mediterranean River. Remote Sens. 11(16,1856).
- GMES PROGRAMME, ESA SENTINEL-2 PROJECT AND SCIENCE DIVISION TEAMS AT ESTEC, NOORDWIJK, NIEDERLANDE; ESA SENTINEL-2 PAYLOAD DATA GROUND SEGMENT AND MISSION MANAGEMENT TEAM AT ESRIN, FRASCATI, ITALIEN & ESA SENTINEL-2 FLIGHT OPERATIONS SEGMENT TEAM AT ESOC, DARMSTADT, DEUTSCHLAND (2012): Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. European Space Agency (ESA), Paris, Frankreich.

- GOETZ, A. F. H., VANE, G., SOLOMON, J. E. & ROCK, B. N. (1985): Imaging Spectrometry for Earth Remote Sensing. Science 228(4704): 1147-1153.
- Gorban, A. N. & Tyukin, I. Y. (2018): Blessing of dimensionality: mathematical foundations of the statistical physics of data. Philos. Trans. Royal Soc. 376(2118,20170237).
- GREBE, M. (2017): Königreiche in Westafrika Früh- und Vorgeschichte von Burkina Faso. In: Runge, J. & Anamou, N. (Edn.): Westafrika Geografische Studienreisen in Bénin, Burkina Faso und Togo. Institut für Physische Geographie, Goethe-Universität Frankfurt am Main, Frankfurt am Main, Deutschland, 155-160.
- Hall, M. A. (1999): Correlation-based Feature Selection for Machine Learning. Department of Computer Science, The University of Waikato, Hamilton, Neuseeland.
- HÄNSCH, R. & HELLWICH, O. (2017): Random Forests. In: Heipke, C. (Ed.): Handbuch der Geodäsie, Photogrammetrie und Fernerkundung. Springer-Verlag GmbH Deutschland, Berlin, Deutschland, 603-643.
- HARDISKY, M. A., KLEMAS, V. & SMART, R. M. (1983): The Influence of Soil Salinity, Growth Form, and Leaf Moisture on the Spectral Radiance of Spartina alterniflora Canopies. Photogramm. Eng. Remote Sens. 49(1): 77-83.
- HASTIE, T., TIBSHIRANI, R. & FRIEDMAN, J. (2009): The Elements of Statistical Learning. Springer, New York, USA.
- **Heidt**, S. (2017): Das Taoudeni Becken in Burkina Faso. In: Runge, J. & Anamou, N. (Edn.): Westafrika Geografische Studienreisen in Bénin, Burkina Faso und Togo. Institut für Physische Geographie, Goethe-Universität Frankfurt am Main, Frankfurt am Main, Deutschland, 19-20.
- **HEUPEL**, K., SPENGLER, D. & ITZEROTT, S. (2018): A Progressive Crop-Type Classification Using Multitemporal Remote Sensing Data and Phenological Information. PFG J. Photogramm. Remote Sens. Geoinf. Sci. 86(2): 53-69.
- HIJMANS, R. J., VAN ETTEN, J., SUMNER, M., CHENG, J., BEVAN, A., BIVAND, R., BUSETTO, L., CANTY, M., FORREST, D., GHOSH, A., GOLICHER, D., GRAY, J., GREENBERG, J. A., HIEMSTRA, P., KARNEY, C., MATTIUZZI, M., MOSHER, S., NOWOSAD, J., PEBESMA, E., PERPINAN LAMIGUEIRO, O., RACINE, E. B., ROWLINGSON, B., SHORTRIDGE, A., VENABLES, B. & WUEEST, R. (2019): Geographic Data Analysis and Modeling (Version: 2.8-19). R Foundation for Statistical Computing, Wien, Österreich.
- **HOLLOWAY**, J. & MENGERSEN, K. (2018): Statistical Machine Learning Methods and Remote Sensing for Sustainable Development Goals: A Review. Remote Sens. 10(9,1365).
- **HOTELLING**, **H.** (1936): Relations Between Two Sets of Variates. Biometrika 28(3/4): 321-377.
- **Hughes**, G. F. (1968): On the Mean Accuracy of Statistical Pattern Recognizers. IEEE Trans. Inf. Theory 14(1): 55-63.
- International Monetary Fund (2008): Burkina Faso: Poverty Reduction Strategy Paper Annual Progress Report. International Monetary Fund, Washington, District of Colombia, USA.
- JEON, B. & LANDGREBE, D. A. (1999): Decision fusion approach for multitemporal classification. IEEE Trans. Geosci. Remote Sens. 37(3): 1227-1233.
- JIANG, Z., HUETE, A. R., DIDAN, K. & MIURA, T. (2008): Development of a two-band enhanced vegetation index without a blue band. Remote Sens. Environ. 112(10): 3833-3845.
- JIN, X., JIE, L., WANG, S., QI, H. J. & LI, S. W. (2018): Classifying Wheat Hyperspectral Pixels of Healthy Heads and Fusarium Head Blight Disease Using a Deep Neural Network in the Wild Field. Remote Sens. 10(3,395).

- Jutzi, B., Meyer, F. J. & Hinz, S. (2017): Aktive Fernerkundungssensorik Technologische Grundlagen und Abbildungsgeometrie. In: Heipke, C. (Ed.): Handbuch der Geodäsie, Photogrammetrie und Fernerkundung. Springer-Verlag GmbH Deutschland, Berlin, Deutschland, 29-64.
- KAHUMBA, J., RASAMIRAVAKA, T., OKUSA, P. N., BAKARI, A., BIZUMUKAMA, L., KALONJI, J.-B., KIENDREBEOGO, M., RABEMENANTSOA, C., EL JAZIRI, M., WILLIAMSON, E. M. & DUEZ, P. (2015): Traditional African medicine: From ancestral knowledge to a modern integrated future. In: Sanders, S. et al. (Eds.): The Art and Science of Traditional Medicine, Part 3: The Global Impact of Traditional Medicine. American Association for the Advancement of Science Custom Publishing Office, American Association for the Advancement of Science, Washington, District of Colombia, 61-63.
- Kasilo, O. M. J., Trapsida, J.-M., Mwikisa, C. N. & Lusamba-Dikas-sa, P. S. (2010): An overview of the traditional medicine situation in the African Region. In: Lusamba-Dikassa, P. et al. (Edn.): The African Health Monitor, African Traditional Medicine Day, 31 August, Special Issue. WHO Regional Office for Africa, Brazzaville, Republik Kongo, 7-15.
- KERR, J. T. & OSTROVSKY, M. (2003): From space to species: ecological applications for remote sensing. TRENDS Ecol. Evol. 18(6): 299-305.
- KHAN, S. S. & MADDEN, M. G. (2014): One-class classification: taxonomy of study and review of techniques. Knowl. Eng. Rev. 29(3): 345-374.
- KILIAN, U. (2020): Müll als Rohstoff. Phys. J. 19(2): 40-41.
- KINDA, P. T., ZERBO, P., GUENNÉ, S., COMPAORÉ, M., CIOBICA, A. & KIENDREBEOGO, M. (2017): Medicinal Plants Used for Neuropsychiatric Disorders Treatment in the Hauts Bassins Region of Burkina Faso. Medicines 4(2,32).
- KNOWELDEN, R. (2019): Copernicus Sentinel Data Access 2018 Annual Report, Issue 1, Revision 0. Serco Gael Consortium, Frascati, Italien.
- KOBAYASHI, N., TANI, H., WANG, X. & SONOBE, R. (2020): Crop Classification using spectral indices derived from Sentinel-2A imagery. J. Inf. Telecommun. 4(1): 67-90.
- KRIGE, D. G. (1951): A Statistical Approach To Some Basic Mine Valuation Problems On The Witwatersrand. J. Chem. Metall. Min. Soc. S. Afr. 52(6): 119-139.
- LARSSON, A., CLAUSEN, L. & VOLUNTEERS (2011): Dia, Eine Anwendung zum Zeichnen von strukturierten Diagrammen, Version: 0.97.2. Free Software Foundation Inc., Boston, Massachusetts, USA.
- LEE, J.-H., PARK, M.-H. & KIM, Y.-I. (1999): An Application of Canonical Correlation Analysis Technique to Land Cover Classification of LANDSAT Images. ETRI J. 21(4): 41-51.
- LEFEBVRE, G., DAVRANCHE, A., WILLM, L., CAMPAGNA, J., REDMOND, L., MERLE, C., GUELMAMI, A. & POULIN, B. (2019): Introducing WIW for Detecting the Presence of Water in Wetlands with Landsat and Sentinel Satellites. Remote Sens. 11(19,2210).
- LETEANE, M. M., NGWENYA, B. N., MUZILA, M., NAMUSHE, A., MWINGA, J., MUSONDA, R., MOYO, S., MENGESTU, Y. B., ABEGAZ, B. M. & ANDRAE-MAROBELA, K. (2012): Old plants newly discovered: Cassia sieberiana D.C. and Cassia abbreviata Oliv. Oliv. root extracts inhibit in vitro HIV-1c replication in peripheral blood mononuclear cells (PBMCs) by different modes of action. J. Ethnopharmacol. 141(1): 48-56.
- LIAW, A. & WIENER, M. (2018): Breiman and Cutler's Random Forests for Classification and Regression (Version: 4.6-14). R Foundation for Statistical Computing, Wien, Österreich.
- MACQUEEN, J. (1967): Some Methods For Classification And Analysis Of Multivariate Observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1: 281-297.
- MAHALANOBIS, P. C. (1936): On The Generalized Distance In Statistics. Proc. Natl. Inst. Sci. India 2(1): 49-55.

Mania, H. (2018): Das Buch der Keltischen Mythen, Von Göttern, Kriegern, Feen und Druiden, Titel der Originalausgabe: The Book of Celtic Myths. From the Mystic Might of the Celtic Worriors to the Magic of the Fey Folk, the Storied History and Folklore of Ireland, Scotland, Brittany, and Wales (Adams Media, Simon & Schuster Incorporated). - Anaconda Verlag GmbH, Köln, Deutschland.

MARANISE, A. M. J. (2012): Investigating the Syncretism of Catholicism and Voodoo in New Orleans. - J. Relig. Soc. 14.

MERKT, F. & QUACK, M. (2011): Molecular Quantum Mechanics and Molecular Spectra, Molecular Symmetry, and Interaction of Matter with Radiation. - In: Merkt, F. & Quack, M. (Edn.): Handbook of High-Resolution Spectroscopy Volume 1. - John Wiley & Sons, Hoboken, New Jersey, USA, 1-55.

MINISTÈRE DE L'AGRICULTURE ET DE L'HYDRAULIQUE (2011): Monographie Regionale Hauts-Bassins, Enquête nationale sur l'accès des ménages aux ouvrages d'assainissement familial - 2010. - Direction générale de l'assainissement des eaux usées et excreta, Ministère de l'Agriculture et de l'Hydraulique, Ouagadougou, Burkina Faso.

MINISTÈRE DE L'ECONOMIE ET DES FINANCES (2008): Recensement General de la Population et de L'Habitation de 2006, Resultats Definitifs. - Bureau Central du Recensement, Comite National du Recensement, Ministère de l'Economie et des Finances, Ouagadougou, Burkina Faso.

MINISTÈRE DE L'ENVIRONNEMENT ET DU CADRE DE VIE (2007): Situation Des Forêts Classées Du Burkina Faso Et Plan De Réhabilitation. - Direction Des Forêts, Direction Générale De La Conservation De La Nature, Secretariat Général, Ministère de l'Environnement et du Cadre de Vie, Ouagadougou, Burkina Faso.

Mokgobi, M. G. (2014): Understanding traditional African healing. - Afr. J. Phys. Health Educ. Recreat. Dance 20(Suppl. 2): 24-34.

NOBLE, S. D. & BROWN, R. B. (2009): Plant species discrimination using spectral/spatial descriptive statistics. - Image Analysis for Agricultural Products and Processes, 15. Workshop Computer-Bildanalyse in der Landwirtschaft: 82-92.

**OLTHOF, S.** (2019): Den Energieniveaus auf der Spur. - Phys. J. 18(8/9): 61-65.

OLUTAYO, O., MICHAEL, A., JOHN, A. A. & OLUSOLA, A. (2012): Antimicrobial activity and Elemental analysis of Cassia sieberiana leaves Using Atomic Absorption Spectrometer. - J. Nat. Prod. Plant Resour. 2(1): 9-18.

OMG (OBJECT MANAGEMENT GROUP) (2017): OMG(R) Unified Modeling Language (R) (OMG UML(R)), Version 2.5.1. - Object Management Group Incorporated, Needham, Massachusetts, USA.

OZIOMA, E.-O. J. & CHINWE, O. A. N. (2019): Herbal Medicines in African Traditional Medicine.
- In: Builders, P. (Ed.): Herbal Medicine. - IntechOpen Limited, London, England, 191-214.

PAHLOW, M. (2013): Das große Buch der Heilpflanzen, Gesund Durch Die Heilkräfte Der Natur. - Gräfe und Unzer Verlag GmbH, München, Deutschland.

PEBESMA, E., BIVAND, R., ROWLINGSON, B., GOMEZ-RUBIO, V., HIJMANS, R., SUMNER, M., MACQUEEN, D., LEMON, J., O'BRIEN, J. & O'ROURKE, J. (2018): Classes and Methods for Spatial Data (Version: 1.3-1). - R Foundation for Statistical Computing, Wien, Österreich.

PODA, J.-N., BELEM/OUEDRAOGO, M. & DIBLONI, O. T. (2014): Mare aux Hippopotames Biosphere Reserve. - In: Schaaf, T., Cardenas, M. R. & Lee, C. (Edn.): Innovative Ways For A Sustainable Use Of Drylands, Final Report Of The SUMAMAD Project. - United Nations Educational, Scientific and Cultural Organization (UNESCO), Paris, Frankreich, 22-39.

PROSPERE, K., McLaren, K. & Wilson, B. (2014): Plant Species Discrimination in a Tropical Wetland Using In Situ Hyperspectral Data. - Remote Sens. 6(9): 8494-8523.

PUJIONO, E., KWAK, D.-A., LEE, W.-K., SULISTYANTO, KIM, S.-R., LEE, J. Y., LEE,

- S.-H., PARK, T. & KIM, M.-I. (2013): RGB-NDVI color composites for monitoring the change in mangrove area at the Maubesi Nature Reserve, Indonesia. For. Sci. Technol. 9(4): 171-179.
- QGIS DEVELOPMENT TEAM (2018): QGIS Geographic Information System ("Madeira" 3.4.2). Open Source Geospatial Foundation Project.
- RAMDANI, F., RAHMAN, S. & SETIANI, P. (2015): Inexpensive Method to Assess Mangroves Forest through the Use of Open Source Software and Data Available Freely in Public Domain. J. Geogr. Inf. Syst. 7(1): 43-57.
- RAMIREZ, C. R. (2007): Ethnobotany and the Loss of Traditional Knowledge in the 21st Century. Ethnobot. Res. Appl. 5: 245-247.
- R CORE TEAM (2019): R: A Language and Environment for Statistical Computing (Version 3.5.3 (2019-03-11) "Great Tuth"). R Foundation for Statistical Computing, Wien, Österreich.
- R CORE TEAM AND CONTRIBUTORS (2019): The R Stats Package (Version: 3.5.3). R Foundation for Statistical Computing, Wien, Österreich.
- **REDDY, C. S.** (2015): Applications of Remote Sensing in Plant Sciences: An Overview. In: Bahadur, B., Rajam, M. V., Sahijram, L. & Krishnamurthy, K. V. (Edn.): Plant Biology and Biotechnology: Volume 1: Plant Diversity, Organization, Function and Improvement. Springer India, Neu Delhi, Indien, 713-727.
- REGIONALER PLANUNGSVERBAND MECKLENBURGISCHE SEENPLATTE (2017): Regionales Entwicklungskonzept Mecklenburgische Seenplatte. Regionaler Planungsverband Mecklenburgische Seenplatte, Neubrandenburg, Deutschland.
- RICHARDS, J. A. & JIA, X. (2006): Remote Sensing Digital Image Analysis, An Introduction. Springer-Verlag Berlin Heidelberg, Berlin, Deutschland.
- Rosswog, S. (2020): Strontium aus kosmischen Kollisionen. Phys. J. 19(1): 22-23.
- ROTTENSTEINER, F. (2017): Kontextbasierte Ansätze in der Bildanalyse. In: Heipke, C. (Ed.): Handbuch der Geodäsie, Photogrammetrie und Fernerkundung. Springer-Verlag GmbH Deutschland, Berlin, Deutschland, 555-602.
- RUESS, K. A. & MÜLLER, M. (2017): Tropische Böden, Verwitterung und Landnutzung in Westafrika.
   In: Runge, J. & Anamou, N. (Edn.): Westafrika Geografische Studienreisen in Bénin, Burkina Faso und Togo. Institut für Physische Geographie, Goethe-Universität Frankfurt am Main, Frankfurt am Main, Deutschland, 57-65.
- SCHAEPMAN, M. E. (2009): Imaging Spectrometers. In: Warner, T. A., Nellis, M. D. & Foody, G. M. (Edn.): The SAGE Handbook of Remote Sensing. SAGE Publications, Thousand Oaks, Kalifornien, USA, 166-178.
- SHARMA, R. K. (2017): Various Spectroscopic Techniques. In: Gurjar, B. R., & Kumar, P. (Edn.): Environmental Pollution Monitoring, Modeling and Control 1st Edition. Studium Press Limited Liability Company, Houston, Texas, USA, 181-206.
- STATISTISCHES AMT MECKLENBURG-VORPOMMERN (2019): Bevölkerungsstand der Kreise, Ämter und Gemeinden in Mecklenburg-Vorpommern. Statistisches Amt Mecklenburg-Vorpommern, Schwerin, Deutschland.
- STOOP, N., VERPOORTEN, M. & DECONINCK, K. (2019): Voodoo, Vaccines and Bed Nets. Econ. Dev. Cult. Change 67(3): 493-535.
- SUTCLIFFE, J. & DUIN, N. (1992): A History Of Medicine. Barnes & Noble Books, New York, USA.
- TAN, S.-Y. (2008): Satellite Biomodelling: Use Of Remote Sensing For Predicting Plant Biodiversity. Proceedings of the International Conference "Studying, Modeling and Sense Making Of Planet Earth".
- TARANTINO, C., CASELLA, F., ADAMO, M., LUCAS, R., BEIERKUHNLEIN, C. & BLONDA,

- **P.** (2019): Ailanthus altissima mapping from multi-temporal very high resolution satellite images. ISPRS J. Photogramm. Remote Sens. 147: 90-103.
- THILAGAVATHI, K. & VASUKI, A. (2019): Dimension Reduction Methods For Hyperspectral Image: A Survey. Int. J. Eng. Adv. Technol. 8(2S): 160-167.
- THIOMBIANO, A., SCHMIDT, M., DA, S., HAHN-HADJALI, K., ZIZKA, G. & WITTIG, R. (2010a): The current state of biodiversity of Burkina Faso 6.3 Vascular plants: Flowering plants. In: Thiombiano, A. & Kampmann, D. (Edn.): Biodiversity Atlas of West Africa Volume II Burkina Faso. Project management agency of the German Aerospace Center (PT-DLR), Environment, Culture, Sustainability, Bonn, Deutschland, 184-192.
- THIOMBIANO, A., SCHMIDT, M., ZIZKA, A., KÖNIG, K. & NACOULMA, B. M. I. (2010b): The current state of biodiversity of Burkina Faso 6.11 Woody plants. In: Thiombiano, A. & Kampmann, D. (Edn.): Biodiversity Atlas of West Africa Volume II Burkina Faso. Project management agency of the German Aerospace Center (PT-DLR), Environment, Culture, Sustainability, Bonn, Deutschland, 232-250.
- Toth, С. & Jutzi, В. (2017): Plattformen und Sensoren für die Fernerkundung und deren Geopositionierung. In: Heipke, С. (Ed.): Handbuch der Geodäsie, Photogrammetrie und Fernerkundung. Springer-Verlag GmbH Deutschland, Berlin, Deutschland, 29-64.
- TRAORE, S. & ANNE, C. A. T. (2010): Principal environmental factors of Burkina Faso 4.3 Soils. In: Thiombiano, A. & Kampmann, D. (Edn.): Biodiversity Atlas of West Africa Volume II Burkina Faso. Project management agency of the German Aerospace Center (PT-DLR), Environment, Culture, Sustainability, Bonn, Deutschland, 130-133.
- TRAUTMANN, T. (2017): Tuareg in Burkina Faso und Mali Ein Gottesstaat im Sahel?. In: Runge, J. & Anamou, N. (Edn.): Westafrika Geografische Studienreisen in Bénin, Burkina Faso und Togo. Institut für Physische Geographie, Goethe-Universität Frankfurt am Main, Frankfurt am Main, Deutschland, 163-168.
- UMWELTMINISTERIUM MECKLENBURG-VORPOMMERN (2003): Gutachtliches Landschaftsprogramm Mecklenburg-Vorpommern. Referat Landschaftsplanung und integrierte Umweltplanung, Umweltministerium Mecklenburg-Vorpommern, Schwerin, Deutschland.
- UUDEBERG, K., ANSKO, I., PÕRU, G., ANSPER, A. & REINART, A. (2019): Using Optical Water Types to Monitor Changes in Optically Complex Inland and Coastal Waters. Remote Sens. 11(19,2297).
- UURTIO, V., MONTEIRO, J. M., KANDOLA, J., SHAWE-TAYLOR, J., FERNANDEZ-REYES, D. & ROUSU, J. (2017): A Tutorial on Canonical Correlation Methods. ACM Comput. Surv. 50(6,95).
- WHO (2001): Legal Status of Traditional Medicine and Complementary / Alternative Medicine: A Worldwide Review. World Health Organization (WHO), Genf, Schweiz.
- WORLD BANK GROUP (2018): Country Partnership Framework For Burkina Faso For the Period FY18-FY23. World Bank Group, Washington, District of Colombia, USA.
- XIE, Y., SHA, Z. & YU, M. (2008): Remote sensing imagery in vegetation mapping: a review. J. Plant Ecol. 1(1): 9-23.
- Xue, J. & Su, B. (2017): Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. J. Sens. 2017.
- YUAN, H., MA, Q., YE, L. & PIAO, G. (2016): The Traditional Medicine and Modern Medicine from Natural Products. Molecules 21(5,559).
- **ZEIDLER, E.** (2013): Die Maximum-Likelihood-Methode zur Gewinnung von Parameterschätzungen. In: Zeidler, E. (Ed.): Springer-Taschenbuch der Mathematik. Springer Spektrum, Vieweg+Teubner Verlag, Wiesbaden, Deutschland, 867-868.
- ZHANG, T., Su, J., LIU, C., CHEN, W.-H., LIU, H. & LIU, G. (2017): Band Selection in

Sentinel-2 Satellite for Agriculture Applications. - Proceedings of the 23rd International Conference on Automation & Computing.

#### Datenquellen

BUNDESAMT FÜR KARTOGRAPHIE UND GEODÄSIE (BKG) (2018): Digitales Landschaftsmodell 1:250000 DLM250. - Bundesamt für Kartographie und Geodäsie, Leipzig, Deutschland.

ESA (2019): Sentinel 2 Level-2A Imagery. - European Space Agency (ESA), Paris, Frankreich.

FOOD AND AGRICULTURE ORGANIZATION (FAO) (2008): Global Administrative Unit Layers (GAUL) Level 1, WorldMap. - Harvard University, Cambridge, Massachusetts, USA.

GBIF (2020): GBIF Occurrence. - Global Biodiversity Information Facility Secretariat, Copenhagen, Dänemark.

GEOFABRIK GMBH & OPENSTREETMAP CONTRIBUTORS (2020): data extracts from the OpenStreetMap project. - OpenStreetMap Foundation, Geofabrik GmbH, Karlsruhe, Deutschland.

NASA JET PROPULSION LABORATORY (NASA JPL) (2013): NASA Shuttle Radar Topography Mission Global 1 arc Second Version 3. - NASA Earth Observing System Data and Information System (EOSDIS) Land Processes Distributed Active Archive Center (DAAC), NASA JPL, Pasadena, Kalifornien, USA.

RAMSAR SECRETARIAT (2020): Ramsar Sites Information Service. - The Secretariat of the Convention on Wetlands (Ramsar Secretariat), Gland, Schweiz.

#### Abbildungsverzeichnis

- Abbildung 1: Übersichtskarte für das Untersuchungsgebiet bei Kraase; Datengrundlagen:
  Verwaltungsgrenzen, Gewässer, Flüsse, Städte, Straßennetz: (Bundesamt für
  Kartographie und Geodäsie, 2018), DEM: (NASA JET PROPULSION LABORATORY,
  2013); bereitgestellt mit QGIS (QGIS DEVELOPMENT TEAM, 2018)
- Übersichtskarte für das Untersuchungsgebiet bei Bobo Dioulasso; Datengrundlagen: Verwaltungs- und Staatengrenzen: (Food and Agriculture Organization, 2008), Ortschaften, Gewässer, Téré Wald, Straßennetz und Eisenbahnlinien: (Geofabrik Gmbh, 2020), Samendéni Stausee: berechnet mit WIW (Lefebvre, 2019) in R (R Core Team, 2019) auf Grundlage einer Sentinel 2 Aufnahme vom 13.01.2019 (ESA, 2019), Mare aux Hippopotames Naturschutzgebiet: (Ramsar Secretariat, 2020), DEM: (NASA Jet Propulsion Laboratory, 2013); bereitgestellt mit QGIS (QGIS Development Team, 2018)
- Abbildung 3: Ausschnitt aus einer Aufnahme des MSI der Sentinel 2 Mission vom 26.07.2019 (ESA, 2019) mit Spektrum für ein Gebiet auf dem kleinen Varchentiner See und typischen Spektren für Wasser aus hoch aufgelösten Messungen (UUDEBERG, 2019); Karte zusammengestellt mit QGIS (QGIS DEVELOPMENT TEAM, 2018); Darstellung des Spektrums mit R (R CORE TEAM, 2019)
- Abbildung 4: Spektralbänder des MSI von Sentinel 2 aus (GMES PROGRAMME, 2012)
- Abbildung 5: Sentinel 2 Produkttypen aus (ESA, 2015)
- Abbildung 6: Preprocessing Prozesskette; UML-Aktivitätsdiagramm nach (OMG, 2017); Zusammenstellung mit Dia (Larsson, 2011)
- Abbildung 7: Streudiagramme für Merkmale aus Beispielaufnahme, Detail mit Histogrammen und Pearson'schen Korrelationskoeffizienten; Datengrundlage: (ESA, 2019); zusammengestellt mit R (R CORE TEAM, 2019)
- Abbildung 8: Workflow Merkmalsauswahl; UML-Aktivitätsdiagramm nach (OMG, 2017); Zusammenstellung mit Dia (LARSSON, 2011)
- Abbildung 9: Ausschnitte aus Resultaten von Einklassenfall ML-Klassifikationen mit und ohne Merkmalsauswahl nach minimaler Korrelation unter einem Schwellwert mit RGB-Echtfarbenkomposit der Ausgangsdaten; Datengrundlage: (ESA, 2019); Karten zusammengestellt mit QGIS (QGIS DEVELOPMENT TEAM, 2018)
- Abbildung 10: Entwicklung des Stichprobenumfangs zur Ermittlung des Korrelationskoeffizienten; berechnet und dargestellt mit R (R CORE TEAM, 2019)
- Abbildung 11: absolute Fehler von aus Stichproben unterschiedlichen Umfangs ermittelten Korrelationskoeffizienten für unterschiedlich stark korrelierte Merkmale; Datengrundlage: (ESA, 2019); berechnet und dargestellt mit R (R CORE TEAM, 2019)
- Abbildung 12: Beispiel k-Means Clustering einer Punktwolke; berechnet und dargestellt mit R (R CORE TEAM, 2019)
- Abbildung 13: Workflow k-Means Clustering; UML-Aktivitätsdiagramm nach (OMG, 2017); Zusammenstellung mit Dia (LARSSON, 2011)
- Abbildung 14: Workflow EM Clustering; UML-Aktivitätsdiagramm nach (OMG, 2017); Zusammenstellung mit Dia (LARSSON, 2011)
- Abbildung 15: grundlegender Workflow für Klassifizierungsalgorithmen; UML-Aktivitätsdiagramm nach (OMG, 2017); Zusammenstellung mit Dia (LARSSON, 2011)
- Abbildung 16: Begrenzungen von Klassen bei der MLC aus (Ahmad, 2012)
- Abbildung 17: Beispiel für das Resultat einer MLC im Merkmalsraum; Datengrundlage: (ESA, 2019); berechnet und dargestellt mit R (R CORE TEAM, 2019)

- Abbildung 18: Regressionsgerade für Merkmalsvektoren zweier Bildpunkte mit Ausschnitt aus der Beispielaufnahme und Positionen der Bildpunkte; Datengrundlage: (ESA, 2019); berechnet und dargestellt mit R (R CORE TEAM, 2019); Karte zusammengestellt mit QGIS (QGIS DEVELOPMENT TEAM, 2018)
- Abbildung 19: Beispiel für das Resultat einer SCM Klassifikation im Merkmalsraum; Datengrundlage: (ESA, 2019); berechnet und dargestellt mit R (R CORE TEAM, 2019)
- Abbildung 20: Beispiel für das Resultat einer Klassifikation mithilfe der kanonischen Korrelationsanalyse im Merkmalsraum;

  Datengrundlage: (ESA, 2019); berechnet und dargestellt mit R
  (R CORE TEAM, 2019)
- Abbildung 21: Beispiel für das Resultat einer Random Forest Klassifikation im Merkmalsraum; Datengrundlage: (ESA, 2019); berechnet und dargestellt mit R (R CORE TEAM, 2019)
- Abbildung 22: Workflow Fernerkundung von Nadel- und Mischwald; UML-Aktivitätsdiagramm nach (OMG, 2017); Zusammenstellung mit Dia (LARSSON, 2011)
- Abbildung 23: Überblick über das Waldgebiet östlich von Möllenhagen mit Bildern von der Situation vor Ort; Datengrundlage: (ESA, 2019); bereitgestellt mit QGIS (QGIS DEVELOPMENT TEAM, 2018)
- Abbildung 24: Clusterverteilungen und Trainingsgebiete für die Aufnahme vom 26.7.2019; Datengrundlage: (ESA, 2019); berechnet mit R (R CORE TEAM, 2019); dargestellt mit QGIS (QGIS DEVELOPMENT TEAM, 2018)
- Abbildung 25: Ausschnitte aus den unitemporalen Einzelergebnissen und dem multitemporalen Gesamtergebnis der multitemporalen Einklassenfall ML Klassifikation für immergrünen Nadelwald; Datengrundlage: (ESA, 2019); berechnet mit R (R CORE TEAM, 2019); dargestellt mit QGIS (QGIS DEVELOPMENT TEAM, 2018)
- Abbildung 26: Histogramme für Ausschnitte aus den unitemporalen Einzelergebnissen und dem multitemporalen Gesamtergebnis der multitemporalen Einklassenfall ML Klassifikation für immergrünen Nadelwald; berechnet und dargestellt mit R (R CORE TEAM, 2019)
- Abbildung 27: Histogramme der unitemporalen Einzelergebnisse der multitemporalen Einklassenfall ML Klassifikation für immergrünen Nadelwald; berechnet und dargestellt mit R (R CORE TEAM, 2019)
- Abbildung 28: Resultat der multitemporalen Einklassenfall ML Klassifikation für immergrünen Nadel- und Mischwald mit RGB-Echtfarbenkomposit der Aufnahme vom 16.2.2019; Datengrundlage: (ESA, 2019); berechnet mit R (R CORE TEAM, 2019); Zusammenstellung mit QGIS (QGIS DEVELOPMENT TEAM, 2018)
- Abbildung 29: Resultat der multitemporalen RF Klassifikation für immergrünen Nadel- und Mischwald mit RGB-Echtfarbenkomposit der Aufnahme vom 16.2.2019; Datengrundlage: (ESA, 2019); berechnet mit R (R CORE TEAM, 2019); Zusammenstellung mit QGIS (QGIS DEVELOPMENT TEAM, 2018)
- Abbildung 30: Trainingsgebiete für die Röhricht-Analyse; Datengrundlage: (ESA, 2019); Zusammenstellung mit QGIS (QGIS DEVELOPMENT TEAM, 2018)
- Abbildung 31: Resultat der multitemporalen Einklassenfall ML Klassifikation für Röhricht mit RGB-Echtfarbenkomposit der Aufnahme vom 16.2.2019 im Hintergrund; Datengrundlage: (ESA, 2019); berechnet mit R (R CORE TEAM, 2019); Zusammenstellung mit QGIS (QGIS DEVELOPMENT TEAM, 2018)
- Abbildung 32: Detail aus Abbildung 31 für den Kleinen Varchentiner See;
  Zusammenstellung mit QGIS (QGIS DEVELOPMENT TEAM, 2018)

- Abbildung 33: Resultat der multitemporalen RF Klassifikation für Röhricht mit RGB-Echtfarbenkomposit der Aufnahme vom 16.2.2019; Datengrundlage: (ESA, 2019); berechnet mit R (R CORE TEAM, 2019); Zusammenstellung mit QGIS (QGIS DEVELOPMENT TEAM, 2018)
- Abbildung 34: Trainingsgebiete für die Weiden-Analyse; Datengrundlage: (ESA, 2019); Zusammenstellung mit QGIS (QGIS DEVELOPMENT TEAM, 2018)
- Abbildung 35: Ausschnitte aus den Resultaten der multitemporalen Einklassenfall ML und RF Klassifikationen für Weiden, Hintergrund: RGB-Echtfarbenkomposit der Aufnahme vom 16.2.2019; Datengrundlage: (ESA, 2019); berechnet mit R (R CORE TEAM, 2019); Zusammenstellung mit QGIS (QGIS DEVELOPMENT TEAM, 2018)
- Abbildung 36: Ausschnitte aus den Resultaten der multitemporalen Euklidischen (Spektralen)
  Distanz Klassifikation für Weiden als Einklassenfall mit und ohne Berücksichtigung
  des Priori Wissens, Hintergrund: RGB-Echtfarbenkomposit der Aufnahme
  vom 16.2.2019; Datengrundlage: (ESA, 2019); berechnet mit R (R CORE TEAM,
  2019); Zusammenstellung mit QGIS (QGIS DEVELOPMENT TEAM, 2018)
- Abbildung 37: Ausschnitt aus einer Clusterverteilung für das Resultat der multitemporalen Euklidischen (Spektralen) Distanz Klassifikation für Weiden, Hintergrund: RGB-Echtfarbenkomposit der Aufnahme vom 16.2.2019; Datengrundlage: (ESA, 2019); berechnet mit R (R CORE TEAM, 2019); Zusammenstellung mit QGIS (QGIS DEVELOPMENT TEAM, 2018)
- Abbildung 38: Ausschnitt aus einer Clusterverteilung für das Resultat der multitemporalen SCM Klassifikation für Weiden, Hintergrund: RGB-Echtfarbenkomposit der Aufnahme vom 16.2.2019; Datengrundlage: (ESA, 2019); berechnet mit R (R CORE TEAM, 2019); Zusammenstellung mit QGIS (QGIS DEVELOPMENT TEAM, 2018)
- Abbildung 39: Resultat multitemporale MLC für Cassia sieberiana mit Beobachtungspositionen aus GBIF-Daten, Hintergrund: RGB-Echtfarbenkomposit der Aufnahme vom 22.2.2019; Datengrundlagen: (ESA, 2019), (GBIF, 2020); berechnet mit R (R CORE TEAM, 2019); Zusammenstellung mit QGIS (QGIS DEVELOPMENT TEAM, 2018)
- Abbildung 40: Rekonstruktion der Auftrittswahrscheinlichkeit von Vitellaria paradoxa innerhalb des Untersuchungsgebietes aus einer Karte aus dem Biodiversitätsatlas, Hintergrund: RGB-Echtfarbenkomposit der Aufnahme vom 22.2.2019; Datengrundlagen: (ESA, 2019), (Thiombiano, 2010b); berechnet mit R (R CORE TEAM, 2019); Zusammenstellung mit QGIS (QGIS DEVELOPMENT TEAM, 2018)
- Abbildung 41: Rekonstruktion der Auftrittswahrscheinlichkeit von Vitellaria paradoxa innerhalb des Untersuchungsgebietes mithilfe einer multitemporalen RFC aus der Verteilung der Auftrittswahrscheinlichkeit aus dem Biodiversitätsatlas, Datengrundlagen: (ESA, 2019), (Thiombiano, 2010b); berechnet mit R (R Core Team, 2019); Zusammenstellung mit QGIS (QGIS Development Team, 2018)
- Abbildung 42: Resultat multitemporale MLC für Vitellaria paradoxa mit Beobachtungspositionen aus GBIF-Daten, Hintergrund: RGB-Echtfarbenkomposit der Aufnahme vom 22.2.2019; Datengrundlagen: (ESA, 2019), (GBIF, 2020); berechnet mit R (R CORE TEAM, 2019); Zusammenstellung mit QGIS (QGIS DEVELOPMENT TEAM, 2018)

#### Formelverzeichnis

- (1): Pearson'scher Korrelationskoeffizient (Bronstein, 2012)
- (2): erste Näherung Stichprobenumfang zur Bestimmung von Korrelationskoeffizienten (BONETT, 2000)
- (3): erste partielle Ableitung von Gleichung 2 nach  $\tilde{\theta}$
- (4): zweite partielle Ableitung von Gleichung 2 nach  $\tilde{\theta}$
- (5): Jeffries-Matusita Abstand (RICHARDS, 2006)
- (6): Jeffries-Matusita Abstand für multivariate Normalverteilungen (RICHARDS, 2006)
- (7): Bhattacharyya Abstand (RICHARDS, 2006)
- (8): Euklidischer Abstand zwischen Clusterzentrum und Primitiv im Merkmalsraum
- (9): Schwerpunktvektor für Cluster aus zugeordneten Primitiven
- (10): vollständige Likelihood-Funktion für EM-Clustering mit multivariaten Gauß'schen Normalverteilungen; abgewandelt und übertragen auf Problemstellung nach (HASTIE, 2009)
- (11): logarithmierte vollständige Likelihood-Funktion für EM-Clustering mit multivariaten Gauß'schen Normalverteilungen; abgewandelt und übertragen auf Problemstellung nach (HASTIE, 2009)
- (12): Gewichte im E-Schritt beim EM-Clustering; abgewandelt und übertragen auf Problemstellung nach (HASTIE, 2009)
- (13): neuer Schätzwert für Mittelwertsvektor im M-Schritt beim EM-Clustering; abgewandelt und übertragen auf Problemstellung nach (HASTIE, 2009)
- (14): neuer Schätzwert für Varianz-Kovarianz-Matrix im M-Schritt beim EM-Clustering; abgewandelt und übertragen auf Problemstellung nach (HASTIE, 2009)
- (15): neue Priori-Wahrscheinlichkeiten für Cluster im M-Schritt beim EM-Clustering; abgewandelt und übertragen auf Problemstellung nach (HASTIE, 2009)
- (16): Satz von Bayes bei der MLC; abgewandelt nach (Ahmad, 2012)
- (17): Satz von der totalen Wahrscheinlichkeit bei MAP Klassifikation; abgewandelt nach (Ahmad, 2012)
- (18): logarithmierte Likelihood-Funktion bei MLC; abgewandelt nach (Albertz, 2009), (Ahmad, 2012) und (Hastie, 2009)
- (19): quadrierter Mahalanobis Abstand; abgewandelt nach (Albertz, 2009)
- (20): Ansatz zur Herleitung der Klassengrenzen bei der Mahalanobis Abstand Klassifikation; Ansatz nach (Ahmad, 2012), Mahalanobis Abstand abgewandelt nach (Albertz, 2009)
- (21): Gleichung für Klassengrenzen bei der Mahalanobis Abstand Klassifikation; Mahalanobis Abstand abgewandelt nach (Albertz, 2009)
- (22): Pearson'scher Korrelationskoeffizient in Vektorform für SCM Klassifikation; Pearson'scher Korrelationskoeffizient nach (Bronstein, 2012)
- (23): Ansatz zur Herleitung der Klassengrenzen bei der SCM Klassifikation; Ansatz nach (Ahmad, 2012), Pearson'scher Korrelationskoeffizient abgewandelt nach (Bronstein, 2012)
- (24): Gleichung für Klassengrenzen bei der SCM Klassifikation; Pearson'scher Korrelationskoeffizient abgewandelt nach (Bronstein, 2012)
- (25-26): Linearkombinationen bei der kanonischen Korrelationsanalyse; Kanonische Korrelationsanalyse nach (Uurtio, 2017), Übertragung auf die Klassifikation von Fernerkundungsdaten nach (Lee, 1999)
- (27-29): Varianz-Kovarianz-Matrizen bei der kanonischen Korrelationsanalyse; Kanonische Korrelationsanalyse nach (Uurtio, 2017), Übertragung auf die Klassifikation von Fernerkundungsdaten nach (Lee, 1999)

#### **Tabellenverzeichnis**

- Tabelle 1: spektrale Indizes zur Merkmalsraumerweiterung
- Tabelle 2: Pearson'sche Korrelationskoeffizienten für Merkmale aus Beispielaufnahme (ESA, 2019); berechnet mit R (R CORE TEAM, 2019)
- Tabelle 3: statistische Kenngrößen für resultierende Verteilungen der Likelihood-Funktionen für immergrünen Nadelwald innerhalb der Trainingsgebiete bei der multitemporalen Einklassenfall ML Klassifikation; berechnet mit R (R CORE TEAM, 2019)
- Tabelle 4: statistische Kenngrößen für resultierende Verteilungen der Likelihood-Funktionen für immergrünen Nadelwald innerhalb der Trainingsgebiete bei der multitemporalen Einklassenfall ML Klassifikation (relative Werte); berechnet mit R (R CORE TEAM, 2019)
- Tabelle 5: Spektralbänder und spektrale Indizes der Merkmalsauswahl für die Aufnahmen bei der Röhricht-Analyse
- Tabelle 6: statistische Kenngrößen für unitemporale Verteilungen der Likelihood-Funktionen für Röhricht inner- und außerhalb der nach dem Endergebnis mit Röhricht bewachsenen Gebiete; berechnet mit R (R CORE TEAM, 2019)
- Tabelle 7: statistische Kenngrößen für multitemporale Verteilungen der Likelihood-Funktionen für Röhricht inner- und außerhalb der nach dem Endergebnis mit Röhricht bewachsenen Gebiete; berechnet mit R (R CORE TEAM, 2019)
- Tabelle 8: Spektralbänder und spektrale Indizes der Merkmalsauswahl für die Aufnahmen bei der Fernerkundung von Cassia sieberiana
- Tabelle 9: Jeffries-Matusita Abstände für die drei unterschiedlichen Klassen der Auftrittswahrscheinlichkeit von Vitellaria paradoxa in den Trainingsdaten
- Tabelle 10: Wahrheitsmatrix mit Genauigkeitsmaßen für das Resultat der Maximum Likelihood Klassifikation; berechnet mit R (R CORE TEAM, 2019)
- Tabelle 11: Wahrheitsmatrix mit Genauigkeitsmaßen für das Resultat der Mahalanobis Abstand Klassifikation; berechnet mit R (R CORE TEAM, 2019)
- Tabelle 12: Wahrheitsmatrix mit Genauigkeitsmaßen für das Resultat der Euklidischen (Spektralen) Distanz Klassifikation; berechnet mit R (R CORE TEAM, 2019)
- Tabelle 13: Wahrheitsmatrix mit Genauigkeitsmaßen für das Resultat der Korrelationskoeffizienten-basierten Klassifikation; berechnet mit R (R CORE TEAM, 2019)
- Tabelle 14: Wahrheitsmatrix mit Genauigkeitsmaßen für das Resultat der Klassifikation mithilfe der Kanonischen Korrelationsanalyse; berechnet mit R (R CORE TEAM, 2019)
- Tabelle 15: Wahrheitsmatrix mit Genauigkeitsmaßen für das Resultat der Random Forest Klassifikation; berechnet mit R (R CORE TEAM, 2019)
- Tabelle 16: Wahrheitsmatrix mit Genauigkeitsmaßen für das Resultat der multitemporalen Random Forest Klassifikation; berechnet mit R (R CORE TEAM, 2019)

# Eidesstattliche Erklärung

Hiermit versichere ich, die vorliegende Masterarbeit ohne Hilfe Dritter und nur mit den angegebenen Quellen und Hilfsmitteln angefertigt zu haben. Alle Stellen, die aus den Quellen entnommen wurden, sind als solche kenntlich gemacht worden. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Neubrandenburg, den