



HOCHSCHULE  
NEUBRANDENBURG  
University of Applied Sciences

Hochschule Neubrandenburg

Studiengang Geoinformatik

# Erfassung von dreidimensionalen Gewässerumringen aus Fernerkundungsdaten

**Bachelorarbeit**

Vorgelegt von: Ols de Wilde

Zum Erlangen des akademischen Grades

**„Bachelor of Engineering“ (B.Eng.)**

Erstprüfer: Prof. Dr. Frank Günther

Zweitprüfer: Dipl.-Ing. Juliane Peters

Eingereicht am: 02.01.2026

URN: urn:nbn:de:gbv:519-thesis2025-0247-9



# Eigenständigkeitserklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe.

Alle Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen sind, wurden unter Angabe der Quelle kenntlich gemacht.

Die Arbeit ist nie in gleicher oder ähnlicher Fassung, auch nicht auszugsweise, für eine andere Prüfungs- oder Studienleistung verwendet worden.

Ols de Wilde

Schwerin, 02.01.2026

Ort, Datum

# Kurzfassung

In der vorliegenden Arbeit wird die automatisierte Erfassung von Gewässerausdehnungen und der anschließenden Ableitung von Gewässeroberflächenhöhen für diese beschrieben. Die zweidimensionale Erfassung basiert grundlegend auf Digitalen Orthophotos, wohingegen die dreidimensionale Ableitung mit Hilfe von Airborne-Laserscanning (ALS) Daten erfolgt. Die Methode wird beispielhaft auf drei landschaftlich verschiedene Untersuchungsgebiete angewendet.

Die zweidimensionale Methodik basiert auf einer überwachten Klassifikation, die anhand manuell erstellter Referenzflächen bewertet wird. Bei der Gewässerklassifikation werden bewachsene und verdeckte Bereiche ausgeschlossen, sodass ausschließlich die Ausdehnung von unbedeckten Wasserflächen extrahiert wird. Die Genauigkeitsbewertung erfolgt mit Hilfe einer Konfusionsmatrix und anschließenden Leistungsmetriken wie Precision, Recall und F1.

Die dreidimensionale Methodik unterscheidet zwischen festen Stand- und variablen Fließgewässerrhöhen. Die Höhendaten werden anhand eines aus ALS-Punkten erstellten Digitalen Geländemodells (DGM) den Gewässern zugewiesen. Die Bewertung erfolgt anhand einer stichprobenbasierten Höhenanalyse zwischen der abgeleiteten Gewässerrhöhe und dem offiziellen DGM des Amtes für Geoinformation, Vermessungs- und Katasterwesen. Anschließend werden Bias, MAE und RMSE als Leistungsmetriken berechnet.

Die 2D-Ergebnisse zeigen eine sehr hohe Gesamtgenauigkeit von  $> 99\%$ , jedoch ist die Metrik aufgrund starker Klassenungleichverteilung ungeeignet. Der F1-Score ist zur Gesamtbewertung besser geeignet. Die Genauigkeiten unterscheiden sich je nach Vegetation, Uferstruktur und Trübung der Gewässer, betragen jedoch mindestens  $\approx 97\%$ .

Die 3D-Ergebnisse weisen zum Teil deutliche Unterschiede zwischen den Untersuchungsgebieten auf. In eher naturnahen Standgewässern können Höhengenaugkeiten von wenigen einstelligen Zentimetern erreicht werden, wohingegen in urbanen und Hafengebieten größere Abweichungen auftreten können.

Der automatisierte Workflow ermöglicht eine zuverlässige Erfassung unbedeckter Wasserflächen. Eine realistische Abschätzung der Höhen ist im Idealfall auf wenige Zentimeter möglich. Besonders bei großflächigen Analysen erleichtert der Workflow die manuelle Arbeit bedeutend. Einschränkungen gibt es bei stark bewachsenen Gewässern oder komplexen Landschaftsstrukturen.

# Abstract

The aim of this thesis is the automated mapping of water bodies and the resulting determination of water levels for these bodies. The two-dimensional mapping is based primarily on digital orthophotos, whereas the three-dimensional determination is carried out using airborne laser scanning data. The method is applied to three different study areas as examples.

The two-dimensional method is based on supervised classification, which is then evaluated using manually created reference areas. When classifying water bodies, vegetated and covered areas are excluded so that only the extent of uncovered water is extracted. Accuracy is assessed using a confusion matrix and resulting performance metrics such as precision, recall, and F1.

The three-dimensional method differentiates between fixed standing water levels and variable flowing water levels. The elevation data is assigned to the water bodies using a DEM created from ALS points. The evaluation is based on a sample-based elevation analysis between the derived water level and the official DEM of the AfGVK. Bias, MAE, and RMSE are then calculated as performance metrics.

The 2D results show a very high accuracy of  $> 99\%$ , but the metric is unsuitable due to strong class imbalance. The F1 score is more suitable for overall evaluation. Accuracy varies depending on vegetation, shoreline structure, and water clarity, but is at least  $\approx 97\%$ .

The 3D results show significant differences between the study areas in some cases. In relatively natural standing waters, height accuracies of a few single-digit centimeters can be achieved, whereas larger discrepancies can occur in urban and harbor areas.

The automated workflow allows for reliable detection of uncovered water areas. Ideally, a height estimate accurate to within a few centimeters is possible. The workflow significantly reduces manual work, especially for large-scale analyses. There are limitations in areas with dense vegetation or complex landscape structures.

# Inhaltsverzeichnis

Abkürzungsverzeichnis . . . . .	vi
Abbildungsverzeichnis . . . . .	vii
Tabellenverzeichnis . . . . .	viii
<b>1 Einführung . . . . .</b>	<b>1</b>
1.1 Problemstellung . . . . .	1
1.2 Motivation und Zielsetzung . . . . .	2
1.3 Aufbau der Arbeit . . . . .	2
<b>2 Theoretischer Hintergrund . . . . .</b>	<b>3</b>
2.1 Grundlagen der Fernerkundung . . . . .	3
2.1.1 Spektrale Signaturen . . . . .	3
2.1.2 Aktive und passive Systeme . . . . .	5
2.1.3 Auflösungen der Aufnahmesysteme . . . . .	5
2.2 Verfahren der Gewässerklassifikation . . . . .	7
2.2.1 Indizes . . . . .	7
2.2.2 (Un-)überwachte Klassifikationsverfahren . . . . .	9
2.2.3 Künstliche Intelligenz . . . . .	11
2.3 Relevanz der Höhenbestimmung . . . . .	13
2.4 Stand der Forschung . . . . .	14
<b>3 Untersuchungsgebiet und Datengrundlage . . . . .</b>	<b>15</b>
3.1 Beschreibung der Untersuchungsgebiete . . . . .	15
3.1.1 Stralsund . . . . .	15
3.1.2 Bützow . . . . .	16
3.1.3 Feldberger Seenlandschaft . . . . .	17
3.2 Eingesetzte Datensätze . . . . .	18
3.2.1 Digitales Orthophoto . . . . .	18
3.2.2 Digitales Oberflächenmodell . . . . .	19
3.2.3 Airborne Laserscanning . . . . .	19
3.2.4 Digitale Landschaftsmodelle . . . . .	20
3.2.5 LUNG-Gewässer . . . . .	20
<b>4 Methodik . . . . .</b>	<b>21</b>
4.1 Übersicht der Verarbeitungskette . . . . .	21
4.2 Ableitung der Gewässerumringe (2D-Erfassung) . . . . .	22
4.2.1 Vorverarbeitung . . . . .	23
4.2.2 Klassifikation der Gewässerflächen . . . . .	23
4.2.3 Eliminierung von Schattenfehlklassifikationen . . . . .	26
4.2.4 Nachbearbeitung . . . . .	29
4.2.5 Trennung von Fließ- und Standgewässern . . . . .	32
4.3 Ableitung der Gewässerhöhen (3D-Erfassung) . . . . .	33
4.3.1 Vorverarbeitung . . . . .	33

4.3.2	Standgewässer . . . . .	34
4.3.3	Fließgewässer . . . . .	35
<b>5</b>	<b>Auswertung der Ergebnisse . . . . .</b>	<b>38</b>
5.1	Stralsund . . . . .	41
5.2	Bützow . . . . .	43
5.3	Feldberger Seenlandschaft . . . . .	45
<b>6</b>	<b>Diskussion der Ergebnisse . . . . .</b>	<b>47</b>
6.1	Einfluss der Referenzflächen . . . . .	47
6.2	Einordnung der 2D-Ergebnisse . . . . .	48
6.3	Einordnung der 3D-Ergebnisse . . . . .	50
<b>7</b>	<b>Fazit und Ausblick . . . . .</b>	<b>54</b>
	<b>Literaturverzeichnis . . . . .</b>	<b>55</b>
<b>A</b>	<b>Bilder und Grafiken . . . . .</b>	<b>59</b>
A.1	Model Builder-Workflow . . . . .	59
<b>B</b>	<b>Python Skripte . . . . .</b>	<b>60</b>
B.1	Ausschnitt aus der iterativen Schattenberechnung . . . . .	60
B.2	Ausschnitt aus der iterativen DGM-Verschmelzung . . . . .	64

# Abkürzungsverzeichnis

**AfGVK** Amt für Geoinformation, Vermessungs- und Katasterwesen

**ALS** Airborne-Laserscanning

**DGM** Digitales Geländemodell

**DLM** Digitales Landschaftsmodell

**DOM** Digitales Oberflächenmodell

**DOP** Digitales Orthophoto

**GIS** Geoinformationssystem

**GNSS** Global Navigation Satellite System

**INS** Inertial Navigation System

**LAiV** Landesamt für innere Verwaltung

**LUNG** Landesamt für Umwelt, Naturschutz und Geologie

**MNDWI** Modified Normalized Difference Water Index

**NDVI** Normalized Difference Vegetation Index

**NDWI** Normalized Difference Water Index

# Abbildungsverzeichnis

1	Reflexionsverhalten ausgewählter Oberflächen auf der Erde (de Lange, 2013) . . . . .	4
2	Aktive und passive Aufnahmesysteme (Klaus, 2023) . . . . .	5
3	Elektromagnetisches Spektrum (Albertz, 2007) . . . . .	6
4	„Signaturkurven dreier Oberflächen und Darstellung von Pixeln zu drei Landbedeckungstypen im dreidimensionalen Raum der Reflexionsgrade dreier Landsat-TM-Kanäle“ (de Lange, 2013) . . . . .	9
5	Übersichtsdarstellung der drei Untersuchungsgebiete (eigene Darstellung) . . . . .	15
6	Vereinfachtes Flussdiagramm des automatisierten QGIS Workflows (eigene Darstellung) . . . . .	22
7	Übersicht über Trainingsgebiete im Untersuchungsgebiet Feldberger Seenlandschaft (eigene Darstellung) . . . . .	24
8	Beispielhafter Vergleich einer unüberwachten (links) zu einer überwachten Klassifikation (rechts) (eigene Darstellung) . . . . .	25
9	Vergleich der linearen und logistischen Regression (DataCamp, o. D.[a])	26
10	Beispielhafte Schattenberechnung (eigene Darstellung) . . . . .	27
11	Zwischenergebnisse der 2D-Gewässerklassifikation (Leserichtung von oben links nach unten rechts): DOP als Ausgangsdaten, überwachte Klassifikation, Schattenpolygone, schatten- und lochbereinigte Gewässerflächen sowie flächengefilterte Gewässerpolygone (eigene Darstellung) . . . . .	31
12	Veranschaulichung des Algorithmus „Attribute nach nächstem verknüpfen“ (eigene Darstellung) . . . . .	36
13	Workflow-Ergebnis in Stralsund (eigene Darstellung) . . . . .	41
14	Workflow-Ergebnis in Bützow (eigene Darstellung) . . . . .	43
15	Workflow-Ergebnis in der Feldberger Seenlandschaft (eigene Darstellung) . . . . .	45
16	Beispiel einer Referenzfläche in Stralsund (eigene Darstellung) . . . . .	47
17	Ausschnitt des Klassifikationsergebnisses bei Stralsund (eigene Darstellung) . . . . .	49
18	Ausschnitt des Klassifikationsergebnisses bei Bützow (eigene Darstellung) . . . . .	49
19	Darstellung der absoluten Höhenabweichungen in Stralsund (eigene Darstellung) . . . . .	50
20	Darstellung der absoluten Höhenabweichungen in Bützow (eigene Darstellung) . . . . .	51
21	Ausschnitt aus der 3D-Betrachtung des klassifizierten Flusslayers in Bützow (eigene Darstellung) . . . . .	52
22	Darstellung der absoluten Höhenabweichungen in der Feldberger Seenlandschaft (eigene Darstellung) . . . . .	53
23	Vollständiger Workflow im QGIS Model Builder (eigene Darstellung)	59

# Tabellenverzeichnis

1	Allgemeine Konfusionsmatrix . . . . .	38
2	Konfusionsmatrix von Stralsund . . . . .	41
3	2D-Genauigkeitsbewertung in Stralsund . . . . .	42
4	3D-Genauigkeitsbewertung in Stralsund . . . . .	42
5	Konfusionsmatrix von Bützow . . . . .	43
6	2D-Genauigkeitsbewertung in Bützow . . . . .	43
7	3D-Genauigkeitsbewertung in Bützow . . . . .	44
8	Konfusionsmatrix der Feldberger Seenlandschaft . . . . .	45
9	2D-Genauigkeitsbewertung in der Feldberger Seenlandschaft . . . . .	46
10	3D-Genauigkeitsbewertung in der Feldberger Seenlandschaft . . . . .	46

# 1 Einführung

Gewässer stellen eine zentrale Rolle hinsichtlich des Wasserhaushaltes der Erde dar, sind jedoch auch unverzichtbarer Bestandteil von infrastrukturellen, hydrologischen und ökologischen Prozessen. Sowohl die zweidimensionale Repräsentation in Form von Gewässerausdehnungen als auch die erweiterte dreidimensionale Abbildung durch Oberflächenhöhen oder Gefälle bilden die Grundlage für Hochwassermodellierungen, Infrastrukturplanungen oder den Naturschutz. Die aktuellen Fernerkundungssysteme liefern temporal und räumlich hoch aufgelöste Daten. Jedoch erfolgt die weitere Verarbeitung dieser großen Datenmengen noch häufig in manueller Arbeitsweise.

## 1.1 Problemstellung

Die vorliegende Arbeit wurde in Abstimmung mit dem Fachbereich 322 (FB 322) „Digitale Luftbild- und Fernerkundungsdaten, ATKIS-DGM“ des Amtes für Geoinformation, Vermessungs- und Katasterwesen (AfGVK) Mecklenburg-Vorpommerns erstellt. Der Fachbereich benötigt verschiedene Gewässerdaten wie die Ausdehnung oder die Höhen als Grundlage für die Weiterverarbeitung von Fernerkundungsdaten sowie die Ableitung von Geobasisdaten.

Aktuell erfolgt die Zuweisung von Höhenwerten für die Gewässeroberflächen zur Erstellung von Digitalen Geländemodellen (DGM) manuell. Bei dem Vorgehen wird von vornherein im Dreidimensionalen digitalisiert, indem die Airborne-Laserscanning (ALS) Punkte, die die Gewässeroberfläche darstellen, auf einen Höhenwert festgelegt werden. Im Anschluss erfolgt die DGM-Berechnung auf Grundlage der ALS Daten. Das manuelle Vorgehen, insbesondere die Höhenwertzuweisung, wird dabei durch subjektive Entscheidungen beeinflusst.

Für die Berechnung von Digitalen Orthophotos DOP aus Luftbildern erfolgt eine manuelle zweidimensionale Digitalisierung der Gewässerumringe. Die Ausdehnungen der Gewässer werden auf Grundlage neuer Befliegungsdaten kartiert. Der Grund für die Erstellung solcher Wassermasken sind Artefakte auf den Wasseroberflächen, die durch das Aufnahmeverfahren entstehen und mit Hilfe der Masken entfernt werden. Dieser Ansatz ist jedoch sehr zeitintensiv und abhängig von subjektiven Entscheidungen.

## 1.2 Motivation und Zielsetzung

Die Motivation für die Untersuchung in dieser Arbeit liegt in der hohen Anzahl manueller Arbeitsschritte für die Erfassung der Gewässerausdehnung und -höhe. Neben dem hohen Zeitaufwand ist die auf Dauer repetitive Arbeit auch nur schwer reproduzierbar. Insbesondere im Hinblick auf die jährlich stattfindenden Befliegungen steigt der Bedarf an der Reduktion manueller Arbeitsschritte, um ein effizienteres und möglichst automatisiertes Verfahren einsetzen zu können. Zusätzlich bietet ein automatisiertes Verfahren die Möglichkeit, die 2D-Detektion sowie die 3D-Ableitung innerhalb eines Workflows zu kombinieren. Ein weiterer Vorteil ergibt sich aus den objektiveren Ergebnissen im Vergleich zu den manuell abgeleiteten Daten.

Das Ziel der Arbeit ist die Entwicklung eines möglichst automatisierten Verfahrens zur zweidimensionalen Erfassung von Gewässerumringen und der anschließenden dreidimensionalen Ableitung von Gewässerhöhen aus Fernerkundungsdaten. Der Fokus liegt hierbei auf der Reduktion manueller Arbeitsschritte und somit der Unterstützung des bestehenden Arbeitsablaufes.

## 1.3 Aufbau der Arbeit

Die vorliegende Arbeit ist insgesamt in sieben Kapitel gegliedert. Im Anschluss an die Einführung wird in Kapitel 2 der theoretische Hintergrund erläutert. Dazu gehören einige Grundlagen der Fernerkundung, die Vorstellung verschiedener Verfahren zur Gewässerklassifikation sowie die Relevanz der Höhenbestimmung als auch ein Überblick über den aktuellen Stand der Forschung. Im folgenden dritten Kapitel werden die verschiedenen Untersuchungsgebiete beschrieben sowie ein Überblick über die im Methodikteil verwendeten Datensätze gegeben. In Kapitel 4 wird die Methodik des erstellten Workflows erläutert. Dabei wird zunächst ein allgemeiner Überblick über den Arbeitsablauf gegeben. Anschließend wird getrennt voneinander das Vorgehen für die zwei- als auch dreidimensionale Erfassung detailliert beschrieben. Im darauffolgenden fünften Kapitel werden die Ergebnisse der Auswertung für jedes Untersuchungsgebiet dargelegt. Kapitel 6 diskutiert die Ergebnisse, ordnet diese kritisch ein und beschreibt Limitationen. Zum Abschluss der Arbeit werden die wesentlichen Erkenntnisse in Form eines Fazits zusammengefasst sowie ein Ausblick auf mögliche Verbesserungspotenziale gegeben.

## 2 Theoretischer Hintergrund

### 2.1 Grundlagen der Fernerkundung

Um einen ersten Einblick in das Aufgabengebiet der Fernerkundung zu erhalten hilft es bereits, den Begriff Fernerkundung zu betrachten. Es wird versucht, aus einer gewissen Distanz, also einem räumlichen Abstand (=Fern) Untersuchungen und Analysen (=Erkundung) an einem Objekt vorzunehmen, um daraus Erkenntnisse ableiten zu können. „Fernerkundung im umfassenden Sinne ist die Aufnahme oder Messung von Objekten, ohne mit diesen in körperlichen Kontakt zu treten, und die Auswertung dabei gewonnener Daten oder Bilder zur Gewinnung quantitativer oder qualitativer Informationen über deren Vorkommen, Zustand oder Zustandsänderung und gegebenenfalls deren natürliche oder soziale Beziehungen zueinander“ (Hildebrandt, 1996).

Häufig fällt im Zusammenhang mit der Fernerkundung auch der Begriff Photogrammetrie. Die Internationale Gesellschaft für Photogrammetrie und Fernerkundung (ISPRS) definiert die Begriffe wie folgt: „*Photogrammetry is the science and technology of extracting reliable three-dimensional geometric and thematic information, often over time, of objects and scenes from image and range data. Remote sensing is the science and technology of capturing, processing and analysing imagery, in conjunction with other physical data of the Earth and the planets, from sensors in space, in the air and on the ground*“ (International Society for Photogrammetry and Remote Sensing, 2022). Die Photogrammetrie verfolgt also das Ziel der Gewinnung zuverlässiger dreidimensionaler geometrischer und thematischer Informationen, wohingegen die Fernerkundung die Erfassung, Verarbeitung und Analyse von Bilddaten in Verbindung mit anderen physikalischen Daten bestrebt. Laut Heipke (2017) kann die Photogrammetrie heute als Teil der Fernerkundung angesehen werden, da beide Disziplinen sich mit der „[...] automatischen, geometrischen und semantischen Bildauswertung“ beschäftigen. Folgend wird in der Arbeit nur noch von Fernerkundung gesprochen.

#### 2.1.1 Spektrale Signaturen

Laut de Lange (2013) basiert das Grundprinzip der Fernerkundung „[...] auf physikalischen Strahlungsvorgängen in der Atmosphäre“. Die hauptsächlich kurzwellige Sonneneinstrahlung wird durch Objekte zum Teil reflektiert und teils absorbiert. Letzteres führt dann zu einer Erwärmung der Erdoberfläche und anschließender

Emission langwelliger Wärmestrahlung in die Atmosphäre. Da jedes Objekt beziehungsweise jede Erdoberfläche ein charakteristisches Reflexions- und Emissionsverhalten besitzt, können so Rückschlüsse auf die Art und den Zustand der beobachteten Objekte geführt werden (de Lange, 2013). Diese spezifischen Strahlungsmuster werden auch Reflexionskurven oder „spectral signatures“ genannt.

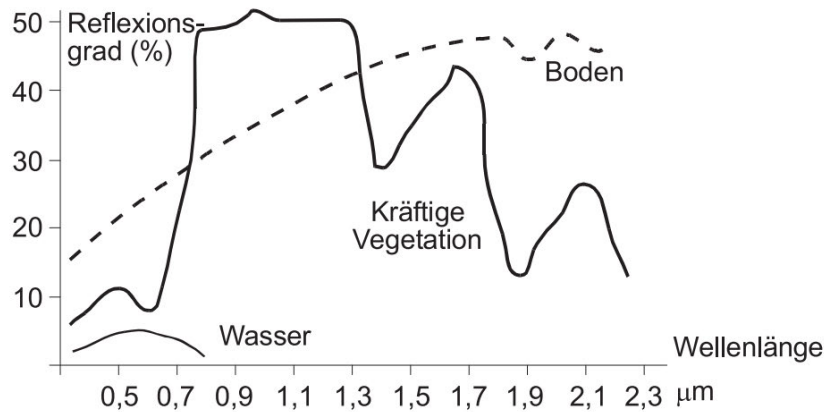


Abbildung 1: Reflexionsverhalten ausgewählter Oberflächen auf der Erde (de Lange, 2013)

Einige Beispiele für diese spezifischen Strahlungsmuster sind in Abbildung 1 dargestellt. Nach de Lange (2013) ist hier der sprunghafte Anstieg des Reflexionsgrades von Vegetation bei einer Wellenlänge von ca.  $0,8 \mu\text{m}$  besonders auffällig. Dieser Anstieg auf über 50% erfolgt beim Übergang vom sichtbaren Licht ( $0,4 - 0,7 \mu\text{m}$ ) in den nahen Infrarotbereich und trägt den Namen „Red Edge“. Er findet häufig Anwendung bei Vegetationsindizes, „z.B. zur Identifizierung des Vitalitätsgrades von Pflanzen“ (de Lange, 2013) (siehe Kapitel 2.2.1). Im Gegensatz dazu steht das Strahlungsmuster von Wasser. Es besitzt eine maximale Reflexionsrate von unter 10% im Bereich des sichtbaren Lichts und weist schon ab dem nahen Infrarot fast keine Reflexion mehr auf.

Der Autor de Lange (2013) beschreibt jedoch auch, dass die Spektralsignaturen nicht allgemein übertragbar sind. Das bedeutet, dass eine bestimmte Oberflächenart (z.B. Wasser) nicht immer die gleiche Reflexionskurve aufweist. Äußere Bedingungen wie Schattenwurf, Jahreszeit oder die Trübung verändern die Signatur teils deutlich. Um dem entgegenzuwirken müssen laut de Lange (2013) homogene Trainingsflächen für jedes Untersuchungsgebiet erstellt werden und somit die Signaturen kalibriert werden. Des Weiteren ist „Die Identifizierung [...] vor allem von der geometrischen und spektralen Auflösung des Aufnahmesystems abhängig“ (de Lange, 2013). Das hat den Grund, dass sich in dicht durchmischten Kulturlandschaften mit Gebäuden,

Straßen und Vegetation mehrere Oberflächen innerhalb eines Pixels überlagern und somit die Fläche sich nicht mehr eindeutig über eine einzelne Signaturkurve erkennen lässt (de Lange, 2013). Je kleiner die einzelnen Pixel sind, desto unwahrscheinlicher ist die Entstehung solcher Mischformen.

### 2.1.2 Aktive und passive Systeme

Bei der Datenerfassung unterscheidet man zwischen aktiven und passiven Aufnahmesystemen (siehe Abbildung 2). Laut Klaus (2023) differenzieren sie sich aufgrund ihrer Strahlungsquellen. Aktive Systeme besitzen die Fähigkeit eigene Strahlung auszusenden, sind also zugleich Sender als auch Empfänger. Die ausgesandte Strahlung wird beim Auftreffen auf Oberflächen und Objekten reflektiert beziehungsweise absorbiert und anschließend vom selben System wieder aufgezeichnet. Hierzu gehören z.B. die Radarsatelliten der Sentinel-1 Mission oder auch ALS Befliegungen.

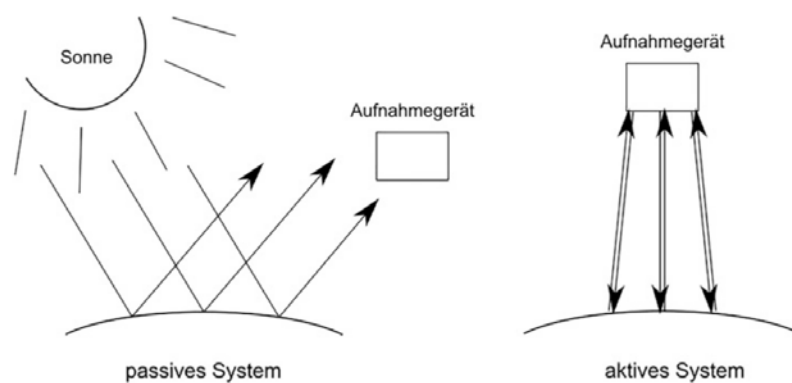


Abbildung 2: Aktive und passive Aufnahmesysteme (Klaus, 2023)

Passive Systeme hingegen, beschreibt Klaus (2023), sind auf eine externe Strahlungsquelle, in der Regel die Sonne, angewiesen. Die Strahlung durchdringt die Atmosphäre und wird ebenso wie bei aktiven Systemen von der Geländeoberfläche bzw. auf ihr befindlichen Objekten reflektiert. Im Anschluss erfassen „Die Sensoren der Kamera [...] die Intensität der reflektierten Strahlung und speichern diese auf einem Datenträger.“ (Klaus, 2023). Beispiele für passive Aufnahmesysteme sind Luftbildkameras oder auch die Sensoren der Sentinel-2 Satelliten.

### 2.1.3 Auflösungen der Aufnahmesysteme

Die folgenden Ausführungen beruhen auf den Überlegungen von de Lange (2013), die davon ausgehen, dass man zwischen vier verschiedenen Arten der Auflösung

differenziert, um die Leistungsfähigkeit eines Aufnahmesystems festzustellen und um auch diverse Systeme untereinander für die unterschiedlichsten Anwendungszwecke zu vergleichen.

Die räumliche Auflösung beschreibt die geometrischen Eigenschaften eines Aufnahmesystems. „Sie gibt die Größe eines Pixels in Meter an, d.h. die Seitenlänge der Fläche des Bodenelements, die durch ein System bei einer bestimmten Flughöhe erfasst wird.“ (de Lange, 2013). Die Sentinel-2 Satelliten erreichen beispielsweise eine räumliche Auflösung von bis zu 10 m (European Space Agency (ESA), o. D.), wohingegen DOPs eine Auflösung von bis zu 0,1 m (Landesamt für innere Verwaltung Mecklenburg-Vorpommern, o. D.[e]) aufweisen.

Multispektrale Sensoren erfassen die Strahlungsreflektionen in mehreren, getrennten Spektralbändern. Die Anzahl dieser sogenannten Kanäle beschreibt die spektrale Auflösung. Die Lage als auch Breite dieser Kanäle im elektromagnetischen Spektrum (siehe Abbildung 3) bestimmen, wie gut sich die unterschiedlichen Oberflächenarten innerhalb einer Aufnahme differenzieren lassen. Im Gegensatz zu multispektralen Sensoren arbeiten panchromatische Sensoren breitbandig und sind über den gesamten sichtbaren Bereich empfindlich. „Die Abstufung der Grauwerte ist typischen Schwarz-Weiß-Bildern vergleichbar.“ (de Lange, 2013).

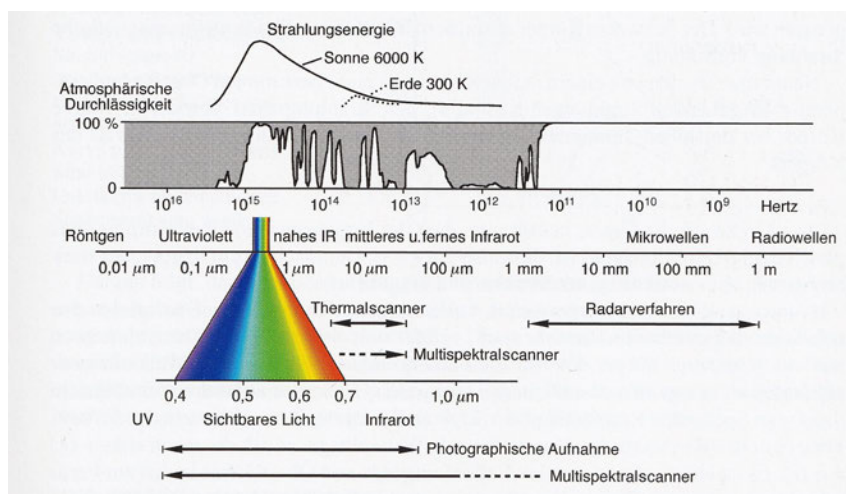


Abbildung 3: Elektromagnetisches Spektrum (Albertz, 2007)

Die radiometrische Auflösung beschreibt, wie fein ein Sensor unterschiedliche Strahlungsintensitäten abstufen kann. Sie wird durch die Anzahl der Grauwerte (Intensitätsstufen) pro Kanal beschrieben. Üblich sind 256 Graustufen pro Kanal, was einer Aufzeichnung mit 8 Bit entspricht. Jedoch weist de Lange (2013) darauf hin, dass dies eine vereinfachte Darstellung ist, da zum Beispiel auch das Signal-Rausch-Verhältnis zur radiometrischen Auflösung zuzuordnen ist.

Die temporale Auflösung (auch Repititionsrate genannt) gibt an, in welchem zeitlichen Abstand ein Gebiet von demselben Aufnahmesensor (beispielsweise Satelliten) erneut erfasst werden kann. Je breiter die Bodenspur des Sensors ist, desto kürzer ist auch die Wiederholungszeit. In nördlicheren und südlicheren Breitengraden werden Gebiete durch die Überlappung benachbarter Orbits polar umlaufender Satelliten noch häufiger überflogen, was die temporale Auflösung zusätzlich steigert (de Lange, 2013).

## 2.2 Verfahren der Gewässerklassifikation

Um Gewässerumringe auf Grundlage von Fernerkundungsdaten zu extrahieren, muss zu Beginn eine Gewässerklassifikation erfolgen. Zur Erfüllung dieser Aufgabe können verschiedene Ansätze mit jeweils unterschiedlich komplexen Strukturen gewählt werden. Einige dieser Ansätze sollen im aktuellen Kapitel betrachtet sowie ihre Vor- und Nachteile näher beleuchtet werden.

### 2.2.1 Indizes

Das Konzept der Indizes besteht bereits seit vielen Jahrzehnten, wie die Aussagen von Gao (1996) zeigen. Sie gehören „Zu den Verfahren der Bildtransformation, bei denen aus einem mehrkanaligen Bilddatensatz [...] neue Bilddaten entstehen können“ (de Lange, 2013). Der Normalized Difference Vegetation Index (NDVI) ist der wohl bekannteste und am weitesten verbreitete Index für die Vegetationserkundung aus dem Weltall (Gao, 1996).

Im Hinblick auf die Identifikation von Gewässerflächen spielt der Normalized Difference Water Index (NDWI) eine bedeutende Rolle. Hierbei ist zu beachten, dass unter NDWI zwei verschiedene Indizes verstanden werden können, da keine einheitliche Namensgebung zwischen ihnen herrscht (Kuechly u. a., 2020). So wird der Index nach Gao dazu genutzt, um den Wassergehalt von Vegetation zu untersuchen. Er wird zumeist als Ergänzung und weniger als Ersatz zum NDVI betrachtet (Gao, 1996).

$$\mathbf{NDWI} \text{ (Gao)} = \frac{\mathbf{NIR} - \mathbf{SWIR}}{\mathbf{NIR} + \mathbf{SWIR}} \quad (1)$$

Der NDWI nach McFeeters (1996) wiederum wird zur Untersuchung der Ausdehnung und Verbreitung von Wasserflächen genutzt (Kuechly u. a., 2020). Das Prinzip hinter dem Index basiert auf der vergleichsweise starken Reflexion von grüner Strahlung

und der starken Absorption von Nahinfrarotstrahlung im Wasser (siehe Abbildungen 1, 3). Dieser Ansatz verstärkt den Kontrast zwischen Wasser- und Landflächen deutlich.

$$\text{NDWI (McFeeters)} = \frac{\text{Grün} - \text{NIR}}{\text{Grün} + \text{NIR}} \quad (2)$$

Im Folgenden bezieht sich der NDWI, sofern nicht anders angegeben, auf die Definition nach McFeeters (siehe Gleichung 2). Der NDWI nimmt Werte zwischen  $-1$  und  $+1$  an, wobei Werte  $> 0,3 - 0,5$  in der Regel auf Wasserflächen hindeuten (Kuechly u. a., 2020; Sentinel-Hub, o. D.). Niedrigere Werte stellen Vegetation oder sonstige Arten der Landbedeckung dar.

Der indexbasierte Ansatz bietet den Vorteil, dass das Verfahren sehr einfach implementiert werden kann. Als Ausgangsdaten dienen allein die Satelliten- bzw. Orthophotos und die Indexberechnung kann in jedem beliebigem Geoinformationssystem (GIS) durchgeführt werden. Zudem ist das Verfahren in der Praxis weit verbreitet und anerkannt. Ein weiterer Vorteil ergibt sich aus den vergleichsweise schnellen Berechnungszeiten. Diese variieren in Abhängigkeit nach den gewählten Ausgangsdaten und ihren entsprechenden räumlichen Auflösungen, unterbieten jedoch (un-)überwachte Klassifizierungen (siehe Kapitel 2.2.2) und Deep Learning-basierte Ansätze (siehe Kapitel 2.2.3) aufgrund der dortigen Vorbereitung hinsichtlich der Trainingsgebiete und dem Trainieren der Klassifikationsmodelle.

Die Gewässerklassifikation mit Hilfe des NDWI weist wiederum auch Limitationen auf. So stellt die Schwellenwertbildung eine bedeutende Unsicherheit dar. Wird der Schwellwert zur Extraktion der Gewässer beispielsweise bei einem tendenziell höheren NDWI von  $0,6$  festgelegt, sind False Positive Werte unwahrscheinlich, jedoch steigt das Risiko für False Negative Werte. Setzt man den Schwellwert jedoch niedriger an, so steigt die Wahrscheinlichkeit für False Positive Werte, wiederum sinkt die Häufigkeit von False Negative Werten. Die Herausforderung bei der Arbeit mit Indizes liegt ganz allgemein in der Bildung von Schwellenwerten, um ein möglichst optimales Ergebnis aus True Positive und True Negative Werten zu ermitteln.

Der NDWI weist zudem weitere Schwächen hinsichtlich der Klassifikation auf. Der Index reagiert empfindlich auf bebauten Landflächen, was wiederum zu einer Überschätzung der Wasserflächen führen kann (Sentinel-Hub, o. D.). Vor diesem Hintergrund wurde der Modified Normalized Difference Water Index (MNDWI) entwickelt. Dieser nutzt anstelle des NIR das MIR-Band und verringert Fehlklassifikationen deutlich (Xu, 2006).

## 2.2.2 (Un-)überwachte Klassifikationsverfahren

Der Inhalt des folgenden Kapitels beruht im Wesentlichen auf de Lange (2013), der der rechnergestützten Klassifikation, das heißt dem „[...] Erkennen von Objekten oder Eigenschaften wie z.B. Landbedeckungstypen durch Auswertung mehrerer Kanäle“, eine besondere Bedeutung zuweist.

Das Grundprinzip der Multispektralklassifikation basiert auf den unterschiedlichen Reflexionsverhalten von Oberflächen und Objekten, die in mehreren spektralen Kanälen erfasst werden (siehe Kapitel 2.1.1). Die Reflexionswerte der einzelnen Kanäle bilden einen mehrdimensionalen Raum (sog. Merkmalsraum), in dem die Pixel idealerweise in separaten Bereichen liegen (siehe Abbildung 4). Diese Häufung von Pixeln mit ähnlichen Eigenschaften werden Cluster genannt und müssen mit Hilfe von Klassifikationsverfahren identifiziert werden. Hierbei wird zwischen der unüberwachten und überwachten Klassifikation unterschieden. Diese Cluster können dann wiederum den verschiedenen Landbedeckungstypen (z.B. Wasser, Land) zugeordnet werden.

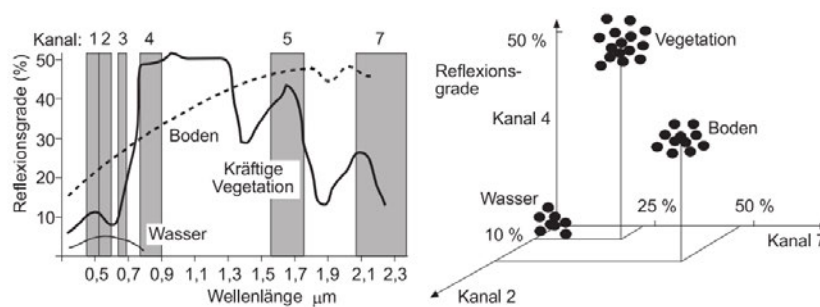


Abbildung 4: „Signaturkurven dreier Oberflächen und Darstellung von Pixeln zu drei Landbedeckungstypen im dreidimensionalen Raum der Reflexionsgrade dreier Landsat-TM-Kanäle“ (de Lange, 2013)

Der Vorteil der unüberwachten gegenüber der überwachten Klassifikation ist, dass keine Voraussetzungen gefordert sind. Es sind keine vorherigen Informationen über die zu ermittelnden Cluster nötig und der einzig benötigte Input des Nutzers ist eine Anzahl der gesuchten Cluster (Ergebnisklassen). „Die Klassifikation erfolgt automatisch, d.h. unüberwacht.“ (de Lange, 2013).

Das k-Means und das Isodata-Verfahren sind zwei häufig genutzte Algorithmen für die unüberwachte Klassifikation, wobei das Isodata-Verfahren auf dem k-Means Algorithmus basiert und seinen Funktionsumfang erweitert. Es besitzt zusätzlich die Möglichkeit, basierend auf den Standardabweichungen der Kanäle und den euklidischen Abständen zwischen den Clustermittelpunkten, Cluster zu teilen und zusammenzulegen (de Lange, 2013).

In der praktischen Anwendung überzeugt das Konzept der unüberwachten Klassifikation aufgrund seiner Simplität in der Implementierung als auch der Anwendung. Da keine Trainingsdaten benötigt werden, spart dieser Ansatz Zeit und Aufwand bei der Datenaufbereitung. Diese Tatsache führt wiederum dazu, dass die Ergebnisse häufig ungenau ausfallen, insbesondere bei komplexen Szenen. Die Initialisierung der Klassifikation in zwei Cluster (Wasser und Land) erzielt regelmäßig großflächige Fehlklassifikationen im Hinblick auf Schattenflächen, was eine Ableitung von Gewässerumringen deutlich verfälscht.

Das Konzept der überwachten Klassifikation verfolgt einen anderen Ansatz, da hier der Nutzer Vorentscheidungen bezüglich der Klassenbildung treffen muss. Voraussetzung ist, dass die Landbedeckungsklassen vorab anhand von Trainingsgebieten bestimmt werden müssen. Die Klassifikation erfolgt „[...] auf der Basis bekannter Oberflächentypen einzelner kleiner Testflächen [...] innerhalb des gesamten digitalen Bildes“ (de Lange, 2013). Wie bereits bei der unüberwachten Klassifikation wird hierbei angenommen, dass die verschiedenen Objekte und Oberflächen ein charakteristisches Reflexionsverhalten aufweisen. Die diversen Signaturkurven (siehe Kapitel 2.1.1) ermöglichen im Anschluss die Identifikation der einzelnen Landbedeckungen. Da diese jedoch nicht allgemeingültig sind und sich je nach Jahreszeit und Oberflächenbeschaffung (z.B. Trübung von Wasser) ändern, werden die Signaturkurven in Form der Trainingsgebiete für jedes Untersuchungsgebiet neu bestimmt.

Die Anforderung an die Trainingsgebiete bestehe darin, möglichst homogen hinsichtlich der Landbedeckung zu sein, da die Spektralsignaturen dieser repräsentativ für den entsprechenden Oberflächentyp im Untersuchungsgebiet seien. Die Signaturkurven definieren die gesuchten Reflexionsklassen für die Klassifikation. Das Ziel der Trainingsgebiete sei nicht die vollständige Abdeckung des Untersuchungsgebiets, sondern eine Auswahl an Pixeln, „[...] deren Bedeutung zweifelsfrei bekannt ist.“ (de Lange, 2013).

Im Anschluss an die Erstellung der Trainingsgebiete folgt die eigentliche Pixelklassifikation. Sie hat die Aufgabe, alle Pixel des gesamten Untersuchungsgebietes den zuvor ermittelten Reflexionsklassen zuzuordnen. Für die Klassifikation stehen diverse Verfahren (Classifier) zur Verfügung, darunter Minimum Distance oder Maximum Likelihood.

Zudem beschreibt de Lange (2013) Schwierigkeiten, die bei der Nutzung der überwachten Klassifikation auftreten können. So gehe das Verfahren von einer eindeutigen Klassifikation aus, was nicht der Realität entspricht. Zum Teil können Pixel sogar verschiedenen Klassen zugeordnet werden. Dies passiert, wenn z.B. mehrere

Klassen relativ heterogen sind und sich in ihren Kennwerten ähneln. Dieses Problem ist unter anderem auf Limitationen einzelner Classifier zurückzuführen. Beispielsweise wird die Trennfunktion des Minimum Distance Verfahren als zu starr betrachtet. Gruppen ließen sich nicht immer so regelhaft trennen. Ein größeres Problem wird den sogenannten Mischpixeln zugeschrieben. Hierbei repräsentieren Pixel mit einer geringen räumlichen Auflösung (vgl. Kapitel 2.1.3) mehrere Landbedeckungen gleichzeitig. Typisch dafür sei beispielsweise der Übergangsbereich zwischen Wasser und Ufer. Aufgrund der Bildung von Mischungen zwischen den spezifischen Strahlungswerten ist eine eindeutige Zuordnung nicht möglich. Infolge dessen sei mit Fehlklassifikationen zu rechnen. de Lange (2013) weist darauf hin, dass das Problem auf niedrig aufgelöste Fernerkundungsdaten zutrifft, jedoch mit der Nutzung von hochaufgelösten Daten entfällt. Nach der pixelbasierten Klassifikation beschreibt de Lange (2013) eine visuelle Unruhe, dem sog. „Salz-und-Pfeffer-Effekt“, welcher sich durch „Einsprengsel“ und „unlogische Pixelwerte“ äußert.

Die Vorteile der überwachten Klassifikation liegen, vereinfacht gesagt, in gezielteren und genaueren Ergebnissen. So bestimmt der Nutzer selbst die Klassen und kann diese auf spezifische Anforderungen anpassen. Des Weiteren werden die Signaturkurven mit Hilfe der Trainingsgebiete für jedes spezifische Untersuchungsgebiet neu bestimmt, was einer Kalibrierung für unterschiedliche Jahreszeiten oder Umgebungsbedingungen entspricht. Wichtigstes Auswahlkriterium für die überwachte Klassifizierung ist jedoch die hohe Genauigkeit. Mit guten Trainingsgebieten und einem für den Anwendungsfall passenden Classifier sind wesentlich bessere Ergebnisse als mit der unüberwachten Klassifizierung zu erzielen. Hinzu kommt die Anpassbarkeit der Trainingsgebiete. Sollte das Ergebnis nicht wie gewünscht ausfallen, können diese verfeinert bzw. erweitert werden, was wiederum die Genauigkeit der Klassifikation erhöht.

### **2.2.3 Künstliche Intelligenz**

In der grundlegend überarbeiteten und erweiterten vierten Auflage des Buches „Geoinformatik in Theorie und Praxis“ beschreibt de Lange die Klassifikation mit Hilfe von künstlichen neuronalen Netzen (de Lange, 2020). Das Grundkonzept basiere auf der Arbeitsweise des menschlichen Gehirns, bestehend aus unzähligen Nervenzellen, den sogenannten Neuronen. Innerhalb der neuronalen Netze stelle das Neuron eine Bearbeitungseinheit dar, die gewichtete Inputs aus anderen Neuronen erhält, diese summiert und bei der Überschreitung eines Schwellwerts ein Signal an das nächste Neuron sendet. Dabei laufe die Verarbeitung vorwärts gerichtet von Eingabe zu

Ausgabe ab.

Das soeben erläuterte, sehr grundlegende Ausgangsmodell müsse nun mit Hilfe des überwachten Lernens trainiert werden. Der Mechanismus beginnt laut de Lange (2020) damit, dass die einzelnen Gewichte der Inputs rein zufällig bestimmt werden. Das Ergebnis dieser Durchführung entspricht, aufgrund der zufällig gewählten Gewichte, nicht der Realität. Deshalb werden die Gewichte und Schwellwerte der Neuronen iterativ angepasst. Dieser sogenannte Backpropagation-Algorithmus bestehe aus drei Schritten. Zunächst werden die Eingabewerte eines Pixels durch das neuronale Netzwerk verarbeitet. Im Anschluss wird das Ergebnis der Ausgabe mit der korrekten Ausgabe verglichen und die Differenz in Form des Fehlers berechnet. Im letzten Schritt wird der berechnete Fehler dem Inputlayer zurückgegeben, woraufhin die Gewichte und Schwellwerte „[...] in Abhängigkeit ihres Einflusses auf den Fehler“ (de Lange, 2020) angepasst werden. Das Ziel des Algorithmus ist es, die Fehlerfunktion zu minimieren, „[...] wobei aber im Allgemeinen lediglich ein lokales Minimum gefunden wird“ (de Lange, 2020).

Neuronale Netze unterscheidet man u.a. anhand ihrer Komplexität. So gibt es laut de Lange (2020) einfache Modelle, wie im ersten Absatz als Grundkonzept beschrieben, welche auch einschichtiges Perzeptron genannt werden. Diese sind jedoch in ihrem Aufbau in der Regel zu simpel, um qualitativ hochwertige Ergebnisse zu liefern. Komplexe Modelle, oder auch mehrschichtige Perzeptrons, bestehen unter anderem aus mehreren Layern (Input, Hidden, Output). Zudem können die Eingabewerte in normalisierter Form (d.h. zwischen 0 und 1) vorliegen. Als Hidden Layer wird eine Schicht von Neuronen bezeichnet, die zwischen Eingabe- und Ausgabeschicht liegt. In ihr findet die eigentliche Verarbeitung der Eingabedaten statt. Ein Deep Learning Modell zeichnet sich, anders als bei einem Simple Neuronal Network, durch viele solcher Hidden Layers aus.

Convolutional Neural Networks (CNN) stellen laut de Lange (2020) eine Weiterentwicklung gegenüber neuronaler Netze hinsichtlich der Bildverarbeitung dar. Ein CNN besteht aus verschiedenen Schichten, darunter Convolutional, Pooling und der vollständig vermaschten Schicht. Bei diesem Verfahren liegen nur „lokal vermaschte Teilnetze“ (de Lange, 2020) vor, was einen reduzierten Speicherbedarf und eine kürzere Trainingszeit mit sich bringt.

Der Vorteil in der Nutzung neuronaler Netze liegt darin, dass diese die Fähigkeit besitzen, zu generalisieren. Neue Inputs, die den Trainingsdaten ähneln, führen meist zu korrekten Klassifikationen. Zudem müssen die Inputgrößen keine spezifische Häufigkeitsverteilung haben, sondern können unabhängig von ihrer Statistik herangezo-

gen werden. Ein dritter Vorteil besteht laut de Lange (2020) darin, dass neuronale Netze eine hohe Robustheit aufweisen. Sie sind tolerant gegenüber Rauschen in Trainingsdaten, da der Ausfall einzelner Neuronen das Ergebnis nicht bemerkenswert beeinflusst (de Lange, 2020).

Jedoch beschreibt de Lange (2020) auch Schwierigkeiten, die mit der Nutzung neuronaler Netze einhergehen. Ein Schwerpunkt liege dabei auf den langen Trainingszeiten. Diese können über mehrere Stunden hinweg andauern, da bei vielen Verbindungen entsprechend viele Gewichte zu testen sind. Des Weiteren erreicht die zuvor erläuterte Backpropagation eventuell nur ein lokales Minimum, was wiederum zu einem suboptimalen Endergebnis führt. Das Problem hängt u.a. damit zusammen, dass das Klassifikationsergebnis grundsätzlich abhängig von den den Neuronen ursprünglich zugewiesenen Gewichten ist. Dies erschwert eine Genauigkeitsvorhersage. Ein weiteres großes Risiko besteht im sogenannten Overfitting. Das bedeutet, dass das Netzwerk überangepasst ist auf die Trainingsdaten, was eine schlechte Generalisierung zur Folge hat. Dem trainierten Muster ähnliche, aber nicht in den Trainingsdaten enthaltene Pixel werden zur Folge schlechter klassifiziert.

## 2.3 Relevanz der Höhenbestimmung

Im AfGVK M-V werden DGM aus ALS-Messdaten erhoben (Landesamt für innere Verwaltung Mecklenburg-Vorpommern, o. D.[a]). Hierbei werden die Gewässerhöhen jedoch nicht direkt aus der Rohpunktwolke übernommen. Das liegt mitunter daran, dass Wasseroberflächen die Laserimpulse zum Teil so reflektieren, dass die reflektierten Strahlen den Sensor nicht wieder erreichen. Das wiederum führt zu lückenhaften oder fehlerbehafteten Ergebnissen. Aus diesem Grund werden die Wasserhöhen im Landesamt modelliert.

Aktuell wird für jedes Standgewässer, wie beispielsweise einem See, eine einheitliche Wasseroberflächenhöhe festgelegt. Dabei werden alle der Wasseroberfläche zugehörigen Punkte der Punkt wolke auf einen konstanten Höhenwert gesetzt. Bei Fließgewässern ist die Anforderung komplexer, da hier zusätzlich das Höhengefälle abgebildet wird. Das bedeutet, dass die Wasseroberflächenhöhe entlang der Fließrichtung interpoliert wird. Die Umsetzung findet manuell auf Basis vorhandener Gewässergeometrien statt. Dieser Schritt erfordert großen händischen Aufwand.

Die manuelle Höhenzuweisung geht mit verschiedenen Schwierigkeiten einher. So ist die Vorgehensweise zeitaufwendig und subjektiv. Die Höhe eines Gewässers ist somit zu einem gewissen Grad von der bearbeitenden Person abhängig. Des Weiteren

kommt es teilweise zu inkonsistenten Ergebnissen bei dem Übergang von Gewässern in andere Bearbeitungsgebiete. Dies erfordert eine zusätzliche Korrektur und Absprache zwischen den bearbeitenden Personen.

Das Potenzial der automatisierten Erfassung von dreidimensionalen Gewässerumringen ergibt sich insbesondere im Hinblick auf die Minimierung der aufgezeigten Schwierigkeiten. So werden die Gewässerumringe mit den zugehörigen Höheninformationen automatisiert abgeleitet. Dies bietet den Vorteil eines durchgängigen Workflows mit reproduzierbaren Ergebnissen. Das Ziel ist nicht, die manuelle Bearbeitung vollständig zu ersetzen, sondern vielmehr den bestehenden Arbeitsablauf zu unterstützen.

## 2.4 Stand der Forschung

In der Forschung ist das Thema der Gewässererkennung mit Hilfe von Fernerkundungsdaten seit vielen Jahrzehnten präsent. Ein simpler, aber etablierter Ansatz nutzt die Indexberechnung. McFeeters (1996) entwickelte beispielsweise im Jahr 1996 mit dem NDWI (vgl. Kapitel 2.2.1) eine zuverlässige Möglichkeit, die Ausdehnung von Gewässern zu erfassen. Andere, bereits schon lange bestehende Ansätze wie die der (un-)überwachten Klassifizierung (vgl. Kapitel 2.2.2) oder auch modernere Verfahren unter dem Einsatz von Künstlicher Intelligenz (vgl. Kapitel 2.2.3) liefern zum Teil qualitativ bessere Klassifikationsergebnisse und sind in der Forschung etabliert. Der Fokus dieser Ansätze liegt jedoch ausschließlich auf zweidimensionalen Gewässermasken, bei der die Extraktion von Gewässerhöhen keinen Bestandteil der Methodik darstellt.

Gleichzeitig gibt es viel Literatur, die sich mit der Nutzung von Höheninformationen zur Analyse von Gewässern befasst. So wurde beispielsweise bereits 2009 eine Arbeit veröffentlicht, die „[...] die hydraulischen Verhältnisse der Donau-Auenlandschaft Lobau im Osten Wiens beschreibt“ (Gschöpf und Blaschke, 2009). In dem Beispiel dient das DGM, erstellt aus ALS-Daten, als Grundlage für die Oberflächengewässermodellierung.

Zusammenfassend zeigt der Stand der Forschung, dass es sowohl etablierte und leistungsstarke Ansätze zur zweidimensionalen Gewässerextraktion als auch diverse Anwendungszwecke von Höheninformationen gibt. Diese werden in der Regel jedoch getrennt voneinander betrachtet. Vor diesem Hintergrund untersucht die vorliegende Arbeit, inwiefern Gewässerumringe mit Höheninformationen möglichst automatisch aus Fernerkundungsdaten abgeleitet werden können.

## 3 Untersuchungsgebiet und Datengrundlage

### 3.1 Beschreibung der Untersuchungsgebiete

Die in dieser Arbeit betrachteten Untersuchungsgebiete befinden sich im Bundesland Mecklenburg-Vorpommern im Nordosten Deutschlands. Abbildung 5 bietet eine Übersicht zur Lage der betrachteten Gebiete. Die gewählten Regionen sollen drei möglichst unterschiedliche landschaftliche als auch hydrologische Gegebenheiten bieten. Stralsund stellt einen eher städtisch geprägten Siedlungsraum an der Ostseeküste dar, wohingegen Bützow einen kleinstädtischen, ländlichen Raum im Binnenland repräsentiert. Das dritte Untersuchungsgebiet in der Feldberger Seenlandschaft stellt in Kontrast dazu einen ausgeprägten Naturraum dar. Diese diversifizierte Aufstellung ermöglicht eine möglichst breite Untersuchung des entwickelten Workflows hinsichtlich der Robustheit und Übertragbarkeit auf weitere Untersuchungsgebiete.



Abbildung 5: Übersichtsdarstellung der drei Untersuchungsgebiete (eigene Darstellung)

#### 3.1.1 Stralsund

Stralsund liegt im Landesteil Vorpommern, hat ca. 54000 Einwohner (Statistisches Amt Mecklenburg-Vorpommern, 2025) und ist die Kreisstadt des Landkreises Vorpommern - Rügen. Mit der Rügenbrücke verbindet Stralsund das Festland mit der Insel Rügen. Das Untersuchungsgebiet rund um die Hansestadt erstreckt sich über eine Ausdehnung von 6 km \* 6 km, also einer Fläche von 36 km<sup>2</sup>. Im Westen beginnt

das Untersuchungsgebiet ab Höhe Grünhufe und im Osten ist sie begrenzt durch die Insel Dänholm im Strelasund (Meeresarm der Ostsee) zwischen Rügen und Stralsund. Der Stadtteil Knieper stellt die nördliche und Groß Lüdershagen die südliche Begrenzung dar.

Die Hansestadt ist hydrologisch geprägt durch ihre Angrenzung an den Strelasund im Osten und der damit gegebenen Anbindung an die Ostsee. Die Binnenseite besitzt mit dem (Großen) Franken-, Knieper- und Moorteich jedoch auch große Seeflächen. Das Untersuchungsgebiet ist weitgehend flach und weist nur geringe Höhenunterschiede auf. Die Besonderheit hier ist, dass eine Mischung aus Meer- und Binnengewässern vorliegt, was eine Identifikation aufgrund von unterschiedlicher Trübungen und gefolgert verschiedener Reflexionswerte erschweren kann.

Landschaftlich ist das Areal grob dreigeteilt. Im Westen liegen z.T. größere landwirtschaftlich geprägte Flächen vor, wohingegen der Osten durch den Strelasund eingenommen wird. Zentral liegt das Stadt- und Hafengebiet Stralsunds. Es ist geprägt durch einen recht hohen Versiegelungsgrad und eine kompakte Stadtstruktur. Aufgrund des Hafens und der Rügenbrücke liegt eine eher technische Infrastruktur vor. Dies kann eine Klassifikation weiter erschweren, da der Schattenwurf zu Fehlklassifikationen führen kann.

### **3.1.2 Bützow**

Bützow liegt im Landesteil Mecklenburg, zentral zwischen der Landeshauptstadt Schwerin und der Großstadt Rostock. Mit rund 7700 Einwohnern zählt Bützow zu den 39 Kleinstädten in Mecklenburg - Vorpommern (Statistisches Amt Mecklenburg-Vorpommern, 2025). Die Stadt ist umgeben von kleineren Wäldern und Ackerflächen. Das Untersuchungsgebiet erstreckt sich von der Gemeinde Rühn im Südwesten Bützows bis hin zum Oetteliner See im Nordosten der Stadt.

Das Untersuchungsgebiet rund um Bützow erstreckt sich ebenfalls auf einer Fläche von 36 km<sup>2</sup>. Aufgrund der Lage der Kleinstadt im Warnowtal weist die Topographie ein leicht welliges Relief auf. Im Nordwesten werden Höhen von bis zu 50 m ü. NHN (Normalhöhennull) erreicht.

Die Landnutzung im Untersuchungsgebiet ist stark geprägt durch die Landwirtschaft. Insbesondere im Nordwesten der Kleinstadt befinden sich große Ackerflächen. Im Südosten befindet sich ein Gewerbegebiet mit Solarparks, umgeben von kleineren Waldgebieten. Das Untersuchungsgebiet weist abseits der kleineren Siedlungsgebiete viel Vegetation auf, was wiederum mit Schwierigkeiten bei der Gewässerklassifika-

tion einhergeht, da dieser Uferbewuchs regelmäßig Schatten und Überdeckungen erzeugt.

Bützow ist hydrologisch durch das Warnowtal geprägt. Der Fluss Warnow erstreckt sich vom Südwesten, trifft bei Bützow auf die Nebel und verläuft weiter Richtung Nordosten. Ein Seitenarm der Warnow bildet einen Zufluss in den Bützower See, aus dem die Temse als Abfluss zurück zur Warnow dient. Neben den vielen Flüssen, Kanälen und Gräben gibt es jedoch auch einige Standgewässer. Dazu gehört der Große Rühner See oder der Lange See im Südwesten. Besonders interessant ist dieses Untersuchungsgebiet im Hinblick auf die dreidimensionalen Gewässerumringe, da hier die Höheninformationen entlang des Flussgefälles eine wichtige Rolle spielen. Zudem kann hier die Trennung zwischen Stand- und Fließgewässern anschaulich gemacht werden.

### **3.1.3 Feldberger Seenlandschaft**

Das dritte und letzte Untersuchungsgebiet befindet sich im südöstlichen Teil Mecklenburg - Vorpommerns und beginnt ca. 6 km südöstlich von Neustrelitz. Es erstreckt sich ebenfalls auf einer Strecke von 6 km \* 6 km mit einer Fläche von 36 km<sup>2</sup>. Der Großteil der Fläche reicht in den östlichen Teil des Müritz Nationalparks hinein (Müritz-Nationalpark, 2025). Die Feldberger Seenlandschaft ist geprägt durch eine eiszeitliche Hügellandschaft mit vielen, teils tiefen Seen (Landesamt für Umwelt, Naturschutz und Geologie Mecklenburg-Vorpommern, 2025).

Landschaftlich ist das Untersuchungsgebiet ganz klar durch Wälder geprägt. Im Südwesten und Südosten gibt es mit Fürstensee und Grammertin vereinzelt kleine Siedlungen. Im Südosten sind zudem noch Ackerflächen vorhanden. Das starke Vorkommen von Waldflächen, insbesondere in Ufernähe, kann die Gewässerklassifikation in diesem Gebiet durch zusätzliche Schattenbildung deutlich erschweren.

Aufgrund der eiszeitlichen Moränenbildung weist die Feldberger Seenlandschaft ein stärker gegliedertes Relief auf. Im Norden des Untersuchungsgebietes liegen die Serrahner Berge mit einer Höhe von über 120 m ü. NHN. Die Seen, darunter der Große Fürstenseer See, der Lutowsee und der Schweingartensee sind zudem gekennzeichnet durch verhältnismäßig steile Ufer.

Hydrologisch dominieren im Untersuchungsgebiet, aufgrund der eiszeitlichen Entwicklung, die Seen. Diese sind durch ihre verschiedenen Tiefen, Trübungen und die sehr reich gegliederten Uferbereiche ein guter Indikator dafür, wie zuverlässig die Gewässerklassifikation Uferlinien trotz starkem Bewuchs erkennen kann.

## 3.2 Eingesetzte Datensätze

Innerhalb des Workflows werden eine Vielzahl von unterschiedlichen Datensätzen genutzt. Dabei stellen die Daten eine zentrale Rolle dar, da diese die Nutzereingabe definieren und somit den einzigen Schnittpunkt zwischen Nutzer und Workflow bieten. Die Daten werden innerhalb des Workflows in mehreren verketteten Schritten verarbeitet, gefolgt von dem finalen Produkt der Gewässerpolygone mit Höheninformationen. Die verschiedenen Datensätze und ihr Nutzen in der Verarbeitungskette werden im Folgenden genauer erläutert.

### 3.2.1 Digitales Orthophoto

Der wichtigste Datensatz, der maßgeblich für die Qualität der Gewässerpolygone ist, ist das DOP. Bereitsteller dieser Daten ist die Abteilung 3 - Amt für Geoinformation, Vermessungs- und Katasterwesen (AfGVK) im Landesamt für innere Verwaltung (LAIv) Mecklenburg - Vorpommerns (Landesamt für innere Verwaltung Mecklenburg-Vorpommern, o. D.[f]).

„Digitale Orthophotos (DOP) sind hochauflösende, verzerrungsfreie Luftbilder, die einen einheitlichen Bildmaßstab und einen exakten Bezug zum Landeskoordinatensystem besitzen.“ (Landesamt für innere Verwaltung Mecklenburg-Vorpommern, o. D.[f]). Sie zeigen den Ist-Zustand einer Landschaft zu einem bestimmten Zeitpunkt, „ohne dass diese bereits selektiert oder strukturiert worden sind.“ (Landesamt für innere Verwaltung Mecklenburg-Vorpommern, o. D.[f]). Die DOP liegen im Rasterformat als GeoTIFF vor.

Die im Workflow verwendeten DOP weisen eine Bodenauflösung von 0,1 m pro Pixel auf. Sie erfassen die Erdoberfläche in vier spektralen Bändern: Rot, Grün, Blau und Nahes Infrarot (RGBI). Zudem weisen die verwendeten DOP die Qualitätsstufe True Orthophotos (TrueDOP) auf, welche sich dadurch auszeichnen, „[...] dass die sichttoten Räume weitestgehend beseitigt sind und keine Umklappeffekte (Verkipnungen) von Objekten über Geländehöhen verbleiben.“ (Landesamt für innere Verwaltung Mecklenburg-Vorpommern, o. D.[f]).

Im Workflow dienen die Digitalen Orthophotos als Input für die überwachte Klassifizierung (siehe Kapitel 4.2.2). Die DOP sind integraler Bestandteil der Gewässerklassifikation und bieten beispielsweise gegenüber Sentinel-2 Daten den Vorteil der sehr hohen räumlichen Auflösung, was die Klassifikation kleinerer Flüsse und Seen ermöglicht.

### 3.2.2 Digitales Oberflächenmodell

„Digitale Oberflächenmodelle (DOM) beschreiben die Erdoberfläche inklusive aller festen Objekte.“ (Landesamt für innere Verwaltung Mecklenburg-Vorpommern, o. D.[d]). Anders als beim Digitalen Geländemodell (DGM) werden hierbei auch die Höhen von beispielsweise Wäldern und Bauwerken, wie Gebäuden und Brücken, erfasst. Das DOM liegt mit einer Gitterweite von 1 m vor und wird aus dem bDOM (Landesamt für innere Verwaltung Mecklenburg-Vorpommern, o. D.[g]) erstellt. Die DOM liegen im Rasterformat, als GeoTIFF vor. Wie bereits bei den DOP ist das AfGVK M-V der Bereitsteller der Daten.

Das Digitale Oberflächenmodell dient im Workflow zur Fehlerminimierung der Gewässerklassifikation durch das DOP. Das DOM wird zur Berechnung der Verschattung zum Zeitpunkt der DOP-Aufnahme benötigt. Das heißt, dass basierend auf dem Sonnenstand zum Aufnahmezeitpunkt der DOP mit Hilfe des DOM berechnet wird, welche Bereiche im Schatten liegen (siehe Kapitel 4.2.3). Hierdurch kann sichergestellt werden, dass die Wasserpolygone tatsächlich Wasser darstellen und nicht falsch identifizierte Polygone in Schattenbereichen von beispielsweise Gebäuden.

### 3.2.3 Airborne Laserscanning

Die Airborne-Laserscanning Daten spielen die zentrale Rolle in der Ableitung der Höheninformationen für die Gewässerpolygone (siehe Kapitel 4.3). Die Daten werden in einzelnen Kacheln organisiert und vom AfGVK im LAS bzw. LAZ-Format (komprimiertes LAS) bereitgestellt.

Airborne-Laserscanning (ALS) wird aktuell als gängige Methode beschrieben, um großflächig Höhendaten der Erdoberfläche zu erfassen. „Die Höheninformation wird hierbei aus der Position des Flugzeuges mit Hilfe des GNSS, der Fluglage des Flugzeuges über INS [Inertial Navigation System] und der Laufzeitlänge der Laserstrahlen vom Laserscanner ermittelt.“ (Landesamt für innere Verwaltung Mecklenburg-Vorpommern, o. D.[b]). Das Ergebnis liegt im Anschluss in Form einer Punktwolke vor, welche in Bodenpunkte und Nichtbodenpunkte klassifiziert ist. Die Punktdichte beläuft sich aktuell auf 5 Punkte/m<sup>2</sup> (Landesamt für innere Verwaltung Mecklenburg-Vorpommern, 2025).

Im Workflow werden die ALS-Daten genutzt, um dem See- als auch dem Flusslayer Höheninformationen zu übergeben. Die genaue Vorhergehensweise dafür wird in Kapitel 4.3 genauer betrachtet. Grundlegend wird über die bereits klassifizierten

Bodenpunkte ein DGM gerechnet, aus welchem sich dann, mittels simpler Statistik, eine feste Höhe auf Seen bzw. variable Höhen auf Flüsse übertragen lassen können.

### **3.2.4 Digitale Landschaftsmodelle**

Das Digitale Landschaftsmodell (DLM) spielt in der Verarbeitungskette des Workflows nur eine untergeordnete Rolle, erfüllt aber eine wichtige Funktion. Als DLM „[...] wird ein objektstrukturierter, topographischer Vektordatenstand bezeichnet.“, welcher „[...] Landschaftsobjekte wie Verkehrswege oder Vegetationsflächen nach Lage und Form in einem Landeskoordinatensystem sowie in ihren Eigenschaften durch Attribute [...]“ beschreibt (Landesamt für innere Verwaltung Mecklenburg-Vorpommern, o. D.[c]). Das DLM wird vom AfGVK M-V bereitgestellt.

Der spezifische Einsatzzweck des DLM wird in Kapitel 4.2.3 genauer erläutert. Vereinfacht dargestellt, wird das DLM mit den zuvor berechneten Schattenflächen verschnitten. Das hat den Hintergrund, dass sich Schatten zum Teil auch auf Gewässern befinden (beispielsweise aufgrund von Kreuzfahrtschiffen), welche wiederum in der weiteren Verarbeitung zu fehlenden Gewässerflächen führen. Das DLM wird somit zur Entfernung von detektierten Schattenflächen auf Gewässern genutzt.

### **3.2.5 LUNG-Gewässer**

Das Landesamt für Umwelt, Naturschutz und Geologie (LUNG) stellt über das Kartenportal Umwelt Mecklenburg - Vorpommern diverse Daten, darunter Daten aus dem Wasserwirtschaftlichen Gewässerkataster kostenfrei zur Verfügung (Landesamt für Umwelt, Naturschutz und Geologie Mecklenburg-Vorpommern, o. D.). Dazu gehören verschiedene Vektordateien wie Fließgewässer (Linie), Standgewässer (Polygon) und Küstengewässer (Polygon).

Zentrale Aufgabe mit Hilfe dieser Daten ist es, das zweidimensionale Klassifikationsergebnis für die unterschiedlichen Vorgehensweisen der Höhenzuweisung für Seen und Flüsse zu trennen. Das Ergebnis dieser Operation sollen also zwei getrennte Layer mit den Inhalten Standgewässer (Seen, Meere) und Fließgewässer (Flüsse) sein. Die genauen Verarbeitungsschritte können in Kapitel 4.2.5 nachvollzogen werden.

## 4 Methodik

### 4.1 Übersicht der Verarbeitungskette

Der in der vorliegenden Arbeit erstellte Arbeitsablauf beinhaltet eine vollständig automatisierte Ableitung von dreidimensionalen Gewässerumringen dar. Der Workflow lässt sich in zwei große Hauptprozesse einteilen. Ein Prozess stellt die grundlegende 2D-Erfassung dar. Hierbei werden die Gewässerflächen in ihrer Ausdehnung erfasst. Der zweite Hauptprozess behandelt die 3D-Erfassung. Den zweidimensional erfassten Gewässerflächen werden hierbei Höheninformationen zugewiesen. Als Eingabedaten dienen überwiegend Fernerkundungsdaten, das Ergebnis des Workflows sind Gewässerpolygone mit Höhenwerten.

In Abbildung 6 ist der Arbeitsablauf in vereinfachter Form schematisch dargestellt. Für die 2D-Erfassung wird zunächst aus DOP-Daten und Trainingsgebieten eine überwachte Klassifikation durchgeführt. Parallel findet aus DOM-Daten eine Schattenberechnung statt. Anschließend wird die Differenz aus den klassifizierten Gewässern und den Schattenflächen berechnet, um Fehlklassifikationen zu minimieren. Es folgen weitere Prozessierungsschritte, wie die Entfernung von Löchern in den Gewässerpolygonen oder die Filterung nach Polygongrößen. Mit Hilfe von Daten des LUNG wird das Klassifikationsergebnis in Fließ- und Standgewässer unterteilt. Für die 3D-Erfassung wird aus ALS-Daten ein DGM erstellt. Bei dem Seelayer wird die Seehöhe anhand des erstellten DGM auf einen festen Wert gesetzt. Für Fließgewässer wiederum wird eine variable Höhe entlang des Flussverlaufes zugewiesen. Das Ergebnis sind zwei dreidimensionale Polygonlayer.

Die Erfassung von herkömmlichen zweidimensionalen Gewässerumringen ist innerhalb des Arbeitsablaufs die Voraussetzung für die Ermittlung von Höheninformationen. Zudem wird angenommen, dass Standgewässer wie Seen eine annähernd konstante Wasseroberfläche vorweisen. Für Fließgewässer wird angenommen, dass sie eine Höhenänderung entlang der Fließrichtung aufweisen.

Der vollständig automatisierte Workflow wurde in QGIS mit Hilfe des Model Builders umgesetzt. Abbildung 23 in Anhang A.1 gibt einen Überblick über den exakten Arbeitsablauf innerhalb des Model Builders. Durch die Nutzung der kostenfreien Open-Source Software (QGIS Development Team, o.D.) im Zusammenhang mit dem Modell kann eine Reproduzierbarkeit des Ergebnisses sichergestellt werden. Zudem ermöglicht die Nutzung eine konsistente Anwendung auf unterschiedliche Untersuchungsgebiete.

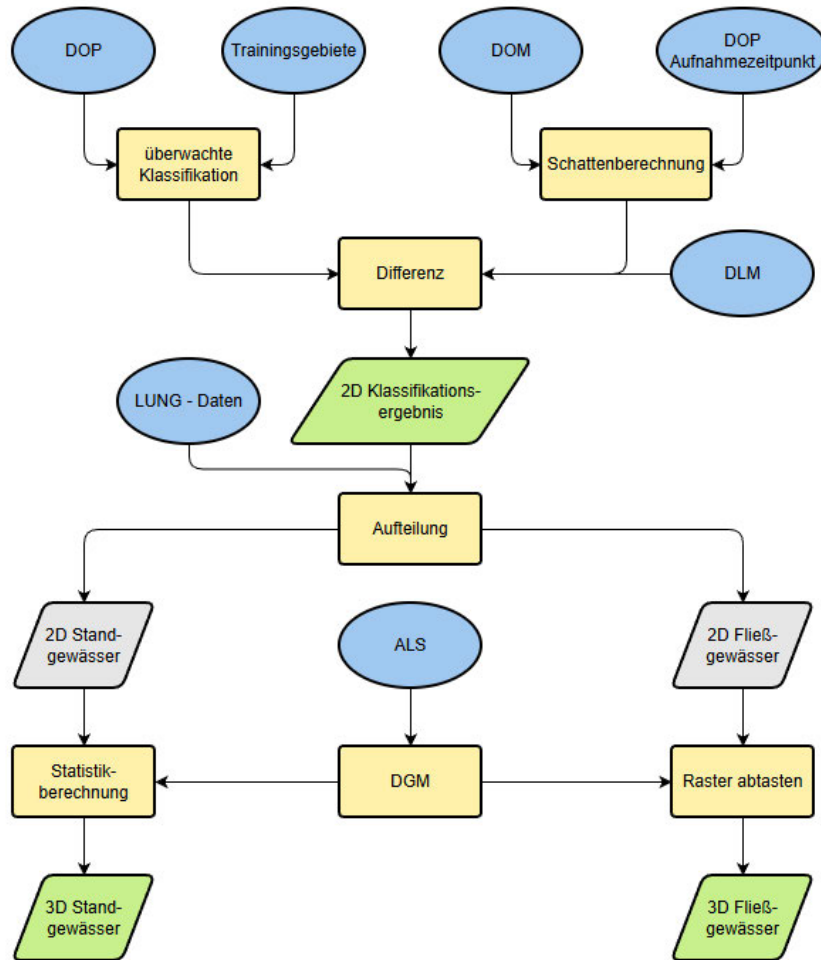


Abbildung 6: Vereinfachtes Flussdiagramm des automatisierten QGIS Workflows (eigene Darstellung)

## 4.2 Ableitung der Gewässerumringe (2D-Erfassung)

Die Ableitung der Gewässerumringe bildet den ersten Teilstrang des automatisierten Arbeitsablaufes. Das Ziel ist die Erfassung der zweidimensionalen Ausdehnung von Gewässern, welche die Grundlage für die 3D-Erfassung darstellt. Im Folgenden werden die einzelnen Prozessierungsschritte zur zweidimensionalen Gewässererkennung beschrieben, einschließlich der verwendeten Eingangsdaten und der Behandlung typischer Problemstellungen.

### 4.2.1 Vorverarbeitung

Zur Vorverarbeitung zählt unter anderem die Herstellung einer konsistenten Datenbasis. Zu den Eingaben gehören zunächst zwei Ordner. Der Erste enthält alle DOP-Tiles des Untersuchungsgebietes. Das DOP besteht aus vier spektralen Bändern (Rot, Grün, Blau, nahes Infrarot). Die Kacheln innerhalb des Ordners werden im Anschluss zu einem sich über das gesamte Untersuchungsgebiet erstreckendes DOP verschmolzen. Der zweite Ordner stellt die DOM-Tiles bereit. Zusammen mit der Eingabe des Aufnahmezeitpunktes des DOP, bildet es den Input für die in Kapitel 4.2.3 näher betrachtete Schattenberechnung zur Fehlerminimierung.

Es ist wichtig zu verstehen, dass die Gesamtausdehnung der bereitgestellten DOM-Tiles identisch der der DOP-Tiles sein muss, da diese Ausdehnung zeitgleich den Umfang des Untersuchungsgebietes darstellt. Andernfalls ist keine korrekte Klassifikation beziehungsweise Schattenberechnung gewährleistet. Des Weiteren sollten die Angaben bezüglich des Aufnahmezeitpunktes des DOP möglichst kleine Abweichungen zum Tatsächlichen aufweisen. Der Zeitstempel dient der Berechnung des Sonnenstandes zum Zeitpunkt der DOP-Aufnahme. Eine fehlerhafte Angabe führt zu einer falschen Berechnung der Schatten, welches wiederum zu Fehlklassifikationen von Wasserflächen in verschatteten Bereichen führt.

Die beschriebenen verwendeten Eingangsdaten sollten optimalerweise eine sehr hohe räumliche Auflösung (vgl. Kapitel 2.1.3) vorweisen. Für das DOP wurden Daten mit einer Auflösung von 0,1 m genutzt. Das DOM liegt mit einer räumlichen Auflösung von 1 m vor. Die gröbere Auflösung des DOM ist ausreichend für die Schattenberechnung, da diese nur für die Minimierung von Fehlklassifizierungen dient. Alle Daten innerhalb des Workflows werden zudem in das Koordinatenreferenzsystem ETRS89 UTM Zone 33N mit dem EPSG-Code 25833 projiziert, um eine einheitliche räumliche Verarbeitung sicherzustellen.

### 4.2.2 Klassifikation der Gewässerflächen

Die Klassifikation der Gewässerflächen stellt den essenziellen Verarbeitungsschritt des automatisierten Workflows dar. Er bildet die Grundlage für die korrekte Ausführung aller darauffolgender Algorithmen. Die Klassifikation lässt sich in zwei Phasen, die Trainings- und in die Vorhersagephase, einteilen.

Bevor jedoch die unterschiedlichen Klassifikationsphasen starten, müssen die im Eingabeordner befindlichen DOP-Tiles verschmolzen werden. Insgesamt bilden neun

Kacheln mit einer Ausdehnung von jeweils 2 km \* 2 km das Untersuchungsgebiet mit einer Fläche von 36 km<sup>2</sup> ab.

Die Trainingsphase greift zunächst auf die Trainingsgebiete zurück. Die Trainingslabel sind manuell digitalisierte Polygone, die vor der eigentlichen Ausführung des automatisierten Workflows für das jeweilige Untersuchungsgebiet erstellt werden müssen. Für die Trainingsgebiete werden zwei Klassen definiert. Polygone der Klasse = 1 stellen Wasserflächen dar, wogegen mit Klasse = 2 definierte Polygone die unterschiedlichen Landbedeckungsformen repräsentieren. Wichtig für die Trainingsphase ist hierbei insbesondere die Benennung des Attributs für die Klassenwerte. Erhält die Spalte, die die Klassenwerte enthält, beispielsweise den Namen „Klasse“, so muss dieser Name mit als Eingabe in den Workflow eingehen. Dies ist erforderlich, damit eine korrekte Klassifikation sichergestellt werden kann. Für die Auswahl beziehungsweise Erstellung der Trainingspolygone sind verschiedene Bedingungen zu beachten. Hierbei sollten die Sample räumlich über das gesamte Untersuchungsgebiet verteilt sein. Zudem sollte ein prozentual ausgewogenes Verhältnis zwischen Wasser- und Landpolygonen vorliegen. Die Anzahl der Trainingsflächen hängt zum einen von der Größe des Untersuchungsgebietes, zum anderen aber auch von der Komplexität und Vielfältigkeit der Landbedeckung ab. Abbildung 7 bietet eine Übersicht über die erstellten Trainingsgebiete für das Untersuchungsgebiet in der Feldberger Seenlandschaft.

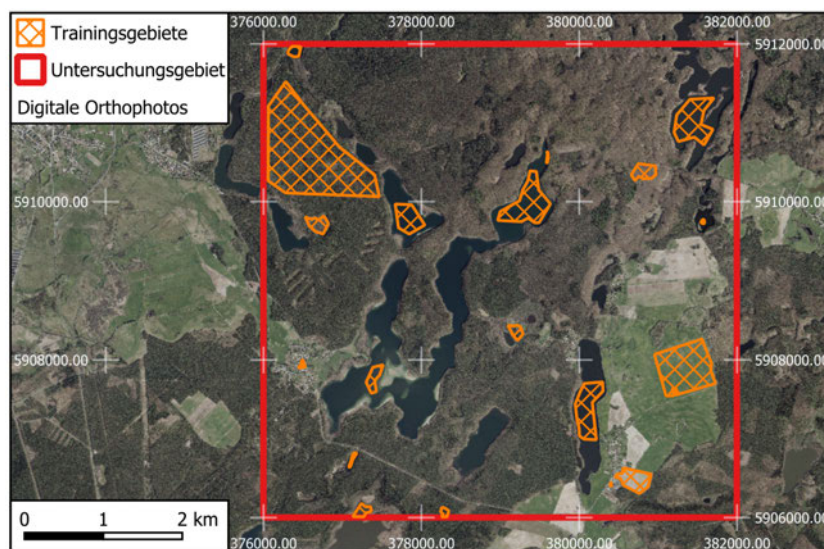


Abbildung 7: Übersicht über Trainingsgebiete im Untersuchungsgebiet Feldberger Seenlandschaft (eigene Darstellung)

Innerhalb des Workflows wird das QGIS Plugin „dzetsaka : Classification tool“ (Karsiak, 2016) zur überwachten Klassifikation genutzt. Das Plugin unterstützt insge-

samt 11 Klassifizierungsalgorithmen (Karasiak, 2016). Im Workflow wird der Algorithmus der Logistic Regression (Logistischen Regression) verwendet. Als Eingabe in das Training dieses Classifiers dienen die soeben beschriebenen Trainingsgebiete, der Name des Klassifikationsattributs sowie das verschmolzene DOP (vgl. Kapitel 4.2.1). Zusätzliche Anpassungen von Parametern abweichend der Standardeinstellungen des Plugins wurden nicht vorgenommen. Von den in Kapitel 2.2.1 und 2.2.3 dargestellten Ansätzen auf Basis von Indizes beziehungsweise Künstlicher Intelligenz wurde im Rahmen der Arbeit Abstand genommen. Ausschlaggebend dafür waren einerseits die qualitativ nicht ausreichenden Ergebnisse der indexbasierten Verfahren innerhalb einer Vorbetrachtung, zum anderen der mit Deep Learning Ansätzen verbundene hohe Entwicklungs- und Implementierungsaufwand, der den zeitlichen Rahmen der Arbeit überschritten hätte. Bei dem weiteren Vergleich zwischen der unüberwachten und der überwachten Klassifikation fällt auf, dass die Überwachte eine höhere geometrische Qualität aufweist (siehe Abbildung 8). Die unüberwachte Klassifikation hat starke Randeffekte, viele Fehlklassifikationen auf Vegetation und Schatten und besitzt zudem eine hohe Pixelstreuung im Uferbereich. Die überwachte Klassifikation zeigt eine eindeutigere Trennung zwischen Wasser und Land und eignet sich dadurch besser für den automatisierten Workflow.

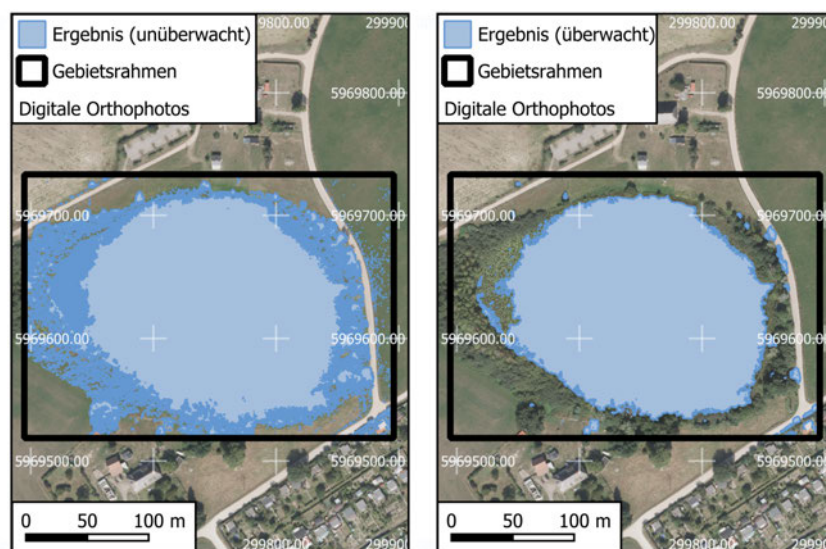


Abbildung 8: Beispielhafter Vergleich einer unüberwachten (links) zu einer überwachten Klassifikation (rechts) (eigene Darstellung)

Die Verwendung einer logistischen Regression als Classifier ist insbesondere dann sinnvoll, wenn binäre Variablen wie 0 und 1, oder „Nicht-Wasser“ und „Wasser“ vorhergesagt werden müssen. Für die Prognose wird die Wahrscheinlichkeit für das Eintreten der Ausprägung 1, also das Vorhandensein des Merkmals geschätzt (numi-

qo Team, o. D.). „Das Besondere an der logistischen Funktion ist, dass sie für Werte zwischen minus und plus unendlich immer nur Werte zwischen 0 und 1 annimmt.“ (numiqo Team, o. D.). Abbildung 9 bietet eine direkte Gegenüberstellung zwischen der klassischen linearen Regression und der in der Arbeit verwendeten logistischen Regression. Da die lineare Regression unbegrenzte kontinuierliche Werte zwischen  $+\infty$  und  $-\infty$  vorhersagt, wird diese mit Hilfe einer Sigmoidfunktion auf den Wertebereich zwischen 0 und 1 eingeschränkt (numiqo Team, o. D.). Dafür wird die in Gleichung 3 gezeigte logistische Funktion verwendet.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

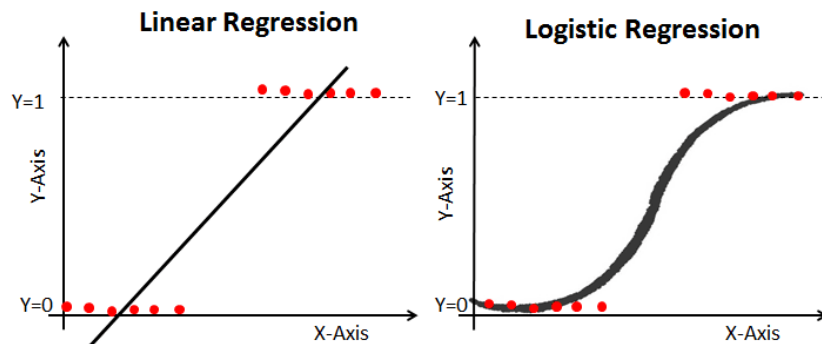


Abbildung 9: Vergleich der linearen und logistischen Regression (DataCamp, o. D.[a])

Die Vorhersagephase erhält als Input das soeben mit Hilfe der Trainingsgebiete vortrainierte Modell sowie das verschmolzene DOP des Untersuchungsgebietes. Die Klassifizierung findet über die gesamte Ausdehnung des zu untersuchenden Gebietes statt. Das Ergebnis der Klassifikation liegt im Rasterformat als GeoTIFF vor. Die räumliche Auflösung von 0,1 m ist abhängig und dadurch identisch mit der des DOP. Der Wertebereich begrenzt sich auf die zwei durch die Trainingsdaten festgelegten Klassen 1 (= Wasser) und 2 (= Land). Dabei gibt es keine Nodata-Werte, da jedes Pixel einer Klasse zugeordnet wird. Im Anschluss an die Durchführung der Klassifikation werden über das Protokoll Genauigkeitsmetriken ausgegeben. Dazu gehören Angaben über die Overall accuracy (Treffergenauigkeit) sowie Werte zu Kappa und F1.

#### 4.2.3 Eliminierung von Schattenfehlklassifikationen

Das Ergebnis der überwachten Klassifizierung weist falsch identifizierte Gewässerflächen auf. Dazu gehören insbesondere Fehler in verschatteten Bereichen. Das liegt

daran, dass Schatten- und Wasserflächen wenig Licht reflektieren und somit zum Teil ähnliche spektrale Signaturen vorweisen. Besonders problematisch sind dabei Baum- und Gebäudeschatten. Die Fehlklassifizierungen haben entsprechende Auswirkungen auf das Ergebnis in Form einer Überschätzung der Gewässerfläche. Die Schattenberechnung verfolgt das Ziel der Eliminierung von diesen Fehlklassifikationen.

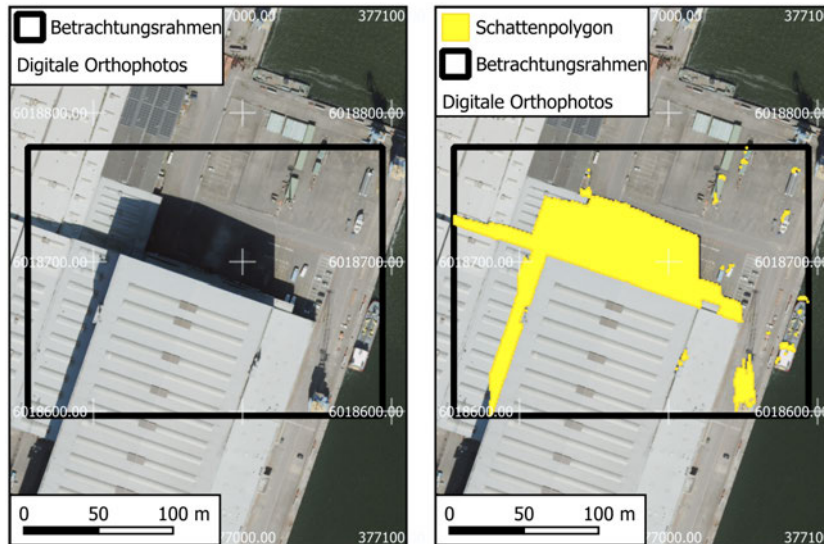


Abbildung 10: Beispielhafte Schattenberechnung (eigene Darstellung)

Die Schattenberechnung wird auf Grundlage des „r.sunmask.datetime“-Algorithmus aus der GRASS GIS Bibliothek durchgeführt (GRASS Development Team, o. D.). Als Eingabedaten dient ein Höhenraster in Form des DOM, welches die Höheninformationen aller Objekte wie Gebäuden und Bäumen enthält, sowie die in Kapitel 4.2.1 erwähnten Daten zum Aufnahmezeitpunkt des DOP. Anhand des Zeitstempels kann der Algorithmus den exakten Sonnenstand mit Hilfe des Zenit- und Azimutwinkels zum Zeitpunkt der DOP-Aufnahme rekonstruieren. Eine niedrig stehende Sonne (kleiner Zenitwinkel) führt dabei zu langen Schatten und großer Zenitwinkel ( $\leq 90^\circ$ ) zu kurzen Schatten. Eine exakte Bestimmung des Zeitpunktes der Befliegung ist essenziell für die Korrektheit der berechneten Schatten. Andernfalls ist eine erfolgreiche Reduktion von Fehlklassifizierungen nicht gegeben. Abbildung 10 zeigt beispielhaft das Ergebnis einer solchen Schattenberechnung. Durch die erstellte Maske können Fehlklassifikationen in Schattenbereichen ausgeschlossen werden.

Innerhalb des automatisierten Workflows wurde der Algorithmus in ein eigen erstelltes Python-Verarbeitungsskript eingebettet. Das ist darauf zurückzuführen, da „r.sunmask.datetime“ als Eingabe eine einzige Rasterdatei erwartet. Insgesamt be-

steht jedes Untersuchungsgebiet jedoch aus 36 einzelnen DOM-Kacheln. Hinsichtlich der Automatisierung ist es in QGIS zum aktuellen Zeitpunkt und Version (3.40) nicht möglich, alle Dateien innerhalb eines Ordners zu verschmelzen. Somit arbeitet das erstellte Verarbeitungsskript iterativ alle Kacheln innerhalb des Eingabeordners ab. Erst im Anschluss werden die berechneten Schatten verschmolzen. Zu Beginn wird geprüft, ob der Eingabeordner existiert und ob in diesem Dateien im GeoTIFF-Format vorliegen. Ist das gegeben, so wird die erste DOM-Kachel bearbeitet. Zunächst wird das Koordinatenreferenzsystem auf EPSG: 25833 festgesetzt. Anschließend werden die im „r.sunmask.datetime“-Algorithmus geforderten Mittelpunktkoordinaten der Kachel berechnet. Wurden diese Schritte erfolgreich abgeschlossen, erfolgt die eigentliche Schattenberechnung mit Hilfe des erwähnten Algorithmus. Nach der Bearbeitung aller 36 im Eingabeordner befindlichen Kacheln erfolgt die Verschmelzung zu einer GeoTIFF, die der Ausdehnung des Untersuchungsgebietes entspricht. Die Konzeption des Python Skripts erfolgte manuell durch den Verfasser dieser Arbeit. Die tatsächliche programmtechnische Umsetzung geschah mit Hilfe des KI-Sprachmodells ChatGPT (OpenAI) (OpenAI, 2025). Der auf Grundlage des Konzepts generierte Code wurde im Anschluss durch den Verfasser geprüft und in den Workflow implementiert. Der Code 1 in Anhang B zeigt den zentralen Verarbeitungsteil des Skripts.

Die erstellte Schattenmaske wird im nächsten Schritt in das Vektorformat umgewandelt, sodass als Ergebnis die berechneten Schattenflächen in Form von Polygonen vorliegen. Diese sogenannte Vektorisierung ist für den nächsten Verarbeitungsschritt notwendig. Hierbei werden die Schattenpolygone mit einem speziellen Datenteil des Basis DLM verschnitten. Das Ziel dieser Verarbeitung ist die Entfernung von detektierten Schatten auf Wasserflächen. Diese treten insbesondere in Hafenanlagen durch den Schattenwurf von Schiffen und Booten auf. Ein markantes Beispiel dafür sind die Kreuzfahrtschiffe in Warnemünde. Die Entfernung ist daher so wichtig, weil im folgenden Prozessierungsschritt die Differenz aus den in Kapitel 4.2.2 klassifizierten Gewässerflächen und den Schattenflächen gebildet wird, um Fehlklassifizierungen auf Landflächen zu löschen. Sobald sich jedoch Schatten mit korrekt identifizierten Gewässern schneiden, werden auch diese gelöscht. Dieser Umstand erfordert den Einsatz des speziellen Datenteils des DLM.

Das DLM (vgl. Kapitel 3.2.4) des AfGVK setzt sich aus insgesamt 42 einzelnen Shapefiles zusammen. Jede Ebene bildet dabei andere Inhalte ab. Es wird zusätzlich zwischen Punkt-, Linien- und Polygoneometrien differenziert. Der für den Verarbeitungsschritt erforderliche Datenstand setzt sich aus neun einzelnen Polygonlayern (veg01\_f, veg02\_f, veg03\_f, ver01\_f, ver03\_f, ver04\_f, ver05\_f, sie03\_f, sie02\_f)

zusammen, die die gesamte Landesfläche abbilden. Gewässer werden im Hinblick auf die Verwendung des Datensatzes nicht abgebildet. Die neun Shapefiles wurden im Zuge der Vorbereitung vereinigt und stehen im Vektorformat als GeoPackage Datei zur Anwendung für jedes Untersuchungsgebiet innerhalb des Bundeslandes Mecklenburg - Vorpommern zur Verfügung.

Das Ergebnis der Eliminierung von Schattenfehlklassifikationen auf Grundlage des „r.sunmask.datetime“-Algorithmus und der erweiterten Korrektur mit Hilfe eines angepassten DLM Datensatzes ist ein Vektorlayer, bestehend aus Gewässerumringen. Dieser weist bedeutend weniger Fehlklassifikationen auf als im Vergleich zuvor. Die Qualität der Ergebnisse wird im Workflow noch weiter optimiert, beschrieben im folgenden Kapitel.

#### 4.2.4 Nachbearbeitung

Durch die vorherige Differenzberechnung zwischen den Gewässerumringen und den Schattenflächen bleiben vereinzelt nicht entfernte Reste von Umringen zurück. Diese Reste gehören häufig, zumindest formal nach der Attributtabelle, noch zu ein und demselben Polygon, obwohl sie räumlich getrennt voneinander auftreten. Für spätere Verarbeitungsschritte ist es wichtig, diese zu trennen. Abhilfe schafft dabei die Funktion „Mehr- zu einteilig“.

Im Anschluss werden die jeweiligen Flächeninhalte der jetzt einzeln vorliegenden Polygone berechnet. Dazu wird im Feldrechner ein neues Feld „flaeche“ angelegt. Der Ausdruck zur Berechnung des Flächeninhaltes lautet „\$area“. Dieser Wert wird unter anderem für eine Selektion im Anschluss an den folgenden Verarbeitungsschritt benötigt. Die Funktion „Löcher löschen“ bewirkt die Schließung beziehungsweise Auffüllung kleinerer Löcher innerhalb der Wassermasken. Im automatisierten Workflow werden Löcher mit einer Fläche kleiner als 5 m<sup>2</sup> gelöscht. Die Auffüllung wesentlich größerer Löcher hätte zur Folge, dass fälschlicherweise auch kleinere Schiffe zur Wasseroberfläche hinzugezählt werden. Durch die Auffüllung der kleineren Löcher wird unter anderem der in Kapitel 2.2.2 angesprochene „Salz-und-Pfeffer-Effekt“ verringert.

Den abschließenden Prozessierungsschritt der Nachbearbeitung stellt eine Selektion nach dem Flächeninhalt der Wassermasken dar. Hierbei wird die Funktion „Nach Attribut extrahieren“ genutzt, um ausschließlich Gewässer zu behalten, die eine Fläche von mindestens 10 m<sup>2</sup> besitzen. Kleinere Polygone werden gelöscht. Durch diesen Verarbeitungsschritt kann sichergestellt werden, dass auch Reste von Fehlklassifikationen aus der Verschneidung mit den Schattenflächen entfernt werden. Die Wahl

eines größeren Flächeninhalts hätte zur Folge, dass eventuell kleinere Teiche entfernt werden. Eine beispielhafte Übersicht über die visuellen Auswirkungen der aufeinanderfolgenden zweidimensionalen Klassifizierungsschritte bietet Abbildung 11.

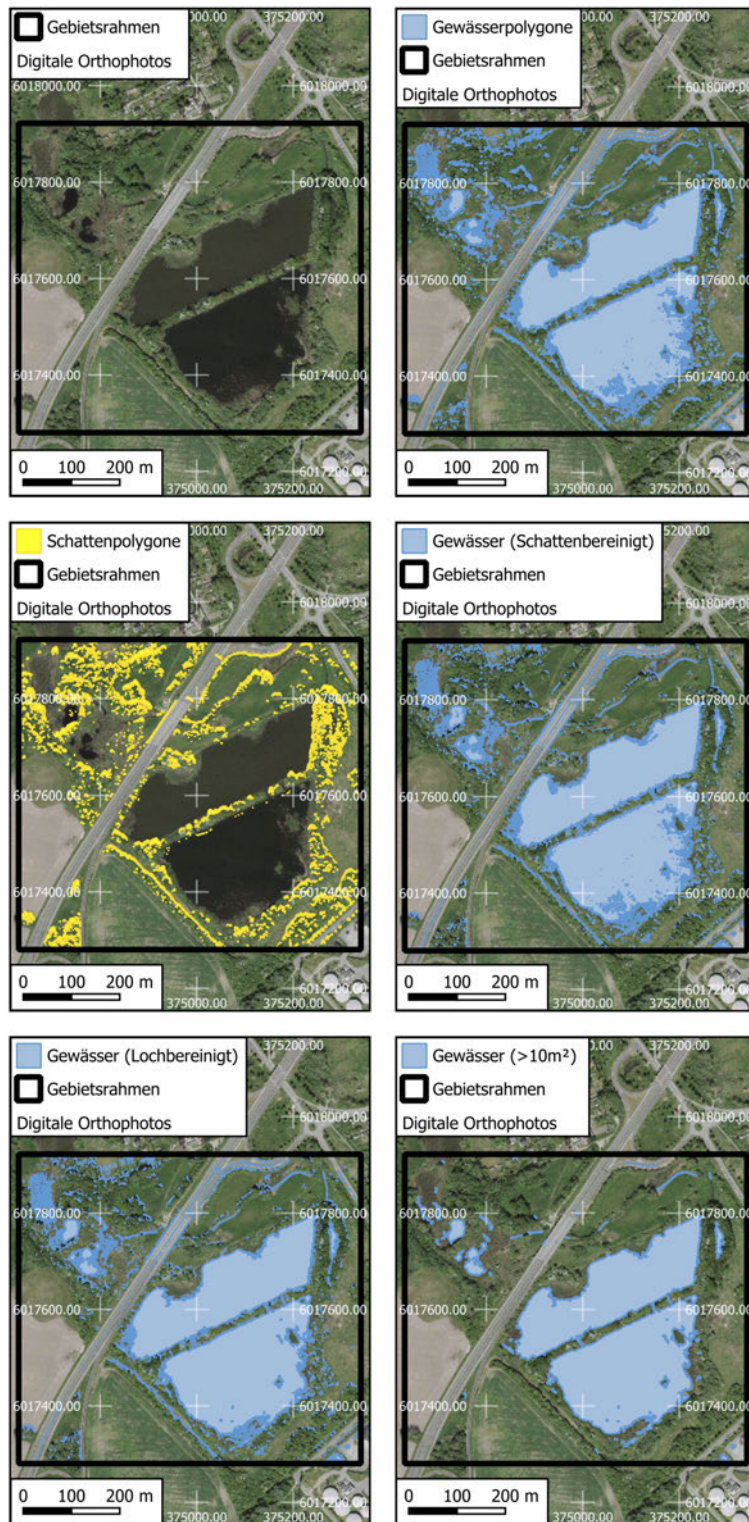


Abbildung 11: Zwischenergebnisse der 2D-Gewässerklassifikation (Leserichtung von oben links nach unten rechts): DOP als Ausgangsdaten, überwachte Klassifikation, Schattenpolygone, schatten- und lochbereinigte Gewässerflächen sowie flächengefilterte Gewässerpolygone (eigene Darstellung)

#### 4.2.5 Trennung von Fließ- und Standgewässern

Das Klassifikationsergebnis bildet aktuell Gewässerumringe jeglicher Art in Form von Polygonen ab. Dabei sind die Gewässer mindestens 10 m<sup>2</sup> groß. Im Hinblick auf die Höhenzuweisung (vgl. Kapitel 4.3) ist eine Trennung notwendig, da an Stand- und Fließgewässer unterschiedliche Anforderungen gestellt werden, die wiederum eine verschiedene Verarbeitung voraussetzen. Für die Aufteilung sind, neben dem Klassifikationsergebnis im Vektorformat, noch drei weitere Eingabedaten notwendig. Das LUNG stellt diese Daten bezüglich Fließ-, Stand- und Küstengewässern bereit (vgl. Kapitel 3.2.5).

Die Extraktion von Fließgewässern beginnt mit der Funktion „Nach Ausdruck extrahieren“. Hierbei werden die Fließgewässer-Referenzdaten des LUNG ausschließlich auf Gewässer erster Ordnung begrenzt. Dafür muss die Bedingung „*Widmung* = 1501 or *Widmung* = 1502“ erfüllt sein. Im Anschluss wird der Linienlayer auf einen Abstand von 10 m gepuffert. Das erfüllt den Zweck, dass bei der folgenden Verschneidung auch die nicht optimal unter der Linie befindlichen Gewässerpolygone einen Schnittpunkt besitzen. Der Verarbeitungsschritt „Nach Position extrahieren“ erhält die Flussmittellinie des LUNG sowie das vektorisierte Klassifikationsergebnis als Input. Hierbei werden alle Polygone ausgewählt, die einen Schnittpunkt mit der gepufferten Mittellinie aufweisen. Die gewählten Polygone stellen somit Fließgewässer dar.

Da es Untersuchungsgebiete geben kann, die keine Fließgewässer enthalten, wird nachfolgend eine „Bedingte Verzweigung“ ausgeführt. Diese prüft, ob es Flüsse innerhalb des Untersuchungsgebietes gibt und leitet je nach Ausgang der Bedingung zu einem unterschiedlichen Prozessierungsschritt weiter. Die Verwendung solcher Verzweigungen ist das Resultat aus der Verwendung unterschiedlicher Untersuchungsgebiete. Während Bützow sowohl Fließ- als auch Standgewässer enthält, stoppte der Workflow unerwartet in Stralsund oder der Feldberger Seenlandschaft. Diese Gebiete enthalten keine Fließgewässer, doch die folgenden Prozessierungsschritte des Workflows erwarteten diese als zwingende Eingabe. Durch die Verwendung bedingter Verzweigungen ist der Arbeitsablauf nun so flexibel, dass er auch in Gebieten ohne Fließgewässer funktioniert.

Die Bedingung „*layer\_property(@Nach\_Position\_extrahieren\_OUTPUT; feature\_count)* = 0“ prüft, ob bei der vorherigen Verschneidung ein leeres Ergebnis existiert und somit keine Flüsse im Untersuchungsgebiet vorliegen. Ist dies der Fall wird davon ausgegangen, dass ausschließlich Standgewässer im Untersuchungsgebiet befindlich sind. In dem Fall wird für die weitere Höhenzuweisung das Ergebnis aus

Kapitel 4.2.4 verwendet, da es das Gesamtklassifikationsergebnis darstellt. Die Erstellung des spezifischen Seelayers ist somit nicht vonnöten und auch technisch nicht umsetzbar, da dafür der spezifische Flusslayer erforderlich ist.

Mit der Erfüllung der Bedingung „*layer\_property(@Nach\_Position\_extrahieren\_OUTPUT,'feature\_count') > 0*“ ist der Umstand gegeben, dass innerhalb des Untersuchungsgebietes Fließgewässer vorhanden sind. Die Bedingung ist Voraussetzung für die Ausführung einer Differenzberechnung zwischen den extrahierten Flusspolygonen und einem zusammengeführten Vektorlayer, bestehend aus Küsten- und Standgewässerdaten des LUNG. Der Seelayer kann nun als Differenz aus dem Gesamtklassifikationsergebnis und dem finalen zweidimensionalen Flusslayer erstellt werden.

Mit den drei Ergebnissen Gesamtklassifikation, Fluss- sowie Seelayer ist die zweidimensionale Verarbeitungskette des automatisierten Workflows abgeschlossen. Diese Daten dienen im gleichen Moment als Input für die im folgenden Kapitel 4.3 erläuterte 3D-Erfassung.

### 4.3 Ableitung der Gewässerhöhen (3D-Erfassung)

Je nach Gewässerart sind unterschiedliche Prozessierungsabfolgen zur Höhenbestimmung nötig. Unabhängig davon sind jedoch Vorbereitungsschritte zu bearbeiten, die universal benötigt werden. Der grundlegende Dateninput zur Bestimmung der Gewässerhöhen ist der Ordner, der die ALS-Daten des Untersuchungsgebietes enthält. Für jedes Untersuchungsgebiet sind insgesamt 36 LAZ-Files mit einer jeweiligen Ausdehnung von 1km \* 1km erforderlich.

#### 4.3.1 Vorverarbeitung

Der erste Bearbeitungsschritt besteht aus der Erstellung eines DGM aus den ALS-Befliegungsdaten. Dafür wird der Algorithmus „*las2dem (folder)*“ der LAStools der deutschen rapidlasso GmbH verwendet (rapidlasso GmbH, o.D.[b]). Das Tool erstellt aus den ALS-Punkten ein temporäres Triangulated Irregular Network (TIN) und rastert dieses anschließend auf ein DGM (rapidlasso GmbH, o.D.[a]). Als Input dient der ALS-Eingabeordner. Weitere Einstellungen des Tools beziehen sich unter anderem auf das Dateiformat des Inputs, welches in Übereinstimmung mit den Eingabedaten auf „\*.laz“ (vgl. Kapitel 3.2.3) festgelegt wurde. Des Weiteren wurde als Filter „*keep\_class 2*“ verwendet. Dadurch werden in die Erstellung des DGM

ausschließlich als Bodenpunkte klassifizierte Messdaten einbezogen. Das bietet den Vorteil, dass Ausreißer durch unklassifizierte bzw. Nichtbodenpunkte entfernt werden. Die Rasterauflösung des zu erstellenden DGM wird auf 1 m festgelegt. Das Tool arbeitet iterativ alle im Inputordner befindlichen ALS-Dateien ab und erstellt jeweils ein eigenes DGM. Die Dateien werden in einem zu definierenden Ausgabeordner im GeoTIFF-Format abgespeichert.

Im Anschluss sollen alle, im DGM-Ausgabeordner befindlichen, erstellten Kacheln zu einem Rasterlayer verschmolzen werden. Im Hinblick auf die Automatisierung ist es, wie bereits in Kapitel 4.2.3 erläutert, nicht möglich, alle Dateien innerhalb eines Ordners zu verschmelzen. Aufgrund dessen wurde der „gdal:merge“ Algorithmus in ein eigen erstelltes Python-Verarbeitungsskript implementiert. Das Python-Skript erwartet einen Ordner als Input. Als Output wird ein Rasterlayer im GeoTIFF-Format erstellt. Zu Beginn wird geprüft, ob der Eingabeordner existiert und ob eine Outputdatei angegeben wurde. Ohne Eingabeordner scheitert der Prozess und ohne angegebenen Speicherpfad wird mit einem Temporärlayer gearbeitet. Zusätzlich wird geprüft, ob und wie viele Rasterdateien im GeoTIFF-Format der Ordner enthält. Ist auch diese Bedingung erfüllt, wird die eigentliche Verschmelzung auf Grundlage des „gdal:merge“ Algorithmus durchgeführt. Im Anschluss wird zusätzlich überprüft, ob die Ausgabedatei erstellt wurde. Ist dies nicht geschehen, so zeigt sich dem Nutzer eine Fehlermeldung. Das Ergebnis des Python-Verarbeitungsskriptes ist ein DGM mit der räumlichen Ausdehnung des gesamten Untersuchungsgebietes. Wie bereits in Kapitel 4.2.3 umgesetzt, wurde auch hier die Konzeption durch den Verfasser der Arbeit übernommen. Im Anschluss wurde der Code durch das KI-Sprachmodell ChatGPT (OpenAI) (OpenAI, 2025) auf Grundlage des Konzepts generiert, durch den Verfasser überprüft und in den Arbeitsablauf implementiert. Der Code 2 in Anhang B zeigt den zentralen Verarbeitungsteil des Skripts.

Der letzte Schritt der Vorverarbeitung besteht darin, dem soeben erstellten DGM die korrekte Projektion ETRS89 UTM Zone 33N mit dem EPSG-Code 25833 zuzuweisen.

### 4.3.2 Standgewässer

Der Prozess zur Höhenermittlung für Standgewässer ist aufgrund seiner niedrigeren Anforderung im Vergleich zu den Fließgewässern bedeutend simpler. Die Standgewässer können aus zwei unterschiedlichen vorherigen Prozessierungsschritten stammen. Der genaue Grund hierfür wurde in Kapitel 4.2.5 genauer erläutert. Dies erfordert zwei separate Prozessierungsabfolgen, die exakt dieselben Algorithmen abarbei-

ten, jedoch unterschiedliche Eingabeparameter besitzen. Ist die Bedingung erfüllt, dass im Untersuchungsgebiet keine Fließgewässer vorkommen, wird das Gesamtklassifikationsergebnis als Input zur weiteren Berechnung genutzt. Sind jedoch Fließgewässer im Untersuchungsgebiet, wird der aus der Aufteilung stammende Seelayer als Input verwendet.

Die eigentliche Höhenberechnung wird mit Hilfe des SAGA Tools „Raster Statistics for Polygons“ durchgeführt. Als Eingabeparameter dient einerseits das aus den ALS Daten errechnete DGM sowie einer der eben genannten Polygonlayer. Das Tool bietet verschiedenste Statistikberechnungen. Innerhalb des Workflows wird das Minimum, das Maximum, der Mittelwert, die Standardabweichung und die 25. Perzentile für jedes Polygon innerhalb des Eingabelayers berechnet. Die Ausgabe erfolgt als Vektorlayer mit den berechneten Werten, angelegt als neue Felder.

Die berechneten Höhendaten liegen nun in Form von Werteeinträgen innerhalb der Attributtabelle vor. Jedoch sind die eigentlichen Polygone nach wie vor zweidimensional, besitzen also ausschließlich X und Y-Werte. Um diese in die dreidimensionale XYZ-Ebene zu versetzen ist eine Zuordnung von Z nötig. Dafür kann die Funktion „Z-Wert setzen“ genutzt werden. Als Eingabe dient dazu der erstellte Vektorlayer mit den berechneten Höhenattributen. Als Z kann beispielsweise ein statischer Wert übergeben werden. Innerhalb des automatisierten Workflows wird als Z-Wert das berechnete Höhenattribut der 25. Perzentile für jedes Polygon verwendet. Die Entscheidung zur Verwendung der 25. Perzentile wurde bewusst getroffen, um eine eher konservative Abschätzung der Gewässerhöhe zu erhalten. Dadurch wird auch der Einfluss von Ausreißern, die überwiegend Abweichungen nach oben darstellen, minimiert. In der Theorie könnten auch andere zuvor berechnete Höhenstatistiken wie der Mittelwert oder das Minimum als Z-Wert genutzt werden. Das finale Ergebnis der Prozessierung der Standgewässer ist ein Vektorlayer im GeoPackage-Format mit einer MultiPolygonZ-Geometrie. Bei der Abfrage der Objekte lässt sich nun die Höhe der einzelnen Standgewässer anhand des Z-Wertes ablesen.

### **4.3.3 Fließgewässer**

Die Verarbeitungskette zur Höhenableitung von Fließgewässern ist aufgrund der Anforderung der Erfassung einer variablen Gewässerhöhe entlang der Fließrichtung komplexer und verfolgt zudem einen anderen Ansatz. Die Durchführung ist zunächst dadurch bedingt, ob Flüsse innerhalb des Untersuchungsgebietes vorliegen. Ist diese Bedingung erfüllt, erfolgt der erste Bearbeitungsschritt. Hierbei wird die Layerausdehnung des aus den ALS-Daten erstellten DGM und somit die Bounding-Box des

Untersuchungsgebietes berechnet.

Die berechnete Ausdehnung dient als Input für die Funktion „Vektor auf Layermaske zuschneiden“. Als weitere Eingabe werden die in Kapitel 4.2.5 auf Gewässer erster Ordnung begrenzten Fließgewässerlinien des LUNG verwendet. Das Ergebnis der Berechnung ist ein Vektorlayer der Mittellinien aller Fließgewässer innerhalb des Untersuchungsgebietes. Die Durchführung des Zuschnitts ist daher von Bedeutung, weil im nachfolgenden Prozessierungsschritt die Funktion „Punkte entlang einer Geometrie“ ausgeführt wird. Im Workflow werden dabei im Abstand von 5 m und mit einem Startversatz von ebenfalls 5 m Punkte entlang der Gewässermittellinie erstellt.

Die erstellten Punkte dienen der Extraktion von Höhenwerten aus dem abgeleiteten DGM. Dazu ist der Algorithmus „Rasterwerte abtasten“ implementiert, der für jeden Punkt innerhalb des Punktlayers den dazugehörigen Höhenwert des DGM-Rasters abfragt. Die Werte werden innerhalb eines neu angelegten Attributs „hoehe“ für jeden Punkt gespeichert. Der nächste Arbeitsschritt des Workflows ermittelt zunächst mit Hilfe des Algorithmus „Stützpunkte extrahieren“ die sogenannten Vertices des Gewässerlayers in Form von Punkten.

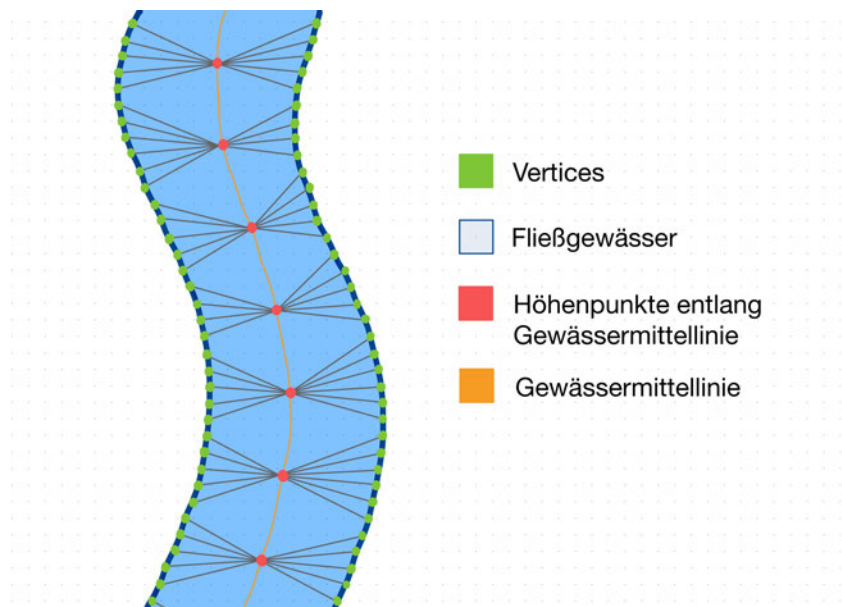


Abbildung 12: Veranschaulichung des Algorithmus „Attribute nach nächstem verknüpfen“ (eigene Darstellung)

Die beiden soeben erstellten Punktlayer bilden die Eingabedaten für die Funktion „Attribute nach nächstem verknüpfen“. Der Algorithmus erweitert einen Inputlayer um zusätzliche Attribute, die aus dem zweiten Input stammen. Dabei werden die Attributwerte immer den nächstgelegenen Objekten entnommen. Konkret bedeutet

das für den Workflow, dass die extrahierten Stützpunkte um das Höhenattribut der Punkte entlang der Flussmittellinie erweitert werden. Dabei wird den Stützpunkten immer der jeweils nächstgelegene Flussmittelpunkt beziehungsweise Höhenwert zugeordnet. Abbildung 12 veranschaulicht schematisch die Funktionsweise. Das Ergebnis der Ausführung des Algorithmus ist ein Punktlayer. Dieser repräsentiert die Stützpunkte der Gewässerumringe und besitzt für jeden Punkt einen Höhenwert.

Den Punktlayer benötigt man im Anschluss, um eine SAGA-Triangulation durchzuführen. Dabei ist neben der Angabe des Inputs auch das Attribut erforderlich, über dessen Werte trianguliert werden soll. Im Workflow wird dafür der zuvor ermittelte Höhenwert als Attribut übergeben. Zudem wird die räumliche Auflösung des Rasters in Übereinstimmung mit der Auflösung des DOP auf 0,1 m festgelegt. Das Ergebnis ist ein Rasterlayer mit zwischen den Stützpunkten triangulierten Höhendaten.

Der letzte Bearbeitungsschritt für die Ableitung der variablen Fließgewässerhöhen besteht darin, die Z-Werte aller Stützpunkte der Fließgewässer auf den aus dem triangulierten Rasterlayer abgetasteten Wert zu setzen. Die beschriebene Funktion wird durch den Algorithmus „Drapieren (Z-Wert von Raster übernehmen)“ umgesetzt. Das Ergebnis ist wie auch bei den Standgewässern (vgl. Kapitel 4.3.2) ein Vektorlayer im GeoPackage-Format mit einer MultiPolygonZ-Geometrie. Die variablen Höhenwerte können über die Objektabfrage ermittelt werden.

## 5 Auswertung der Ergebnisse

Die Auswertung der Ergebnisse erfolgt einzeln für jedes Untersuchungsgebiet. Eine visuelle Übersicht über die erstellten dreidimensionalen Gewässerumringe bieten die Abbildungen in den Kapiteln 5.1, 5.2 sowie 5.3. Die Genauigkeitsbetrachtung wird zunächst für das Klassifikationsergebnis, also die zweidimensionale Erfassung und im Anschluss für die Höhengenaugigkeit (3D-Erfassung), durchgeführt.

Um das Klassifikationsergebnis der zweidimensionalen Gewässerumringe bewerten zu können, wird die Konfusionsmatrix genutzt. Diese „[...] ist ein Instrument zur Bewertung der Leistung eines Modells [...]“ (DataCamp, o. D.[b]) und findet häufig im Bereich des Machine Learning Anwendung. Zur Erstellung einer Konfusionsmatrix ist das eigentliche Klassifikationsergebnis als Vorhersage sowie Ground-Truth Daten als Referenz erforderlich. Der Referenzdatensatz wurde für diese Arbeit manuell für jedes Untersuchungsgebiet erstellt. In der vorliegenden Arbeit ist die Konfusionsmatrix ein pixelbasierter Vergleich zwischen klassifizierter Gewässerfläche und Referenzdatensatz mit den Klassen „Wasser“ (Positive) und „Nicht-Wasser“ (Negative). Die Matrix wird in Form einer Tabelle dargestellt und besteht aus vier Hauptkomponenten (siehe Tabelle 1). True Positive (TP) Pixel sind im Referenzdatensatz Wasser und wurden auch als Wasser klassifiziert. Es stellt somit die richtig erkannten Wasserflächen dar. True Negative (TN) Pixel sind kein Wasser in der Referenz und wurden korrekt als kein Wasser klassifiziert. Sie stellen die richtig erkannten Landflächen dar. False Positive (FP) Pixel sind im Referenzdatensatz kein Wasser, wurden jedoch als Wasser klassifiziert. False Negative (FN) Pixel sind in der Referenz Wasser, wurden jedoch als kein Wasser klassifiziert.

	True Class	
	Positive	Negative
Predicted Class	True Positive (TP)	False Positive (FP)
	False Negative (FN)	True Negative (TN)

Tabelle 1: Allgemeine Konfusionsmatrix

Auf Grundlage der Konfusionsmatrix können verschiedene Leistungsmetriken berechnet werden. Die Genauigkeit (Overall Accuracy) (DataCamp, o. D.[b]) gibt an, welcher Anteil aller Pixel korrekt klassifiziert wurde, unabhängig ob Wasser- oder Nicht-Wasserpixel (siehe Gleichung 4). Je näher der Wert an 1 liegt, desto mehr Pixel wurden korrekt klassifiziert.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (4)$$

Die Precision (DataCamp, o.D.[b]) beschreibt die Zuverlässigkeit der Wasserklassifikation (siehe Gleichung 5). Der Wert gibt den Anteil der korrekt als Wasser klassifizierten Pixel von allen als Wasser klassifizierten Pixeln an. Je näher der Wert für die Precision an 1 liegt, desto weniger Nicht-Wasserpixel wurden fälschlich als Wasser klassifiziert.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

Der Recall (DataCamp, o.D.[b]) beschreibt die Vollständigkeit der Wasserklassifikation (siehe Gleichung 6). Der Wert gibt den Anteil der korrekt klassifizierten Wasserpixel von allen tatsächlich vorhandenen Wasserpixeln an. Je näher der Wert für den Recall an 1 liegt, desto vollständiger wurden die Wasserflächen erfasst.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

Der F1-Wert (DataCamp, o.D.[b]) ist das harmonische Mittel aus Precision und Recall (siehe Gleichung 7). Der Wert stellt die Genauigkeit der Gewässererkennung unter Berücksichtigung von Über- und Unterklassifikationen dar.

$$\text{F1} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

Für die dreidimensionale Auswertung wird die Höhengenaugkeit der ermittelten Gewässerrhöhen untersucht. Dabei kommt das DGM des AfGVK mit der Auflösung von 1 m als Referenzdatensatz zur Anwendung. Die Auswertung erfolgt dabei stichprobenartig mittels einzelner Punkte innerhalb der Gewässerpolygone. Für Standgewässer wird aufgrund der gleichbleibenden Höhe ein Punkt pro Polygon genutzt. Für Fließgewässer werden regelmäßige Punkte in einem Abstand von 2 m über das Gewässer verteilt. Anschließend erfolgt die Ermittlung der punktweisen Höhendifferenzen und daraus die Berechnung verschiedener Leistungsmetriken.

Die Höhendifferenz  $\Delta h_i$  beschreibt die punktweise Höhenabweichung und wird aus der Differenz zwischen der abgeleiteten Gewässerrhöhe und der DGM-Höhe bestimmt (siehe Gleichung 7). Dabei bezeichnet  $n$  die Anzahl der Stichprobenpunkte,  $h_{\text{Gew},i}$  die abgeleitete Gewässerrhöhe und  $h_{\text{DGM},i}$  die Referenzhöhe des Digitalen Gelände-modells.

$$\Delta h_i = h_{\text{Gew},i} - h_{\text{DGM},i} \quad (8)$$

Der sogenannte Bias (auch Verzerrung oder systematischer Fehler) (Wikipedia, o. D.[a]) stellt den arithmetischen Mittelwert der Höhendifferenzen dar (siehe Gleichung 9). Der Wert ist ein Maß für die systematische Abweichung und zeigt Über- oder Unterschätzungen an. Der Bias ist vorzeichenbehaftet, zeigt somit auch die Richtung der Abweichung (positiv oder negativ) an.

$$\text{Bias} = \frac{1}{n} \sum_{i=1}^n \Delta h_i \quad (9)$$

Der Mean Absolute Error (MAE) (Wikipedia, o. D.[b]) ist der Mittelwert der absoluten Höhendifferenzen und stellt ein Maß für die mittlere absolute Abweichung dar (siehe Gleichung 10). Da die Metrik absolute Werte nutzt, ist sie robust gegenüber der Ausgleichung positiver und negativer Abweichungen.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\Delta h_i| \quad (10)$$

Der Root Mean Squared Error (RMSE) (Wikipedia, o. D.[c]) ist die Wurzel des Mittelwertes der quadrierten Höhendifferenzen (siehe Gleichung 11). Aufgrund der Quadrierung liegt eine stärkere Gewichtung auf großen Abweichungen, was wiederum zu einer Sensitivität gegenüber Ausreißern führt.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\Delta h_i)^2} \quad (11)$$

## 5.1 Stralsund

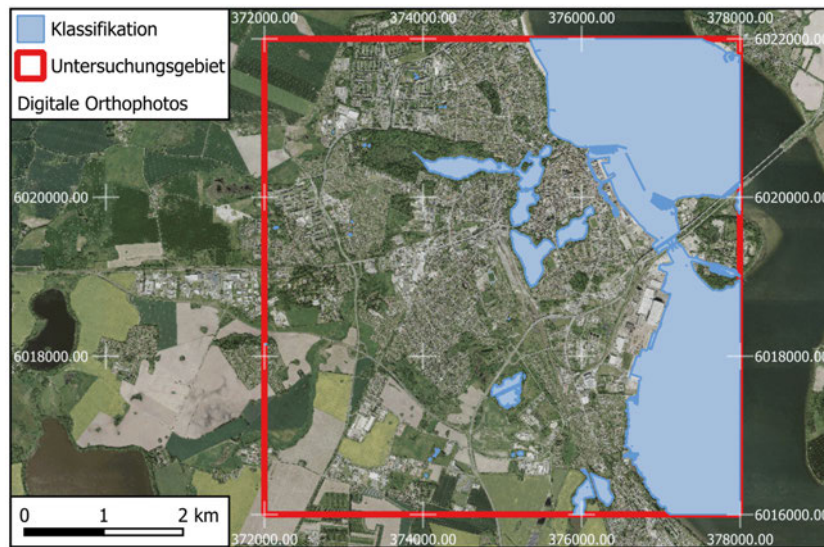


Abbildung 13: Workflow-Ergebnis in Stralsund (eigene Darstellung)

Jedes Untersuchungsgebiet erstreckt sich über eine Ausdehnung von 6 km \* 6 km beziehungsweise 6000 m \* 6000 m. Bei einer Rasterauflösung von 0,1 m ergibt sich eine Ausdehnung von 60 000 \* 60 000 Pixel und eine daraus resultierende Fläche von 3 600 000 000 Pixeln. Durch die Addition aller Pixelwerte innerhalb der Konfusionsmatrix kann die vollständige Zuordnung aller Pixel sichergestellt werden.

		True Class	
		Positive	Negative
Predicted Class	Positive	912 851 042 (TP)	3 239 273 (FP)
	Negative	4 025 101 (FN)	2 679 884 584 (TN)

Tabelle 2: Konfusionsmatrix von Stralsund

Aus der vorliegenden Konfusionsmatrix für das Stralsunder Untersuchungsgebiet (siehe Tabelle 2) ergibt sich, dass  $\approx 74,4\%$  aller Pixel True Negative,  $\approx 25,4\%$  True Positive,  $\approx 0,1\%$  False Negative und  $\approx 0,1\%$  False Positive sind.

Tabelle 3 stellt die Ergebniswerte der zweidimensionalen Genauigkeitsbewertung auf Grundlage der Konfusionsmatrix übersichtlich dar. Hierbei wird eine Accuracy von  $\approx 99,8\%$  erreicht. Die Werte für Precision, Recall und F1 liegen mit  $\approx 99,6\%$  bei einer ähnlichen Genauigkeit.

Metrik	Accuracy	Precision	Recall	F1
Ergebniswert	$\approx 0,997982$	$\approx 0,996464$	$\approx 0,99561$	$\approx 0,996037$

Tabelle 3: 2D-Genauigkeitsbewertung in Stralsund

Im Untersuchungsgebiet von Stralsund wurden ausschließlich Standgewässer detektiert. Für die stichprobenartige dreidimensionale Genauigkeitsbewertung lagen 79 Punkte und somit 79 verschiedene Höhenwerte zugrunde. Der Mittelwert, ausgedrückt durch den Bias, beläuft sich auf  $\approx +9,6\text{cm}$ . Die mittlere absolute Abweichung ist mit  $\approx 30,6\text{cm}$  mehr als dreifach so groß. Der RMSE ist mit einer Abweichung von  $\approx 68\text{cm}$  nochmal mehr als doppelt so groß (siehe Tabelle 4).

Gewässertyp	n-Punkte	Bias [m]	MAE [m]	RMSE [m]
Standgewässer	79	$\approx +0,095937$	$\approx 0,305962$	0,680031

Tabelle 4: 3D-Genauigkeitsbewertung in Stralsund

## 5.2 Bützow

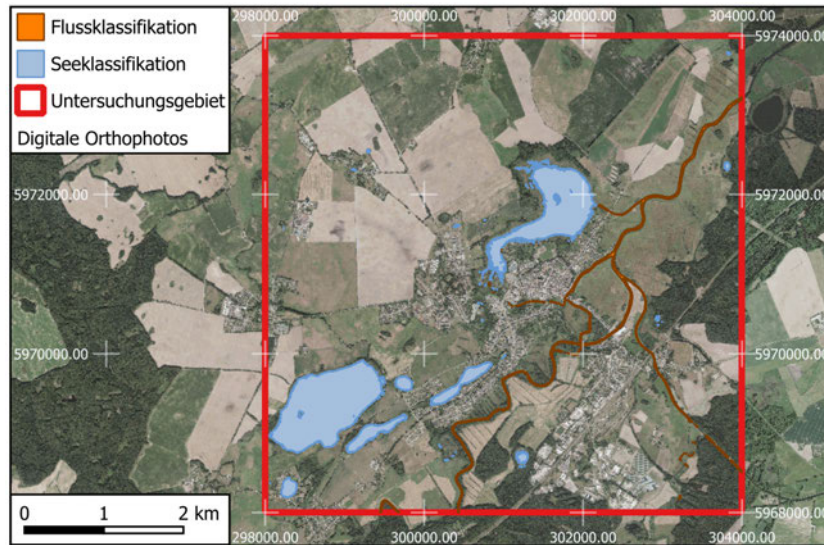


Abbildung 14: Workflow-Ergebnis in Bützow (eigene Darstellung)

Die Addition der Pixelwerte der Konfusionsmatrix des Untersuchungsgebietes im Raum Bützow ergibt, dass alle Pixel vollständig zugeordnet worden sind. Im Vergleich zu Stralsund ist der Anteil der True Negative Pixel mit  $\approx 93,9\%$  nochmal bedeutend größer (siehe Tabelle 5). Nur knapp  $\approx 5,8\%$  der Pixel sind True Positive,  $\approx 0,2\%$  False Positive und  $\approx 0,1\%$  False Negative.

	True Class	
	Positive	Negative
Positive	206 839 066 (TP)	8 472 101 (FP)
Negative	3 823 669 (FN)	3 380 865 164 (TN)

Tabelle 5: Konfusionsmatrix von Bützow

Die zweidimensionale Genauigkeitsbewertung ergibt, dass eine Accuracy von  $\approx 99,7\%$  vorliegt. Jedoch ist die Precision mit  $\approx 96,1\%$  etwa  $\approx 3\%$  niedriger. Der Recall beträgt  $\approx 98,2\%$  und auch der F1-Score liegt mit  $\approx 97,1\%$  deutlich unter der Accuracy (siehe Tabelle 6).

Metrik	Accuracy	Precision	Recall	F1
Ergebniswert	$\approx 0,996585$	$\approx 0,960652$	$\approx 0,981849$	$\approx 0,971135$

Tabelle 6: 2D-Genauigkeitsbewertung in Bützow

Im Untersuchungsgebiet von Bützow wurden Stand- als auch Fließgewässer detektiert. Die Statistikberechnung für die Standgewässer beruht auf 187 verschiedenen

Datenpunkten. Der Bias beläuft sich auf  $\approx +0,8\text{cm}$ . Der MAE ist mit  $\approx 5,2\text{cm}$  nur ein Drittel so groß wie in Stralsund und auch der RMSE liegt mit  $\approx 12,5\text{cm}$  bei etwa einem Fünftel des Wertes von Stralsund. Für Fließgewässer wurden die Daten von knapp 58 000 Punkten ausgewertet. Die mittlere Abweichung beträgt  $\approx +4,4\text{cm}$ . Der MAE ist mit  $\approx 5,1\text{cm}$  nur geringfügig höher. Mit  $\approx 11,75\text{cm}$  ist der RMSE auf einem ähnlichen Niveau wie der Wert für Standgewässer in diesem Untersuchungsgebiet (siehe Tabelle 7).

<b>Gewässertyp</b>	<b>n-Punkte</b>	<b>Bias [m]</b>	<b>MAE [m]</b>	<b>RMSE [m]</b>
Fließgewässer	57 901	$\approx +0,043884$	$\approx 0,050764$	$\approx 0,1175$
Standgewässer	187	$\approx +0,007845$	$\approx 0,052444$	$\approx 0,125227$

Tabelle 7: 3D-Genauigkeitsbewertung in Bützow

### 5.3 Feldberger Seenlandschaft

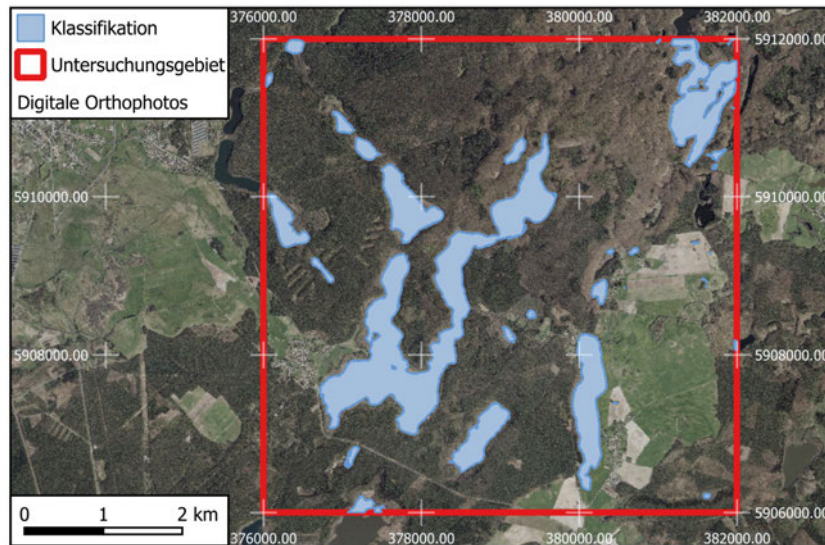


Abbildung 15: Workflow-Ergebnis in der Feldberger Seenlandschaft (eigene Darstellung)

Auch für das letzte Untersuchungsgebiet in der Feldberger Seenlandschaft lief die Zuordnung der 3 600 000 000 Pixel vollständig ab. Ähnlich zu der Werteverteilung in Stralsund und Bützow liegt der Anteil an True Negative Werten mit  $\approx 87,4\%$  in der absoluten Mehrheit. True Positive Werte machen einen Anteil von  $\approx 12,1\%$  aus, gefolgt von False Negative ( $\approx 0,26\%$ ) und False Positive ( $\approx 0,24\%$ ).

	True Class	
	Positive	Negative
Positive	437 014 670 (TP)	8 828 788 (FP)
Negative	9 391 736 (FN)	3 144 764 806 (TN)

Tabelle 8: Konfusionsmatrix der Feldberger Seenlandschaft

Die zweidimensionalen Genauigkeitsbewertung der Feldberger Seenlandschaft ergibt, dass die Accuracy mit  $\approx 99,5\%$  ähnlich den Werten der vorherigen Untersuchungsgebiete ist. Die Precision ist im Vergleich zur Accuracy mit  $\approx 98\%$  knapp  $\approx 1,5\%$  niedriger. Der Recall bewegt sich mit einem Wert von  $\approx 97,9\%$  auf einem ähnlichen Niveau wie die Precision. Der F1-Score als harmonisches Mittel zwischen Precision und Recall liegt bei  $\approx 98\%$ .

<b>Metrik</b>	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1</b>
<b>Ergebniswert</b>	$\approx 0,994939$	$\approx 0,980198$	$\approx 0,978961$	$\approx 0,979579$

Tabelle 9: 2D-Genauigkeitsbewertung in der Feldberger Seenlandschaft

In der Feldberger Seenlandschaft wurden ausschließlich Standgewässer detektiert. Die Statistikberechnung basiert auf 86 unterschiedlichen Datenpunkten. Als einziges Untersuchungsgebiet ergibt sich ein negativer Bias von  $\approx -1\text{cm}$ . Der MAE beläuft sich auf eine Abweichung in der gleichen Größenordnung von  $\approx 1\text{cm}$ . Auch der RMSE ist mit  $\approx 3,9\text{cm}$  der niedrigste Wert von allen Untersuchungsgebieten.

<b>Gewässertyp</b>	<b>n-Punkte</b>	<b>Bias [m]</b>	<b>MAE [m]</b>	<b>RMSE [m]</b>
Standgewässer	86	$\approx -0,009512$	$\approx 0,010488$	$\approx 0,038629$

Tabelle 10: 3D-Genauigkeitsbewertung in der Feldberger Seenlandschaft

## 6 Diskussion der Ergebnisse

### 6.1 Einfluss der Referenzflächen

Vor der eigentlichen Diskussion der zwei- und dreidimensionalen Ergebniswerte muss eine Einschätzung über den Einfluss der Referenzflächen auf die Auswertung erfolgen. Da die Gewässerausdehnungen manuell nach visuellen Kriterien auf Grundlage des DOP erstellt worden sind, ist es von relevanter Bedeutung zu wissen, nach welchen Kriterien Wasser definiert ist.

Als Referenz wurden eindeutig identifizierbare und sichtbare Wasserflächen gewählt. Dabei sind stark bewachsene Bereiche (durch Seerosen und Schilf) oder von Vegetation überdeckte Abschnitte ausgeschlossen. In Abbildung 16 ist ein Beispiel für eine Referenzfläche abgebildet. Durch die Transparenz des Layers kann nachvollzogen werden, inwiefern bewachsene Gewässerabschnitte aus der Referenzerstellung ausgeschlossen wurden. Das Klassifikationsergebnis wird dadurch nicht gegen die „wahre“, hydrologisch korrekte Gewässerausdehnung geprüft, sondern gegen das Detektionsziel des Algorithmus.

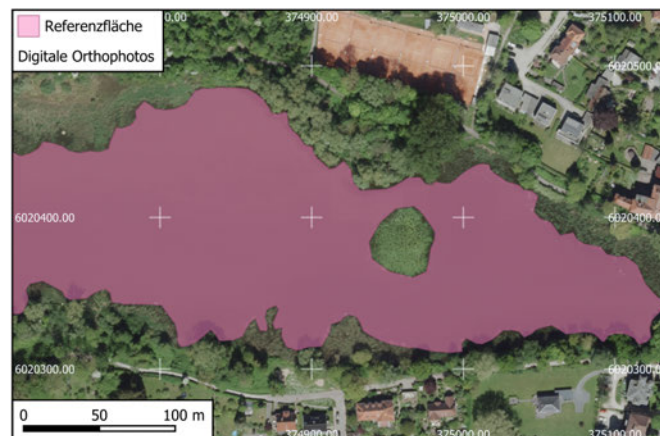


Abbildung 16: Beispiel einer Referenzfläche in Stralsund (eigene Darstellung)

Diese Vorgehensweise bietet den Vorteil, dass das Ergebnis des Algorithmus (als Klassifikation von „sichtbarem“ Wasser) mit der Referenz als Ziel des Algorithmus verglichen werden kann. Jedoch können mit diesem Ansatz keine Aussagen zur vollständigen hydrologischen Gewässerausdehnung gemacht werden. Hohe Genauigkeitswerte bedeuten also nicht automatisch eine vollständige Erfassung aller Gewässer. Eine große Übereinstimmung zwischen der Klassifikation und den visuell erstellten Ausdehnungen beschreibt somit eher die vollständige Erfassung aller dem Algorithmus möglich erkennbaren Wasserflächen.

Weitere Einflüsse auf die Genauigkeiten sind auf die manuelle Erstellung der Referenzdaten zurückzuführen. Die sogenannten „Ground Truth“-Daten sind während der manuellen Erfassung immer interpretationsabhängig und subjektiv, stellen in der vorliegenden Arbeit also nur eine visuelle Referenz dar. Diese Referenzflächen spiegeln aufgrund dessen nicht die objektive Wahrheit wider. Beispielsweise sind Übergangsbereiche zwischen der offenen Wasserfläche und Bewuchs nicht eindeutig und können zu einer variablen Abgrenzung der Gewässerränder führen. Das hat insbesondere Einfluss auf die False Positives und False Negatives der Konfusionsmatrix (vgl. Kapitel 5) und in Folge dessen auch auf die Genauigkeitswerte Precision und Recall. Eine konservative Referenz bewirkt dadurch beispielsweise einen Anstieg der Precision.

## 6.2 Einordnung der 2D-Ergebnisse

Bei der Betrachtung der Accuracy fällt über alle drei Untersuchungsgebiete hinweg auf, dass dieser mit Werten von  $> 99\%$  sehr hoch ist. Die frühzeitige Annahme eines nahezu perfekten Klassifikationsergebnisses entspricht jedoch nicht den Tatsachen. Die Berechnung der Accuracy wird stark beeinflusst durch den hohen Anteil an True Negative-Pixeln, also korrekt klassifizierten Landflächen. In den Untersuchungsgebieten hat dieser mit Werten von  $\approx 75 - 94\%$  aller Bildpixel jedoch eine absolute Mehrheit. Durch diese Klassenungleichverteilung wird bereits eine hohe Accuracy allein durch die Klassifizierung von Nicht-Wasserpixeln erreicht. Für die Einschätzung der Klassifikationsgenauigkeit eignen sich eher die Werte für Precision und Recall beziehungsweise der sich daraus erschließende F1-Score (vgl. Kapitel 5), da diese spezifisch die Erfassung der Gewässerausdehnungen bewerten.

In Stralsund weist der F1-Score mit einem Wert von  $\approx 99,6\%$  eine sehr hohe Genauigkeit auf. Das kann unter anderem auf das vermehrte Vorkommen von eindeutigen Grenzen zwischen Wasser und Land aufgrund der Hafenanlagen, den vorhandenen Kaimauern und anderen Uferbauwerken zurückgeführt werden. Ein Beispiel dafür lässt sich in der Nahbetrachtung des Klassifikationsergebnisses in Abbildung 17 finden. Die Hafensinsel in Stralsund bietet sehr kontrastreiche Grenzen zwischen Wasser und Land und erleichtert dadurch eine genaue Extraktion von Gewässerumringen. Des Weiteren sind die detektierten Standgewässer im Untersuchungsgebiet überwiegend groß (Ostsee) und nur vereinzelt durch Vegetation eingeschränkt.

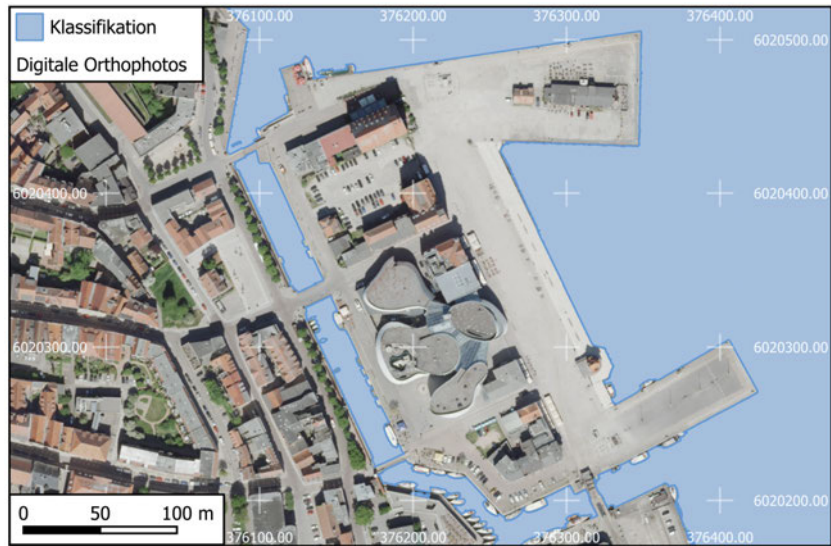


Abbildung 17: Ausschnitt des Klassifikationsergebnisses bei Stralsund (eigene Darstellung)

Dazu weist das Untersuchungsgebiet bei Bützow einen vergleichsweise niedrigen Wert von  $\approx 96,1\%$  für die Precision auf, wohingegen der Recall mit  $\approx 98,2\%$  noch relativ hoch ausfällt. Ursache hierfür sind vermutlich die im Vergleich zu Stralsund für die Klassifikation schwierigeren Landschaftseigenschaften. Bützow besitzt als einziges Untersuchungsgebiet schmale Fließgewässer. Zudem sind die Uferbereiche zum Teil stark bewachsen und der Übergang von Wasser zu Land erschwert zusätzlich die Erstellung von Referenzflächen (siehe Abbildung 18).



Abbildung 18: Ausschnitt des Klassifikationsergebnisses bei Bützow (eigene Darstellung)

Das Untersuchungsgebiet in der Feldberger Seenlandschaft liegt mit Genauigkeitswerten von  $\approx 98\%$  für F1 zwischen den Ergebnissen von Stralsund und Bützow. Dies lässt sich auf gut erkennbare Wasserflächen mit deutlichen Kontrasten zwischen Vegetation und Wasser zurückführen. Des weiteren weisen die Standgewässer nur geringen Bewuchs durch beispielsweise Seerosen oder sonstige Hydrophyten auf. Das Ergebnis könnte besser ausfallen, jedoch ist eine Wasserfläche vollständig undetektiert geblieben. Dies ist möglicherweise auf eine starke Trübung und in Folge dessen auf eine ungewöhnliche Reflexionskurve (vgl. Kapitel 2.1.1) zurückzuführen. Insgesamt zeigen die Ergebnisse, dass eine sehr gute zweidimensionale Klassifikation möglich ist, diese jedoch von den lokalen landschaftlichen Gegebenheiten und Gewässerstrukturen abhängt.

### 6.3 Einordnung der 3D-Ergebnisse

Bei dem Vergleich der dreidimensionalen Genauigkeitsauswertung zwischen den drei Untersuchungsgebieten fällt auf, dass große Unterschiede bestehen. Für weitere Rückschlüsse ist eine gemeinsame Betrachtung des Bias, MAE und RMSE notwendig.

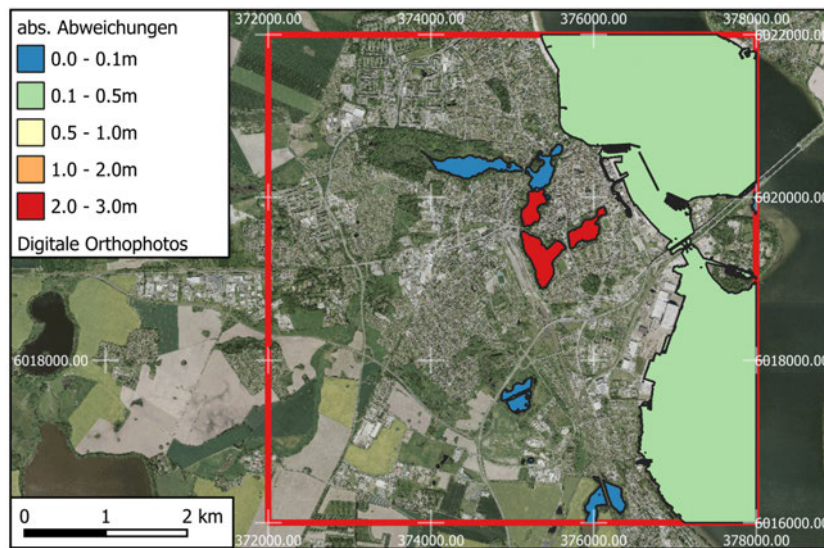


Abbildung 19: Darstellung der absoluten Höhenabweichungen in Stralsund (eigene Darstellung)

Das Untersuchungsgebiet, das die besten zweidimensionalen Genauigkeiten aufweisen konnte, besitzt die ungenauesten ermittelten Höheninformationen. Schon mit einem moderaten Bias von  $\approx +10\text{cm}$  zeigt Stralsund die höchste Verzerrung. Die Ungenauigkeit nimmt bei der Betrachtung von MAE und RMSE deutlich zu. Mit

einem Wert von  $\approx 68\text{cm}$  liegt der RMSE mehr als doppelt so hoch als der MAE. Das deutet auf vereinzelte, jedoch sehr starke Abweichungen hin. Grund dafür sind wahrscheinlich Gewässerfehldetektionen in Ufernähe durch Schiffe oder sonstige Hafenanbauwerke, die Höhenunterschiede von mehreren Metern verursachen. Abbildung 19 bietet eine Übersicht über die Verteilung der absoluten Abweichungen der Gewässerhöhen. Bei der Betrachtung fällt insbesondere die große Höhenabweichung von 2 – 3 m im Knieper- und Frankenteich auf. Das ist auf den Fakt zurückzuführen, dass die Seen im DGM eine Höhe von 2,6 m aufweisen, jedoch die Berechnung der 25. Perzentile einen Höhenwert von 0 m ergibt. Hinzu kommt die niedrige Anzahl an Datenpunkten, die den Einfluss der Ausreißer steigern.

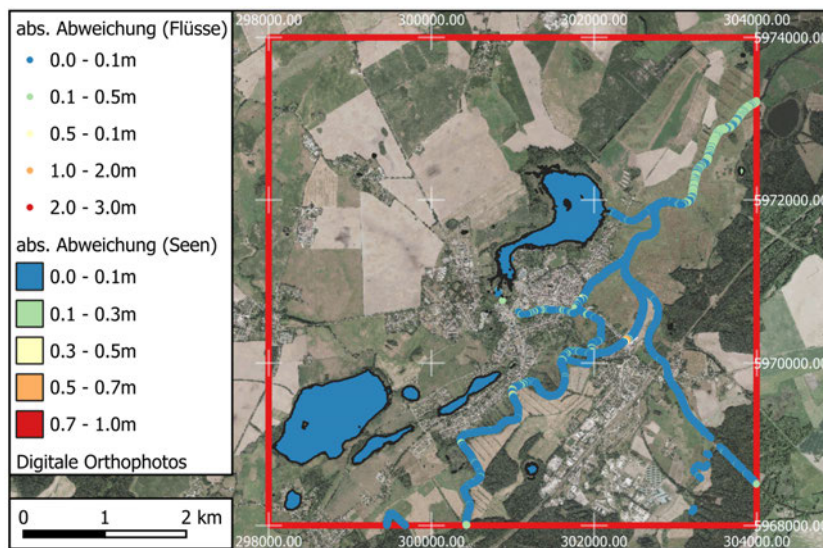


Abbildung 20: Darstellung der absoluten Höhenabweichungen in Bützow (eigene Darstellung)

Im Gegensatz dazu bietet Bützow bedeutend niedrigere Höhenabweichungen. Der Bias und MAE bei Stand- als auch Fließgewässern beläuft sich auf mittlere einstellige Zentimeter. Auch der RMSE ist mit Werten von  $\approx 12\text{cm}$  nicht wesentlich höher, was auf nur wenige stärkere Ausreißer hinweist. Zusätzlich profitiert die Höhenbewertung der Fließgewässer mit  $\approx 58\,000$  Punkten von einer großen Datenmenge zur statistischen Auswertung. Diese regelmäßige Punktverteilung minimiert zufällige Abweichungen in den variablen Höhen. Bei der Übersicht über die Verteilung der Höhenabweichungen in Abbildung 20 fallen die vereinzelt auftretenden Ausreißer in den Fließgewässern auf. Die Abweichungen der Oberflächenhöhen der Standgewässer sind sehr gering.

Durch den in Kapitel 4.3.3 genauer erläuterten Arbeitsablauf zur Höhenermittlung der Fließgewässer entstehen qualitative Schwächen. Abbildung 21 zeigt diese Schwä-

chen anschaulich. Durch die Entnahme der Höhen aus zufällig gewählten Datenpunkten entlang der Flusslinie und der anschließenden Übertragung dieser Höhen auf die nahegelegenen Vertices entstehen Stufen innerhalb des Fließgewässers. Diese Stufen erstrecken sich quer zur Fließrichtung und weisen Höhendifferenzen von bis zu  $\approx 10\text{cm}$  auf. Eine erste simple Möglichkeit zur Reduktion der Stufen wäre der Einsatz einer glättenden Interpolation zwischen den Datenpunkten.



Abbildung 21: Ausschnitt aus der 3D-Betrachtung des klassifizierten Flusslayers in Bützow (eigene Darstellung)

Die besten Genauigkeitswerte werden in der Feldberger Seenlandschaft erzielt. Der Bias weist hier als einziges Untersuchungsgebiet einen negativen Wert auf, was eine Unterschätzung der Gewässershöhe bedeutet. Dieser Wert befindet sich jedoch mit nur  $\approx -1\text{cm}$  in einer sehr kleinen Größenordnung. Auch die absolute Abweichung (MAE) und der RMSE weisen lediglich Abweichungen im niedrigen einstelligen Zentimeterbereich auf. Das kann möglicherweise auf die Einheitlichkeit der Landschaft und die nur vereinzelt vorkommenden Bauwerke in Gewässernähe zurückzuführen sein. In Abbildung 22 werden die niedrigen Höhenabweichungen der Standgewässer des Untersuchungsgebietes anschaulich dargestellt.

Insgesamt unterscheiden sich die dreidimensionalen Genauigkeiten stark von der vorliegenden Landschaftsstruktur. Hinzu kommt der Einfluss von Abweichungen durch niedrige Anzahlen von Stichprobenpunkten. Die Ermittlung genauer Gewässershöhen auf Basis des automatisierten Workflows ist jedoch grundsätzlich möglich.

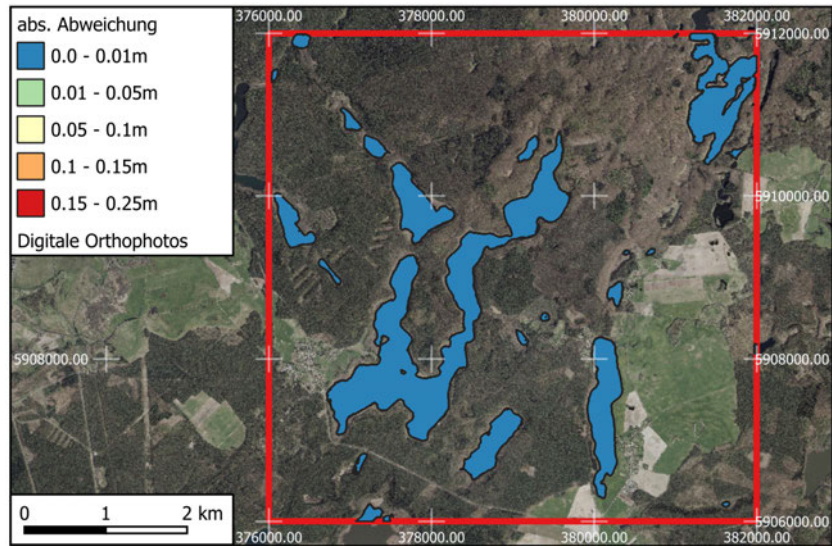


Abbildung 22: Darstellung der absoluten Höhenabweichungen in der Feldberger Seenlandschaft (eigene Darstellung)

## 7 Fazit und Ausblick

Abschließend lässt sich feststellen, dass das Ziel der Arbeit, die Erstellung eines automatisierten Workflows zur Erfassung dreidimensionaler Gewässerumringe, erreicht wurde. Der erstellte Arbeitsablauf verbindet dabei verschiedene Eingaben, darunter Daten von passiven als auch aktiven Aufnahmeverfahren und kombiniert diese zur Nutzung verschiedener Nachbearbeitungsschritte. Die Gewässerklassifikation liefert gute Ergebnisse, insbesondere bei kontrastreichen Gewässergrenzen. Unzuverlässigkeiten der automatisierten Auswertung entstehen hinsichtlich schmaler und bewachsener Geometrien mit unklarem Übergang von Wasser zu Land. Die Qualität der Höhenabweichung variiert stark von den betrachteten Gebieten. Hohe Abweichungen entstehen insbesondere in Bereichen mit starker Bebauung in unmittelbarer Ufernähe, wogegen naturnahe und homogene Landschaften niedrige Höhenabweichungen aufweisen. Unabhängig von dem bestehenden Arbeitsablauf besteht eine gewisse Unsicherheit in der Bewertung hinsichtlich der manuell erstellten Referenzflächen, da deren Qualität einen direkten Einfluss auf die Bewertung hat. Insgesamt stellt die Methode jedoch eine Möglichkeit zur halbautomatisierten Erfassung von Gewässerumringen dar. Die beschriebenen Limitationen können durch manuelle Korrekturen, im Vergleich zur vollständig manuellen Erfassung, schnell behoben werden.

Als Ausblick bietet der Einsatz von überwacht trainierten Deep Learning-Modellen zur Gewässerklassifikation Potenzial zur stabileren Erfassung schmaler und stark bewachsener Gewässergeometrien. Hierbei muss zwischen dem höheren Implementierungsaufwand und der möglichen höheren Qualität der Ergebnisse abgewogen werden. Ein weiteres Augenmerk sollte in Zukunft auf die variable Höhenzuweisung für Fließgewässer gelegt werden. Aktuell werden aus einzelnen, zufällig gewählten Punkten entlang der Gewässermittellinie die Höhenwerte interpoliert. Hierbei könnte es zu starken lokalen Abweichungen kommen, die großen Einfluss auf die dreidimensionale Abbildung hätten. Die Weiterentwicklung der variablen Gewässerhöhenmodellierung sollte dabei eine höhere Robustheit gegenüber Ausreißern aufweisen.

# Literaturverzeichnis

- Albertz, Jörg (2007). *Einführung in die Fernerkundung: Grundlagen der Interpretation von Luft- und Satellitenbildern*. German. 3., aktualisierte und erw. Aufl. Literaturverz. S. 229 - 242. Darmstadt: Wissenschaftliche Buchgesellschaft (wbg). ISBN: 3534198786.
- DataCamp (o. D.[a]). *Supervised Machine Learning*. Online, <https://www.datacamp.com/de/blog/supervised-machine-learning>. [abgerufen am 2025-12-30].
- (o. D.[b]). *What Is a Confusion Matrix in Machine Learning?* Online, <https://www.datacamp.com/de/tutorial/what-is-a-confusion-matrix-in-machine-learning>. [abgerufen am 2025-12-22].
- de Lange, Norbert (2013). *Geoinformatik: in Theorie und Praxis*. German. 3., vollst. überarb. u. akt. Aufl. 2013. SpringerLink Bücher. Description based upon print version of record. Berlin, Heidelberg: Springer. ISBN: 9783642348075. DOI: 10.1007/978-3-642-34807-5.
- (2020). *Geoinformatik in Theorie und Praxis: Grundlagen von Geoinformationssystemen, Fernerkundung und digitaler Bildverarbeitung*. German. 4., grundlegend überarbeitete und erweiterte Auflage. Lehrbuch. Literaturangaben. Berlin: Springer. ISBN: 9783662607091. DOI: 10.1007/978-3-662-60709-1.
- European Space Agency (ESA) (o. D.). *Sentinel-2: factsheet*. [https://esamultimedia.esa.int/docs/EarthObservation/SENTINEL-2\\_sheet\\_170125.pdf](https://esamultimedia.esa.int/docs/EarthObservation/SENTINEL-2_sheet_170125.pdf). [abgerufen am 2025-12-05].
- Gao, Bo-Cai (1996). „NDWI A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space“. In: *Remote Sensing of Environment* 58.3, S. 257-266. DOI: 10.1016/S0034-4257(96)00067-3.
- GRASS Development Team (o. D.). *r.sunmask*. GRASS GIS Manual. [abgerufen am 2025-12-09]. URL: <https://grass.osgeo.org/grass-stable/manuals/r.sunmask.html>.
- Gschöpf, Christine und Alfred P. Blaschke (2009). „Unterstützung gekoppelter hydraulischer Modelle durch Fernerkundungsdaten“. In: *Österreichische Wasser- und Abfallwirtschaft* 61.7-8, S. 111-115. DOI: 10.1007/s00506-009-0096-2.
- Heipke, Christian, Hrsg. (2017). *Photogrammetrie und Fernerkundung: Handbuch der Geodäsie, herausgegeben von Willi Freeden und Reiner Rummel*. German. Springer Reference Naturwissenschaften. herausgegeben von Christian Heipke. Berlin, Heidelberg: Springer. ISBN: 9783662470947. DOI: 10.1007/978-3-662-47094-7.
- Hildebrandt, Gerhard (1996). *Fernerkundung und Luftbildmessung für Forstwirtschaft, Vegetationskartierung und Landschaftsökologie*. Heidelberg: Wichmann.

- International Society for Photogrammetry and Remote Sensing (2022). *Statutes*. <https://www.isprs.org/documents/statutes.aspx>. [abgerufen am 2025-12-02].
- Karasiak, Nicolas (2016). *Dzetsaka Qgis Classification plugin*. DOI: 10.5281/zenodo.2552284. URL: <https://github.com/nkarasiak/dzetsaka>.
- Klaus, Josefine, Hrsg. (2023). *Geomatik: Eine Einführung*. German. 1st ed. 2023. herausgegeben von Josefine Klaus. Berlin, Heidelberg: Springer. ISBN: 9783662662748. DOI: 10.1007/978-3-662-66274-8.
- Kuechly, H. u. a. (2020). *Grundlagen der Fernerkundung*. Inforeihe SAPIENS: Satellitendaten für Planung, Industrie, Energiewirtschaft und Naturschutz. Potsdam: Deutsches GeoForschungsZentrum GFZ, S. 25. DOI: 10.2312/sapiens.2020.001.
- Landesamt für innere Verwaltung Mecklenburg-Vorpommern (2025). *Übersicht Verfügbarkeit ALS - Daten*. Online-PDF. [abgerufen am 2025-12-09]. URL: [https://www.geodaten-mv.de/dienste/karten\\_akt\\_uebersicht\\_file/Uebersicht\\_Verfuegbarkeit\\_ALS-Daten.pdf](https://www.geodaten-mv.de/dienste/karten_akt_uebersicht_file/Uebersicht_Verfuegbarkeit_ALS-Daten.pdf).
- (o. D.[a]). *Geländemodelle*. Online. [abgerufen am 2025-12-09]. URL: <https://www.laiv-mv.de/Geoinformation/Geobasisdaten/Gelaendemodelle/>.
- (o. D.[b]). *Geotopographie ALS-Produktbeschreibung*. Online, <https://laiv.geodaten-mv.de/afgvk/Geotopographie/Beschreibung?produkt=ALS>. [abgerufen am 2025-12-09].
- (o. D.[c]). *Geotopographie DLM-Produktbeschreibung*. Online, <https://laiv.geodaten-mv.de/afgvk/Geotopographie/Beschreibung?produkt=DLM>. [abgerufen am 2025-12-09].
- (o. D.[d]). *Geotopographie DOM-Produktbeschreibung*. Online, <https://laiv.geodaten-mv.de/afgvk/Geotopographie/Beschreibung?produkt=DOM>. [abgerufen am 2025-12-09].
- (o. D.[e]). *Luftbilder*. <https://www.laiv-mv.de/Geoinformation/Luftbilder/>. [abgerufen am 2025-12-05].
- (o. D.[f]). *Luftbilder DOP - Produktbeschreibung*. <https://laiv.geodaten-mv.de/afgvk/Luftbilder/Beschreibung?produkt=DOP>. [abgerufen am 2025-12-09].
- (o. D.[g]). *Oberflächenmodelle*. Online, <https://www.laiv-mv.de/Geoinformation/Geobasisdaten/Oberflaechenmodelle/>. [abgerufen am 2026-01-01].
- Landesamt für Umwelt, Naturschutz und Geologie Mecklenburg-Vorpommern (2025). *Geotope Feldberg (PDF)*. Online-PDF. [abgerufen am 2025-12-08]. URL: <https://www.laiv-mv.de/Geoinformation/Geobasisdaten/Geotope/Feldberg/PDF>.

- [//www.lung.mv-regierung.de/static/LUNG/dateien/fachinformationen/Geologie/geotope/04.feldberg.ng.june25.pdf](https://www.lung.mv-regierung.de/static/LUNG/dateien/fachinformationen/Geologie/geotope/04.feldberg.ng.june25.pdf).
- Landesamt für Umwelt, Naturschutz und Geologie Mecklenburg-Vorpommern (o. D.). *Umweltkarten Mecklenburg-Vorpommern interaktive Kartenanwendung*. Online, <https://www.umweltkarten.mv-regierung.de/script/>. [abgerufen am 2025-12-09].
- McFeeters, S. K. (1996). „The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features“. In: *International Journal of Remote Sensing* 17.7, S. 1425–1432. DOI: 10.1080/01431169608948714.
- Müritz-Nationalpark (2025). *Startseite*. <https://www.mueritz-nationalpark.de/>. [abgerufen am 2025-12-08].
- numiqo Team (o. D.). *Logistische Regression*. Online, <https://www.datacamp.com/de/blog/supervised-machine-learning>. [abgerufen am 2025-12-30].
- OpenAI (2025). *ChatGPT*. <https://chat.openai.com/>. KI-gestütztes Sprachmodell zur Unterstützung bei der Code-Generierung.
- QGIS Development Team (o. D.). *QGIS Geographic Information System*. Open-Source-Software. [abgerufen am 2025-12-09]. URL: <https://qgis.org/>.
- rapidlasso GmbH (o. D.[a]). *las2dem README*. Online, [https://downloads.rapidlasso.de/html/las2dem\\_README.html](https://downloads.rapidlasso.de/html/las2dem_README.html). [abgerufen am 2025-12-10].
- (o. D.[b]). *Product Overview*. Online, <https://rapidlasso.de/product-overview/>. [abgerufen am 2025-12-09].
- Sentinel-Hub (o. D.). *Sentinel-3 NDWI Custom Script Documentation*. <https://custom-scripts.sentinel-hub.com/custom-scripts/sentinel-3/ndwi/>. [zuletzt abgerufen am 2025-12-10].
- Statistisches Amt Mecklenburg-Vorpommern (2025). *Bevoelkerungsentwicklung der Kreise und Gemeinden in Mecklenburg-Vorpommern (Faktoren der Bevoelkerungsentwicklung), Kennziffer A113 2024*. Statistische Berichte A. URL: <https://www.laiv-mv.de/Statistik/Veroeffentlichungen/Statistische-Berichte/A/>.
- Wikipedia (o. D.[a]). *Arithmetisches Mittel*. Online, [https://de.wikipedia.org/wiki/Arithmetisches\\_Mittel](https://de.wikipedia.org/wiki/Arithmetisches_Mittel). [abgerufen am 2025-12-23].
- (o. D.[b]). *Mittlerer absoluter Fehler*. Online, [https://de.wikipedia.org/wiki/Mittlerer\\_absoluter\\_Fehler](https://de.wikipedia.org/wiki/Mittlerer_absoluter_Fehler). [abgerufen am 2025-12-23].
- (o. D.[c]). *Root mean square deviation*. Online, [https://en.wikipedia.org/wiki/Root\\_mean\\_square\\_deviation](https://en.wikipedia.org/wiki/Root_mean_square_deviation). [abgerufen am 2025-12-23].

Xu, Hanqiu (2006). „Modification of Normalised Difference Water Index (NDWI) to Enhance Open Water Features in Remotely Sensed Imagery“. In: *International Journal of Remote Sensing* 27.14, S. 3025–3033. DOI: 10.1080/01431160600589179.



## B Python Skripte

### B.1 Ausschnitt aus der iterativen Schattenberechnung

```
1     def processAlgorithm(self, parameters: dict[str, Any],
2       context: QgsProcessingContext, feedback:
3       QgsProcessingFeedback) -> dict[str, Any]:
4         dom_folder = self.parameterAsFile(parameters, self.
5           DOM_FOLDER, context)
6         output_folder = self.parameterAsFile(parameters, self
7           .OUTPUT_FOLDER, context)
8         year = self.parameterAsInt(parameters, self.YEAR,
9           context)
10        month = self.parameterAsInt(parameters, self.MONTH,
11          context)
12        day = self.parameterAsInt(parameters, self.DAY,
13          context)
14        hour = self.parameterAsInt(parameters, self.HOUR,
15          context)
16        minute = self.parameterAsInt(parameters, self.MINUTE,
17          context)
18        second = self.parameterAsInt(parameters, self.SECOND,
19          context)
20        timezone = self.parameterAsInt(parameters, self.
21          TIMEZONE, context)
22        output_merged = self.parameterAsFile(parameters, self
23          .OUTPUT_MERGED, context)
24
25        if not os.path.isdir(dom_folder):
26            raise QgsProcessingException(f"Input folder does
27              not exist: {dom_folder}")
28
29        os.makedirs(output_folder, exist_ok=True)
30
31        tif_files = [f for f in os.listdir(dom_folder) if f.
32          lower().endswith('.tif')]
33        if not tif_files:
34            raise QgsProcessingException(f"No TIF files found
35              in {dom_folder}")
36
37        feedback.pushInfo(f"Found {len(tif_files)} TIF files
38          to process.")
39
40        schatten_ausgaben = []
41        total_steps = len(tif_files)
42
43        for current, tif in enumerate(tif_files):
44            if feedback.isCanceled():
```

```

29         break
30
31     dom_pfad = os.path.join(dom_folder, tif)
32     raster_layer = QgsRasterLayer(dom_pfad, '
33         tmp_raster', 'gdal')
34     if not raster_layer.isValid():
35         feedback.reportError(f"Invalid raster: {tif}"
36             )
37         continue
38
39     # --- FORCE CRS 25833 ---
40     feedback.pushInfo(f"Forcing CRS EPSG:25833 for: {
41         dom_pfad}")
42     dom_25833 = os.path.join(output_folder, f"
43         crs25833_{tif}")
44
45     processing.run(
46         "gdal:warp",
47         {
48             'INPUT': dom_pfad,
49             'SOURCE_CRS': raster_layer.crs().authid()
50             ,
51             'TARGET_CRS': 'EPSG:25833',
52             'RESAMPLING': 0, # nearest neighbour
53             'OUTPUT': dom_25833
54         },
55         context=context,
56         feedback=feedback
57     )
58
59     dom_pfad = dom_25833
60     raster_layer = QgsRasterLayer(dom_pfad, '
61         tmp_raster', 'gdal')
62     # -----
63
64     extent = raster_layer.extent()
65     center_x = (extent.xMinimum() + extent.xMaximum()
66         ) / 2
67     center_y = (extent.yMinimum() + extent.yMaximum()
68         ) / 2
69
70     feedback.pushInfo(
71         f"Processing {current + 1}/{total_steps}: {
72             tif} (Mittelpunkt: {center_x:.2f}, {
73             center_y:.2f})"
74     )
75
76     output_name = os.path.splitext(tif)[0] + '
77         _schatten.tif'

```

```

67         output_pfad = os.path.join(output_folder,
68                                     output_name)
69     try:
70         result = processing.run(
71             "grass7:r.sunmask.datetime",
72             {
73                 'elevation': dom_pfad,
74                 'east': center_x,
75                 'north': center_y,
76                 'year': year,
77                 'month': month,
78                 'day': day,
79                 'hour': hour,
80                 'minute': minute,
81                 'second': second,
82                 'timezone': timezone,
83                 'output': output_pfad,
84             },
85             context=context,
86             feedback=feedback,
87         )
88         schatten_ausgaben.append(result['output'])
89         feedback.setProgress(int((current + 1) /
90                                 total_steps * 100))
91     except Exception as e:
92         feedback.reportError(f"Error processing {tif
93                               }: {str(e)}")
94         continue
95
96 if not schatten_ausgaben:
97     raise QgsProcessingException("No shadow rasters
98     were created successfully.")
99
100 feedback.pushInfo(f"Merging all {len(
101                    schatten_ausgaben)} shadow rasters to {
102                    output_merged} ...")
103
104 try:
105     processing.run(
106         'gdal:merge',
107         {
108             'INPUT': schatten_ausgaben,
109             'OUTPUT': output_merged
110         },
111         context=context,
112         feedback=feedback
113     )
114     feedback.pushInfo(f"Merged raster successfully
115                       saved to {output_merged}")

```

```
109     except Exception as e:
110         feedback.reportError(f"Error during merging: {str(e)}")
111         raise QgsProcessingException(f"Error during merging: {str(e)}")
112
113     return {self.OUTPUT_MERGED: output_merged}
```

Listing 1: Zentraler Verarbeitungsteil der Schattenberechnung

## B.2 Ausschnitt aus der iterativen DGM-Verschmelzung

```
1  def processAlgorithm(  
2      self,  
3      parameters: dict[str, Any],  
4      context: QgsProcessingContext,  
5      feedback: QgsProcessingFeedback,  
6  ) -> dict[str, Any]:  
7      """  
8      Here is where the processing itself takes place.  
9      """  
10  
11     # Retrieve input folder and output file path  
12     input_folder = self.parameterAsFile(parameters, self.  
13         INPUT_FOLDER, context)  
14     output_file = self.parameterAsFile(parameters, self.  
15         OUTPUT_FILE, context)  
16  
17     # If no output file specified, use temporary location  
18     if not output_file:  
19         temp_dir = context.temporaryFolder()  
20         output_file = os.path.join(temp_dir, "  
21             merged_raster.tif")  
22         feedback.pushInfo("No output path specified.  
23             Using temporary location.")  
24  
25     # Validate input folder  
26     if not os.path.isdir(input_folder):  
27         raise QgsProcessingException(  
28             f"Input folder does not exist: {input_folder}"  
29             "  
30             )  
31  
32     # Find all .tif/.tiff files in the folder  
33     tif_files = sorted(  
34         glob.glob(os.path.join(input_folder, "*.tif")) +  
35         glob.glob(os.path.join(input_folder, "*.tiff"))  
36     )  
37  
38     # Check if any TIF files were found  
39     if not tif_files:  
40         raise QgsProcessingException(  
41             f"No GeoTIFF files (.tif/.tiff) found in: {  
42                 input_folder}"  
43             )  
44  
45     feedback.pushInfo(f"Found {len(tif_files)} raster  
46         file(s) to merge")
```

```

41 # Ensure output directory exists
42 output_dir = os.path.dirname(output_file)
43 if output_dir and not os.path.exists(output_dir):
44     os.makedirs(output_dir, exist_ok=True)
45
46 try:
47     feedback.pushInfo("Running gdal:merge algorithm")
48
49     # Use QGIS's built-in gdal:merge algorithm
50     result = processing.run(
51         "gdal:merge",
52         {
53             "INPUT": tif_files,
54             "PCT": False,
55             "SEPARATE": False,
56             "NODATA_INPUT": None,
57             "NODATA_OUTPUT": None,
58             "DATA_TYPE": 5, # Float32
59             "OUTPUT": output_file,
60         },
61         context=context,
62         feedback=feedback,
63         is_child_algorithm=True,
64     )
65
66     output_file = result.get("OUTPUT")
67
68     # Verify output file was created
69     if not output_file or not os.path.exists(
70         output_file):
71         raise QgsProcessingException(
72             "Output file was not created by gdal:
73             merge"
74         )
75
76     file_size_mb = os.path.getsize(output_file) /
77     (1024**2)
78     feedback.pushInfo(f"Successfully merged {len(
79         tif_files)} raster file(s)")
80     feedback.pushInfo(f"Output file size: {
81         file_size_mb:.2f} MB")
82     feedback.pushInfo(f"Output: {output_file}")
83
84 except QgsProcessingException as e:
85     raise QgsProcessingException(
86         f"Merge failed: {str(e)}"
87     )
88 except Exception as e:
89     raise QgsProcessingException(

```

```
85         f"Error during merge process: {str(e)}"  
86     )  
87  
88     # Return the output file path  
89     return {self.OUTPUT_FILE: output_file}
```

Listing 2: Zentraler Verarbeitungsteil der DGM-Verschmelzung