

Zusammenfassung der Arbeiten im Berichtszeitraum 7.11.2018 bis 30.08.2020

**SOIL-DE – Entwicklung von Indikatoren zur Bewertung
der Ertragsfähigkeit, Nutzungsintensität und Vulnerabilität
landwirtschaftlich genutzter Böden in Deutschland**

Arbeitspaket AP 5

Dr. Markus Möller

Halle, 7. Februar 2021

TERRASYS geodatenanalyse (<https://terrasys.github.io>) im Auftrag des Deutschen
Zentrums für Luft- und Raumfahrt (DLR)

Inhaltsverzeichnis

1	Bodenprognose	1
1.1	Prozesskette	1
1.1.1	Vorprozessierung	3
1.1.2	Parametrisierung	3
1.1.3	Maschinelles Lernen	4
1.2	Prognose von Bodenparametern	5
1.2.1	Humusgehalt	7
1.2.2	Bodenart	10
2	Bodenbedeckung und Bodenerosionsmodellierung	16
2.1	Tägliche Starkniederschläge	16
2.2	Phasenspezifische Bodenbedeckung	17
	Literatur	20

Der Fokus des Arbeitspaketes AP 5 (Bodenerosionsmodellierung und Parametrisierung) im Rahmen des Verbundprojektes SOIL-DE liegt in

1. der maßstabsspezifischen Prognose der Bodenparameter *Humusgehalt* und *Bodenart* (Kap. 1) sowie
2. der zeitlich dynamischen Ableitung von schlag- und phasenspezifischen Bodenbedeckungsinformationen und Bodenerosionsmodellierung (Kap. 2).

1 Bodenprognose

Die Arbeiten im Bereich der Bodenprognose umfassen die Entwicklung einer Prozesskette (Kap. 1.1), die eine vergleichende Bewertung des Einflusses erklärender Variablen auf die Prognose von Bodenparametern und -klassen erlaubt (Kap. 1.2). Dabei sollen folgende Fragestellungen untersucht werden:

1. Analyse des Einflusses von Spektralinformationen auf die maßstabsspezifische Prognosegüte der Bodenparameter *Humusgehalt* und *Bodenart* sowie
2. Vergleich von Berechnungsvarianten der Bodenparameter *Humusgehalt* und *Bodenart* in Abhängigkeit von unterschiedlichen Datenquellen.

1.1 Prozesskette

Die inhaltliche Aufbau der Prozesskette ist in Abbildung 1 dargestellt und lässt sich in die folgenden Schritte gliedern:

1. Als bodenkundliche Datengrundlagen (S) dienen Aufschlüsse (Samples) oder Attribute bodenkundlicher Kartenwerke (Maps).
2. Multi-skalare Reliefparameter ergeben sich aus der Anwendung von Reliefanalysealgorithmen (TA) auf digitale Höhenmodelle.
3. Multi-temporale Reflektanzen (SR) sind das Ergebnis des Algorithmus "Soil Composite Mapping Processor" (SCMaP; Rogge et al., 2018), der im Zuge der Bearbeitung des Arbeitspaketes AP 4 zur Anwendung kommt. Die Berechnung der SCSMaP-Daten ist für den Bezugszeitraum 1984 bis 2014 vorgenommen worden. Die resultierenden Komposite haben jeweils 15 Kanäle. Die Kanäle 1-7 repräsentieren die „normalen“ gemittelten Landsatreflektanzen, die Kanäle 8 bis 14 die pro Szene normalisierten gemittelten Landsatreflektanzen.
4. Die Boden-, Relief- und Spektralinformationen werden mit multi-skalaren Bezugseinheiten (RU) durch die Anwendung eines Algorithmus der zonalen Statistik gekoppelt.

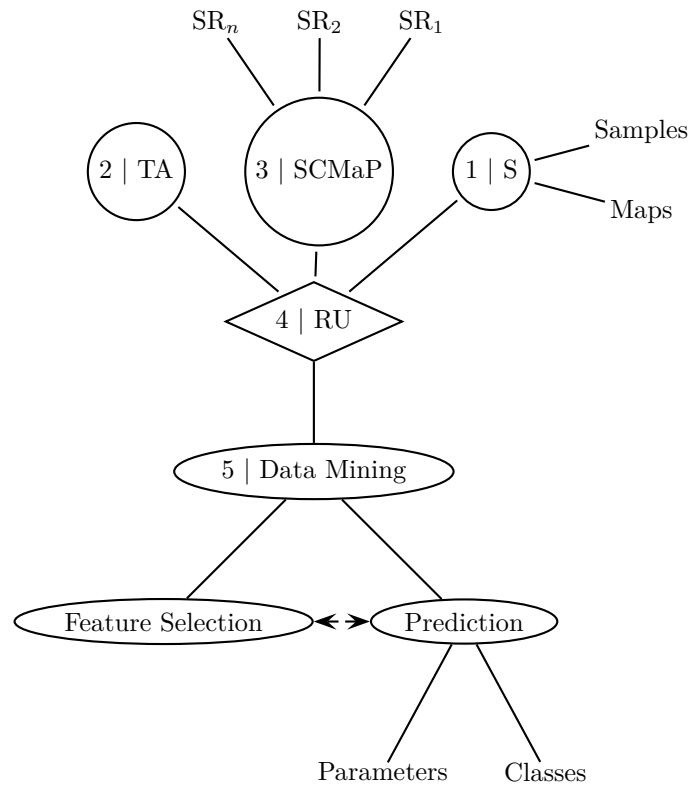


Abbildung 1: Fließschema zur maßstabsspezifischen Prognose von Bodeneigenschaften. SCMaP – Soil Composite Mapping Processor | S – Bodendaten | TA – Reliefanalyse | RU – Bezugseinheit | SR – multi-temporale Spektralreflektanzen.

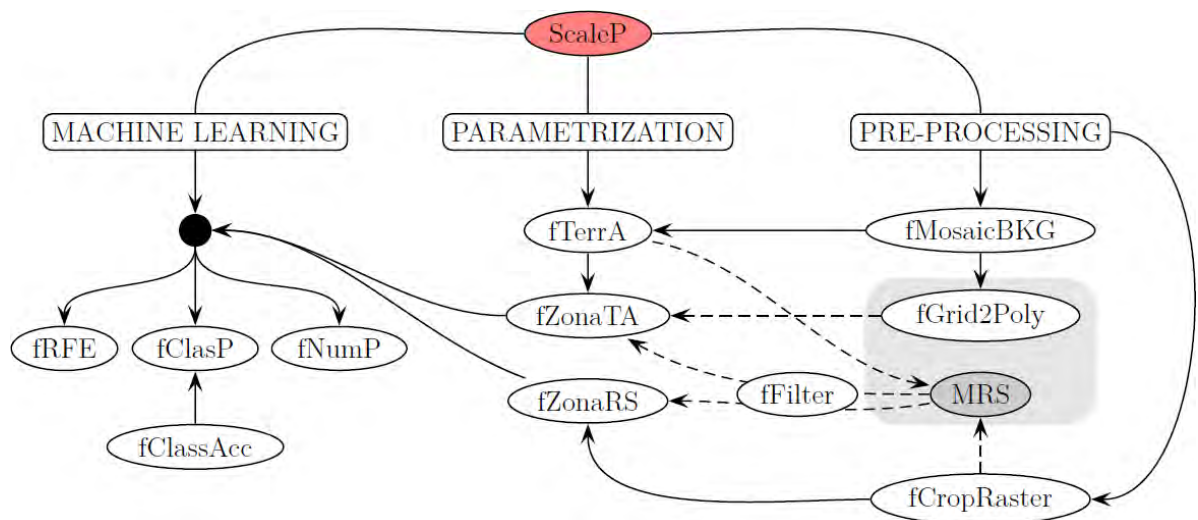


Abbildung 2: Struktur und Abhängigkeiten der Funktionen zur Prognose von numerischen und thematischen Bodenparametern. Eingangsdaten, Funktionsparameter und Ergebnisdaten sind im [ScaleP-Wiki](#) beschrieben.

5. Mit “Data Mining” werden Algorithmen des maschinellen Lernens zusammengefasst, die die Prognose von Bodenklassen und numerischen Parametern sowie die Detektion von optimalen Merkmalskombinationen (Feature Selection) für die Bodenprognose beinhalten.

Alle Schritte der Prozesskette sind – mit Ausnahme von MRS – innerhalb der Programmumgebung **R** in Form von Funktionen umgesetzt ([{R Core Team}, 2020](#)). Die Funktionen sind im GitHub-Repositorium [ScaleP](#) abgelegt und dokumentiert. Abbildung 2 veranschaulicht die Struktur und Abhängigkeiten der Funktionen, die sich in die Kategorien “Vorprozessierung” (Kap. 1.1.1), “Parametrisierung” (Kap. 1.1.2) und “Maschinelles Lernen” (Kap. 1.1.3) einordnen lassen.

1.1.1 Vorprozessierung

Zur Kategorie “Vorprozessierung” gehören die Funktionen der Bearbeitung von Relief- und Stellitenbilddaten. [fMosaicBKG](#) erlaubt die Mosaikierung und Aggregation von Einzeldateien des Digitalen Geländemodells [DGM10](#) des Bundesamtes für Kartographie und Geodäsie (BKG), [fCropRaster](#) das Ausschneiden von ein- oder mehrdimensionalen Rasterdaten und [fGrid2Poly](#) die Aggregation und Vektorisierung von Rasterdaten. MRS steht für den innerhalb der Programmumgebung [eCognition](#) implementierten “multi-resolution segmentation”-Algorithmus ([Benz et al., 2004](#)), mit dem Bezugseinheiten (RU) in Form von multi-hierarchischen Vektordatensätzen generiert werden können ([Möller et al., 2008, 2012](#); [Möller & Volk, 2015](#)). Die Funktion [ffilter](#) erlaubt die Filterung der Segmentierungsergebnisse in Abhängigkeit von parameterspezifischen Schwellenwerten.

1.1.2 Parametrisierung

Bei der Parametrisierung werden zum einen bodenkundlich relevante Reliefattributen mit der Funktion [fTerra](#) abgeleitet, die verschiedene Aggregations- bzw. Maßstabsniveaus des Reliefs repräsentieren (Tab. 1). Das betrifft insbesondere Varianten der Reliefattribute NH und TPI, der Berechnung mit verschiedenen “Moving Window”-Größen vorgenommen worden ist. Die Berechnungsvarianten der Reliefattribute VDC und TCI basieren auf verschiedenen Aggregationsniveaus des Tiefenliniennetzes. Die MBI-Versionen sind schließlich Ausdruck von Varianten der Differenzierbarkeit von dominanten und subdominanten Reliefformen. Alle Algorithmen der Reliefattributableitung sind in der [fTerra](#) zusammengefasst. Diese enthält in erster Linie Funktionen des **R**-Paketes `RSAGA` ([Brenning et al., 2018](#)), worüber Funktionalitäten der Reliefanalysesoftware SAGA-GIS angesprochen werden können ([Conrad et al., 2015](#)).

Tabelle 1: Reliefattribute

Abkürzung	Bedeutung	Quelle
FILL	Digitales Höhenmodell (= <i>Digital Elevation Model</i>) mit gefüllten Senken	Planchon & Darboux (2002)
SLP	Hangneigung	Zevenbergen & Thorne (1987)
CA	Fließakkumulation	Quinn et al. (1991)
VDC	Vertical Distance above Channel network	Conrad et al. (2015)
TCI	Terrain Classification Index (TCI)	Conrad et al. (2015)
TWI	Topographic Wetness Index	Beven & Kirkby (1979)
MBI	Mass Balance Index	Möller et al. (2008, 2012) ; Möller & Volk (2015)
TOP	Topographic (positive) Openness	Yokoyama et al. (2002)
TON	Topographic (negative) Openness	Yokoyama et al. (2002)
NH	Relative Hangposition	Conrad et al. (2015)
TPI	Topographic Position Index	Guisan et al. (1999)

Die eigentliche Parametrisierung der Bezugseinheiten mit (eindimensionalen) Reliefattributen bzw. mehrdimensionalen SCMaP-Rasterdaten wird mit den Funktionen [fZonaTA](#) und [fZonaRS](#) umgesetzt, über die zonale Statistikalgorithmen angesprochen werden.

1.1.3 Maschinelles Lernen

In der Prozesskette des maschinellen Lernens sind drei Funktionen vereint. Mit der Funktion [fRFE](#) können optimale Kombinationen erklärender Variablen bei der Prognose der Bodenparameter durch die Anwendung sogenannten [Recursive Feature Elimination](#)-Prozedur abgeleitet werden. Die Prognose von numerischen und thematischen Bodenparametern wird durch die Funktionen [fNumP](#) und [ClasP](#) realisiert, über die verschiedene Data Mining-Algorithmen angesprochen werden können¹. Im Rahmen der Arbeiten des Arbeitspaketes AP 5 kam der im bodenkundlichen Kontext weit verbreitete und etablier-

¹<https://topepo.github.io/caret/train-models-by-tag.html>

te Entscheidungsbaumalgorithmus “Random Forest” zur Anwendung (Möller & Volk, 2015; Behrens et al., 2018; Taghizadeh-Mehrjardi et al., 2020). Das Verfahren teilt unter optionaler Berücksichtigung von thematischen Klassen den n-dimensionalen Merkmalsraum von erklärenden Variablen solange, bis der höchste statistische Zusammenhang bei minimaler Varianz erreicht wird (Breiman, 2001).

Zur Validierung der Klassifikationsergebnisse wird der Gesamtdatensatz unter Berücksichtigung der Zielklassen- bzw. Zielparameterverteilung in einen Trainings- und Testdatensatz von 75 % bzw. 25 % geteilt. Die Modellbildung basiert auf dem Trainingsdatensatz. Auf dessen Grundlage wird auch eine Kreuzvalidierung durchgeführt, aus der sich die Gesamtgenauigkeit des Modells ergibt. Der Testdatensatz dient der unabhängigen Validierung (Khaledian & Miller, 2020), auf den das trainierte Modell angewendet wird. Wie bei der in Möller & Volk (2015) beschriebenen Vorgehensweise dienen zur Bewertung der numerischen Parameter das Bestimmtheitsmaß R^2 und das Streuungsmaß RMSE (= Root Mean Square Error). Die Beurteilung der prognostizierten Klassen wird mithilfe der klassenbezogenen F1-Werte und der Gesamtgenauigkeit vorgenommen (vgl. Gerstmann et al., 2016b), die in der Funktion `fClasAcc` auf der Grundlage von Konfusionsmatrizen berechnet werden (Stehman, 1997). Der F1-Wert stellt das gewichtete harmonische Mittel aus Genauigkeit (“precision”) und Trefferquote (“recall”) einer jeden Klasse dar (Manning et al., 2008): “precision” beschreibt das Verhältnis richtig klassifizierter Fälle zur Gesamtzahl der simulierten Fälle einer Klasse. “recall” kennzeichnet das Verhältnis richtig klassifizierter Fälle zur Gesamtzahl der tatsächlichen Fälle einer Klasse. Die Gesamtgenauigkeit berechnet sich wiederum aus dem Verhältnis aller korrekten Treffer und der Gesamtzahl der Fälle.

1.2 Prognose von Bodenparametern

Die Prozesskette zur Prognose von Bodenparametern wird am Beispiel eines bayrischen Testgebietes und den Betrachtungszeitraum 1984 bis 2014 veranschaulicht, für das Bodendaten der Bayerischen Landesanstalt für Landwirtschaft (LfL) vorliegen (Abb. 3). Dabei handelt es sich um Dauerbeobachtungsflächen, deren Einzelmessungen für den Betrachtungszeitraum durch den Auftraggeber gemittelt worden sind.

Abbildung 6 zeigt Beispiele erklärender Variablen. Dabei handelt es sich um den SCMaP-Datensatz (Kanalkombination 4-5-3), der für den Zeitraum 1984 bis 2014 berechnet worden ist (Rogge et al., 2018), sowie das Reliefattribut TWI (vgl. Tab. 1). Für das multihierarchische Attribut NH sind zwei Varianten dargestellt, die unterschiedliche Aggregierungsstufen repräsentieren. Durch die Anwendung des MRS-Segmentierungsalgorithmus auf die Reliefattribute NH, MBI, SLP und TWI sowie den SCMaP-Datensatz sind

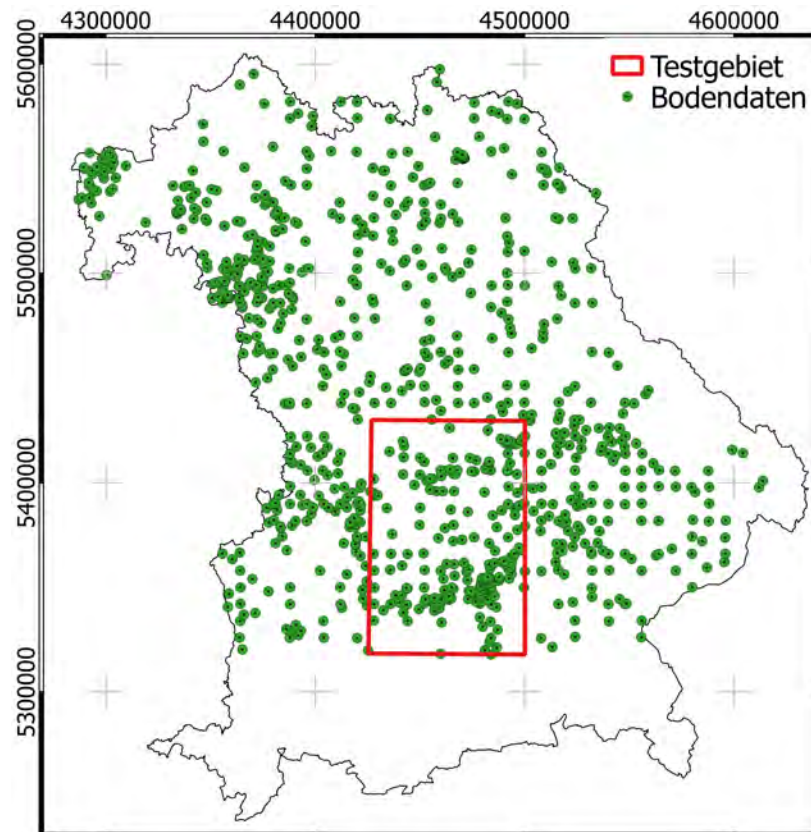


Abbildung 3: Lage des bayrischen Testgebietes und der Bodendaten der Bayerischen Landesanstalt für Landwirtschaft (Projektion: [EPSG 31468](#)).

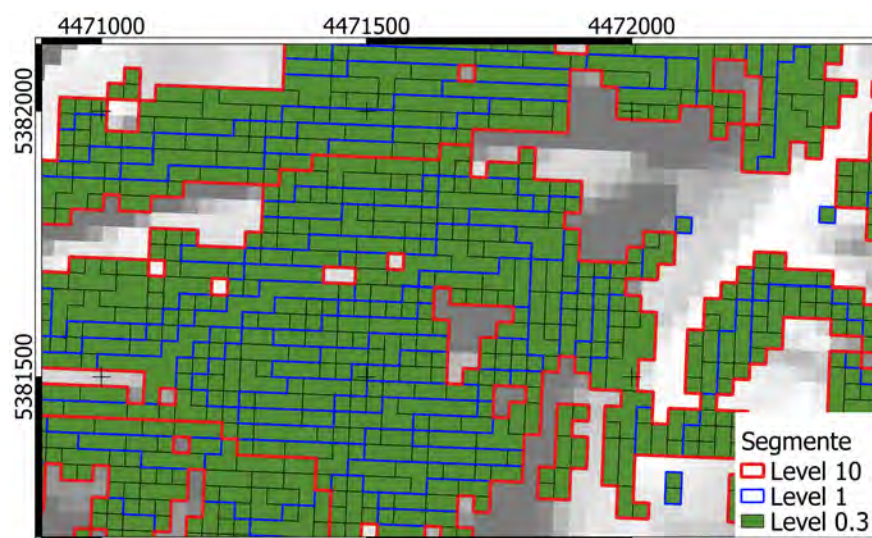


Abbildung 4: Visualisierung von drei Segmentierungsstufen in einem Testgebietsausschnitt (Projektion: [EPSG 31468](#)).

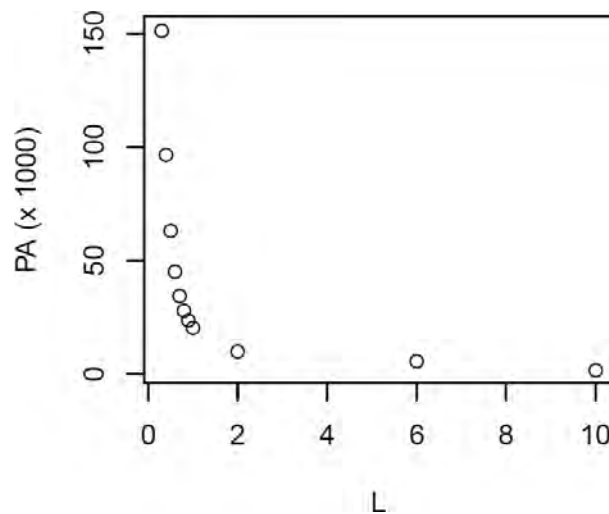


Abbildung 5: Zusammenhang zwischen Polygonanzahl (PA) und Segmentierungsebene bzw. *Scale Level* (L) für die Segmentierungsergebnisse des bayrischen Testgebietes.

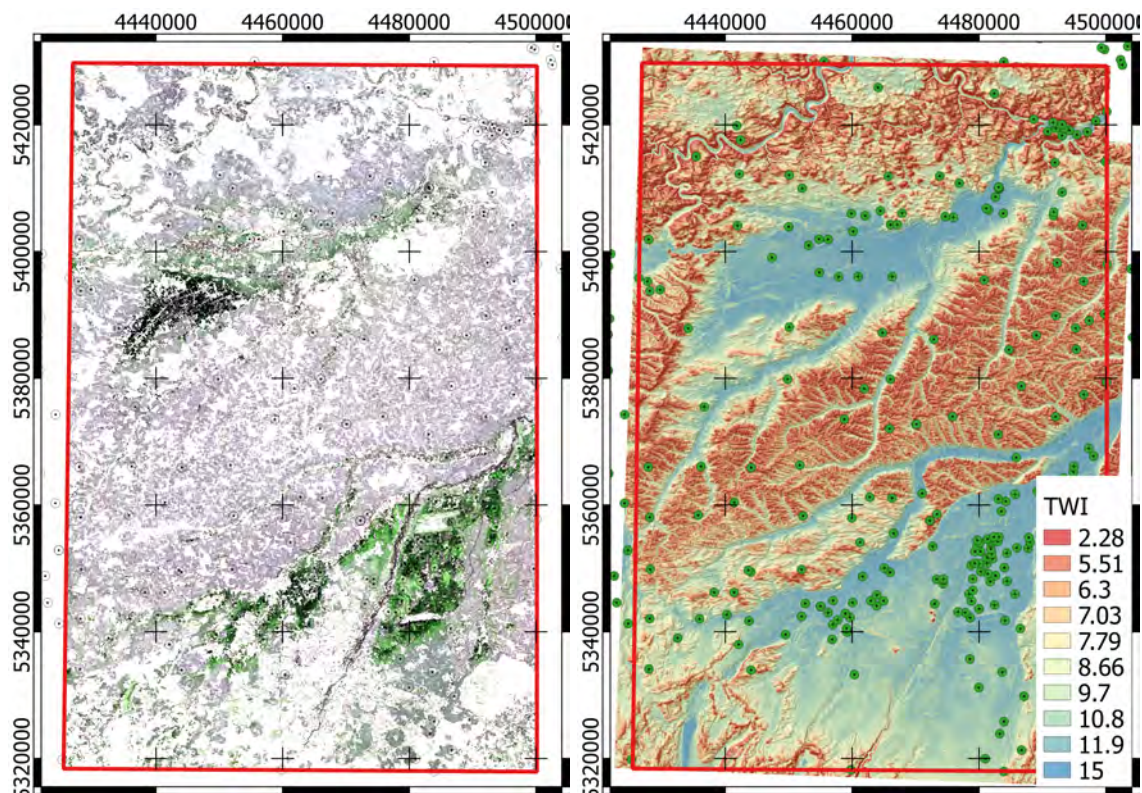
Segmentierungsebenen in Form von Vektordatensätzen abgeleitet worden. Abbildung 4 veranschaulicht am Beispiel eines Testgebietsausschnittes drei Segmentierungsvarianten. Danach sind die “Level 0.3”-Polygone Teilflächen der “Level 1”-Polygone, die wiederum die “Level 10”-Polygone untergliedern. Die Benennung der Segmentierungsebenen folgt der Angabe des Parameters *Scale Level* L , der im Zuge der Anwendung des MRS-Algorithmus variiert wurde und die Merkmalsheterogenität und damit indirekt die Objektgrößen der resultierenden Segmente bzw. Polygone steuert (Abb. 5; vgl. Möller & Volk, 2015).

In den folgenden Abschnitten werden Ergebnisse der Funktionen `fNumP`, `ClasP` und `fRFE` dargestellt und für die Zielparameter *Humusgehalt* (Masse-%), *Bodenart* und *Humusgehaltsklasse* anhand ausgewählter Segmentierungsebenen erläutert.

1.2.1 Humusgehalt

Abbildung 7 veranschaulicht die Parameterwerteverteilung des Trainings- und Testdatensatzes in Form einer Dichte- und Summenfunktion. Danach ist der Großteil der Werte im Bereich zwischen 0 und 5 Masse-% zu finden. Der Kolomogorov Smirnow- bzw. KS-Test zeigt, dass die Werteverteilungen des Trainings- und Testdatensatzes durch ein hohes Maß an statistischer Ähnlichkeit gekennzeichnet sind (vgl. Möller et al., 2012, 2013).

Die Anwendung der trainierten Modelle auf die Testdatensätze resultiert in maßstabsspezifischen Bestimmtheitsmaßen der Humusgehaltsprognosen für alle Segmentierungs-



(a) SCMaP (1984-2014), Kanalkombination 4-5-3

(b) TWI

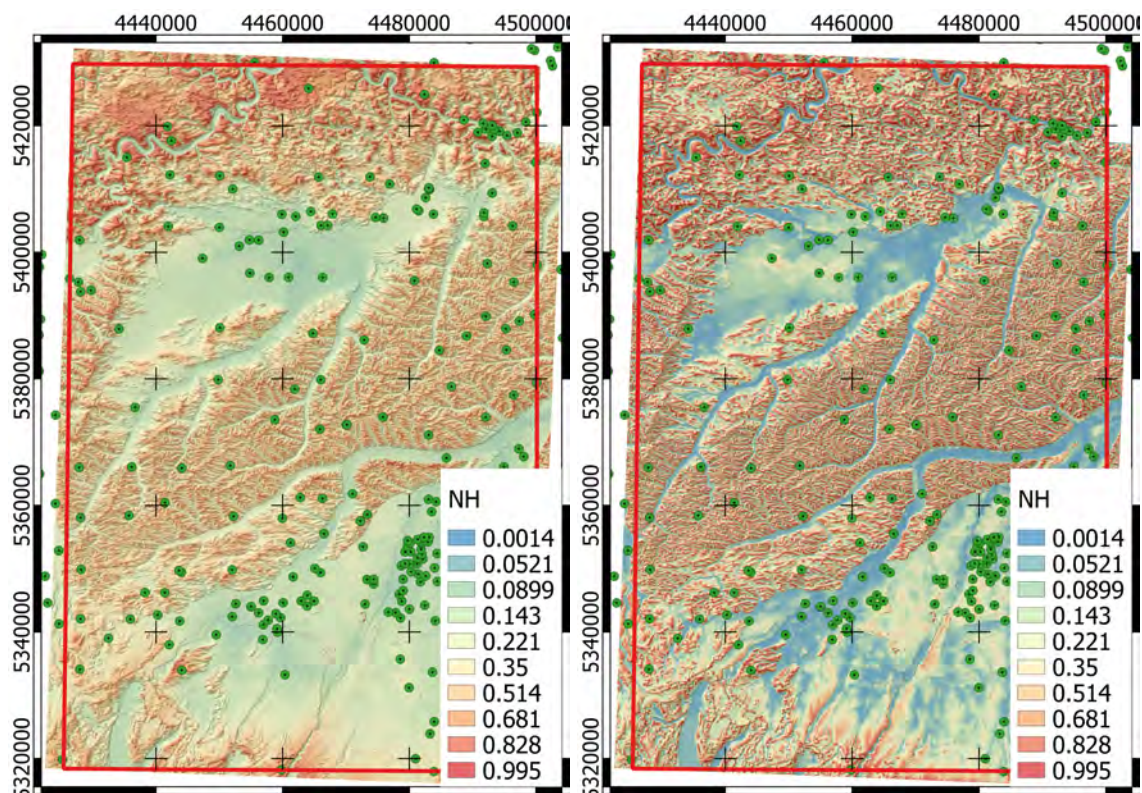
(c) NH mit $T = 2$ (siehe Funktion `fTerrA`)(d) NH mit $T = 1000$ (siehe Funktion `fTerrA`)

Abbildung 6: Beispielhafte Visualisierung erklärender Variablen (Projektion: [EPSG 31468](#)).

Tabelle 2: Maßstabsspezifische Bestimmtheitsmaße (R^2) bzw. Gesamtgenauigkeiten (OA) für die Prognose des Humusgehaltes [Masse-%] bzw. der Bodenarten auf der Grundlage von Reliefattributen (A), SCMaP-Daten (B) und beiden Datengrundlagen (C). PA – Polygonanzahl | TA – Reliefattribute | SM – SCMaP-Daten.

L	PA	R_A^2	R_B^2	R_C^2	OA_A	OA_B	OA_C
0,3	1515513	0,57	0,65	0,67	0,33	0,10	0,31
0,4	965994	0,54	0,82	0,82	0,27	0,10	0,22
0,5	631794	0,56	0,80	0,83	0,29	0,20	0,34
0,6	450939	0,52	0,82	0,85	0,27	0,15	0,27
0,7	345398	0,51	0,78	0,79	0,32	0,17	0,29
0,8	280084	0,53	0,77	0,81	0,34	0,20	0,27
0,9	235484	0,51	0,74	0,74	0,34	0,22	0,24
1	203379	0,47	0,79	0,77	0,32	0,17	0,29
2	98630	0,57	0,72	0,73	0,22	0,24	0,24
3	74575	0,44	0,58	0,61	0,29	0,17	0,24
4	65018	0,44	0,45	0,45	0,22	0,24	0,27
5	60338	0,53	0,55	0,56	0,29	0,20	0,24
6	57718	0,63	0,63	0,64	0,24	0,20	0,29
7	56113	0,61	0,58	0,63	0,24	0,22	0,20
8	55011	0,58	0,54	0,61	0,27	0,20	0,27
9	54306	0,58	0,53	0,51	0,27	0,20	0,24
10	18183	0,52	0,41	0,47	0,20	0,15	0,28

Tabelle 3: Normierte und sortierte Bedeutung (R = Rang) der besten 20 erklärenden Variablen auf das Prognoseergebnis (VI) für die Variante $L = 0,6$ und $TA + SM$ (vgl. Tab. 2 u. Abb. 8) des Zielparameters *Humusgehalt*.

	SM12	SM2	SM10	SM9	SM3	SM4	SLP	SM13	VDC_TH129155	SM5	VDC_TH359381	VDC_TH599484	SM8	VDC_TH46416	TOP	SM1	SM11	NH4	DEM	SM15
VI	100	93	58	49	46	39	38	38	35	30	29	28	28	26	26	26	24	23	22	20

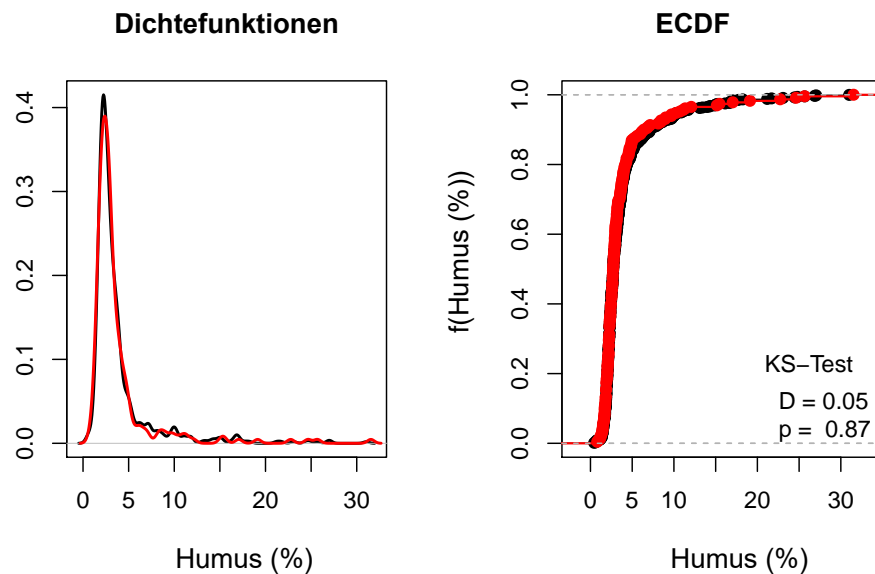


Abbildung 7: Dichte- und Summenfunktionen (ECDF) des Zielparameters *Humusgehalt* [Masse %] für den Trainings- und Testdatensatz. KS: Kolmogorov Smirnov-Test | D: Distanz zwischen Summenfunktionen | p: statistische Signifikanz.

ebenen sowie die drei Variablenkombinationen, die in Tabelle 2 zusammengefasst sind. Danach wird die Prognosegüte sowohl vom Aggregationsniveau als auch von der Variablenkombination beeinflusst. Insgesamt führt – im Vergleich zu den Reliefattributen – die Verwendung von SCMaP-Daten zu einer deutlichen Verbesserung der Prognosegüte. Die besten Ergebnisse werden bei der kombinierten Verwendung von Reliefattributen und SCMaP-Daten erzielt. Das betrifft insbesondere die Segmentierungsvariante $L = 0,6$ mit $R^2 = 0,85$, wofür in Abbildung 8 die Prognosegenauigkeiten bezogen auf den Trainings- und Testdatensatz dargestellt sind.

Der Einfluss der Variablen auf das Prognoseergebnis ist in Tabelle 3 dokumentiert. Dort sind die 20 Variablen mit der größten Bedeutung aufgelistet. Den größten Einfluss haben demnach die SCMaP-Kanäle. Das betrifft vor allem die Kanäle SM12 und SM2. Als wichtigste Reliefparameter haben sich die Neigung sowie verschiedene VDC-Varianten erwiesen. Zusammen mit Tabelle 3 zeigt schließlich Abbildung 9, dass die Kombination der vier Variablen SM12, SM2, SM10 und SM3 zu einer Prognose mit minimalen RMSE-Werten führt.

1.2.2 Bodenart

In Abbildung 11a und b sind die Anteile der Bodenartenklassen an der Grundgesamtheit des Trainings- (696 Stichproben) und Testdatensatzes (218 Stichproben) dargestellt. Bei-

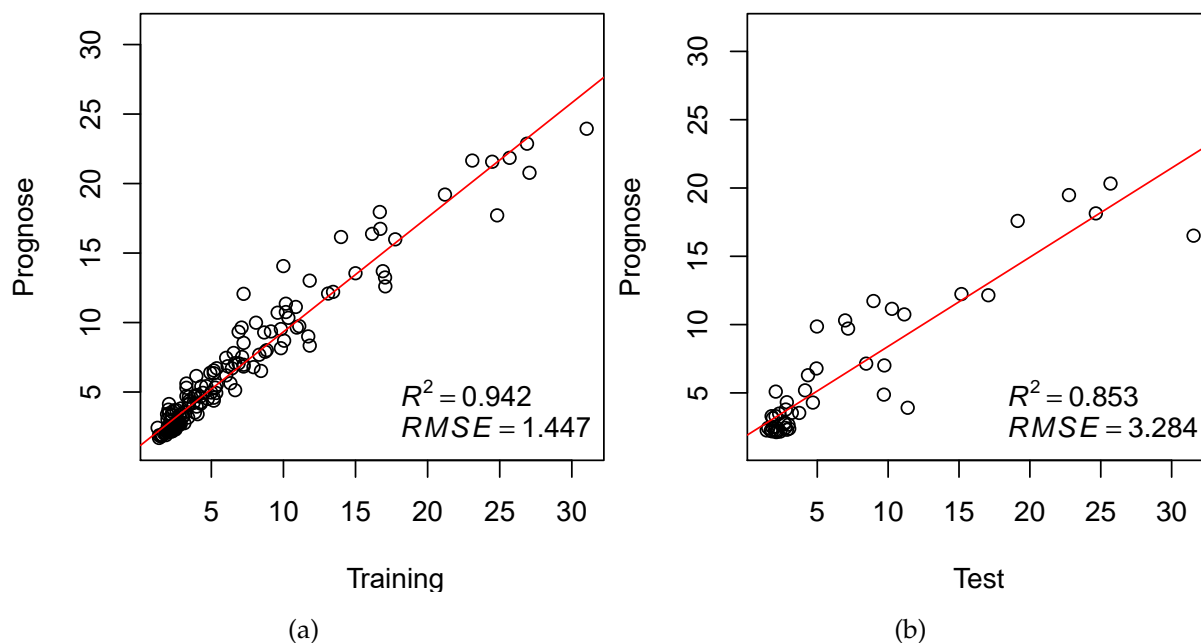


Abbildung 8: Streudiagramm der Prognose- und Trainings- bzw. Testdaten des Zielparameters *Humusgehalt* für die Segmentierungsebene $L = 0,6$ und $TA + SM$.

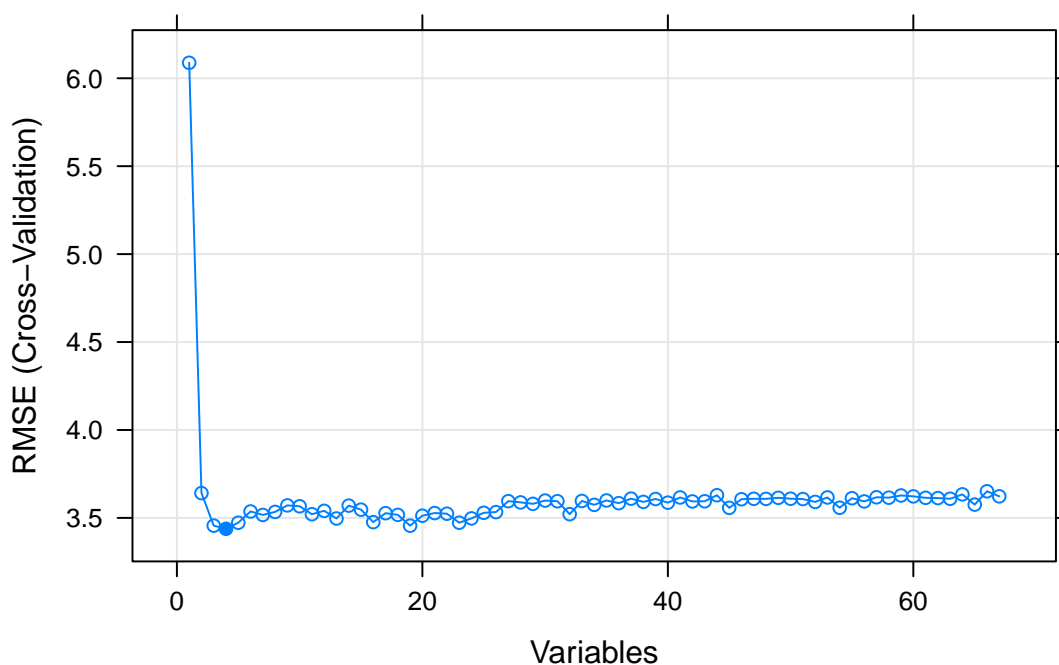


Abbildung 9: RMSE-Werte für die Humusgehaltsprognosevariante $L = 0,6$ und $TA + SM$ in Abhängigkeit von Variablenkombinationen.

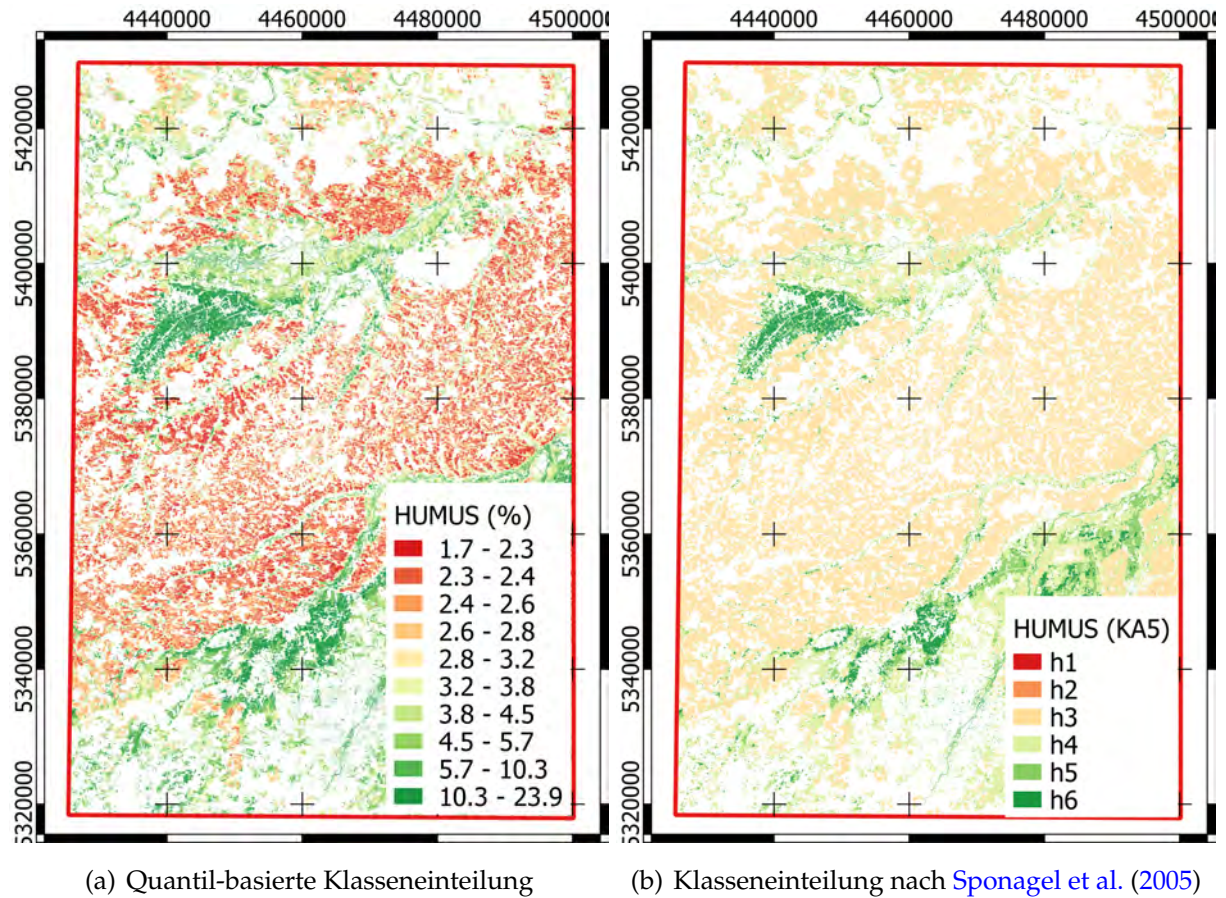


Abbildung 10: Visualisierung der Prognosevariante $L = 0,6$ und $TA + SM$ (vgl. Tab. 2 u. Abb. 8) des Zielparameters *Humusgehalt*.

de Datensätze bilden die Grundlage für die Prognose der Bodenart, deren Ergebnisse als Gesamtgenauigkeiten (OA) in Tabelle 2 zusammengefasst sind und insgesamt als gering eingeschätzt werden können. Die Ergebnisdateien werden am Beispiel der Prognosevariante $L = 0,34$ und $TA+SM$ mit $OA = 0,34$ erläutert. Die Konfusionsmatrix und die darauf basierenden F1-Werte sind in Tabelle 4 dokumentiert. Lediglich die vier Klassen $Lt2$, Lu , $Tu4$ und $Ut3$ werden durch F1-Werte charakterisiert. In Abbildung 11c können die Anteile der prognostizierten Bodenartenklassen mit den Anteilen der Trainingsdaten verglichen werden. Dabei fällt auf, dass einige Klassen nicht prognostiziert werden, was mit dem unbalanzierten Trainingsdatensatz im Zusammenhang steht (Taghizadeh-Mehrjardi et al., 2020). Die Variablenbedeutung für jede Klasse und der Mittelwert für alle Klassen (VI) ist in Tabelle 5 zusammengefasst. In Abbildung 12 wird das Prognoseergebnis schließlich als Karte dargestellt.

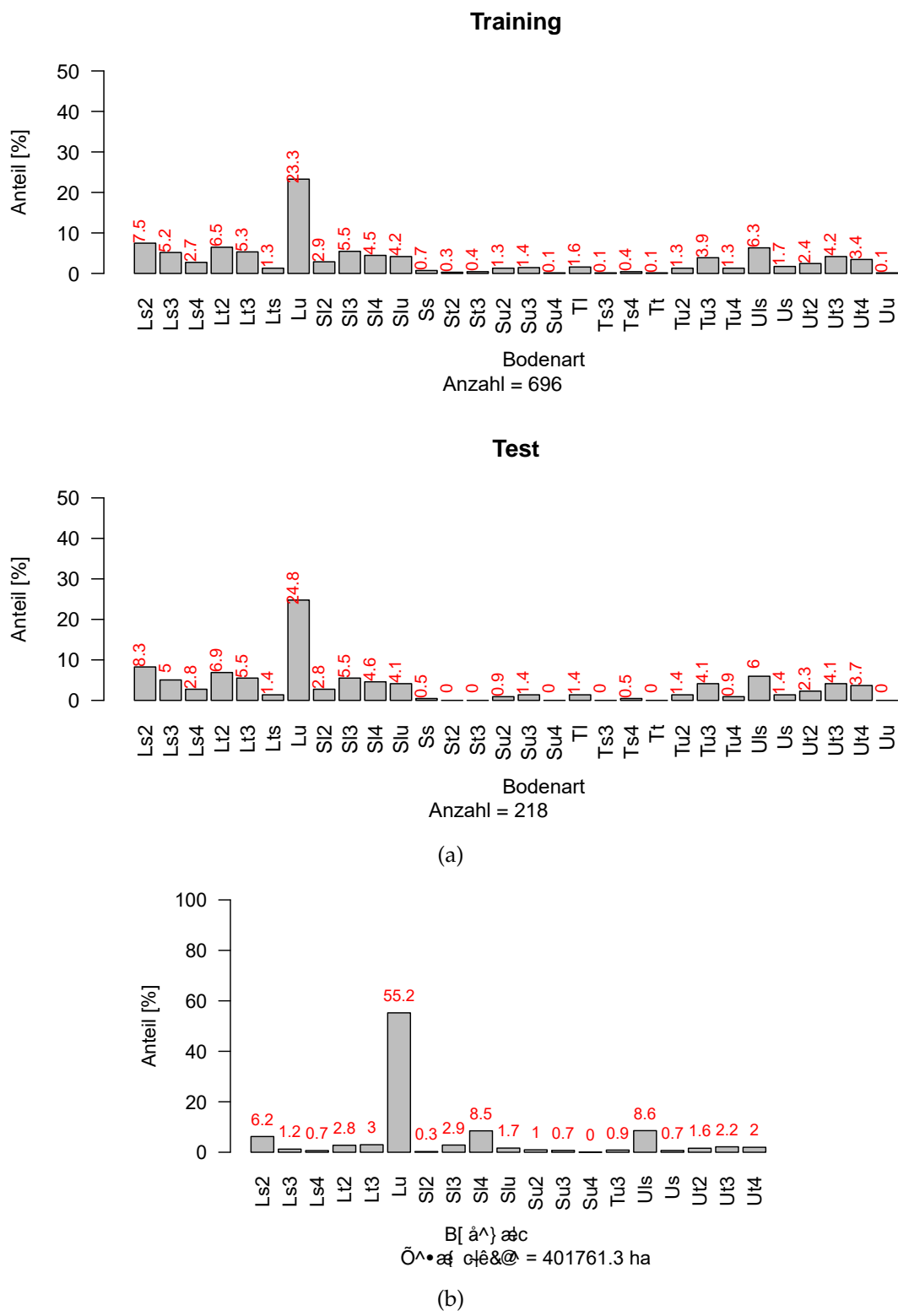


Abbildung 11: Anteile der Bodenartenklassen im Trainings- und Testdatensatz (a) sowie Prognoseergebnis (b; vgl. Tab. tab:VI-Humusgehalt-L6-TASM).

Tabelle 4: Konfusionsmatrix, F1-Werte und Gesamtgenauigkeit für die Bodenart-Prognose (Variante $L = 0,5$ und $TA + SM$; vgl. Tab. ??).

	Ls2	Ls3	Ls4	Lt2	Lt3	Lts	Lu	Sl2	Sl3	Sl4	Slu	Su3	Tu4	Uls	Ut2	Ut3	Ut4	Σ
Ls2	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	2
Ls3	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2
Ls4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lt2	0	0	0	1	0	0	3	0	0	0	0	0	0	0	1	0	0	5
Lt3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
Lts	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
Lu	0	0	0	1	0	0	5	0	0	0	0	0	0	0	0	0	0	6
Sl2	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	3
Sl3	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	2
Sl4	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
Slu	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	3
Su3	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
Tu4	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
Uls	0	0	0	0	0	0	1	0	0	0	1	1	0	6	0	0	0	9
Ut2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
Ut3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
Ut4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2
Σ	1	0	1	3	1	0	16	0	1	1	1	2	0	10	3	1	0	
F1	NA	NA	NA	0,25	1,00	NA	0,45	NA	NA	NA	NA	NA	NA	0,63	NA	1,00	NA	

Tabelle 5: Normierte und sortierte Bedeutung (R = Rang) der besten 20 erklärenden Variablen auf das Prognoseergebnis (VI) für die Variante $L = 0,5$ und $TA + SM$ (vgl. Tab. 2) des Zielparameters *Bodenart*.

	SM9	SM10	SM12	DEM	TOP	MBI6	SM11	TCL_TH10000	SM14	TCL_TH16681	TWI	NH2	SLP	TCL_TH359381	TCL_TH46416	MBI1	TCL_TH215443	MBI2	SM6	MBI7
Ls2	99	92	98	95	89	79	89	91	99	85	83	90	95	77	86	81	77	73	95	78
Ls3	82	85	81	81	76	79	79	73	79	72	72	67	68	74	68	76	77	73	73	72
Ls4	71	66	68	82	73	73	68	79	77	69	71	66	70	68	72	67	67	57	58	68
Lt2	86	95	77	84	84	84	76	89	77	82	78	76	81	78	74	86	74	81	77	78
Lt3	18	12	8	12	19	23	12	21	23	28	16	19	3	23	22	23	15	34	5	36
Lu	42	49	18	20	18	19	42	32	15	46	45	53	14	52	35	15	49	27	11	20
Sl2	74	68	67	65	71	61	66	65	67	67	56	58	54	62	61	62	54	58	64	57
Sl3	86	82	100	94	88	83	89	74	76	81	89	75	74	76	73	72	75	78	77	78
Sl4	94	85	95	92	91	90	81	81	89	83	85	79	89	75	74	96	82	84	83	87
Slu	80	87	86	84	78	85	76	83	81	87	92	87	91	86	91	87	85	89	86	78
Su2	71	71	69	69	67	65	73	58	62	65	60	69	66	62	64	60	60	56	68	57
Su3	72	68	74	76	74	75	73	63	70	64	70	64	72	70	66	63	69	62	84	71
Su4	47	48	37	39	45	43	35	34	37	37	37	36	50	44	35	38	39	38	35	36
Tu3	65	59	53	58	63	62	54	58	59	49	57	58	53	52	54	56	55	50	53	57
Uls	83	84	90	86	82	74	85	83	79	96	88	88	78	81	82	81	83	81	75	81
Us	60	63	59	57	52	61	61	63	49	52	54	58	58	54	53	55	57	60	59	57
Ut2	78	78	99	79	80	84	72	84	81	79	83	80	99	84	84	81	79	88	79	75
Ut3	82	84	89	85	84	77	80	79	81	67	73	75	85	76	78	77	75	80	79	76
Ut4	75	71	77	71	75	72	68	67	70	61	60	66	63	62	66	62	59	61	67	60
VI	72	71	71	70	69	68	67	67	67	67	67	67	67	66	65	65	65	65	65	64

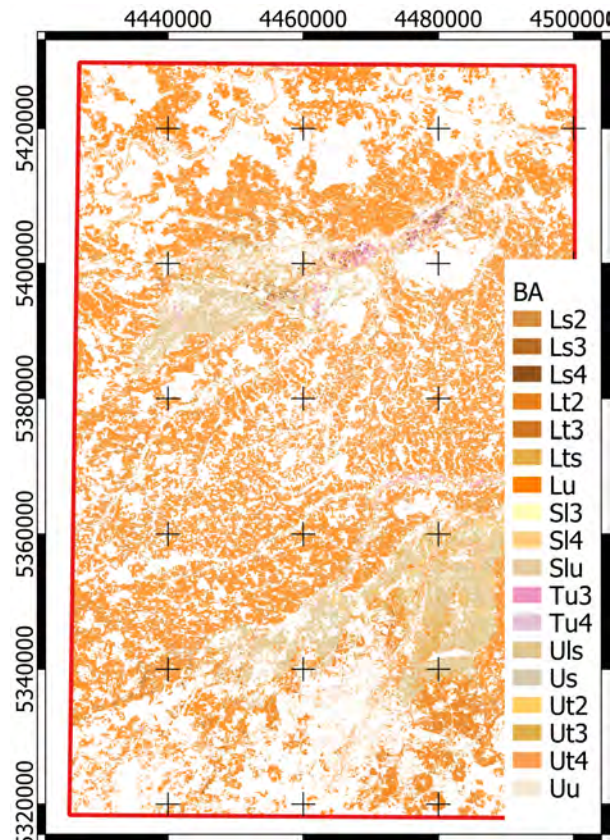


Abbildung 12: Visualisierung der Prognosevariante $L = 0,5$ und $TA + SM$ (vgl. Tab. 2) des Zielparameters *Bodenart*.

2 Bodenbedeckung und Bodenerosionsmodellierung

Bodenerosion ist das Ergebnis einer Störung des Faktorengleichgewichtes zwischen Niederschlag, Relief, Boden und Bodenbedeckung, wobei zeitlich stabile (Bodenart, Relief) und variable Faktoren (Niederschlag, Bodenbedeckung) unterschieden werden können (Volk et al., 2010). Die folgenden Abschnitte beschreiben die Ableitung der variablen Faktoren "Starkniederschlag" und "phasen-spezifische Bodenbedeckung".

2.1 Tägliche Starkniederschläge

Zur Charakterisierung von Starkniederschlagssituationen wird der Starkniederschlagsindex PI verwendet, der sich aus Bodenradarmessungen des Deutschen Wetterdienstes (DWD) ableitet². Die sogenannten RADOLAN-Daten zeichnen sich durch eine räumli-

²ftp://opendata.dwd.de/climate_environment/CDC/grids_germany/5_minutes/radolan

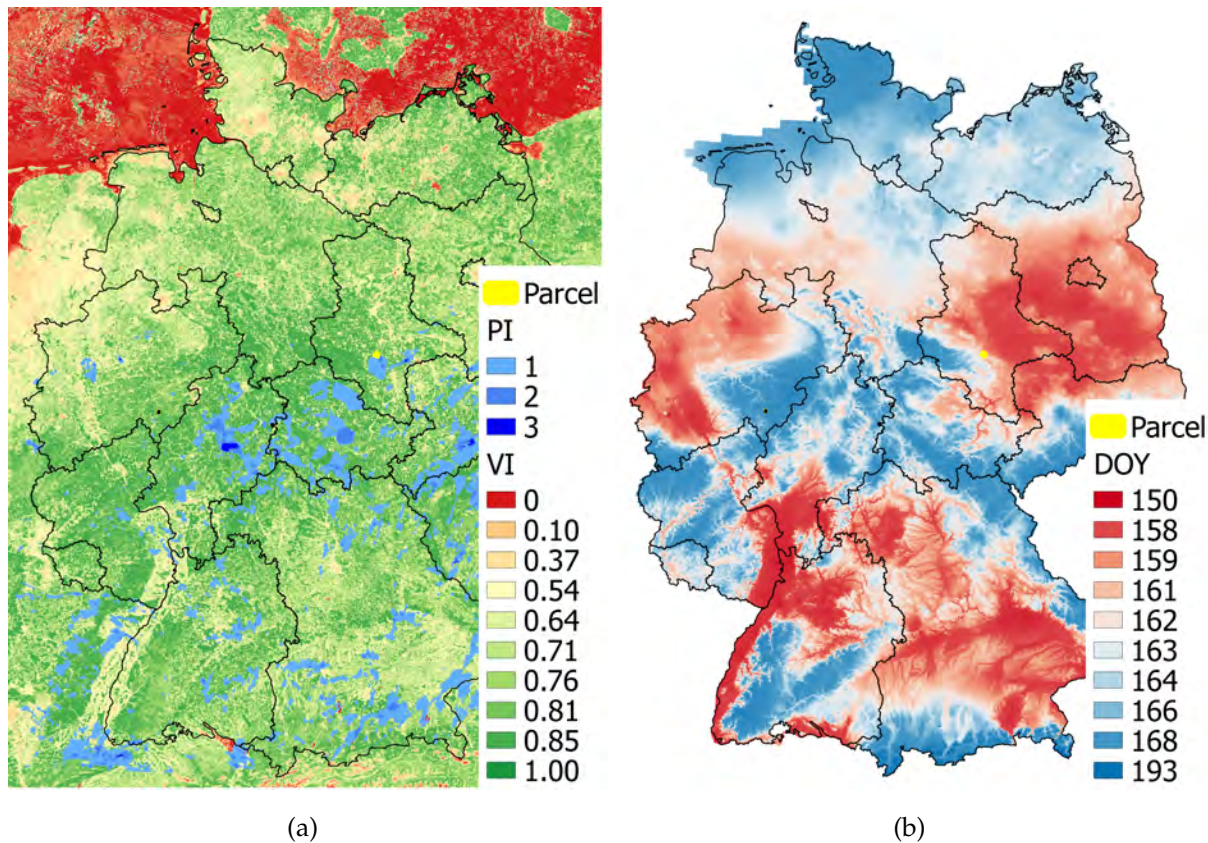


Abbildung 13: Starkniederschlags- (PI) und Vegetationsindex (VI) für den Kalendertag 160 (a) sowie interpolierte phänologische Mais-Phase "Schossen" des Jahres 2013.

che Auflösung von 1 km^2 und eine zeitliche Auflösung von 5 min aus. Der Starkniederschlagsindikator *SNI* berechnet sich aus der Anzahl der Stunden pro Tag mit Niederschlagssummen größer 10 mm, der in Abbildung 13a für Kalendertag 160 des Jahres 2013 dargestellt ist.

2.2 Phasenspezifische Bodenbedeckung

Flächendatensätze phänologischer Phasen leiten sich aus der räumlichen Interpolation von historischen und aktuellen phänologischen Beobachtungen (Gerstmann et al., 2016a; Möller et al., 2020) ab, die im Rahmen des langjährigen phänologischen Beobachtungsprogrammes des Deutschen Wetterdienstes (DWD) erhoben werden (Kaspar et al., 2015). Abbildung 13b zeigt für Mais am Beispiel des Jahres 2013 die interpolierten Starttermine der Phase "Schossen".

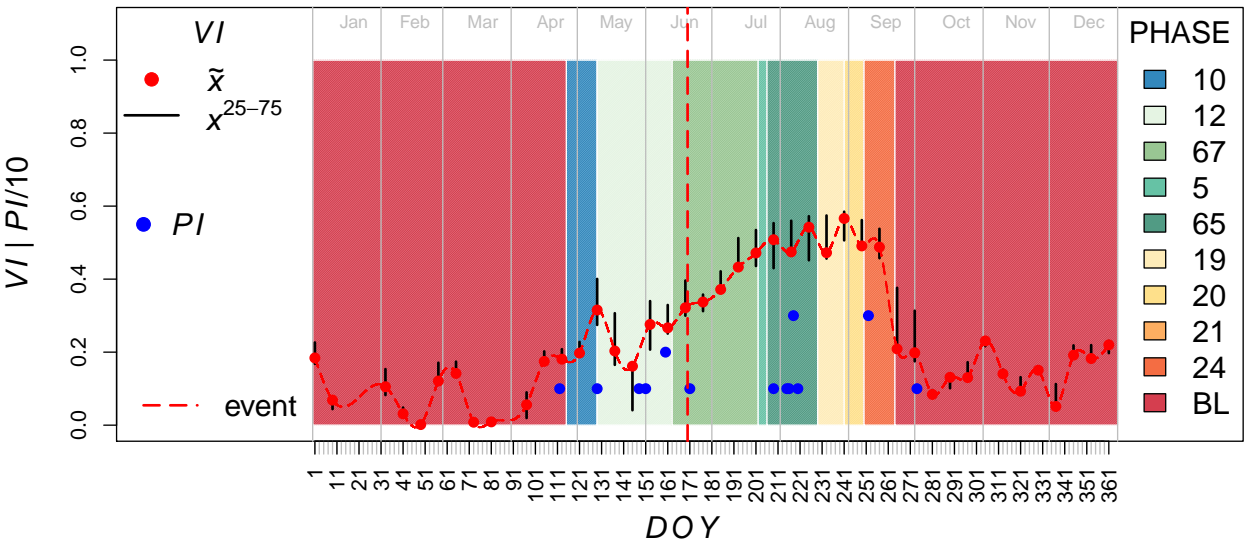


Abbildung 14: Schlagspezifische phänologische Fenster der Ackerfrucht Mais für PI/VI-Zeitserien des Jahres 2013 (Tab. 7).

Tabelle 6: Beziehung zwischen phänologischen Phasen und Bodenbedeckungen (BB; nach Schmidt et al., 1996, vgl. Tab. 7, SB – Saatbett, B – Brache).

PHASEN	5	10	12	14	15	17	18	19	20	21	24	65	67	SB	B
BB (WW)	NA	0	7	NA	60	NA	87	93	NA	93	84	NA	NA	NA	50
BB (WR)	46	3	8	79	NA	95	NA	NA	NA	NA	95	84	NA	94	30
BB (MS)	82	0	2	NA	NA	NA	NA	82	85	85	85	60	12	0	30

Vegetationsindizes können während vitaler phänologischer Phasen als Indikator der Bodendeckung angesehen werden (Möller et al., 2017). Abbildung 14 zeigt ein Kopp- lungsergebnis von Starkniederschlags- und Vegetationsindex für einen in Abbildung 14 lokalisierten Beispielschlag. Danach korrespondieren niedrige VI-Werte während der Brache. Die VI-Werte steigen bis zur Phase “Ährenschieben”, um dann mit zunehmender Reifung zu sinken.

Alternativ kann die Zuweisung von Bodenbedeckungsinformationen zu phänologi- schen Phasen auf der Grundlage von Expertenwissen erfolgen. Ein Beispiel ist die Ver- knüpfungsregel 3.18 der Methodendokumentation Bodenkunde, in der der “Bedeckungs- grad der durch Pflanzen und Pflanzenrückstände bedeckten Bodenoberfläche” aufgeführt ist (Schmidt et al., 1996). Neben den beobachteten phänologischen Phasen werden noch die Kategorien “Saatbett” (SB) und “Brache” (B) berücksichtigt, die nicht beobachtete

Tabelle 7: Namen und Zahlencodes beginnender phänologischer DWD-Phasen ([DWD, 2015](#)) sowie korrespondierende BBCH-Codes ([Bleiholder et al., 2001](#))

Phasename	DWD-Code	BBCH-Code
Aussaat	10	0
Auflaufen	12	10
Rosettenbildung	14	14
Schossen/Längenwachstum	15/67	31
Knospenbildung	17	50
Ährenschieben	18	51
Milchreife	19	75
Teigreife	20	83
Gelbreife	21	87
Vollreife	22	–
Ernte	24	–
Fahnenschieben	65	53

Zeiträume repräsentieren:

- Als Saatbett wird der Zeitraum vor der Bestellung von Mais bezeichnet. Dabei wird davon ausgegangen, dass die Bodenbearbeitung bereits im Herbst des Vorjahres stattgefunden hat und der Boden unbedeckt ist.
- Brache ist der Zeitraum nach der Ernte und vor Beginn der Bodenbearbeitung. Der Boden ist von Stoppeln und Pflanzenresten bedeckt, wobei für Mais und Winterweizen unterschiedliche Bedeckungsgrade angenommen werden.

Literatur

- Behrens, T., Schmidt, K., MacMillan, R. A., & Viscarra Rossel, R. A. (2018). Multi-scale digital soil mapping with deep learning. *Scientific Reports*, 8(1).
- Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I., & Heynen, M. (2004). Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(3-4), 239–258.
- Beven, K. J. & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Bulletin*, 24(1), 43–69.
- Bleiholder, H., Weber, E., Feller, C., Hess, M., Wicke, H., Meier, U., Boom, T. v. d., Lancashire, P. D., Buhr, L., Hack, H., Hack, H., Stauss, R., & Meier, U. (2001). *Growth stages of mono- and dicotyledonous plants*. Technical report, Federal Biological Research Centre for Agriculture and Forestry, Brunswick, Germany. Edition: 2.
- Breiman, L. (2001). [No title found]. *Machine Learning*, 45(1), 5–32.
- Brenning, A., Bangs, D., & Becker, M. (2018). *RSAGA: SAGA Geoprocessing and Terrain Analysis*.
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., & Böhner, J. (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, 8(7), 1991–2007.
- DWD (2015). *Vorschriften und Betriebsunterlagen für die phänologischen Beobachter des Deutschen Wetterdienstes*. Offenbach, Germany: Deutscher Wetterdienst.
- Gerstmann, H., Doktor, D., Gläßer, C., & Möller, M. (2016a). PHASE: A geostatistical model for the Kriging-based spatial prediction of crop phenology using public phenological and climatological observations. *Computers and Electronics in Agriculture*, 127, 726–738.
- Gerstmann, H., Möller, M., & Gläßer, C. (2016b). Optimization of spectral indices and long-term separability analysis for classification of cereal crops using multi-spectral RapidEye imagery. *International Journal of Applied Earth Observation and Geoinformation*, 52, 115–125.
- Guisan, A., Weiss, S. B., & Weiss, A. D. (1999). GLM versus CCA spatial modeling of plant species distribution}. *Plant Ecology*, 143(1), 107–122.
- Kaspar, F., Zimmermann, K., & Polte-Rudolf, C. (2015). An overview of the phenological observation network and the phenological database of Germany's national meteorological service (Deutscher Wetterdienst). *Advances in Science and Research*, 11(1), 93–99.
- Khaledian, Y. & Miller, B. A. (2020). Selecting appropriate machine learning methods for digital soil mapping. *Applied Mathematical Modelling*, 81, 401–418.

- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. New York: Cambridge University Press. OCLC: ocn190786122.
- Möller, M., Birger, J., Gidudu, A., & Gläßer, C. (2013). A framework for the geometric accuracy assessment of classified objects. *International Journal of Remote Sensing*, 34(24), 8685–8698.
- Möller, M., Boutarfa, L., & Strassemeyer, J. (2020). PhenoWin – An R Shiny application for visualization and extraction of phenological windows in Germany. *Computers and Electronics in Agriculture*, 175, 105534.
- Möller, M., Gerstmann, H., Gao, F., Dahms, T. C., & Förster, M. (2017). Coupling of phenological information and simulated vegetation index time series: Limitations and potentials for the assessment and monitoring of soil erosion risk. *CATENA*, 150, 192–205.
- Möller, M., Koschitzki, T., Hartmann, K.-J., & Jahn, R. (2012). Plausibility test of conceptual soil maps using relief parameters. *CATENA*, 88(1), 57–67.
- Möller, M. & Volk, M. (2015). Effective map scales for soil transport processes and related process domains — Statistical and spatial characterization of their scale-specific inaccuracies. *Geoderma*, 247–248, 151–160.
- Möller, M., Volk, M., Friedrich, K., & Lymburner, L. (2008). Placing soil-genesis and transport processes into a landscape context: A multiscale terrain-analysis approach. *Journal of Plant Nutrition and Soil Science*, 171(3), 419–430.
- Planchon, O. & Darboux, F. (2002). A fast, simple and versatile algorithm to fill the depressions of digital elevation models. *CATENA*, 46(2-3), 159–176.
- Quinn, P., Beven, K., Chevallier, P., & Planchon, O. (1991). The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrological Processes*, 5(1), 59–79.
- {R Core Team} (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rogge, D., Bauer, A., Zeidler, J., Mueller, A., Esch, T., & Heiden, U. (2018). Building an exposed soil composite processor (SCMaP) for mapping spatial and temporal characteristics of soils with Landsat imagery (1984–2014). *Remote Sensing of Environment*, 205, 1–17.
- Schmidt, J., Von Werner, M., & Michael, A. (1996). *EROSION 2D/3D: Ein ComputermodeLL zur Simulation der Bodenerosion durch Wasser*. Technical report, Sächsische Landesanstalt für Landwirtschaft, Dresden, & Sächsisches Landesamt für Umwelt und Geologie.
- Sponagel, H., hoc-Arbeitsgruppe Boden der Staatlichen Geologischen Dienste und der Bundesanstalt für Geowissenschaften und Rohstoffe, A., & für Geowissenschaften und Rohstoffe, B., Eds. (2005). *Bodenkundliche Kartieranleitung: mit 103 Tabellen und 31 Lis-*

- ten. Stuttgart: E. Schweizerbart'sche Verlagsbuchhandlung (Nägele und Obermiller), 5., verbesserte und erweiterte auflage edition. OCLC: 76756137.
- Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, 62(1), 77–89.
- Taghizadeh-Mehrjardi, R., Schmidt, K., Eftekhari, K., Behrens, T., Jamshidi, M., Davatgar, N., Toomanian, N., & Scholten, T. (2020). Synthetic resampling strategies and machine learning for digital soil mapping in Iran. *European Journal of Soil Science*.
- Volk, M., Möller, M., & Wurbs, D. (2010). A pragmatic approach for soil erosion risk assessment within policy hierarchies. *Land Use Policy*, 27(4), 997–1009.
- Yokoyama, R., Shlrasawa, M., & Pike, R. J. (2002). Visualizing Topography by Openness: A New Application of Image Processing to Digital Elevation Models Ryuzo Yokoyama, Mlchlo Shlrasawa, and Richard J. Pike. *Photogrammetric Engineering and Remote Sensing*, 68, 257–265.
- Zevenbergen, L. W. & Thorne, C. R. (1987). Quantitative analysis of land surface topography. *Earth Surface Processes and Landforms*, 12(1), 47–56.