

Zusätzliche Informationen zu

„pH-Werte deutscher Böden auf Wald- und Agrarflächen“, Thünen à la carte 9, Oktober 2021

Teil B: ERWEITERTE STATISTISCHE AUSWERTUNG

Zur Auswertung der erstellten Karten wurden statistische Kennwerte und Verteilungen herangezogen (Tab.B1 und Abb.B1). Die Mittelwerte der Modellergebnisse entsprechen den Mittelwerten der gemessenen Werte. Die Verteilungen der modellierten Werte für beide Bodenschichten sind jedoch enger als die der gemessenen. Weiterhin unterscheiden sich gemessene und modellierte Werte in den Formen ihrer Verteilungen. Die gemessenen Werte ergeben für beide Bodenschichten trimodale Verteilungen (drei lokale Maxima), während die modellierten Werte bimodal verteilt sind (zwei lokale Maxima). Die Aufschlüsselung nach Waldstandorten und agrarischen Standorten zeigt, dass die Trimodalität der gemessenen pH-Werte sich aus einer Überlagerung beider Standortklassen mit je bimodaler Verteilung ergibt (Abb. B1, links). Allerdings gelingt es bei beiden Klassen nicht vollständig, bimodale Verteilungen zu modellieren (Abb. B1, rechts). Stattdessen ergeben sich unimodale Verteilungen. Diese Unterschiede in den Verteilungen entstanden, weil anteilig weniger Extremwerte modelliert als gemessen wurden. Eine Ursache dafür könnte sein, dass das genutzte Modell *Quantile Regression Forest* (Meinshausen, 2006), wie vom *Random Forest* Ansatz stammenden Modelle allgemein, Extremwerte generell unterschätzen und auch nicht darüber hinaus extrapolieren können. Der Grund hierfür ist, dass jede Modellschätzung stets ein Mittelwert der Schätzungen aller einzelnen Entscheidungsbäume im Modell ist, und auch diese Einzelschätzungen maximal den gelernten Extremwerten im Trainingsdatensatz entsprechen können.

Der hauptsächliche Grund für die engere Verteilung der modellierten Werte ist jedoch Überanpassung: Kombinationen der vielen Prädiktoren erklären zwar die seltenen extremen gemessenen pH-Werte, wiederholen sich so aber kaum, weshalb diese Extremwerte selten an Rasterzellen, in denen es keine Messwerte gibt, auftreten. Eine Lösungsansatz zur Bewältigung dieses Problems könnte darin liegen, Informationen von temporär nicht genutzten Datensätzen bei der Modellbildung nach *Random Forest* - Prinzip zu nutzen. Solche nicht genutzten Datensätze (*Out-of-Bag Data*) werden dabei gemäß *Bootstrapping* Methode immer wieder generiert. Die Residuen zwischen Messwerten und den auf die *Out-of-Bag Data* angewandten Modellergebnissen (*Out-of-Bag Error*) sollten an den Rändern des pH-Wertespektrums am höchsten sein. Eine Möglichkeit diese Residuen in die Karte zu integrieren wäre eine geostatistische Nachbearbeitung des Modells (*Regression-Kriging*). Die für die vorliegenden Karten entstandenen Abweichungen können jedoch in Kauf genommen werden, da (1) diese bei der Klassifizierung der pH-Werte nicht stark ins Gewicht fallen und (2) eine genaue statistische Bewertung der Prädiktoren und des Modells nicht im Vordergrund stand.

Der für die Verteilungen der Oberflächennutzung angewandte *Corine Land Cover*-Datensatz (CLC, BKG, 2012) enthielt eine Waldfläche von ca. 107.000 km². Die durch die Bundeswaldinventur ermittelte Gesamtwaldfläche Deutschlands beträgt ca. 115.000 km², was einer Abweichung von ca. 7 %

entspricht (Thünen-Institut, 2012). Die landwirtschaftlich genutzte Fläche nach CLC liegt mit über 218.500 km² ca. 18% über der von Tietz et al (2012) ermittelten Fläche von ca. 185.000 km². Gründe für diese Abweichungen können sein, dass zum einen Fernerkundungsdaten nicht eindeutig zwischen bestimmten Nutzungsformen unterscheiden und dass nicht eindeutige Nutzungsformen bei Tietz et al (2012) und nach CLC womöglich unterschiedlich zugeordnet werden und dass zum anderen Rasterzellenwerte über den Modalwert bestimmt wurden (Abschnitt „Räumliche Modellierung“). Unter Berücksichtigung dessen wurden die Abweichungen im Sinne der Zielstellung als vertretbar angesehen.

		n	1.Quantil	Median	3. Quantil	Mittelwert
BZE gesamt	0 – 30 cm	4.954	4.64	5.92	6.81	5.86
	30 – 100 cm	4.897	4.95	6.14	7.2	6.16
Raster gesamt	0 – 30 cm	325.473	4.67	6.06	6.7	5.86
	30 – 100 cm		5.16	6.31	6.96	6.16
BZE-Wald	0 – 30 cm	1.851*	4.24	4.45	4.86	4.8
	30 – 100 cm	1.809	4.52	4.76	5.58	5.3
Raster Wald	0 – 30 cm	106.886	4.35	4.43	4.66	4.65
	30 – 100 cm		4.63	4.75	5.19	5.12
BZE-LW	0 – 30 cm	3.103*	5.85	6.44	7.13	6.48
	30 – 100 cm	3.088	5.93	6.66	7.47	6.67
Raster LW	0 – 30 cm	218.587	5.98	6.48	6.88	6.45
	30 – 100 cm		6.17	6.68	7.12	6.66

Tabelle B1. Statistische Kenngrößen der gemessenen und modellierten pH-Werte. Zeilen BZE: Bodenzustandserhebungen, Zeilen Raster: An das Corine Land Cover (Tabelle A1, Zielwerte des CLC) angelehnte Einordnung der modellierten Werte, Raster LW: CLC Klassen 211-243, Raster Wald: CLC Klassen 311, 312, 313 und 324. Spalte n: Anzahl der Messpunkte (BZE) oder Anzahl der Rasterzellen. *Abweichungen von Angaben im Haupttext stammen daher, dass sich einige wenige Messpunkte nicht für die Modellierung eigneten.

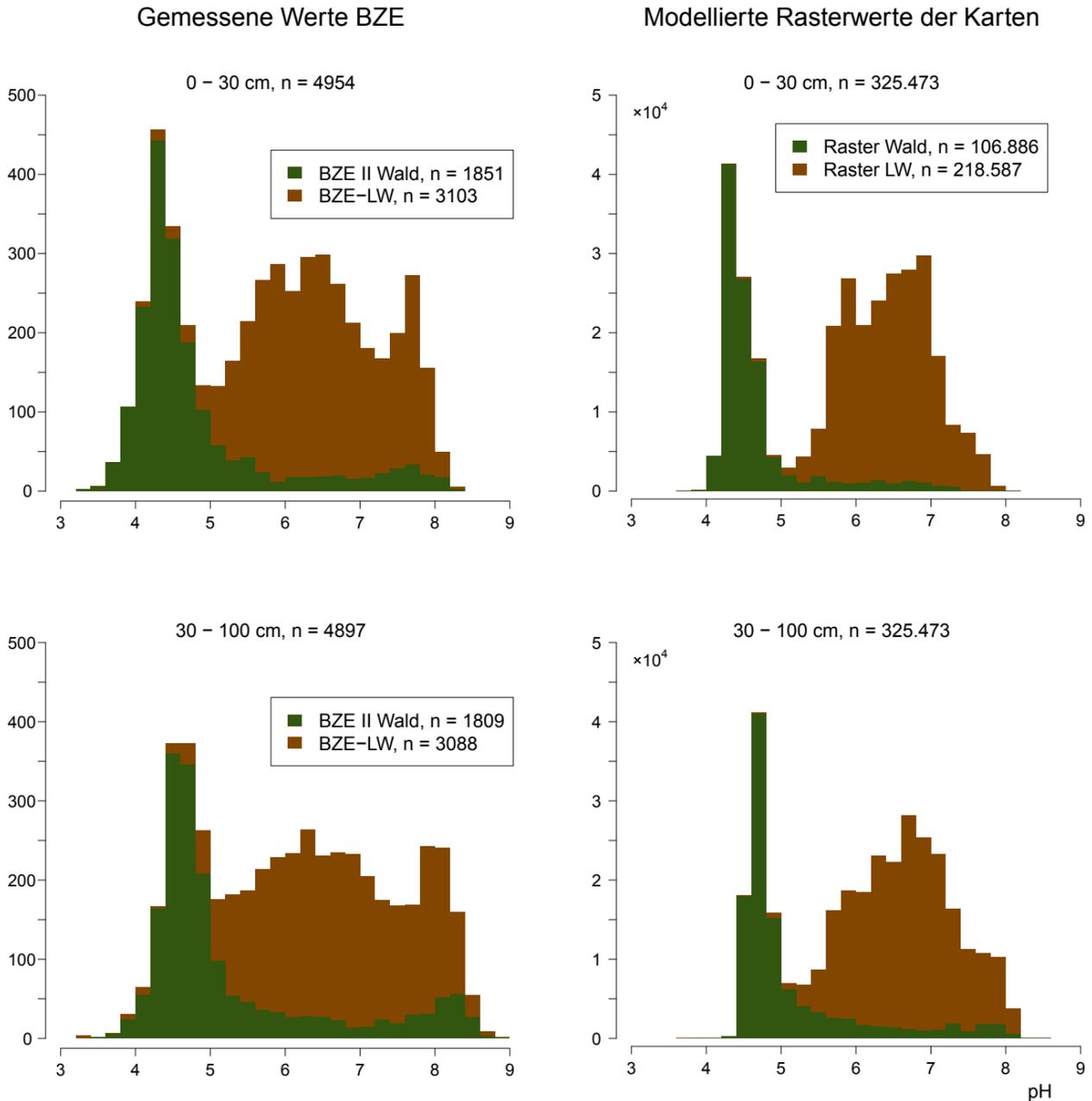


Abbildung B1: Verteilungen der gemessenen pH-Werte für die Punkte der Bodenzustandserhebung im Wald (BZE II-Wald) und auf landwirtschaftlich genutzten Flächen (BZE-LW) sowie für die modellierten pH-Werte auf Wald- und Landwirtschaftlichen Flächen des *Corine Land Cover*-Datensatzes (Tabelle A1, Zielwerte des CLC). Als Waldflächen wurden die CLC Klassen: 311, 312, 313 und 324 definiert, als landwirtschaftliche Flächen die CLC Klassen: 211-343, n BZE: Anzahl der Messpunkte, n Karte: Anzahl der Rasterzellen.

Danksagung

Wir möchten Nikolai Knapp (Institut für Waldökosysteme des J.H.von Thünen-Instituts) für seine Beratung danken, ohne welche diese erweiterte statistische Auswertung nicht möglich gewesen wäre.

Quellenangaben

BKG (2012) CORINE Land Cover 10 ha (CLC10). Bundesamt für Kartographie und Geodäsie.

Thünen-Institut (2012) Dritte Bundeswaldinventur - Ergebnisdatenbank. Verfügbar unter: <https://bwi.info>, Aufruf am: 29.04.2021, Auftragskürzel: 77Z1JI_L101of_2012, Archivierungsdatum: 2014-8-13 16:42:23.590, Überschrift: Waldfläche [ha] nach Land und Waldspezifikation, Filter: Jahr=2012.

Tietz, A, Bathke, M, Osterburg, B (2012) Art und Ausmaß der Inanspruchnahme landwirtschaftlicher Flächen für außerlandwirtschaftliche Zwecke und Ausgleichsmaßnahmen. Arbeitsberichte aus der vTI-Agrarökonomie.