

Scherstjanoi, M., Grüneberg, E., Wellbrock, N.

Zusätzliche Informationen zu

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

Teil A: ERWEITERTE METHODENBESCHREIBUNG

Die Erstellung beider Karten (Bodentiefen 0-30 cm und 30-100 cm) basierte auf einer von der FAO (Omuto et al, 2020) empfohlenen maschinellen Lernmethode (Quantile Regression Forest, Meinshausen, 2006). Diese Methode wurde auf je 17 räumliche Vorhersagevariablen angewandt (Tabelle A1), wobei die Auswahl der Prädiktoren auf einem Leitfaden der FAO basierte. Die räumliche Auflösung der Rasterzellen von Karten und Prädiktoren beträgt 1 km². Die Prädiktoren mit metrischen Variablen wurden aus den jeweiligen Basiskarten für die gewünschten Rasterzellen interpoliert und so für die gesamte Modellierung verwendet. Für Prädiktoren mit nominalen Variablen wurden die Modalwerte innerhalb der Rasterzellen verwendet. Wenn jedoch eine Rasterzelle einen pH-Messpunkt enthielt, wurde zusätzlich nach einem Quell- und einem Zielwert unterschieden. Der Zielwert entspricht dem Modalwert der Zelle und ist für die Berechnung an dem Rasterpunkt relevant. Der Quellwert ist für den Abgleich mit den Messwerten relevant und entspricht dem Wert, den die Prädiktoren an den Stellen der Messpunkte haben. Ohne diese Unterscheidung trafen relevante Korrelationen zwischen Messpunkten und Prädiktoren nicht zu und das Gesamtergebnis wäre stärker verfälscht. Wegen der hohen Fragmentierung des Prädiktors Oberflächenbedeckung wurde zur Bestimmung seines Quellwertes eine Basiskarte mit 1 ha Auflösung herangezogen. Für den Zielwert wurde eine Basiskarte mit 10 ha Auflösung für ausreichend befunden.

Die gesamte Umrechnung der Eingangsdaten und der Rasterkartenwerte wurde mit der statistischen Programmierumgebung R (v.4.0.2, R Core Team, 2020) und den entsprechenden Softwarepaketen (Tabelle A2) durchgeführt. Nach Tests mit der *regmodelSuit* Funktion des R *soilassessment* packages (Omuto et al, 2020; v.0.2.1, Omuto, 2021) erwies sich das *Quantile Regression Forest* Modell (Meinshausen, 2006) aufgrund der geringsten Fehlerwahrscheinlichkeit am geeignetsten für die maschinelle Lernmethode. Die Mittelung der gemessenen Werte auf die in den Karten abgebildeten Bodentiefen erfolgte wie von der FAO vorgeschlagen (Omuto et al, 2020) mit einer Spline Interpolation (Bishop *et al.*, 1999, Smoothing Parameter 0.21) unter Nutzung des R packages *GSIF: Global Soil Information Facilities* (Hengl et al, 2019).

Karte / Daten	Prädiktor	Typ	Methode	Funktion
Digitales Höhenmodell ¹				
	Geländehöhe	1	Rasterprojektion mit GRASS ⁸ QGIS ⁷ tools	r.resamp.interp, bilinear interpolation
	Hang	1b		Basic Terrain Analysis
	LS-Faktor	1b		Basic Terrain Analysis
	Taltiefe	1	Abgeleitet aus Geländehöhe mit SAGA ⁹ QGIS tools	Basic Terrain Analysis
	Längskrümmung	1		Slope, aspect and curvature
	Kanalnetzwerkabstand	1		Basic Terrain Analysis
Bodenübersichtskarte ²				
	Bodentyp	2		
CORINE Land Cover 1ha (Quellwerte) und 10 ha (Zielwerte) ³				
	Oberflächenbedeckung	2		
			1. Gerastert mit QGIS	Rastern (Vektor nach Raster) 100 m Auflösung,
	Zielwerte		2a. Rasterprojektion mit GDAL QGIS tools ¹⁰	Transformieren, 1 km Auflösung, Abtastmethode: Modus
	Quellwerte		2b. Wert bei Messpunkt an Raster übergeben, mit GDAL QGIS tools	Rasterisieren (Mit Attribute überschreiben)
Monatliche Klimadaten 1990-2018 ⁴				
	Mittlerer Jahresniederschlag	1b		
	Minimale mittlere Monatstemperatur	1	Rasterprojektion mit GRASS QGIS tools	r.resamp.interp, bilinear interpolation
	Maximale mittlere Monatstemperatur	1		
Modis Daten ⁵				
	Eigenwerte der ersten sechs Komponenten nach Hauptkomponentenanalyse mit den Vegetationsindizes SII-6, SAVI, VSSI, NDSI, NDVI, SR, CRSI, BI ⁶	1	Berechnung der Vegetationsindizes aus Modis Bändern 1-4, 6 und 7 ⁶ , Erstellung der Bänder mit QGIS Semi-Automatic Classification Plugin ¹¹ , Merging und Resampling mit SAGA QGIS tools, und der Mosaic Raster Layers function, bicubic spline interpolation, mean overlapping areas, match: none.	

Tabelle A1: Übersicht der genutzten Prädiktoren, ihrer Quellen, angewandter Methoden und Funktionen. Alle Prädiktoren wurden als tiff Rasterkarten mit 1km² Auflösung berechnet und mit den Ländergrenzen Deutschlands unter Nutzung der GDAL QGIS tools und der *Clip Raster by Layer Mask* Funktion verschnitten. Typ 1: Kardinalwerte. Typ 1b: log-transformierte Kardinalwerte. Typ 2: Nominalwerte. ¹U.S. Geological Survey (2020a), ²BGR (2020), ³BKG (2012), ⁴Fick und Hijmans (2017), ⁵U.S. Geological Survey (2020b), ⁶Omuto et al. (2020), ⁷QGIS.org (2020), v.3.14.1-Pi, ⁸GRASS Development Team (2017), v.7.8.2, ⁹Conrad et al (2015) und Conrad et al (2019), v.7.3.0, ¹⁰GDAL/OGR contributors (2020), v.3.0.4, ¹¹Congedo (2020) v.6.4.6.

R-Pakete	Funktionen	Beschreibung, Anwendungsbereich
rgdal ¹	readGDAL	Einlesen von Rasterdaten
sp ²		Bearbeitung räumlicher Daten
soilassessment ³	imageIndices	Transformieren der Fernerkundungsdaten
aqp ⁴	depths	Übernehmen der gemessenen Bodenschichten
GSIF ⁵	mpspline	Spline Interpolation zu Zielbodenschichten
car ⁶ , carData ⁷	powerTransform	Box-Cox Transformation
soilassessment ³	regmodelSuit	Auswahl der maschinellen Lernmethode
caret ⁸ , lattice ⁹ , ggplot2 ¹⁰	trainControl, train	Bestimmung des Modells mit maschineller Lernmethode
stats ¹¹	predict	Berechnung der Ergebnisse auf Basis des Modells

Tabelle A2: Genutze R-Pakete. ¹Bivand et al (2020), v.1.5-15, ²Pebesma et al (2020), v.1.4-2, ³Omuto et al (2021), v.0.2.1, ⁴Beaudette und Roudier (2020), v.1.19, ⁵Hengl et al (2019), v.0.5-5.1, ⁶Fox et al (2020a), v.3.0-8, ⁷Fox et al (2020b), v.3.0-4, ⁸Kuhn (2008, 2020), v.6.0.86, ⁹Sarkar (2008, 2020), v.0.20.41, ¹⁰Wickham (2016) und Wickham et al (2020), v.3.3.2, ¹¹R Core Team (2020), v.4.0.2.

Quellenangaben

BGR (2020) Bodenübersichtskarte der Bundesrepublik Deutschland 1:1.000.000 (BÜK1000). Bundesanstalt für Geowissenschaften und Rohstoffe. Verfügbar unter: <https://www.bgr.bund.de>.

Bishop, TFA, McBratney, AB, Laslett, GM (1999) Modelling soil attribute depth functions with equal-area quadratic smoothing splines, *Geoderma*, 91(1–2), S. 27–45. doi:10.1016/S0016-7061(99)00003-8.

Bivand, R, Keitt, T, Rowlingson, B (2020) rgdal: Bindings for the Geospatial Data Abstraction Library. Verfügbar unter: <https://CRAN.R-project.org/package=rgdal>.

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

Congedo, L (2020) Semi-automatic classification plugin. Verfügbar unter: <https://plugins.qgis.org/plugins/SemiAutomaticClassificationPlugin/>.

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), S. 1991–2007. doi:10.5194/gmd-8-1991-2015.

Conrad, O, Bechtel, B, Bock, M, Dietrich, H, Fischer, E, Gerlitz, L, Wehberg, J, Wichmann, V, Böhner, J (2019) System for automated geoscientific analyses (SAGA). Verfügbar unter: <http://www.saga-gis.org/>.

Fick, SE, Hijmans, RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas, *International journal of climatology*, 37(12), S. 4302–4315. doi:10.1002/joc.5086.

Fox, J, Weisberg, S, Price, B (2020a) car: Companion to Applied Regression. Verfügbar unter: <https://r-forge.r-project.org/projects/car/>, <https://CRAN.R-project.org/package=car>, <https://socialsciences.mcmaster.ca/jfox/Books/Companion/index.html>.

Fox, J, Weisberg, S, Price, B (2020b) carData: Companion to Applied Regression Data Sets. Verfügbar unter: <https://r-forge.r-project.org/projects/car/>, <https://CRAN.R-project.org/package=carData>, <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion/index.html>.

GDAL/OGR contributors (2020) GDAL/OGR Geospatial data abstraction software library. Open Source Geospatial Foundation. Verfügbar unter: <https://gdal.org>.

GRASS Development Team (2017) Geographic Resources Analysis Support System (GRASS) Software. Open Source Geospatial Foundation. Verfügbar unter: <http://grass.osgeo.org>.

Hengl, T, Kempen, B, Heuvelink, G, Malone, B (2019) GSIF: Global Soil Information Facilities. Verfügbar unter: <https://gsif.r-forge.r-project.org/>.

Kuhn, M (2008) Building predictive models in R using the caret package, *Journal of Statistical Software*, 28(5), S. 1–26.

Kuhn, M (2020) caret: Classification and Regression Training. Verfügbar unter: <https://CRAN.R-project.org/package=caret>.

Meinshausen, N (2006) Quantile regression forests., Journal of Machine Learning Research. Herausgegeben von G. Ridgeway, 7(6), S. 983–999.

Omuto, CT, Vargas, R, Viatkin, K, Yigini, J (2020) Mapping of salt-affected soils: Technical manual. Rome: FAO. DOI:10.4060/ca9215en

Pebesma, E, Bivand, R, Rowlingson, B, Gómez-Rubio, V (2020) sp: Classes and Methods for Spatial Data. Verfügbar unter: <https://CRAN.R-project.org/package=sp>.

QGIS.org (2020) QGIS Geographic Information System. QGIS Association. Verfügbar unter: <http://qgis.org>.

R Core Team (2020) R: A language and environment for statistical computing. R foundation for Statistical Computing, Vienna, Austria.

Sarkar, D (2008) Lattice: multivariate data visualization with R. Springer Science & Business Media.

Sarkar, D (2020) lattice: Trellis Graphics for R. Verfügbar unter: <http://lattice.r-forge.r-project.org/>.

U.S. Geological Survey (2020a) Earth Explorer Maps GMTED2010N30E000 & GMTED2010N50E000. Verfügbar unter: <https://earthexplorer.usgs.gov>.

U.S. Geological Survey (2020b) Earth Explorer Maps MOD09A1.A2020217.h18v03.006 & MOD09A1.A2020217.h18v04.006. Verfügbar unter: <https://earthexplorer.usgs.gov>.

Wickham, H (2016) ggplot2-Elegant Graphics for Data Analysis. Springer-Verlag New York. Verfügbar unter: <https://ggplot2.tidyverse.org>.

Wickham, H, Chang, W, Henry, L, Pedersen, TL, Takahashi, K (2020) ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. Verfügbar unter: <https://ggplot2.tidyverse.org>, <https://cran.r-project.org/web/packages/ggplot2/>.