

# Entwicklung von Pedotransferfunktionen zur Bestimmung von hydrologischen Bodeneigenschaften in Österreich

## *Development of pedotransfer functions for the assessment of soil hydrological properties in Austria*

Florian Darmann\*, Irene Winkler, Monika Kumpan, Peter Strauss, Thomas Weninger

### **Einleitung**

Die Erfassung von hydrologischen Bodeneigenschaften ist kostenintensiv und mit hohem zeitlichem Aufwand verbunden. Um dennoch eine flächenhafte Aussage über die entsprechenden Eigenschaften treffen zu können, werden diese oft indirekt aus einfacher zu bestimmenden Parametern über sogenannte Pedotransferfunktionen (PTFs) abgeschätzt. Die Methoden zur Ableitung solcher PTFs reichen von einfachen linearen Regressionen über multiple Regressionen bis hin zur Anwendung von Machine Learning Algorithmen (Vereecken et al., 2010). In der Regel werden PTFs jedoch für bestimmte Regionen oder Einzugsgebiete entwickelt. In dieser Studie liegt der Fokus auf der Ableitung neuer PTFs für österreichische Böden mit landwirtschaftlicher Nutzung. Unter der Anwendung der Random Forest Methode werden Vorhersagemodelle zur Bestimmung von hydrologischen Bodeneigenschaften wie der Feldkapazität (FC), dem Permanenten Welkepunkt (PWP) und der gesättigten hydraulischen Leitfähigkeit (Ks) entwickelt.

### **Material und Methoden**

Die Ableitung der PTFs erfolgte auf Basis der Random Forest Methode. Diese wird den Machine Learning Algorithmen zugeordnet und gilt als robuster Vorhersagealgorithmus (Boulesteix et al., 2012). Dabei wird durch die Kombination mehrerer, einzelner Entscheidungsbäume eine nichtlineare Beziehung zwischen den Variablen modelliert. Um zusätzlich Informationen über die Unsicherheiten der Vorhersage zu erhalten, kann die Random-Forest-Methode in Kombination mit der Methode der Quantilsregression verwendet werden. Während das reguläre Random Forest Modell den Mittelwert aller Einzelbäume ermittelt, wird durch die Quantilsregression die gesamte Verteilung aller vorhergesagten Werte bestimmt (Meinshausen, 2006). Dadurch ist es möglich, auf Basis des errechneten Interquartilsabstandes (Q90-Q10) auch die Unsicherheit bei der Prognose der hydrologischen Bodeneigenschaften zu ermitteln und eine Aussage über die Qualität der Vorhersage zu treffen.

Bodendaten: Die Grundlage zum Aufbau des Random-Forest Modells bildet ein Datensatz des Bundesamts für Wasserwirtschaft – Institut für Kulturtechnik und Wasserhaushalt (BAW-IKT). Dieser umfasst Bodendaten von diversen Projekten und repräsentiert dabei verschiedenste Bodentypen und Landschaften aus ganz Österreich. Insgesamt beinhaltet der hier verwendete Datensatz 2278 Proben von 518 landwirtschaftlichen Standorten, wobei jedes Bodenprofil zumindest eine Messung von hydraulischen Bodenparametern enthält. Eine Übersicht über alle Probenahmestellen ist in Abbildung 1 dargestellt.

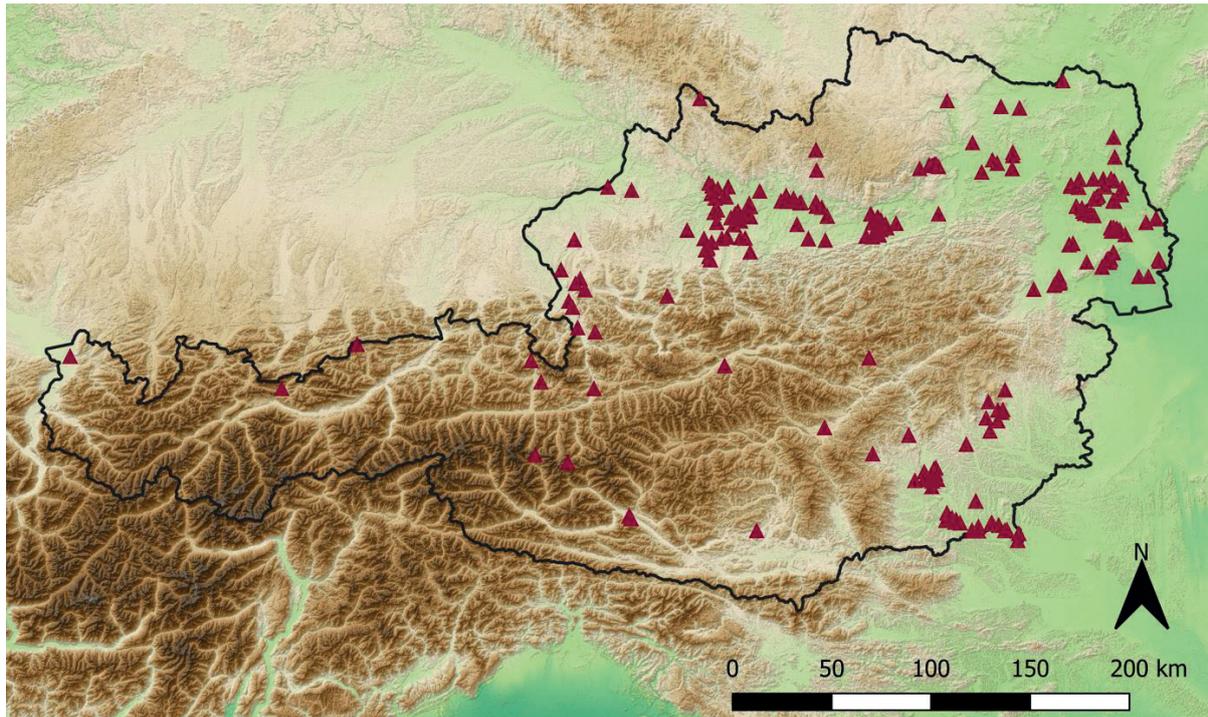


Abbildung 1: Übersicht über die im Datensatz des BAW-IKT enthaltenen Probenahmestellen

Bei der Datenaufbereitung wurden ausschließlich solche Proben berücksichtigt, die mit Standardmethoden analysiert wurden. Der Datensatz enthält Informationen zu grundlegenden Bodeneigenschaften wie Korngrößenverteilung, Lagerungsdichte, Humusgehalt, und hydrologischen Bodeneigenschaften wie den Wassergehalt bei verschiedenen Matrixpotentialen (Tabelle 1)

Tabelle 10: Verfügbare Daten inklusive Minimum, Maximum, Mittelwert und Median

	n	Min	Max	Mean	Median
Number of profiles	518				
Number of samples	2278				
Bulk density [g/cm <sup>3</sup> ]	1177	0,1	2,2	1,4	1,4
Soil organic carbon content [Mass-%]	1424	0,0	34,9	1,1	0,9
Soil organic matter [Mass-%]	1424	0,0	60,0	1,8	1,5
Calcium carbonate content [Mass-%]	1144	< 0,95	82,6	13,2	11,7
Porosity [Vol.-%]	557	0,4	68,7	42,0	44,3
pH-value	961	4,1	8,1	7,0	7,4
kF [cm/d]	883	0,0	27624,4	547,7	103,8
<b>Particle size distribution</b>					
Sand content [Mass-%]	1518	1,3	98,3	36,7	29,7
Silt content [Mass-%]	1518	1,1	86,0	45,9	49,0
Clay content [Mass-%]	1518	0,0	59,8	17,4	18,0
<b>Water retention curve by method</b>					
Pressure plate apparatus	956				
HYPROP	299				

**Anwendung der Pedotransferfunktionen:** Während der Datensatz des BAW-IKT zur Modellentwicklung verwendet wurde, wurden für die flächendeckende Zuordnung der abgeleiteten bodenhydrologischen Eigenschaften die offiziellen Bodendaten aus der österreichischen Bodenkarte (eBOD) verwendet. Diese beinhaltet Daten zu allen landwirtschaftlichen Flächen Österreichs und enthält Information zu Bodentextur, organischer Substanz oder pH-Wert. Dementsprechend wurde bei der Entwicklung der PTFs die Verfügbarkeit der Input-Daten berücksichtigt. Als Zielvariablen wurden in dieser Arbeit zwei

Ansätze für die Feldkapazität festgelegt ( $pF = 1,8$  entspricht dem Wasseranteil bei einem Matrixpotential von  $-60\text{hPa}$  und  $pF = 2,5$  bei  $-300\text{ hPa}$ ), sowie der Permanente Welkepunkt und die gesättigte hydraulische Leitfähigkeit. Für jede Zielvariable wurden mehrere mögliche Kombinationen an Input-Variablen getestet, wobei die Bodentextur und die Entnahmetiefe als fixe Variablen definiert wurden. Lagerungsdichte, Humusgehalt, Grobstoffanteil, pH-Wert und  $\text{CaCO}_3$  wurden in unterschiedlichen Varianten kombiniert. Die weitere Beurteilung der abgeleiteten Modelle erfolgte anhand eines Rangsystems, welches auf den resultierenden Qualitätsparametern  $R^2$  (Bestimmtheitsmaß) und RMSE (Root Mean Square Error) sowohl beim Training, als auch beim Validieren basiert. Dabei wurden für jede Zielgröße die drei besten PTFs zur weiteren Verwendung ausgewählt. Um die Anwendbarkeit der abgeleiteten PTFs sicherzustellen, wurde ein Vergleich mit bereits bestehenden PTFs durchgeführt. Dazu wurden die ausgewählten IKT-Modelle sowie die bestehenden PTFs auf den Validierungsdatensatz angewendet und die Ergebnisse anhand statischer Parameter verglichen. Zusätzlich zum  $R^2$  und RMSE wurden dabei der MSE (Mean Square Error) und der SDPE (Standard Deviation of Prediction Error) berücksichtigt.

### Ergebnisse und Diskussion

Die Ableitung der PTFs resultiert in 27 unterschiedlichen PTFs mit variablen Input-Parametern für jede Zielvariable, wobei die Ergebnisse exemplarisch für die Feldkapazität bei  $pF 2,5$  erläutert werden. Anhand des Rangsystems wurden die drei am besten gereihten Modelle zur weiteren Verwendung ausgewählt (IKT\_23, IKT\_26, IKT\_27). Diese zeigten sowohl beim Training als auch der Validierung die besten Ergebnisse. Die drei Modelle berücksichtigen die chemischen Parameter pH-Wert und  $\text{CaCO}_3$  sowie zusätzlich die Lagerungsdichte in allen drei Varianten. Der Humusgehalt ist bei zwei von drei Varianten vorgesehen, der Grobstoffanteil nur einmal bei IKT\_27 vertreten. Um eine bessere Vergleichbarkeit zu schaffen, wurden zusätzlich zwei Modelle mit abweichenden Inputparametern ausgewählt. Dies umfasst die Variante IKT\_01, die ausschließlich die beiden fixen Parameter Bodentextur und Entnahmetiefe beinhaltet, sowie IKT\_19, bei dem die Kombination Lagerungsdichte, Humusgehalt und Grobstoff berücksichtigt werden. Anders als bei den drei besten Modellen beinhalten diese Varianten keine chemischen Parameter als mögliche Inputs.

Tabelle 1: Ergebnisse der fünf ausgewählten PTFs zur weiteren Verwendung für Feldkapazität bei  $pF 2,5$ .  
 PSD = Korngrößenverteilung Sand/Schluff/Ton, Depth = Entnahmetiefe der Probe,  
 BD = Lagerungsdichte, SOM = Humusgehalt, Coarse = Grobstoff

Feldkapazität - $pF 2,5$								
PTF	Input	Train			Test			Overall
		n	RMSE	$R^2$	n	RMSE	$R^2$	Rank
IKT_01	PSD + DEPTH	703	5,45	0,67	165	4,53	0,72	27
IKT_19	PSD + DEPTH + BD + SOM + COARSE	617	4,63	0,76	165	4,09	0,77	20
IKT_23	PSD + DEPTH + $\text{CaCO}_3$ + pH + BD	394	4,32	0,78	100	3,78	0,80	3
IKT_26	PSD + DEPTH + SOM + BD + $\text{CaCO}_3$ + pH	394	4,29	0,79	100	3,84	0,79	2
IKT_27	PSD + DEPTH + SOM + BD + $\text{CaCO}_3$ + COARSE + pH	394	4,31	0,79	100	3,77	0,80	1

Beim Vergleich mit bestehenden PTFs zeigte sich, dass alle fünf IKT-Modelle die entsprechenden hydrologischen Bodenkennwerte mit höherer Genauigkeit ableiten (Abb. 2). Die fünf IKT-Modelle weisen nur geringe Schwankungen beim Vergleich untereinander auf (MSE zwischen 14,2 bis 19,0, SDPE zwischen 28,4 und 36,4). Im Vergleich dazu liegt der MSE für die anderen PTFs zwischen 23,9 und 58,2, der SDPE reicht von 41,4 bis 73,2. Ein Grund für die bessere Performance der IKT-Modelle ist die zu erwartende Regionalität bei der Ableitung neuer PTFs, nachdem die Anwendung der jeweiligen Funktionen auf spezifische Gebiete oder Einzugsgebiete limitiert ist.

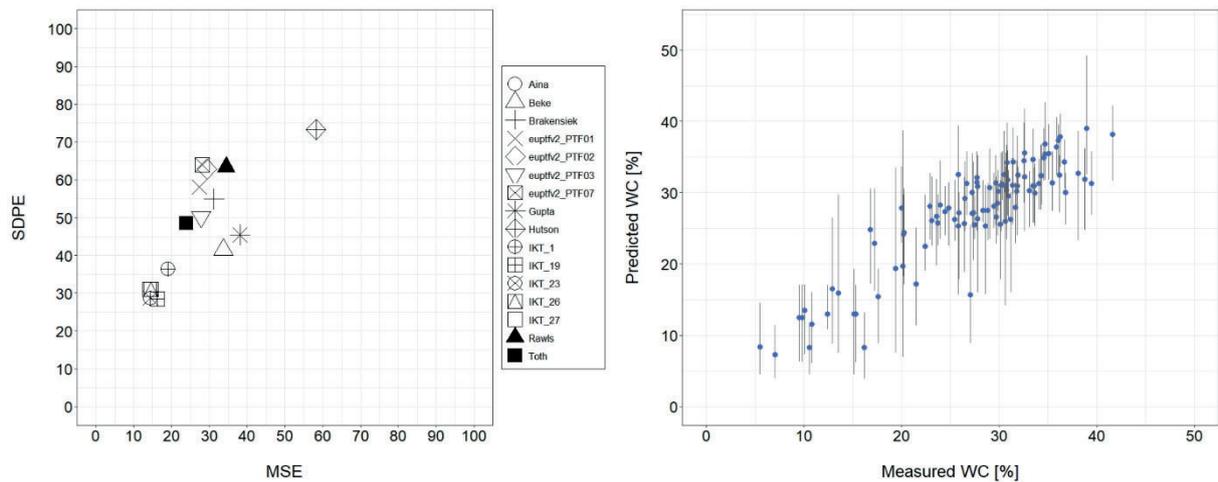


Abbildung 2: Links: Bewertung der verschiedenen Pedotransferfunktionen anhand der statistischen Kennwerte SDPE und MSE. Rechts: Gemessener versus vorhergesagter Wassergehalt von IKT\_27. Die blauen Punkte stellen den ermittelten Wassergehalt dar, die grauen Linie den Interquartilsabstand als Maß für die Unsicherheiten

Auf Basis der eBOD-Daten wurden durch die Anwendung der auf Rang 1 gereihten PTF die hydrologischen Bodeneigenschaften für alle landwirtschaftlichen Flächen Österreichs ermittelt. Die Ergebnisse werden in der nachfolgenden Abbildung 3 in Form des errechneten Wassergehalts bei Feldkapazität (pF 2,5), sowie der resultierenden Unsicherheiten in der Vorhersage dargestellt. Insgesamt weisen 67 % der geschätzten Fläche einen IQR bis 15 % auf, während für 30,4 % der Fläche Unsicherheiten zwischen 15 und 21 % ermittelt wurden. 3,6 % entfällt auf die Klasse größer als 21 %.

Dabei ist ersichtlich, dass die größten Unsicherheiten in den Talsohlen des nördlichen Alpenrandes zu beobachten sind, wo schwere Böden mit einem hohen Tonanteil typisch sind. Die Ergebnisse zeigen weiters, dass der Interquartilsbereich umso größer ist, je höher der vorhergesagte Wassergehalt ist.

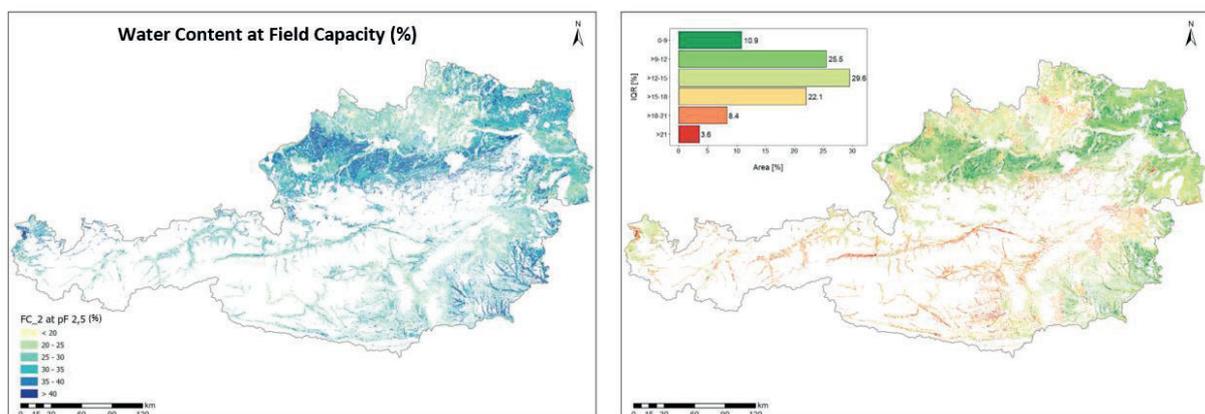


Abbildung 3: Links: Wassergehalt bei Feldkapazität (pF 2,5) für alle landwirtschaftlichen Flächen in Österreich auf Basis der eBOD. Rechts: Darstellung der Unsicherheiten anhand des IQR (Q90-Q10).

## Zusammenfassung

Die direkte Messung von hydrologischen Bodeneigenschaften ist oft zeitaufwändig und kostenintensiv. Dementsprechend werden diese häufig indirekt aus leichter zu messenden Parametern durch sogenannte Pedotransferfunktionen (PTFs) bestimmt. In dieser Studie wurden neue PTFs für österreichische Böden entwickelt, wobei die Grundlage hierfür die Datenbank des BAW-IKT bildet. Diese umfasst Informationen zu verschiedenen Bodeneigenschaften und wird im Rahmen diverser Projekte fortlaufend erweitert. Unter Anwendung der Random Forest Methode erfolgt auf Basis dieser Daten die Ableitung neuer Vorhersagemodelle. Konkret beinhaltet dies die Abschätzung der Feldkapazität (bei pF 1,8 und pF 2,5), des Permanenten Welkepunkts sowie der gesättigten hydraulischen Leitfähigkeit (Ks). Die entwickelten Berechnungsmodelle wurden auf die vorhandenen eBod-Daten angewendet, womit eine Abschätzung der

entsprechenden Zielvariablen für alle landwirtschaftlichen Flächen in Österreich möglich ist. Durch die implementierte Ausweisung der Modellunsicherheit lässt sich eine Aussage über die Güte der Berechnung angeben und über verschiedene räumliche Skalen statistisch auswerten.

### ***Abstract***

As the measurement of soil hydraulic properties is time-consuming and expensive, they are often computed from easily measurable soil properties via pedotransfer functions (PTFs). There are plenty of existing PTFs which were mainly derived for specific regions or catchments. In our study, new PTFs for Austrian soils were developed to estimate soil hydraulic properties such as field capacity or permanent wilting point. The PTFs were built by applying the random forest method, including the prediction of uncertainty measures. The used database includes soil physical data from different projects and monitoring campaigns all over Austria and represents many different landscapes and soil types within the country. We applied the newly derived PTFs to data from the official Austrian soil survey to get nationwide information on the estimated soil hydrological properties.

### ***Literatur***

BOULESTEIX A-L., JANITZA S., KRUPPA J., KÖNIG I.R., 2012: Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *WIREs Data Mining Knowl. Discov.* 2012, 2: 493–507. doi: 10.1002/widm.1072  
MEINSHAUSEN N., 2006: Quantile Regression Forests. *Journal of Machine Learning Research* 7 (2006) 983-999  
VERECKEN H., WEYNANTS M., JAVAUX M., PACHEPSKY Y., SCHAAP M.G., VAN GENUCHTEN M.TH., 2010: Using Pedotransfer Functions to Estimate the van Genuchten – Mualem Soil Hydraulic Properties: A Review. *Vadose Zone J.* 9:795-820. doi:10.2136/vzj2010.0045

### ***Adresse der Autoren***

Bundesamt für Wasserwirtschaft, Institut für Kulturtechnik und Bodenwasserhaushalt, Pollnbergstraße 1, A-3252 Petzenkirchen

\* Ansprechpartner: DI Florian DARMANN, [florian.darmann@baw.at](mailto:florian.darmann@baw.at)