

Tagungsbeitrag zu:
Jahrestagung der DBG, Kommission V.
Titel der Tagung:
Böden – eine endliche Ressource.
Veranstalter:
DBG, 5 – 13. September 2009, Bonn.
Berichte der DBG
(nicht begutachtete online Publikation)
<http://www.dbges.de>

Hochauflösende Vorhersage von Waldbodeneigenschaften in der Nordwestschweiz

Philipp Herbst und Thomas Mosimann

Zusammenfassung

Flächendeckend hochauflösende Karten von Waldbodeneigenschaften sind eine wichtige Grundlage für die Waldentwicklung und schonende Waldbewirtschaftung. Die Bodeneigenschaften müssen dabei modellgestützt prognostiziert werden.

Rund 450 Bodenstandorte liefern die Basis für die Entwicklung detaillierter entscheidungsbasierter Modelle auf der Grundlage von über 30 Vorhersagegrößen. Diese ermöglichen die Erstellung konzeptionell neuer Bodenkarten, die neben der Vorhersage der Bodeneigenschaften auch deren Wahrscheinlichkeit flächendeckend ausweisen.

Parallel erfolgt die Prognose ausgewählter Bodeneigenschaften unter der Anwendung von Random Forest.

Die wissensbasierten Prognosen zeigen, dass je nach Bodeneigenschaft eine Vorhersage in vier bis fünf Ausprägungsklassen möglich ist. Die Random Forest-Modelle sind grundsätzlich anwendbar, bisher aber nur in Kombination mit einem wissensbasierten Modell genügend zuverlässig. Die Datenbasis ist für einen rein automatisierten Modellansatz fast immer zu gering.

Schlüsselwörter

Waldbodeneigenschaften / Hochauflösende Waldbodenkarten / Wahrscheinlichkeitsbasierte Prognosemodelle / Classification Trees / Random Forest / GIS

Leibniz Universität Hannover
Institut für Physische Geographie
und Landschaftsökologie
Schneiderberg 50, 30167 Hannover
e-mail: herbst@phygeo.uni-hannover.de

Problemstellung

Für weite Gebiete Europas gibt es keine flächendeckend hoch auflösenden Karten wichtiger Waldbodeneigenschaften. Dies gilt auch für die meisten Teile der Schweiz. Großmaßstäbige Waldbodenkarten für die Waldbewirtschaftung und die Entwicklung seiner ökologischen Funktionen werden jedoch verstärkt benötigt. Gleichzeitig können detaillierte Feldkartierungen durch die zuständigen Institutionen kaum mehr finanziert werden. Die Bodeneigenschaften müssen prognostiziert werden (MOSIMANN 2005). Die modellgestützte Generierung von hochauflösenden Bodenkarten gewinnt deshalb zusehends an Bedeutung.

Im Rahmen des Projekts „Modellierung von Waldbodeneigenschaften im Kanton Basel-Landschaft“ erfolgt für die gesamte Waldfläche des Kantons (ca. 200 km²) eine entscheidungsbasierte empirisch-statistische Vorhersage der Waldbodeneigenschaften Gründigkeit, Skelettgehalt, Vernässung, Azidität und Wasserspeichervermögen. Da eine solche Modellierung mit Unsicherheiten behaftet ist, werden konzeptionell neue Karten benötigt, die neben einer Prognose der Ausprägung der Bodeneigenschaften auch Aussagen zur Sicherheit der Informationen enthalten.

Seit einigen Jahren gelangen neben empirisch-statistischen Modelansätzen halb- und vollautomatisierte Prognosemodelle verstärkt zum Einsatz (LAGACHERIE ET AL. 2007). Mit dem rasch wachsenden Umfang der Datenbestände, können diese *Data-Mining* Verfahren eine zeit- und kosteneffektive Alternative für die Bereitstellung hoch auflösender Bodenkarten darstellen (BEHRENS & SCHOLTEN 2006).

Dabei mangelt es jedoch an Studien, bei denen automatisierte Prognosen einer flächendeckenden Überprüfung im Extrapolationsraum unterzogen werden.

Deshalb werden im Projekt parallel zu den expertenbasierten Modellvorhersagen, die Bodeneigenschaften Gründigkeit und Skelettgehalt unter der Anwendung von Random Forest (BREIMAN 2001) prognostiziert. Die Modellergebnisse werden flächendeckend validiert.

Methodik

Für die Modellentwicklung stehen rund 450 im Bodenaufschluss beschriebene Waldbodenprofile des Kanton Basel-Landschaft und der unmittelbaren Nachbargebiete zur Verfügung. Von diesen profil- und horizontbezogenen Informationen zu den Waldböden wurden etwa 100 Standorte im Projekt feldbodenkundlich neu aufgenommen.

Basierend auf einer mehrstufigen hierarchischen Reliefgliederung, erfolgt unter ArcGIS und SAGA zunächst die weitgehend automatische Gliederung des Reliefs in rund 20 komplexe Relieftypeinheiten. Diese bestimmen die Entwicklung und Ausprägung der Bodeneigenschaften und des standörtlichen Wasserhaushaltes. Sie sind damit als Vorhersagegrößen (Prädiktoren) bodenphysikalischer Kennwerte unverzichtbar.

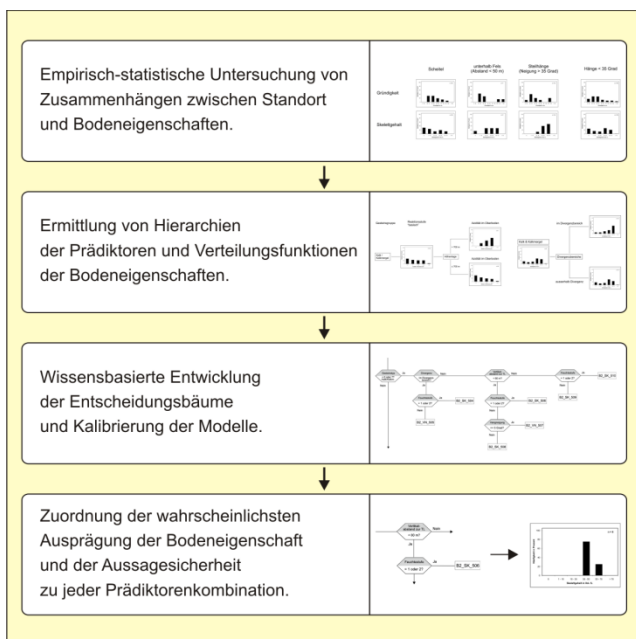


Abb. 1 Von Zusammenhängen zwischen Standort- und Bodeneigenschaften zum wahrscheinlichkeitsbasierten Modell.

Neben den Reliefparametern bilden Informationen u.a. zum Gestein, zur Waldgesellschaft und besonderen Bestandstypen, zur Höhenlage und weiteren Positionsparametern die Grundlage für die empirisch-statistische Untersuchung von Zusammenhängen zwischen Standortfaktoren und Bodeneigenschaften (s. Abb. 1).

Hieraus resultiert die Entwicklung der detaillierten Entscheidungsbäume auf Basis eines großen Umfangs von über 30 Prädiktoren. In den entscheidungsbasierten Modellen wer-

den empirisch-statistische Zusammenhänge mit Expertenwissen kombiniert. Es erfolgt die Zuordnung der wahrscheinlichsten Ausprägung der Bodeneigenschaften und der Aussagesicherheit zu jeder Prädiktorenkombination. Nicht prognostizierbare Flächen werden getrennt ausgewiesen.

Dieser unter ArcGIS umgesetzte Modellansatz lässt damit die Erstellung konzeptionell neuer digitaler Bodenkarten zu, die neben der hochauflösenden Vorhersage der Bodeneigenschaften auch deren Eintrittswahrscheinlichkeit flächendeckend ausweisen.

Die Gründigkeit- und Skelettgehaltsvorhersagen der wissensbasierten Modelle dienen zudem zur flächendeckenden Validierung halbautomatisierter Prognosen unter der Anwendung von Random Forest (BREIMAN 2001). Der Random Forest Algorithmus ist ein Klassifizierungs- und Regressionsmodell in Form eines Ensembles einzelner voneinander unabhängiger Entscheidungsbäume (BREIMAN & CUTLER 2004, LIAW & WIENER 2002).

Für die Bodeneigenschaften Gründigkeit und Skelettgehalt werden unabhängige Random Forest Modelle entwickelt. Als Prognose werden im Untersuchungsgebiet die Merkmalsausprägungen ausgewiesen, welche auf Basis der erlernten Klassifizierungsregeln am häufigsten auftreten.

Zur flächendeckenden Validierung der Random Forest Prognosen anhand der wahrscheinlichkeitsbasierten Modellergebnisse erfolgt neben einer Gegenüberstellung räumlicher Verteilungsmuster die lagebezogene Prüfung der Klassendifferenzen zwischen den Modellvorhersagen. Die abweichenden Modellergebnisse werden bilanziert. Die Umsetzung erfolgt unter ArcGIS.

Ergebnisse

Die wissensbasierten Prognosen im Pilotprojekt zeigen, dass je nach Bodeneigenschaft eine Vorhersage in vier bis fünf Ausprägungsklassen möglich ist. Der Anteil der nicht prognostizierbaren Flächen liegt dabei auf Grundlage des Datenbestandes in der Pilotphase für die Eigenschaften Gründigkeit, Vernässung und Azidität zwischen 2-10 %, für den Skelettgehalt im Jura bei ca. 25 %.

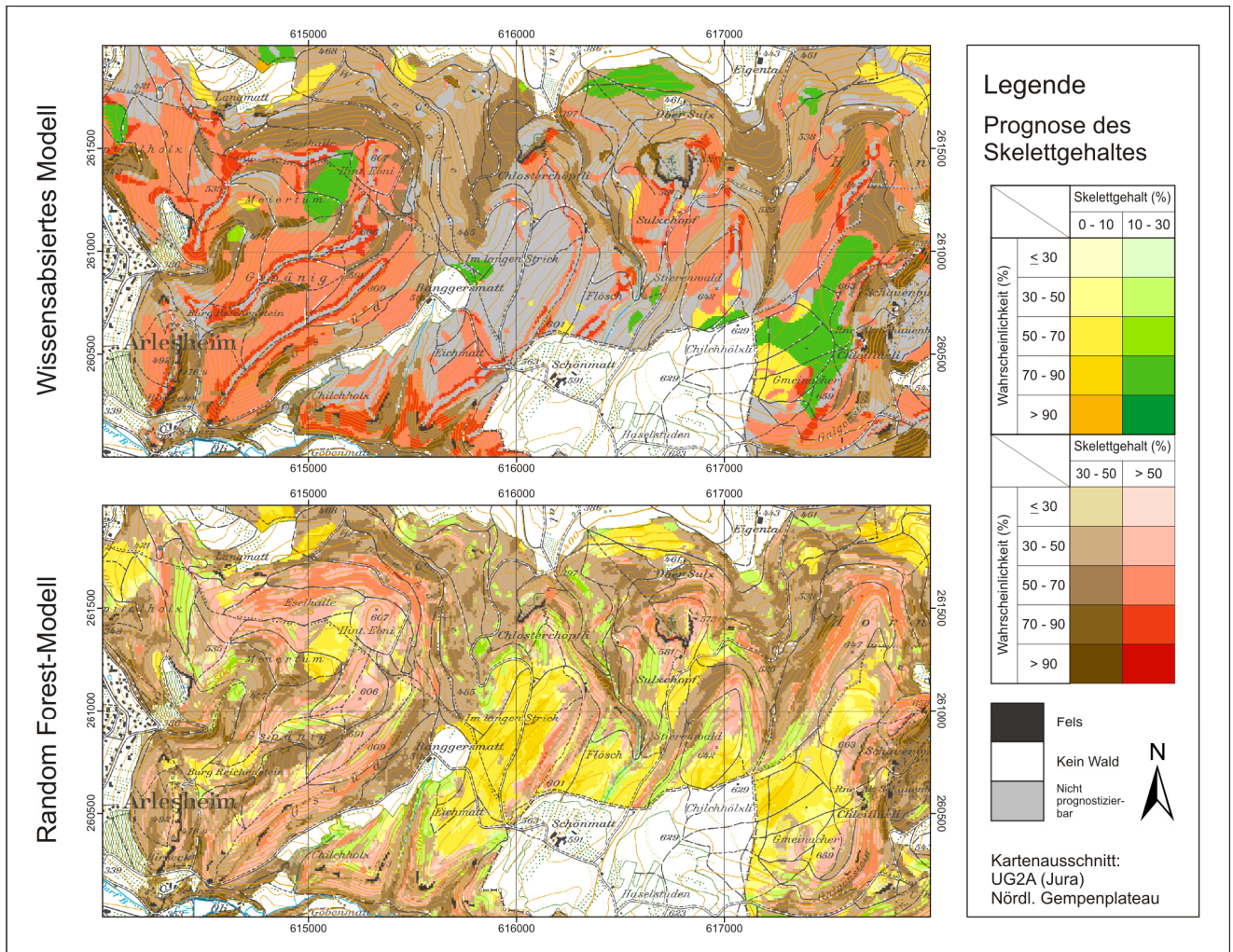


Abb. 2 Gegenüberstellung der Skelettgehaltsprognosen für ein Pilotgebiet im Tafeljura.

Nicht prognostizierbare Fälle treten v.a. auf, wenn für die Prädiktorenkombinationen unscharfe oder zu kleine Stichproben vorliegen oder wenn Felduntersuchungen belegen, dass bei bestimmten Standortbedingungen die Bodeneigenschaften nicht vorhersagbar sind (MOSIMANN ET AL. 2007).

Die Prüfung der Random Forest Modelle zeigt, dass die grundlegenden räumlichen Vorhersagemuster der expertenbasierten Modelle erkennbar sind. Großflächige Verteilungsmuster werden in sehr ähnlicher Struktur ausgewiesen.

Die Flächenanteile mit identischen Modellprognosen variieren im gesamten Modellierungsraum aber erheblich. Während die deutlichsten Übereinstimmungen im Lösshügelland festzustellen sind, weisen die Skelettgehaltsprognosen im Jura die geringste Übereinstimmung auf (s. Abb. 2).

Folgend werden die Ergebnisse anhand des Pilotgebietes mit den markantesten Modell-differenzen beispielhaft dargestellt.

Bei den wissensbasierten Modellen fallen die Prognosen der Bodeneigenschaften insgesamt deutlich strukturierter aus, als es bei Random Forest der Fall ist.

Die Random Forest Modelle weisen innerhalb der Großstrukturen häufig kleinflächig variierende Vorhersagen auf. Dabei sind nicht nachvollziehbare Hangmuster der prognostizierten Bodeneigenschaften feststellbar. Sie sind der Grund dafür, dass hier Differenzen der Modellprognosen verstärkt auftreten.

Im Gegensatz zu den wissensbasierten Vorhersagen, ist in solchen Bereichen die Prognosequalität der automatisierten Random Forest Modelle stark eingeschränkt.

Abb. 3 zeigt, dass der Anteil der Waldflächen mit identischen Vorhersagen von den Skelettgehaltsprognosen im Lösshügelland (> 98 %), über die Gründigkeitsprognosen (ca. 75 %), bis hin zu den Skelettgehaltsprognosen im Jura (ca. 40 %) deutlich abnimmt.

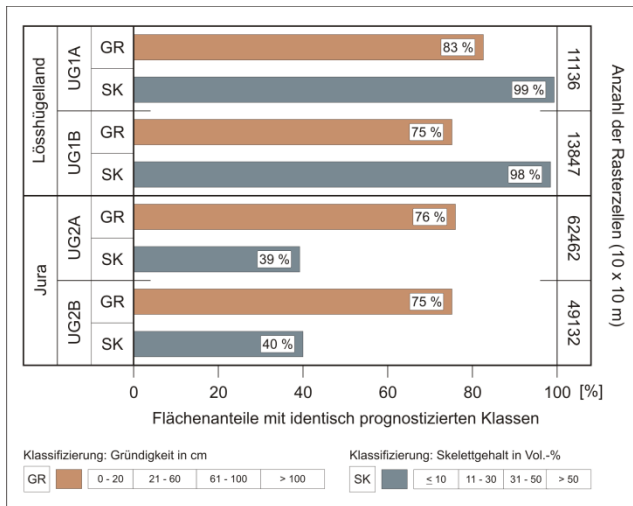


Abb. 3 Flächenanteile in den Pilotgebieten mit identischen Prognosen der wissensbasierten Modelle und Random Forest.

Als wichtige Ursachen sind hierfür zu nennen:

- Zu geringe Stichproben für einen Teil der Prädiktoren des Skelettgehalts im Jura.
- Hohe räumliche und Tiefenvariation der Skelettgehalte in den Gehängeschutten.
- Vielfältige Reliefstrukturen und kleinräumige Wechsel der Prädiktoren im Jura.

Die Ergebnisse verdeutlichen damit insgesamt, dass v.a bei den Skelettgehaltsprognosen im Jura von einer geringen lagebezogenen Qualität der Random Forest Vorhersagen ausgegangen werden muss.

Fazit

Die vorliegenden Ergebnisse belegen die grundsätzliche Anwendbarkeit der Random Forest Modelle, zeigen aber auch deren Grenzen auf.

Es hat sich gezeigt, dass die Qualität der Datenbasis ist für einen rein automatisierten Modellansatz zu gering ist. Damit liefert Random Forest auch Vorhersagen bei ungenügend gesicherten Zusammenhängen. Entsprechend wichtig sind die lagebezogenen Plausibilitätsprüfungen dieser Modellprognosen.

Die Random Forest Prognosen sind bisher nur in Kombination mit einem wissensbasierten Modell genügend zuverlässig. Deshalb sind weitere Feldaufnahmen zur gezielten Ergänzung der Stichprobe unerlässlich, um verbesserte Vorhersagen zu ermöglichen.

Literatur

BEHRENS, T. & T. SCHOLTEN (2006): Digital soil mapping in Germany – a review. In: Journal of Plant Nutrition and Soil Science, Vol. 169, S. 434 – 443.

BREIMAN, L. & A. CUTLER (2004): Random Forest Description. URL http://www.stat.berkeley.edu/users/breiman/RandomForests/cc_home.htm

BREIMAN, L. (2001): Random Forests. In: Machine Learning 45, Boston, S. 5 – 32.

LAGACHERIE, P., MCBRATNEY, A. & M. VOLTZ [HRSG.] (2007): Digital Soil Mapping – An Introductory Perspective. In: Developments in Soil Science, Vol. 33, Amsterdam, 600 S.

LIAW, A. & M. WIENER (2002): Classification and Regression by randomForest. In: R News, Vol. 2/3, S. 18 - 22.

MOSIMANN, TH., MEER, U. & J. GROß (2007): Dokumentation der Struktur der entscheidungsbasierten Modelle für die Vorhersage von Bodeneigenschaften. Unveröff. Dokumentation, Inst. f. Physische Geographie u. Landschaftsökologie, Hannover.

MOSIMANN, TH. (2005): Waldböden des Kantons Basel-Landschaft. Geosynthesis Sonderband, Inst. f. Physische Geographie u. Landschaftsökologie, Hannover, 269 S.