

Tagungsbeitrag: Kommission V der DGB
 Titel der Tagung: Böden verstehen, Böden nutzen, Böden fit machen.
 DBG Jahrestagung 3.-9. September 2011 in Berlin.

Mitteilungen der DGB (nicht begutachtete online-Publikation)

<http://www.dbges.de>

Robuste geostatistische Methoden zur räumlichen Analyse und Kartierung von Bodeneigenschaften

Keller¹ A., Rehbein¹ K., C. Schwierz² und A. Papritz²

1. Einleitung

Primäres Ziel des Projekts Bodeninformationssystem NABODAT ist die Zusammenführung aller in der Schweiz verfügbaren digitalen Bodenschadstoffdaten in eine Datenbank und die Verknüpfung der Datenbank mit einem GIS. Durch die zentrale und einheitlich strukturierte Bereitstellung der Daten können Bodeninformationen, z.B. über Bodeneigenschaften, in Zukunft leicht abgefragt werden. Die Ergebnisse der Abfragen können mit einfachen GIS-Funktionen visualisiert werden. Die Verknüpfung der Messdaten von einzelnen Messstellen mit einem GIS stellt zwar für die Kantone und andere Benützer der NABODAT einen bedeutenden Mehrwert dar, jedoch können mit den punktuellen Messungen der Bodeneigenschaften nicht ohne weiteres flächenhafte Aussagen gemacht werden. Flächenhafte Bodeninformationen sind allerdings erforderlich, um dem Boden bei den vielfältigen Vollzugs- und Planungsaufgaben im Gewässer- und Bodenschutz, der Raumplanung sowie anderen

fachverwandten Disziplinen fachlich gerecht zu werden. Bislang gibt es weder auf kantonaler noch auf Bundesebene geeignete Werkzeuge für die räumliche Analyse und Kartierung von digital vorliegenden Bodendaten. Es besteht daher der Bedarf geeigneter statistischer Methoden für themenbezogene, räumliche Auswertungen und für die flächenhafte Interpolation von Bodeneigenschaften zu entwickeln. Hierbei kommt der robusten statistischen Methode eine besondere Rolle zu.

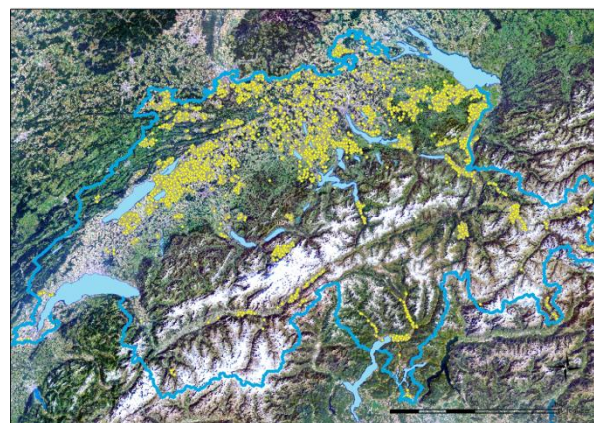


Abb. 1: Räumliche Verteilung der ca. 14.000 Bodenprofile, die im Zeitraum 1963-1996 von der damaligen FAP Zürich erhoben wurden.

Im Rahmen der räumlichen Auswertung stellen robuste (geo)statistische Verfahren sicher, dass ein Messstandort, der lokal stark von der restlichen Stichprobe abweicht, das Resultat einer regionalen Kartierung nicht verfälscht. Robuste statistische Verfahren haben gegenüber den Standardverfahren den Vorteil, dass einzelne, extreme Messwerte die Resultate von statistischen Analysen nicht stark beeinflussen. Robuste Verfahren erlauben zudem, abweichende Messungen zu identifizieren und damit nach den Ursachen zu suchen, welche für die Abweichungen verantwortlich sind. Damit erscheinen sie gut geeignet, um z. B. bei einer regionalen Kartierung Bodenmesswerte aus einer lokalen Belastung zu erkennen und angemessen zu berücksichtigen. Im Unterschied zu den relativ homogenen Medien Wasser und Luft variieren Bodendaten oft sehr stark und deshalb erscheint die Ver-

¹Nationale Bodenbeobachtung Schweiz (NABO); ART Agroscope, Zürich.

armin.keller@art.admin.ch

²Institut für terrestrische Ökosysteme, ETH Zürich

wendung von robusten (geo)statistischen Verfahren bei themenbezogenen, räumlichen Auswertungen besonders wichtig. Weil die Resultate tolerant gegen Anomalien in den Datensätzen sind, eignen sich robuste Verfahren auch für die halbautomatisierte Analyse von Daten und haben zudem den Vorteil, durch wenig erfahrene Benutzer und Benutzerinnen erfolgreich(er) angewendet werden zu können.

2. Methoden

2.1 Ansatz robustes Universal Kriging

In der Literatur finden sich zahlreiche Ansätze und Algorithmen für die robuste Auswertung von Umweltdaten. Eine Einführung in die robuste Statistik gibt unter anderem Maronna et al. (2006). Lark (2000) gibt in seinem Review eine Übersicht zur robusten Variogramm-Schätzung, zuvor wurde dies von einigen Autoren thematisiert (Cressie and Hawkings 1980; Genton 1998) bzw. später umgesetzt (Marchant and Lark 2007; Marchant et al. 2010).

Die Theorie und Entwicklung der robusten Universal Kriging Methode wurde von Künsch et al. (2010) an der ETH Zürich entwickelt und wird hier nicht im Detail beschrieben. Die Zielvariable Y am Ort s wird gemäss folgendem Ansatz mit einer Strukturmatrix (Hilfsvariablen) und einer autokorrelierten Gauss-Variablen mit Erwartungswert Null und Kovarianzmatrix modelliert.

$$Y(s) = x^T(s)\beta + Z(s) + \varepsilon(s)$$

- s : Ort
- $x(s)\beta$: Strukturdaten (Regression)
- $Z(s)$: autokorrelierte Gauss-Variablen mit $N(0, V_\theta)$
- $\varepsilon(s)$: unabhängiger Fehler (nugget)

Die Grundidee dieses Ansatzes besteht darin, dass das räumliche Variationsmuster der Zielvariablen anhand verfügbarer Umweltvariablen erklärt werden kann. Dies kann zum besseren Prozessverständnis und der Erklä-

rung von Zusammenhängen beitragen. Die Robustheit des mathematischen Ansatzes wird durch die Einführung einer Gewichtungsfunktion $\psi(r)$ für die Residuen $r = (y - X\beta - z)$ erreicht (siehe Maronna et al 2006). Die Schätzung der Regressionskoeffizienten sowie die Parameter der Variogrammschätzung erfolgt simultan.

Die Erfahrungen aus der Fallstudie „Digitale Kartierung von Schwermetallgehalten in den Böden des Kantons Thurgau“ (Rehbein und Keller, 2007) zeigen, dass die (geo)statistische Funktionalität von „state-of-the-art“ GIS-Software nicht ausreicht, um eine halbwegs befriedigende, geostatistische Analyse von Bodenschadstoffdaten durchzuführen. In der Fallstudie wurde deshalb neben der GIS- auch diverse Statistik-Software eingesetzt. Dabei stellte sich heraus, dass sich für die (geo)statistischen Analysen, die für NABODAT in Frage kommen, die Open Source Statistik-Softwareumgebung R (R Development Core Team, 2011) gut eignet.

2.2 Fallstudie Fribourg

Die robuste Methode wurde im Kanton Fribourg (1600 km²) für die Regionalisierung der organischen Substanz im Boden angewendet. Die Region erstreckt sich vom Schweizer Mittelland (400 m.ü.M) bis zu den Voralpen (1800 m.ü.M). Für die Fallstudie lagen Messdaten für drei verschiedene Tiefenbereiche (0-20 cm, 30-60 cm, 60-90 cm) von 250 Standorten des kantonalen Monitoring-Messnetzes Fribourg vor (Rossier et al. 2007). Es wurden die Messdaten für die organische Substanz 0-20 cm der vierten Erhebungsrunde im Monitoring-Messnetz verwendet, welche im Zeitraum 2001 bis 2006 erhoben wurden (Abb. 2). Das Messnetz repräsentiert in etwa ein 2 x 2 km² Gitter für die landwirtschaftliche Nutzfläche von etwa 1000 km². Mehrheitlich werden die Dauerbeobachtungsflächen genutzt für den Ackerbau (n=136), Dauergrünflächen (67) oder in höheren Lagen als

Alpweide (47). Zusammenhänge zwischen der organischen Bodensubstanz an den Monitoring-Standorten und räumlichen Hilfsvariablen wie Landnutzung (Arealstatistik), Klimadaten (Niederschlag, Jahresmitteltemperatur), Höhe, Terrain Attribute (Exposition) sowie Informationen aus der Bodeneignungskarte 1:200.000 für 100 m x 100 m Raster modelliert und als erklärende Variablen in einem robusten Universal Kriging Ansatz verwendet. Weitere erklärende Variable können aus weiteren Datenquellen aufbereitet werden (z.B. Remote Sensing).

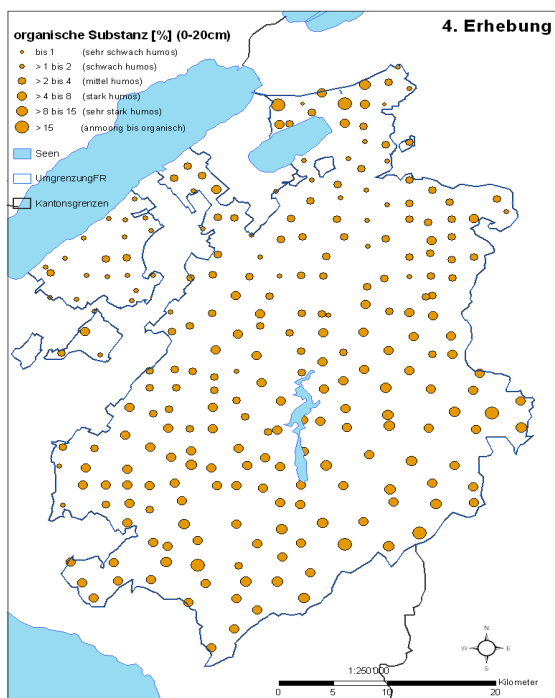


Abb. 2: Verteilung der Messwerte der organischen Substanz im Oberboden (0-20 cm) im kantonalen Bodenmessnetz Fribourg.

3. Ergebnisse

Für die Trendschätzung mit dem Regressionsmodell zeigten die Höhe (oder Niederschlag), die Exposition, die Landnutzung sowie physiografische Klassen der Bodeneignungskarte einen statistisch signifikanten Zusammenhang mit der organischen Substanz im Oberboden (Abb. 3). Das Regressionsmodell erklärte etwa Zweidrittel der räumlichen Variation der Zielvariable ($R^2 = 0.6$ bis 0.7). Das experimentelle Variogramm konnte mit einem spherical Modell am Besten modelliert werden.

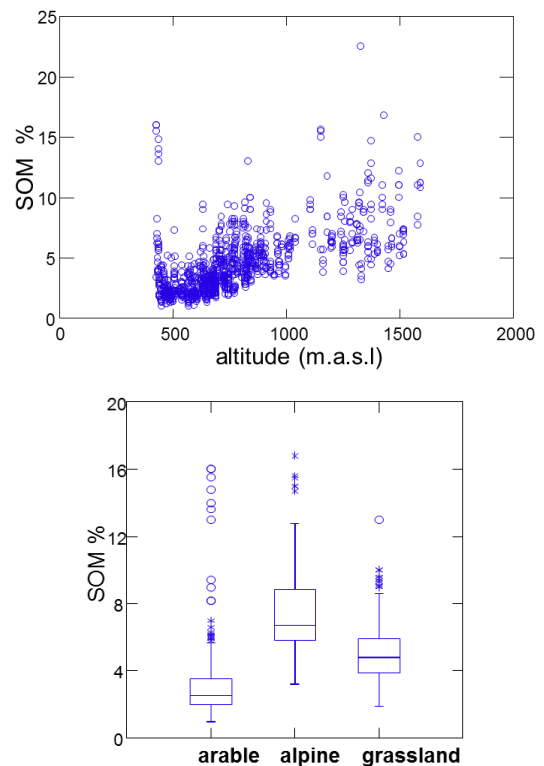


Abb. 3: Explorative Analyse der Hilfsvariablen zur Trendmodellierung der organischen Substanz (SOM) im Oberboden, gemessen an 250 Dauerbeobachtungsflächen

Die Korrelation der organischen Substanz im Boden mit der Höhe spiegelt zugleich die enge Beziehung zur Jahresmitteltemperatur und dem Jahresniederschlag wieder. Abbildung 3 zeigt zudem deutlich die Unterschiede der Gehalte für die unterschiedlichen Landnutzungen. In Abbildung 4 sind die interpolierten Schätzwerte für das 1 ha Raster dargestellt.

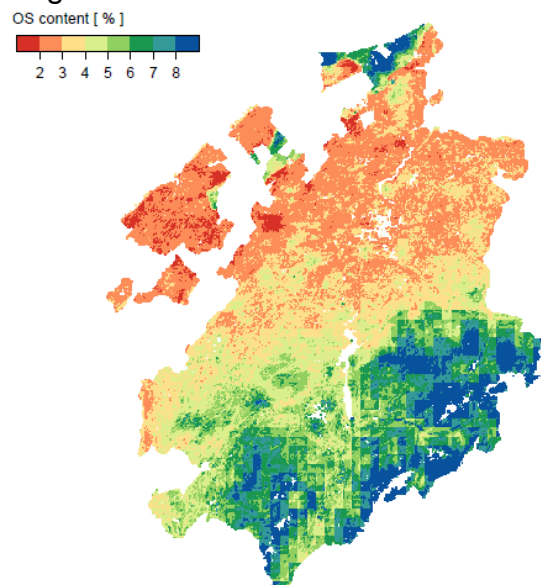


Abb.4: Schätzwerte der organischen Substanz im Oberboden (1 ha Raster, spherical Modell)

Auf Basis der explorativen Analyse des Datensatzes und der Analyse der Zusammenhänge der Zielvariable mit den erklärenden Variablen und der räumlichen Struktur der Residuen kann die Gewichtungsfunktion für die Begrenzung des Einflusses extremer Residuen gewählt werden. Die Berechnung wurde für verschieden starke Beschränkungen extremer Residuen durchgeführt. Abbildung 5 zeigt die Auswirkung der robusten Methode gegenüber der klassischen Kriging-Methode auf (Verhältnis Schätzwert klassisch zu Schätzwert robust).

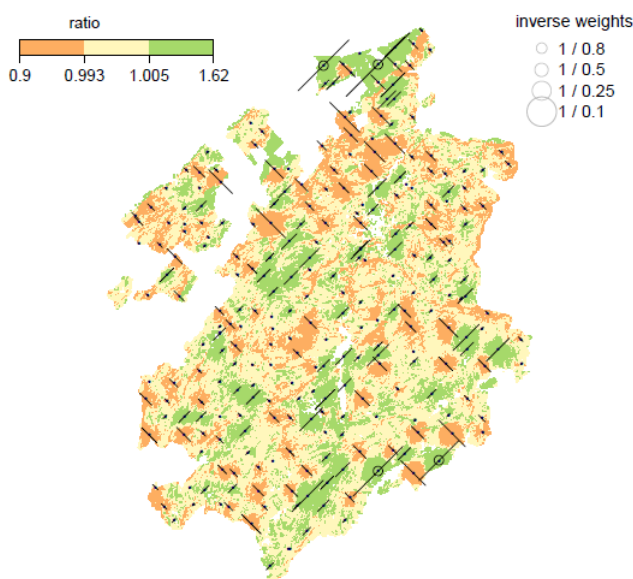


Abb.5: Vergleich der robusten mit der klassischen Universal-Kriging Methode. Der Einfluss von Messungen mit hohen Residuen wird in der robusten Method beschränkt (grosse Kreise). (ratio = klassisch / robuste Schätzmethode)

Für Gebiete mit gelbem Hintergrund zeigten beide Verfahren quasi gleiche Schätzwerte für die organische Substanz, während für die grünen Bereiche die klassische Methode die Gehalte überschätzt und für die orangen Bereiche unterschätzt.

5. Zusammenfassung

Die von Künsch et al. (2010) entwickelte robuste geostatistische Methode erlaubt die räumliche Analyse von Zusammenhängen mit der Zielvariablen (Strukturmatrix). Die Schätzung der Regressionskoeffizienten der räumlichen Trendmodellierung erfolgt simul-

tan mit der robusten Schätzung der Kovarianzfunktion. Der Einfluss lokaler „Extremwerte“ kann hierbei wahlweise mehr oder weniger stark mit einer Gewichtungsfunktion eingeschränkt werden. Vor dem Hintergrund heterogener Datensätze für Bodeneigenschaften, welche aus unterschiedlichen Datenquellen stammen, liefert die robuste Schätzmethode zuverlässigere Schätzwerte (und Schätzvarianzen) als die klassische Kriging-Methode.

Schlüsselworte: Geostatistik, robust, Universal Kriging, Bodeneigenschaften, digitale Kartierung, Digital Soil Mapping.

Cressie, N. and Hawkins, D. M. (1980). Robust estimation of the variogram: I. *Mathematical Geology*, 12, 115–125.

Genton, M. (1998). Highly robust variogram estimation. *Mathematical Geology*, 30(2), 213–220.

Marchant, B.P. and Lark, R. M. (2007). Robust estimation of the variogram by residual maximum likelihood. *Geoderma*, 140, 62–72.

Marchant, B.P., N.P.A. Saby, R.M. Lark, P.H. Bellamy, C.C. Jolivet and D. Arrouays. (2010). Robust analysis of soil properties at the national scale: cadmium content of French soils. *European Journal of Soil Science*, February 2010, 61, 144–152.

Rossier, N., J. Altermath, T. Niggli. (2007). FRI-BO – Réseau fribourgeois d'observation des sols. Institut agricole de l'Etat de Fribourg. Grangeneuve. 1725 Posieux. Suisse. S. 156.

Maronna, R.A., Martin, D. R. und Yohai, V. J. 2006. *Robust Statistics Theory and Methods*. John Wiley & Sons, Chichester.

R Development Core Team. 2011. <http://www.R-project.org>, Wien.

Rehbein, K. und Keller, A. 2007. Digitale Kartierung von Schwermetallgehalten in den Böden des Kantons Thurgau. Teil 1 Methodik. Berichtsentwurf. Forschungsanstalt Agroscope Reckenholz-Tänikon ART, Zürich.

Lark, R.M. (2000). A comparison of some robust estimators of the variogram for use in soil survey. *European Journal of Soil Science*. 51 : 137-157.

Künsch H.R., Papritz A., Schwierz C., and W.A. Stahel. (2010). Robust estimation of the external drift and the variogram of spatial data. ETH Zürich. Seminar für Statistik. 8092 Zürich. Switzerland.