SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

Evaluation of digital soil mapping approaches with large sets of environmental covariates

Madlene Nussbaum[1], Kay Spiess[1], Andri Baltensweiler[2], Urs Grob[3], Armin Keller[3], Lucie Greiner[3], Michael E. Schaepman[4], and Andreas Papritz[1]

[1]Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, Universitätstrasse 16, CH-8092 Zürich, Switzerland
[2]Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland
[3]Research Station Agroscope Reckenholz-Taenikon ART, Reckenholzstrasse 191, CH-8046 Zürich, Switzerland
[4]Remote Sensing Laboratories, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

Correspondence to: M. Nussbaum (madlene.nussbaum@env.ethz.ch)

Abstract.

Spatial assessment of soil functions requires maps of basic soil properties. Unfortunately, these are either missing for many regions or are not available at the desired spatial resolution or down to required soil depth. Field based generation of large soil data sets and of conventional soil maps remains costly. Meanwhile, soil legacy data and comprehensive sets of spatial environmental data are available for many regions.

Digital soil mapping (DSM) approaches – relating soil data (responses) to environmental data (covariates) – are facing the challenge to build statistical models from large sets of covariates originating for example from airborne imaging spectroscopy or multi-scale terrain analysis. We evaluated six approaches for DSM in three study regions in Switzerland (Berne, Greifensee, ZH forest) by mapping effective soil depth available to plants (SD), pH, soil organic matter (SOM), effective cation exchange capacity (ECEC), clay, silt, gravel content and bulk density for four soil layers (totalling 48 responses). Models were built from 300-500 environmental covariates by selecting linear models by 1) grouped lasso and by an ad-hoc stepwise procedure for 2) robust external-drift kriging (EDK). For 3) geoadditive models we selected penalized smoothing spline terms by component-wise gradient boosting (geoGAM). We further used two tree-based methods: 4) boosted regression trees (BRT) and 5) Random Forest (RF). Lastly, we computed 6) weighted model averages (MA) from predictions obtained from methods 1–5.

Lasso, georob and geoGAM successfully selected strongly reduced sets of covariates (subsets of 3-6 % of all covariates). Differences in predictive performance, tested on independent validation data, were mostly small and did not reveal a single best method for 48 responses. Nevertheless, RF was on average often best among methods 1–5 (28 of 48 responses), but was outcompeted by MA for 14 of these 28 responses. RF tended to over-fit the data. Performance of BRT was slightly worse than RF. GeoGAM performed poorly on some responses and was only best for 7 of 48 responses. Predictive precision of lasso was intermediate. All models generally had small bias. Only the computationally very efficient lasso had slightly larger bias likely because it tended to under-fit the data. Summarizing, although differences were small, the frequencies of best and worst performance clearly favoured RF if a single method is applied MA if multiple prediction models can be developed.

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

1 Introduction

Human well-being depends on numerous services that soils provide in agriculture, forestry, natural hazards, water protection, resources management and other environmental domains. The capacity of soils to deliver services is largely determined by its functions, e.g. regulation of water, nutrient and carbon cycles, filtering of compounds, production of food and biomass or providing habitat for plant species biodiversity and soil fauna (Haygarth and Ritz, 2009; Robinson et al., 2013). The assessment of the multi-functionality of soils commonly depends on soil datasets of its chemical, physical and biological properties (Calzolari et al., 2016). Greiner et al. (in rev.) compiled a set of approved assessment methods for soil functions from the applied soil science community that cover the multi-functionality of soils (Table 1). This set of soil functions can be assessed with 12 basic soil properties. Unfortunately, spatial assessment of soil functions is often hindered because precise maps of soil properties are missing in many countries of the world (Hartemink et al., 2013; Rossiter, 2016). However, for many regions legacy data on soil properties (responses) and comprehensive spatial environmental data (covariates) are available and can be linked by digital soil mapping (DSM) techniques (e.g. McBratney et al., 2003; Scull et al., 2003).

Many recent DSM studies used relatively small sets of no more than 30 covariates (e.g. Adhikari et al., 2013; Vaysse and Lagacherie, 2015; Were et al., 2015; Mulder et al., 2016; Lacoste et al., 2016; Taghizadeh-Mehrjardi et al., 2016; Yang et al., 2016). Geodata availability and deemed importance often determine what covariates are used for DSM. However, Brungard et al. (2015) showed that a priori preselection of covariates using pedological expertise might result in a decreased precision of soil class predictions. Using comprehensive environmental geodata for DSM likely improves predictive precision because soil forming factors are better covered using more covariates. Derivatives of geological or legacy soil maps (Nussbaum et al., 2014), multi-scale terrain analysis (Behrens et al., 2010a, b, 2014; Miller et al., 2015), wide ranges of climatic parameters (Liddicoat et al., 2015) and imaging spectroscopy (Mulder et al., 2011; Poggio et al., 2013; Fitzpatrick et al., 2016) all contribute to generate high-dimensional sets of partly multi-collinear covariates. We presume that DSM techniques benefit from such large number of covariates. A versatile DSM model building strategy faces therefore the challenge to deal with (very) large covariate sets. If, in addition, many responses have to be mapped, such an approach should

1. efficiently build models without much user interaction, even if there are more covariates p than observations n ($n < p$),
2. cope with numerous multi-collinear and likely noisy covariates,
3. result in predictions with good precision and
4. avoid over-fitting the calibration data.

Besides, the method should fulfil basic DSM requirements like modelling nonlinear and nonstationary relations between response and covariates, considering spatial autocorrelation, allowing to check pedological plausibility of the relationships between soil properties and soil forming factors and modelling predictive uncertainty.

DSM approaches used in the past can broadly be grouped in 1) linear regression models (LM), 2) variants of geostatistical approaches, 3) generalized additive models (GAM), 4) methods based on single trees like classification and regression trees

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

(CART), 5) machine learning methods as support vector machines (SVM) or artificial neural networks (ANN), 6) ensemble machine learners like boosted regression trees (BRT) or random forest (RF), and 7) averaging predictions of any of the mentioned methods (model averaging, MA).

LM (e.g. Meersmans et al., 2008; Wiesmeier et al., 2013) can not be fitted for $n < p$ and estimates of coefficients become unstable with collinear covariates. Liddicoat et al. (2015) and Fitzpatrick et al. (2016) used lasso (least absolute shrinkage and selection operator), a form of penalized LM suitable for large correlated covariate sets. Lasso models only linear relations between response and covariates, but performs model building computationally very efficiently. Fitzpatrick et al. (2016) found that different stepwise LM selection procedures were clearly outperformed by lasso. Geostatistical approaches are generally popular in DSM (McBratney et al., 2003), and they have clear advantages over other methods: They allow for change-of-support and predictive uncertainty that follows straightforward from the kriging variances. Similar to LM external-drift kriging (EDK) requires a parsimonious linear trend model. Nussbaum et al. (2014) used lasso for initial covariate selection, but subsequent manual model building steps were needed. Lacoste et al. (2016) found that even a computationally optimized form of kriging was not suitable for fitting small sets of covariates p on numerous observations n. Nonlinear additive modelling through GAM also relies on covariate selection for stable trend estimation. Poggio et al. (2013) used a covariate selection procedure with a random component, and Nussbaum et al. (in rev.) applied componentwise gradient boosting to preselect relevant covariates. Unless combined with either lasso or boosting, large sets of covariates are difficult to process by LM, EDK or GAM. But these methods allow for simple model interpretation by partial effects plots (Faraway, 2005, p. 73)

Generally, more complex approaches seem to yield more precise predictions than simpler DSM methods (Brungard et al., 2015). Tree-based methods model effects of interactions between covariates on responses. Single trees (CART, see references in McBratney et al., 2003) tend to be noisy (large variance), but have small bias. Cubist, an extension of CART with LM at the terminal nodes of the tree, was so far mostly applied to small covariate sets (e.g. Adhikari et al., 2013; Lacoste et al., 2016; Mulder et al., 2016). Forming ensembles of trees aims at reducing their variance. BRT combines trees in a stagewise forward manner by componentwise gradient boosting, and RF averages de-correlated fully grown trees. RF seems stable for large sets of covariates (Behrens et al., 2010a, b, 2014). BRT was compared to RF by Yang et al. (2016) yielding similar model precision.

SVM outperformed LM and cubist in a study by Somarathna et al. (2016), where SVM were further improved by a spatial ensemble approach. SVM and ANN were compared by Were et al. (2015) and Taghizadeh-Mehrjardi et al. (2016): They were about equally good and outperformed the ensemble learner RF. Li et al. (2011), on the other hand, found RF to outcompete SVM and were able to improve RF by residual kriging. Vaysse and Lagacherie (2015) also found that subsequent kriging of residuals improved RF.

Lastly, averaging predictions from different models (MA) can be seen as a way of dimension reduction. Building new models by combining predictions by several methods possibly reduces prediction variance (Hastie et al., 2009, pp. 288). Malone et al. (2014) explored different weighing strategies for MA, but it became not clear from the study whether MA was indeed better than predictions by a single method because predictions were not validated by independent data. Li et al. (2011) averaged only predictions computed by the best performing (very similar) models, and this did result in no advantage of MA over single models.

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

All of the studies mentioned above used only small sets of covariates ($p < 30$), tested only few approaches or, with exception of Vaysse and Lagacherie (2015), did not extend the evaluation to several soil properties or study regions. It is therefore currently not clear how well models can be built form large covariate sets by popular DSM methods. Empirical evidence is still too limited to rate DSM methods with respect to the criteria 1–4 listed above. In particular, it is not known whether methods can be identified that are more prone to over-fit soil data or that yield precise predictions more often than others.

The objectives of this study were to evaluate for a rather broad choice of currently often used DSM methods how well they cope with requirements 1–4 listed above. We compared in our study a) lasso, b) robust EDK (georob), c) spatial GAM with model selection based on boosting (geoGAM), two ensemble tree-methods d) BRT and e) RF as well as f) weighted MA. In more detail, our objectives were to

i) automatically build models by methods (a)–(e) and compute MA of (a)–(e) for numerous responses from large sets of covariates (300–500),

ii) evaluate predictive performance of these models with independent validation data,

iii) evaluate over-fitting behavior and practical usage of approaches,

iv) and, lastly, to briefly compare accuracies of DSM predictions and predictions derived from a legacy soil map 1:5 000.

We focused on three study regions in Switzerland: A forested region and two regions covering agricultural land, where harmonized legacy soil data and in latter airborne imaging spectrometer data were available. For the agricultural land soil properties required for assessing regulation, habitat and production functions were mapped (Table 1). For forests we had less diverse soil data and mapped only properties to assess acidification status (Zimmermann et al., 2011).

2 Materials

2.1 Study regions

We chose three study regions on the Swiss Plateau with contrasting patterns regarding land use, geology, soil types and availability of airborne remote sensing images (Fig. 1, Table 2). Agricultural land north of the city of Berne and around Lake Greifensee in the Canton of Zurich was selected within the outline of imaging spectroscopy data gathered by the APEX spectrometer in the years 2013 and 2014 (Schaepman et al., 2015). Agricultural land was defined as the area not covered by any areal features extracted from the Swiss topographic landscape model (swissTLM3D, Swisstopo, 2013a), hence wetlands, forests, parks, gardens and developed areas were excluded.

The majority of Berne study region (80 %) was covered by crop land and 15 % by permanent grassland. In the Greifensee region crop land covered roughly half of the area and one third was permanent grassland. The remaining areas were orchards, vineyards, horticultural areas or mountain pastures (Hotz et al., 2005).

The third study region comprise forested areas of the Canton of Zurich (ZH forest) derived from the topographic landscape model (swissTLM3D, Swisstopo, 2013a). Two thirds of the forested area are dominated by conifers (FSO, 2000b). In all three

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

Table 1. Basic soil properties needed for spatial soil function assessment in the three study regions of the PMSoil project, which was part of the Swiss National Research Programme "Sustainable use of Soil as a Resource" (NRP68). Most soil functions required data on further, expensive-to-measure soil properties that were inferred by pedotransfer functions (PTF) from the basic soil properties (see Greiner et al., in rev., BD: soil density, SOM: soil organic matter, d_w depth of stagnic or gleyic horizon, d_c drainage class [d_w and d_c modelled in Nussbaum et al., in rev.], BS: base saturation, ECEC: effective cation exchange capacity, BC/Al: ratio of sum of basic cations to aluminium).

Static soil(sub)function	clay	silt	gravel	BD	SOM	pH	ECEC	BS	BC/Al	SD	d_w	d_c
Regulation function												
Capacity of a soil for water infiltration and storage (Danner et al., 2003)	*	*	*	*[3]	*					*	*	
Nutrient cycling (Lehmann et al., 2013)	*	*	*	*[3]	*	*				*	*	
Binding capacity for inorganic contaminants (DVWK, 1988)	*		*	*[3]	*	*				*		
Binding and decomposition capacity of a soil for organic contaminants (Litz, 1998)	*	*	*	*[3]	*	*				*	*	*
Filtering of pollutants and acidity buffering (Bechler and Toth, 2010)	*		*	*[3]	*	*				*	*	
C storage (SOC-stock to 1 m soil depth, Greiner et al. unpublished)			*	*[3]	*					*		
Capacity for plant nutrient retention (against percolation and overland flow, Jäggli et al., 1998)	*	*	*	*[3]	*					*	*	*
Acidity state of forest soils, resilience to acidification and risk of aluminum toxicity (Zimmermann et al., 2011)			*[1]	*		*	*	*[2]	*[2]			
Habitat function												
Soils with extreme properties allowing a non-standard plant community (Siemer et al., 2014)	*	*		*[3]	*	*				*	*	
Habitat for plants (Greiner et al., unpublished)	*	*		*[3]	*	*				*		
Production function												
Agricultural production (Jäggli et al., 1998)	*	*	*	*[3]	*	*				*	*	*

[1] Only 50 sites with gravel estimates available, mean content per soil layer used.

[2] Limited data for BS and BC/Al (topsoil 300, subsoil 210 sites), no independent validation possible, therefore not included in this publication.

[3] For Berne and Greifensee computed by PTF which used SOM to predict BD.

Table 2. Description of three study regions (a: area, h: elevation, Swisstopo, 2013b; p: mean annual precipitation; t: mean annual temperature, Zimmermann and Kienast, 1999).

name	land use	a [km^2]	h [m]	p [mm]	t [°C]
Berne	agriculture	235	430–910	960–1440	6.8–9.3
Greifensee	agriculture	170	390–840	1040–1590	7.5–9.1
ZH forest	forest	507	340–1170	880–1780	6.1–9.1

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

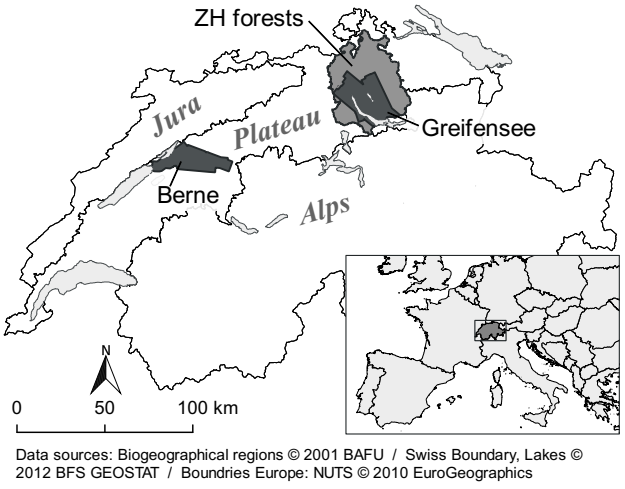

Figure 1. Location of study regions Berne and Greifensee (agricultural soils) and Canton of Zurich (forest soils).

study regions soils formed mostly on weathered Molasse formations and quaternary sediments dominantly from last glaciation typical for the Swiss Plateau. In the northeastern part Jura foothills with limestone rocks reach into ZH forest (Hantke, 1967). In the western part of the Berne study region alluvial plains with silty sediments or peat formations prevail (Swisstopo, 2005).

Soils are rather young in all study regions ($< 20\,000$ years old) as they mostly formed after the end of the last glaciation. Typical soils are Cambisols and Luvisols (calcaric to dystric), Gleysols and Fluvisols (reflecting frequent wet conditions) and Histosols (on former peatlands). Shallower soils are often Regosols (FSO, 2000a).

2.2 Soil data

2.2.1 Origin of soil data and data harmonization

We gathered and harmonized legacy soil data from various soil surveys performed between 1960 and 2014. In Berne data was collected mostly before 1980 in local soil mapping projects for land improvement. Data for Greifensee and ZH forest originate from long-term soil monitoring of the Canton of Zurich (KaBo), a soil pollutant survey (Wegelin, 1989), field surveys for creating soil maps of the agricultural land (scale 1:5 000, Jäggli et al., 1998) or soil investigations in the course of forest vegetation surveys by the Swiss Federal Institute for Forest, Snow and Landscape Research (WSL, Walthert et al., 2004). Hence, the compiled soil database comprised data of soil properties that were measured or estimated for pedogenic soil horizon at soil profiles or measured at fixed depth layers from bulked soil samples. Sites for pollution surveying were chosen on a regular grid. The remaining sites were selected purposively by field surveyors to best represent soils typical for the given landform. The sites of WSL were chosen purposively according to the aims of the project. Collating these soil data from different sources implicated that soil data were not directly comparable, and tailored harmonisation procedures were required to provide consistent soil datasets. The heterogeneity of soil legacy data resulted among others from several standards of soil

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

description and soil classification, different data keys, different analytical methods and in particular, often missing metadata for a proper interpretation of the datasets. Therefore, we elaborated a general harmonisation scheme that covers performance steps required to merge different soil legacy data into one common consistent database (Walthert et al., 2016). Sampling sites were recorded in the field on topographic maps (scale 1:25 000), hence the accuracy of the coordinates is about $\pm$ 25 m.

Horizon-based (and non-fixed depth) soil property data was converted to common depth-layers of 0–10, 10–30, 30–50 and 50–100 cm soil depth for Berne and Greifensee and of 0–20 and 40–60 cm depth for ZH forest. The latter intervals were chosen because at the majority of forest sites only these layers had been sampled. Values for layer t were computed from horizon (or layer) data of o_i by

$$o_t = \sum_{i=1}^{h} w_i \, o_i, \tag{1}$$

with w_i given by the product of the fraction of the thickness of horizon/layer i within t and its soil density ρ_i. The w_i were normalized to sum to 1. Because we lacked estimates of volumetric gravel content for the majority of samples we assumed that it was constant. ρ_i was partly derived by pedotransfer functions (PTF, Table S1 in Supplement).

Soil properties were either measured by standard laboratory procedures, estimated in the field or calculated by pedotransfer functions (PTF, see overview in Table S1 in Supplement). We accounted for temporal changes over the long period over which the data had been collected and for possible differences between laboratory measurements, field estimates and PTF predictions by statistical modelling. We included categorical covariates (factors) in the statistical models that coded separately for laboratory measurements, field estimates and PTF predictions the period when the data had been gathered. For Berne three periods (years 1968–1974, 1975–1978 and 1979–2010) were coded separately for laboratory measurements and field estimates. For Greifensee and ZH forest coding required more care because we had replicate samples from soil monitoring. Instead of only using mean or median values per site this coding allowed us to use all individual observations. For Greifensee we coded the years of 1960–1989, 1990–1994 and 1995–1999 separately for laboratory and field data and 2000–2014 for laboratory measurement only. For ZH forest we distinguished the periods 1985–1994, 1995–1999, 2000–2004, 2005–2009, 2010–2014 for laboratory measurements and further two levels for predictions by PTF or pH measurements on field-moist samples (see Table S1 in Supplement). Older (or newer) data on pH, soil organic matter (SOM) and effective cation exchange capacity (ECEC) than reported above was discarded. To compute model predictions for mapping we used the most recent time period and laboratory measurements as reference level.

2.2.2 Soil properties

For the agricultural land (Berne, Greifensee) we modelled clay and silt, gravel content, pH, SOM and effective soil depth available to plants (SD) and for ZH forest ECEC, pH and density of the fine soil fraction ($\leq$ 2 mm, BD). For Berne and Greifensee (possibly incomplete) soil data was available for 1052 and 2050 sites respectively, and for ZH forest we had 2379 sites with soil data (Fig. S1 to S3 in Supplement). We used roughly 20 % of the sampled sites for independent model validation. Depending on data availability, this resulted in 120–300 validation sites that were chosen by weighted random sampling. We

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

ensured an even distribution of validation sites over the study regions by assigning to each site a sampling weight that was proportional to respectively the forested and agricultural area within its Dirichlet polygon (Dirichlet, 1850).

Models for properties of agricultural soils were calibrated with data of 700–900 sites. Only for SOM there were more topsoil sites available (1140), but in the subsoil we had only data of 400 (Greifensee) and 530 (Berne) sites, respectively. For ZH forest topsoil chemical properties were available for 1055 (ECEC) to 1470 (pH) sites, but for subsoils data was again scarce (ECEC 380 and pH 690 sites). For modelling BD we had only 550 (topsoil) to 370 (subsoil) sites. On average we calibrated the models with the following spatial data densities: Berne 2.9–3.6, Greifensee 4.2–5.1 and ZH forest 1.2–1.8 observations per km^2.

Tables S3 to S7 in Supplement report per depth layer descriptive statistics of all soil properties. In general, soils in the Greifensee region were richer in clay (mean clay content 26 %) than in Berne (17–19 %) and had larger gravel content (8–13 % vs. 3–5 %). In both agricultural study regions, large SOM content was occasionally found (> 40 %) as drained organic soils were sampled at some sites. Topsoil pH showed in Berne and Greifensee similar variation (mean of 6.3–6.7 and standard deviation of 0.7–0.9), because agricultural management likely evens out pedogenic differences. ZH forest soils were more acid (mean topsoil pH 4.7) and pH varied more strongly (minimum pH 2.6).

2.3 Covariates for statistical modelling

To represent soil forming factors we used data from 28 sources, totalling to roughly 480 covariates for Berne and Greifensee and 330 for ZH forest where APEX imaging spectrometer data was not available (Table 3 and Table S2 in Supplement). Exact numbers of covariates used depended on soil properties. When sampling density of soil data was small we excluded covariates that showed hardly any spatial variation (e.g. coarse-gridded climate data) or that resulted only in few data points per factor level. Wherever possible, we aggregated factor levels based on pedological knowledge to obtain at least 20 observations per level.

3 Methods

The large number of responses – 21 for each of Berne and Greifensee, 6 for ZH forest – and of covariates (Table 3) required that statistical models could be automatically built without user interaction. Hence, we used five approaches: lasso (Sect. 3.1) and robust external-drift kriging (georob, Sect. 3.2), geoadditive modelling (geoGAM, Sect. 3.3) as well as two tree-based machine learning procedures (boosted regression trees [BRT], Sect. 3.4 and random forest [RF], Sect. 3.5). The predictions by the five methods were moreover combined by weighted averaging (MA, Sect. 3.6).

For parametric methods (Sect. 3.1 to 3.3) we transformed strongly positively skewed responses $Y(s)$ (see Tables S4, S6 and S7 in Supplement for skewness). Transformation by natural logarithm was applied to soil organic carbon (SOM) and effective cation exchange capacity (ECEC) while gravel content was transformed by square root (sqrt). Predictions of log-transformed data were unbiasedly backtransformed according to Cressie (2006, Eq. (20), see also Nussbaum et al., in rev.) and for sqrt-transformed data we used

$$\tilde{Y}(s) = \hat{f}(\mathbf{x}(\mathbf{s}))^2 + \hat{\sigma}^2 - \text{Var}[\hat{f}(\mathbf{x}(\mathbf{s}))] \quad (2)$$

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

Table 3. Overview of geodata sets and derived covariates (for more information see Table S2 in Supplement, r: pixel size for raster datasets or scale for vector datasets, a: limited to study region Be: Berne, Gr: Greifensee or Zf: ZH forest, n: number of covariates per dataset, NDVI: normalized differenced vegetation index, TPI: topographic position index, TWI: topographic wetness index, MRVBF: multi-resolution valley bottom flatness).

geodata set	r	a	n	covariate examples
Soil				physiographic units, historic wetlands, presence of drainage networks or soil amelioration
Soil overview map (FSO, 2000a)	1:200 000		8	
Wetlands Wild maps (ALN, 2002)	1:50 000	Gr	1	
Wetlands Siegfried maps (Wüst-Galley et al., 2015)	1:25 000	Gr	1	
Agricultural suitability (LANAT, 2015)	1:25 000	Be	1	
Anthropogenic soil interventions (AWEL, 2012)	1:5 000	Gr	1	
Drainage networks (ALN, 2014b)	1:5 000	Gr	2	
Parent material				(aggregated) geological units, ice level during last glaciation, aquifers, areas suitable for gravel exploitation
Geological overview map (Swisstopo, 2005)	1:500 000	Be	4	
Map of last glacial maximum (Swisstopo, 2009)	1:500 000		1	
Geotechnical map (BFS, 2001; BAFU and GRID-Europe, 2010)	1:200 000		2	
Geological map (ALN, 2014a)	1:50 000		7	
Geological maps (Swisstopo, 2016), roughly harmonized	1:25 000	Be	1	
Groundwater occurrence (AWEL, 2014; AWA, 2014b)	1:25 000	Gr	2	
Hydrogeological infiltration zones (AWA, 2014a)	1:25 000	Be	2	
Mineral raw materials (AGR, 2015)	1:25 000	Be	1	
Climate				mean annual/monthly temperature and precipitation, radiation, continentality index, site water balance, NH_3 concentration in air
MeteoSwiss 1961–1990 (Zimmermann and Kienast, 1999)	25/100 m		33	
MeteoTest 1975–2010 (Remund et al., 2011)	250 m		38	
Air pollutants (BAFU, 2011)	500 m	Zf	2	
NO_2 immissions (AWEL, 2015)	100 m	Gr	3	
Vegetation				band ratios, NDVI, imaging spectroscopy bands, aggregated vegetation units, canopy height
Landsat7 scene (USGS EROS, 2013)	30 m		9	
DMC mosaic (DMC, 2015)	22 m		4	
SPOT5 mosaic (Mathys and Kellenberger, 2009)	10 m	Zf	12	
APEX spectrometer mosaics (Schaepman et al., 2015)	2 m	Gr,Be	180	
Share of coniferous trees (FSO, 2000b)	25 m	Zf	1	
Vegetation map (Schmider et al., 1993)	1:5 000	Zf	2	
Species composition data (Brassel and Lischke, 2001)	25 m	Zf	1	
Digital surface model (Swisstopo, 2011)	2 m	Zf	1	
Topography				slope, curvature, northness, TPI, TWI, MRVBF (various radii/resolutions)
Digital elevation model (Swisstopo, 2011)	25 m		62	
Digital terrain model (Swisstopo, 2013b)	2 m		134	

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

with $\hat{f}(\mathbf{x}(\mathbf{s}))^2$ being the prediction of the sqrt-transformed response, $\hat{\sigma}^2$ the estimated residual variance of the fitted model and $\text{Var}[\hat{f}(\mathbf{x}(\mathbf{s}))]$ the variance of $\hat{f}(\mathbf{x}(\mathbf{s}))$ as provided again by the final model. Predictions by group lasso (Sect. 3.1) were backtransformed by $\exp(\cdot)$ or $(\cdot)^2$ because $\text{Var}[\hat{f}(\mathbf{x}(\mathbf{s}))]$ was not known.

For tree-based models (Sect. 3.4 and 3.5) responses were not transformed. Clay and silt were modelled independently, and sand was computed as the remainder to 100 %. Additive log-ratio transformation (ALR) for compositional data (Aitchison, 1986, pp. 113) was tested for geoGAM (Sect. 3.3), but as ALR had no advantage, we preferred to model textural components on their original scale.

To find optimal tuning parameters, we minimized root mean squared error (RMSE, Eq. (3)) in 10-fold cross-validation using the same cross-validation subsets for all methods in Sect. 3.1 to 3.4. For RF (Sect. 3.5) root mean squared error (RMSE) was computed for out-of-bag predictions. All computations were done in R (R Core Team, 2016) using the functions reported below.

3.1 Group lasso

The lasso (least absolute shrinkage and selection operator) is a shrinkage method that likely excludes non-relevant covariates and is therefore an attractive framework for highdimensional covariate selection. Lasso estimates coefficients of a linear model by minimizing a penalized residual sum of squares, with the penalty being equal to the weighted sum of absolute values of the estimated coefficients. By increasing the weight λ of the penalty term a kind of continuous subset selection is performed. Covariates with coefficients shrunken exactly to zero are excluded from the model (Hastie et al., 2009, Sect. 3.4).

We used the grouped lasso which jointly shrinks all coefficients of a factor (R package *grpreg*, Breheny and Huang, 2015). The optimal λ was chosen such that we obtained the least complex model with cross-validation mean squared error (MSE) one standard error (SE) larger than the optimal MSE (Hastie et al., 2009, p. 62).

3.2 Robust external-drift kriging (georob)

As geostatistical method we applied external-drift kriging (EDK) with robustly estimated trend coefficients and exponential variogram parameters (R package *georob*, Papritz, 2016; Nussbaum et al., 2014).

Building a parsimonious trend model from a large number of covariates p was challenging for EDK. We built trend models by concatenating several covariate selection steps. First, we did a pre-selection by finding common covariates in repeated lasso cross-validation runs (32 repetitions, optimal λ from $\text{argmin}_i(\text{MSE}_i) + 1\,\text{SE}$, R package *glmnet*, Friedman et al., 2010).

Then we reduced and expanded this initial covariate set by repeated stepwise covariate selection (models were reduced by *step* function minimizing Bayesian information criterion [BIC], and enlarged by adding covariates with $p \leq 0.05$ in Wald tests). Covariates with inflated coefficients due to multi-collinearity had to be removed manually form the final models (40 % of responses). We used a robustness parameter ψ equal to 1.75. When the robust estimation algorithm did not converge, we first increased ψ and fitted the model non-robustly if this did not help (8 % of responses, see Tables S10 and S11 in Supplement).

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

3.3 Boosted geoadditive model (geoGAM)

Additive models accommodate besides linear effects smooth nonlinear effects of continuous covariates. Spatial auto-correlation can be represented in geoGAM by a smooth function of the spatial coordinates (smooth spatial surface), and nonstationary effects are modelled by interactions between smooth spatial functions and covariates. We based model building for geoGAM on componentwise gradient boosting, a slow stagewise additive model building algorithm. At each stage base procedures are fitted to residuals of the previous model and the best fitting base procedure is retained to update the model by a small step size v. We used non-parametric penalized smoothing splines for continuous covariates and linear base procedures for factors. After boosting further model reduction was achieved by stepwise removal of covariates and aggregation of factor levels. Optimal number of boosting iterations m_{stop} and parameters for further model reduction were found by minimizing cross-validation RMSE. For more details on the model building procedure, see Nussbaum et al. (in rev.) and R package *geoGAM* (Nussbaum, 2017).

Nonstationary effects were added for all continuous covariates, but cross-validation RMSE did not substantially decrease, and we preferred the simpler stationary models throughout. Maximum boosting iterations m_{max} were kept on default 300 iterations (*geoGAM*, Nussbaum, 2017), except if visual inspection of the sequence of cross-validation RMSE values suggested that RMSE had not yet levelled off (20 % of the responses).

3.4 Boosted regression trees (BRT)

Classification and regression trees (CART) are based on recursive binary partitioning of the covariates and can capture complex interaction structures in a dataset. Generally, single trees tend to be noisy (large variance), but to have small bias. Combining trees by ensemble methods aims at reducing their variance. One such approach uses regression trees as base procedures in componentwise gradient boosting (Sect. 3.3).

The optimal number of trees (= number of boosting iterations) n_{trees} and the number of splits per tree i_{d} (representing interaction depth) was found by cross-validation by iterating through a grid of $n_{\text{trees}} = 2, 4, 8, .., 200, 210, .., 800$ and $i_d = 1, 2, .., 12, 14, .., 50$ (R package *gbm*, Southworth, 2015, optimization done using R package *caret*, Kuhn, 2015).

Learning rate was kept similarly small as for geoGAM with $v = 0.1$ (Sect. 3.3, Hastie et al., 2009, Chapt. 10), and minimal number of observations in each end node was set to 5 as in RF (Sect. 3.5).

3.5 Random forest (RF)

RF (Breiman, 2001), another method of balancing instability of CART, averages a committee of fully grown trees. Two mechanisms are used to de-correlate trees and, consequently, reduce the variance of the predictions: 1) bootstrap sampling (bagging) creates a different response vector for each tree, 2) at each node only $m_{\text{try}} < p$ randomly selected covariates are tested as candidates for binary splitting. Predictions are simple means of all fitted n_{tree} trees.

Tuning parameters are the number of trees n_{tree}, the minimal number of observations at terminal nodes n_{min} and the number of tested covariates m_{try} at each split. Tests with five different responses confirmed that tuning n_{tree} and n_{min} did not reduce

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

out-of-bag RMSE substantially (Spiess, 2016). Therefore, we used default values of $n_{\text{tree}} = 500$ and $n_{\text{min}} = 5$ for all RF fits (R package *randomForest*, Liaw and Wiener, 2002). To find optimal m_{try} we minimized out-of-bag RMSE by iterating through $m_{\text{try}} = 1, 2, .., p$.

3.6 Model averaging (MA)

The five methods described above likely represent different aspects of the covariates p and can be seen as different means of reducing the high-dimensional covariate input. Hence, combining predictions of several models to an ensemble possibly improves predictive performance over single methods as large variance of individual models is reduced through averaging (Hastie et al., 2009, Sect. 8.8). We computed weighted sums of the predictions by our five digital soil mapping (DSM) procedures with weights proportional to the inverse cross-validation or out-of-bag RMSE (Tables S8, S10 and S11 in Supplement).

3.7 Legacy soil map

For Greifensee region a legacy soil map 1:5 000 was available, which reported classes of clay and gravel content for top- and subsoil and effective soil depth available to plants (SD, Jäggli et al., 1998). Experienced soil surveyors assigned to each class or combination of classes a typical value of these soil properties (Nussbaum and Papritz, 2017), and we used these as predictions when we computed the statistics for the validation sets (Sect. 3.8).

The map defined topsoil by pedogenetic A horizon without indicating a particular depth. We compared therefore predictions for topsoil to values observed in 0–10 and 10–30 cm depth and predictions for subsoil to observations in 30–50 and 50–100 cm. Inhomogeneous mapping units (complex polygons with multiple soil units assigned) were excluded from the validation. Since all the sites in the validation sets had been used to create the map, statistics rather report "goodness-of-fit" than rigorous validation measures for the legacy map.

3.8 Evaluating predictive performance

The precision of predictions by the six statistical DSM approaches and the legacy soil map was evaluated by comparing predicted $\tilde{Y}(\mathbf{s}_i)$ with observed $Y(\mathbf{s}_i)$ soil properties for all locations $\mathbf{s}_i$ of the validation sets. To rate the methods, we used RMSE and mean squared error skill score (SS_{mse}, Wilks, 2011, p. 359):

$$\text{RMSE} = \left(\frac{1}{n} \sum_{i=1}^{n} \left(Y(\mathbf{s}_i) - \tilde{Y}(\mathbf{s}_i) \right)^2 \right)^{1/2}, \tag{3}$$

$$\text{SS}_{\text{mse}} = 1 - \frac{\sum_{i=1}^{n} \left(Y(\mathbf{s}_i) - \tilde{Y}(\mathbf{s}_i) \right)^2}{\sum_{i=1}^{n} \left(Y(\mathbf{s}_i) - \frac{1}{n} \sum_{i=1}^{n} Y(\mathbf{s}_i) \right)^2}. \tag{4}$$

SS_{mse} has the same interpretation as the R^2 which is occasionally reported, with $\text{SS}_{\text{mse}} = 1$ for perfect predictions (RMSE = 0), $\text{SS}_{\text{mse}} = 0$ if predictions have the same variance as the data of the validation set and $\text{SS}_{\text{mse}} < 0$ for predictions with larger variance. Note, however, that some DSM studies report R^2 values, where R is the Pearson correlation coefficient of $Y(s_i)$ and $\hat{Y}(s_i)$ (e.g. Behrens et al., 2014; Somarathna et al., 2016). Such R^2 values differ, except for linear models fitted by ordinary

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

least squares, from SS_{mse}. Since computation of reported R^2 is sometimes not clear, we call the statistic SS_{mse}, which makes it clear that it is a skill score.

4 Results and Discussion

4.1 Model building

Grouped lasso, robust external-drift kriging (georob) and boosted geoadditive models (geoGAM) successfully selected strongly reduced sets of covariates. On average, lasso models had 21, georob 27 and geoGAM only 12 covariates in the final models. This corresponds to only 3-6 % of all covariates. Boosted regression trees (BRT) performed weak covariate selection. The stagewise forward algorithm selected on average 43 % of all covariates (covariates with importance > 0) for its models. Nonetheless, complexity of BRT models varied quite strongly with 12 % of covariates selected for the smallest and 86 % for the largest model. The number of covariates in final lasso, geoGAM, georob and BRT models was positively correlated over the responses (Pearson correlation between methods 0.43–0.58). RF included all available covariates in its models (all covariates with importance > 0). Having models that depend only on a reduced set of the initial input covariates is desirable, because computing predictions is then less demanding and checking of effect feasibility becomes easier. For three responses we checked therefore, whether covariate importance (Hastie et al., 2009, p. 368) can be used to select covariates for random forest (RF). We either selected $q = 10, 20, .., 50$ most important covariates or selected covariates by stepwise recursive elimination of the least important covariate. For given q, both approaches selected similar sets (correspondence 60–90 %), and root mean squared error (RMSE) computed with independent validation data did not much change by the selection. For example for effective cation exchange capacity (ECEC) 0–20 cm, RMSE increased only by 0.5 $\mathrm{mmol_c\,kg^{-1}}$ for a model with 50 instead of 325 covariates. This increase was clearly within normal fluctuations of RMSE by bagging and random covariate selection (Spiess, 2016). Brungard et al. (2015) even improved predictive precision of RF by recursive covariate elimination.

Optimal values of m_{try} were quite large, hence trees were not strongly de-correlated. Out of 48 models, 32 tuned fits had $m_{\text{try}} > \frac{p}{3}$ which is the software default (Liaw and Wiener, 2002). However, the gain obtained by optimizing m_{try} was generally small. On average RMSE of models fitted with default m_{try} were only 1.015 times larger than RMSE of models with optimized m_{try}. The largest relative benefit of tuning was found for topsoil density of the fine soil fraction (BD) in ZH forest where optimal m_{try} reduced out-of-bag RMSE from 0.052 to 0.049 $\mathrm{Mg\,m^{-3}}$.

In contrast, BRT profited more from tuning its parameters n_{trees} and i_{d}. In particular, optimizing n_{trees} resulted in some reduction of cross-validation RMSE. 69 % of the fits had smaller optimal n_{trees} than the software default (100). Tuning n_{trees} reduced RMSE on average by a factor of 0.941. Optimizing the interaction depth i_{d} (mean optimal value = 10, default = 1), decreased RMSE on average by a factor of 0.982. Tuning n_{trees} and i_{d} had the largest effect for subsoil ECEC and pH in ZH forest where cross-validation RMSE was reduced from 47.3 to 40.9 $\mathrm{mmol_c\,kg^{-1}}$ and 0.91 to 0.75 pH units, respectively.

Residual spatial autocorrelation of georob models was much weaker than autocorrelation of the original responses (Tables S4, S6 and S7 in Supplement). Effective ranges for Greifensee and ZH forest were less than 300 m for most models, and for Berne effective ranges varied between 1–10 km with rather large nugget effects (around 50 % of the total sill). Only 5 of

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

48 final geoGAM models contained a smooth spatial surface, they seemed often too smooth to represent small-scale residual spatial autocorrelation.

Since cross-validation and out-of-bag RMSE did not vary much between the five methods, model averaging (MA) weights did in general not differ much from $1/5$ (interquartile range of weights: 0.18 - 0.21). Only for subsoil soil organic matter (SOM, 50–100 cm) cross-validation RMSE of parametric models were larger compared to BRT and RF and resulted in somewhat larger differences between MA weights. A complete list of model parameters and MA weights is given in Tables S10 and S11 in Supplement.

Summing up, lasso, georob and geoGAM and partly BRT effectively selected relevant covariates from a large number. Reduction of covariates in RF seems – tested on few responses – promising. The benefit of tuning model parameters was sometimes only small, but remains relevant over all responses.

4.2 Evaluation of model performance

4.2.1 General performance

Table 4 reports RMSE and mean squared error skill score (SS_{mse}) of all models for independent validation data, and Fig. 2 summarizes SS_{mse} by method and study region. Overall, the models accounted only for a moderate part of the variance of the validation data (median SS_{mse} of best performing method per response: 0.257). SOM in 10–30 cm soil depth in study region Berne was best predicted with SS_{mse} of 0.677. Soil properties in Greifensee were in general more difficult to predict, yielding for some responses negative SS_{mse} for all methods. For pH in 30–50 cm soil depth, for example, lasso performed "best" with SS_{mse} of -0.089 which is undoubtedly a bad result. In general, topsoil properties were predicted more precisely than subsoil properties (Fig. 3). We are aware of the limitation that we did not validate the methods with data collected by a randomized statistical design (Brus et al., 2011). This is a common drawback if digital soil mapping (DSM) is based on legacy soil data and thus represents a typical situation. Other studies that validated DSM methods with independent data on several soil properties from multiple depths – and likely did not suppress evidence for poor performance – reported similar R^2 values: Negative up to 0.75 (Vaysse and Lagacherie, 2015), 0.1 to 0.48 (Mulder et al., 2016), 0.6 to 0.68 (Kempen et al., 2011), 0.36 to 0.52 (Viscarra Rossel et al., 2015) and 0.26 to 0.55 (Adhikari et al., 2013). Also, these studies found that R^2 values of predictions of topsoil properties were generally larger than R^2 related to subsoils.

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

Table 4. Precision of predictions of soil properties by study region and soil depth computed with independent validation data (RMSE: root mean squared error, SS_{mse} : mean squared error skill score according to Eq. (4), legacy map: legacy soil map 1:5 000, lasso: grouped least absolute shrinkage and selection operator, georob: robust external-drift kriging, geoGAM: boosted geoadditive model, BRT: boosted regression trees, RF: random forest, MA: model averaging, NA: no convergence of georob algorithm).

	depth	legacy map		lasso		georob		geoGAM		BRT		RF		MA	
		RMSE	SS_{mse}	RMSE	SS_{mse}	RMSE	SS_{mse}	RMSE	SS_{mse}	RMSE	SS_{mse}	RMSE	SS_{mse}	RMSE	SS_{mse}
Berne															
clay	0-10			6.698	0.230	5.928	0.396	5.776	0.427	5.897	0.403	6.096	0.365	5.838	0.417
	10-30			7.666	0.162	6.974	0.307	7.450	0.209	6.812	0.339	6.638	0.367	6.717	0.352
	30-50			9.056	0.090	8.743	0.152	9.990	-0.108	8.733	0.153	8.619	0.175	8.729	0.154
	50-100			9.001	-0.002	9.706	-0.165	9.458	-0.106	9.050	-0.013	8.922	0.016	8.871	0.027
silt	0-10			12.954	0.001	12.245	0.107	12.031	0.138	12.19	0.115	11.369	0.230	11.644	0.192
	10-30			11.810	0.115	11.606	0.145	11.391	0.176	11.135	0.213	10.493	0.304	10.778	0.266
	30-50			14.231	0.143	14.151	0.153	14.163	0.151	14.263	0.139	13.809	0.193	13.701	0.206
	50-100			15.604	0.081	15.661	0.074	15.829	0.054	15.108	0.139	15.161	0.136	14.923	0.163
gravel	0-10			2.582	0.129	2.595	0.120	2.567	0.139	2.769	-0.002	2.635	0.113	2.522	0.188
	10-30			3.280	0.200	3.277	0.201	3.281	0.199	3.311	0.185	3.299	0.200	3.143	0.274
	30-50			4.846	0.207	4.767	0.232	4.462	0.328	4.852	0.205	4.843	0.224	4.641	0.287
	50-100			6.146	0.144	6.343	0.088	6.582	0.018	6.367	0.081	6.040	0.173	5.992	0.186
SOM	0-10			4.528	0.634	5.456	0.469	5.137	0.529	5.291	0.501	4.698	0.608	4.742	0.601
	10-30			4.167	0.677	4.981	0.539	4.648	0.599	5.235	0.491	4.910	0.554	4.431	0.636
	30-50			7.817	0.096	7.627	0.139	9.167	-0.243	7.174	0.239	8.379	-0.025	6.562	0.371
	50-100			12.871	-0.015	19.284	-1.279	14.518	-0.296	11.817	0.144	10.629	0.308	9.958	0.392
pH	0-10			0.564	0.549	0.569	0.542	0.547	0.577	0.564	0.549	0.554	0.565	0.536	0.593
	10-30			0.601	0.495	0.591	0.511	0.609	0.482	0.616	0.469	0.601	0.494	0.581	0.527
	30-50			0.715	0.408	0.762	0.327	0.725	0.390	0.722	0.395	0.691	0.447	0.690	0.448
	50-100			0.769	0.425	0.811	0.361	0.791	0.392	0.763	0.434	0.761	0.437	0.728	0.484
SD	–			31.413	0.094	32.61	0.023	33.286	-0.017	31.039	0.115	30.543	0.143	31.014	0.117
Greifensee															
clay	0-10	6.241	0.206	6.208	0.214	6.208	0.214	6.095	0.243	6.296	0.192	6.129	0.234	5.958	0.277
	10-30	6.397	0.293	6.637	0.239	6.662	0.233	6.474	0.276	6.813	0.198	6.575	0.253	6.412	0.289
	30-50	8.478	-0.123	7.651	0.085	7.402	0.144	7.488	0.124	7.286	0.170	7.177	0.195	7.129	0.206
	50-100	8.972	0.037	8.741	0.086	7.944	0.245	9.356	-0.047	8.183	0.199	8.031	0.228	8.048	0.225
silt	0-10			6.624	0.062	6.375	0.131	7.385	-0.167	6.322	0.145	6.007	0.228	6.225	0.171
	10-30			6.676	0.047	6.479	0.102	6.785	0.015	6.360	0.135	6.310	0.148	6.309	0.149
	30-50			7.959	0.021	8.512	-0.120	8.160	-0.030	8.429	-0.099	8.071	-0.007	8.039	0.001
	50-100			9.189	-0.026	10.006	-0.217	9.817	-0.171	9.253	-0.041	9.091	-0.005	9.251	-0.040
gravel	0-10	6.440	-0.128	5.896	0.059	5.549	0.167	5.431	0.202	5.300	0.240	5.326	0.233	5.218	0.263
	10-30	5.831	0.184	6.086	0.116	6.066	0.121	5.335	0.321	5.454	0.290	5.560	0.264	5.438	0.296
	30-50	8.655	0.049	8.778	0.027	8.346	0.120	8.089	0.173	7.991	0.193	7.887	0.214	7.945	0.203
	50-100	9.811	0.314	11.77	0.018	10.821	0.170	10.402	0.233	10.373	0.237	10.696	0.189	10.407	0.232
SOM	0-10			3.504	0.078	3.210	0.226	3.219	0.222	3.244	0.209	3.202	0.230	3.158	0.251
	10-30			3.675	0.028	3.349	0.192	3.455	0.141	3.282	0.224	3.315	0.191	3.258	0.218
	30-50			5.838	-0.072	5.599	0.014	5.900	-0.095	5.352	0.099	5.259	0.130	5.481	0.055
	50-100			7.536	-0.223	NA	NA	11.917	-2.058	6.090	0.201	6.512	0.087	6.620	0.056
pH	0-10			0.714	0.043	0.701	0.077	0.742	-0.035	0.707	0.061	0.700	0.081	0.693	0.097
	10-30			0.700	0.078	0.720	0.024	0.718	0.031	0.708	0.056	0.691	0.102	0.683	0.121
	30-50			0.751	-0.089	0.810	-0.266	0.830	-0.332	0.790	-0.205	0.752	-0.092	0.756	-0.103
	50-100			0.750	-0.085	0.856	-0.412	0.799	-0.228	0.753	-0.092	0.747	-0.075	0.750	-0.083
SD	–	11.076	0.763	19.345	0.278	21.009	0.148	20.928	0.155	19.511	0.265	18.820	0.316	18.858	0.314
Zurich															
ECEC	0–20			75.382	0.356	83.040	0.261	74.900	0.365	73.378	0.423	72.548	0.436	72.294	0.440
	40–60			55.926	0.240	83.238	-0.683	69.113	-0.160	54.681	0.274	51.369	0.359	54.531	0.278
pH	0–20			0.871	0.406	0.913	0.348	0.928	0.325	0.870	0.407	0.856	0.426	0.839	0.448
	40–60			1.122	0.268	1.248	0.093	1.452	-0.227	1.138	0.246	1.093	0.305	1.107	0.287
bd	0–20			0.052	0.203	0.048	0.334	0.055	0.128	0.050	0.271	0.047	0.343	0.046	0.389
	40–60			0.047	0.283	0.061	-0.221	0.051	0.148	0.045	0.336	0.043	0.400	0.044	0.373

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

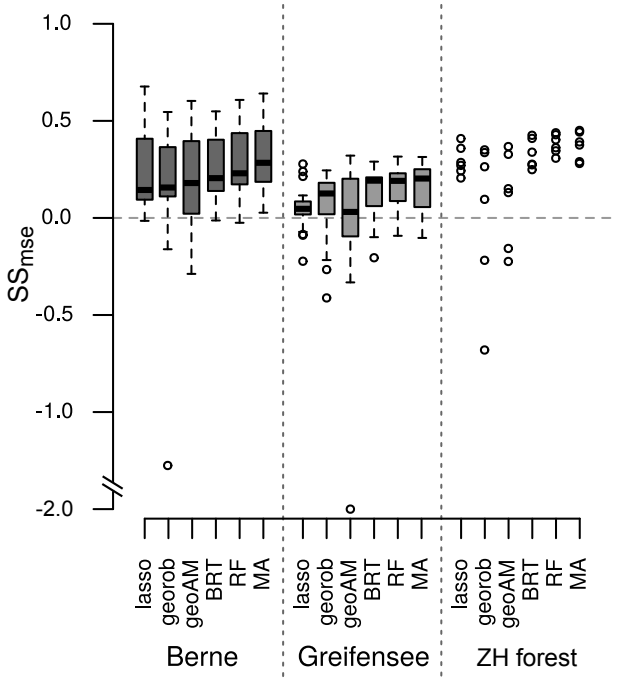

Figure 2. Boxplots of SS_{mse} (for independent validation data) grouped by method and study region. Boxplots summarize SS_{mse} values of n = 21 soil properties for study regions Berne and Greifensee (20 for georob in Greifensee). For ZH forest SS_{mse} are individually shown for $n = 6$ soil properties (lasso: grouped least absolute shrinkage and selection operator, georob: robust external-drift kriging, geoGAM: boosted geoadditive model, BRT: boosted regression trees, RF: random forest, MA: model averaging).

4.2.2 Performance of methods

There was no method that consistently performed best for all soil properties, soil depths and study regions. Each of the tested methods (lasso, georob, geoGAM, BRT, RF) performed best for at least one response, and SS_{mse} varied more strongly between responses than methods. Although no method consistently outperformed the others, Fig. 2 and 3 suggest that the tree-based methods BRT and in particular RF performed on average best. For 28 out of 48 responses, RF hat maximum SS_{mse}, and it never had minimum SS_{mse}. In contrast, georob and geoGAM most often fared worst (for 15 and 14 out of 48 responses, respectively) and were best only for two (georob) and five responses (geoGAM). Lasso ranked between these two methods and BRT.

MA further improved on RF: For 14 out of the 28 responses for which RF was best, MA resulted in even larger SS_{mse} and MA was best for another 9 of the 20 remaining responses. Hence, for 23 out of 48 responses MA had overall largest SS_{mse}.

Apart from overall precision as captured by RMSE and SS_{mse} also bias matters for choosing a DSM method. In general, marginal bias was small (median $bias^2$-to-MSE-ratio < 6 %, Fig. 4, Table S9 in Supplement). Bias contributed more to mean squared error (MSE) when SS_{mse} was small (methods lasso, georob, geoGAM, study region Greifensee), except for the tree-based methods RF and BRT which often had very small $bias^2$-to-MSE-ratios. BRT had slightly lower $bias^2$-to-MSE-ratios

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

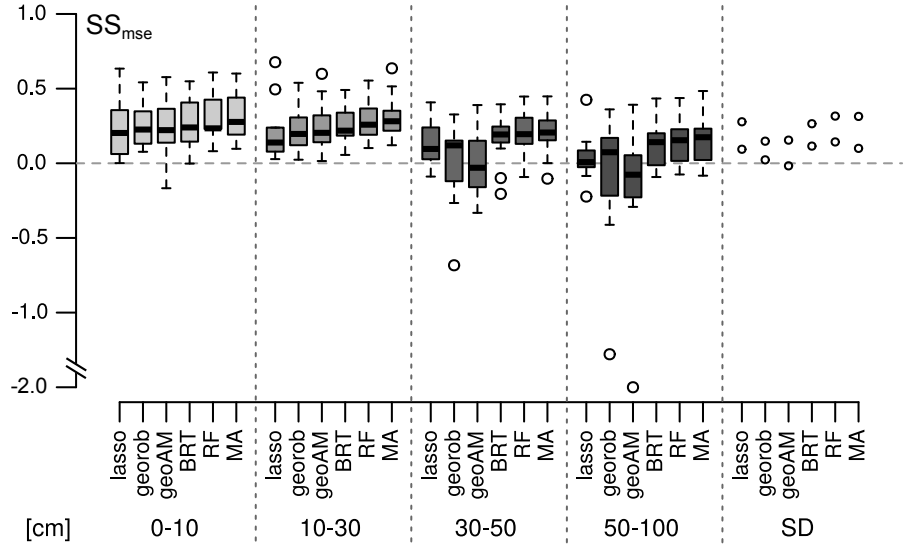

Figure 3. Boxplots of SS_{mse} (for independent validation data) grouped by method and soil depth. Statistics of 0–10 and 0–20 cm soil layers and 20–40 and 30–50 cm were pooled. (SD: effective soil depth available to plants, lasso: grouped least absolute shrinkage and selection operator, georob: robust external-drift kriging, geoGAM: boosted geoadditive model, BRT: boosted regression trees, RF: random forest, MA: model averaging).

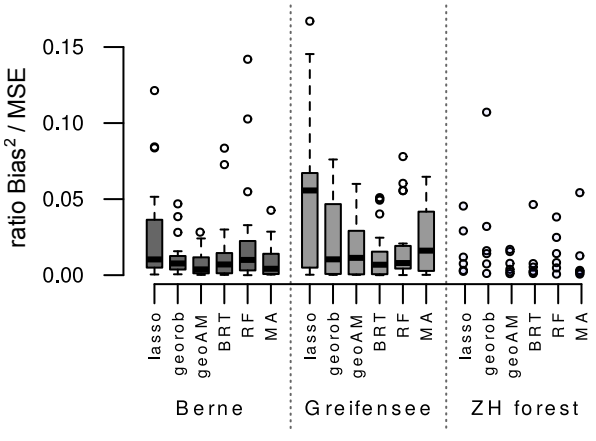

Figure 4. Boxplots of bias^2-to-MSE-ratio (for independent validation data) grouped by method and study region. Boxplots summarize ratios of n = 21 soil properties for study regions Berne and Greifensee (20 for georob in Greifensee). For ZH forest ratios are individually shown for $n = 6$ soil properties (lasso: grouped least absolute shrinkage and selection operator, georob: robust external-drift kriging, geoGAM: boosted geoadditive model, BRT: boosted regression trees, RF: random forest, MA: model averaging).

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

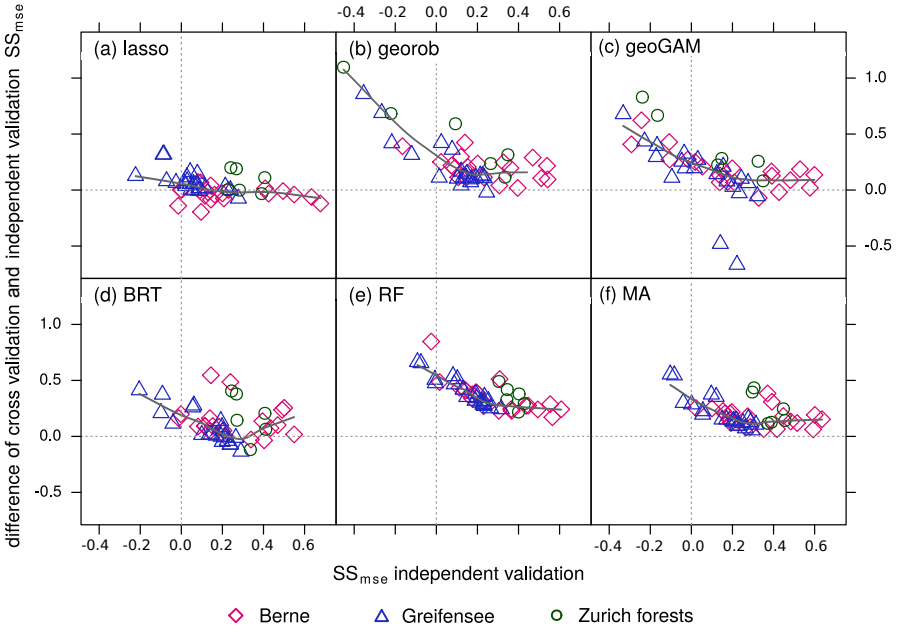

Figure 5. Difference of 10-fold cross-validation and independent validation SS_{mse} plotted against independent validation SS_{mse}, grouped by method. (lasso: grouped least absolute shrinkage and selection operator, georob: robust external-drift kriging, geoGAM: boosted geoadditive model, BRT: boosted regression trees, RF: random forest, MA: model averaging, $SS_{mse} < -1$ were omitted).

compared to RF, confirming that boosting reduces bias in an adaptive way while bagging in RF lowers only variance but not bias (Hastie et al., 2009, p. 588). Largest bias^2-to-MSE-ratios were most often found for lasso, and they were especially large (12 to 17 %, Table S9 in Supplement) for predicting gravel content in Greifensee and SOM in Berne in 50–100 cm depth. Shrinkage methods such as lasso trade reduced variance of predictions for increased bias (Hastie et al., 2009, Chapt. 3). Also RF resulted occasionally in biased predictions, for example for SOM 30–50 cm in Berne. Conditional bias – distortion of predictions conditional on the observed values (Wilks, 2011, p. 304) – did not differ between methods. Predictions were only conditionally biased if overall precision was small.

Lastly, we evaluated whether the various methods tended to over-fit the data by computing differences between cross-validation (CV) or out-of-bag (OOB, RF) SS_{mse} and independent validation SS_{mse} (Fig. 5). We interpret positive (negative) differences in the sequel as indications of over-(under-)fitting, although we cannot exclude that differences between calibration and validation datasets contributed to discrepancies in SS_{mse}. In particular, replicated observations from a given site were not always assigned to the same CV subset, and this possibly contributed to overly optimistic CV results. Except for lasso all methods ([b] to [f] in Fig. 5) often had larger CV or OOB than independent validation SS_{mse}. As also found by Liddicoat et al. (2015) lasso partly under-fitted the data, likely because we penalized the residual sum of squares by the "optimum plus 1 standard error" rule (Sect. 3.1). Why BRT tended partially to under-fit the data remained unclear. Georob and RF tended to over-fit the data most, and geoGAM was intermediate. For all the methods differences in SS_{mse} were largest for poorly

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

performing models (small SS_{mse} in independent validation). For georob this was most pronounced. For ZH forest ECEC (40–60 cm) CV yielded SS_{mse} of 0.70 and independent validation SS_{mse} of -0.683. Hence, repeated covariate selection steps based on BIC and Wald test tend to over-fit the data when responses only weakly depend on covariates.

4.2.3 Factors controlling predictive performance

We explored whether characteristics of the (spatial) empirical distributions of the responses were in some way related to variations of predictive performance observed between responses. We checked whether SS_{mse} and $bias^2$-to-MSE-ratios depended on spatial sampling density, skewness, (robust) coefficient of variation, strength of spatial autocorrelation and tuning parameters of methods (Tables S3 to S7 and S10 to S11 in Supplement), but no clear relationships became evident. Particularly, we could not find any relationships between predictive performance and strength of autocorrelation as measured by spatially structured variance ratios (Vaysse and Lagacherie, 2015, $1 - \text{nugget}/\text{sill}_{\text{total}}$) or spatial ranges of response variograms.

Only for extremely positively skewed responses (SOM below 30 cm in Greifensee) we found that BRT and RF were clearly better than lasso, georob and geoGAM, likely because log-transformation was too weak to fully account for skewness. For skewness < 2 the advantage of tree-based methods disappeared.

4.2.4 Performance legacy soil map

RMSE and SS_{mse} of the legacy soil map (Table 4) were mostly within the range of values observed for DSM methods. Only for subsoil gravel (50–100 cm) and the effective soil depth available to plants (SD) predictions by the legacy soil map were better than DSM predictions. (Note, however, that RMSE and SS_{mse} of the legacy soil map are rather "goodness-of-fit" than rigorous validation measures, because all data of the validation set had been used to create the soil map.) Vaysse and Lagacherie (2015) also found that a legacy soil map predicted only SD more precisely than DSM methods. To create the legacy soil map at a scale of 1:5 000 many auger samples were taken to delineate map units (Jäggli et al., 1998), but this data was not recorded and therefore unavailable for DSM. This might explain why the legacy map modelled SD substantially better (SS_{mse} 0.76) than DSM methods (SS_{mse} 0.15–0.32).

4.3 Evaluation of covariate relevance

4.3.1 Covariate importance

To characterize "predictive skill" of covariates by topic, we computed weighted averages of RF covariate importance (Hastie et al., 2009, p. 368), weighing importance of covariates by validation SS_{mse} (Fig. 6). Overall, terrain attributes were important covariates. For Greifensee they were the main source of information for modelling soil properties. None of the other covariate groups was able to capture much of the variation of soil properties for this study region. Likely, this explains why DSM generally performed poorly here (Sect. 4.2) and indicates that the performance of DSM depends also on regional specific conditions. In the study region Berne climatic covariates were important for chemical but not for physical soil properties, in topsoil more than in subsoil. Additionally, geology, information on soil, sampling period and type of data had moderate

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

importance for this study region, also for physical properties. Similarly, for ZH forest covariate importance differed between chemical (pH, ECEC) and physical properties (BD). Vegetation was very influential for modelling pH and ECEC whereas for BD spatial location, sampling period and type of data were important as well.

Sampling period and type of soil data was important for many responses (Fig. 6). As mentioned also by Mulder et al. (2016) this emphasizes the necessity to compensate temporal changes and differences in analytics when using legacy soil data. Topsoil pH and SOM in Berne – among the responses predicted best in this study – were mostly "explained" by maps of mean monthly and yearly precipitation, a geological overview map, an agricultural suitability map and topographic wetness indices smoothed by different radii (7–60 m). The geological and soil overview maps were also important for modelling SD in Berne. Unlike Greifensee, terrain attributes did not contribute much to modelling SD in Berne. In Greifensee, a map of historic wetlands and distances to water bodies were in addition to terrain attributes (mostly indicating local depressions) important for modelling SD. Predictions of physical and chemical properties in Greifensee relied mainly on vertical and horizontal distance to water bodies, local topographic indices and curvatures (50–90 m radii) as well as the multi-resolution valley bottom flatness (MRVBF). For ZH forest by far the most important covariate was the vegetation map accounting for nearly half of predictive skill in topsoil for pH and ECEC and for one third in subsoil. Terrain attributes were important for ZH forest predictions on both small (variation of slope in 20 m radius) and large scale (topographic indices in radii 50 and 125 km).

Overall, APEX covariates had very small importance (average rank of covariate importance of 168 for RF and 48 for BRT). Differences of reflectance intensities between autumn and spring flights and in between agricultural lands with various crops most likely obscured relations between surface reflectance of vegetation and bare soil. Preprocessing using co-kriging with data from bare soil areas possibly improves predictive capabilities for the present study regions (Lagacherie et al., 2012).

4.3.2 Covariate interpretation

Besides studying covariate importance, we evaluated for all five methods by partial effects or dependence plots what effects single covariates had on selected responses. Figure 7 shows as example how MRVBF (continuous covariate) and the factor for different sampling period and type of soil data (legacy data correction) affected topsoil clay content (0–10 cm, Greifensee). The effect of MRVBF on clay content was similar for all five methods. Large MRVBF values point to accumulation sites in the landscape (Gallant and Dowling, 2003), and such sites often have larger clay contents. BRT and RF partial dependence plots suggests that the relation is nonlinear with a sharp transition at MRVBF equal to 4. Patterns of estimated differences between sampling periods and type of data were similar for the five methods, which further strengthens the evidence that such differences should be compensated when one uses legacy soil data.

4.4 Mapping

In addition to the reported analysis, we visually inspected the soil property maps generated by the six DSM methods. Figure 8 shows for a section of the Greifensee study region DSM maps of topsoil clay content (0–10 cm) along with a map of clay content derived from the legacy soil map. Recall that topsoil clay content could be predicted fairly well for the Greifensee study region (Table 4). All methods, including the soil map predicted soils rather rich in clay with clay content > 20 % for

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

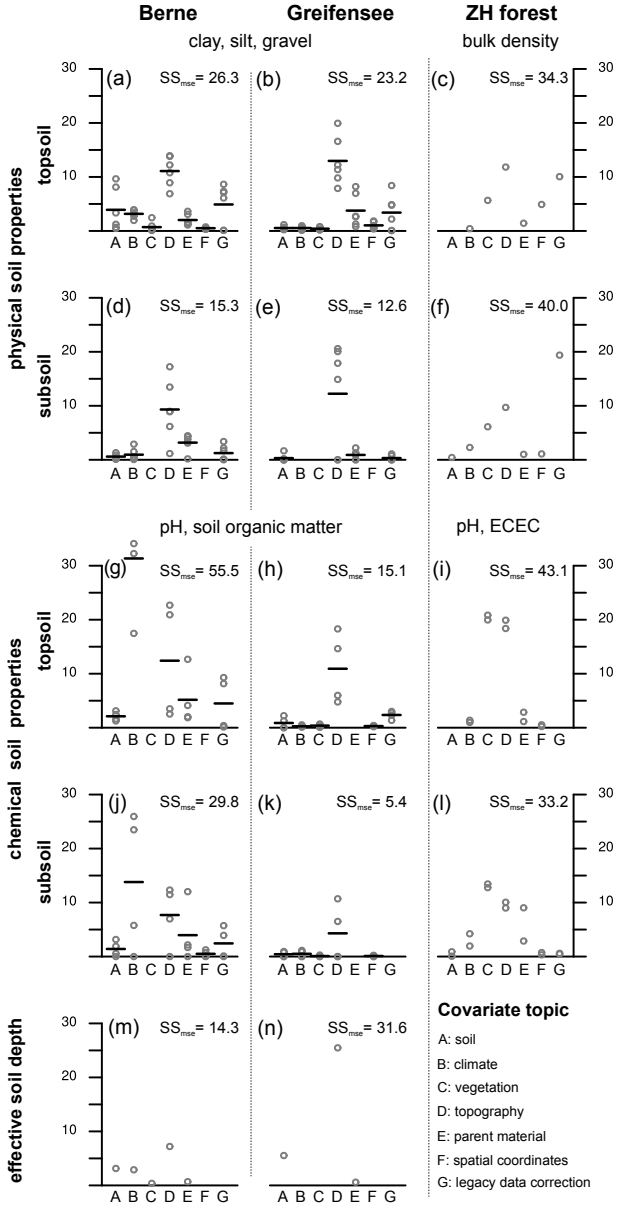

Figure 6. Mean predictive skill [%] of covariates (weighted averages of covariance importance, Hastie et al., 2009, p. 368) grouped by covariate topic (see Table 3, for legacy data correction see Sect. 2.2). Predictive skill is reported separately for study region (Berne, Greifensee and ZH forest), top- (0–30 cm) and subsoils (30–100 cm) and type of response (physical, chemical soil properties, effective soil depth). Mean predictive skill was computed from 30 most important covariates and summed up by topic. The resulting value was weighed by validation SS_{mse} and plotted as grey dots for each response. Mean validation SS_{mse} [%] per covariate topic is given by black horizontal lines (if responses $n > 2$).

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

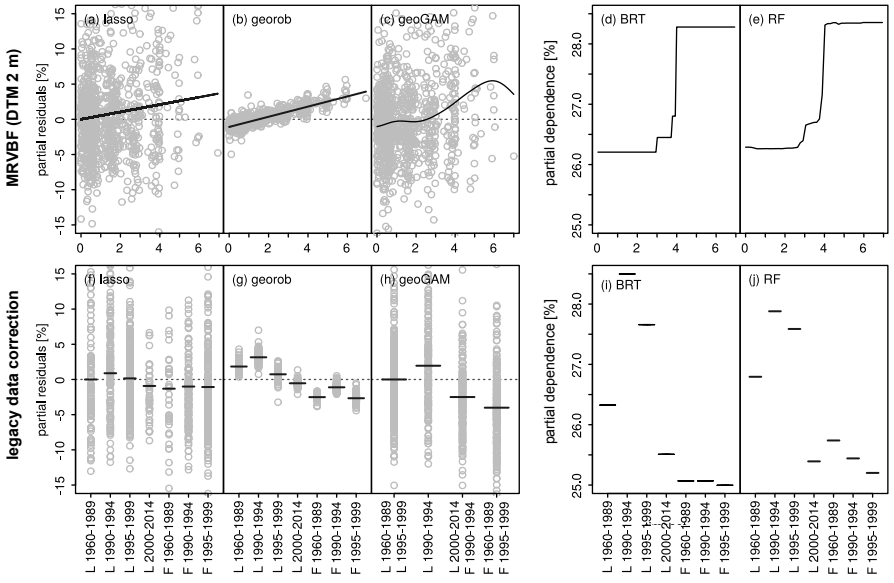

Figure 7. Example partial residual plots (e.g. Faraway, 2005, p. 72) for lasso, georob and geoGAM (panels (a) to (c), (f) to (h)) and partial dependence plots (e.g Hastie et al., 2009, pp. 369) for tree-based methods (panels (d), (e), (i), (j)) for two covariates that were present in lasso, georob and geoGAM and had large importance in BRT and RF for the response clay 0–10 cm in Greifensee (MRVBF 2 m: multi-resolution valley bottom flatness [Gallant and Dowling (2003)], legacy data correction: factor accommodating sampling period and type of soil data, see Sect. 2.2, L: Laboratory measurements, F: field estimates, lasso: grouped least absolute shrinkage and selection operator, georob: robust external-drift kriging, geoGAM: boosted geoadditive model, BRT: boosted regression trees, RF: random forest).

most sites which agrees with available observations (coloured dots, [g] in Fig. 8). Modelled patterns of the maps were similar, but lasso and particularly RF predictions were very smooth. In contrast, predictions by georob, geoGAM and BRT varied more with larger clay content on valley bottoms to the east. MA performed best for this response (SS_{mse} 0.28, Table 4) and, being a weighted average of (a) to (e), showed smoother spatial predictions than georob and geoGAM because RF had highest model averaging weight (0.24, Table S11 in Supplement). The legacy soil map predicted for most polygons the class *sandy loam to loam* with 10–30 % clay to which we assigned a typical clay content of 20 %, which is less than most DSM predictions. As typical for polygon maps small areas with deviating clay content were delineated. The DSM methods were not able to map clay content with similar detail because calibration data was too scarce (Sect. 4.2.4). According to the legacy soil map, there were organic soils in depressions (darkgreen polygons, [g] in Fig. 8), but these had not been sampled.

Maps of georob and BRT predictions showed artefacts (single pixels [georob] or bands [BRT]) with very large predicted values. In the MA map outlying predictions were smoothed out. Outlying georob predictions were caused by the multiflow specific catchment area (2 m resolution), an extremely positively skewed terrain attribute. This covariate was not chosen for the geoGAM model, in lasso its coefficient was strongly shrunken, and BRT and RF do not create extrapolation errors for extreme

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

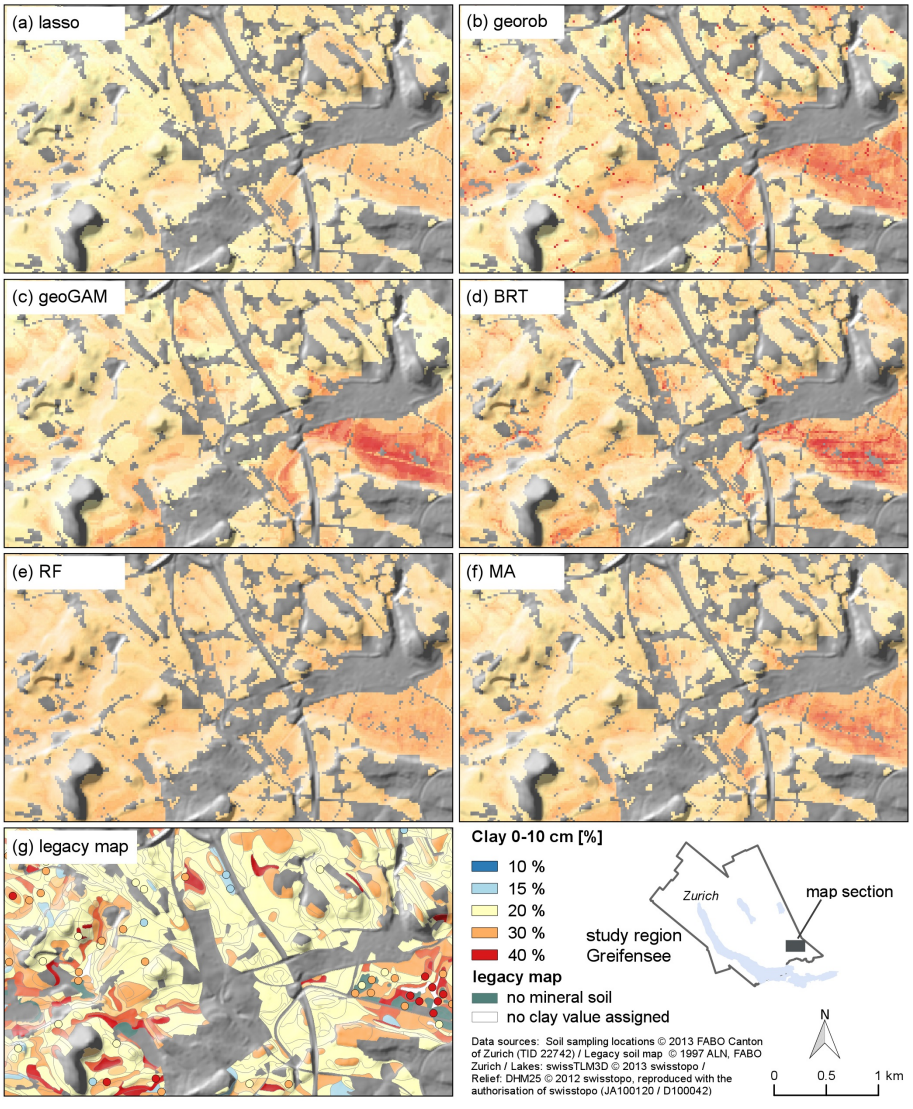

Figure 8. Predictions of clay content [%] in 0–10 cm soil depth computed by six DSM methods (a to f) on a grid of 20 m resolution and by a legacy soil map 1:5 000 (g) for a section of Greifensee study region north of village of Bubikon. The legacy soil map (g) predicted texture classes to each of which we assigned a typical clay content displayed here. For complex polygons the texture class of the main unit is shown. Dots in (g) depict observations of clay content used for calibrating (a) to (f) and for creating the soil map (g).

values of covariates. The cause of the artefact in BRT was impossible to spot because the BRT model contained 148 covariates (Table S11 in Supplement).

Besides creating extrapolation errors, parametric methods (lasso, georob, geoGAM) predicted physically impossible values (e.g. clay content < 0 % or > 100 %) that we had to eliminate. In contrast, trees do not extrapolate beyond the range of observed values of the response when computing predictions.

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

4.5 Practical use of statistical methods

All tested DSM methods were able to process large sets of factors and continuous covariates. Although RF more often performed best and MA even improved on that, the advantage measured in validation SS_{mse} was small (Sect. 4.2). Hence other reasons than precision might become more decisive for choosing a particular approach.

In our study residual spatial autocorrelation was weak or short-ranged. For a response with strong residual autocorrelation a geostatistical approach might still offer an advantage. The smooth spatial surface of geoGAM is possibly too coarse to capture short-ranged autocorrelation. BRT and RF include spatial coordinates as covariates, but if the response depends only weakly on other covariates, spatial coordinates become overly important. Repeated recursive splitting on coordinates likely leads to "chessboard type" artefacts.

All methods allowed interpretation of modelled relationships (Fig. 7), but a large number of remaining covariates in a model hinders interpretation of partial effects or dependencies. The most parsimonious models were chosen by geoGAM with only 12 remaining covariates remaining on average (lasso: 21, georob: 27). For BRT and RF a covariate selection scheme would still need to be implemented and tuned. Preliminary results suggest that this might be well worth the effort (Sect. 4.1). But even without covariate selection, BRT and RF allowed to analyse the importance of the covariates (Fig. 6).

R packages are readily available for all methods used in this study. Lasso and geoGAM optimize their tuning parameters directly without any further input to the software while RF and BRT require specification of parameter ranges to be tested. The number of parameters to tune influences computing times considerably. Using default m_{try} for RF (Sect. 4.1) and coarse grids for finding optimal BRT parameters (Sect. 3.4) might be a good compromise to balance computing efforts with good predictive performance. Computational effort was especially large for georob, where there is no established efficient procedure for building models from large sets of covariates. Lasso, based on a coordinate descent algorithm, built models most quickly (see also Fitzpatrick et al., 2016) while computational effort for geoGAM model building was quite variable depending mainly on number of observations and number of covariates selected by boosting step (Nussbaum et al., in rev., Sect. 2.2).

Moreover, ease of modelling predictive uncertainty is another factor relevant for choice of a DSM method. In georob uncertainties can be directly derived from the kriging variances. For RF, conditional quantiles of predictive distributions can be estimated directly at the cost of a larger memory requirement (R package *quantregForest*, Meinshausen, 2015). For lasso, geoGAM and BRT model-based bootstrapping can be used to simulate predictive distributions (Nussbaum et al., in rev.), but bootstrapping involves quite some computational effort.

Responses for DSM are not always continuous soil properties. Binary, multinomial (e.g. soil types) or ordinal (e.g. drainage classes) responses are sometimes relevant. Grouped lasso is availble for binary (R package *grpreg*, Breheny and Huang, 2015) and nominal responses (R package *glmnet*, Friedman et al., 2010). Logistic geostatistical models could be fitted in the generalized linear mixed model framework (R package *geoRGLM*, Christensen and Ribeiro Jr, 2002; Diggle and Ribeiro Jr, 2002; Pringle et al., 2014), but this is practical only for small datasets. INLA (Integrated Nested Laplace Approximation, Rue et al., 2009; Lindgren et al., 2011) could be viable alternative. geoGAM accommodate binary and ordinal responses (Nussbaum, 2017), but extension to nominal responses would be straightforward. Classification for binary and nominal responses are easily

SOIL Discuss., doi:10.5194/soil-2017-14, 2017
Manuscript under review for journal SOIL
Discussion started: 9 May 2017

fitted by RF (R package *randomForest*, Liaw and Wiener, 2002) and BRT (R package *gbm*, Southworth, 2015) while ordinal response BRT could be implemented by R package *mboost* (Hothorn et al., 2015) with slightly larger effort on model specification.

5 Conclusions

We applied – to a total of 48 soil responses observed in three study regions in Switzerland – six statistical digital soil mapping (DSM) methods: grouped lasso (least absolute shrinkage and selection operator), robust external-drift kriging (georob), boosted geoadditive models (geoGAM), boosted regression trees (BRT) and random forest (RF). Models were build from 300–500 environmental covariates, and performance was assessed by comparing model predictions with independent validation data.

From this study we conclude:

- All methods were successfully building models automatically from large sets of covariates. The applied ad hoc procedure to find a parsimonious trend model for georob was however very inefficient.
- Except for lasso, cross-validation and out-of-bag precision measures were sometimes better than actually observed for the validation data. This suggests that the methods partly tended to over-fit the data and underpins the necessity of model evaluation with independent data.
- Admittedly, the best performing method frequently had not much larger mean squared error skill score (SS_{mse}) than its closest competitors, and the empirical distributions of SS_{mse} did not differ much for BRT, RF and MA (Figures 2 and 3). Nevertheless, the frequencies of best and worst performance clearly favoured RF if only one method is used. Applying model averaging (MA) of several approaches likely even improves on RF.
- Correcting for sampling period and soil data type by adding a factor to the models turned out to be important. Legacy soil data is inherently heterogeneous for various reasons, but one can (and should) compensate this variation by careful statistical modelling.

Code availability. The geoGAM model building procedure was published as R package `geoGAM` (Nussbaum, 2017).

Data availability. The soil data of the Canton of Zurich were used under a non-public data licence (Canton of Zurich, contract number TID 22742; WSL) and could not be published. Data from Berne study region was partly published as test data *berne* and *berne.grid* in R package `geoGAM` (Nussbaum, 2017).

Author contributions. A. Papritz proposed the comparison of the selected DSM approaches and defined the model selection procedure for georob. M. Nussbaum implemented the DSM approaches for the three study regions and evaluated the results. K. Spiess explored influence of tuning parameters and covariate selection on RF for selected responses. A. Baltensweiler heavily contributed to the computation of the multi-scale terrain attributes and U. Grob and A. Keller harmonized the soil data with collaborators. L. Greiner defined the demand for soil properties to be mapped and proposed the derivation of SD from horizon qualifiers of Swiss soil classification data. M. Nussbaum prepared the manuscript with considerable input from A. Papritz and further contributions from all co-authors.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We thank the Swiss National Science Foundation SNSF for funding this work in the frame of the National Research Program "Sustainable use of Soil as a Resource" (NRP 68)" and the "Swiss Earth Observatory Network" (SEON) for funding aerial surveys using APEX and Sanne Diek for preprocessing the imagery. The contribution of Michael E. Schaepman was supported by the University of Zurich Research Priority Program on Global Change and Biodiversity (URPP GCB). Special thanks go to WSL and the cantonal agencies for soil protection of Zurich and Berne sharing their soil data to make this study possible. Moreover, we are grateful to soil surveyors Peter Schwab, Martin Zürrer and Alexander Lehmann for their support to compare the legacy soil map with DSM results.

References

Adhikari, K., Kheir, R., Greve, M., Bøcher, P., Malone, B., Minasny, B., McBratney, A., and Greve, M.: High-resolution 3-D mapping of soil texture in Denmark, Soil Sci Soc Am J, 77, 860–876, doi:10.2136/sssaj2012.0275, 2013.

AGR: Geoprodukt Geologische Rohstoffkarte ADT, Metadaten komplett. Amt für Gemeinden und Raumordnung des Kantons Bern, www.be.ch/geoportal, last access: 04.04.2017, 2015.

Aitchison, J.: The statistical analysis of compositional data, Chapman & Hall, doi:10.1007/978-94-009-4109-0, 1986.

ALN: Historische Feuchtgebiete der Wildkarte 1850. Amt für Landschaft und Natur des Kantons Zürich, http://www.aln.zh.ch/internet/baudirektion/aln/de/naturschutz/naturschutzdaten/geodaten.html, last access 29.03.2017, 2002.

ALN: Geologische Karte des Kantons Zürich nach Hantke et. al 1967, GIS-ZH Nr. 41. Amt für Landschaft und Natur des Kantons Zürich, http://www.gis.zh.ch/Dokus/Geolion/gds_41.pdf, last access: 15.02.2015, 2014a.

ALN: Meliorationskataster des Kantons Zürich, GIS-ZH Nr. 148. Amt für Landschaft und Natur des Kantons Zürich., http://www.geolion.zh.ch/geodatensatz/show?nbid=387, last access 29.03.2017, 2014b.

AWA: Geoprodukt Versickerungszonen VSZ, Metadaten komplett. Amt für Wasser und Abfall des Kantons Bern, www.be.ch/geoportal, last access: 04.04.2017, 2014a.

AWA: Geoprodukt Grundwasserkarte GW25, Metadaten komplett.Amt für Wasser und Abfall des Kantons Bern, www.be.ch/geoportal, last access: 04.04.2017, 2014b.

AWEL: Hinweisflächen für anthropogene Böden, GIS-ZH Nr. 260. Amt für Abfall, Wasser, Energie und Luft des Kanton Zürich, http://www.geolion.zh.ch/geodatensatz/show?nbid=985, last access 29.03.2017, 2012.

AWEL: Grundwasservorkommen, GIS-ZH Nr. 327. Amt für Abfall, Wasser, Energie und Luft des Kanton Zürich, http://www.geolion.zh.ch/geodatensatz/show?nbid=723, last access 29.03.2017, 2014.

AWEL: NO_2-Immissionen, GIS-ZH Nr. 82, Amt für Abfall, Wasser, Energie und Luft des Kanton Zürich, http://geolion.zh.ch/geodatensatz/show?nbid=783, last access 29.03.2017, 2015.

BAFU: Luftbelastung: Karten Jahreswerte, Ammoniak und Stickstoffdeposition, Jahresmittel 2007 (modelliert durch METEOTEST), http://www.bafu.admin.ch/luft/luftbelas-tung/schadstoffkarten, last access 15.02.2015, 2011.

BAFU and GRID-Europe: Swiss Environmental Domains. A new spatial framework for reporting on the environment, Environmental studies 1024, Federal Office for the Environment FOEN, Berne, http://www.bafu.admin.ch/publikationen/publikation/01564/index.html?lang=en, 2010.

Bechler, K. and Toth, O.: Bewertung von Böden nach ihrer Leistungsfähigkeit, Leitfaden für Planungen und Gestattungsverfahren, LUBW Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg, http://www.fachdokumente.lubw.baden-wuerttemberg.de/servlet/is/99474/Bodenschutz_23_Lesefassung_aktuell.pdf?command=downloadContent&filename=Bodenschutz_23_Lesefassung_aktuell.pdf&FIS=199, 2. Auflage, last access: 04.04.2017, 2010.

Behrens, T., Schmidt, K., Zhu, A. X., and Scholten, T.: The ConMap approach for terrain-based digital soil mapping, Eur J Soil Sci, 61, 133–143, doi:10.1111/j.1365-2389.2009.01205.x, 2010a.

Behrens, T., Zhu, A., Schmidt, K., and Scholten, T.: Multi-scale digital terrain analysis and feature selection for digital soil mapping, Geoderma, 155, 175–185, doi:10.1016/j.geoderma.2009.07.010, 2010b.

Behrens, T., Schmidt, K., Ramirez-Lopez, L., Gallant, J., Zhu, A.-X., and Scholten, T.: Hyper-scale digital soil mapping and soil formation analysis, Geoderma, 213, 578–588, doi:10.1016/j.geoderma.2013.07.031, 2014.

BFS: GEOSTAT Benützerhandbuch, Bundesamt für Statistik, Bern, 2001.

Brassel, P. and Lischke, H., eds.: Swiss National Forest Inventory: Methods and models of the second assessment, Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, 2001.

Breheny, P. and Huang, J.: Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors, Stat Comput, 25, 173–187, doi:10.1007/s11222-013-9424-2, 2015.

Breiman, L.: Random Forests, Machine Learning, 45, 5–32, 2001.

Brungard, C. W., Boettinger, J. L., Duniway, M. C., Wills, S. A., and Jr., T. C. E.: Machine learning for predicting soil classes in three semi-arid landscapes, Geoderma, 239–240, 68–83, doi:10.1016/j.geoderma.2014.09.019, 2015.

Brus, D. J., Kempen, B., and Heuvelink, G. B. M.: Sampling for validation of digital soil maps, Eur J Soil Sci, 62, 394–407, doi:10.1111/j.1365-2389.2011.01364.x, 2011.

Calzolari, C., Ungaro, F., Filippi, N., Guermandi, M., Malucelli, F., Marchi, N., Staffilani, F., and Tarocco, P.: A methodological framework to assess the multiple contributions of soils to ecosystem services delivery at regional scale, Geoderma, 261, 190–203, doi:10.1016/j.geoderma.2015.07.013, 2016.

Christensen, O. and Ribeiro Jr, P.: geoRglm – A package for generalised linear spatial models, R-NEWS, 2, 26–28, http://cran.R-project.org/doc/Rnews, iSSN 1609-3631, last access 04.04.2017, 2002.

Cressie, N.: Block Kriging for Lognormal Spatial Processes, Math Geol, 38, 413–443, doi:10.1007/s11004-005-9022-8, 2006.

Danner, C., Hensold, C., Blum, P., Weidenhammer, S., Aussendorf, M., Kraft, M., Weidenbacher, A., Holleis, P., and Kölling, C.: Das Schutzgut Boden in der Planung, Bewertung natürlicher Bodenfunktionen und Umsetzung in Planungs- und Genehmigungsverfahren, Bayerisches Landesamt für Umweltschutz, Bayerisches Geologisches Landesamt, http://www.lfu.bayern.de/boden/bodenfunktionen/ertragsfaehigkeit/doc/arbeitshilfe_boden.pdf, last access 29.03.2017, 2003.

Diggle, P. and Ribeiro Jr, P.: Bayesian inference in gaussian model-based geostatistics, Geographical and Environmental Modelling, 6, 129–146, doi:10.1080/1361593022000029467, 2002.

Dirichlet, G. L.: Über die Reduction der positiven quadratischen Formen mit drei unbestimmten ganzen Zahlen, Journal für die reine und angewandte Mathematik, 40, 209–227, doi:10.1017/cbo9781139237345.005, 1850.

DMC: Disaster Monitoring Constellation International Imaging, http://www.dmcii.com, last access: 03.02.2015, 2015.

DVWK: Filtereigenschaften des Bodens gegenüber Schadstoffen. Teil I: Beurteilung der Fähigkeit von Böden, zugeführte Schwermetalle zu immobilisieren. DVWK-Merkblätter zur Wasserwirtschaft, Bericht, Deutscher Verband für Wasserwirtschaft und Kulturbau (DVWK), 1988.

Faraway, J. J.: Linear Models with R, vol. 63 of *Texts in Statistical Science*, Chapman & Hall/CRC, Boca Raton, 2005.

Fitzpatrick, B. R., Lamb, D. W., and Mengersen, K.: Ultrahigh Dimensional Variable Selection for Interpolation of Point Referenced Spatial Data: A Digital Soil Mapping Case Study, PLoS One, 11, 1–19, doi:10.1371/journal.pone.0162489, 2016.

Friedman, J., Hastie, T., and Tibshirani, R.: Regularization paths for generalized linear models via coordinate descent, J Stat Softw, 33, 1–22, doi:10.18637/jss.v033.i01, 2010.

FSO: Swiss soil suitability map. BFS GEOSTAT. Swiss Federal Statistical Office, http://www.bfs.admin.ch/bfs/portal/de/index/dienstleistungen/geostat/datenbeschreibung/digitale_bodeneignungskarte.html, last access 15.02.2015, 2000a.

FSO: Tree composition of Swiss forests. BFS GEOSTAT. Swiss Federal Statistical Office, http://www.bfs.admin.ch/bfs/portal/de/index/dienstleistungen/geostat/datenbeschreibung/waldmischungsgrad.html, last access 15.02.2015, 2000b.

Gallant, J. C. and Dowling, T. I.: A multiresolution index of valley bottom flatness for mapping depositional areas, Water Resour Res, 39, doi:10.1029/2002WR001426, 2003.

Greiner, L., Keller, A., Grêt-Regamey, A., and Papritz, A.: Soil function assessment methods for quantifying the contributions of soils to ecosystems services, Submitted to Ecosyst Serv, in rev.

Hantke, R. u.: Geologische Karte des Kantons Zürich und seiner Nachbargebiete, Kommissionsverlag Leemann, Zürich, Sonderdruck aus *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich*, 112(2): 91–122, 1967.

Hartemink, A. E., Krasilnikov, P., and Bockheim, J.: Soil maps of the world, Geoderma, 207–208, 256–267, doi:10.1016/j.geoderma.2013.05.003, 2013.

Hastie, T., Tibshirani, R., and Friedman, J.: The Elements of Statistical Learning; Data Mining, Inference and Prediction, Springer, New York, 2 edn., 2009.

Haygarth, P. M. and Ritz, K.: The future of soils and land use in the UK: Soil systems for the provision of land-based ecosystem services, Land Use Policy, 26, Supplement 1, S187–S197, doi:10.1016/j.landusepol.2009.09.016, land Use Futures, 2009.

Hothorn, T., Buehlmann, P., Kneib, T., Schmid, M., and Hofner, B.: mboost: Model-Based Boosting, http://CRAN.R-project.org/package=mboost, R package version R package version 2.4-2, last access 29.03.2017, 2015.

Hotz, M.-C., Weibel, F., Ringgenberg, B., Beyeler, A., Finger, A., Humbel, R., and Sager, J.: Arealstatistik Schweiz Zahlen – Fakten – Analysen, Bericht, Bundesamt für Statistik (BFS), Neuchâtel, 2005.

Jäggli, F., Peyer, K., Pazeller, A., and Schwab, P.: Grundlagenbericht zur Bodenkartierung des Kantons Zürich, Tech. rep., Volkswirtschaftsdirektion des Kantons Zürich und Eidg. Forschungsanstalt für Agrarökologie und Landbau Zürich Reckenholz FAL, 1998.

Kempen, B., Brus, D., and Stoorvogel, J.: Three-dimensional mapping of soil organic matter content using soil type-specific depth functions, Geoderma, 162, 107–123, doi:10.1016/j.geoderma.2011.01.010, 2011.

Kuhn, M.: caret: Classification and Regression Training, https://CRAN.R-project.org/package=caret,https://github.com/topepo/caret, R package version 6.0-71, last access: 04.04.2017, 2015.

Lacoste, M., Mulder, V., de Forges, A. R., Martin, M., and Arrouays, D.: Evaluating large-extent spatial modeling approaches: A case study for soil depth for France, Geoderma Regional, 7, 137–152, doi:10.1016/j.geodrs.2016.02.006, 2016.

Lagacherie, P., Bailly, J. S., Monestiez, P., and Gomez, C.: Using scattered hyperspectral imagery data to map the soil properties of a region, Eur J Soil Sci, 63, 110–119, doi:10.1111/j.1365-2389.2011.01409.x, 2012.

LANAT: Geoprodukt Landwirtschaftliche Eignungskarte LWEK74, Metadaten komplett. Amt für Landwirtschaft und Natur, Kanton Bern, http://files.be.ch/bve/agi/geoportal/geo/lpi/LWEK74_1974_01_LANG_DE.PDF, last access: 04.04.2017, 2015.

Lehmann, A., David, S., and Stahr, K.: TUSEC — Bilingual-Edition: Eine Methode zur Bewertung natürlicher und anthropogener Böden (Deutsche Fassung), Hohenheimer Bodenkundliche Hefte 86, Institut für Bodenkunde und Standortslehre, Universität Hohenheim, Stuttgart, https://soil.uni-hohenheim.de/uploads/media/TUSEC_2.Aufl_03.pdf, 2. Auflage, last access: 07.06.2016, 2013.

Li, J., Heap, A. D., Potter, A., and Daniell, J. J.: Application of machine learning methods to spatial interpolation of environmental variables, Environ Modell Software, 26, 1647–1659, doi:10.1016/j.envsoft.2011.07.004, 2011.

Liaw, A. and Wiener, M.: Classification and Regression by randomForest, R News, 2, 18–22, http://CRAN.R-project.org/doc/Rnews/, last acces: 04.04.2017, 2002.

Liddicoat, C., Maschmedt, D., Clifford, D., Searle, R., Herrmann, T., Macdonald, L., and Baldock, J.: Predictive mapping of soil organic carbon stocks in South Australia's agricultural zone, Soil Res, 53, 956–973, doi:10.1071/SR15100, 2015.

Lindgren, F., Rue, H., and Lindström, J.: An explicit link between gaussian fields and gaussian markov random fields: The stochastic partial differential equation approach, J Roy Stat Soc B, 73, 423–498, doi:10.1111/j.1467-9868.2011.00777.x, 2011.

Litz, N.: Schutz vor Organika, in: Handbuch der Bodenkunde, edited by Blume, H.-P., vol. 5, chap. 7.6.6, p. 28, Wiley-VCH, Landsberg, 1998.

Malone, B. P., Minasny, B., Odgers, N. P., and McBratney, A. B.: Using model averaging to combine soil property rasters from legacy soil maps and from point data, Geoderma, 232–234, 34–44, doi:10.1016/j.geoderma.2014.04.033, 2014.

Mathys, L. and Kellenberger, T.: Spot5 RadcorMosaic of Switzerland, Tech. rep., National Point of Contact for Satellite Images NPOC: Swisstopo; Remote Sensing Laboratories, University of Zurich, Zurich, 2009.

McBratney, A. B., Mendonça Santos, M. L., and Minasny, B.: On Digital Soil Mapping, Geoderma, 117, 3–52, doi:10.1016/S0016-7061(03)00223-4, 2003.

Meersmans, J., De Ridder, F., Canters, F., De Baets, S., and Van Molle, M.: A multiple regression approach to assess the spatial distribution of Soil Organic Carbon (SOC) at the regional scale (Flanders, Belgium), Geoderma, 143, 1–13, doi:10.1016/j.geoderma.2007.08.025, 2008.

Meinshausen, N.: quantregForest: Quantile Regression Forests, https://CRAN.R-project.org/package=quantregForest, R package version 1.3-5, last access 29.03.2017, 2015.

Miller, B. A., Koszinski, S., Wehrhan, M., and Sommer, M.: Impact of multi-scale predictor selection for modeling soil properties, Geoderma, 239–240, 97–106, doi:10.1016/j.geoderma.2014.09.018, 2015.

Mulder, V., Lacoste, M., de Forges, A. R., and Arrouays, D.: GlobalSoilMap France: High-resolution spatial modelling the soils of France up to two meter depth, Sci Total Environ, 573, 1352–1369, doi:10.1016/j.scitotenv.2016.07.066, 2016.

Mulder, V. L., de Bruin, S., Schaepman, M. E., and Mayr, T. R.: The use of remote sensing in soil and terrain mapping – A review, Geoderma, 162, 1–19, doi:10.1016/j.geoderma.2010.12.018, 2011.

Nussbaum, M.: geoGAM: Select Sparse Geoadditive Models for Spatial Prediction, https://CRAN.R-project.org/package=geoGAM, R package version 0.1-2, last access 29.03.2017, 2017.

Nussbaum, M. and Papritz, A.: Validierung von konventionellen Bodenkarten mit unabhängigen Bodendaten – Methodik mit Fallstudie, unpublished, 2017.

Nussbaum, M., Papritz, A., Baltensweiler, A., and Walthert, L.: Estimating soil organic carbon stocks of Swiss forest soils by robust external-drift kriging, Geosci Model Dev, 7, 1197–1210, doi:10.5194/gmd-7-1197-2014, 2014.

Nussbaum, M., Walthert, L., Fraefel, M., Greiner, L., and Papritz, A.: Mapping of soil properties at high resolution in Switzerland using boosted geoadditive models, submitted to SOIL, in rev.

Papritz, A.: georob: Robust Geostatistical Analysis of Spatial Data, https://cran.r-project.org/web/packages/georob/index.html, R package version 0.3-1, last access: 04.04.2017, 2016.

Poggio, L., Gimona, A., and Brewer, M.: Regional scale mapping of soil properties and their uncertainty with a large number of satellite-derived covariates, Geoderma, 209–210, 1–14, doi:10.1016/j.geoderma.2013.05.029, 2013.

Pringle, M., Zund, P., Payne, J., and Orton, T.: Mapping depth-to-rock from legacy data, using a generalized linear mixed model, in: GlobalSoilMap: Basis of the global spatial soil information system, edited by Arrouays, D., McKenzie, N., Hempel, J., Richer de Forges, A., and McBratney, A., pp. 295–299, CRC Press, doi:10.1201/b16500-55, 2014.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/, last access: 29.03.2017, 2016.

Remund, J., Frehner, M., Walthert, L., Kägi, M., and Rihm, B.: Schätzung standortspezifischer Trockenstressrisiken in Schweizer Wäldern, 2011.

Robinson, D., Hockley, N., Cooper, D., Emmett, B., Keith, A., Lebron, I., Reynolds, B., Tipping, E., Tye, A., Watts, C., Whalley, W., Black, H., Warren, G., and Robinson, J.: Natural capital and ecosystem services, developing an appropriate soils framework as a basis for valuation, Soil Biol Biochem, 57, 1023–1033, doi:10.1016/j.soilbio.2012.09.008, 2013.

Rossiter, D.: Digital Soil Mapping Across Paradigms, Scales and Baunaries, chap. Digital Soil Resource Inventories: Status and Prospects in 2015, pp. 275–286, Springer Environmental Science and Engineering, 2016.

Rue, H., Martino, S., and Chopin, N.: Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations, J Roy Stat Soc B, 71, 319–392, doi:10.1111/j.1467-9868.2008.00700.x, 2009.

Schaepman, M., Jehle, M., Hueni, A., D'Odorico, P., Damm, A., Weyermann, J., Schneider, F., Laurent, V., Popp, C., Seidel, F., Lenhard, K., Gege, P., Küchler, C., Brazile, J., Kohler, P., Vos, L., Meuleman, K., Meynart, R., Schläpfer, D., and Itten, K.: Advanced radiometry measurements and Earth science applications with the Airborne Prism Experiment (APEX), Remote Sens Environ, 158, 207–219, doi:10.1016/j.rse.2014.11.014, 2015.

Schmider, P., Küper, M., Tschander, B., and Käser, B.: Die Waldstandorte im Kanton Zürich Waldgesellschaften, Waldbau Naturkunde, vdf Verlag der Fachvereine an den schweizerischen Hochschulen und Techniken, Zürich, 1993.

Scull, P., Franklin, J., Chadwick, O. A., and McArthur, D.: Predictive Soil Mapping: A review, Prog Phys Geog, 27, 171–197, doi:10.1191/0309133303pp366ra, 2003.

Siemer, B., Obmann, L., Hinrichs, U., Penndorf, O., Pohl, M., Schürer, S., Schulze, P., and Seiffert, S.: Bodenbewertungsinstrument Sachsen, Tech. rep., Sächsisches Landesamt für Umwelt, Landwirtschaft und Geologie, Dresden, 2014.

Somarathna, P., Malone, B., and Minasny, B.: Mapping soil organic carbon content over New South Wales, Australia using local regression kriging, Geoderma Regional, 7, 38–48, doi:10.1016/j.geodrs.2015.12.002, 2016.

Southworth, H.: gbm: Generalized Boosted Regression Models, https://CRAN.R-project.org/package=gbm, R package version 2.1.1, last access: 04.04.2017, 2015.

Spiess, K.: Vorhersage von Bodeneigenschaften mit Quantile Regression Forest, Validierung und Vergleich mit den Vorhersagen aus geoadditiven Modellen, BSc Thesis, Departement für Umweltsystemwissenschaften der ETH Zürich, Zürich, 2016.

Swisstopo: Geologische Karte der Schweiz 1:500000, http://www.swisstopo.admin.ch/internet/swisstopo/de/home/products/maps/geology/geomaps/gm500.html, last access: 07.06.2016, 2005.

Swisstopo: Switzerland during the Last Glacial Maximum 1:500 000, http://www.swisstopo.admin.ch/internet/swisstopo/en/home/products/maps/geology/geomaps/LGM-map500.html, last access: 07.06.2016, 2009.

Swisstopo: Höhenmodelle, http://www.swisstopo.admin.ch/internet/swisstopo/de/home/products/height.html, last access: 07.06.2016, 2011.

Swisstopo: swissTLM3D: Topographic Landscape Model 3D. Version 1.1, http://www.swisstopo.admin.ch/internet/swisstopo/de/home/products/landscape/swissTLM3D.html, last access: 08.03.2016, 2013a.

Swisstopo: swissAlti3D. Das hoch aufgelöste Terrainmodell der Schweiz, http://www.swisstopo.admin.ch/internet/swisstopo/de/home/products/height/swissALTI3D.html, last access: 07.06.2016, 2013b.

Swisstopo: GeoCover, Zugang zu flächendeckende geologische Datensätze für alle, https://shop.swisstopo.admin.ch/de/products/maps/geology/GC_VECTOR, last access: 14.11.2016, 2016.

Taghizadeh-Mehrjardi, R., Nabiollahi, K., and Kerry, R.: Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran, Geoderma, 266, 98–110, doi:10.1016/j.geoderma.2015.12.003, 2016.

USGS EROS: USGS Land Remote Sensing Program, Landsat 7 Scene 01.09.2013. U.S. Geological Survey's Earth Resources Observation and Science Center, 2013.

Vaysse, K. and Lagacherie, P.: Evaluating Digital Soil Mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France), Geoderma Regional, 4, 20–30, doi:10.1016/j.geodrs.2014.11.003, 2015.

Viscarra Rossel, R., Chen, C., Grundy, M., Searle, R., Clifford, D., and Campbell, P.: The Australian three-dimensional soil grid: Australia's contribution to the GlobalSoilMap project, Soil Res, 53, 845–864, doi:10.1071/SR14366, 2015.

Walthert, L., Zimmermann, S., Blaser, P., Luster, J., and Lüscher, P.: Waldböden der Schweiz. Band 1. Grundlagen und Region Jura, Eidg. Forschungsanstalt WSL and Hep Verlag, Birmensdorf and Bern, 2004.

Walthert, L., Bridler, L., Keller, A., Lussi, M., and Grob, U.: Harmonisierung von Bodendaten im Projekt "Predictive mapping of soil properties for the evaluation of soil functions at regional scale (PMSoil)" des Nationalen Forschungsprogramms Boden (NFP 68), Bericht, Eidgenössische Forschungsanstalt WSL und Agroscope Reckenholz, Birmensdorf und Zürich, doi:10.3929/ethz-a-010801994, 2016.

Wegelin, T.: Schadstoffbelastung des Bodens im Kanton Zürich Resultate des kantonalen Bodenrasternetzes, Bericht, Amt für Gewässerschutz und Wasserbau Fachstelle Bodenschutz, Zürich, 1989.

Were, K., Bui, D. T., Dick, Ø. B., and Singh, B. R.: A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape, Ecol Indic, 52, 394–403, doi:10.1016/j.ecolind.2014.12.028, 2015.

Wiesmeier, M., Prietzel, J., Barthold, F., Spörlein, P., Geuss, U., Hangen, E., Reischl, A., Schilling, B., von Lützow, M., and Kögel-Knabner, I.: Storage and drivers of organic carbon in forest soils of southeast Germany (Bavaria) – Implications for carbon sequestration, For Ecol Manage, 295, 162–172, doi:10.1016/j.foreco.2013.01.025, 2013.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Academic Press, 3 edn., 2011.

Wüst-Galley, C., Grünig, A., and Leifeld, J.: Locating organic soils for the Swiss greenhouse gas inventory, Agroscope Science 26, Agroscope, Zurich, https://www.bafu.admin.ch/dam/bafu/en/dokumente/klima/klima-climatereporting-referenzen-cp2/wuest-galley_c_gruenigaleifeldj2015.pdf.download.pdf, last access: 29.03.2017, 2015.

Yang, R.-M., Zhang, G.-L., Liu, F., Lu, Y.-Y., Yang, F., Yang, F., Yang, M., Zhao, Y.-G., and Li, D.-C.: Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem, Ecol Indic, 60, 870–878, doi:10.1016/j.ecolind.2015.08.036, 2016.

Zimmermann, N. E. and Kienast, F.: Predictive mapping of alpine grasslands in Switzerland: Species versus community approach, J Veg Sci, 10, 469–482, doi:10.2307/3237182, 1999.

Zimmermann, S., Widmer, D., and Mathis, B.: Bodenüberwachung der Zentralschweizer Kantone (KABO ZCH): Säurestatus und Versauerungszustand von Waldböden, Bericht im Auftrag der Zentralschweizer Umweltdirektionen (ZUDK), Eidg. Forschungsanstalt für Wald, Schnee und Landschaft WSL, 2011.