Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt
Fachgebiet für Landschaftsökologie

Spatial Prediction Methods for the Assessment and Mapping of Forest Site Characteristics

Examples from Bavarian Forests

Dipl. Geogr. Tim Häring

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Johannes Kollmann

Prüfer der Dissertation:

1. Univ.-Prof. Dr. Boris Schröder
2. Prof. Dr. Jörg Ewald (Hochschule Weihenstephan-Triesdorf)
3. Priv.-Doz. Dr. Jörg Müller

Die Dissertation wurde am 30.07.2013 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 06.11.2013 angenommen.

Summary

Sustainable forest management in times of changing environmental conditions requires, among other things, detailed knowledge about site ecological conditions. Only upon this information sustainable decisions could be made today which also correspond approximately to predicted future climatic conditions, e.g. regarding planting of tree species or forest management strategies.
Traditional site classification in Germany describes site characteristics (e.g. soil texture, soil moisture, soil chemical properties, landscape position) by using a qualitative 3-numerals code. Such site maps have been criticized recently as being static and not reproducible. In addition, traditional mapping approaches are very time consuming and expensive. Therefore, the development of dynamic, homogeneous, area-wide and digital site information system have been initiated in many forest administrations in Germany.
Aim of this study is to apply methods from the area of digital soil mapping and species distribution modelling for the assessment and mapping of forest site characteristics. In the following sections of this study I will present different application examples of existing methods as well as the development of new approaches for the spatial prediction of site characteristics. My aim is to present a set of modelling approaches for the ongoing and future projects on site mapping in forestry.
This study is a cumulative dissertation, meaning it consists mainly of three publications which accrue from two research projects at the Bavarian State Institute of Forestry (*KLIP4 - Maps For The Future*) and the University of Applied Sciences Weihenstephan-Triesdorf (*Waldinformationssystem Nordalpen - WINALP*).
Two topics constitute the context of my dissertation, namely site classification and spatial prediction models. The development and the basic concept of both will be introduced first.
The main part of my dissertation consists of three thematically related articles that are published in ISI-listed scientific journals. In the first article I present the development and application of a modelling approach for spatial disaggregation of complex soil map units, i.e. map units aggregating two or more soil types which

may be characterized with differing soil properties. The method has been applied to the entire area of the 1:25k soil map of Bavaria and validated with more than 2000 sampling locations. The field data results in a prediction accuracy of more than 70%. The disaggregated soil map illustrates a spatially explicit and with regards to content refined database for the assessment of site conditions.
In the second and third article I present the application of the species distribution modelling framework to Ellenberg indicator values in order to estimate the spatial distribution of the effective moisture content and the soil reaction properties. In both publications I used 1505 vegetation plot records from the Bavarian Alps to fit different geoadditive regression models for regionalization. The resulting maps show detailed and diverse spatial pattern of effective soil moisture and soil reaction. The maps could be regarded as a major improvement regarding input data for site classification in the Bavarian Alps. For the regionalization of soil reaction I used a new, multivariate regression approach to model ordinal-scaled indicator values (proportional-odds model). In contrast to the common practice in vegetation ecology to model *average* indicator values the approach predicts the probability distribution *over all nine* indicator values. I could show that in areas with heterogeneous indicator spectra the new approach is a strong improvement compared to Gaussian regression models.
Subsequent to the summary of these articles I will present three additional studies which are targeted on the modelling and creation of input data for site classification: (1) The official soil map of Bavaria has been used in both projects as primary source of information on the spatial distribution of soil properties in Bavaria. To fill existing spatial gaps in the soil map, methods from the area of digital soil mapping have been applied to extrapolate soil map units. (2) Subsequent to modelling, field validation data have been collected (4500 samples). The analysis of the validation data has shown that the statistical model error of the predicted soil map units gives no reliable indication regarding prediction accuracy. Detailed petrographic information is needed e.g. on slope sediment layering or alluvial deposits to achieve more accurate predictions. (3) Homogeneous terrain units play an important role for the model based delineation of site classes. By using methods of digital image analysis (image segmentation) I have delineated multi-scale landscape objects based on digital elevation models and terrain parameters, which have been used in the *Maps For The Future* project for different applications (distribution of sampling locations, spatial disaggregation, assessment of effective soil moisture).
The described modelling approaches can be used as individual methods for dif-

ferent kinds of applications in soil science and spatial ecology, e.g. the refinement of a land use map or the prediction of an ordinal-scaled ecological parameters. However, altogether the methods assembled here provide a set of new approaches for the development of digital, quantitative, homogeneous, and high resolution maps of site characteristics of the new generation.

Zusammenfassung

Nachhaltige Waldbewirtschaftung in Zeiten sich verändernder Umweltbedingungen benötigt - nebst anderen Dingen - räumlich hoch aufgelöste und inhaltsdifferenzierte Informationen über Standortbedingungen. Nur auf Grundlage detaillierter Karten können heute nachhaltige Entscheidung hinsichtlich Baumartenwahl, Baumartenzusammensetzung und Waldbewirtschaftung getroffen werden, die sowohl den derzeitigen als auch den zukünftigen Umweltbedingungen entsprechen.

Die traditionelle Standortklassifikation in Deutschland beschreibt Standorteigenschaften (bspw. Bodentextur, Wasserhaushalt, Bodenchemie, Lage in der Landschaft) mit einem qualitativen 3-Ziffern System. Diese Karten wurden in jüngster Zeit jedoch kritisiert, da sie sich als zu statisch und nicht reproduzierbar erwiesen haben. Darüber hinaus ist die Erstellung von Standortskarten sehr zeitaufwendig und deshalb teuer. Die Entwicklung und Implementierung dynamischer, homogener, landesweiter und digitaler Standortinformationssysteme wird deshalb in vielen Landesforstverwaltungen in Deutschland vorangetrieben.

Ziel dieser Arbeit soll sein, räumliche Prognosemethoden aus den Bereichen der digitalen Bodenkartierung und der Habitatmodellierung für die Kartierung von Wald-Standortseigenschaften zu verwenden. In den Kapiteln dieser Arbeit präsentiere ich verschiedene Anwendungsbeispiele bestehender Methoden auf neue Fragestellungen als auch die Entwicklung neuer Ansätze für die räumliche Prognose von Standorteigenschaften. Ziel dabei ist es, leicht anwendbare und adaptierbare Methoden für die aktuellen und zukünftigen digitalen Standortkartierungs-Projekte der Forstverwaltung bereit zu stellen.

Die vorliegende Arbeit ist eine kumulative Dissertation, d.h. sie setzt sich hauptsächlich aus drei Veröffentlichungen zusammen, die während zwei Forschungsprojekten an der Bayerischen Landesanstalt für Wald und Forstwirtschaft (*KLIP4 - Karten für die Zukunft*) und der Hochschule Weihenstephan-Triesdorf (*Waldinformationssystem Nordalpen - WINALP*) entstanden sind.

In der Einleitung der Arbeit werden zunächst die beiden grundlegenden Themen eingeführt, die den Rahmen meiner Forschungsarbeit darstellen: das Thema

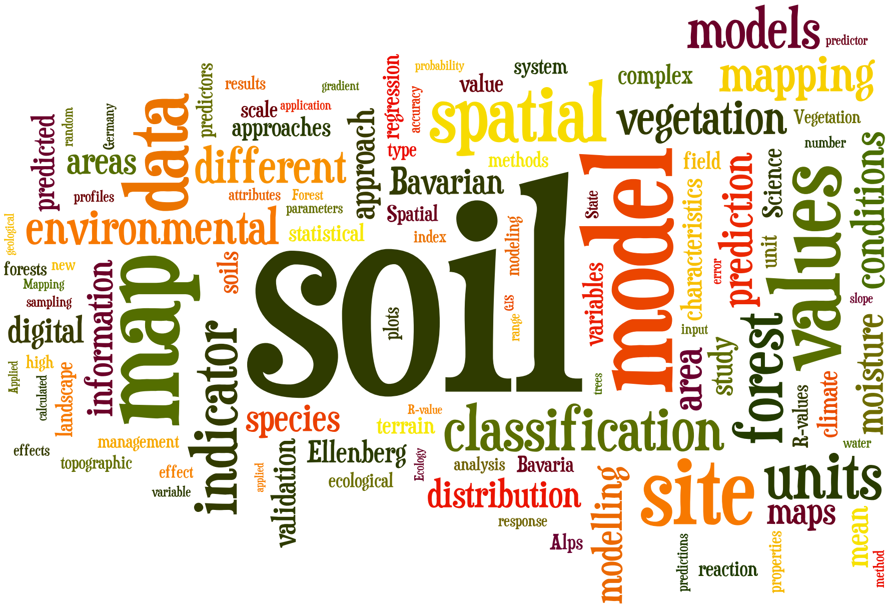
Standortkartierung und räumliche Prognosemethoden. Als Hauptteil meiner Dissertation finden sich daran anschließend drei Artikel, die in ISI gelisteten, wissenschaftlichen Zeitschriften veröffentlicht wurden. In der ersten Publikation beschreibe ich die Entwicklung und Anwendung eines Modells zur räumlichen Disaggregierung von Bodenkomplexeinheiten, d.h. Karteneinheiten, die mehrere Bodentypen mit unterschiedlichen standörtlichen Eigenschaften zusammen fassen. Die Methode wurde auf die gesamte verfügbare Fläche der Bodenkarte im Maßstab 1:25.000 angewendet und an über 2000 Punkten im Gelände validiert. Die Geländedaten ergaben eine Prognosegenauigkeit von über 70%. Die disaggregierte Bodenkarte ist eine räumlich und bodenkundlich verfeinerte Datengrundlage für die Abschätzung von Standortbedingungen.

In den verbleibenden zwei Publikationen präsentiere ich die Anwendung von Habitatmodellen auf Ellenberg Zeigerwertpflanzen. Ziel dabei ist eine räumlich differenzierte Abschätzung des effektiven Wasserhaushaltes und der Bodenreaktion zu erhalten. In beiden Publikationen habe ich 1505 Vegetationsaufnahmen aus den Bayerischen Alpen mit verschiedenen geoadditiven Regressionsmodellen regionalisiert. Die Ergebniskarten illustrieren räumlich stark differenzierte Muster der ökologischen Feuchte und Bodenreaktion und stellt somit eine große Verbesserung der Datengrundlage für Standortkartierungen in den Bayerischen Alpen dar. Für die Modellierung der Bodenreaktion verwende ich einen neuen, multivariaten Regressionsansatz, um ordinal-skalierte Zeigerwerte zu regionalisieren. Im Gegensatz zum üblichen Ansatz in der Vegetationsökologie schätzt das neue Modell nicht einen *mittleren* Zeigerwert, sondern die Wahrscheinlichkeitsverteilung *aller neun* Zeigerwerte. Ich konnte zeigen, dass in Untersuchungsgebieten mit stark heterogenen Zeigerwertspektren dieser Ansatz eine deutliche Verbesserung im Vergleich zur Mittelwerts-Modellierung darstellt.

Anschließend an die Zusammenfassung dieser Veröffentlichungen präsentiere ich weitere Studien, die darauf abzielen, Eingangsdaten für die Standortsklassifizierung zu modellieren: (1) Die Übersichtsbodenkarte ÜBK25 wurde in beiden Forschungsprojekten als Basisinformation über die räumliche Verbreitung der Böden in Bayern verwendet. Um Bodendaten in bisher nicht kartierten Bereichen zu erhalten, habe ich mit Methoden der digitalen Bodenkartierung Karteneinheiten extrapoliert. (2) Anschließend an die Modellierung wurden Validierungsdaten im Gelände erhoben (4500 Stichproben). Die Auswertung der Validierungsdaten zeigt, dass der statistische Modellfehler bei der durchgeführten Bodenprognose keine verlässliche Information über die Prognosegüte liefert. Petrographische Informationen z.B. über quartäre Deckschichten oder alluviale Sedimente sind

erforderlich, um Bodenkarten verlässlicher zu prognostizieren. (3) Homogene Geländeeinheiten spielen für die Ableitung von Standortseinheiten eine bedeutende Rolle. Mit Methoden der digitalen Bildverarbeitung (Bildsegmentierung) habe ich auf Grundlage von digitalen Geländemodellen und daraus abgeleiteten Reliefparametern multi-skalige Landschaftsobjekte abgeleitet, die im Laufe des *Maps For The Future* Projektes für verschiedene Anwendung verwendet wurden (stratifizierte räumliche Stichprobennahme, Bewertung des Bodenwasserhaushaltes, Disaggregierung der Bodenkarte).

Die dargestellten Methoden sind für unterschiedliche Fragestellungen in Bodenkunde und Ökologie anwendbar. Zusammengenommen ergeben sie ein Set an neuen Methoden für die Entwicklung digitaler, quantitativer, homogener und hoch aufgelöster Standortskarten der neusten Generation.



soil
model
map
site
spatial
values
classification
data
units
forest
mapping
models
vegetation
environmental
different
indicator
information
prediction
species
distribution
conditions
moisture
area
areas
approach
approaches
Bavarian
Bavaria
Ellenberg
digital
predicted
maps
characteristics
modelling
modeling
mean
climate
study
variables
complex
Vegetation
Science
unit
landscape
validation
statistical
terrain
ecological
topographic
management
effect
effects
soils
forests
Spatial
regression
system
value
scale
results
predictors
methods
field
Alps
reaction
properties
predictions
response
analysis
index
plots
type
State
attributes
Forest
parameters
profiles
sampling
Mapping
new
high
calculated
Applied
applied
geological
random
Germany
gradient
probability
application
accuracy
predictor
number
slope
input
error
range
GIS
trees
R-value
R-values
water
method
variable
Ecology

Contents

List of Figures

List of Tables

terra reddit, quod accipit

(Xenophon)

Chapter 1.

State of the Art and Objectives

> The topmost step, which lies nearest Småland, is mostly covered with poor soil and small stones, and no trees except birches and bird-cherry and spruce which can stand the cold on the heights, and are satisfied with little can thrive up there...
>
> *(The Wonderful Adventures of Nils by Selma Lagerloef)*

1.1. Mapping of forest site characteristics - past and present

1.1.1. The term “forest site”

Why do we have such a fascinating tapestry of forests on Earth? To understand the diversity and distribution of forests we sooner or later come across the term “**forest site**”. There is a diversity of definitions of site, but in a nutshell the term can be defined as the sum of growth conditions at a location (Pfadenhauer 1997). The German term “**Standort**”, which is a terminus technicus in forestry and agriculture, has this comprehensive definition. Usually, the English term forest site is narrowed down to mean a description of the physiograhic conditions which have an influence on the growth of plants, namely soil, topography, and climate (Arbeitskreis Standortskartierung 2003), even though biotic processes are also present in soil (mineralization, humus) and climate (micro-climate). Grey (1980) defined a site as a natural unit, a spatial entity, which can be described, classified, recorded and mapped but cannot be further subdivided without the loss of some intrinsic characteristic. Bailey et al. (1978) defined forest site as the function of the interplay between climate, topography, parent material and vegetation over a specific time. The term is mainly utilized in agriculture and forestry in the context of manufacturing resources for crop plants, e.g. timber or arable crops. Forest site has therefore a strong use-oriented meaning, for example Schulze et al. (2005) and Nentwig et al. (2012) relate the term to the field of applied ecology.

Beside the term site, other scientific terms also relate to the physiographic environment. There is a huge overlap between those terms, yet with different focus. The term “biotope” literally means an area where life lives (Pfadenhauer 1997). The concept was introduced initially to define a complex of factors that determines physical conditions of existence of a species, a “biocenosis”. The difference to the term site is the strict relation of a biotope (the abiotic environment) to a specific biocenosis (the biotic community) as the “biotope of a biocenosis”, cf. Dahl (1908) and Olenin and Ducrotoy (2006). Thus the biotope is considered as the abiotic part of an ecosystem. Nowadays, the term biotope is mainly used in nature conservation and environmental planning literature (CORINE 1991; Kaule 2002; Pfadenhauer 1997). The term “ecosystem” relates to the functional description of processes and interactions between plants and animals and the physical-chemical environment. “Ecosystems are thus networks of interrelations between organisms and their environment in a defined space” (Schulze et al. 2005).

Ecosystems have no fixed boundaries; instead their parameters are set according to the scientific, management, or policy question being examined. Depending upon the purpose of the analysis, a single lake, a watershed, or an entire region could be considered an ecosystem.

Confusingly, there are some studies in forestry which use the term forest site interchangeably with "forest site quality" and refer to the productivity potential of a given site ("site productivity"), i.e. an interpretation of the physiographic conditions for timber production (cf. Aertsen et al. 2010; Altun et al. 2008; Carmean 1975; Pokharel and Dech 2011). However, in this thesis I use the term forest site to refer to the classification and mapping of the physiographic environment.

1.1.2. The classification of forest sites

Because of the importance of the forest site, being the irreplaceable natural resource of forestry, the development of different approaches to assess and map site characteristics has a long history. First approaches for site classifications in Germany date back to 1888 (Roth 1916; Watson 1917). However, the main intention of these early approaches was to classify "lands into good and poor" or "more fertile and less fertile ones" (Roth 1916) and the volume of timber produced was the main measure. "The only final criterion of site quality is the current annual cubicfoot increment of a fully stocked stand of the species under consideration" (Bates 1918). Growth height was also used as an index of site quality (Frothingham 1921; Roth 1916; Watson 1917). But also at that time the correlation of site productivity on the one hand and soil and climate on the other had been identified as an important parameter. Recognizing the obvious variation in wood productivity and quality associated with habitat variance made it imperative to strive for means of evaluating environmental potentials as a guide in forest management. Subsequent approaches paid increasing attention to the physiographic environment for site classification following the general principle of sustainable forest management and became less driven by timber yield and capital investment criteria (c.f. Carmean 1975; Daubenmire 1976; Kohm and Franklin 1997; Lindenmayer et al. 2000; Wijewardana 2008).

A direct measurement of driving forces, processes, and interactions of ecosystems such as energy flow, nutrient cycles, light, temperature, water, and physicochemical parameters above the field scale is not feasible Schulze et al. (2005). However, the indirect assessment of *site characteristics* (German: Standortsmerkmal) based on the observable nature and properties of the physical environment has been proven as an effective approach (Arbeitskreis Standortskartierung 2003). The rationale behind that approach is the ecosystem concept (Bailey 2009; Begon et al. 2006; Nentwig et al. 2012): vegetation and soils are products of climate, organisms, topography, parent material, and time. Plants and soil, considered simultaneously, integrate all ecosystem components and reflect ecosystem functioning.

In principal there are two fundamental ways to delineate site classes, i.e. areas with more or less homogeneous site-ecological conditions, from the physiographic environment (Ellenberg 1967): First, soil properties, climate or topographical characteristics can be estimated and their concurrence used to delineate site classes or second, plants are used as indirect indicators for their site. Reviews of methods of site classification and site productivity have been provided by Ellenberg (1967), Bailey et al. (1978), Spurr and Barnes (1980) and Gauer et al. (2011). The various approaches to forest site classification have stressed either landform (Wertz and Arnold 1975), soil (Ad-hoc-AG Boden 2005; Soil Survey Staff 1999), or vegetation (Daubenmire 1976; Pfister and Arno 1980).

Mapping site characteristics based on soil information alone is a special application of effective soil classification (Schlichting 1970). A soil classification or map is not an ecological classification if the relationship of the classes to the vegetation of the area is unknown (Barnes et al. 1982). Therefore, the mapper has to make an estimate for the site regarding its suitability for a specific crop. The resulting maps are therefore restricted to a single crop only. Since the suitability of a site to grow a certain crop depends also on additional parameters, the classification of a soil type regarding its site characteristics could vary in space. This classification scheme is therefore not applicable to an area outside of the area where it was defined. Therefore, as one group of scientists collaborating on methods of site evaluation concluded: "Soil mapping was not separately evaluated, because today purely edaphic data are no longer generally considered sufficient for site characterization" (Ellenberg 1967).

Site classification methods based on plants and vegetation have been widely used to classify vegetation types, e.g. purely phyto-sociological approaches following Braun-Blanquet (1964), or the habitat type approach, which is also known as mapping of potential natural or climax vegetation (Daubenmire 1976; Frehner 1967; Pfister and Arno 1980; Schönhar 1993).

1.1.3. Ellenberg Indicator Values

In addition to these two vegetation based site classification methods species indicator values have been used widely in central Europe for the assessment of forest site characteristics. The occurrence and abundance of different plant species enables ecologists to make statements about the prevailing environmental conditions - they make visible what is not immediately perceptible without conducting measurements (Diekmann 2003). For the assessment of site characteristics they are particularly useful because they summarize complex environmental factors (e.g. groundwater level, soil moisture content, precipitation, humidity, etc.) in one figure. In addition, plants do not refer to conditions at a specific moment but represent integrated expressions of the values that may fluctuate in time.
One formalized system of indicator values is that developed by the German Ecologist Heinz Ellenberg (Ellenberg and Leuschner 2010; Ellenberg et al. 2001), which is probably the most widely applied approach for site diagnostics (Ewald 2007). They assigned indicator values to most species occurring in western central Europe (2726 vascular plant species, also numbers of bryophytes and lichens), with respect to moisture, soil nitrogen status, soil reaction (acidity/lime content), soil chloride concentration, light regime, temperature and continentality. The values were developed mainly on the basis of field experience, and quantification generally follows a nine-point scale (Diekmann 2003; Schaffers and Sykora 2000). The indicator values reflect the ecological behavior of species, not their physiological preferences (Ellenberg et al. 2001).
For a further overview of Ellenberg indicator values see Diekmann (2003), Pfadenhauer (1997) and Dierschke (1994).

1.1.4. Site classification at present

Nowadays, most approaches follow a combined approach of site-classification, i.e. conclusions on site characteristics are drawn from several environmental compartments in parallel. This holistic view of the physiographic environment allows delineation of sufficiently ecologically homogeneous site classes (Arbeitskreis Standortskartierung 2003). The complex gradients of an area are broken down into site units that recur in the landscape-units that can be distinguished by major differences in physiography, soils, and vegetation. Each of these three ecosystem factors provides information for building the classification and mapping the site units. Physiography often determines micro-climate and water movement, certain landforms are highly correlated with soil conditions, and major landforms or features (aspect, slope position) can be identified from aerial photographs and digital elevation models. Soil factors, particularly soil moisture, nutrients, and pH, strongly control plant and animal composition, size, and productivity. Thus, the three major factors and their interrelationships most clearly distinguish the local ecosystems (Bailey 2009; Barnes et al. 1982; MacMillan et al. 2007).

1.1.5. Example 1: The biogeoclimatic ecosystem classification of British Columbia

A widely-recognized approach for holistic site classification is the biogeoclimatic ecosystem classification (BEC) of British Columbia (Canada) (Barnes et al. 1982; Pojar et al. 1987). It was initially developed at the University of British Columbia from 1949 to the 1970s and was systematized by the Ministry of Forests in B.C. after 1975. Since that time the BEC has been used to estimate current and future site potential for sustaining forested ecosystems. The system incorporates primarily (local) climate, soil, and vegetation data and provides a framework for resource management, as well as for scientific research (Haeussler 2011; MacMillan et al. 2007; Pojar et al. 1987).

The BEC system is a hierarchical classification scheme. On a large scale broad biogeoclimatic units based on zonal climate classification were characterised, whereas on a finer scale biogeoclimatic units or "ecosystems" were disaggregated in major forest and range sites (site classification).

- **Zonal Climate Classification:** In the BEC system, regional climate has been identified as the most important determinant of the nature of terrestrial ecosystems (Meidinger and Pojar 1991). Therefore, broad geographical areas, called biogeoclimatic units, were delineated which represent classes of ecosystems under the influence of the same regional climate. In the BEC system, there is a hierarchy of biogeoclimatic units, with the biogeoclimatic subzone being the basic unit characterized by a distinct climax (or near-climax) plant association. Subzones could be aggregated into zones which in turn could be aggregated into regions and formations (highest level of hierarchy), if there are affinities in climatic characteristics, webs of energy flow and nutrient cycling and typical patterns of vegetation and soil characteristic, prevailing soil-forming processes, and one or more typical, major, climax species of tree, shrub, herb, and/or moss. In the BEC system 14 biogeoclimatic zones were defined. The subzone could also be divided into

variants to account for further differences in regional climate and is generally recognized for areas that are slightly drier, wetter, snowier, warmer, or colder than other areas in the subzone. Variants can then also be divided into phases to accommodate the variation in the regional climate, resulting from local relief. They are also useful in designating significant, extensive areas of ecosystems that are, for topographic or topoedaphic reasons, atypical for the regional climate.
In the BEC system, the different climatic units were coded with letters or numbers. The example

ICH mc 1 a

refers to the coastal (a) phase, of the Nass (1) variant, of the Moist Cold (mc) subzone, of the Interior Cedar Hemlock (ICH) zone.

- **Site Classification:** In the local and more detailed hierarchical level *site units* were delineated within each biogeoclimatic units. These units represent sites with homogeneous environmental conditions and the same potential vegetation. The differentiation from a site unit to one another is based on a range of environmental properties. Beside the potential vegetation, the BCS use a edatopic grid to evaluate the moisture and nutrient regime of a site relative to a regional climate (see Table 1.1). Therefore, the average amount of soil water annually available for evapotranspiration and the amount of essential soil nutrients that are available to vascular plants over several years were adopted to nine classes (soil moisture: from 'Very Xeric' to 'Hydric') and five classes (soil nutrients: from 'Very Poor' to 'Very Rich') respectively.

  As for the biogeoclimatic units, there is also hierarchy of units for the local level (site group > site association > site series > site type), whereas the site association is conceptually similar to the forest type of several European classifications (Jahn 1982). Partitioning of site units is possible to create units with climatic and/or edaphically more consistent units.

Since the implementation of the biogeoclimatic framework for forest management in British Columbia in the 1970s and 1980s the method is recognized widely. The BEC system is a sophisticated approach though flexible and transferable to

Table 1.1.: Edatopic grid of the BEC system for relative estimation of soil moisture and nutrient regime. The grid is used for differentiate between site units within the same local climate.

	Soil nutrient regime				
	A	B	C	D	E
Soil moisture regime	very poor	poor	medium	rich	very rich
Very Xeric - 0					
Xeric - 1					
Subxeric - 2					
Submesic - 3					
Mesic - 4					
Subhygric - 5					
Hygric - 6					
Subhydric - 7					
Hydric - 8					

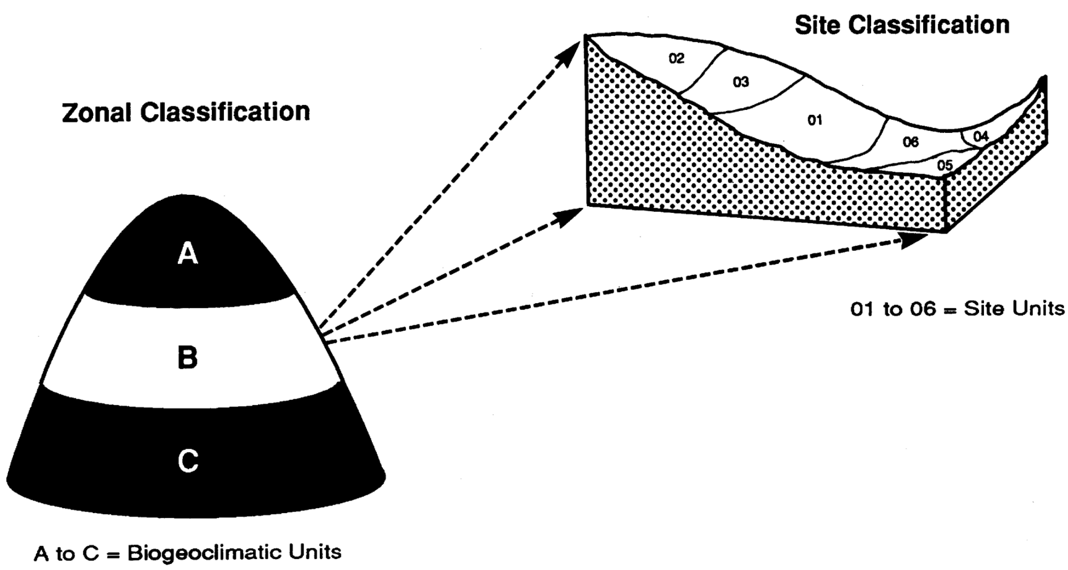


Figure 1.1.: Schematic relation between zonal and site classification in the BEC system (Source: Meidinger and Pojar 1991)

other regions. It is also well documented. There are several examples of site classification approaches which are based on the BEC system, e.g. in the Yukon Territory of Canada (Griesbauer and Green 2012), northeast Asia (Krestov et al. 2006; Nakamura et al. 2007), Britain (Pyatt 1995; Quine et al. 2002), USA (Bailey et al. 1994; Hessburg et al. 2000), China (Liu et al. 1998), and Japan (Kojima 1991).

The BEC system itself has been further developed over the years. Main enhencement was the refinement of the maps based on new data (e.g. remote sensing or high resolution climate data DeLong et al. 2010; Fitterer et al. 2012; Hamann and Wang 2006) and spatial modelling approaches. MacMillan et al. (2007) predicts ecological site types with success using knowledge-based routines for automated polygon extraction and classification (cf. subsection 1.3.1).

1.1.6. Example 2: Site classification in Bavaria

Similar to the BEC system the site classification in Bavaria follows also a two-step classification approach, see Figure 1.2.

The basic framework is a *regional* stratification of Bavaria following ecological and physiographic aspects. So called growth areas ('Wuchsgebiete') are identified on the basis of climate, geology, and vegetation. These growth areas are fairly heterogeneous and are, therefore, subdivided into growth districts to account for

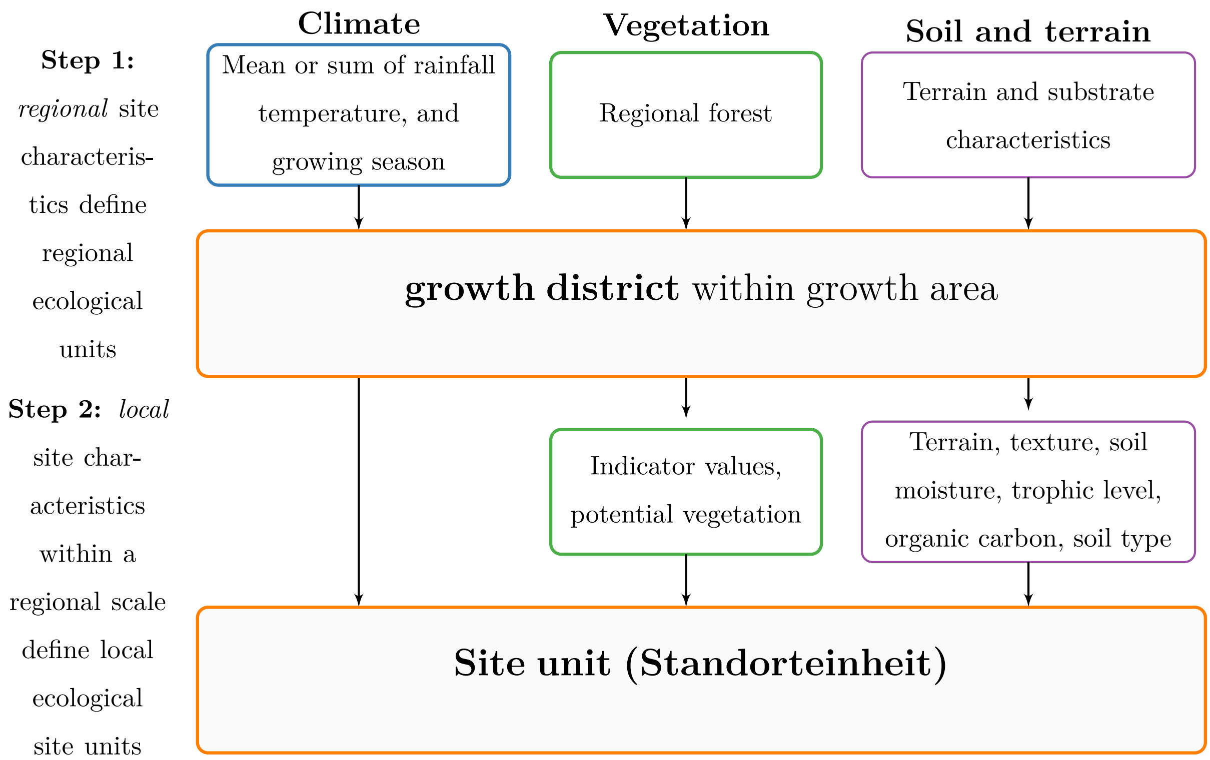


Figure 1.2.: Two-step site classification approach of Bavaria. (c.f. Arbeitskreis Standortskartierung 2003, modified)

local climate conditions ('Wuchsbezirke'), see Figure 1.3 and B.1. Therefore, a *Wuchsbezirk* represents regional physiographic units with similar climatic and geological conditions and a specific regional potential vegetation (c.f. Walentowski et al. 2001).

In the second, *local* step, these regional units were subdivided following ecological criteria to delineate homogeneous site classes ('Standortseinheiten'), being the basic site-ecological entity. These criteria were (c.f. Arbeitskreis Standortskartierung 2003)

1. **Geological substrate/parent material:** This parameter is the primary mapping criteria at the local scale. The differentiation is carried out taking soil texture, parent material, layering and structure into account. In addition, soil physical and chemical properties of the site were also considered following the German guideline for soil mapping (Ad-hoc-AG Boden 2005) and thereby soil temperature and aeration were also taken into account.

2. **Soil moisture (influenced by topography):** The assessment of soil moisture for sites without additional water input (e.g. inferflow, groundwater) is carried out using a semi-quantitative method. The determining factor is the duration, in which soil water is available for assimilation without additional rainfall assuming an uptake of $3\frac{mm}{day}$. Based on this assessment, sites were classified on a relative scale within the growth-area.
   Sites with a temporary or permanent surplus of water were treated separately. Relevant criteria are the origin (stagnating surface water or groundwater) and the duration of the surplus and also the slope position.

3. **Terrain characteristics and landforms:** Evaluating site conditions at a detailed local scale, terrain characteristics and landforms play an important role, mainly because of their influence on the flow of water, soil erosion and the accumulation and spatial distribution of nutrients. In areas with strong relief, elevation, slope and aspect could become the most important parameters for site classification.

4. **Vegetation:** Indicator values of plant species and also the potential vegetation of a site represent easy and quick assessment tools for soil moisture, temperature, aeration, and nutrition and are therefore integrated in the site classification system.

Figure 1.3.: Growth areas (*Wuchsgebiete*) and growth districts (*Wuchsbezirke*) of Bavaria. These physiographic areas represent the first step of the Bavarian site classification system (Source: Walentowski et al. 2001).

5. **Local site-specifics:** Here, additional information regarding important site characteristics, which were not covered so far, were addressed: trophic level, clay content, impermeable soil horizon, content of calcium carbonate, slope position, or organic carbon content.

Single site units were coded and mapped using three numerals. The first numeral indicates the geologic substrate, the second is an expression of the trophic level and local specifics and the third numeral defines the soil moisture conditions, see Table 1.2. On top of the site class, local site-specifics were indicated with symbols or hachures.
Figure 1.4 illustrates a recent example of a Bavarian site map from the forestry district of Kehlheim (belonging to the state-owned company *Bayerische Staatsforsten*). The map shows the growth district *Westliches Niederbayerisches Tertiärhügelland* within the growth area *Tertiäres Hügelland (Wuchsgebiet 12)* and is mapped with the typical scale of 1:10.000

In addition, a comprehensive silvicultural guide, the so called *Operat*, is prepared to accompany the site map of each state, city, or private forest. Each site unit is described in detail, and the potential productivity of each species or species mixture is given in a color coded table (*Baumarteneignungtabelle*).

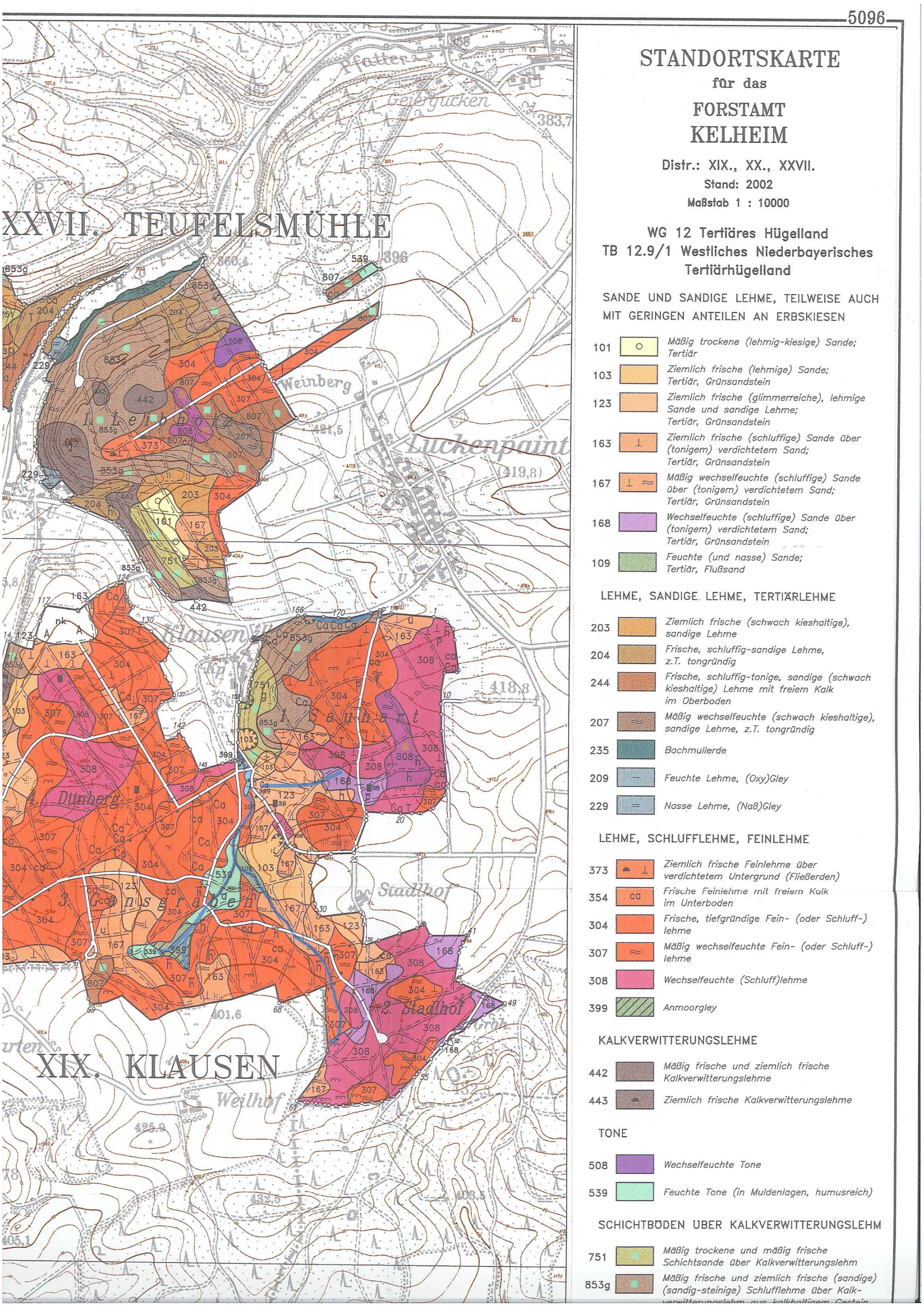


Figure 1.4.: Example site map of Bavaria based on the three-numerals-code. The cutout is from the mapsheet of Kehlheim (scale 1:10.000, Courtesy of Bayerische Staatsforsten, Mr. Kay Müller)

Table 1.2.: Site classification system of Bavaria. The system is based on a code of three numerals. Courtesy: LWF

	1. Ziffer: Bodenart und Schichtung		**2. Ziffer:** Besondere Standortmerkmale		**3. Ziffer:** Wasserhaushalt
0	**Sand**, Kies, Felsmosaik, Blockmosaik, Humuskarbonatboden	0	**normal**	0	(sehr) **trocken**
1	**lehmiger Sand, sandiger Lehm**	1	**nährstoffarm**, podsoliert	1	**mäßig trocken**, grundwechseltrocken
2	**Lehm**	2	**nährstoffreich**	2	**mäßig frisch**, mäßig grund-/ hangfrisch
3	**Feinlehm, Schlufflehm**, Schluff	3	**humusreich**[2]	3	**ziemlich frisch**, frühjahrsfrisch, grundfrisch
4	**milder Ton**, Tonlehm, Kalkverwitterungslehm[1]	4	**Karbonat oberhalb 50 cm**	4	(sehr) **frisch**, hangfrisch, grundfeucht
5	**strenger Ton**[1]	5	**Karbonat in 50 - 100 cm Tiefe**	5	**hangwasserzügig**, quellfrisch, wechselnd hangfeucht,(mäßig) hang(wechsel)feucht
6	**Decksand und -lehm** (Ton-/ Tonlehm höher 30 cm)	6	**tongründig**[3]	6	(mäßig) wechseltrocken
7	**Schichtsand** (Ton-/ Tonlehm in 30 - 60 cm Tiefe)	7	**Verdichtung im Unterboden**[4]	7	**mäßig wechselfeucht**, schwach wechselfeucht
8	**Schichtlehm** (Ton-/ Tonlehm in 30 - 60 cm Tiefe)	8	**Hanglage**	8	(stark) **wechselfeucht**, wechselnd feucht
9	**Moor**	9	**anmoorig**[5]	9	**feucht** (naß, staunaß, quellnaß)

[1] Deckschicht bis 10(20)cm möglich

[2] **humusreich:** Anteil organischer Substanz im Boden 5 - 30% (Feuchtböden max. 15%) und Humus-Mineralbodenhorizont mind. 10 cm mächtigDeckschicht bis 10(20)cm möglich

[3] **tongründig**: Ton(lehm) tiefer als 60 cm

[4] **Unterboden**: tiefer als 60 cm

[5] **anmoorig**: Gehalt an organischer Substanz im Boden 15 - 30%

1.2. Spatial Prediction of Ecological Parameters

Coming from these well-tried and traditional approaches, which were applied worldwide for the assessment and mapping of the spatial distribution of site, soil, and vegetation characteristics, the transition to the digital arena entailed fundamental changes also to mapping techniques. The explosion in information and computation technology in combination with vast amounts of data and the tools to work with it enable protagonists in all fields of ecology and environmental sciences to facilitate environmental management in an unexpected dimension: the *quantitative* characterization of the nature and property of the human environment.
Soil and vegetation databases at all spatial scales were generated and became available through the world-wide-web. To get an understanding of these large stores of data powerful and flexible statistical models, data-mining and machine learning methods have been developed (cf. Hastie et al. 2009) and made accessible e.g. through the R language and environment for statistical computing (R Core Team 2013). In addition, the increasing power of tools such as geographic information systems (GIS), GPS, remote and proximal sensors and data sources such as those provided by digital elevation models (DEMs) are suggesting new ways forward (Boettinger et al. 2010; McBratney et al. 2003; Mulder et al. 2011; Viscarra-Rossel et al. 2010).

On the other hand, there is an increasing need to use local measurements to assess change at landscape, regional and global scale, and to conduct a sustainable management of the environment. In the last three decades the development and application of spatial prediction models has become a research area with fast growing popularity. In principal, spatial prediction models can be described as the application of a statistical model to a set of digital data (observational data of a parameter of interest and environmental variables thought to influence the spatial distribution thereof) to predict a continuous or most probable categorical environmental attribute in space. Statistical models could be any kind of method available, ranging from classical linear least square regression or generalized linear model (GLM, McCullagh and Nelder 1989) to semi-parametric generalized additive models (GAM, Hastie and Tibshirani 1990), non-parametric algorithms from the area of machine-learning (tree-based methods, ensemble techniques like bagging or boosting, artificial neural networks, or support vector machines, Breiman 1996, 2001b; Breiman et al. 1984; Elith et al. 2008; Hastie et al. 2009; Schölkopf and Smola 2002) or any kind of geostatistics (kriging, regression kriging, Hengl et al. 2007; Webster and Oliver 2007). Data could be any type of data used in a geographic information system (GIS) like point (e.g. sampling location, weather station), line (e.g. street or river network) or polygon (e.g. soil map, physiographic or administrative area) data as well as raster data (e.g. climate maps, remote sensing data). In addition, different types of non-spatial data like databases, description of vegetation plots or laboratory analysis of soil profiles can be used.

The basic and also simple assumption behind predictive modelling in environmental sciences is the same as that behind statistical modelling in general, namely that the spatial distribution of an environmental parameter can be related to a set of different readily available, spatially exhaustive ancillary data. For example, the spatial distribution of soil properties at landscape scale is largely influenced by topography that produces gradients in moisture, energy and nutrients across the landscape (Dobos and Hengl 2009; McKenzie and Ryan 1999; Moore et al. 1993), the presence or absence of a tree species is largely influenced by different climatic variables or variables characterizing the habitat (Elith and Leathwick 2009; Franklin 2010; Guisan and Thuiller 2005) or the vulnerability of groundwater with regard to pesticide leaching is largely controlled by parameters influencing pesticide behavior in the environment like water flux, soil organic matter content and compound properties (Tiktak et al. 2006).
Thus in principle, spatial prediction is characterized by two steps. First, a statistical model has to be fitted based on a representative sample of the parameter of interest. This step also includes model validation. Second, to predict the parameter of interest at unobserved locations, the model has to be applied to the GIS layers of predictor variables.

Most environmental parameters develop progressively over time through a process of change called ecological succession (for vegetation) or soil development (for soil). However, statistical models could - in a strict sense - only be applied to parameters which are in a condition of 'dynamic equilibrium', 'steady state' or 'climax ecosystem' (Elith and Leathwick 2009). Meidinger and Pojar (1991) defines this condition as 'a stable, permanent occupant of the landscape, self-perpetuating unless disturbed by outside forces or modifying factors. The living components (...) are in equilibrium with the prevailing factors of the physical environment, and the member species are in dynamic balance with one another'.

For instance, wrong conclusions could be drawn when modeling the habitat of a species or the spatial distribution of a soil parameter in an area with a strong history of disturbance like wind throw, flooding, fires, insect outbreak, strong soil erosion or landslides. Especially in forests, which are strongly influenced by humans, anthropogenic disturbances such as planting of exotic tree species or forest clearing should also be considered.

1.2.1. Digital Soil Mapping

Soil, being mostly an inert environmental compartment at human time scales, is a good example of a steady state variable[1]. Spatial modelling approaches for soils and soil attributes, known as *digital soil mapping* or *predictive soil mapping*, date back to the 1960s (cf. the review of early soil modelling approaches in McBratney et al. 2003). Since then, the research area of digital soil mapping has been developed, which has been defined as the creation and population of spatial soil information by the use of field and laboratory observational methods coupled with spatial and non-spatial soil inference systems (Carre et al. 2007; McBratney et al. 2003; Scull et al. 2003)

In principal, digital soil mapping has its roots in the famous state-factor approach of Jenny (1941), which relates soil development to climate, organisms (including humans), relief, parent material and time. A recent study of Florinsky (2012) showed that the central idea of the soil-forming factors was defined as early as 1886 by Vasily Dokuchaev. This mechanistic model has been used by innumerable surveyors all over the world as a qualitative list for understanding the factors that may be important for producing the soil pattern within a region (McBratney et al. 2003; Yaloon 1975). In addition to these state factor-based approaches, purely spatial approaches have been used extensively to interpolate between soil observation locations - mainly geostatistics and related methods.

McBratney et al. (2003) reviewed existing digital soil mapping studies and proposed a generalized and formalized approach based on the state-factors approach as soil spatial prediction function:

[1]Of course there are many proesses in soils which are highly dynamic in time and space, e.g. soil hydrological processes. However, regarding processes influencing soil development soil could be refer as a steady state variable

$$S_a = f(s, c, o, r, p, a, n)$$
$$S_c = f(s, c, o, r, p, a, n)$$

with

S_a : soil attribute;

S_c : soil class;

s : soil, other properties of the soil at a point;

c : climate, climatic properties of the environment at a point;

o : organisms, vegetation or fauna or human activity;

r : topography, landscape attributes;

p : parent material, lithology;

a : age, the time factor

n : space, spatial position.

Discussions of state-of-the-art digital soil mapping applications at different extents, geographic settings, and model resolutions were provided by Lagacherie et al. (2007), Hartemink and McBratney (2008), Grunwald (2009), and Boettinger et al. (2010). Current research areas have been developed in several directions, e.g. the prediction of soil properties and classes at different spatial (from field and coarse landscape to even global scale, cf. the establishment of the GlobalSoilMap.net consortium which aims to deliver a new digital soil map of the world at fine resolution (Hartemink et al. 2010; Sanchez et al. 2009)) and temporal scales, the incorporation of remote and soil sensor technology (which is however restricted to application mainly outside of forests), the incorporation of legacy soil data, soil sampling, as well as calibration and validation of soil prediction models.

1.2.2. Species Distribution Modelling

Based on the same spatial modelling principles as for digital soil mapping, the modeling of species distribution in space and time is another important application area of spatial prediction models. "Species distribution models" (SDM) or, when restricting the application the the spatial distribution of vegetation, "predictive vegetation mapping" have been used to characterize the fundamental (potential) or realized (actual) niche of a species, called *niche models* (Elith and

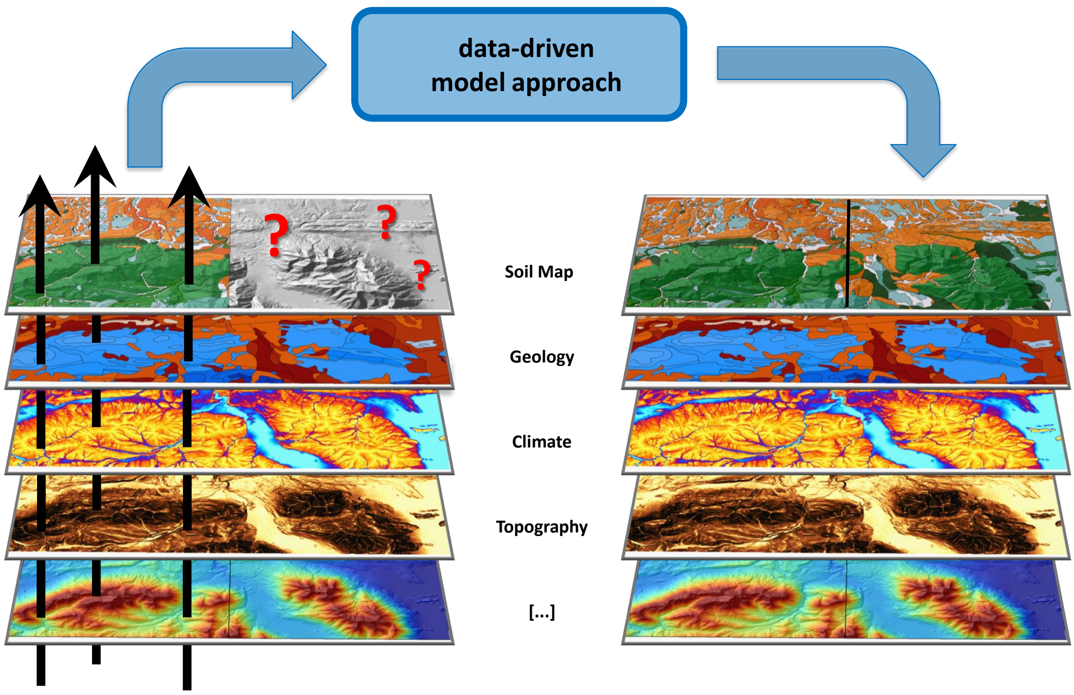


Figure 1.5.: In digital soil mapping quantitative relationships between soil properties or soil map units and environmental attributes were fitted in a data-driven model approach to predict soil information into non-mapped areas

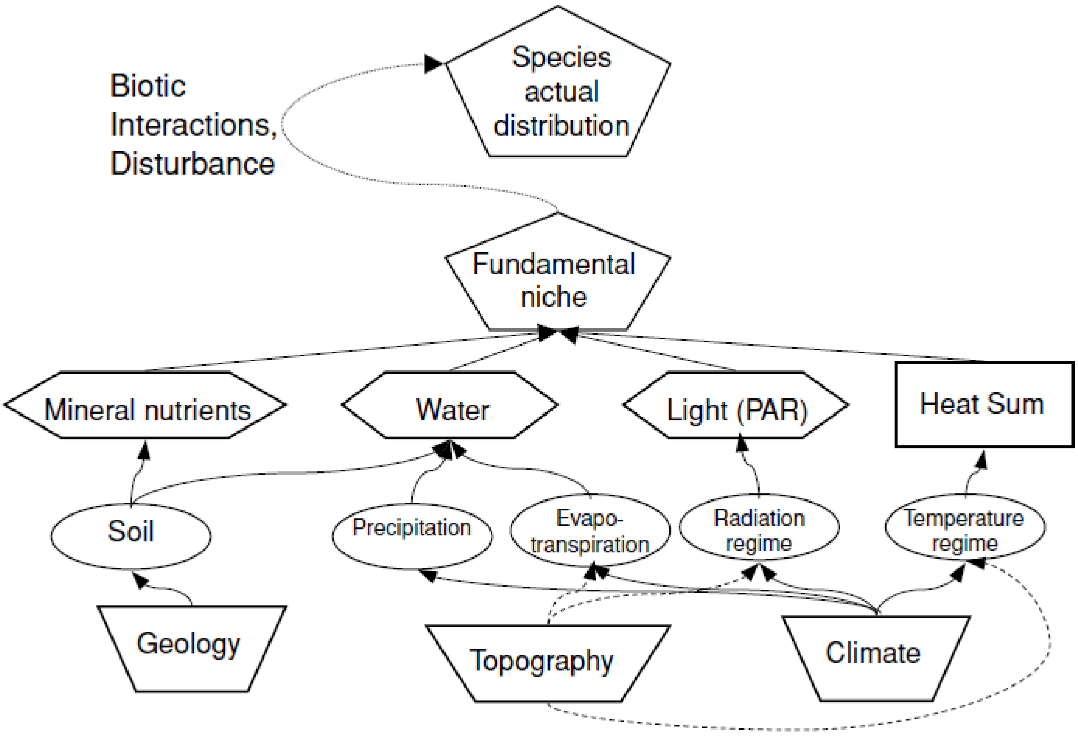


Figure 1.6.: Conceptual model of SDM: Environmental factors controlling the primary environmental regimes circumscribing the fundamental species niche (Source: Franklin 2010).

Leathwick 2009). SDMs are also referred to as *habitat suitability models*, describing the suitability of habitat to support a species and to predict the probability of species presence at a location (Franklin 2010).

Application areas of SDMs are manifold. The aim is to understand the relationship between a species and its abiotic and biotic environment based on observations for the purpose of ecological inference, or to test ecological or biogeographical hypotheses about species distributions and ranges. See for a comprehensive overview of theory and practice of SDMs Elith and Leathwick (2009); Guisan and Thuiller (2005); Peterson et al. (2011).

1.3. Both approaches: Spatial prediction methods for the assessment of site-ecological conditions

1.3.1. Challenges and opportunities of site classification

The similarity between the traditional approach to map site characteristics (section 1.1) and computer-based spatial prediction techniques (section 1.2) is obvious. Both approaches aim to delineate site characteristics by analyzing the observable environment - either outdoors or in the digital representation. Therefore, it is unsurprising that in the last years both approaches have been increasingly combined to complement each other. Spatial prediction techniques have become increasingly accepted by environmental resource managers and also by authorities and state institutes (Carre et al. 2007).
McBratney et al. (2003) and Carre et al. (2007) specify advantages of spatial prediction methods compared to traditional approaches, such as costs, consistency and documentation, the possibility to update when new data become available and the capability of deriving uncertainties for predicted outcomes.

As illustrated in subsection 1.2.1 and 1.2.2 the availability of new and high-resolution environmental GIS data, the development of spatial prediction methods and the fast growing application of these techniques have entailed fundamental changes to natural resource management in soil science and ecology. Unsurprisingly, resource management in forest science in terms of site classification and site mapping has now also entered in a period of transition. Therefore, forest site classification is currently challenged for several reasons.

- Classical mapping approaches are very time consuming and expensive which results in slow progress in times of decreasing public funds.

- From a scientific point of view, the most significant problem of these approaches is their lack of reproducibility. Site characteristics are derived indirectly by the interpretation of the observable nature around the surveyor (cf. section 1.1). In addition, the system is based on a qualitative and ordinal schematisation of sites (cf. Table 1.2). The consequences thereof are site maps which are not comparable, e.g. the class “frisch” does not have a consistent meaning throughout Bavaria. Site classes have a non-stationary meaning and are to a certain extent subjective. On the one hand, this was the intention of the developers of the mapping approach (hierarchical

structure of the system), on the other hand, this makes the map static and inflexible. Large scale analyses across borders of maps or growth districts are therefore biased.
The refinement and adaptation of the system to new applications is not possible based on the qualitative descriptions of sites, although this is an important topic in forestry at the moment due to changing environmental conditions.

- In addition, Haeussler (2011) points out that there is an ongoing replacement of the generation of specialists well-trained in this classical approaches in government, academia, and the private sector. The teaching of field work and field mapping techniques is also today an important part in the education of young environmental scientists, but in addition the teaching of GIS modelling is nowadays also a relevant topic in Universities.

Recent studies illustrate the enhancements of site classification systems such as the BEC system of British Columbia (cf. subsection 1.1.5) by means of integration of spatial prediction methods (Haeussler 2011). MacMillan et al. (2007) present a hybrid of automated, semi-automated and manual procedures that develop and apply heuristic, rule-based conceptual models of ecological-landform and soil-landform relationships in a manner similar to the digital soil mapping approach (SCORPAN, cf. subsection 1.2.1). The innovation regarding site classification in MacMillan et al. (2007) is the attempt to directly translate the official filed guide to forest site identification and interpretation into a GIS-based framework. The rationale for the enhancement of the BEC system was to increase the rate of production of ecological maps by adopting an automated predictive ecosystem mapping approach, but also to reduce costs and to improve consistency and replicabililty.
An additional study to enhance the BEC system in British Columbia, but at a broader spatial scale, has been presented by Fitterer et al. (2012). They use remote sensing data as well as terrain indices to delineate homogeneous ecosystem units at the larger scale, i.e. the Biogeoclimatic Units in the BEC system or the growth district in the Bavarian classification approach. Their motivation to update and modify the boundaries of the biogeoclimatic polygons of the BEC system is to mitigate the shift between (former) mapped ecosystems and current conditions due the ongoing loss of habitats and environmental degradation resulting from anthropogenic activities and changing environmental conditions (Hamann and Wang 2006).

The challenges which motivated MacMillan et al. (2007) and Fitterer et al. (2012) to conduct their studies are also present in Germany. Many different research activities in the area of forest site mapping demonstrate that the current challenges to site mapping and the possible improvement of site mapping through the application of spatial prediction techniques have been recognized by decision-makers. For example, in Bavaria two research projects have been initiated to support forest managers with decision support systems which draw on a wide set of site information for the maintenance of forests in the 21st century.
The research project *KLIP 4 - Maps for the future* has been initiated at the Bavarian state institute of forestry LWF in 2008 to develop an area-wide, comprehensive, and homogeneous digital site information system (*Standortinformationssystem*) for the Bavarian forest administration (Beck et al. 2012). The aim is to provide maps of soil properties as well as suitability maps for different tree species for current and future climate conditions for forest managers. The approach was to combine existing soil maps and soil profile information as well as extensive field data with GIS modelling and spatial prediction methods. The *KLIP4 - Maps for the future* was successfully completed at the end of 2012. The new site information system will be implemented in 2013 in the online web-GIS plattform BayWIS (Bayerisches Waldinformationssystem).
The second example from Bavaria is the project *Waldinformationssystem Nordalpen WINALP*. Within this project a new approach has been developed to systematically assess the distribution of forest types in the Bavarian Alps (Reger and Ewald 2011). The approach is called the TRM-model (temperature-reaction-moisture model) and combines traditional site classification approaches (e.g. code with three numerals) with spatial prediction methods applied to site-ecological parameters (Reger et al. 2014). The resulting map is provided to forest managers through an online web-GIS application - `http://arcgisserver.hswt.de/winalp/`.
Also in other federal states of Germany similar projects are ongoing to support forest managers (Asche and Schulz 2004; Beck et al. 2012; Gauer 2010; Walentowski and Bergmeier 2009; Zirlewagen and Wilpert 2011) - cf. the recent symposium on modelling based site classification approaches in German ferdaral states - link to the presentations.

The methods and studies for this thesis have been developed in the course of these two research projects.

1.3.2. Research objectives and outline of the study

Based on the ongoing development of new mapping techniques in environmental sciences the research objective of this study is to develop and apply spatial prediction techniques for the assessment of forest site characteristics. The intention of this work is to introduce different new modelling approaches for site mapping in order to support ongoing and future forest management in Bavaria with a toolbox of mapping techniques. It is not the intention to replace the well-tried and established mapping approaches, but to illustrate different ways on how to include computational and geospatial technology in the assessment of forest site characteristics. The aim is to improve site mapping in order to facilitate forest management in several directions:

- from a qualitative to a more quantitative description of site characteristics
- to make site assessment more reproducible
- to make descriptions of site characteristics more dynamic, i.e. the opportunity to easily update maps when new data become available
- the capability of deriving uncertainties for predicted outcomes thus allowing the tracking of error propagation through the whole process
- to make site mapping more consistent across borders of growth districts and growth areas (Carre et al. 2007)
- to make site mapping more cost-effective

The main part of the thesis consists of three peer-reviewed articles published in ISI-listed scientific journals. In those publications new methods to enhance site mapping by means of spatial prediction methods are presented. Summaries of the published studies are presented in chapter 2 and the full papers can be found in Appendix A.

The publications address the following topics:

1. **Spatial disaggregation of complex soil mapping units**
   Traditional soil maps very often serve as data sources for site mapping. However, in many soil maps different soil types were combined in single *complex map units* which are characterized with different site characteristics. In this publication, a method is presented to disaggregate such complex map units.

2. **Spatial prediction of Ellenberg indicator values**
   Characterizing forest sites based on Ellenberg indicator values is a widely used approach in site classification. These values are available in many vegetation databases. The aim of this publication was to determine whether spatial prediction methods could be applied to regionalize Ellenberg indicator value and thereby generate maps of site characteristics.

3. **Multivariate prediction of Ellenberg indicator spectra**
   Usually, single plant species are aggregated to an averaged indicator value of a vegetation plot. However, compared to the range and shape of a indicator spectrum of a plot this could lead to a loss of information when applying statistical modelling. In the last part of this study a Proportional Odds model has been applied to Ellenberg soil reaction values in order to predict the spatial distribution instead of an average value.

In addition to this main part of my thesis, different application examples and method developments for spatial modelling of site characteristics are presented in chapter 3. The results of these studies have been used in the KLIP4 and WINALP projects either as input for modelling of site characteristics (Häring et al. 2012; Osenstetter et al. 2013; Reger et al. 2014) or to facilitate traditional site classification in the field (Osenstetter et al. 2013). Thus it‘s all about the **application of spatial prediction methods to create input data for site ecological mapping**.

Digital soil mapping based on the SCORPAN approach has been used to extrapolate soil map units into areas where no soil maps were available during the KLIP4 and WINALP projects. In addition to statistical validation of the prediction models, digital soil maps were validated with extensive field measurements. These data have been used to analyze the options and limits of digital soil mapping approaches in Bavaria.

Besides digital soil mapping, digital terrain analysis and image segmentation algorithms have been used to delineate homogeneous terrain objects. Following a hierarchical structure of disaggregating the landscape into more homogeneous areas, this concept follows the same idea as traditional site classification approaches (cf. the zonal climate classification in the BEC system - subsection 1.1.5 or the regional stratification in growth area and growth district in Bavaria - subsection 1.1.6).

Chapter 2.

Summary of published studies

> ... but on the middle step there is better soil, and it does not lie bound down under such severe cold, either. This one can see at a glance, since the trees are both higher and of finer quality.
> There you'll find maple and oak and linden and weeping-birch and hazel trees growing, but no cone-trees to speak of...
>
> *(The Wonderful Adventures of Nils*
> *by Selma Lagerloef)*

2.1. Publication 1: Spatial disaggregation of complex soil map units

Published as:

Häring, T., Dietz, E., Osenstetter, S., Koschitzki, T. & Schröder, B. (2012) Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils. *Geoderma*, 185-186: 37-47.

Own contribution to the publication:

- development of the method
- implementation of the method in a modelling and GIS environment
- primary responsibility for writing the article and accompany the publication process

Summary

Forest site classification relies beside field work frequently on available soil maps, because these are often the only information about soils in an area of interest. However, depending on the map scale or the detail of the map legend soil mapping units rarely comprise single soil types, but usually consist of a combination of a dominant soil with minor associated soils. When the various soils of a soil mapping unit occur in a recognizable geographical pattern in defined proportions they constitute *complex soil map units*. However, aggregating soil types with different site characteristics into one map unit could be too imprecise for site-specific forest management or land evaluation.
The publication introduces an approach to disaggregate complex soil map units into more homogeneous map sub-units containing only one soil type. The overall goal of the study was the refinement on the official 1:25.000 soil map of Bavaria to make it more useful for site-classification within the Maps-for-the-future project. The method follows the traditional top-down approach in soil mapping and site classification to divide an existing map unit in more homogeneous sub-units. By using spatial prediction models we could draw new boundaries inside the map

polygons to represent a single soil type and no longer a mixture of several soil types.

The basic idea for the method is the functional relationship between soil types and topographic position as formulated in the concept of the catena. We use a comprehensive soil profile database (9924 soil profiles) and topographic attributes derived from a 10m digital elevation model as input data for modelling. We apply random forest models for the classification of soil types based on the hypothesis that soil types could be discriminated based on their 'topographic fingerprint'. We grouped all complex map units with the same combination of soil types together. Each of these groups of map units were modeled separately to apply comparably simple but class-specific models for the delineation of sub-areas.

To account for uncertainty in the models and in the data we identify map subunits only above a defined threshold of 70% predicted probabilities. In areas where the probability is below this threshold we assign the area as "indifferent" because the model only makes unspecific classification in these cases. The results show a significant spatial refinement of the original soil polygons.

In addition to the statistical validation of each model we initiated a field campaign to collect independent validation data. Validation of our predictions was estimated based on 1812 independent soil profiles and gave an overall accuracy of 70%. Map units, in which shallow soils were grouped together with deep soils could be separated best. Also Histosols could be predicted successful. The highest error rates were found in map units in which Gleysoils were grouped together with deep soils or Anthrosols. To check for validity of the results the black box random forest model has been illuminated by calculating the variable importance for each predictor variable and plotting response surfaces. The results corroborate our hypothesis defined above.

The refined soil map is a clear improvement regarding the ability to classify forest sites. It has been used within the Maps-for-the-future project to assess site-characteristics by assigning representative soil profiles.

2.2. Publication 2: Predicting Ellenbergs soil moisture indicator value in the Bavarian Alps using additive georegression

Published as:

Häring, T., Reger, B., Ewald, J., Hothorn, T. & Schröder, B. (2013) Predicting Ellenberg's soil moisture indicator value in the Bavarian Alps using additive georegression. *Applied Vegetation Science*, 16: 110 - 121.

Own contribution to the publication:

- co-development of the concept
- preparation of the data
- execution of mboost modelling and application to GIS data
- primary responsibility for writing the article and accompany the publication process

Summary

Soil hydrology is one of the major ecological driving forces for site characteristics. Therefore, detailed knowledge of the spatial variation thereof is essential for sustainable and site-specific ecosystem management. In contrast to current approaches in spatially-distributed hydrological modelling to quantify spatial variations in soil water availability, in this publication we present a new approach for spatial modelling, which focuses more on ecologically effective hydrological conditions than on the simulation of the temporal change in soil water content.
The approach in this study is based on vegetation data, namely Ellenberg indicator values for moisture (EIV). EIV have been used as response variable in a sophisticated statistical modelling framework. We used mean soil moisture values as found on 1505 forest plots from the database WINALPecobase and extrapolate these values based on topographic, climatic and soil variables to the area of the Bavarian Alps. We adopted methods developed in species distribution

modelling to regionalize EIV. We use the additive georegression framework for spatial prediction of EIV with the R-library `mboost`, which is a feasible way to consider environmental effects, spatial autocorrelation, predictor interactions and non-stationarity simultaneously in our data. The framework is much more flexible than established statistical and machine-learning models in species distribution modelling. We estimated five different mboost models reflecting different model structures on 50 bootstrap samples in each case.
Median R^2 values calculated on independent test samples ranged from 0.28 to 0.45. Our results show a significant influence of interactions and non- stationarity in addition to environmental covariates. Unweighted mean indicator values can be modelled better than abundance-weighted values, and the consideration of bryophytes did not improve model performance. Partial response curves indicate meaningful dependencies between moisture indicator values and environmental covariates. However, mean indicator values <4.5 and >6.0 could not be modelled correctly, since they were poorly represented in our calibration sample. The final map represents high-resolution information of site hydrological conditions.
We can conclude that EIV offer an effect-oriented alternative to physically-based hydrological models to predict water-related site conditions, even at landscape scale. The presented approach is applicable to all kinds of EIV. Therefore, it is a significant step toward a new generation of models of forest site types and potential natural vegetation. The predicted map of soil moisture EIV has been used in the WINALP project to generate a map of potential natural vegetation (Reger et al. 2014).

2.3. Publication 3: Regionalising indicator values for soil reaction in the Bavarian Alps - from averages to multivariate spectra

Published as:

Häring, T., Reger, B., Ewald, J., Hothorn, T. & Schröder, B. (2014) Regionalising indicator values for soil reaction in the Bavarian Alps - from averages to multivariate spectra. *Folia Geobotanica*, doi:10.1007/s12224-013-9157-1.

Own contribution to the publication:

- co-development of the concept
- preparation of the data
- execution of mboost modelling and application to GIS data
- primary responsibility for writing the article and accompany the publication process

Summary

Besides temperature and moisture, soil reaction is the third pillar in forest site classification (cf. section 1.1). In this study we present an approach to produce maps of Ellenberg values for soil reaction (R-value) in the Bavarian Alps. Eleven meaningful environmental predictors covering GIS-derived information on climatic, topographic and soil conditions were used to predict R-values. As dependent variables, EIV for soil reaction were derived from plot records in the vegetation database WINALPecobase. We used an additive georegression model, which combines complex prediction models and the increased prediction accuracy of a boosting algorithm. In addition to environmental predictors we included spatial effects into the model in order to account for spatial autocorrelation.
In contrast to my second paper (modelling of Ellenberg soil moisture values, cf.

section 2.2), in this study I was particularly interested in the usefulness of *averaged* R-values for spatial prediction. EIV spectra are often skewed, uniform or even bimodal, which leads to misleading estimates of site-ecological conditions when applying Gaussian regression models. We apply two different models to analyse the influence of averaging indicator spectra: (1) a geo-additive regression model which estimates mean R-values (the "classical" approach) and (2) as an alternative a proportional-odds-model predicting species indicator values, which, instead of the average indicator value, estimates the probability distribution over the range of R-values 1 to 9 for a given set of environmental gradients. We hypothesize that we should get a more realistic estimate of site-ecological conditions by using a multivariate regression model. By using the mboost regression framework we are able to estimate a mean indicator value (Gaussian regression model) as well as an ordinal value distribution (proportional-odds-model) with the same set of vegetation data and environmental predictors.
We found meaningful dependencies between the R-value and our predictors. Both models produced the same spatial pattern of predictions. Spatial effects had an impact only in the first model. The main drawback of mean R-values is the oversimplification of complex conditions of soil reaction which is entailed in the averaging and regression to mean values. Therefore, regionalised average indicator values provide only limited information on site-ecological characteristics. Model 1 failed to predict the range and shapes of original indicator spectra precisely. In contrast, the second model provided a more sophisticated picture of soil reaction. To make the multivariate output of model 2 comparable to the output of model 1, we propose to plot the distribution in a three dimensional color-space. In addition, comparison of both models based on a multivariate linear regression model result in a R^2 of 0.93. The proportional-odds-model is a promising approach also for other indicator values and different regions as well as for other ordinal-scaled ecological parameters.
The predicted map of soil reaction EIV has been used in the WINALP project to generate a map of potential natural vegetation (Reger et al. 2014).

Chapter 3.

Additional studies: Creating input data for site classification

> …but the very lowest step is the best of all. It is covered with good rich soil; and, where it lies and bathes in the sea, it hasn't the slightest feeling of the Småland chill.
> Beeches and chestnut and walnut trees thrive down here; and they grow so big that they tower above the church-roofs…
>
> *(The Wonderful Adventures of Nils by Selma Lagerloef)*

The following sections present additional application examples and method developments for spatial modelling of site characteristics which have been developed and applied in the KLIP4 and WINALP projects. The results of these studies have been used either as input for subsequent modelling of site characteristics (Häring et al. 2012; Osenstetter et al. 2013; Reger et al. 2014) or to facilitate traditional site classification in the field (Osenstetter et al. 2013).

3.1. Spatial prediction of soil mapping units

Site-classification and mapping at regional and landscape scale always depends - in addition to climatic data - on soil data. However, the availability of small-scale soil maps in Bavaria is as fragmentary as in many other European countries (Jones et al. 2005). There are huge areas where no soil maps were available.
However, when mapping site characteristics, especially when applying modelling based mapping approaches, a homogeneous and coherent soil database is necessary. Therefore, digital soil mapping as described in subsection 1.2.1 has been applied to predict soil map units of the official soil map *ÜBK25* (Übersichtsbodenkarte) for areas which were not mapped so far. The areas of Bavaria which have to be predicted are widely scattered and comprise heterogeneous physiographic conditions. In addition, high-resolution geological maps, which serve as one of the main environmental predictors, were also not available for all prediction areas. Therefore, the unmapped area is subdivided into 16 prediction areas (see Figure 3.1).

The ÜBK25 is mapped with a scale of 1:25.000 and covers[1] 64 071 km^2 which is about 80% of the Bavarian State territory. The map contains more than 700 different map units and follows the official German guideline for soil mapping (Ad-hoc-AG Boden 2005). Soil mapping in Germany follows the concept of substrate-systematic mapping, which means that every soil map unit contains information about the soil type and its parent material (geologic substrate).

As environmental predictors for digital soil mapping the following parameters have been selected:

- where available, the geological map with a scale of 1:25,000. In the remaining area the 1:200,000 geological map.

[1]It should be mentioned that soil mapping in Bavaria is an ongoing task. Since the beginning of KLIP4 the coverage of the ÜBK has increased considerably. However, the status of the ÜBK25 at the beginning of the project was the starting point for further project planning.

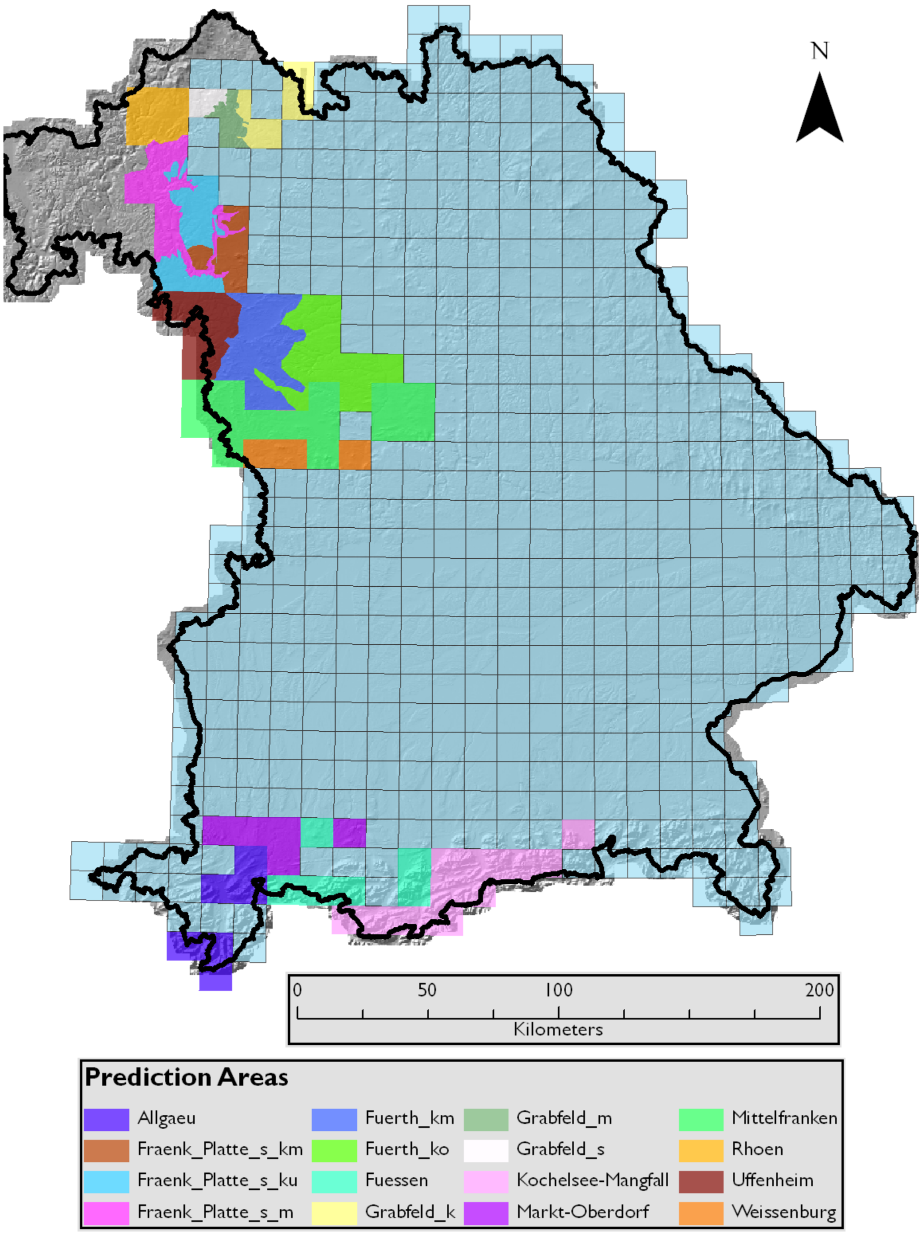


Figure 3.1.: The coverage of the 1:25.000 soil map of Bavaria (Übersichtsbodenkarte, light blue map sheets, at the beginning of KLIP4) and 16 different prediction areas to fill the gaps in the soil map. The delineation of the prediction areas is based on the environmental conditions of the area, data availability and data quality (Source: Bavarian Environmental Agency, Bavarian Topographical Survey).

- a 10m digital elevation model was re-sampled to 20m to align the cell size to the map resolution (Häring et al. 2009). From this DEM a set of terrain attributes were delineated.

- climatic variables such as mean annual temperature and mean annual precipitation (Hera et al. 2012).

Prediction of soil map units was performed with random forest (Breiman 2001a). Random forest is an ensemble method in which many different classification trees are combined to produce a more stable and accurate classification compared to a single decision tree (Bauer and Kohavi 1999; Breiman 1996; Dietterich 2000). Each tree is built on a bootstrap sample of the given data. To form the ensemble, the different trees are combined using bagging (bootstrap aggregating). The resulting *forest* is a *random* forest because at each split only a random subset of the candidate predictors is considered for the binary partition (Elith and Graham 2009). This de-correlates the trees, improves the variance reduction and finally leads to more accurate predictions (Bühlmann and Yu 2002; Strobl et al. 2009). The predictions of each single tree are combined using a majority vote to get a final ensemble prediction. In recent years, random forests have been widely used in digital soil mapping (e.g. Häring et al. 2012; Liess et al. 2012; Roecker et al. 2010; Stum et al. 2010; Wiesmeier et al. 2011).
Validation of the random forest models was performed using the out-of-bag error. The predictive performance of the model is calculated on those observations which were not included in the learning sample for a specific decision tree, i.e. those observations which were not part of the bootstrap sample of the original data set. Using those out-of-bag observations, we have independent test samples for computing the prediction accuracy. It could be shown that the out-of-bag error is a conservative estimate (Strobl et al. 2009).

Model calibration and validation was conducted with already available map sheets of the ÜBK25. Spatial prediction of soil maps is only feasible for areas with similar physiographic conditions as in the area used for the model calibration (*training area*). Thus substantial areas of Bavaria could not be predicted with digital soil mapping techniques due to the lack of suitable training areas (cf. Figure 3.1).

In order to get reliable predictions, an independent area for model calibration as well as for model validation is needed. Therefore, the available ÜBK map sheets have to be divided into two subareas: training area and validation area. In both

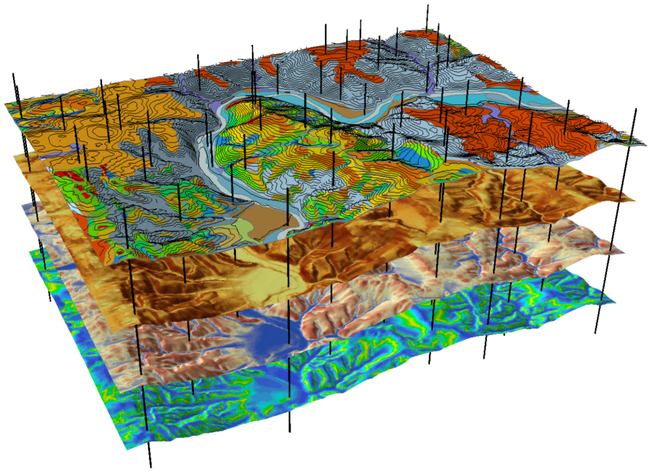


Figure 3.2.: Sampling of environmental variables (Source: Bavarian Topographical Survey).

areas an independent sample of the soil mapping units and the environmental predictors has to be drawn by collecting all values at a certain location from the stack of maps (Figure 3.2)

An example of a predicted soil map is shown in Figure 3.3. The mapped area is located between Füssen, Kaufbeuren and the Walchensee in southern Bavaria.

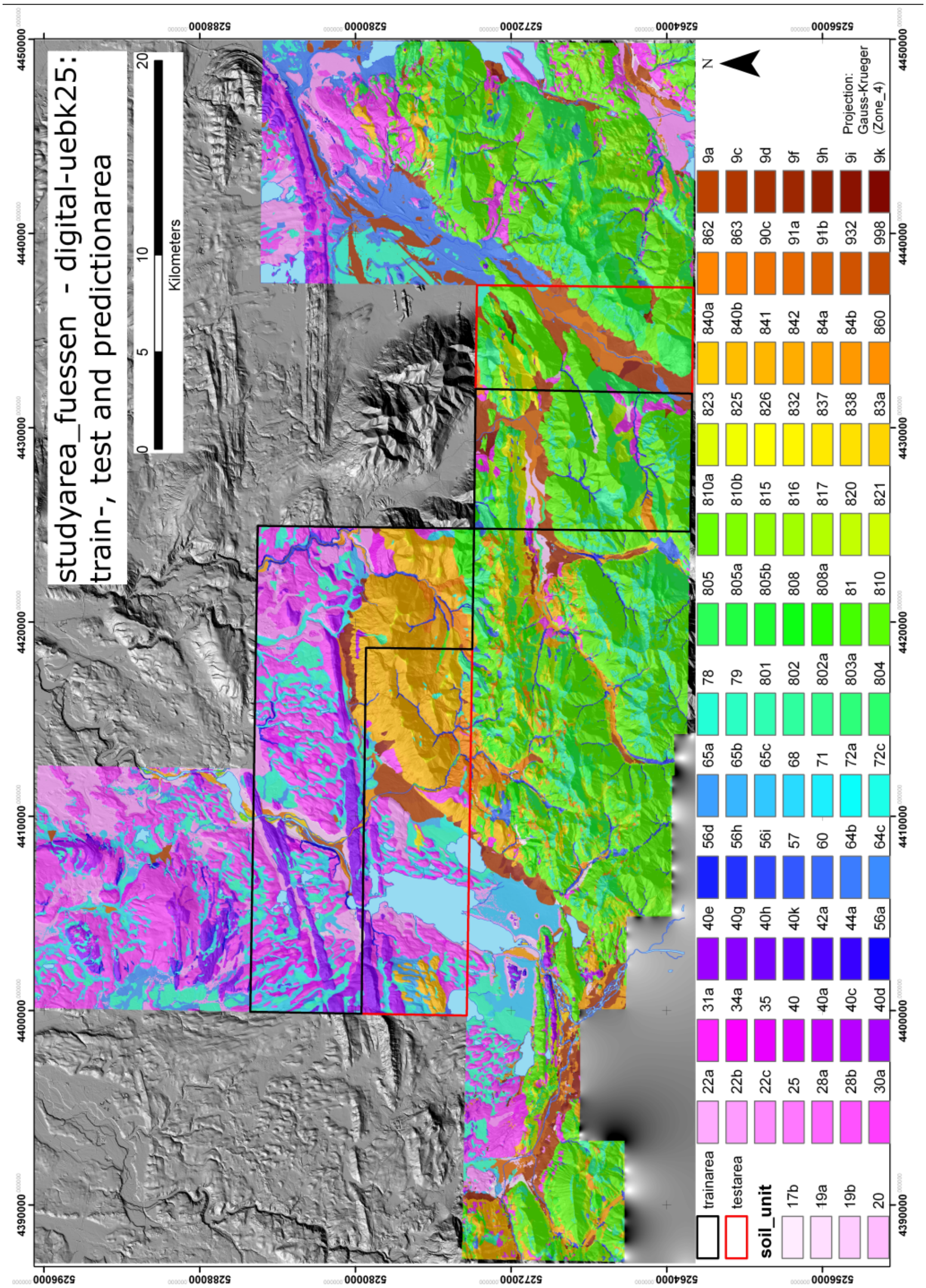


Figure 3.3.: Digital soil map around the foothills of the Bavarian Alps. The map contains soil mapping units of the ÜBK25. Model calibration and validation was conducted with available map sheets of the ÜBK25: black encircled areas indicate the training area, red encircled area indicates the validation area. The remaining area indicates the predicted soil map.

3.2. Validation of digital soil maps and field measurements

Similar to conventional soil maps, digital soil maps contain errors. McBratney et al. (2003) refer two reasons for that: (i) local variation of soils at whatever spatial resolution, which is especially the case for nominal map units (cf. Häring et al. 2012) and (ii) the uncertainties of the environmental layers of predictor variables can propagate errors. Therefore, the estimation of quality of digital soil maps is an integral part of each modelling area.

The quality estimation based on statistical validation measures such as the classification accuracy of a classification model or the R^2 of a regression model are self-evident. Their calculation is very straightforward and mostly they were already computed by default by the statistical software. Analysis within the KLIP4 project however have shown that an independent validation of predicted maps with field validation data is necessary, primarily for quality estimation of the map produced (Brus et al. 2011), but also to gain "user acceptance" where the user has little experience of modelling based mapping approaches. Quality estimation based on statistical validation alone could be misleading. This could be shown with a review of statistical model error with field validation data.

3.2.1. Confronting statistical validation with field validation data

Central Franconia was subdivided into four different modelling areas due to heterogeneous environmental conditions as well as data availability (cf. Figure 3.1). After the creation of digital soil maps, soil experts went out in the forests to collect in total 4500 soil samples. The spatial distribution of all validation points is illustrated in Figure 3.4.
The four modelling areas were not only very different with regard to environmental conditions, but also with regard to the complexity of the soil prediction model. The number of different soil map units as well as the number of geological map units is illustrated in Figure 3.2.1.

[1]This section is based on
Häring, T. & Schröder, B. (2010) A review of model-error in digital soil mapping: Confronting statistical soil landscape models with large-scale field validation data. *Geophysical Research Abstracts*, 12, EGU2010-12757

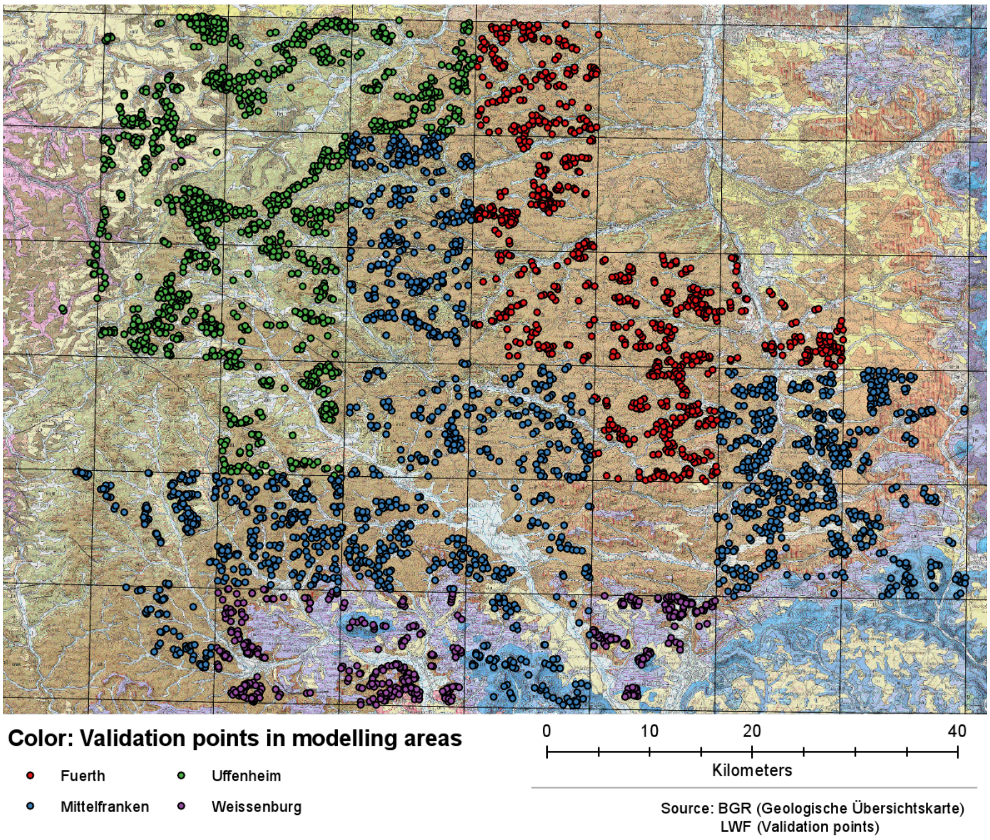


Figure 3.4.: Validation points in central Franconia. The colors indicate four different modelling areas.

	n soil units	n geological units	geology binarized?
Mittelfranken	92	37	X
Weissenburg	50	25	X
Uffenheim	63	19	
Fuerth	56	15	

- **Mittelfranken** is the most complex model in this analysis. The model comprises a large number of soil units as well as geologic map units. Due to the large number of geological units, the parameter was binarized[2]. The map scale of the geological map is 1:200.000.
- **Weissenburg** is the only area where high-resolution geologic maps were available. The model has an intermediate complexity.

[2]For technical reasons, the maximum number of values within a nominal predictor is 32

- **Uffenheim** also uses mid-resolution geologic map. The parameter was not binarized.

- **Fuerth** is characterized by homogeneous geological conditions.

During the validation process different parameters had to be estimated by the soil expert in the field. The substrate based soil classification allows the validation of the predicted soil map units in different dimensions. The overall validation parameter is the exact text in the map legend for a map unit. However other further parameters were also requested: (1) Is the soil *type* predicted correctly? (2) Is the geologic substrate correct? (3) Is the geologic stratigraphic unit correct? and (4) Is the number of horizontal strata correct?
Using these different parameters a comprehensive review of the statistical model error as well as the map extrapolation could be carried out.

How reliable is the statistical model error?

First, I analyzed the reliabiliy of statistical model error, i.e. the proportion of correctly predicted map units by the model. In this case this is the out-of-bag error of the random forest model. This value was plotted against the misclassification error calculated with the field validation data. The misclassification error of the field data was calculated as the proportion of correctly predicted map units within the field data. Thus the reliability of the statistical error could be checked. The out-of-bag error is calculated for each soil map unit separately, which means we know which map units could be discriminated well by the model and which not. The question is, can map units with a low model error be predicted more successfully than those with a high model error?
The misclassification rate of the field validation data was also calculated for each map unit separately. The number of correctly predicted samples for a specific map unit was divided by the total number of validation samples for the same map unit. The scatter plot of both error values is illustrated in Figure 3.5.

For each modelling area a separate plot was created. The misclassification rate calculated on field validation data is displayed on the x-axis (*field error*), the out-of-bag error of the random forest model on the y-axis. The size of the bubbles indicates the number of validation points for a map unit, i.e. the larger the bubble the more validation points for a map unit were available. The color of the bubbles indicates the number of training samples available for model fitting. In all four plots the bubbles show a widely scattered pattern and by no means

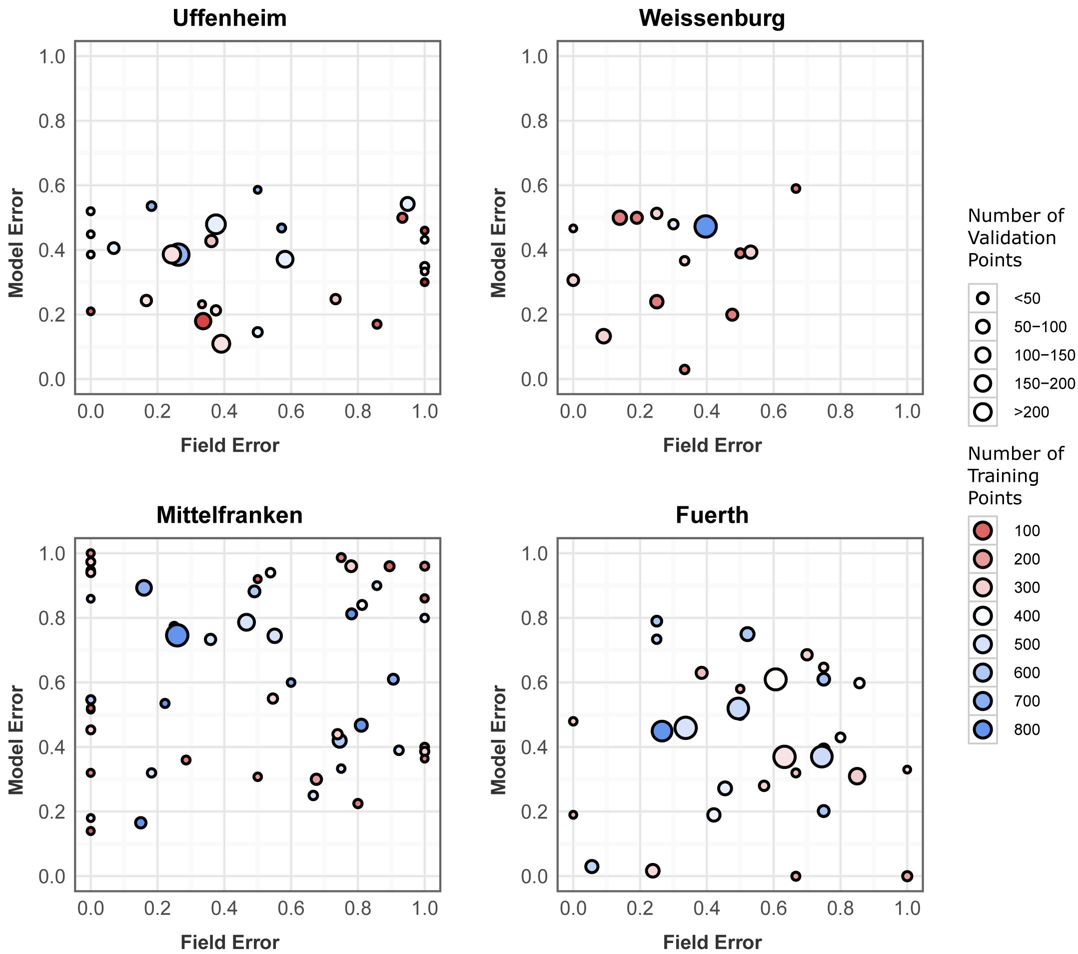


Figure 3.5.: Out-of-bag model error of single soil map units plotted against the misclassification rate in the field. Color of the bubbles indicates the number of training samples available for model fitting, the size of the bubble indicates the number of validation points for a map unit.

a correlation between model error and field error. There are map units with a very low model error but still a high misclassification rate in the field. On the other hand, there are also map units which could be discriminated only poorly in the model but show a low field error. In addition, the number of samples used for model fitting as well as the number of validation points show no correlation with the error values. It can be concluded that the statistical model error gives only a very poor indication of the actual prediction accuracy and is therefore not recommendable as a single quality measure for digital soil mapping.

Which predictions are fine - which are not?

To analyze which parameter of the legend text of a map unit could be predicted successfully, bar charts indicating the relative frequency of true and false predictions of single map units in the field data have been created (Figure 3.6 to Figure 3.8). The plots have been grouped according to the four modelling areas to see whether the model complexity and data quality have an impact on model performance.

Figure 3.6 indicates the true and false prediction of the legend text (Is soil type as well as geologic material as well as horizontal layering correct?). The plot shows a scattered pattern over all map units and also over all modelling areas. The best results were achieved for Weissenburg, but overall the performance of the models is not satisfactory, as seen also in Figure 3.5.

The overall picture in Figure 3.7 becomes more greener, which means the prediction accuracy becomes better when considering only the soil type, regardless the geologic substrate on which the soil type has developed.

The best results could be achieved in Figure 3.8, where only the geological stratigraphy is considered. Clearly the mapping of map units to the same geologic stratigraphy could be done successfully.

Some conclusions

What are the reasons for these findings? Figure 3.9 shows the geological stratigraphic system for the Triassic and Jurassic eras, which is the main geological characteristic in the modelling areas. The soil map legend has a hierarchical structure which is oriented in the geologic stratigraphic system. Furthermore, map units of geological maps in Germany adhere strictly to the stratigraphic system. However, soil map units representing the combination of soil type and geological substrate, the so-called *Bodenform*, e.g. cambisol on sandy-loamy ma-

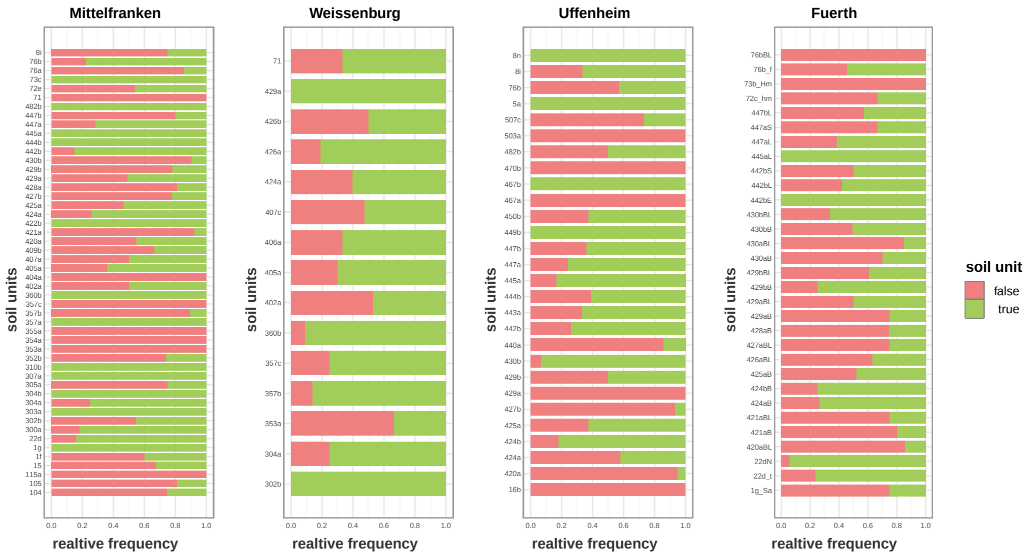


Figure 3.6.: Relative frequency of true and false predictions of the legend text of soil map units in all four modelling areas.

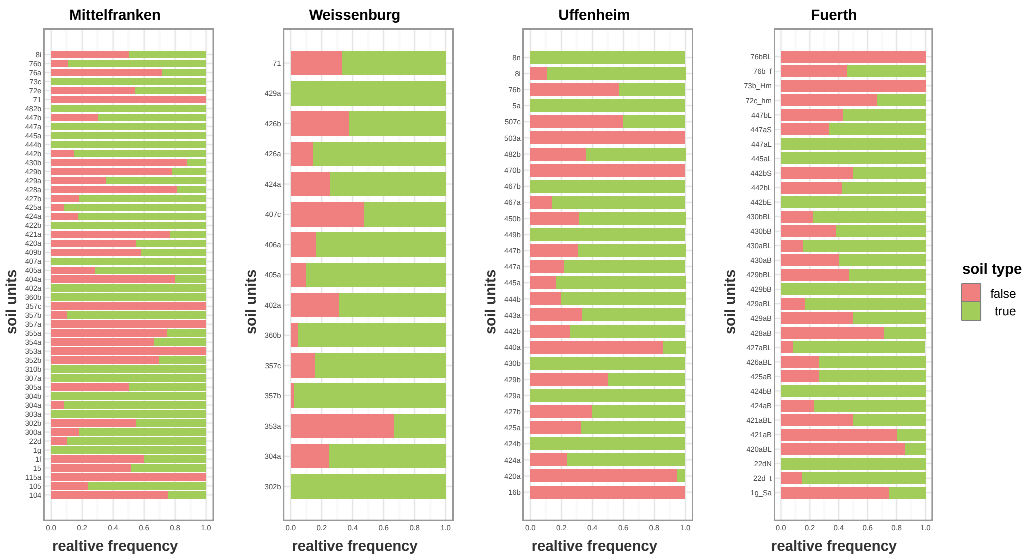


Figure 3.7.: Relative frequency of true and false predictions of soil types for several map units in all four modelling areas.

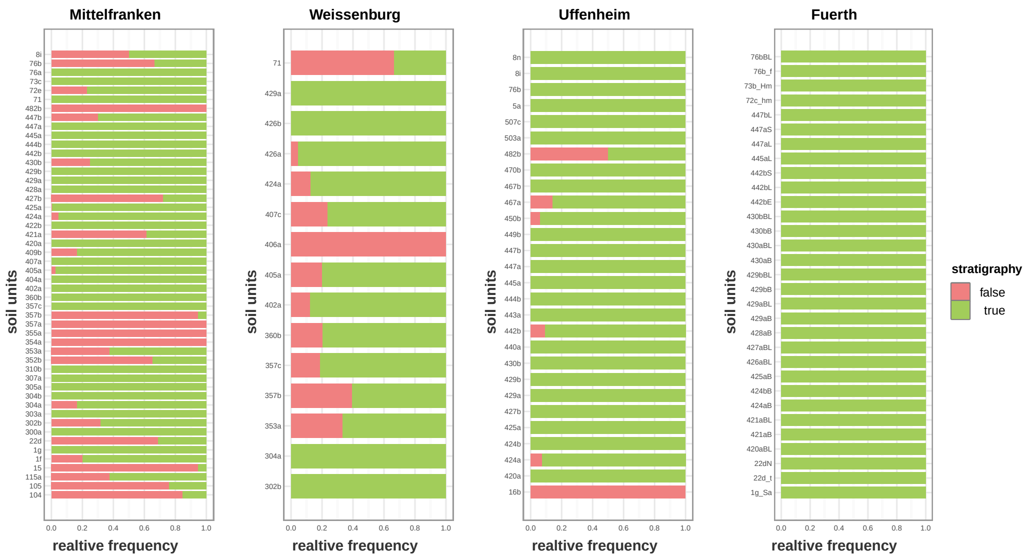


Figure 3.8.: Relative frequency of true and false predictions of geologic stratigraphy for several map units in all four modelling areas.

terial of the Burgsandstein or on clayey to silty material of the Blasensandstein, are more detailed than purely geological maps. This is also indicated by the block diagram in the center of Figure 3.9, which shows a high petrographical diversity within one stratigraphic unit (different sandstones or clays within one group of soil map units).

We could see that soil map units could be predicted successfully to the correct stratigraphy (Figure 3.8), for example soil map units 420a to 430b within the area of Blasen-, Coburg- and the Burgsandstein. This could be achieved by using geologic maps as predictor variables. Soil types could also be predicted with success (Figure 3.7), which could be explained mainly by the strong correlation of some soil types with topographic parameters. The influence of topography on the spatial distribution of soil types at field to landscape scale was first formulated in the catena concept of Milne (1935). This relationship has been used in numerous digital soil mapping studies (Behrens et al. 2010; Deumlich et al. 2010; McBratney et al. 2003; Möller et al. 2008) and serve as the hypothesis of the method presented in Häring et al. (2012).

By analyzing the distribution of different soil types over a set of topographic gradients a specific "topographic fingerprint" of each soil type could be derived. This is shown in Figure 3.10 in which 7887 soil samples were analyzed regarding their topographic location and soil type. The analysis was restricted to eight soil types which could be found frequently in Bavaria (Braunerde, Pelosol, Vega, Gley-

Braunerde, Podsol-Braunerde, Pseudogley, Pseudogley-Braunerde). The plots show significant differences in the appearance of these soil types on topographic gradients.

However, the analysis of the 7887 soil types confirms the findings of Figure 3.5 and 3.6. Neither official geological maps nor topographical information contain sufficient information to discriminate between *Bodenformen*, i.e. to discriminate between similar soil types within a stratigraphic unit having different petrographic conditions (sandy, silty, loamy, clayey). There is no difference between the three different texture classes for one soil type in Figure 3.10.

To conclude: the combination of topographic attributes with geological maps illustrate sufficient predictor combination to predict soil types as well as stratigraphic units. However, detailed petrographic information is needed on slope sediment layering or alluvial deposits, but these data were rarely available.

Based on these findings the predicted digital soil maps have been tested against the field validation data. In order to use reliable digital soil maps within the KLIP4 and WINALP projects we manually adjusted the predicted map units regarding their geological substrates. Predicted map units which were correct regarding soil type and stratigraphy but with the wrong stratigraphy were adjusted according to the informatiom from the field data.

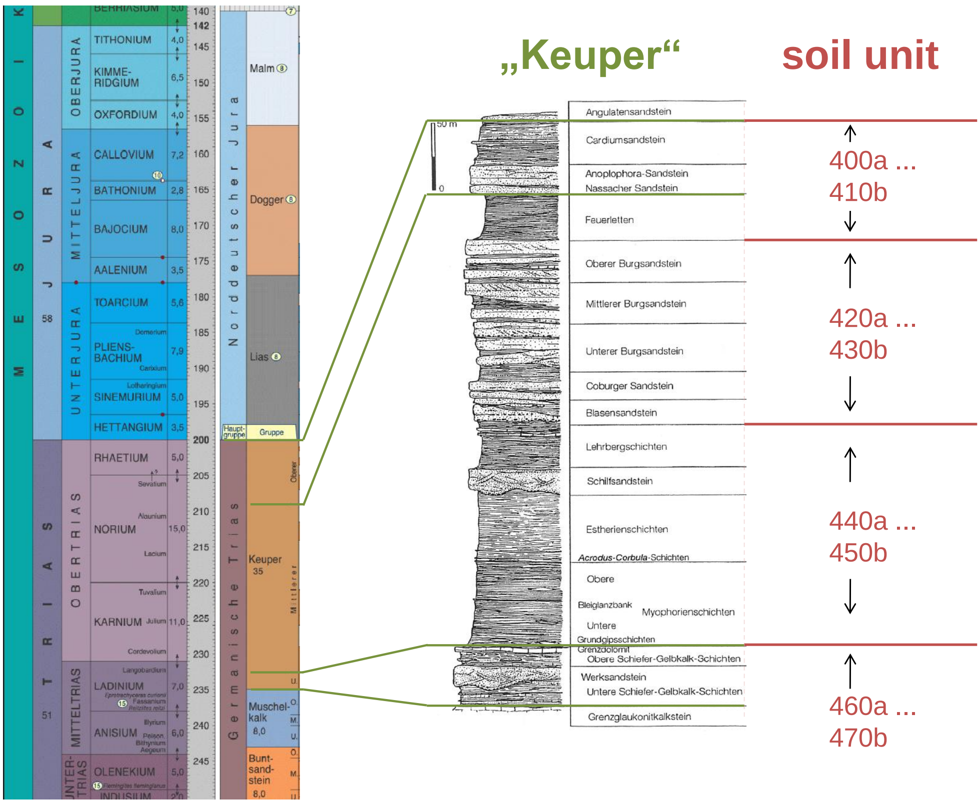


Figure 3.9.: The organization of map legend follows the geological stratigraphy (i.e. geological time steps). Petrographic information is not included in the geological map.

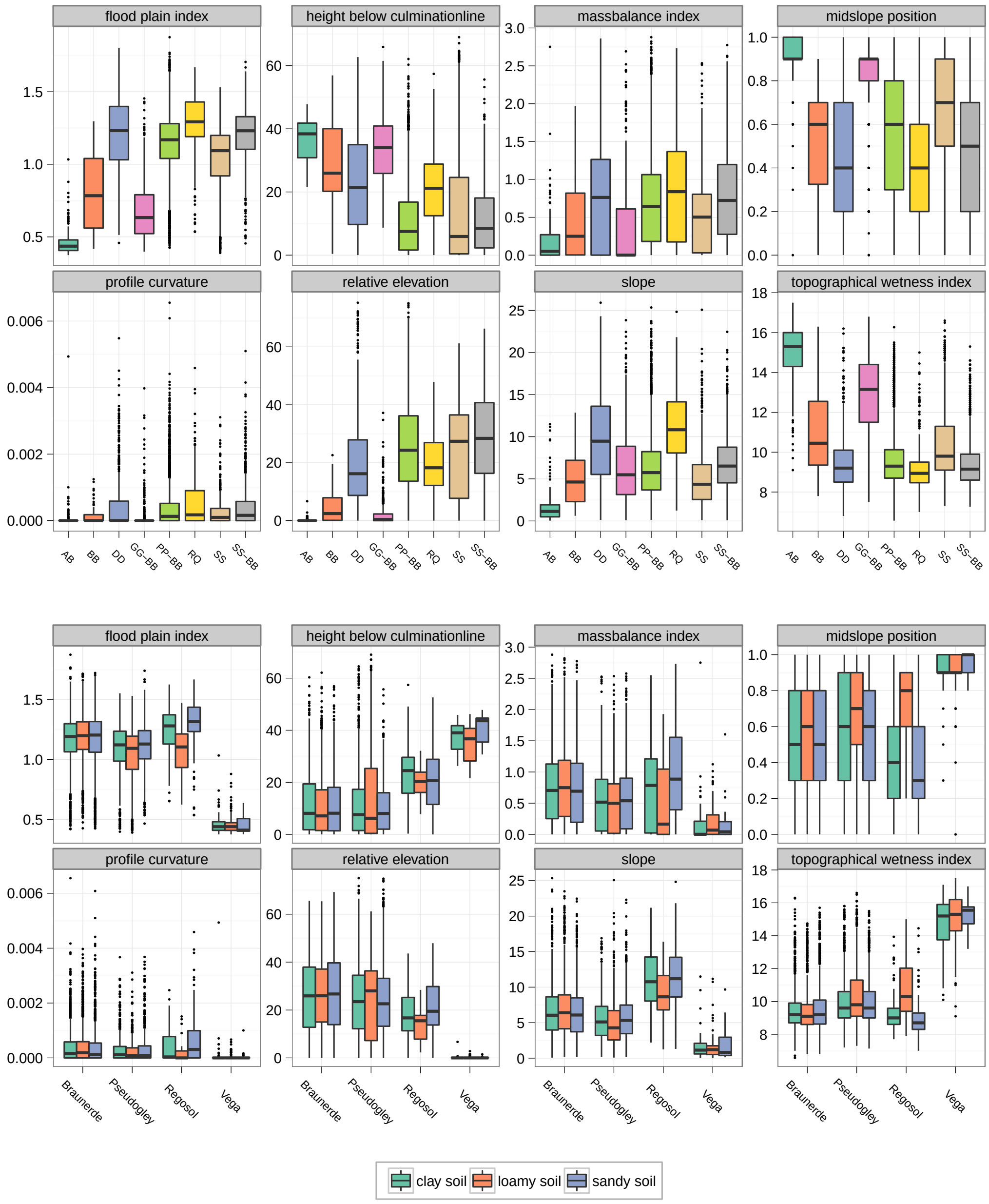


Figure 3.10.: Topographical fingerprint of different soil types (upper part) delineated from 7887 soil profiles (AB = Vega, BB = Braunerde, DD = Pelosol, GG = Gley, PP = Podsol, RQ = Regosol, SS = Pseudogley). No variance can be detected when considering soil texture class (lower part).

3.3. Delineation of homogeneous terrain objects

3.3.1. Landscape segmentation

Considering site characteristics at landscape scale, terrain characteristics and landforms play an important role, mainly because of their influence on the flow of water, soil erosion and accumulation and spatial distribution of nutrients. In areas with strong relief, elevation, slope and aspect could become the most important parameters for site classification (Hengl and MacMillan 2009).
Topography is functionally related to the characteristics, processes, and spatial arrangement of soils (Huggett and Cheesman 2002; Moore et al. 1991). This finding dates back to the beginnings of soil science as Hilgard (1921) and Jenny (1941) recognized different factors (including topography) influencing soil formation. However, topography as being a continuous gradient is an abstract term to handle. Mapping site-characteristics considering a site-classification system therefore benefits from using discrete landscape objects with boundaries delineated from multivariate topographic gradients. Moreover, and especially in hierarchically organized and stepwise site-classification systems such as in Bavaria (site-units within growth districts, growth districts within growth areas, growth areas within Bavaria, cf. Figure 1.2), the delineation of homogeneous and discrete terrain objects perpetuates the philosophy of disaggregating nature into more and more uniform entities with regard to site characteristics.

The main concept behind the method in this section is to divide an area of interest into terrain segments or units that can be used subsequently as strata for mapping or modelling purposes, e.g. sampling. These segments should be as homogeneous as possible in terms of the environmental attributes used for segmentation in order to create tangible terrain objects with characteristic, recognizable shapes, e.g. uniform slopes, fluvial terraces, or floodplains (MacMillan and Shary 2009). The delineation of landform elements is a central task in terrain analysis and geomorphometry (Hengl and Reuter 2009). Several authors use the term segmentation for geomorphometric analysis however with different meanings (Drâgut and Blaschke 2006; Martin and Timmer 2006; Minár and Evans 2008; Möller et al. 2008; Pennock 2003; Pennock and Corre 2001; Stepinski and Bagaria 2009). Here we use segmentation in the sense of the application of methods from

[2]This section is based on
Häring, T. & Schröder, B. (2010) Sampling Optimization using Image Segmentation. *Proceedings of the 4^{th} International Workshop on Digital Soil Mapping*, 24.-26.05.2010, Rome.

the field of digital image analysis (Drâgut and Blaschke 2006).

3.3.2. Object-based image analysis

Image segmentation was developed to identify objects of interest in digital images, such as remote sensing data or medical images. The advantage of using image analysis for landscape analysis compared to per-pixel analyses is the consideration of spatial connection of individual pixels (Evans et al. 2009). It is possible to take into account the sizes, shapes and relevant positions of real-world image objects like terrain elements or landscape patterns across a stack of several environmental gradients (Blaschke and Strobl 2001). These gradients in geomorphometry were first provided by digital elevation models as the fundamental source of information and parameters like primary and secondary terrain attributes extracted from them. Besides these, remote sensing images and discrete environmental information like geological maps can be used for geomorphometric analysis. Although image segmentation is not new, the idea to create objects based on topological and shape information for landscape modeling is not common, although there are promising results in several recently published studies (Burnett and Blaschke 2003; Drâgut and Blaschke 2006; Drâgut et al. 2009; Möller et al. 2012, 2008)
There are different segmentation methods like histogram-based methods, edge detection, watershed segmentation or graph partitioning methods (Shapiro and Stockman 2001). Here, a multi-resolution image segmentation algorithm as introduced by Baatz and Schäpe (2000) was used, which is implemented in the commercial software environment eCognition. The algorithm is a bottom-up region-merging technique starting with each pixel forming one image object and merging adjacent objects in a pairwise fashion to larger objects in subsequent steps if a specific criterion, called degree of fitting, is matched. The degree of fitting is a measure of homogeneity of a single image object. Beyond the multivariate feature space the size, smoothness, compactness, and shape of image objects are considered. Thus one can modify several parameters to get suitable segmentation results for a certain image data stack and a considered application. A detailed description of the algorithm can be found in Baatz and Schäpe (2000) and Benz et al. (2004).
The image analysis within eCognition follows a hierarchical network. It successively creates image objects on different spatial scales. The objects can be linked together resulting in a topological structure of image objects within different hierarchical levels.

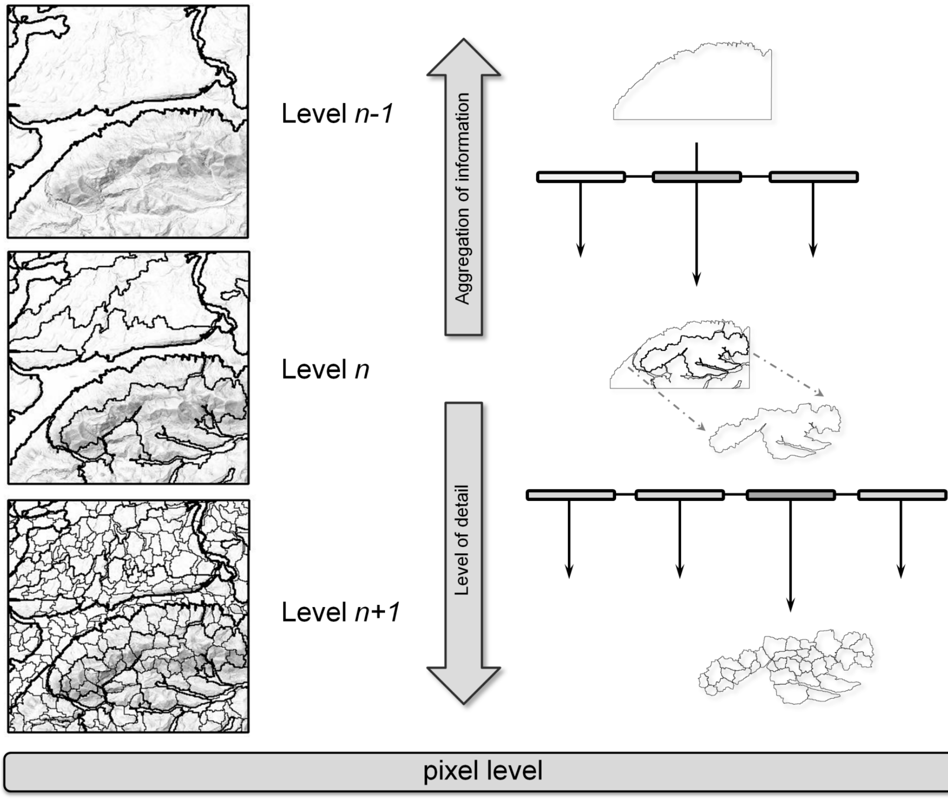


Figure 3.11.: Landscape segmentation using object oriented image analysis follows the idea of a hierarchically structured landscape. Beginning with individual pixels the information is aggregated into image objects. Depending on different adjustments considering the degree of homogeneity, size, smoothness, compactness, and shape, the approach generates more or less homogeneous image objects at different spatial scales (coinciding with hierarchical levels). The arrangement in different levels constitutes a topological network: each image object in level n belongs to a super-object in level (n-1), is connected to adjacent image objects on the same level n, and each image object in level n contains sub-objects in level (n+1) (cf. Burnett and Blaschke 2003).

This follows concepts of hierarchically structured landscapes in which large spatial units (super-objects) arise from the significant alteration of landscape-related attributes and the arrangement of small-scale landscape objects within hierarchical super-objects (Steinhardt and Volk 2003; Urban et al. 1987; Wu and David 2002). In the example in Figure 3.11, first an area of interest is partitioned in large-scale topographic regions like floodplains, mountainous regions, and glacial deposits (super-object, Level 1) which are then refined in successive steps into finer landscape objects until a desired terrain decomposition is achieved (Level 2, Level 3,...).

3.3.3. Application example: Spatial sampling

In order to understand trends and patterns in natural resource management and environmental sciences we often need to draw inferences from samples to entire populations. Samples therefore play an important role in gaining understanding of our environment. They were needed as inputs for modeling as well as for validation of the model output.

Based on the disaggregated landscape in homogeneous terrain objects, I developed a sampling method to distribute sampling locations in an area of interest (Häring and Schröder 2010b). These locations have been used for a large-scale field validation campaign in order to estimate the prediction accuracy of digital soil maps (cf. Häring et al. 2012; Häring and Schröder 2010a). The method has also been used to distribute "sampling" locations on an available soil map to generate a training dataset for spatial prediction models for digital soil mapping (Häring and Schröder 2010b).

Bearing the importance of topography on soil development in mind, a sample should be representative for all combinations of soil type and terrain objects which occur within the study area, i.e. a range of different environmental gradients should be represented. In addition, the distribution of sampling *locations* should be well-balanced over the study area. The sampling method based on terrain objects aims at considering both the spatially-balanced distribution in geographic space and the well-balanced distribution in the multivariate feature space.

Using terrain objects as strata to distribute sampling locations has been proven to be very flexible. The method is an iterative sampling procedure. This means that one can incorporate several environmental variables either in a stepwise fashion or simultaneously to delineate homogeneous environmental units. Secondly, because the method is iterative and considers multivariate environmental gradients very flexibly, it can take into account different spatial scales. In a first iteration, the study area is divided into comparatively coarse-grained environmental units, e.g. fluvial plains or mountain areas. Then these units are subsequently subdivided so that the scale of examination becomes finer with each iteration step. In addition, the sampling method does not need to consider any specific preconditions such as the shape of the study area or the known form of the trend or the spatial structure of the residuals. Lastly, the method is able to treat different subareas differently, so one can sample subareas with higher sampling densities (i.e., number of sampling points per unit area) than others.

3.3.4. Application example: Geomorphographic map of Bavaria

In the *Maps for the future* project a map of geomorphographic units has been prepared based on the application of image segmentation to terrain attributes (cf. Koschitzki et al. 2011). The map has been used as modelling input for the spatial disaggregation of complex soil map unit as introduced in Häring et al. (2012).

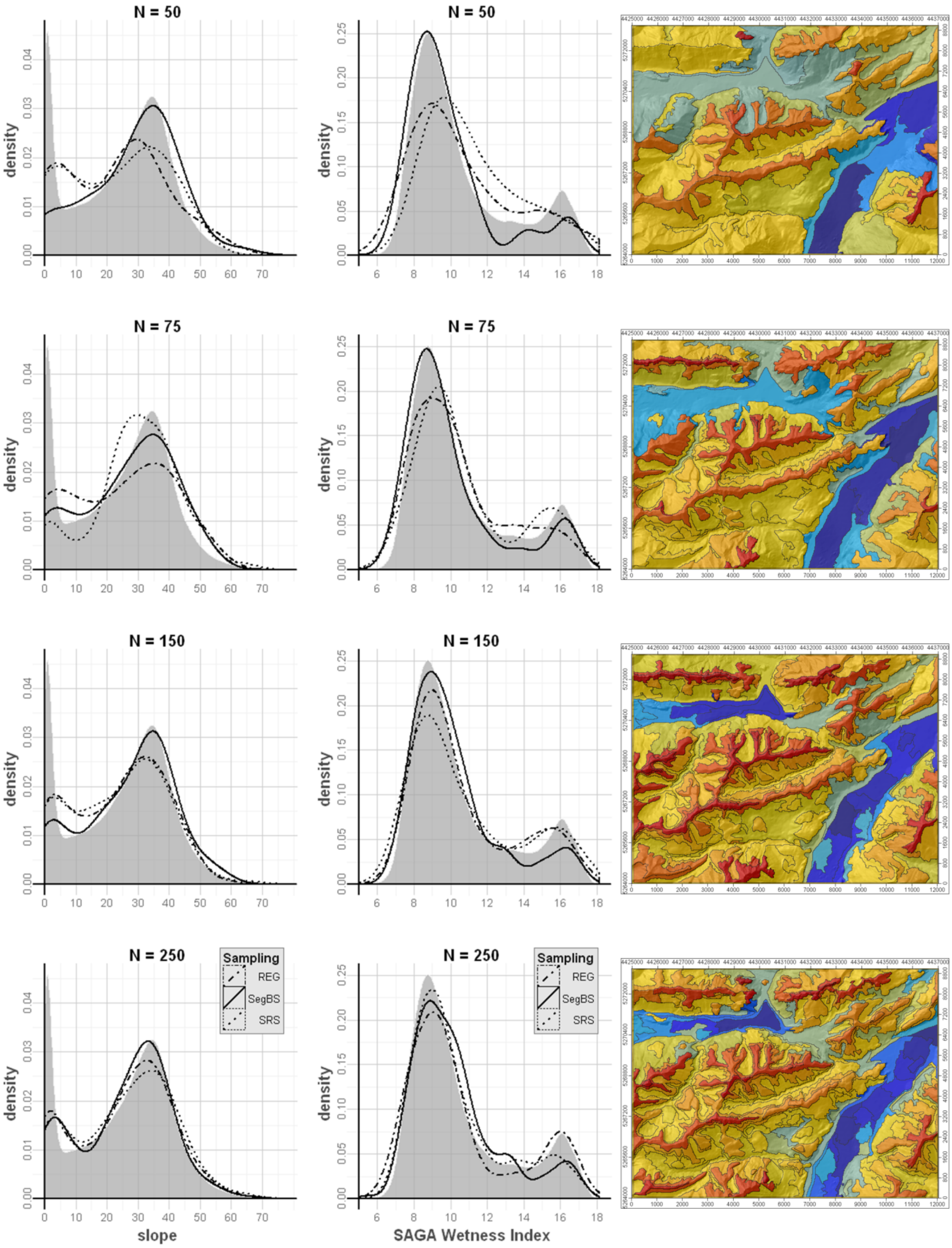


Figure 3.12.: Density plots for slope (column 1) and SAGA Wetness Index (column 2) for sampling based on image segmentation (SegBS), simple random sampling (SRS), and regular sampling (REG) and for four sampling intensities (row one to four). The solid dark grey shadow indicates the underlying population (entire grid). The third column shows the single segments for every particular sampling intensity.

Chapter 4.

Concluding summary

"You must tell me if the humans are doing anything for the forest," said the fox.
"Yes, you may be sure they are!" said Karr. "They are working as hard as they can."

(The Wonderful Adventures of Nils
by Selma Lagerloef)

Forest ecosystems are very likely to be influenced by climate change (Engelbrecht 2012; Kellomäki and Leinonen 2005; Maracchi et al. 2005). The impact of changing climate conditions on forest will be manifold, so that forest scientists additionally use the term *site change* besides the term climate change (cf. Bolte et al. 2010). Climate change will affect especially areas that are highly vulnerable to changing environmental conditions, like mountainous regions (Loarie et al. 2009; Parry et al. 2007). The adaptation of today‘s forest to future climate conditions is therefore an important concern for forest science (Bolte et al. 2009, 2010; Canadell and Raupach 2008; Engelbrecht 2012; Spittlehouse 2005). But also beyond forest science this adaptation is an important concern, e.g. to mitigate global climate change (cf. Canadell and Raupach 2008; Chazdon 2008). Forest management therefore requires, among other things, detailed information on site ecological conditions to adapt forests to future conditions in a sustainable way (Seidl et al. 2011).

The major objective of this thesis was to develop spatial prediction methods derived in the field of digital soil mapping and species distribution modelling for forest site classification, particularly focusing on Bavarian forest sites. All methods and examples in this thesis result from research in two strongly application-oriented projects within the Bavarian forest administration (Beck et al. 2012; Reger and Ewald 2011). Spatial prediction methods have been used for the assessment and mapping of forest site characteristics in Bavaria (Falk and Mellert 2011; Reger et al. 2014, 2011), but also elsewhere (DeLong et al. 2010; Herbst et al. 2012; Jansen et al. 2002; MacMillan et al. 2007; Peters et al. 2011; Schwärzel et al. 2009, 2011; Zirlewagen and Wilpert 2011). Environmental resources management has arrived in the digital world, not only at Universities and research institutes but also in the daily work of authorities (Asche and Schulz 2004; Beck et al. 2012; Gauer 2010; Zirlewagen and Wilpert 2011). Spatial prediction methods are now widely used across different terrestrial realms (e.g. soil attributes, vegetation, or animals), but also freshwater, and marine realms (Elith and Leathwick 2009). However, there are still some areas of critical discussion in the literature, e.g. the use of spatial modelling for extrapolation of species distribution into future or past climates or to new and unsampled geographic areas, as well as to gain ecological understanding (De Marco et al. 2008; Dormann 2007; Elith and Leathwick 2009; Guisan et al. 2007; Midgley et al. 2006). Nevertheless, information technology and computational power in combination with soil and vegetation databases at all spatial scales and the increasing power of tools such as GIS, GPS, remote

and proximal sensing helps also ecologists, forest, and soil scientists in forest site classification (Jansen et al. 2002; Schwärzel et al. 2011).
The development and science of early forest site classification at the beginning of the last century and two exemplary holistic site classification systems, the biogeoclimatic ecosystem classification of British Columbia and the site classification in Bavaria, have been described in the state of the art section of this study (section 1.1). These traditional approaches constitute the basic principle of site-classification until today. However, it has been explained that these approaches are rather qualitative frameworks to classify site ecological conditions. The consequent implementation of these frameworks in maps, which was identified as the heart of classification (Barnes et al. 1982), is rare or at least remains in the qualitative implementation of rule-sets in a GIS (Clare and Ray 2001; Skidmore et al. 1991).
The urgent need for quantitative and high-resolution site information on the one hand and the possible improvement of site classification along with computer-based spatial prediction techniques (section 1.2) on the other lead to challenges for traditional forest site classification. To face these challenges the research objective of this study was to develop and apply spatial prediction techniques for the assessment of forest site characteristics. The intention of this thesis was to introduce different new modelling approaches for site mapping in order to support ongoing and future forest management in Bavaria with a toolbox of mapping techniques. It was not the intention to replace the well-tried and established mapping approaches, but to illustrate different ways in which to include computational and geospatial technology in the assessment of forest site characteristics. The aim was to improve site mapping and move towards a new generation of site classification.

Whereas the purpose of traditional site classification was the assessment and illustration of site characteristics, the new generation will not be pure site maps any more. The application oriented world requires answers for specific questions, e.g. the suitability of a site for planting tree species today and in the future, the possibility to use heavy vehicles, the vulnerability regarding pest infestation, or biomass utilization. Different and heterogeneous sources of information have to be considered for a complex representation of the environment (cf. Dobos and Hengl 2009). Therefore, open and dynamic information systems are nowadays the preferred method (Haeussler 2011). The opportunity now exists to use geospatial technology for grouping assemblages of site data. Newly developed statistical frameworks facilitate the transparent characterization of site units while minimizing the need for subjective decisions (Holt et al. 2013). Thereby it is not all about the creation of *new* data, but often also to reuse, improve, or update existing site data (cf. examples from different fields of ecology at all spatial scales in Fitterer et al. 2012; Fränzle et al. 2006; Häring et al. 2012; Holt et al. 2013; Kempen et al. 2009).

Following strictly the methodological framework developed in the realm of digital soil mapping and species distribution modelling, which form the methodological starting point for my thesis, the modelling approaches presented here are based on statistics. It could be shown that each different method offers a valuable improvement for site mapping approaches and the implementations in the projects *Maps for the future* and *WINALP* have demonstrated their real-world feasibility. However, two alternatives to purely statistical approaches for spatially distributed modelling of site characteristics have been presented. Herbst et al. (2012), Mosimann et al. (2011), and Mosimann and Herbst (2013) present an approach to combine statistical modelling with an empirical-knowledge based approach to assess forest soil properties in northern Switzerland. They define different rule-sets of empirically derived relationships between terrain characteristics and soil properties and apply these relationships to a statistical modelling framework. Within the *WINALP* project a similar approach was chosen to model forest types of special sites like raised bogs, fens, karstic plateaus, and alluvial forests. Reger et al. (2014) applied empirically derived rule-sets to spatially modelled input data. The third approach to model site characteristics, mainly to assess soil hydrological conditions, is the combination of process based hydrological models with statistical models. Schultze et al. (2005) and Falk et al. (2008) describe the approach of the Bavarian state institute of forestry to model and regionalize the soil water available for tree species (transpiration difference). They modelled numerous combinations of site conditions with the hydrologic model LWF-Brook 90 to determine the transpiration difference T_{diff}, the difference between potential and actual transpiration. After running the process model they use a linear multiple regression model to regionalize T_{diff}. Later on, within the *KLIP4* project, boosted regression trees have been used for regionalization (Osenstetter et al. 2013). Similar to Falk et al. (2008), Falk et al. (2011) present a concept to estimate the soil oxygen deficiency. Similarly, Schwärzel et al. (2009) present an approach from Saxony by combining LWF-Brook 90 with GIS data to to simulate the daily water fluxes and soil moisture status. Peters et al. (2011)

apply fuzzy inference systems to regionalize LWF Brook 90 results.

All these approaches have their pros and cons and have to be regarded as different opportunities for spatial modelling of forest sites. Empirical and rule-based approaches have a benefit in their ability to integrate local expert knowledge into the modelling framework. Spatial prediction results can be modified by hand if necessary. In addition, these approaches are more comprehensible than purely data-driven modelling and thereby in some cases more acceptable because they consist of understandable decision rules. On the other hand, due to their detailed adaptation to local conditions which results in only limited generalization of the model, expert and rule-set based approaches are difficult to transfer to other study areas. Approaches which integrate process based hydrological modelling are the most complex ones. They can be used to address very specific questions regarding site characteristics (eg. moisture deficiency for Norway spruce during vegetation period, water logging that persists for a certain length of time). However, because of their specificity it may be difficult to find spatially explicit data to extrapolate from point to the area of interest. Hydrological models like LWF-Brook 90 require detailed and varied data for calibration, which are not always available for the entire study area. Moreover, the more complex the parameter of interest, the more difficult it is to validate the model after extrapolation. Statistical modelling approaches, as presented in this thesis, offer a comparatively easy way for spatially distributed modelling, even though the statistical model itself can be very complex and even a black box model. Modelling results are reproducible and comprehensive though not simple, cf. the different model effects in Häring et al. (2013). However, the quality of statistical modelling results rely mostly on the quality of input data. The application of advanced statistical models to GIS data and the integration of such approaches to site information systems has become very straightforward. This is mainly due to recent developments in the availability of free and open-source modelling tools and the growing GIS capabilities of these tools, e.g. the R modelling language (R Core Team 2013), which has been used throughout this thesis.

The site-information system of the *WINALP* project could be regarded as an exemplary implementation of a modern and holistic approach for the assessment of site characteristics. The system is based on different types and sources of environmental information, ranging from a substantial amount of field data on vegetation, soil and geomorphology, existing digital maps on geology, topography,

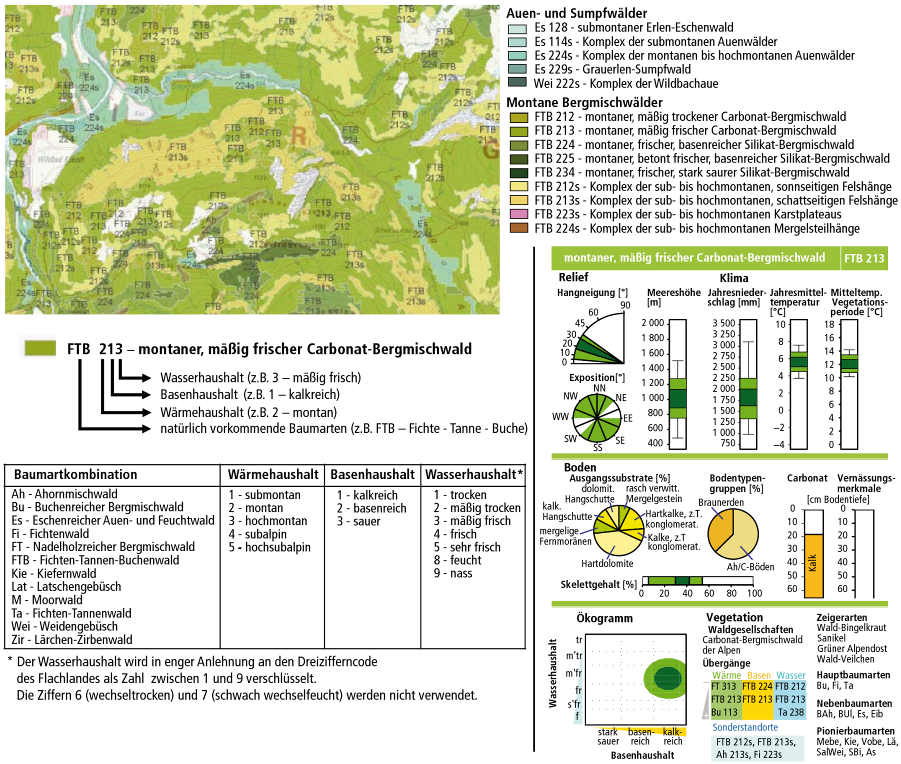


Figure 4.1.: Example of the intermediate-scale map of potential natural forest vegetation of the Bavarian Alps, scale 1:25,000 with the example of FTB 213 - Montane mixed mountain forest on submesic calcareous soils (Source: (Reger and Ewald 2011)).

soils, and climate to advanced spatial modelling applications, as well as expert knowlegde. These data have been prepared to fit into the well-tried qualitative site classification code (Reger and Ewald 2011), cf. Figure 4.1.

My two studies on regionalizing Ellenberg indicator values in section 2.2 and section 2.3 could contribute directly to this site information system. Spatial prediction methods have been used in addition to soil reaction (Häring et al. 2014) and soil moisture (Häring et al. 2013), as well as for temperature (Reger et al. 2011). All three parameters have been used as data input for the three-dimensional temperature-reaction-moisture system (TRM model) introduced by Reger et al. (2014). These three ecological factors enable the establishment of a three-dimensional "site cube", in which a forest type can be defined as homogeneous ecological unit. Field validation at 6358 validation sites results in a total accuracy of the modelled forest types of about 71%.

The map of potential natural forest vegetation can be freely accessed via an online webgis application (`http://arcgisserver.hswt.de/winalp`) and may be of practical use in forestry as it can serve as planning instrument in the regional context. The intermediate spatial resolution of 1:25,000, however, remains a limiting factor for local application. For a detailed management of mountain forests, however, human expertise is still needed. The map will further support research that aims to outline areas potentially sensitive to climate change or highly relevant for nature conservation.

List of References

Ad-hoc-AG Boden (2005). *Bodenkundliche Kartieranleitung.* Bundesanstalt für Geowissenschaften und Rohstoffe in Zusammenarbeit mit den Staatlichen Geologischen Diensten der Bundesrepublik Deutschland, Hannover, 5. edition. 5, 12, 36

Aertsen, W., Kint, V., van Orshoven, J., Ozkan, K., and Muys, B. (2010). Comparison and ranking of different modelling techniques for prediction of site index in mediterranean mountain forests. *Ecological Modelling*, 221(8):1119–1130. 4

Altun, L., Başkent, E., Bakkaloğlu, M., Günlü, A., and Kadioğullari, A. (2008). Comparing methods for determining forest sites: a case study in Gümüşhane-Karanlikdere forest. *European Journal of Forest Research*, 127:395–406. 4

Arbeitskreis Standortskartierung (2003). *Forstliche Standortsaufnahme. Begriffe, Definitionen, Einteilungen, Kennzeichnungen, Erläuterungen.* IHW-Verlag, Eching, 6. edition. 3, 5, 6, 11, 12

Asche, N. and Schulz, R. (2004). Standortsklassifikation mit digitalen Werkzeugen und die digitale Waldtypenkarte in Nordrhein-Westfalen. *Forst und Holz*, 59:428–431. 25, 58

Baatz, M. and Schäpe, A. (2000). Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. In Strobl, J., Blaschke, T., and Griesebner, G., editors, *Angewandte Geographische Informations-Verarbeitung XII*, pages 12–23. Wichmann Verlag, Karlsruhe. 52

Bailey, R., Avers, P., King, T., and McNab, W., editors (1994). *Ecoregions and subregions of the United States (map). 1:7,500,000.* U.S. Department of Agriculture, Forest Service, Washington, DC. With supplementary table of map unit descriptions, compiled and edited by W.H.McNab & R.G. Bailey. 10

Bailey, R., Pfister, R., and Henderson, J. (1978). Nature of Land and Resource classification - review. *Journal of Forestry*, 76(10):650–655. 3, 5

Bailey, R. G. (2009). *Ecosystem geography: from ecoregions to sites*. Springer, New York [u.a.], 2. edition. 5, 7

Barnes, B. V., Pregitzer, K. S., Spies, T. A., and Spooner, V. H. (1982). Ecological forest site classification. *Journal of Forestry*, 80(8):493–498. 5, 7, 59

Bates, C. G. (1918). Concerning site. *Journal of Forestry*, 16(4):383–388. 4

Bauer, E. and Kohavi, R. (1999). An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine Learning*, 36(1-2):105–139. 38

Beck, J., Dietz, E., and Falk, W. (2012). Digitales Standortinformationssystem für Bayern. *LWF aktuell*, 87:20–23. 25, 58

Begon, M., Townsend, C. R., and Harper, J. L. (2006). *Ecology: from individuals to ecosystems*. Blackwell Publ., Malden, MA [u.a.], 4. edition. 5

Behrens, T., Schmidt, K., Zhu, A. X., and Scholten, T. (2010). The ConMap approach for terrain-based digital soil mapping. *European Journal of Soil Science*, 61(1):133–143. 47

Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I., and Heynen, M. (2004). Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS J. Photogr. Remote Sens.*, 58(3-4):239–258. 52

Blaschke, T. and Strobl, J. (2001). What‘s wrong with pixels? Some recent developments interfacing remote sensing and GIS. *GIS - Zeitschrift für Geoinformationssysteme*, 6:12–17. 52

Boettinger, J., Howell, D., Kienast-Brown, S., Hartemink, A. E., and Moore, A. (2010). *Digital Soil Mapping: Bridging Research, Environmental Application, and Operation*. Springer. 17, 20

Bolte, A., Ammer, C., Löf, M., Madsen, P., Nabuurs, G.-J., Schall, P., Spathelf, P., and Rock, J. (2009). Adaptive forest management in central europe: Climate change impacts, strategies and integrative concept. *Scandinavian Journal of Forest Research*, 24(6):473–482. 58

Bolte, A., Ammer, C., Löf, M., Nabuurs, G.-J., Schall, P., and Spathelf, P. (2010). Adaptive forest management: A prerequisite for sustainable forestry in the face of climate change. In Spathelf, P., editor, *Sustainable Forest Management in a Changing World*, volume 19 of *Managing Forest Ecosystems*, pages 115–139. Springer Netherlands. 58

Braun-Blanquet, J. (1964). *Pflanzensoziologie*. Springer, Wien. 5

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2):123–140. 17, 38

Breiman, L. (2001a). Random forests. *Machine Learning*, 45(1):5–32. 38

Breiman, L. (2001b). Statistical modeling: The two cultures. *Statistical Science*, 16(3):199–231. 17

Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). *Classification and regression trees*. Wadsworth, Montery, CA. Classification and regression trees. 17

Brus, D., Kempen, B., and Heuvelink, G. (2011). Sampling for validation of digital soil maps. *European Journal of Soil Science*, 62(3):394–407. 41

Bühlmann, P. and Yu, B. (2002). Analyzing bagging. *The Annals of Statistics*, 30(4):927–961. 38

Burnett, C. and Blaschke, T. (2003). A multi-scale segmentation/object relationship modelling methodology for landscape analysis. *Ecological Modelling*, 168:233–249. 52, 53

Canadell, J. G. and Raupach, M. R. (2008). Managing forests for climate change mitigation. *Science*, 320(5882):1456–1457. 58

Carmean, W. (1975). Forest site quality evaluation in the United States. *Advances in Agronomy*, 27:209–270. 4

Carre, F., McBratney, A. B., Mayr, T., and Montanarella, L. (2007). Digital soil assessments: Beyond DSM. *Geoderma*, 142(1-2):69–79. 19, 23, 26

Chazdon, R. L. (2008). Beyond deforestation: Restoring forests and ecosystem services on degraded lands. *Science*, 320(5882):1458–1460. 58

Clare, J. and Ray, D. (2001). A spatial model of ecological site classification for forest management in Britain. In *Proceedings of the 4th AGILE Conference on Geographic Information Science Brno*. 59

CORINE (1991). *Biotopes Manual. A method to identify and describe consistently sites of major importance for nature conservation data specifications*. European Communities - Commission EUR 12587. 3

Dahl, F. (1908). Grundsätze und Grundbegriffe der biocönotischen Forschung. *Zoologischer Anzeiger*, 33:349–353. 3

Daubenmire, R. (1976). The use of vegetation in assessing the productivity of forest lands. *Botanical Review*, 42(2):115–143. 4, 5

De Marco, Jr., P., Felizola Diniz-Filho, J. A., and Mauricio Bini, L. (2008). Spatial analysis improves species distribution modelling during range expansion. *Biology Letters*, 4(5):577–580. 58

DeLong, C., Griesbauer, H., Mackenzie, W., and Foord, V. (2010). Corroboration of biogeoclimatic ecosystem classification climate zonation by spatially modelled climate data. *BC Journal of Ecosystems and Management*, 10(3):49–64. 10, 58

Deumlich, D., Schmidt, R., and Sommer, M. (2010). A multiscale soil-landform relationship in the glacial-drift area based on digital terrain analysis and soil attributes. *Journal of Plant Nutrition and Soil Science*, 173(6):843–851. 47

Diekmann, M. (2003). Species indicator values as an important tool in applied plant ecology - a review. *Basic and Applied Ecology*, 4(6):493–506. 6

Dierschke, H. (1994). *Pflanzensoziologie*. Verlag Eugen Ulmer, Stuttgart. 6

Dietterich, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning*, 40(2):139–157. 38

Dobos, E. and Hengl, T. (2009). Soil Mapping Applications. In Hengl, T. and Reuter, H., editors, *Geomorphometry - Concepts, Software, Applications*, pages 461–480. Elsevier, Amsterdam. 18, 59

Dormann, C. (2007). Promising the future? Global change projections of species distributions. *Basic and Applied Ecology*, 8:387–97. 58

Drăgut, L. and Blaschke, T. (2006). Automated classification of landform elements using object-based image analysis. *Geomorpholohy*, 81:330–344. 51, 52

Drăgut, L., Schauppenlehner, T., Muhar, A., Strobl, J., and Blaschke, T. (2009). Optimization of scale and parametrization for terrain segmentation: An application to soil-landscape modeling. *Computers & Geosciences*, 35(9):1875–1883. 52

Elith, J. and Graham, C. H. (2009). Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography*, 32(1):66–77. 38

Elith, J. and Leathwick, J. R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, 40:677–697. 18, 20, 22, 58

Elith, J., Leathwick, J. R., and Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4):802–813. 17

Ellenberg, H., editor (1967). *Vegetations- und bodenkundliche Methoden der forstlichen Standortskartierung: Ergebnisse eines internationalen Methodenvergleichs im Schweizer Mittelland*. Veröffentlichungen des Geobotanischen Instituts der Eidgenössischen Technischen Hochschule, Stiftung Rübel, 39, Zürich. 5

Ellenberg, H. and Leuschner, C. (2010). *Vegetation Mitteleuropas mit den Alpen in ökologischer, dynamischer und historischer Sicht*. UTB für Wissenschaft. Ulmer, Stuttgart. 6., vollst. neu bearb. und stark erw. Aufl. 6

Ellenberg, H., Weber, H., Düll, R., Wirth, V., and Werner, W. (2001). *Zeigerwerte von Pflanzen in Mitteleuropa*. Verlag Erich Goltze KG, Göttingen. 6

Engelbrecht, B. M. J. (2012). Forests on the brink. *Nature*, 491(7426):675–677. 58

Evans, I. S., Hengl, T., and Gorsevski, P. (2009). Applications in geomorphology. In Hengl, T. and Reuter, H. I., editors, *Geomorphometry. Concepts, Software, Applications*, pages 497–526. Elsevier, Amsterdam. 52

Ewald, J. (2007). Beurteilung von Waldstandorten und Waldgesellschaften mit Zeigerarten-Ökogrammen. *Tuexenia*, 27:7–18. 6

Falk, W., Dietz, E., Grünert, S., Schultze, B., and Kölling, C. (2008). Wo hat die Fichte genügend Wasser? *LWF aktuell*, 66:21–25. 60

Falk, W. and Mellert, K. (2011). Species distribution model as a tool for forest management planning under climate change: risk evaluation of Abies alba in Bavaria. *Journal of Vegetation Science*, 22(4):621–634. 58

Falk, W., Osenstetter, S., and Dietz, E. (2011). Konzept für ein quantitatives Verfahren zur Kennzeichnung des Lufthaushaltes von Waldböden in Bayern. *Waldökologie, Landschaftsforschung und Naturschutz*, 12:93–99. 60

Fitterer, J. L., Nelson, T. A., Coops, N. C., and Wulder, M. A. (2012). Modelling the ecosystem indicators of British Columbia using Earth observation data and terrain indices. *Ecological Indicators*, 20:151 – 162. 10, 24, 25, 60

Florinsky, I. (2012). The dokuchaev hypothesis as a basis for predictive digital soil mapping (on the 125th anniversary of its publication). *Eurasian Soil Science*, 45:445–451. 19

Franklin, J. (2010). *Mapping Species Distributions. Spatial Inference and Prediction.* Cambridge University Press. 18, 22

Fränzle, O., Müller, F., and Schröder, W., editors (2006). *Handbuch der Umweltwissenschaften. Grundlagen und Anwendungen der Ökosystemforschung*, volume 17. Erg. Lfg. Wiley VCH. 60

Frehner, H.-K. (1967). Kartierung der Waldgesellschaften des V. aargauischen Forstkreises Zofingen. In Ellenberg, H., editor, *Vegetations- und bodenkundliche Methoden der forstlichen Standortskartierung: Ergebnisse eines internationalen Methodenvergleichs im Schweizer Mittelland*, pages 135–149. Veröffentlichungen des Geobotanischen Instituts der Eidgenössischen Technischen Hochschule, Stiftung Rübel, 39. 5

Frothingham, E. H. (1921). Classifying forest sites by height growth. *Journal of Forestry*, 19(4):374–381. 4

Gauer, J. (2010). Gis-gestützte Weiterentwicklung von forstlichen Standortskarten. *AFZ - Der Wald*, 64:32–35. 25, 58

Gauer, J., Feger, K. H., and Schwärzel, K. (2011). Erfassung und Bewertung des Wasserhaushalts von Waldstandorten in der forstlichen Standortskartierung:

Gegenwärtiger Stand und künftige Anforderungen. *Waldökologie, Landschaftsforschung und Naturschutz*, 12:7–16. 5

Grey, D. (1980). On the concept of site in forestry. *South African Forestry Journal*, 113(1):81–83. 3

Griesbauer, H. P. and Green, D. S. (2012). Geographic and temporal patterns in white spruce climate-growth relationships in Yukon, Canada. *Forest Ecology and Management*, 267:215–227. 10

Grunwald, S. (2009). Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma*, 152(3-4):195–207. 20

Guisan, A. and Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8(9):993–1009. 18, 22

Guisan, A., Zimmermann, N. E., Elith, J., Graham, C. H., Phillips, S., and Peterson, A. T. (2007). What matters for predicting the occurrences of trees: Techniques, data, or species' characteristics? *Ecological Monographs*, 77(4):615–630. 58

Haeussler, S. (2011). Rethinking biogeoclimatic ecosystem classification for a changing world. *Environmental Reviews*, 19:254–277. 7, 24, 59

Hamann, A. and Wang, T. (2006). Potential effects of climate change on ecosystem and tree species distribution in British Columbia. *Ecology*, 87(11):2773–2786. 10, 24

Häring, T., Dietz, E., and Kölling, C. (2009). Zusammenhang zwischen Rastergröße und Modellgüte für die Prognose von Bodenkarten im Maßstab 1 : 25.000. In *Jahrestagung der Deutschen Bodenkundlichen Gesellschaft: Böden - eine endliche Ressource, September 2009, Bonn, Berichte der DBG*. 38

Häring, T., Dietz, E., Osenstetter, S., Koschitzki, T., and Schröder, B. (2012). Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils. *Geoderma*, 185-186:37–47. 27, 36, 38, 41, 47, 54, 55, 60

Häring, T., Reger, B., Ewald, J., Hothorn, T., and Schröder, B. (2013). Predicting Ellenbergs soil moisture indicator value in the Bavarian Alps using additive georegression. *Applied Vegetation Science*, 16:110–121. 61, 62

Häring, T., Reger, B., Ewald, J., Hothorn, T., and Schröder, B. (2014). Regionalising indicator values for soil reaction in the Bavarian Alps - from averages to multivariate spectra. *Folia Geobotanica*, DOI 10.1007/s12224-013-9157-1. 62

Häring, T. and Schröder, B. (2010a). A review of model-error in digital soil mapping: Confronting statistical soil landscape models with large-scale field validation data. *Geophysical Research Abstracts*, 12:EGU2010–12757. 54

Häring, T. and Schröder, B. (2010b). Sampling optimization using image segmentation. In *Proceedings of the 4th International Workshop on Digital Soil Mapping*, 24.-26.05.2010, Rome. 54

Hartemink, A. E., Hempel, J., Lagacherie, P., McBratney, A., McKenzie, N., MacMillan, R. A., Minasny, B., Montanarella, L., Mendonça Santos, M. L., Sanchez, P., Walsh, M., and Zhang, G.-L. (2010). Globalsoilmap.net - a new digital soil map of the world. In Boettinger, J. L., Howell, D. W., Moore, A. C., Hartemink, A. E., Kienast-Brown, S., Hartemink, A. E., and McBratney, A. B., editors, *Digital Soil Mapping*, volume 2 of *Progress in Soil Science*, pages 423–428. Springer Netherlands. 20

Hartemink, A. E. and McBratney, A. (2008). A soil science renaissance. *Geoderma*, 148:123–129. 20

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall. 17

Hastie, T. J., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Series in Statistics. Springer, New York. 17

Hengl, T., Heuvelink, G. B. M., and Rossiter, D. G. (2007). About regression-kriging: From equations to case studies. *Computers & Geosciences*, 33(10):1301–1315. 18

Hengl, T. and MacMillan, R. A. (2009). Geomorphometry - a key to landscape mapping and modelling. In Hengl, T. and Reuter, H. I., editors, *Geomorphometry. Concepts, Software, Applications*, volume 33, pages 433–460. Elsevier. 51

Hengl, T. and Reuter, H. I. (2009). *Geomorphometry. Concepts, Software, Applications*. Elsevier. 51

Hera, U., Rötzer, T., Zimmermann, L. Schulz, C., Maier, H., Weber, H., and Kölling, C. (2012). Klima en détail. Neue, hochaufgelöste Klimakarten bilden wichtige Basis zur klimatischen Regionalisierung Bayerns. *LWF aktuell*, 86:34–37. 38

Herbst, P., Gross, J., Meer, U., and Mosimann, T. (2012). Geomorphographic terrain classification for predicting forest soil properties in Northwestern Switzerland. *Zeitschrift für Geomorphologie*, 56(1):1–22. 58, 60

Hessburg, P. F., Salter, R. B., Richmond, M. B., and Smith, B. G. (2000). Ecological subregions of the Interior Columbia Basin, USA. *Applied Vegetation Science*, 3(2):163–180. 10

Hilgard, E. (1921). *Soils: Their formation, properties, composition, and plant growth in the humid and arid regions*. Macmillan, London. 51

Holt, B. G., Lessard, J.-P., Borregaard, M. K., Fritz, S. A., Araùjo, M. B., Dimitrov, D., Fabre, P.-H., Graham, C. H., Graves, G. R., Jønsson, K. A., Noguès-Bravo, D., Wang, Z., Whittaker, R. J., Fjeldså, J., and Rahbek, C. (2013). An update of wallace‘s zoogeographic regions of the world. *Science*, 339(6115):74–78. 60

Huggett, R. and Cheesman, J. (2002). *Topography and the Environment*. Prentice Hall, Harlow. 51

Jahn, G., editor (1982). *Application of vegetation science to forestry*, volume 12 of *Handbook of vegetation science*. Junk, The Hague [u.a.]. 8

Jansen, M., Judas, M., and Saborwski, J., editors (2002). *Spatial modelling in forest ecology and management*. Springer, Berlin. 58, 59

Jenny, H. (1941). *Factors of soil formation: A system of quantitative pedology*. McGraw-Hill, Mineola, N.Y. 19, 51

Jones, R. J., Houškovà, B., Bullock, P., and Montanarella, L., editors (2005). *Soil Ressources of Europe*. European Soil Bureau Research Report No.9, EUR 20559 EN, Office for Official Publications of the European Communities, Luxembourg. 36

Kaule, G. (2002). *Umweltplanung*. Ulmer, Stuttgart. 3

Kellomäki, S. and Leinonen, S., editors (2005). *Management of European Forests Under Changing Climatic Conditions.* Tiedonantoja/Research Notes No. 163. University of Joensuu, Faculty of Forestry, Finnland. 58

Kempen, B., Brus, D., Heuvelink, G. B. M., and Stoorvogel, J. (2009). Updating the 1:50,000 dutch soil map using legacy soil data: A multinomial logistic regression approach. *Geoderma*, 151:311–326. 60

Kohm, K. A. and Franklin, J. F., editors (1997). *Creating a forestry for the 21st century: the science of ecosystem management.* Island Press, Washington, D.C. [u.a.]. 4

Kojima, S. (1991). Classification and ecological characterization of coniferous forest phytogeocoenoses of Hokkaido, Japan. *Vegetatio*, 96(1):25–42. 10

Koschitzki, T., Möller, M., and Häring, T. (2011). Automatische Reliefklassifikation - eine kurze Einführung. In *Posterpresentation, workshop on digital soil mapping, Kommission V der DBG, 06/2011*. 55

Krestov, P., Song, J., Nakamura, Y., and Verkholat, V. (2006). A phytosociological survey of the deciduous temperate forests of mainland Northeast Asia. *Phytocoenologia*, 36(1):77–150. 10

Lagacherie, P., McBratney, A. B., and Voltz, M. (2007). *Digital Soil Mapping. An Introductory Perspective*, volume 31. Elsevier, Amsterdam. 20

Liess, M., Glaser, B., and Huwe, B. (2012). Uncertainty in the spatial prediction of soil texture. Comparison of regression tree and Random Forest models. *Geoderma*, 170:70–79. 38

Lindenmayer, D., Margules, C., and Botkin, D. (2000). Indicators of biodiversity for ecologically sustainable forest management. *Conservation Biology*, 14(4):941–950. 4

Liu, Q., Kondoh, A., and Takeuchi, N. (1998). Study of changes in life zone distribution in north-east China by climate-vegetation classification. *Ecological Research*, 13(3):355–365. 10

Loarie, S. R., Duffy, P. B., Hamilton, H., Asner, G. P., Field, C. B., and Ackerly, D. D. (2009). The velocity of climate change. *Nature*, 462(7276):1052–1055. 58

MacMillan, R. A., Moon, D. E., and Coupe, R. A. (2007). Automated predictive ecological mapping in a forest region of BC, Canada, 2001-2005. *Geoderma*, 140(4):353–373. 7, 10, 24, 25, 58

MacMillan, R. A. and Shary, P. A. (2009). Landforms and landform elements in geomorphometry. In Hengl, T. and Reuter, H. I., editors, *Geomorphometry. Concepts, Software, Applications*, pages 227–254. Elsevier, Amsterdam. 51

Maracchi, G., Sirotenko, O., and Bindi, M. (2005). Impacts of present and future climate variability on agriculture and forestry in the temperate regions: Europe. *Climatic Change*, 70:117–135. 58

Martin, W. K. E. and Timmer, V. R. (2006). Capturing spatial variability of soil and litter properties in a forest stand by landform segmentation procedures. *Geoderma*, 132(1-2):169–181. 51

McBratney, A. B., Mendonça Santos, M. L., and Minasny, B. (2003). On digital soil mapping. *Geoderma*, 117(1-2):3–52. 17, 19, 23, 41, 47

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. Chapman and Hall, London. 17

McKenzie, N. J. and Ryan, P. J. (1999). Spatial prediction of soil properties using environmental correlation. *Geoderma*, 89(1-2):67–94. 18

Meidinger, D. and Pojar, J. (1991). Ecosystems of British Columbia. Technical report, B. C. Ministry of Forests Special Report Series 6. 7, 10, 18

Midgley, G., Hughes, G., Thuiller, W., and Rebelo, A. (2006). Migration rate limitations on climate change induced range shifts in cape proteaceae. *Diversity and Distributions*, 12:555. 58

Milne, G. (1935). Some suggested units of classification and mapping particularly for east african soils. *Soil Research*, 4:193–198. 47

Minár, J. and Evans, I. S. (2008). Elementary forms for land surface segmentation: The theoretical basis of terrain analysis and geomorphological mapping. *Geomorphology*, 95(3-4):236–259. 51

Möller, M., Koschitzki, T., Hartmann, K.-J., and Jahn, R. (2012). Plausibility test of conceptual soil maps using relief parameters. *Catena*, 88(1):57–67. 52

Möller, M., Volk, M., Friedrich, K., and Lymburner, L. (2008). Placing soil-genesis and transport processes into a landscape context: A multiscale terrain-analysis approach. *Journal of Plant Nutrition and Soil Science*, 171(3):419–430. 47, 51, 52

Moore, I., Gessler, P., Nielsen, G., and Peterson, G. (1993). Soil attribute prediction using terrain analysis. *Soil Science Society of America Journal*, 57(2):443–452. 18

Moore, I., Grayson, R., and Ladson, A. (1991). Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrological Processes*, 5(1):3–30. 51

Mosimann, T., Herbst, M., Gross, J., and Meer, U. (2011). Wissensbasierte Modellierung von Waldbodeneigenschaften - Möglichkeiten, Grenzen, Herausforderungen. *Bulletins der Bodenkundlichen Gesellschaft der Schweiz*, 32:43–50. 60

Mosimann, T. and Herbst, P. (2013). Flächenhafte Modellierung von Waldbodeneigenschaften in der Nordwestschweiz. *Schweizer Zeitschrift fü Forstwesen*, 164(1):10–22. 60

Mulder, V. L., de Bruin, S., Schaepman, M. E., and Mayr, T. R. (2011). The use of remote sensing in soil and terrain mapping - A review. *Geoderma*, 162(1-2):1–19. 17

Nakamura, Y., Krestov, P. V., and Omelko, A. M. (2007). Bioclimate and zonal vegetation in Northeast Asia: first approximation to an integrated study. *Phytocoenologia*, 37(3-4):443–470. 10

Nentwig, W., Bacher, S., and Brandl, R. (2012). *Ökologie kompakt*. Spektrum Akademischer Verlag, Heidelberg, 3. edition. 3, 5

Olenin, S. and Ducrotoy, J. (2006). The concept of biotope in marine ecology and coastal management. *Marine Pollution Bulletin*, 53(1-4):20–29. 3

Osenstetter, S., Falk, W., Reger, B., and Beck, J. (2013). Wasser, Luft und Nährstoffe - alles, was ein Baum zum Leben braucht. *LWF aktuell*, 94:12–18. 27, 36, 60

Parry, M., Canziani, O., Palutikof, J., van der Linden, P., and Hanson, C. H., editors (2007). *Climate change 2007*. Cambridge Univ. Press, Cambridge [u.a.], 1. publ. edition. 58

Pennock, D. J. (2003). Terrain attributes, landform segmentation, and soil redistribution. *Soil and Tillage Research*, 69(1-2):15–26. 51

Pennock, D. J. and Corre, M. D. (2001). Development and application of landform segmentation procedures. *Soil and Tillage Research*, 58(3-4):151–162. 51

Peters, R., Schwärzel, K., and Feger, K. H. (2011). Fuzzy-Inference-Systeme zur Regionalisierung des Standortswasserhaushaltes von Wäldern. *Waldökologie, Landschaftsforschung und Naturschutz*, 12:111–117. 58, 60

Peterson, A., Soberón, J., Pearson, R., Anderson, R., Martínez-Meyer, E., Nakamura, M., and Araújo, M. (2011). *Ecological Niches and Geographic Distributions*. Monographs in Population Biology 49. Princeton University Press. 22

Pfadenhauer, J. (1997). *Vegetationsökologie: ein Skriptum*. IHW-Verlag, Eching, 2. edition. 3, 6

Pfister, R. and Arno, S. (1980). Classifying forest habitat types based on potential climax vgetation. *Forest Science*, 26:52–70. 5

Pojar, J., Klinka, K., and Meidinger, D. (1987). Biogeoclimatic ecosystem classification in british columbia. *Forest Ecology and Management*, 22:119–154. 7

Pokharel, B. and Dech, J. P. (2011). An ecological land classification approach to modeling the production of forest biomass. *Forestry Chronicle*, 87(1):23–32. 4

Pyatt, G. (1995). An ecological site classification for forestry in great britain. Technical report, Forestry Commission Research Information Note 260. 10

Quine, C., Humphrey, J., Purdy, K., and Ray, D. (2002). An approach to predicting the potential forest composition and disturbance regime for a highly modified landscape: a pilot study of Strathdon in the Scottish Highlands. *Silva Fennica*, 36(1):233–247. 10

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. 17, 61

Reger, B. and Ewald, J. (2011). Waldtypenkarte Bayerische Alpen. Eine neue Planungshilfe für die Forstpraxis. *AFZ - Der Wald*, 24:14–16. 25, 58, 62

Reger, B., Häring, T., and Ewald, J. (2014). The TRM-model of potential natural vegetation in mountain forests. *Folia Geobotanica*, DOI 10.1007/s12224-013-9158-0. to appear. 25, 27, 32, 34, 36, 58, 60, 62

Reger, B., Kölling, C., and Ewald, J. (2011). Modelling effective thermal climate for mountain forests in the bavarian alps: Which is the best model? *Journal of Vegetation Science*, 22(4):677–687. 58, 62

Roecker, S. M., Howell, D. W., Haydu-Houdeshell, C. A., and Blinn, C. (2010). A qualitative comparison of conventional soil survey and digital soil mapping approaches. In Boettinger, J. L., Howell, D. W., Moore, A. C., Hartemink, A. E., and Kienast-Brown, S., editors, *Digital Soil Mapping*, volume 2, pages 369–384. Springer Netherlands. 38

Roth, F. (1916). Concerning site. *Journal of Forestry*, 14(1):3–12. 4

Sanchez, P. A., Ahamed, S., Carre, F., Hartemink, A. E., Hempel, J., Huising, J., Lagacherie, P., McBratney, A. B., McKenzie, N. J., Mendonça-Santos, M. D., Minasny, B., Montanarella, L., Okoth, P., Palm, C. A., Sachs, J. D., Shepherd, K. D., Vagen, T. G., Vanlauwe, B., Walsh, M. G., Winowiecki, L. A., and Zhang, G. L. (2009). Digital soil map of the world. *Science*, 325(5941):680–681. 20

Schaffers, A. P. and Sykora, K. V. (2000). Reliability of Ellenberg indicator values for moisture, nitrogen and soil reaction: a comparison with field measurements. *Journal of Vegetation Science*, 11(2):225–244. 6

Schlichting, E. (1970). Bodenklassifikation nach Genese oder Effekt. *Mitteilungen der Deutschen Bodenkundlichen Gesellschaft*, 10:13–16. 5

Schölkopf, B. and Smola, A. J. (2002). *Learning with kernels : support vector machines, regularization, optimization, and beyond*. Adaptive computation and machine learning. MIT Press, Cambridge, Mass. [u.a.]. 17

Schönhar, S. (1993). Die Waldbodenvegetation als Standortsweiser. *Allg. Forst- und Jagdzeitung*, 164:173–180. 5

Schultze, B., Kölling, C., Dittmar, C., Rötzer, T., and Elling, W. (2005). Konzept für ein quantitatives Verfahren zur Kennzeichnung des Lufthaushaltes von Waldböden in Bayern: Modellierung - Regression - Regionalisierung. *Forstarchiv*, 76:155–163. 60

Schulze, E.-D., Beck, E., and Müller-Hohenstein, K. (2005). *Plant Ecology*. Springer Verlag Heidelberg. 3, 4

Schwärzel, K., Feger, K. H., Häntzschel, J., Menzer, A., Spank, U., Clausnitzer, F., Kostner, B., and Bernhofer, C. (2009). A novel approach in model-based mapping of soil water conditions at forest sites. *Forest Ecology and Management*, 258(10):2163–2174. 58, 60

Schwärzel, K., Feger, K. H., Köstner, B., and Bernhofer, C. (2011). Wasserhaushalt von Wäldern. *Waldökologie, Landschaftsforschung und Naturschutz*, 12:5–6. 58, 59

Scull, P., Franklin, J., Chadwick, O. A., and McArthur, D. (2003). Predictive soil mapping: a review. *Progress in Physical Geography*, 27(2):171–197. 19

Seidl, R., Rammer, W., and Lexer, M. (2011). Climate change vulnerability of sustainable forest management in the eastern alps. *Climatic Change*, 106:225–254. 58

Shapiro, L. and Stockman, G. (2001). *Computer Vision*. Prentice-Hall, New Jersey. Computer Vision. 52

Skidmore, A. K., Ryan, P. J., Dawes, W., Short, D., and O'Loughlin, E. (1991). Use of an expert system to map forest soils from a geographical information system. *International Journal of Geographical Information Systems*, 5:431–445. 59

Soil Survey Staff (1999). *Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys*. Agriculture handbook 436. U.S. Gov. Print., Washington, 2. edition. 5

Spittlehouse, D. (2005). Integrating climate change adaptation into forest management. *Forestry Chronicle*, 81(5):691–695. 58

Spurr, S. H. and Barnes, B. V. (1980). *Forest ecology*. John Wiley and Sons, Inc, New York, 3. edition. 5

Steinhardt, U. and Volk, M. (2003). Meso-scale landscape analysis based on landscape balance investigations: problems and hierarchical approaches for their resolution. *Ecological Modelling*, 168(3):251–265. 54

Stepinski, T. F. and Bagaria, C. (2009). Segmentation-based unsupervised terrain classification for generation of physiographic maps. *IEEE Geoscience and Remote Sensing Letters*, 6(4):733–737. 51

Strobl, C., Malley, J., and Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, 14(4):323–348. 38

Stum, A. K., Boettinger, J. L., White, M. A., and Ramsey, R. D. (2010). Random forests applied as a soil spatial predictive model in arid utah. In Boettinger, J. L., Howell, D. W., Moore, A. C., Hartemink, A. E., and Kienast-Brown, S., editors, *Digital Soil Mapping. Bridging Reasearch, Environmental Application, and Operation*, volume 2, pages 179–190. Springer Netherlands. 38

Tiktak, A., Boesten, J. J. T. I., van der Linden, A. M. A., and Vanclooster, M. (2006). Mapping ground water vulnerability to pesticide leaching with a process-based metamodel of EuroPEARL. *Journal of Environmental Quality*, 35(4):1213–1226. 18

Urban, D. L., Oneill, R. V., and Shugart, H. H. (1987). Landscape ecology. *Bioscience*, 37(2):119–127. 54

Viscarra-Rossel, R., McBratney, A., and Minasny, B., editors (2010). *Proximal Soil Sensing*. Progress in Soil Science. Springer. 17

Walentowski, H. and Bergmeier, E. (2009). "Wald im Klimawandel" - Aktivitäten der Arbeitsgemeinschaft Forstliche Standorts- und Vegetationskunde (AFSV). *Forst und Holz*, 64:38. 25

Walentowski, H., Gulder, H.-J., Kölling, C., Ewald, J., and Türk, W., editors (2001). *Die regionale natürliche Waldzusammensetzung Bayerns*. LWF Berichte 32. 12, 13, 135

Watson, R. (1917). Site determination, classification, and application. *Journal of Forestry*, 15(5):552–563. 4

Webster, R. and Oliver, M. (2007). *Geostatistics for Environmental Scientists*. Wiley, West Sussex. 18

Wertz, W. and Arnold, J. (1975). Land stratification for land-use planning. In Bernier, B. and Winget, C., editors, *Forest Soils and Forest Land Management*, pages 617–629. Laval Univ. Press, Quebec. 5

Wiesmeier, M., Barthold, F., Blank, B., and Koegel-Knabner, I. (2011). Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. *Plant and Soil*, 340(1-2, SI):7–24. 38

Wijewardana, D. (2008). Criteria and indicators for sustainable forest management: The road travelled and the way ahead. *Ecological Indicators*, 8(2):115–122. 4

Wu, J. and David, J. L. (2002). A spatially explicit hierarchical approach to modeling complex ecological systems: theory and applications. *Ecological Modelling*, 153(1-2):7–26. 54

Yaloon, D. (1975). Conceptual models in pedogenesis: can soil forming functions be solved? *Geoderma*, 14:189–205. 19

Zirlewagen, D. and Wilpert, K. v. (2011). Regionalisierung bodenphysikalischer Eingangsgrössen für bodenhydraulische Pedotransferfunktionen. *Waldökologie, Landschaftsforschung und Naturschutz*, 12:85–91. 25, 58

Appendix A.

Full papers

A.1. Publication I

Häring, T., Dietz, E., Osenstetter, S., Koschitzki, T. & Schröder, B. (2012) Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils. *Geoderma*, 185-186: 37-47.

Geoderma 185–186 (2012) 37–47

Contents lists available at SciVerse ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils

Tim Häring [a,b,*], Elke Dietz [c], Sebastian Osenstetter [c], Thomas Koschitzki [d], Boris Schröder [b,e]

[a] BASF SE, Environmental Fate-Modelling, APD/EF, Speyerer Str. 2, 67117 Limburgerhof, Germany
[b] Landscape Ecology, Technische Universität München, Emil-Ramann-Strasse 6, 85350 Freising, Germany
[c] Bavarian State Institute of Forestry, Hans-Carl-von-Carlowitz-Platz 1, 85354 Freising, Germany
[d] Geoflux GbR, Schleiermacherstraße 15, 06114 Halle (Saale), Germany
[e] University of Potsdam, Karl-Liebknecht-Str. 24–25, 14476 Potsdam, Germany

ARTICLE INFO

Article history:
Received 27 April 2011
Received in revised form 8 March 2012
Accepted 7 April 2012
Available online xxxx

ABSTRACT

Detailed knowledge on the spatial distribution of soils is crucial for environmental monitoring, management, and modeling. However soil maps with a finite number of discrete soil map units are often the only available information about soils. Depending on the map scale or the detailing of the map legend this information could be too imprecise. We present a method for the spatial disaggregation of map units, namely the refinement of complex soil map units in which two or more soil types are aggregated. Our aim is to draw new boundaries inside the map polygons to represent a single soil type and no longer a mixture of several soil types. The basic idea for our method is the functional relationship between soil types and topographic position as formulated in the concept of the catena. We use a comprehensive soil profile database and topographic attributes derived from a 10 m digital elevation model as input data for the classification of soil types with random forest models. We grouped all complex map units which have the same combination of soil types. Each group of map units is modeled separately. For prediction of the soil types we stratified the soil map into these groups and apply a specific random forest model only to the associated map units. In order to get reliable results we define a threshold for the predicted probabilities at 0.7 to assign a specific soil type. In areas where the probability is below 0.7 for every possible soil type we assign a new class "indifferent" because the model only makes unspecific classification there. Our results show a significant spatial refinement of the original soil polygons. Validation of our predictions was estimated on 1812 independent soil profiles which were collected subsequent to prediction in the field. Field validation gave an overall accuracy of 70%. Map units, in which shallow soils were grouped together with deep soils could be separated best. Also histosols could be predicted successful. Highest error rate were found in map units, in which Gleysoils were grouped together with deep soils or Anthrosols. To check for validity of our results we open the black box random forest model by calculating the variable importance for each predictor variable and plotting response surfaces. We found good confirmations of our hypotheses, that topography has a significant influence on the spatial arrangement of soil types and that these relationships can be used for disaggregation.
© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Knowledge on the spatial distribution of soils and soil attributes is crucial for many tasks in environmental management, monitoring, and modeling. In forestry, for instance, high-resolution spatial information on soils is required in order to conduct sustainable management of forests which concerns the site-specific environmental conditions in relation to species-specific requirements (e.g. Falk and Mellert, 2011; Thwaites and Slater, 2000). Up to now, soil maps representing categorical soil units in finite number of map entities were the most common source of spatial soil information in environmental authorities (Hartemink et al., 2010). In order to allot soil properties to the soil map, it is common to assign several representative soil profile data to the different map units (Ad-hoc-AG Boden, 2005; Legros, 2006; Soil Atlas of Europe, 2005). Soil properties can then be derived from soil polygon maps by calculating area-weighted or non-weighted averages across the different soil profiles in each map unit.

Discretizing soils into several soil units is a challenge for the mapper, since the spatial distribution of soils and their associated properties can change significantly within short distances and due to the

* Corresponding author at: Landscape Ecology, Technische Universität München, Emil-Ramann-Strasse 6, 85350 Freising, Germany. Tel.: +49 621 60 27587; fax: +49 621 60 27945.
E-mail addresses: tim.haering@basf.com (T. Häring), elke.dietz@lwf.bayern.de (E. Dietz), sebastian.osenstetter@lwf.bayern.de (S. Osenstetter), info@geoflux.de (T. Koschitzki), boris.schroeder@tum.de (B. Schröder).

0016-7061/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.geoderma.2012.04.001

continuous nature of soil (Heuvelink and Webster, 2001; Webster and Becket, 1968). In nature soils do not occur as discrete bodies with sharp boundaries (Odgers et al., 2011a). Therefore, mapping soils as categorical map units can be criticized from an ontological point of view because it contradicts this situation (e.g. Burrough and Frank, 1995). Nevertheless, it has been proven to be practical and effective, because it allows for structuring our knowledge by classification (Legros, 2006; Webster and Becket, 1968). A common approach in soil mapping is the aggregation of several soil types with different soil properties into one single complex soil map unit depending on the specific mapping scale on the one hand and the small scale heterogeneity of soils across the landscape on the other (Ad-hoc-AG Boden, 2005; IUSS Working Group WRB, 2010; Soil Atlas of Europe, 2005).

The aim of this study is to present an approach for disaggregating complex soil map units. Even though the construction of complex map units is comprehensible from a mapper's point of view, it poses a question regarding site-specific forest management or land evaluation where aggregated soil map units may cause problems. To assign soil physical or soil chemical properties to map units, usually several representative soil profiles have to be selected from an existing soil data base or by analyzing soil material in the laboratory obtained from a soil pit. If there are different soil types combined in one complex map unit, it may cause unrealistic results when calculating areal weighted or non-weighted means. As an example consider following map unit (in which the originally German soil types were translated into the international WRB system):

Soil complex with small scale variation of Stagnosols and Leptosols. Very to extremely blocky-stony, sandy-loamy periglacial detritus of amphibolites, diorites, and gabbros.

Stagnosols are characterized by periodically stagnating surface water leading to mottled color pattern or bleaching due to anaerobic conditions. They develop on a wide variety of unconsolidated materials and can be found in flat or gently sloping terrain positions (IUSS Working Group WRB, 2007). In contrast, Leptosols are very shallow soils and extremely gravelly and/or stony. They can be found on exposed landscape positions with strongly dissected topography (IUSS Working Group WRB, 2007). Clearly, the properties of those two soil types are very different, e.g. with respect to their suitability for planting tree species or to their vulnerability regarding windfall. A typical Stagnosol belonging to this map unit has an available water capacity (AWC) of 167 mm/m^2, an impermeable layer at a depth of 35 cm and a fraction of coarse fragments of 21%. In contrast, a typical Leptosol in the same map unit has an AWC of 45 mm/m^2, no impermeable layer and a fraction of coarse fragments of 76%. Clearly this causes problems in calculating the mean values for these attributes from e.g. ten Stagnosols and seven Leptosols, as the resulting calculated mean will express neither the characteristic properties of Stagnosols nor Leptosols correctly.

Scale issues in soil science either with respect to transforming soil information to finer scales (downscaling, disaggregation) or to coarser scales (upscaling, aggregation) have been addressed in literature (Carre et al., 2008; Heuvelink and Pebesma, 1999; Odgers et al., 2011a, 2011b; Panagos et al., 2011). According to Cheng (2008), downscaling is the process of estimating values for smaller scales without observation of the values available in surrounding locations. Soil distribution at one scale is therefore used to estimate the distribution at another scale.

In order to transform soil information to another scale, spatial prediction techniques have been applied in digital soil mapping literature (Grunwald, 2010). Beside some theoretical considerations for aggregation and disaggregation of soil information, McBratney (1998) proposed three approaches for disaggregation of polygon soil maps: transfer functions, fractal analysis, and pycnophylactic splines. De Bruin et al. (1999) used stepwise image interpretation and inductive learning to formalize soil–landscape relationships. Terrain objects, which were delineated from aerial photographs, were connected with location-specific soil sample data. Bui and Moran (2001) apply decision trees for disaggregation and extrapolation of fluvial facies units into unmapped areas.

In areas where no soil profiles were available and no detailed information on where in the landscape a specific soil type of a complex map unit is located, several studies proposed clustering methods for spatial predictions. Bui and Moran (2001) use k-means clustering to classify soils with Landsat MSS bands, slope position and relief as predictor variables. Yang et al. (2011) used fuzzy clustering to quantify soil–landscape relationships on a 1:20.000 soil map in Canada. The extracted knowledge was used for refined soil mapping using the Soil Land Inference Model SoLIM. Similarly, Smith et al. (2010) disaggregated soil maps in the Canadian province of British Columbia using terrain attributes, landform classes, and ecological subzones as predictor variables for fuzzy classification rules.

However, in cases where representative soil profiles as training data were available, supervised classification is an alternative method for spatial prediction. The benefit from supervised classification is its ability to estimate prediction accuracy and the identification of clearly described map units or subunits. Thereby it is possible to follow the traditional top-down approach in soil mapping to divide an existing map unit in more homogeneous sub-units and leave the former boundaries of map polygons unchanged. This is in contrast to the aforementioned studies which result in completely disaggregated soil maps.

However, there are situations in which dissolving is not intended. Kempen et al. (2009) presented an approach to update the existing 1:50.000 Dutch soil map. This was motivated due to an area-wide transformation of peat soils to other soil types.

Similar to Kempen et al. (2009), we do not alter the boundaries of soil polygons in our study. Even though polygons of soil maps cannot be viewed as 100% correct, soil maps serve as a basis for several applications. We aim to disaggregate not the entire mapped area, but only complex map units. Therefore, the existing methods are not useful for our purpose.

Moreover, regarding the number of different classes it is unfeasible to estimate one model for our entire study area. If the number of classes becomes very high—as in our case 104 groups of map units—one model needs to be either very complex or it is not able to discriminate between all single classes (cf. Bui and Moran, 2001; Kempen et al., 2009). Therefore, we developed an approach applying comparably simple but class-specific models for the delineation of sub-areas. Almost all studies in which categorical map units were disaggregated considered a smaller number of classes (i.e. less than 10 classes: Behrens et al., 2010; Brus et al., 2008; Sun et al., 2011; 10 to 20 classes: Debella-Gilo and Etzelmüller, 2009; Hengl et al., 2007; Kempen et al., 2009; Moonjun et al., 2010; more than 20 classes: Grinand et al., 2008; Smith et al., 2010; Stum et al., 2010). Only Smith et al. (2010) predicted more than 100 classes, however not in a single model but with knowledge-based fuzzy classification rules for every class separately.

Because many map units do not occur on the entire map but only on small subareas a stratification of the study-area to get parsimonious models is also favorable in our case.

In Bavaria, traditional soil maps were the main source of soil information. Even though 9924 soil profiles are available in Bavaria for modeling purposes, it was not possible to generate soil maps using classical spatial interpolation techniques. In many cases, several sampling points were located on representative landscape positions within few hundreds of meters on a catena. Therefore, we encountered a high density of profiles in some areas, whereas in others the sampling is rather sparse. Moreover, soil properties data are available only for a small subset of the sample (e.g. soil chemical properties are available for only 11% of the profiles). For the majority of samples, we only got information on the soil type, coordinates, succession of soil horizons, and texture. To generate high-resolution soil data we needed to strike a new path.

In this study, we present a method to disaggregate soil map units, especially suitable for complex map units in which two or more different soil types were combined. Thereby we follow the traditional top-down approach in soil mapping to divide an existing map unit in more homogeneous sub-units. We use decision tree-based models to quantify the relationship between soil types and topography and use these models to predict the single soil types within the complex map unit.

2. Materials and methods

2.1. Study area

The study area is the forest area of the German federal state Bavaria in the south-east of Germany which is covered by the 1:25.000 soil map (Figure 1). Bavaria has an area of ca. 70 550 km^2 and is characterized by diverse physical-geographic conditions. It measures approximately 366 km in N–S direction and 352 km in E–W direction and has a long altitudinal gradient from Kehl am Main (102 m asl) to Germany's highest mountain in the south (Zugspitze 2962 m asl). The climate is cool humid with a mean annual temperature decreasing from 10.3 °C at lower elevations to −4 °C at the summits and annual precipitation ranging from 483 mm up to 2800 mm. Due to a high geological diversity (from crystalline basement rocks, volcanic rocks, different triassic sedimentary rock, large areas of limestones, tertiary molasse to quaternary fluvial, glacial, and aeolian deposits), Bavaria is also characterized by a rich mosaic of soil types.

2.2. Soil profile database

We established a soil profile database as our main information about soil types for statistical modeling. Our aim was to merge all soil profiles within forests which were available in the forest and environmental administration in the state of Bavaria, inside as well as outside of the mapped area (Figure 2). Since our focus lies on the forest area, we only took soil profiles into account which were located in forests.

We ended up with 9924 soil profiles consisting of 93 different soil types according to the German soil classification system. When using the terms *soil type* and *complex soil map unit*, we refer to the German soil classification system (Ad-hoc-AG Boden, 2005). A soil type is characterized by a specific sequence of soil horizons influenced by soil forming processes. Every soil map unit contains information about the soil type and its parent material. A soil type in the German soil classification system is similar to the reference soil groups in the World Reference Base for Soil Resources (IUSS Working Group WRB, 2007). In cases in which there are two or more soil types associated into one single map unit, we use the term *complex soil map unit*. Within the WRB system, there is a similar approach for the description of complex units (IUSS Working Group WRB, 2010). Throughout this paper, the German soil types were translated into WRB names. All profiles were attributed with geographic coordinates to join the points with topographic observations.

2.3. Soil map

We used the official soil map of Bavaria (Bavarian Environment Agency, *Übersichtsbodenkarte ÜBK25*, http://www.lfu.bayern.de). It is

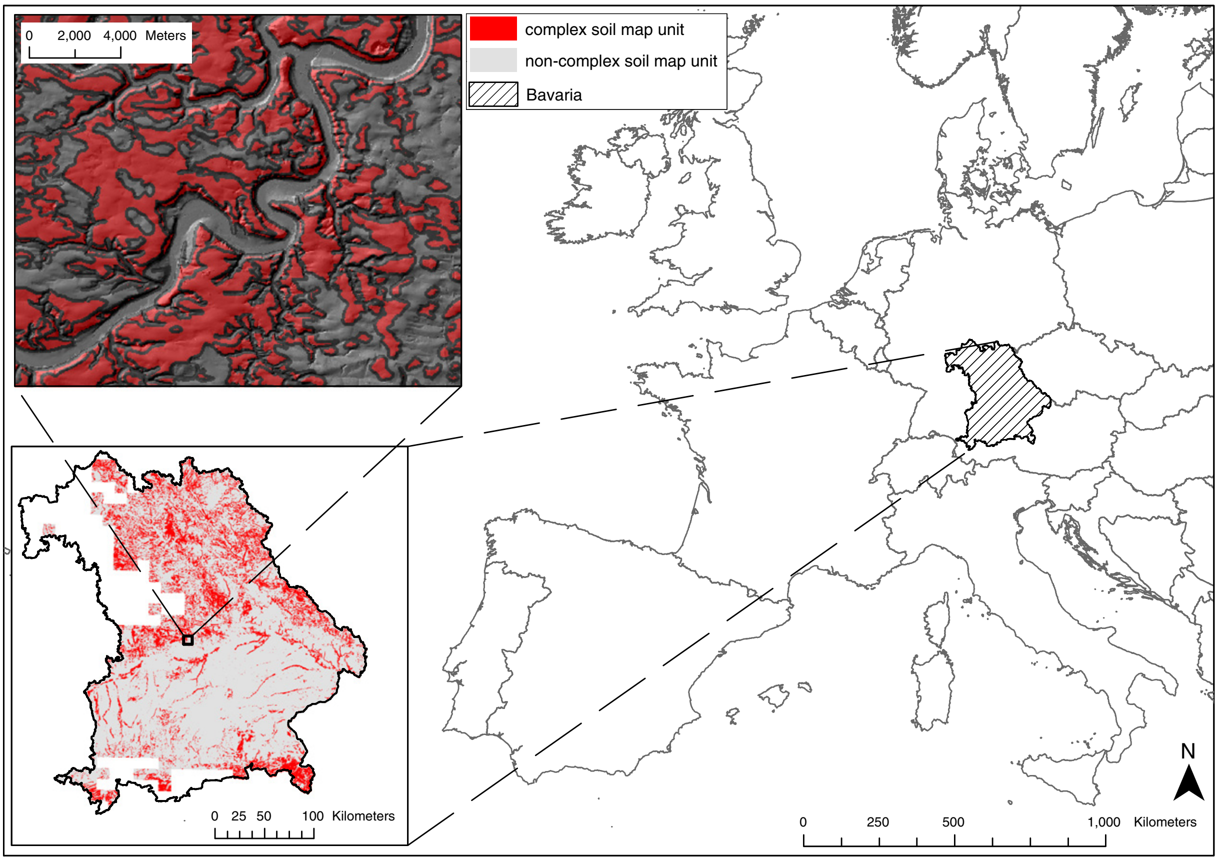


Fig. 1. Complex and non-complex soil map units of the study-area and location in Germany and Europe.

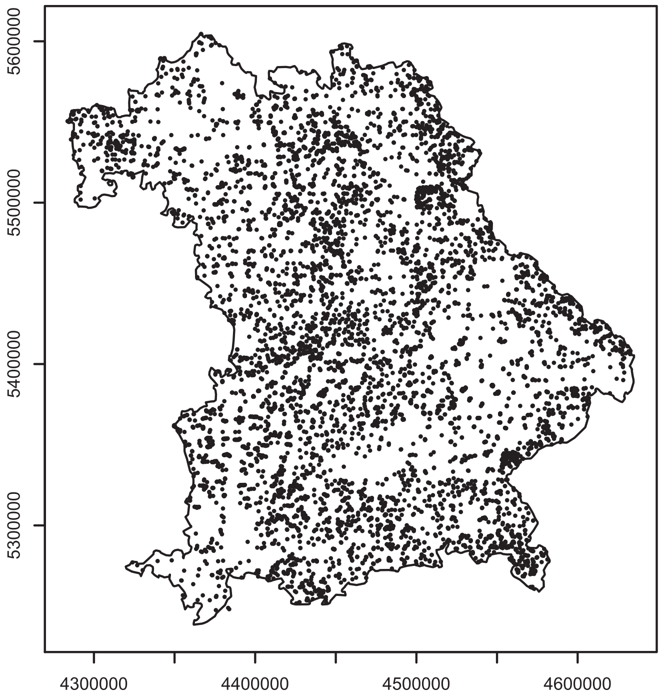


Fig. 2. Distribution of 9924 soil profiles located in forested areas.

mapped with a scale of 1:25.000 and covers 64 071 km^2 which is about 80% of the Bavarian State territory (south-east Germany, Figure 1). The map contains more than 700 different map units, i.e. a clearly defined entity in the map legend, from which more than 450 are complex map units containing more than one single soil type. It accounts for the official German guideline for soil mapping (Ad-hoc-AG Boden, 2005). Soil mapping in Germany follows the concept of substrate-systematic mapping, which means that every soil map unit contains information about the soil type and its parent material.

2.4. Topographic covariates

We use a digital elevation model (DEM) with a cell size of 10 m as basis for the delineation of topographic attributes. It was constructed using airborne laser-scan data by the Bavarian Topographical Survey and has a vertical accuracy of 0.3 m and a positional accuracy of approximately 1 m. Errors and anthropogenic elements like roads or settlements in the DEM were eliminated by the topographical survey before terrain attributes were derived.

We derived a set of 23 terrain attributes. Besides classical local terrain attributes calculated with a 3 × 3 moving window (e.g. slope gradient, curvatures), we derived complex secondary terrain attributes (Pike et al., 2009; Wilson and Gallant, 2000). In addition, we used different window sizes (8 × 8, 15 × 15) to analyze the effect of scales and spatial context (Grinand et al., 2008; Smith et al., 2006). To select the most important variables and to remove highly correlated variables, we applied the feature selection method ReliefF (Kira and Rendell, 1992; results not shown). ReliefF measures the usefulness of terrain attributes based on their ability to distinguish between very similar soil profiles belonging to different soil types. We found a high importance of secondary terrain attributes. Slope gradient (calculated on a 3 × 3 window) was the only important local terrain attribute. Finally, we got seven topographic attributes which were used for modeling (Table 1). All terrain parameter were calculated in SAGA-GIS (Saga Development Team, 2011).

Table 1
Terrain attributes used as topographic covariates for statistical modeling. All parameters were calculated in SAGA-GIS.

Terrain attribute	Definition
Topographical wetness index (twi)	SAGA Wetness Index, implemented in SAGA GIS (Böhner et al., 2002)
relative height (hut)	vertical distance to channel network
floodplain index (fpi)	Indicates flat areas with high flow accumulation and low relative elevation (1 + slope gradient) * (1 - twi) * (1 + relative height)
modified floodplain index (fpi2)	Indicates areas with high flow accumulation and low relative elevation (1 + relative height) * (1 – twi)
mass balance index (mbi)	Indicates areas of erosion and deposition (Möller et al., 2008) (plan-curvature + profile-curvature) * (1 + slope)
Slope gradient	according to Zevenbergen and Thorne (1987)
mid-slope position	The higher the relative vertical distance to the mid slope in valley or crest directions the higher this value. (Böhner and Antonic, 2009) |2 * normalized.height – 1|

2.5. Data preparation

First, all georeferenced soil profiles were attributed with the seven topographic attributes in order to establish a dataset for subsequent modeling. We interpolate our profiles with the terrain parameter using the bilinear interpolation method. Due to the fact that the coordinates of our soil profiles were mainly measured with GPS, we had to deal with a certain degree of spatial uncertainty, because the forest canopy blocks and reflects the satellite signal causing multipath effects and signal losses that lower the accuracy (Mauro et al., 2011). Smoothing values by interpolation attenuates this problem.

Secondly, we grouped the 250 complex soil map units that needed to be disaggregated, according to their soil types. In cases in which two or more map units had the same combination of soil types, they were grouped together, e.g. Calcaric Cambisols and Umbric Leptosols developed on dolomite on the one hand and on limestone on the other. Finally, we got 104 different groups of map units with the same combination of soil types. 89 groups consist of only two different soil types, the remaining 15 groups consist of three different soil types. The number of profiles which were assigned to the 104 groups ranged from 35 to 2668 (median = 339.0, mean = 638.4).

2.6. Soil landscape relationship

Topography is one of the elementary soil forming factors. The influence of relief on the spatial distribution of soils especially on field to landscape scale was first formulated in the catena concept (Milne, 1935). Numerous studies in soil landscape modeling and digital soil mapping used topographic attributes as spatial covariates (see for an overview Behrens et al., 2010; Deumlich et al., 2010; McBratney et al., 2003; Möller et al., 2008). Our method relies on these relationships. We hypothesize that it is possible to separate different soil types within a complex map unit by quantifying the functional relationship between soil type and topography by means of statistical modeling. We expect that we can derive a specific topographic fingerprint for each soil type by investigating the distributions of several topographic attributes for different soil types respectively. If the fingerprint of a specific soil type is different from that of an accompanying soil type, we are able to draw new boundaries inside a soil map polygon.

Fig. 3 shows boxplots of the aforementioned Stagnosols and Leptosols for three topographic attributes.

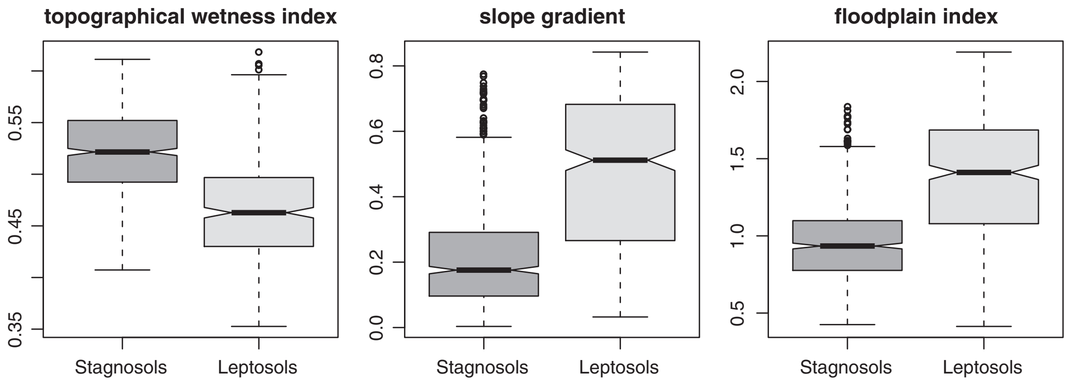


Fig. 3. Boxplots of topographical wetness index, slope gradient, and floodplain index for Stagnosols and Leptosols of our soil profile database. The values for the topographic attributes were transformed. The soil types show a significant difference regarding topographical gradients. We use these differences to classify the soil types with random forest.

The plots show significant differences in the appearance of these soil types on topographic gradients. These differences can be used for classification by decision tree-based classification models.

There are several complex map units in our soil map in which other parameters than topography might have a significant influence on the spatial arrangement of and the differentiation between different soil types like geological, chemical, or hydrological conditions. For example a map unit containing different cambisols developed on tertiary marl or sandstones or a map unit containing calcaric and dystric Gleysols is much more influenced by geology than by topography. Different geologic conditions may also be reflected by topography, but not at all times. In order to produce meaningful results, we selected only those complex map units that can be separated with topographical information according to our expert-knowledge. From the 450 existing complex map units, we selected 250 units for spatial disaggregation with our approach presented here. The area of these 250 complex units is 14776 km^2, which is 23% of the entire soil map. Reduced to the forested area, these 250 map units cover 6150 km^2 (i.e. 30% of the soil map under forest).

2.7. Statistical modeling

Classification of our soil types was performed with random forest (Breiman, 2001). Random forest is an ensemble method in which many different classification trees are combined to produce a more stable and accurate classification compared to a single decision tree (Bauer and Kohavi, 1999; Breiman, 1996; Dietterich, 2000). Each tree is built on a bootstrap sample of the given data. To form the ensemble, the different trees are combined using bagging (bootstrap aggregating). The resulting "forest" is a "random" forest because at each split only a random subset of the candidate predictors is considered for the binary partition (Elith and Graham, 2009). This de-correlates the trees, improves the variance reduction and finally leads to more accurate predictions (Bühlmann and Yu, 2002; Strobl et al., 2009). The predictions of each single tree are combined using a majority vote to get a final ensemble prediction. In recent years, random forests have been widely used in digital soil mapping (e.g. Roecker et al., 2010; Stum et al., 2010).

Widely used decision trees like Breiman et al.'s CART (1984) or Quinlan's C5 (http://www.rulequest.com/, 1993) were built on recursive partitioning and impurity reduction. Entropy measures, like the Gini Index or the Shannon Index, are used to quantify the impurity in each node (Hastie et al., 2009; Strobl et al., 2009). When working with environmental data and in particular topographical data as covariates for statistical modeling, we always have to concern multicollinearity (Graham, 2003; Hengl and MacMillan, 2009). Strobl et al. (2007, 2008) showed that CART-based random forest implementations (like the R-package randomForest, Liaw and Wiener, 2002) are biased when predictor variables are correlated or measured on different scales. Therefore, we used a random forest implementation applying conditional inference trees as base learners, which has been proven to be unbiased (Hothorn et al., 2006; Strobl et al., 2010). The splitting in recursive partitioning in conditional inference trees is based on significance tests of independence between any of the predictors and the response. Such a framework is implemented with the functions ctree() and cforest() in the package party in R (Hothorn et al., 2006; Müller et al., 2009; R Development Core Team, 2010).

Typically, predictions of classification models like random forest are response classes. The predictions are made on a majority vote using the predicted probabilities for the present soil types, i.e. the class with the highest probability is assigned (Strobl et al., 2009). We do not use the predicted classes in our study, but an estimate of the conditional class probabilities. We defined a probability threshold at $P > 0.7$ to allow for unspecified classifications in the model assigning a specific soil type in the prediction only if its probability exceeds 0.7. All areas with a maximum probability for any soil type below 0.7 were classified as "indifferent".

Validation of the random forest models was performed using the out-of-bag error. The predictive performance of the model is calculated on those observations which were not included in the learning sample for a specific decision tree, i.e. those observations which were not part of the bootstrap sample of the original data set. Using those out-of-bag observations, we have independent test samples for computing the prediction accuracy. It could be shown that the out-of-bag error is a conservative estimate (Strobl et al., 2009).

In order to detect the dependencies between predictor and dependent variables and to select the relevant predictors, one can calculate variable importance measures. The extraction of important predictors is calculated on the permutation accuracy importance measure. This measure is estimated by randomly permuting the values of a particular variable. By comparing the prediction accuracy before and after permuting a variable we get a measure of variable importance. For plausibility check in this study we use the permutation importance in the party package because it is a reliable measure even in cases with correlated predictors (Strobl et al., 2010).

Since there are some soil types in our profile database which were very frequent (such as Cambisols, Gleysols, or Luvisols) and others that are rather scarce (e.g. Histosols or soil types with stagnic properties), we often had the problem of extremely unbalanced datasets for modeling. So at least one of the soil types constituted only a very small minority of the data which may cause a limited classification performance (Japkowicz and Stephen, 2002). Therefore, we implemented an if-then condition in our modeling framework: if the number of observations for one soil type in our database is greater than a proportion of 2:1 to the number of profiles of the other soil type in the grouped map unit, then we take a random sample of the former to enforce a proportion of 2:1. The proportion of 2:1 is a compromise between having a more balanced dataset on the one hand and using

Table 2
Classification accuracy for unbalanced and balanced dataset.

			Full data					Reduced data				
Map unit	ST1[1]	ST2	n ST1[2]	n ST2	Acc[3]	TP1[4]	TP2	n ST1	n ST2	Acc	TP1	TP2
70	BB[5]	RQ[6]	2000	35	0.998	1	0	75	35	0.7	0.9	0.26
3	BB	BB/CF[7]	2000	155	0.93	0.99	0.03	310	155	0.77	0.85	0.61
92	SS[8]	GG[9]	741	113	0.9	0.97	0.41	226	113	0.83	0.93	0.62
103	SS	FF[10]	741	85	0.95	0.98	0.67	170	85	0.91	0.94	0.85

[1]soil type1, [2] number of observations for soil type 1, [3] Accuracy (correctly classified instances), [4] TP1 = true positive (fraction of soil type 1, that is actually classified as soil type 1), [5] Cambisol, [6] Anthrosol, [7] Calcaric Cambisol, [8] Stagnosol, [9] Gleysol, [10] Leptosols.

sufficient samples on the other (Japkowicz and Stephen, 2002). The influence of an unbalanced dataset on model performance is presented in Table 2.

Even though the overall model performance calculated on the entire data for the unbalanced dataset is better than for the reduced dataset, the problem lies in the ability of the model to discriminate between the two soil types. The evaluation of the different classes separately is performed with true positive measure (TP). TP is the fraction of a predicted class which is actually this class. Fitting a model on a highly unbalanced dataset which only predicts the over-represented class one gets a high classification accuracy but TP = 0 for the under-represented class (see map unit 70 in Table 2).

The entire modeling framework is illustrated in Fig. 4.

2.8. Field validation

In addition to the statistical model evaluation using the out-of-bag error, we estimated model performance in addition on field validation

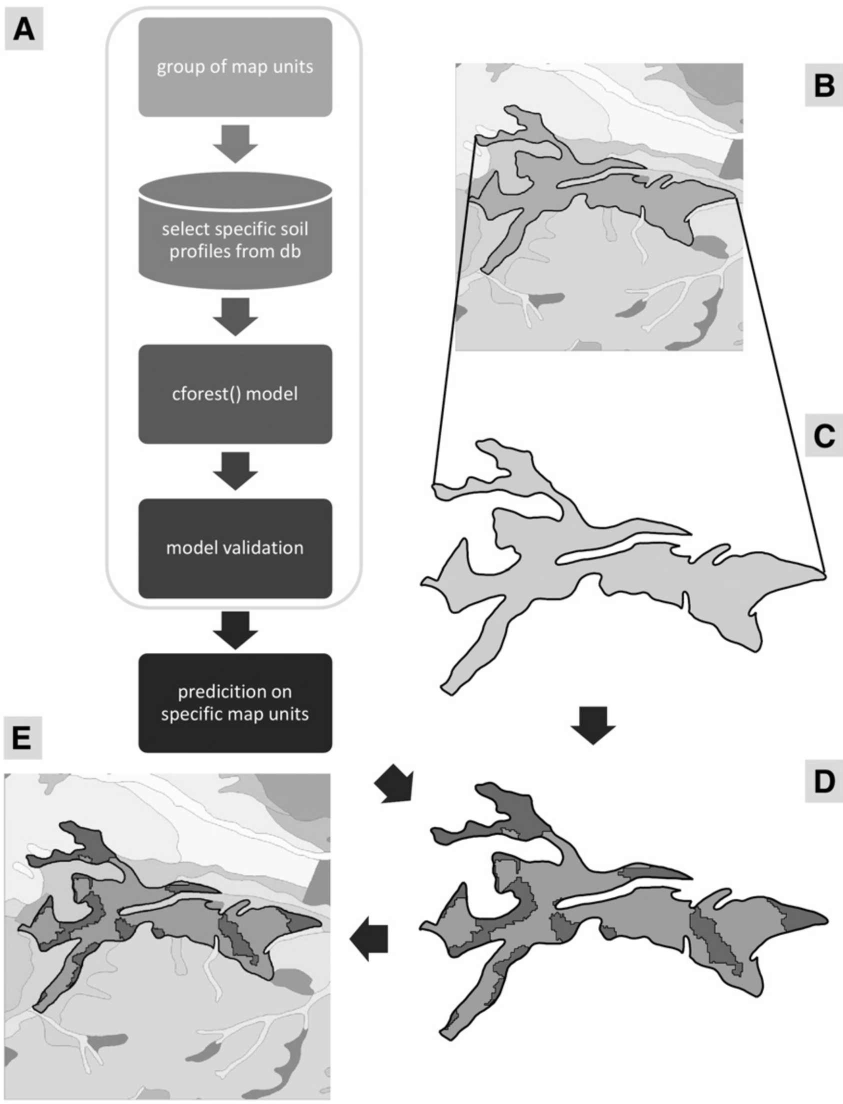


Fig. 4. Illustration of our modeling framework for spatial disaggregation of complex soil map units. We disaggregate only those map units which can be separated with terrain information. The map units were grouped according to their combination of soil types. For classification of soil types we select specific soil samples from a profile database. Statistical modeling is performed with random forest. Model validation is estimated with the out-of-bag error (A). The soil map (B) will be stratified for prediction. We predict soil types only in those areas that belong to a specific group of complex map units (C, D). To generate the final map we merge the single disaggregated parts to one soil map (E).

data as suggested by Brus et al. (2011). Subsequent to modeling, soil experts described augered soil profiles on defined locations. The distribution of sampling locations followed a stratified random sampling design. First, the study area was subdivided based on an official physiographic classification of Bavaria (Wittmann, 1991). Within these areas, sampling locations were randomly distributed over the entire predicted area. In sum we got 1820 validation points.

We grouped the validation data according to the occurrence of soil types in different map units to investigate which soil types could be separated successfully.

3. Results and discussion

3.1. Statistical Modeling

Random forest models are estimated for each of the 104 grouped map units. We calculated 500 trees in every random forest model (the "ntree" argument in cforest). Random forest models with 1000 trees did not improve the performance (results not shown). The models were stable mostly with less than 200 trees. The number of randomly selected variables as candidates at each split (the "mtry" argument in cforest) was three as recommended by Hastie et al. (2009) (mtry = square root of number of predictor variable).

The statistical validation of the models based on the out-of-bag error showed misclassification rates ranging from 0.09 to 0.55 with a median of 0.31. Compared to other digital-soil-class-mapping studies this are quite good results (cf., e.g., Hengl et al., 2007; Kempen et al., 2009; Lemercier et al., 2012; Stum et al., 2010).

3.2. Prediction

After fitting the 104 random forest models, we applied these models on the corresponding regions of the soil map in order to predict the occurrence probability of each soil type that is present in the map unit.

Fig. 5 shows six examples of the final disaggregated soil map. Finally, 57% of the area have been predicted as a specific soil type ($p > 0.7$), whereas 43% have been predicted as 'indifferent'.

3.3. External validation

Estimation of model performance on 1820 field validation points gave an overall accuracy of 70% (1246 correct classified, 540 incorrect, 8 not usable due to erroneous profile descriptions).

The predictive performance depended on the number of available profiles.

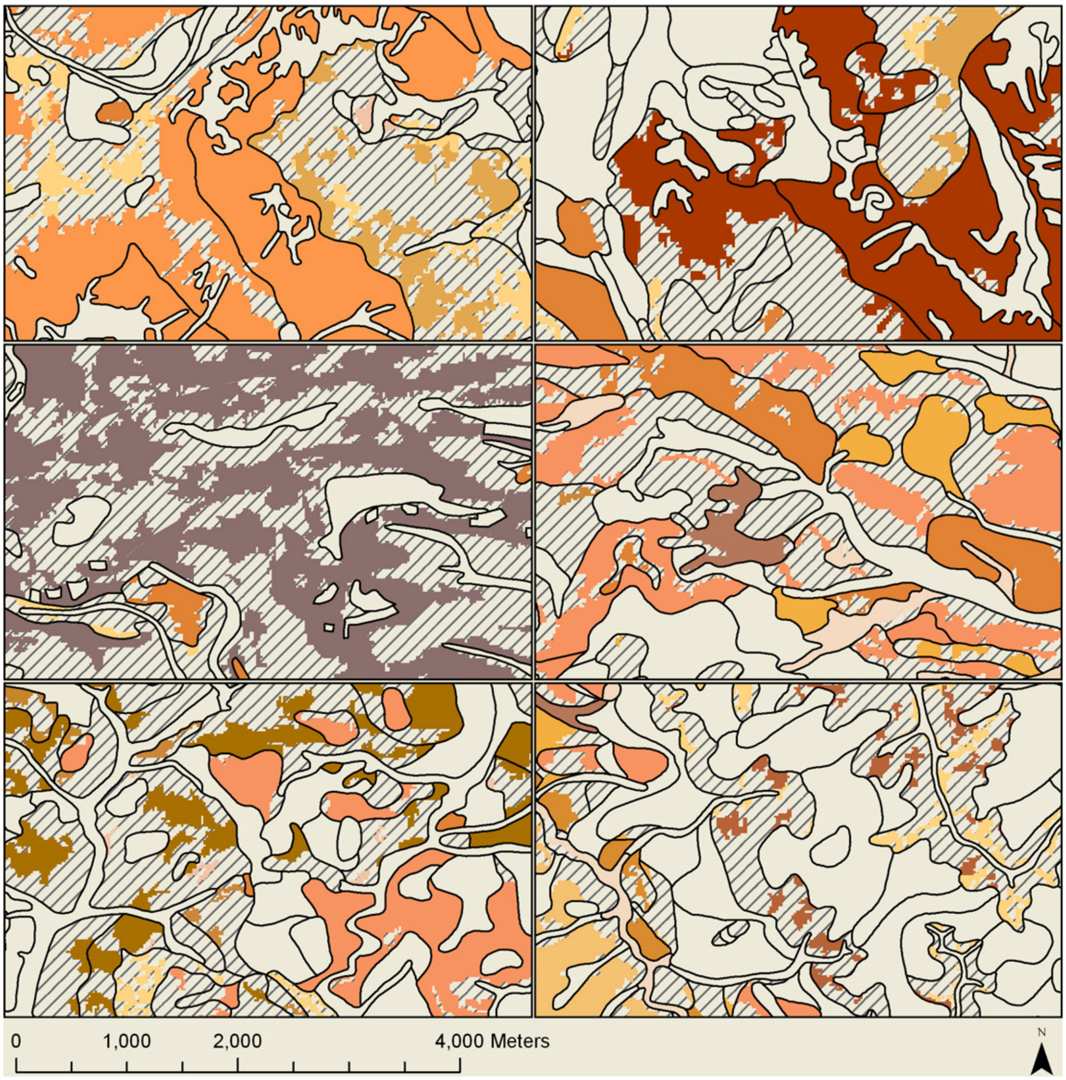


Fig. 5. Six cutouts of our resulting disaggregated soil map. The polygons of the original soil map were shown in solid black lines. Map units which were not considered during downscaling (non-complex map units) were colored in white. Areas which were classified as indifferent (predicted probability for every soil type < 0.7) were hatched.

In Fig. 6 we plot the accuracy calculated on the field validation data over the number of profiles for a given group of map units. We plot only those groups of map units which have more than 15 validation points in order to have reliable accuracy values. Accuracy increases roughly with the number of calibration data. If the number of data points exceeds 350, this correlation disappears and the accuracy values becomes more scattered. However, accuracy values always exceed 0.7. There are two outliers in the plot. One with an accuracy of 0.44 and 1184 calibration profiles which is a map unit in which Stagnosols and Cambisols with stagnic properties were grouped. And a second outlier with an accuracy of 0.5 and 447 calibration profiles, which is a map units in which Gleysols, deep soils and Stagnosols were grouped. Obviously, these two groups of soil types are too similar regarding their topographical properties that they couldn't be separated well.

These findings are very promising and confirm our approach. It seems that even better results are possible, if more profiles are available.

Fig. 7 shows a fluctuation plot, which is a graphical representation of contingency table. The extent of a graph is proportional to count. On the right hand side of the plot the numbers of correct and incorrect predicted validation points are listed.

12 out of 14 groups have more true predictions than false predictions. For these 12 groups, we can conclude that separation between groups of soil types is possible, however with different success, because prediction accuracy differs between the groups. Groups with better reliability are those in which soils are highly influenced by topographic characteristics, which was also reported by Debella-Gilo and Etzelmüller (2009). A high proportion of true predictions can be found in groups which differ in the profile depth (deep soils vs. shallow soil, initial soils vs. shallow soils) due to the strong dependency of profile depth and terrain position. Also Histosols could be predicted very successfully, since there is a strong influence of water availability on their development which is in turn mainly controlled by topography. Our results confirm findings of Seibert et al. (2007) in Swedish forest soils, who could show a strong dependency of Histosols with topography.

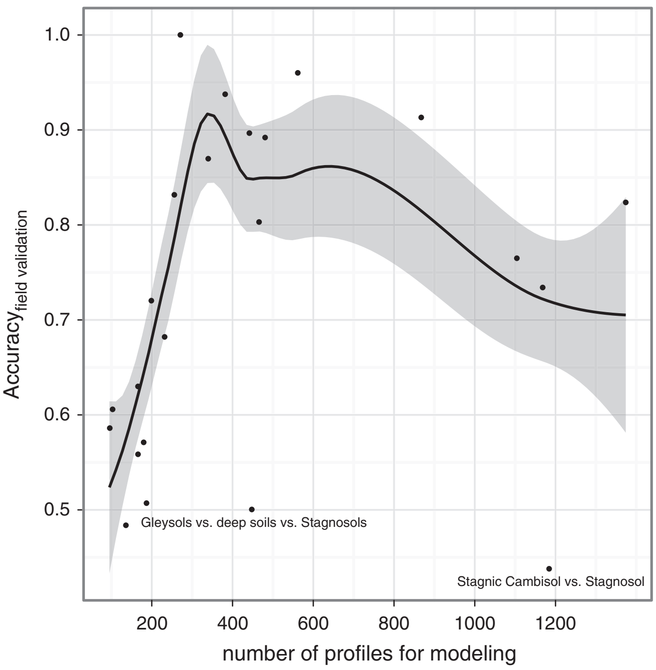


Fig. 6. Classification accuracy calculated on field-validation data plotted over the number of profiles used for training. There is a relationship between model performance and the number of profiles available for calibration. If the number of data points exceeds 350 this correlation disappears and the accuracy values becomes more scattered. However, accuracy values always exceed 0.7.

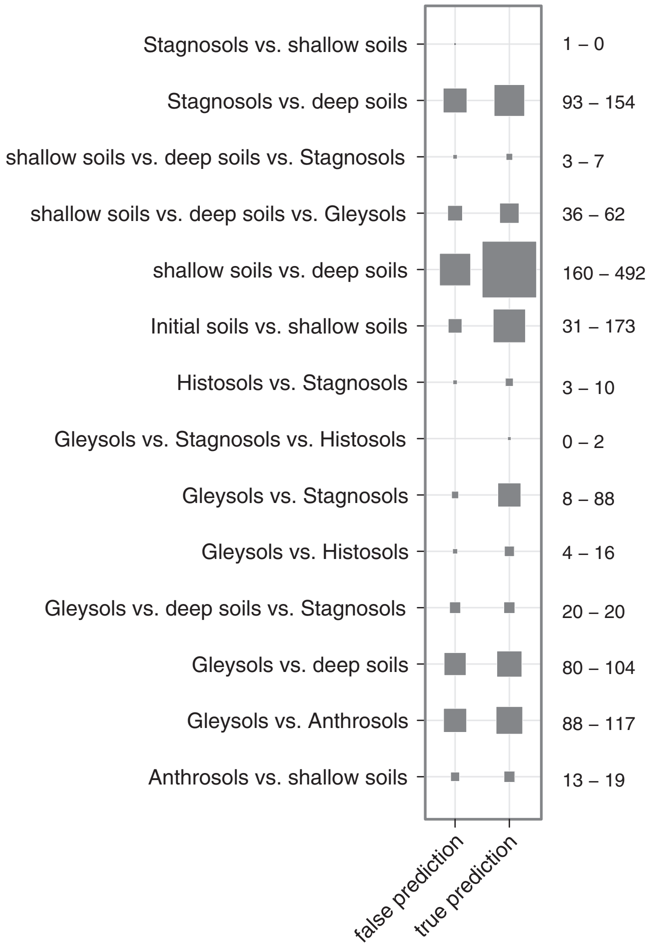


Fig. 7. Fluctuation plot of true and false predicted groups of map units. Numbers on the right hand side indicates the count of true and false. The size of the figures is proportional to the count of true and false which is displayed on the right hand side.

Except from Histosols and Stagnosols map units containing Gleysols could be separated the worst. Mainly because those map units are located in flat valley bottoms where no variability in terrain exists. Therefore, discrimination based on terrain attributes becomes extremely difficult. Similar to Histosols, the discrimination of Gleysols vs. Stagnosols is strongly influenced by the availability of groundwater and surface water respectively, which depends on topography and could therefore be executed successfully.

With the introduction of a threshold at $P > 0.7$ for the prediction of soil types we are able to generate results with high accuracy as shown with the field validation data. On the other hand, 43% of the predicted area classified as “indifferent” is not optimal for our purpose. To capture this problem we might reduce the threshold and thereby minimize the indifferent area. The threshold can be reduced until the indifferent class disappears completely and prediction is made on the highest class probability ($P > 0.5$). Prediction in such a way is done in many studies (e.g. Behrens et al., 2010; Debella-Gilo and Etzelmüller, 2009; Grinand et al., 2008). However, prediction performance has then to be estimated in more detail on additional validation data which were not available at the moment. Response surface plots (Figure 8) indicate less accurate predictions at smaller probabilities.

3.4. Plausibility check

Finally we calculated the variable importance for each predictor variable in all 104 random forest models and visualized response surfaces for every soil type of the models. This procedure provides

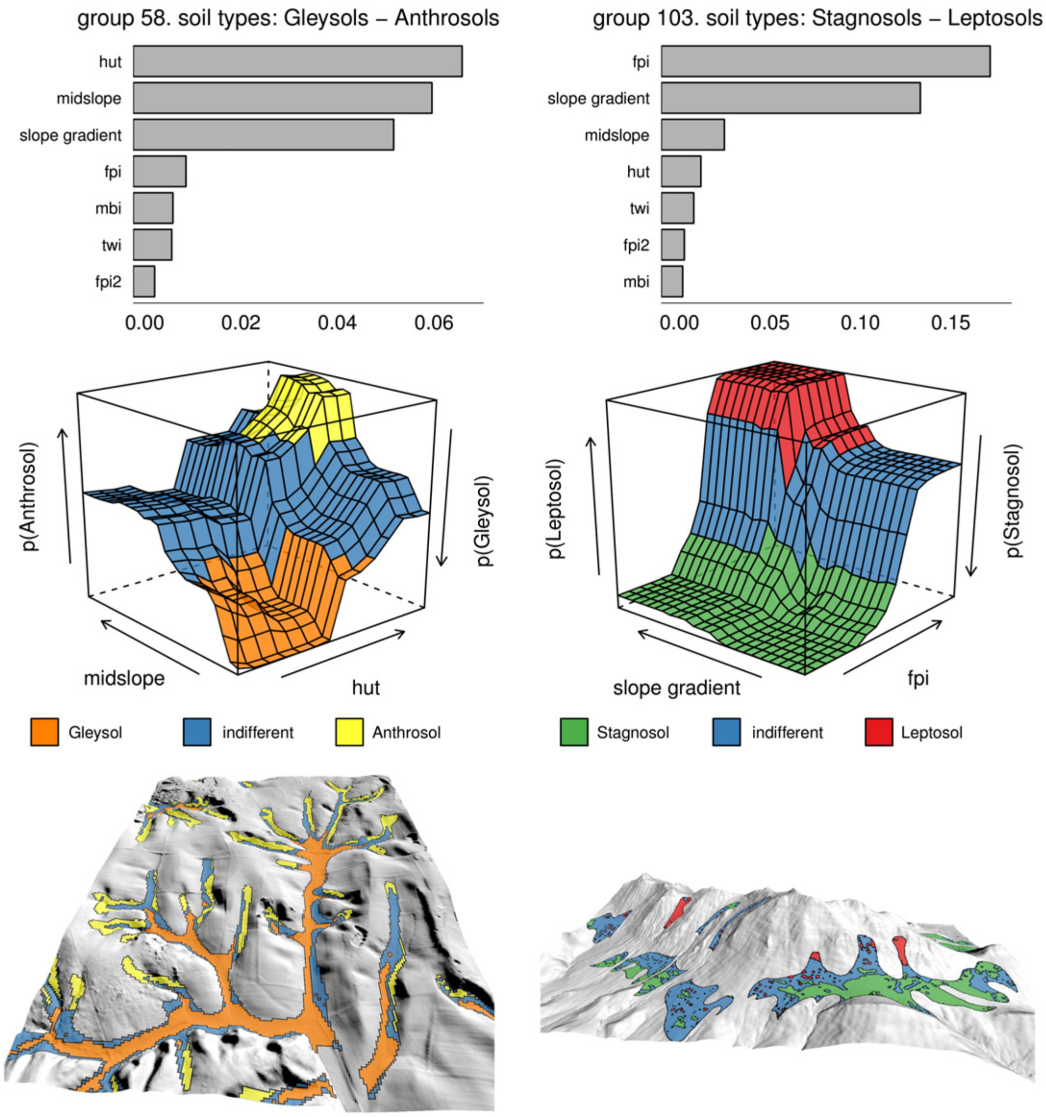


Fig. 8. Variable importance plot (top), response surfaces (middle), and exemplary cutout of the soil map for two randomly selected groups of soil units. In the response surface plots the probability of a specific soil type is plotted over the two most important predictor variables in the model. Coloring of the response surfaces indicates our probability thresholds ($P>0.7$, $0.7>P>0.3$, $P<0.3$). The area of the response surface between the thresholds indicates unspecified predictions ("indifferent"). The plots were used to validate our models. We checked if soil types are located where we expected them following our expert knowledge. The plots confirm our expectations, e.g. shallow Leptosols in exposed position where erosion occur (group 103), Anthrosols at footslopes where soil material is accumulated, and Gleysols in area with low vertical distance to channels (group 58). "fpi" = flood plain index; "fpi2" = modified flood plain index; "hut" = relative height; "twi" = topographic wetness index; "mbi" = mass balance index. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

reasonable insights on how the soil types depend on the topographic attributes, which were the relevant attributes in the model, and where specific thresholds can be found. Response surface plots are an effective tool to display model behavior and thresholds in more than one dimension (Elith and Graham, 2009; Lintz et al., 2011).

During this procedure we also checked whether our soil units were predicted in those landscape positions where expert knowledge would expect them. For example, a map unit consisting of Cambisols and Leptosols we would expect to find Leptosols with shallow soil depth on exposed positions with high slope angle and mass balance index where erosion occur. Cambisols, on the other hand, should be located in flat areas with a low mass balance index, i.e. those areas where the development of Cambisols is not disturbed by erosion processes

Fig. 8 shows the variable importance, response surfaces for the two most important variables in the model, as well as an exemplary cutout of the soil map for that particular group for two specific groups of soil units.

All plots reveal meaningful dependencies between soil types and topography. In group 103, Stagnosols and Leptosols were aggregated. The plots show a high importance of flood plain index (fpi) and slope gradient (see Figure 3). The remaining parameters have only marginal influence on the model. The response surfaces show high probability for Leptosols for high value of flood plain index and slope gradient, i.e. exposed terrain positions such as steep backslopes. Stagnosols are influenced mainly by the flood plain index. Slope gradient has no effect on Stagnosols in group 103. These dependencies can also be identified in the map cutout.

Gleysols and Anthrosols, which were aggregated in group 58, could be discriminated mostly by the vertical distance above a channel (hut), the midslope position, and slope gradient. Anthrosols can be found at footslopes where eroded material is accumulated (high midslope). On the other hand, Gleysols have their highest probability in areas with low vertical distances to channels. In the map cutout, Gleysols were located in the flat and inner areas of the original map polygon. Anthrosols could be found at the bottom of slope gradients

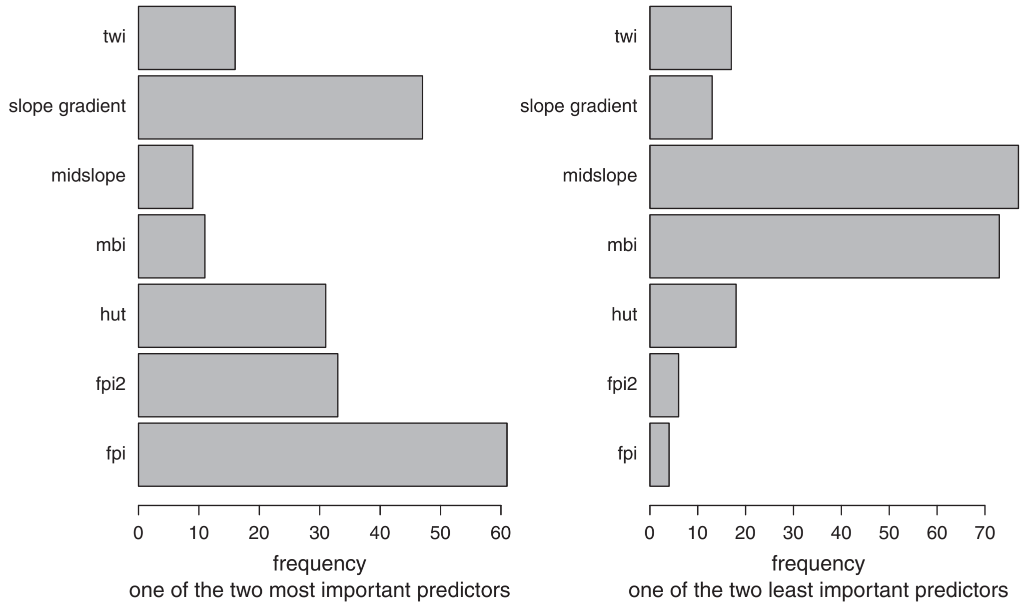


Fig. 9. The overall variable importance estimated over all 104 random forest models. We counted how often a predictor variable is one of the two most and the two least important predictors in all models.

where colluvium has been deposited. The area between these two landscape positions is classified as indifferent.

3.5. Global variable importance

Lastly we evaluated the variable importance of all predictor variables over all 104 random forests. Therefore, we counted how often a variable is one of the two most important predictors in a model and how often a variable is one of the two least important predictors. These frequencies are plotted in Fig. 9.

The plots show patterns which complements one another. Flood plain index (fpi), slope gradient, relative height (hut), and modified flood plain index (fpi2) are frequently one of the two most important predictors in a random forest. Midslope position and mass balance index (mbi) are only selected nine and eleven times respectively as one of the two most important predictors in all models. However, these two predictors are very frequently one of the two least important predictors and the remaining five are all less than 20 times part of this group. We found no preference of using either local parameter (slope gradient, mass balance index) or regional parameters (all the remaining) as important variable. This suggests that both small scale variations as well as landscape scale patterns provide important information in our approach.

4. Conclusions

High-resolution spatial information of soils and soil properties are essential for many application areas in environmental sciences. Soil maps provide the main information on soils. We demonstrate a method for the spatial disaggregation of existing soil maps for providing soil information on higher resolution. Our focus lies on soil map units in which two or more different soil types were aggregated into one map unit.

We found a significant influence of topography on the spatial arrangement of soil types. By comparing different soil types we found a characteristic topographical fingerprint for each soil type. These topographical differences were quantified with unbiased random forest models.

Future work will focus on the selection and assignment of soil profile data with representative soil physical and soil chemical properties to the disaggregated and thereby newly generated map units. Soil property maps will be generated by calculating mean values for each map unit. In those areas, in which the models predict the new “indifferent class”, a mean value of the former entire map units is assigned, as it was before disaggregation.

Acknowledgements

The authors acknowledge funding by the Bavarian State Ministry of Nutrition, Agriculture, and Forestry. This work is part of project “KLIP4—Maps for the future” at the Bavarian State Institute of Forestry. We thank the Bavarian Environment Agency and the Bavarian Topographical Survey for providing the data. We also thank the editor and two anonymous reviewers for providing helpful comments. BS acknowledges funding by the German Science Foundation (grant no. SCHR1000/3-1, SCHR1000/6-1).

References

Ad-hoc-AG Boden, 2005. Bodenkundliche Kartieranleitung KA 5. E. Schweizerbart'sche Verlagsbuchhandlung, Stuttgart.
Bauer, E., Kohavi, R., 1999. An empirical comparison of voting classification algorithms: bagging, boosting, and variants. Machine Learning 36 (1–2), 105–139.
Behrens, T., Zhu, A.X., Schmidt, K., Scholten, T., 2010. Multi-scale digital terrain analysis and feature selection for digital soil mapping. Geoderma 155 (3–4), 175–185.
Böhner, J., Antonic, O., 2009. Land-surface parameters specific to topo-climatology. In: Hengl, T., Reuter, H.I. (Eds.), Geomorphometry. Concepts, Software, Applications. Developments in Soil Science, vol. 33. Elsevier, pp. 195–226.
Böhner, J., Köthe, R., Conrad, O., Ringeler, A., Selige, T., 2002. Soil regionalisation by means of terrain analysis and process parameterisation. In: Micheli, E., Nachtergaele, F., Montanarella, L. (Eds.), Soil Classification 2001: EUR 20398 EN. European Soil Bureau, Research Report No. 7, Luxembourg, pp. 213–222.
Breiman, L., 1996. Bagging predictors. Machine Learning 24 (2), 123–140.
Breiman, L., 2001. Random forests. Machine Learning 45 (1), 5–32.
Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth, Montery, CA.
Brus, D.J., Bogaert, P., Heuvelink, G.B.M., 2008. Bayesian Maximum Entropy prediction of soil categories using a traditional soil map as soft information. European Journal of Soil Science 59, 166–177.
Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. European Journal of Soil Science 62, 394–407.
Bühlmann, P., Yu, B., 2002. Analyzing Bagging. The Annals of Statistics 30 (4), 927–961.
Bui, E.N., Moran, C.J., 2001. Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. Geoderma 103 (1–2), 79–94.

Burrough, P.A., Frank, A.U., 1995. Concepts and paradigms in spatial information: are current geographical information systems truly generic? International Journal of Geographical Information Systems 9 (2), 101–116.

Carre, F., Reuter, H.I., Daroussin, J., Scheurer, O., 2008. From a Large to a Small Scale Soil Map: Top-Down Against Bottom-Up Approaches. Application to the Aisne Soil Map (France). In: Hartemink, A.E., McBratney, A.B., Mendonça Santos, M.L. (Eds.), Digital Soil Mapping with Limited Data. Springer, Dordrecht, pp. 203–212.

Cheng, Q., 2008. Modeling local scaling properties for multiscale mapping. Vadose Zone Journal 7 (2), 525–532.

de Bruin, S., Wielemaker, W.G., Molenaar, M., 1999. Formalisation of soil–landscape knowledge through interactive hierarchical disaggregation. Geoderma 91 (1–2), 151–172.

Debella-Gilo, M., Etzelmüller, B., 2009. Spatial prediction of soil classes using digital terrain analysis and multinomial logistic regression modeling integrated in GIS: examples from Vestfold County, Norway. Catena 77, 8–18.

Deumlich, D., Schmidt, R., Sommer, M., 2010. A multiscale soil–landform relationship in the glacial-drift area based on digital terrain analysis and soil attributes. Journal of Plant Nutrition and Soil Science 173 (6), 843–851.

Dietterich, T.G., 2000. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. Machine Learning 40 (2), 139–157.

Elith, J., Graham, C.H., 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography 32 (1), 66–77.

Falk, W., Mellert, K., 2011. Species distribution model as a tool for forest management planning under climate change: risk evaluation of Abies alba in Bavaria. Journal of Vegetation Science 22 (4), 621–634.

Graham, M.H., 2003. Confronting multicollinearity in ecological multiple regression. Ecology 84 (11), 2809–2815.

Grinand, C., Arrouays, D., Laroche, B., Martin, M.P., 2008. Extrapolating regional soil landscapes from an existing soil map: sampling intensity, validation procedures, and integration of spatial context. Geoderma 143 (1–2), 180–190.

Grunwald, S., 2010. Current State of Digital Soil Mapping and what is next. In: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S. (Eds.), Digital Soil Mapping. Bridging Reasearch, Environmental Application, and Operation. Springer, Netherlands, pp. 3–12.

Hartemink, A.E., Hempel, J., Lagacherie, P., McBratney, A., McKenzie, N., MacMillan, R.A., Minasny, B., Montanarella, L., Mendonça Santos, M.L., Sanchez, P., Walsh, M., Zhang, G.L., 2010. GlobalSoilMap.net—a new digital soil map of the world. In: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S. (Eds.), Digital Soil Mapping. Bridging Reasearch, Environmental Application, and Operation. Springer, pp. 423–428.

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

Hengl, T., MacMillan, R.A., 2009. Geomorphometry— Key to Landscape Mapping and Modelling. In: Hengl, T., Reuter, H.I. (Eds.), Geomorphometry. Concepts, Software, Applications. Elsevier, pp. 433–460.

Hengl, T., Toomanian, N., Reuter, H.I., Malakouti, M.J., 2007. Methods to interpolate soil categorical variables from profile observations: lessons from Iran. Geoderma 140 (4), 417–427.

Heuvelink, G.B.M., Pebesma, E.J., 1999. Spatial aggregation and soil process modeling. Geoderma 89, 47–65.

Heuvelink, G.B.M., Webster, R., 2001. Modelling soil variation: past, present and future. Geoderma 100, 269–301.

Hothorn, T., Hornik, K., Zeileis, A., 2006. Unbiased recursive partitioning: a conditional inference framework. Journal of Computational and Graphical Statistics 15 (3), 651–674.

IUSS Working Group WRB, 2007. World Reference Base for Soil Resources 2006, first update 2007. World Soil Resources Reports No. 103. FAO, Rome.

IUSS Working Group WRB, 2010. Addendum to the World Reference Base for Soil Resources: guidelines for constructing small-scale map legends using the World Reference Base for Soil Resources. http://www.fao.org/fileadmin/templates/nr/images/resources/pdf_documents/WRB_Legend.pdf (last access 21.02.2011).

Japkowicz, N., Stephen, S., 2002. The class imbalance problem: a systematic study. Intelligent Data Analysis 6 (5), 429–449.

Kempen, B., Brus, D.J., Heuvelink, G.B.M., Stoorvogel, J.J., 2009. Updating the 1:50,000 Dutch soil map using legacy soil data: a multinomial logistic regression approach. Geoderma 151, 311–326.

Kira, K., Rendell, L.A., 1992. A Practical Approach to Feature Selection. In: Sleeman, D., Edwards, P. (Eds.), Proc. Intern. Conf. On Machine Learning (Aberdeen, July 1992). Morgan Kaufmann, pp. 249–256.

Legros, J.-P., 2006. Mapping of the Soil. Science Publishers, Enfield.

Lemercier, B., Lacoste, M., Loum, M., Walter, C., 2012. Extrapolation at regional scale of local soil knowledge using boosted classification trees: a two-step approach. Geoderma 171–172, 75–84.

Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. R News 2 (3), 12–22.

Lintz, H.E., McCune, B., Gray, A.N., McCulloh, K.A., 2011. Quantifying ecological thresholds from response surfaces. Ecological Modelling 222 (3), 427–436.

Mauro, F., Valbuena, R., Manzanera, J.A., Garcia-Abril, A., 2011. Influence of Global Navigation Satellite System errors in positioning inventory plots for tree-height distribution studies. Canadian Journal of Forest Research 41 (1), 11–23.

McBratney, A.B., 1998. Some considerations on methods for spatially aggregating and disaggregating soil information. Nutrient Cycling in Agroecosystems 50 (1), 51–62.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117 (1–2), 3–52.

Milne, G., 1935. Some suggested units of classification and mapping particularly for East African soils. Soil Research 4, 193–198.

Möller, M., Volk, M., Friedrich, K., Lymburner, L., 2008. Placing soil-genesis and transport processes into a landscape context: a multiscale terrain-analysis approach. Journal of Plant Nutrition and Soil Science 171 (3), 419–430.

Moonjun, R., Farshad, A., Shresha, D.P., Vaiphase, C., 2010. Artificial Neural Network and Decision Tree in Predictive Soil Mapping of Hoi Num Rin Sub-Watershed. In: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S. (Eds.), Digital Soil Mapping. Bridging Reasearch, Environmental Application, and Operation. Springer, pp. 151–163.

Müller, D., Schröder, B., Müller, J., 2009. Modelling habitat selection of the cryptic Hazel Grouse *Bonasia bonasia* in a montane forest. Journal of Ornithology 150, 719–732.

Odgers, N.P., McBratney, A.B., Minasny, B., 2011a. Bottom-up digital soil mapping. I. Soil layer classes. Geoderma 163, 38–44.

Odgers, N.P., McBratney, A.B., Minasny, B., 2011b. Bottom-up digital soil mapping. II. Soil series classes. Geoderma 163, 30–37.

Panagos, P., Van Liedekerke, M., Montanarella, L., 2011. Multi-scale European Soil Information System (MEUSIS): a multi-scale method to derive soil indicators. Computational Geosciences 15 (3), 463–475.

Pike, R.J., Evans, I.S., Hengl, T., 2009. Geomorphometry: A Brief Guide. In: Hengl, T., Reuter, H.I. (Eds.), Geomorphometry. Concepts, Software, Applications. Elsevier, pp. 3–30.

Quinlan, J.R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann, San Francisco.

R Development Core Team, 2010. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria3-900051-07-0. http://www.R-project.org.

Roecker, S.M., Howell, D.W., Haydu-Houdeshell, C.A., Blinn, C., 2010. A Qualitative Comparison of Conventional Soil Survey and Digital Soil Mapping Approaches. In: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S. (Eds.), Digital Soil Mapping. Bridging Reasearch, Environmental Application, and Operation. Springer, pp. 369–384.

Saga Development Team, 2011. System for Automated Geoscientific Analyses (SAGA GIS). http://www.saga-gis.org/ (last access 20.04.2011).

Seibert, J., Stendahl, J., Sorensen, R., 2007. Topographical influences on soil properties in boreal forests. Geoderma 141 (1–2), 139–148.

Smith, M.P., Zhu, A.X., Burt, J.E., Stiles, C., 2006. The effects of DEM resolution and neighborhood size on digital soil survey. Geoderma 137 (1–2), 58–69.

Smith, S., Bulmer, C., Flager, E., Frank, G., Filatow, D., 2010. Digital soil mapping at multiple scales in British Columbia, Canada. Program and Abstracts, 4th Global Workshop on Digital Soil Mapping, 24–26 May 2010, Rome, Italy, p. 17.

Soil Atlas of Europe, 2005. European Soil Bureau Network European Commission. 128 pp. Office for Official Publications of the European Communities, L-2995 Luxembourg. http://eusoils.jrc.ec.europa.eu/projects/soil_atlas/index.html (last access 05.09.2011).

Strobl, C., Boulesteix, A.-L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinformatics 8 (1), 25.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. BMC Bioinformatics 9 (1), 307.

Strobl, C., Malley, J., Tutz, G., 2009. An Introduction to Recursive Partitioning: Rationale, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests. Psychological Methods 14 (4), 323–348.

Strobl, C., Hothorn, T., Zeileis, A., 2010. Party on! A New, Conditional Variable Importance Measure for Random Forests Available in the party Package. The R Journal 1 (2), 14–17.

Stum, A.K., Boettinger, J.L., White, M.A., Ramsey, R.D., 2010. Random Forests Applied as a Soil Spatial Predictive Model in Arid Utah. In: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S. (Eds.), Digital Soil Mapping. Bridging Reasearch, Environmental Application, and Operation. Springer, pp. 179–190.

Sun, X.L., Zhao, Y.G., Zhang, G.L., Wu, S.C., Man, Y.B., Wong, M.H., 2011. Application of a Digital Soil Mapping Method in Producing Soil Orders on Mountain Areas of Hong Kong Based on Legacy Soil Data. Pedosphere 21 (3), 339–350.

Thwaites, R.N., Slater, B.K., 2000. Soil–landscape resource assessment for plantations - a conceptual framework towards an explicit multi-scale approach. Forest Ecology and Management 138 (1–3), 123–138.

Webster, R., Becket, P.H.T., 1968. Quality and Usefulness of Soil Maps. Nature 219, 680–681.

Wilson, J.P., Gallant, J.C., 2000. Terrain Analysis: Principles and Applications. Wiley, New York.

Wittmann, O., 1991. Standortkundliche Landschaftsgliederung von Bayern. GLA-Fachberichte, 5. Geologisches Landesamt, München.

Yang, L., Jiao, Y., Fahmy, S., Zhu, A.-X., Hann, S., Burt, J.E., Qi, F., 2011. Updating Conventional Soil Maps through Digital Soil Mapping. Soil Science Society of America Journal 75 (3), 1044–1053.

Zevenbergen, L.W., Thorne, C.R., 1987. Quantitaitive Analysis of Land Surface Topography. Earth Surface Processes and Landforms 12, 47–56.

A.2. Publication II

Häring, T., Reger, B., Ewald, J., Hothorn, T. & Schröder, B. (2013) Predicting Ellenberg's soil moisture indicator value in the Bavarian Alps using additive georegression. *Applied Vegetation Science*, 16: 110 - 121.

Applied Vegetation Science **16** (2013) 110–121

Predicting Ellenberg's soil moisture indicator value in the Bavarian Alps using additive georegression

Tim Häring, Birgit Reger, Jörg Ewald, Torsten Hothorn & Boris Schröder

Keywords
Boosting; Mboost; Non-stationarity; Predictive vegetation mapping; Site ecology; Species distribution modelling

Nomenclature
Wisskirchen & Haeupler (1998)

Received 10 October 2011
Accepted 30 April 2012
Co-ordinating Editor: Sándor Bartha

Häring, T. (Corresponding author, tim.haering@basf.com) & **Schröder, B.** (boris.schroeder@tum.de): Landscape Ecology, Technische Universität München, Emil-Ramann-Strasse 6, 85350, Freising, Germany
Häring, T.: BASF SE, Environmental Fate – Modelling, Speyerer Street 2, 67117, Limburgerhof, Germany
Reger, B. (birgit.reger@lwf.bayern.de): Bavarian State Institute of Forestry, Hans-Carl von-Carlowitz-Platz 1, 85354, Freising, Germany
Ewald, J. (joerg.ewald@hswt.de): Faculty of Forest Science and Forestry, University of Applied Sciences Weihenstephan-Triesdorf, Hans-Carl von-Carlowitz-Platz 3, 85354, Freising, Germany
Hothorn, T. (torsten.hothorn@stat.uni-muenchen.de): Institut für Statistik, Ludwig-Maximilians-Universität München, Ludwigstraße 33, 80539, München, Germany
Schröder, B.: University of Potsdam, Karl-Liebknecht-Straße 24/25, 14476, Potsdam, Germany

Abstract

Questions: Can forest site characteristics be used to predict Ellenberg indicator values for soil moisture? Which is the best averaged mean value for modelling? Does the distribution of soil moisture depend on spatial information?

Location: Bavarian Alps, Germany.

Methods: We used topographic, climatic and edaphic variables to model the mean soil moisture value as found on 1505 forest plots from the database WINALPecobase. All predictor variables were taken from area-wide geodata layers so that the model can be applied to some 250 000 ha of forest in the target region. We adopted methods developed in species distribution modelling to regionalize Ellenberg indicator values. Therefore, we use the additive georegression framework for spatial prediction of Ellenberg values with the R-library mboost, which is a feasible way to consider environmental effects, spatial autocorrelation, predictor interactions and non-stationarity simultaneously in our data. The framework is much more flexible than established statistical and machine-learning models in species distribution modelling. We estimated five different mboost models reflecting different model structures on 50 bootstrap samples in each case.

Results: Median R^2 values calculated on independent test samples ranged from 0.28 to 0.45. Our results show a significant influence of interactions and non-stationarity in addition to environmental covariates. Unweighted mean indicator values can be modelled better than abundance-weighted values, and the consideration of bryophytes did not improve model performance. Partial response curves indicate meaningful dependencies between moisture indicator values and environmental covariates. However, mean indicator values <4.5 and >6.0 could not be modelled correctly, since they were poorly represented in our calibration sample. The final map represents high-resolution information of site hydrological conditions.

Conclusions: Indicator values offer an effect-oriented alternative to physically-based hydrological models to predict water-related site conditions, even at landscape scale. The presented approach is applicable to all kinds of Ellenberg indicator values. Therefore, it is a significant step towards a new generation of models of forest site types and potential natural vegetation.

Introduction

Soil hydrological conditions represent an essential ecological gradient controlling plant species composition and distribution. A deficit in soil moisture is often the most important stress factor for vegetation. Therefore, detailed knowledge on the spatial variation of hydrological conditions is essential for sustainable and site-specific ecosystem management. However, realistic estimates of temporal and spatial patterns of soil moisture are challenging, in particular in complex terrain where increased diversity in topographic, land use, soil and climate conditions results in large variations in soil water availability (Jasper et al. 2006). Generally, physical-based hydrological models are used to quantify these variations. A multitude of different approaches and models has been developed in the last

Doi: 10.1111/j.1654-109X.2012.01210.x © 2012 International Association for Vegetation Science

decades (cf. Lane 1998; Praskievicz & Chang 2009 for an overview). To quantify and visualize the spatial variations of soil hydrological conditions several studies aim to link geographic information systems (GIS) with process-based models, e.g. WaSim-ETH (Jasper et al. 2006), the SVAT model PROMET with a modified version of TOPMODEL (Ludwig & Mauser 2000), BROOK90 (Schwärzel et al. 2009) or others (Fitz et al. 1996; Aspinall & Pearson 2000; Schröder et al. 2008).

Although being conceptionally different, all of the above models have in common simulation of the temporal change in soil water content. Their main intention is not to predict spatially explicit patterns of soil moisture, but to dynamically quantify the water balance of river basins. In addition, hydrological models usually operate at the scale of a specific river catchment and are therefore spatially restricted. Moreover, they need a wide set of calibration data, which are seldom available for larger study areas. These characteristics make process-based hydrological models impractical for site-specific ecosystem management, especially if the area of interest covers more than one catchment and if the required data are not available.

In this study, we present a new approach for spatial modelling of ecologically effective hydrological conditions based on vegetation data. We use Ellenberg indicator values for moisture as the response variable in a sophisticated statistical modelling framework. Indicator values have been widely used in vegetation science, forestry and landscape ecology as proxies for environmental conditions (Diekmann 2003). Each indicator species in the Central European flora is assigned an ordinal value on a nine-point scale (Ellenberg et al. 2001). Indicator values of sample plots are calculated by (weighted or non-weighted) averaging moisture values for all species, thus assigning the plot to a relative position on the hydrological gradient. Other indicator values are available, e.g. for light (L), soil reaction/pH (R) and nutrients (N). Applications range from comparison of environmental conditions between sites to the analysis of temporal changes in environmental conditions (cf. Diekmann 2003). Thus, indicator values are usually used to draw conclusions from plant species composition on the site conditions prevailing on a plot.

Regression of indicator values against measured environmental variables is a common method of assessing the quality of indication (Schaffers & Sykora 2000; Wamelink et al. 2002; Diekmann 2003; Ewald 2003), and species response curves of individual species have been used to refine the indicator value system (Wamelink et al. 2005). In both cases, indicator values serve to replace costly on-site measurements: indicator values serve to estimate soil conditions in plot locations.

Predicting indicator values from spatial environmental data has a fundamentally different purpose. Here, indicator values that have been sampled in plots are projected onto larger landscapes based on coarser geodata describing relief, climate and soils, i.e. the best combination of environmental predictors serves to extrapolate maps of the biotic response. So far, spatial modelling of indicator values has rarely been based on advanced statistical approaches, whereas the latter have been widely applied in species distribution modelling (SDM, Elith et al. 2006; Franklin 2010). Due to restricted availability of environmental predictors, studies predicting indicator values have covered only small study areas and/or specific habitats, e.g. mires. Thus, Schmidtlein & Sassin (2004) and Schmidtlein (2005) extrapolated indicator values for soil moisture, soil pH and soil fertility from 46 point observations to an area of about 2.5 km^2 using airborne hyperspectral images with partial least squares regression. Feldmeyer-Christe et al. (2007) and Ecker et al. (2008, 2010) performed several studies on mires in Switzerland using topographic and remote sensing data to model mean indicator values calibrated against vegetation plots.

To our knowledge, the study of Reger et al. (2011) is the only one to model indicator values on a larger landscape with more heterogeneous environmental conditions. They mapped the effects of temperature and radiation on forest vegetation, expressed as indicator values for temperature, in forests of the Bavarian Alps using multiple linear regression with climatic and topographic predictors. In this study we extend the approach of Reger et al. (2011) to model moisture values in the Bavarian Alps. We consider moisture as a far more complex gradient than temperature for two reasons: (1) it involves a much larger set of environmental covariates of relief, climate and soil and their interactions, and (2) it is influenced by different, but overlapping, physiological constraints like shade, drought and waterlogging (Niinemets & Valladares 2006). Thus, modelling moisture values requires a more flexible modelling framework such as additive georegression performed with the R library mboost (R Foundation for Statistical Computing, Vienna, AT). By doing so, we are able to predict the site-specific vegetation response to moisture conditions at the landscape scale and produce a map of Ellenberg mean moisture values.

In the last decades, considerable progress has been made in the development of statistical models especially for application in SDM (Guisan & Zimmermann 2000; Elith et al. 2006; Hastie et al. 2009). In SDM, the distribution of one or several species is related to environmental conditions by applying advanced statistics or machine learning techniques. Since model parameters are estimated from observation data, the quality of SDM depends on the quality, grain, and extent of the data (Guisan et al. 2007; Graham et al. 2008). Besides good data, the statistical approach must deal with the complex nature of environmental data.

Doi: 10.1111/j.1654-109X.2012.01210.x © 2012 International Association for Vegetation Science

Hothorn et al. (2010) developed the boosting-based additive georegression model framework, which has been shown to be an optimal modelling technique for environmental studies because it is the first approach to deal with all possible complexities of SDM at the same time: nonlinear relationships and interactions between predictors, spatial autocorrelation and non-stationarity (Hothorn et al. 2011; Schmid et al. 2011).

In this study, we used this statistical framework to predict indicator values for moisture at high resolution and large extent. Combined with comparable regionalizations of temperature (Reger et al. 2011) and soil nutrient status, the moisture parameter will serve to map forest types with equal site conditions and potential natural vegetation (B. Reger, T. Häring & J. Ewald, Submitted).

Methods

Study area

The study area comprises the Bavarian Alps, a mountain range in Southern Germany with an area of ca. 4600 km^2 (Fig. 1). The mountain range is characterized by a long altitudinal gradient, from the Saalach valley (470 m a.s.l.) to Germany's highest mountain (Zugspitze 2962 m a.s.l.). The climate is cool humid, with mean annual temperature decreasing from 9 °C at lower elevations to −4 °C at the summits and annual precipitation ranging from 1175 up to 2800 mm. Due to geological diversity, the Bavarian Alps are characterized by a rich mosaic of soil types. Thus, limestones give rise to lithic as well as rendzic leptosols, histosols and terrae fuscae, whereas cambisols, stagnosols and gleyosols are widespread on sandstone, mudstone and marl (classification according to IUSS Working Group WRB 2007).

Predictor variables

We used climatic, edaphic and topographic data as predictor variables for statistical modelling (Table 1). The predictor variables were selected for their potential relevance to model soil moisture.

Information on climate conditions was derived from monthly climate maps of Bavaria with a spatial resolution of 50 m (Hera et al. 2012). Climate maps were spatial interpolations based on daily measurements from 82 climate stations for the period 1971 to 2000 obtained from the German Meteorological Service (DWD), of which 14 were located in the Bavarian Alps. Monthly temperature and precipitation maps were used to calculate mean temperature and precipitation in the vegetation period (May–September).

We used the official soil map of Bavaria (Bavarian Environment Agency, '*Übersichtsbodenkarte*' *ÜBK25*, http://www.lfu.bayern.de), with a map scale of 1:25 000, to derive information on soil properties. As the soil map does not cover the entire study area, we used the concept of digital soil mapping to fill the gaps (McBratney et al. 2003; Häring Accepted). In digital soil mapping, statistical models are used to quantify functional relationships between the spatial distribution of soils and different environmental predictors, such as geology and topography. These models can be used to predict soil map units in areas where no soil map is available. Accuracy estimates with independent validation data of the predicted soil maps in our study area range between 50% and 80% correctly classified instances. In addition to the statistical validation, an extensive field campaign revealed the same mean accuracy from ground-truth data (Häring & Schröder 2010). By combining the traditionally mapped soil map of the Bavarian Environment Agency (LfU) and the predicted soil map, we obtained a homogeneous soil map for the entire study area.

Information on soil physical or soil hydrological properties were attached to the soil map by assigning data of representative soil profiles to every single soil map unit. Soil attributes of interest for this study are saturated soil hydraulic conductivity, depth of a slowly permeable horizon, available water capacity and air capacity.

Terrain characteristics were obtained by calculating different topographic attributes (Hengl & Reuter 2009; Table 1) from a digital surface model (DSM) of the Bavarian Topographical Survey with a resolution of 10 m. The DSM is based on airborne laser scan measurements using a LiDAR system. Errors and anthropogenic elements such as roads or settlements in the DSM were eliminated to obtain a digital elevation model before calculating topographic attributes.

Following Ewald (2009) we transformed aspect to an ecologically meaningful variable on a scale from −1 (thermally least favoured, NNW-exposed slopes, approximately 22.5°) to 1 (thermally most favoured, SSE-exposed slopes, approximately 202.5°).

Response variables

Hydrological site conditions were tested against the response of forest vegetation composition towards moisture, as summarized by Ellenberg indicator values for moisture given in Ellenberg et al. (2001) for all vascular plants and bryophytes. Ellenberg indicator values for moisture are part of an expert system, which ranks all Central European plant species according to empirical knowledge on an ordinal nine-point scale (Table 2).

The indicator values were assigned to 904 vascular plants and 238 bryophytes recorded in 1505 georeferenced vegetation plots (Fig. 1) from the database WINALPecobase (Reger et al. in press). The database is an ecological

 Applied Vegetation Science
Doi: 10.1111/j.1654-109X.2012.01210.x © 2012 International Association for Vegetation Science

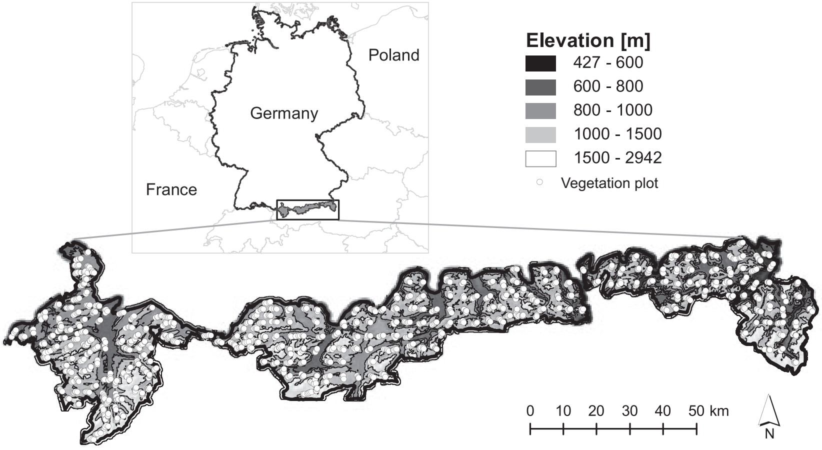


Fig. 1. Location of the 1505 vegetation plots of the database WINALPecobase within the study area (Reger et al. In press). Data source: Digital elevation model from the Bavarian Surveying Administration (LVG).

repository for forest vegetation with a concomitant soil profile description in the Bavarian Alps. The sampling was conducted in a combined systematic and stratified design in order to ensure representative, multivariate, spatially balanced data recording. Vegetation plots were organized in transects of five samples each. Starting positions of the transects were chosen from the grid of the second National Forest Inventory (NFI 2, *Bundeswaldinventur*) in Germany, with points spaced 4 × 4 km or 2 × 2 km in two sub-regions. Each NFI position consists of four coordinate points with a distance of 150 m, of which at least one is located in a forest stand. The first vegetation plot was randomly selected out of one to four possible coordinate points of the NFI 2 square. The selected plot was the starting point of a contour line of ca. 2-km length. The contour line was field-inspected by a joint team consisting of a vegetation and a soil expert and classified into ecologically homogeneous segments according to the forest type classification of Ewald & Binner (2007). Four additional plots were placed to represent the different site types present along the line. Plot locations were chosen in mature stands with tree layer composition and structure corresponding as closely as possible to the natural vegetation of the forest types.

Table 1. Description of the climate, edaphic and topographic variables derived from climate maps, soil maps and digital elevation models.

Predictor variable	Min	Mean (±SD)	Max
Climate variables			
Mean temperature in the growing season (May–September) (°C)	7.68	11.68 (1.45)	15.59
Mean precipitation in the growing season (May–September) (°C)	586	958 (132)	1398
Topographic variables			
Aspect (modified)	−1	−0.07 (0.68)	1
Slope (°)	0.15	25.11 (14.1)	65.05
Flood plain index	0.38	1.54 (0.55)	2.55
Vertical distance to channel network	0	49.06 (155.6)	639.37
Mid-slope position	0	0.52 (0.28)	1
Topographic wetness Index	2.09	5.65 (2.34)	20.77
Plan curvature	0	0.036 (0.0083)	0.077
Profile curvature	0	0.021 (0.0099)	0.037
Soil variables			
Available water capacity (mm)	54.6	128.2 (73.7)	600
Hydraulic conductivity ($cm{\cdot}d^{-1}$)	2.87	31.67 (19.3)	108.68
Aeration capacity (mm)	17.19	78.79 (32.7)	250
Depth of an impermeable layer (cm)	16	127 (25.8)	NA

In plots of 14 × 14 m, all soil-dwelling vascular plants, epigeic bryophytes and lichen species and their cover (six-point scale simplified after Braun-Blanquet 1964) were recorded, with woody species separated into four vertical vegetation layers.

Following common practice (Schaffers & Sykora 2000; Diekmann 2003), Ellenberg values of all species in the plot were averaged by calculating (1) log-abundance weighted averages of indicator values for soil moisture based on vascular plants and bryophytes; (2) log-abundance weighted averages of indicator values for soil moisture based on vascular plants; (3) unweighted averages of indicator values for soil moisture based on vascular plants and bryophytes; and (4) unweighted averages of indicator values for moisture based on vascular plants. The four variants of average indicator values for moisture were highly correlated (ranging between $r = 0.92$ and $r = 0.97$) and were used as alternative response variables in our analyses (cf. Reger et al. 2011).

Table 2. Definition of Ellenberg indicator values for moisture.

Ellenberg value	Description
1	Indicator of strong drought, viable under frequent desiccation and restricted to dry sites
2	Transitional between 1 and 3
3	Absent from moist soils
4	Transitional between 3 and 5
5	Indicator of mesic conditions, optimum on soils of intermediate moisture, neither on wet nor on droughted soils
6	Transitional between 5 and 7
7	Indicator of moist conditions, optimum on moist soils without permanent waterlogging
8	Transitional between 7 and 9
9	Indicator of wetness, ecological optimum under frequent waterlogging and poor aeration

Statistical modelling

According to Hothorn et al. (2011) complexities in environmental data can be classified into three groups: (1) nonlinear relationships and interactions between predictors, (2) spatial autocorrelation, and (3) non-stationarity. Spatial autocorrelation addresses the small-scale variability of observations, i.e. observations that are close in space are more similar to each other than the corresponding environmental conditions would suggest (Legendre 1993). Spatial non-stationarity describes relationships that are not constant across space (Miller & Hanham 2011), meaning that the influence of, e.g. slope on the Ellenberg moisture value is not global but may vary across our study area. All three issues can have an influence on the statistical model if they are present in the data, and thus might lead to biased estimates of model parameters (Dormann et al. 2007) and misleading conclusions on site ecological characteristics. The modelling framework applied here allows for inclusion of terms dealing with all three issues simultaneously by decomposing the predictor into several components (cf. Hothorn et al. 2011). As response variable, we used the weighted and non-weighted mean soil moisture values calculated on vascular plants or on vascular plant in addition to bryophytes as dependent variables, which are on a continuous scale and fit the model by optimizing the squared error loss.

In order to represent all significant effects in the final model, i.e. environmental covariates, spatial effects, interactions and non-stationarity, we fitted four separate regression models to check each effect. In each model, specific model components were included. In the first model, we only used all environmental covariates listed in Table 1 (*covar*), assuming that the regression effects are constant in space. In order to consider non-linear relationships between moisture values and the environmental predictors, we used smoothing splines as base-learners. To account for spatial dependency between neighbouring plots, we included a smooth two-dimensional surface function that quantifies spatial effects in the second model (*covarspatial*) in addition to the environmental predictors. A P-spline tensor product surface depending on the geographic coordinates of the vegetation plots was included as separate base-learner (see Kneib et al. 2008). The third model used decision trees instead of smoothing splines as base-learners (*tree*) and is equal to the well-established boosted regression tree model (Elith et al. 2008). With the tree model we checked for interactions between predictor variables. To allow for non-stationarity of the environmental variables, we included additionally spatially varying effects for every variable in the fourth model (*vary*). Table 3 gives an overview of the structure of the different models.

Our final model, which is used for prediction, combined all significant model components into one model. In order to keep the model as parsimonious as possible, we applied a variable selection procedure called stability selection, as proposed in Meinshausen & Bühlmann (2010). This procedure is also included in the mboost package and has the advantage of improving the intrinsic variable selection properties of boosting, independent of specific assumptions, as well as being comprehensible for the user.

Validations of the regression models were conducted by bootstrap sampling of the 1505 plots. Two-thirds of the data were used as calibration data and one-third for validation. Using the calibration data, we fitted the regression models and calculated the R^2 between observed and predicted mean moisture values of the validation data set. In order to obtain stable results, we repeated this procedure 50 times.

Statistical modelling was carried out with the mboost package (v. 2.0-11) for R (version 2.13.2; R Foundation for Statistical Computing). Spatial prediction was made with the R package raster (v. 1.8-39). Terrain analysis was conducted with SAGA GIS (v. 2.0.7).

Table 3. Model structure, i.e. model components and base-learner, in all five models used in our study.

Model structure	Model components			Base-learner	
	Environmental effects	Spatial effects	Non-stationary effects	Spline	Tree
covar	×			×	
covarspatial	×	×		×	
tree	×				×
vary	×	×	×	×	
final	×		×	×	×

Applied Vegetation Science
Doi: 10.1111/j.1654-109X.2012.01210.x © 2012 International Association for Vegetation Science

Results

We first investigated the prediction accuracy of the five different models, i.e. the ability of the models to predict moisture values for independent validation data. Figure 2 shows R^2 values of the 50 replicates for each of the five regression models and for all four response values. The plots were grouped according to the model structures. Colours of the boxplots indicate the four mean moisture values. Figure 2 illustrates three different findings.

First, there are interactions between the predictors as well as spatially varying effects present in the data. The *covar* model, which neither accounts for spatial autocorrelation nor interactions, nor non-stationarity, exhibits the worst performance. Since the R^2 of the *covarspatial* model is only negligibly better than the *covar* model, we can conclude that there are no effects of spatial autocorrelation in our data. However, interactions are present, because the *tree* model shows a significantly better fit. Moreover, spatially varying effects of the environmental variables seem also to be present in the data that cannot be modelled by the *tree* model. Therefore, the *vary* model outperforms the *tree* model. Our combined final model has the highest R^2 values. Best results were achieved for unweighted mean values calculated on vascular plants alone. We found a significant influence of interaction effects as well as non-stationary effects.

Although the five model approaches differ in the magnitude of their R^2, the overall pattern of the boxplots between these five groups remains. In all five model approaches unweighted mean moisture values can be better modelled than weighted values. Finally, the consideration of bryophytes does not improve model performance. Considering the unweighted values, the R^2 values are even higher for vascular plants compared to vascular plants plus bryophytes.

The final model, with unweighted mean moisture values calculated only for vascular plants as dependent variable, was used for predicting a map of regionalized Ellenberg indicator values. In this final model, topographic, climate and soil parameters are used as predictor variables. The applied variable selection procedure reduced the model effects to nine significant variables. The remaining less- or non-informative variables are not considered. Partial contributions of all selected environmental variables are given in Fig. 3. All selected non-stationary effects are shown in Appendix S1; the resulting map is presented in Fig. 4.

Response curves (Fig. 3) indicate meaningful dependencies between Ellenberg's moisture values and environmental covariates. Light grey lines correspond to partial contributions of the 50 bootstrap samples. Mid-slope position between 0.3 and 0.7, i.e. footslope positions where soil water is accumulated, has a positive effect on moisture values. Areas with lower precipitation and higher temperatures were modelled as sites with lower indicator values. Warmer and drier SSW-exposed slopes (aspect = 1) have a negative effect on moisture conditions, in contrast to the positive effect of NNE-exposed slopes (aspect = −1). Increasing plan curvature values, i.e. exposed topographic positions with increasing convexity like ridges, have a negative effect on moisture values. Floodplains (small values for flood plain index) tend to have higher moisture values. Increasing flood plain index has a negative effect. Steep slopes have a strongly negative effect on moisture conditions. Positive effects were also present in soils with avail-

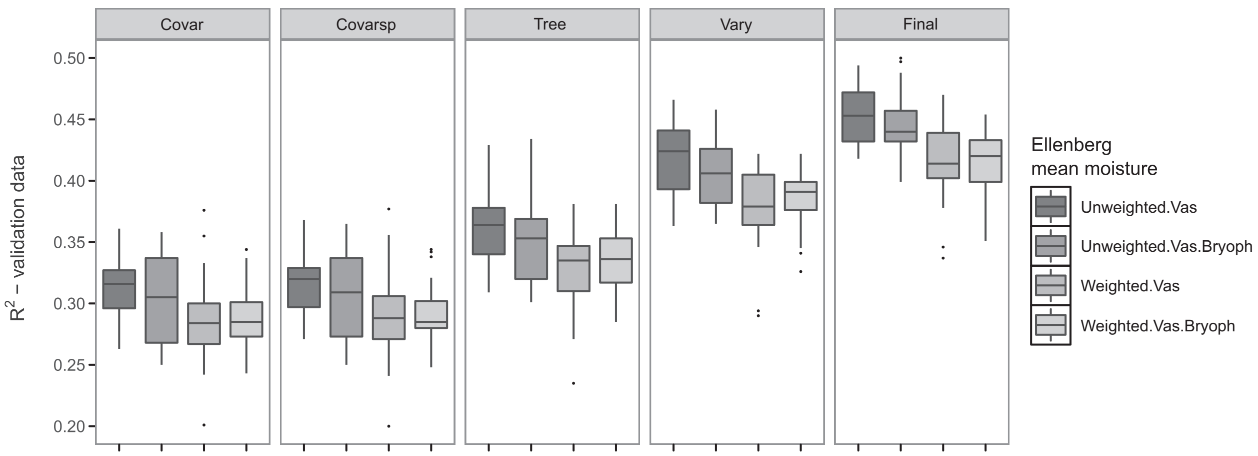


Fig. 2. The R^2 values of 50 replicates grouped according to five different mboost models (*covar*, *covarspatial*, *tree*, *vary*, *final*). Shading of the boxplots indicates the four different mean moisture values. Our final model, which includes environmental covariates, interaction terms (by means of decision trees) and spatially varying model components performs best. Unweighted mean moisture values give better results compared to weighted values. The consideration of bryophytes does not improve model performance.

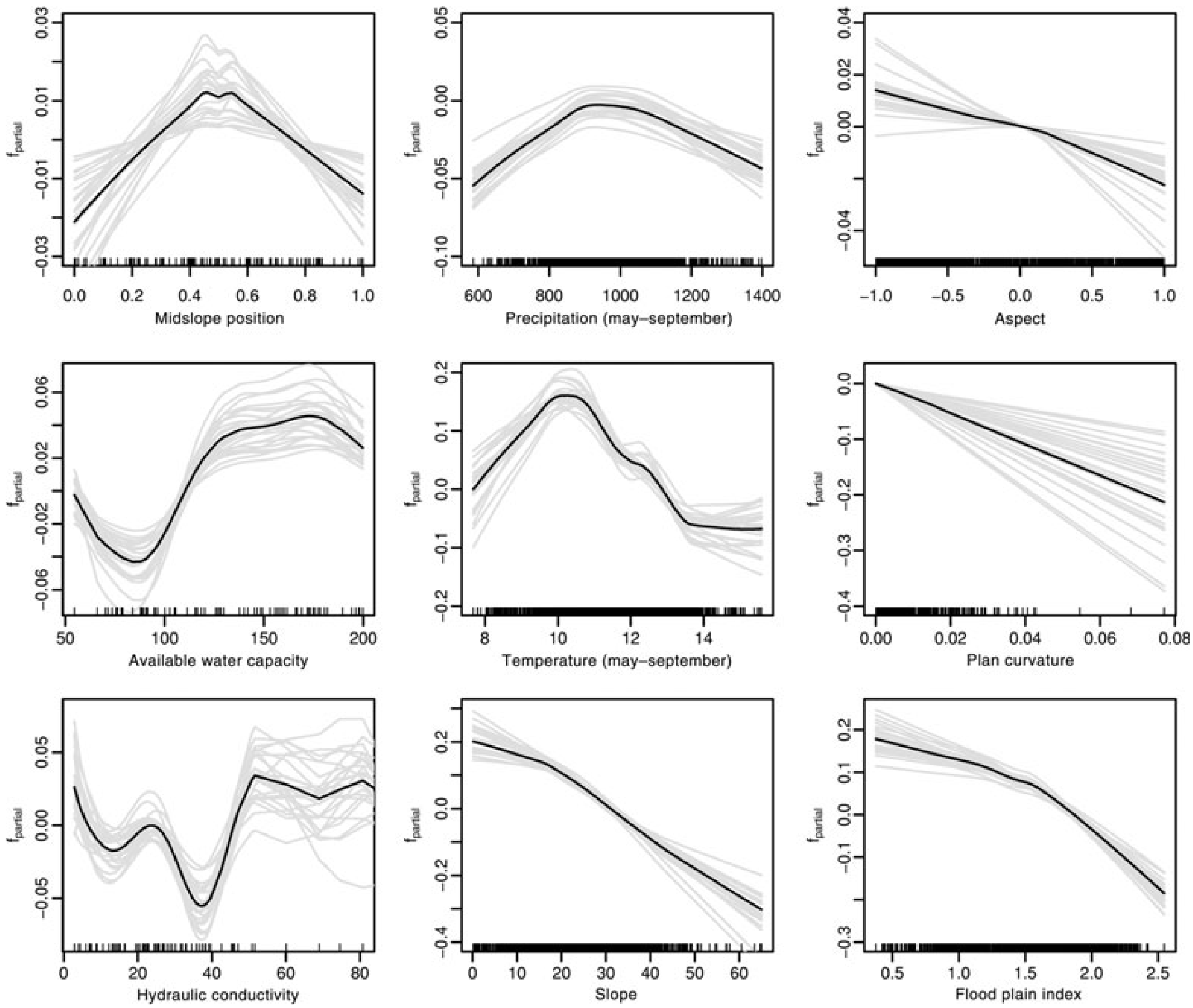


Fig. 3. Response curves for the 50 replicates of the final mboost model with unweighted mean values of vascular plants as response values. The black line indicates the mean response of the 50 bootstrap samples. Higher magnitudes of the coefficients indicate higher importance for modelling moisture values.

able water capacity exceeding 115 mm and hydraulic conductivity >46 mm·d^{-1}. Concerning the magnitude of the marginal effects, slope, plan curvature, flood plain index and temperature had the highest impact in our final model.

In addition to the environmental model components, spatially varying model effects have a significant effect on soil moisture distribution. Appendix S1 shows these effects, in which blue shading represents areas where the corresponding covariates have a positive effect, whereas red shading represents areas where the impact is negative. All maps show high spatial variability, indicating that the soil moisture–environment relationship is dependent on the absolute location in space. Most relationships coincide with physiography (mid-altitudinal ranges in the N part, high altitudes in the S; cf. Fig. 4). Only aspect shows a strong gradient in E–W direction. The probability of high moisture values is higher in the W part of the study area compared to areas with the same exposure in the E, i.e. slopes exposed to the N are moister in the W part of the study area than in the E (assuming all other environmental conditions as constant). Relative elevation has a slight positive effect in high-altitude areas, as well as plan and profile curvature. However, the effects of both curvature covariates are much higher. Small-scale topographic variations play an important role of soil moisture conditions in mountainous regions, where exposed ridges and water-accumulating valleys can alternate within a short distance. Precipitation reveals a strong, spatially varying pattern; however, the magnitude of effect is much less compared to the other covariates.

Doi: 10.1111/j.1654-109X.2012.01210.x © 2012 International Association for Vegetation Science

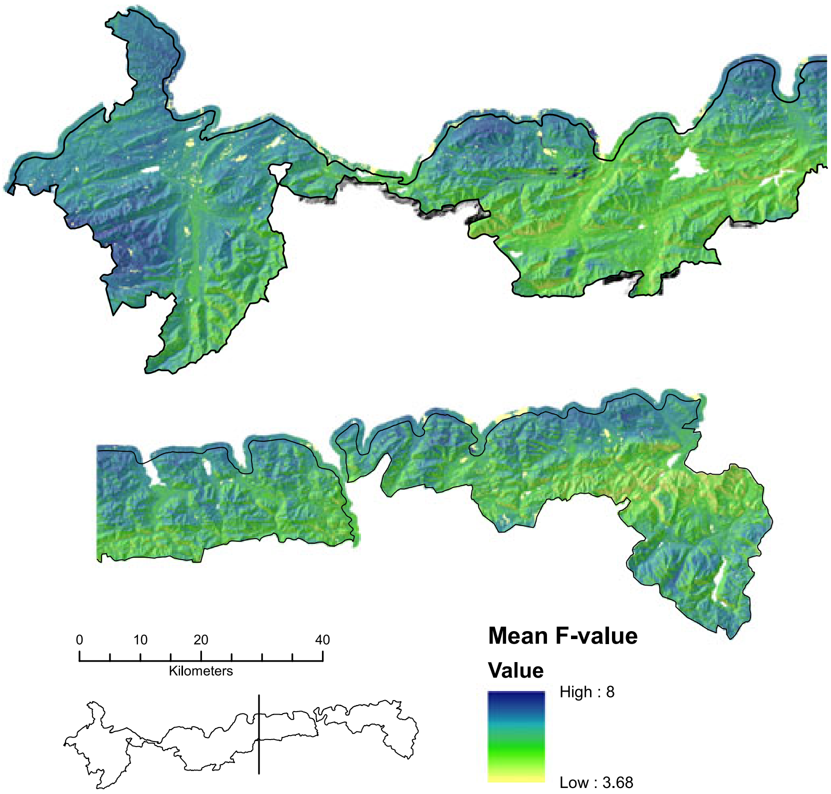


Fig. 4. Predicted mean moisture value (unweighted mean of vascular plants) for the Bavarian Alps. The prediction is made on a 10-m grid and is displayed draped over a hillshade (© Elevation data: Bavarian Topographical Survey).

The resulting map (Fig. 4) shows high variability across different topographic and edaphic conditions. Areas with S and SW exposure, where potential insolation is high, slopes steep and ridges exposed, are modelled as drier areas.

Figure 5 shows the predicted mean moisture values plotted against the observed values for the calibration and validation data. The plot indicates a promising correlation for indicator values between 4.5 and 6.0. However, prediction accuracy decreases for extremely dry and extremely wet areas, which are under- or overestimated. This may be due to a weak sample for these sections of the gradient. Mean moisture for vascular plants has a 5% quantile of 4.64 and a 95% quantile of 6.06 in our data set.

Discussion

Forest ecosystems are very likely to be influenced by climate change (Kellomäki & Leinonen 2005; Maracchi et al. 2005), especially in areas that are highly vulnerable to changing environmental conditions, like mountainous regions (Parry et al. 2007). Future forest management therefore requires, among other things, detailed information on site ecological conditions to adapt forests to future conditions in a sustainable way (Seidl et al. 2011). Various approaches to assess and classify climate and soil conditions are available (Barnes et al. 1982; Pojar et al. 1987; Wilson et al. 2001); however, these approaches are rather qualitative frameworks to classify site ecological conditions. The consequent implementation of these frameworks in maps, which was identified as the 'heart of classification' (Barnes et al. 1982), is rare or at least remains in the qualitative implementation of rule-sets in a GIS (Clare & Ray 2001).

Predictive vegetation mapping (Franklin 1995) or, more generally, SDM, provide methods to generate quantitative and reproducible maps of site conditions. The spatial distribution of soil moisture conditions in the Bavarian Alps could be successfully predicted across space by combining statistical models with spatially explicit data in a GIS environment. However, our results point to the need for a modelling framework that accounts not only for the species–environment relationship, as in the majority of currently available models (Franklin 2010). Small-scale (spatial autocorrelation) and large-scale (non-stationarity) influences of spatial dependencies have proven to be essential components of species–environment relationships (Foody 2004; Miller & Hanham 2011).

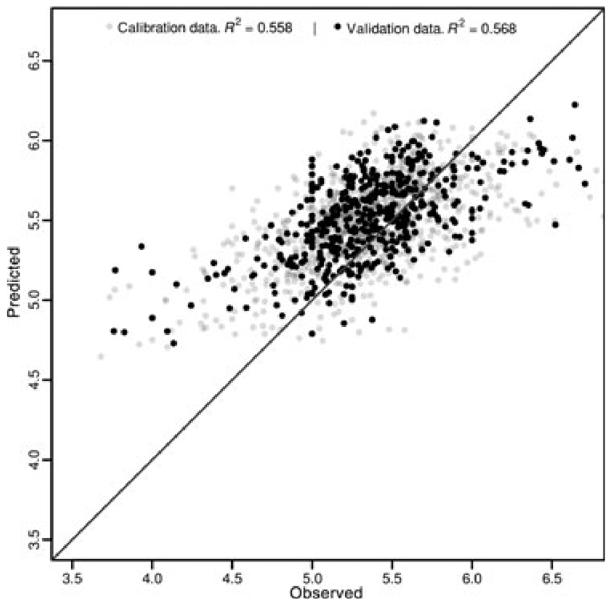


Fig. 5. Predicted mean moisture values plotted against observed values. Grey points are calibration data, black points are independent validation data. The R^2 of validation data is equal to calibration data, showing a good fit to the data.

Figure 2 reveals that best results can be achieved when combining environmental covariates with spatially varying model effects and interaction terms. In the traditional statistical framework for vegetation modelling, vegetation response is related to a set of environmental predictors without considering any spatial information or dependencies. Nevertheless, there are spatial patterns that can be explained by predictor variables in models, like precipitation, temperature and elevation, which can exhibit a high spatial dependency (Miller et al. 2007). However, our results show a clear improvement of model performance when considering spatial information. This implies that not all spatial dependencies of our response could be modelled with environmental predictors alone. One commonly noted feature is that biogeographic patterns appear to vary as a function of space on different scales (Foody 2004; Miller & Hanham 2011). In the feature selection algorithm applied to the final model selected – besides environmental variables in a tree as base-learner – only three variables (mid-slope position, available water capacity, hydraulic conductivity) are included as spline functions (comparable to the traditional GAM) in the model, whereas five variables were selected as being non-stationary, i.e. spatially varying effects (precipitation, aspect, plan curvature, profile curvature, vertical distance to channel network; cf. Appendix S1).

Untreated spatial information might not only lead to lower model performance (Foody 2004; Hothorn et al. 2011) but also to biased estimates (Legendre & Legendre 1998; Lennon 2000). Therefore, applying models in SDM, which accounts for spatial dependencies, like mboost, has provided an important enhancement (Lichstein et al. 2002; Foody 2004; Hothorn et al. 2011; Miller & Hanham 2011).

Our models achieved only slightly higher levels of explained variance than the simple linear regression models used by Ewald (2003, 2009). Although in the same region, both of these studies were based on a different data set, vegetation plots and soil profiles than used in the present study. The first study used available soil water capacity; while the second used a more complex predictor variable that logically combined water storage and topographical position. Strikingly, soil variables appeared much more important than in our study. This could be explained by the fact that we had only extremely coarse soil information available, derived from sample profiles representing entire mapping units. The quality of prediction in our model depends mostly on topographic and climatic variables and their interactions and local weighting, while the rather general soil information makes a minor contribution (cf. magnitude of partial contribution in Fig. 3). Taking this into account, our study demonstrates the considerable predictive power of the mboost model, especially for spatial predictions. However, detailed data sets as used in Ewald (2003, 2009) are not useful for our purpose because they are only available for sampling locations. Spatial prediction studies require environmental variables that are available as spatial GIS layers.

As already shown by Reger et al. (2011) for temperature indicator values, unweighted moisture values can be modelled slightly better with environmental predictors than log abundance-weighted indicator values. The fact that qualitative (presence/absence) information outperforms quantitative data indicates that variation in abundance is partly caused by unmeasured predictors such as disturbance, browsing pressure and successional status (Ewald 2008). Moreover, recording cover is more prone to errors than species identification (Wilson 2011).

The non-significant contribution of bryophytes to indicators based on Ellenberg values of vascular plants has also been reported in Ewald (2009) and Reger et al. (2011) for the Bavarian Alps. Ewald (2009) suggested two reasons for this: (1) bryophytes may respond to environmental variables not commonly measured (e.g. air humidity), and (2) their indicator values may be partly mis-calibrated. We therefore used vascular plant-based indicator values.

Conclusion and outlook

Our modelling approach has two remarkable properties: it finds optimal parameters and interaction terms among a typical set of physiographic predictors by regressing them

Doi: 10.1111/j.1654-109X.2012.01210.x © 2012 International Association for Vegetation Science

onto the moisture-specific component in plant composition. We show that this is possible even at landscape scale in a study area with an extent >4500 km^2, thus offering an effect-oriented alternative to climatological, hydrological and physiological models. It makes the best of those predictors that are available for broad-scale regionalization in GIS and, in the face of the obvious weaknesses of soil predictors, achieves remarkable levels of prediction. The mapping of ecological gradients is a significant step towards a new generation of models for forest site types and potential natural vegetation.

Acknowledgements

The research presented here forms part of the project 'Forest Information System for the Northern Alps' (www.winalp.info), which is funded by the European Fund for Regional Development (EFRE) within the 'INTERREG Bayern – Österreich 2007–2013' programme, the Bavarian Forest Administration and the Bavaria State Forest Enterprise (BaySF). We are indebted to the Bavarian State Institute of Forestry's 'Maps for the future' crew. The digital elevation model was obtained from the Bavarian Topographical Survey (LVG), soil maps were obtained from the Bavarian State Agency for the Environment (LfU) and climate data were obtained from German Meteorological Service (DWD).

References

Aspinall, R. & Pearson, D. 2000. Integrated geographical assessment of environmental condition in water catchments: linking landscape ecology, environmental modelling and GIS. *Journal of Environmental Management* 59: 299–319.

Barnes, B.V., Pregitzer, K.S., Spies, T.A. & Spooner, V.H. 1982. Ecological Forest Site Classification. *Journal of Forestry* 80: 494–498.

Braun-Blanquet, J. 1964. *Pflanzensoziologie*. Springer, Vienna, AT.

Clare, J. & Ray, D. 2001. A Spatial Model of Ecological Site Classification for forest management in Britain. *Proceedings of the 4th AGILE Conference on Geographic Information Science in Brno, CZ*, 19–21 April 2001.

Diekmann, M. 2003. Species indicator values as an important tool in applied plant ecology – a review. *Basic and Applied Ecology* 4: 493–506.

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, D., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M. & Wilson, R. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30: 609–628.

Ecker, K., Kuchler, M., Feldmeyer-Christe, E., Graf, U. & Waser, L.T. 2008. Predictive mapping of floristic site conditions across mire habitats: evaluating data requirements. *Community Ecology* 9: 133–146.

Ecker, K., Waser, L.T. & Kuchler, M. 2010. Contribution of multi-source remote sensing data to predictive mapping of plant-indicator gradients within Swiss mire habitats. *Botanica Helvetica* 120: 29–42.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S. & Zimmermann, N. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129–151.

Elith, J., Leathwick, J.R. & Hastie, T. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77: 802–813.

Ellenberg, H., Weber, H.E., Düll, R., Wirth, V. & Werner, W. 2001. *Zeigerwerte von Pflanzen in Mitteleuropa*. Erich Goltze KG, Göttingen, DE.

Ewald, J. 2003. The sensitivity of Ellenberg indicator values to the completeness of vegetation relevés. *Basic and Applied Ecology* 4: 507–513.

Ewald, J. 2008. Plant species richness in mountain forests of the Bavarian Alps. *Plant Biosystems* 142: 594–603.

Ewald, J. 2009. Epigeic bryophytes do not improve bioindication by Ellenberg values in mountain forests. *Basic and Applied Ecology* 10: 420–426.

Ewald, J. & Binner, S. 2007. Werkzeuge zur Bestimmung der Waldtypen im bayerischen Hochgebirge. *Waldökologie Online* 5: 25–77.

Feldmeyer-Christe, E., Ecker, K., Kuchler, M., Graf, U. & Waser, L. 2007. Improving predictive mapping in Swiss mire ecosystems through re-calibration of indicator values. *Applied Vegetation Science* 10: 183–192.

Fitz, H.C., DeBellevue, E.B., Costanza, R., Boumans, R., Maxwell, T., Wainger, L. & Sklar, F.H. 1996. Development of a general ecosystem model for a range of scales and ecosystems. *Ecological Modelling* 88: 263–296.

Foody, G.M. 2004. Spatial nonstationarity and scale-dependency in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. *Global Ecology and Biogeography* 13: 315–320.

Franklin, J. 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography* 19: 474–499.

Franklin, J. 2010. *Mapping species distributions. Spatial inference and prediction*. Cambridge University Press, Cambridge, UK.

Graham, C.H., Elith, J., Hijmans, R.J., Guisan, A., Peterson, A.T. & Loiselle, B.A. 2008. The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology* 45: 239–247.

Guisan, A. & Zimmermann, N. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147–186.

Guisan, A., Zimmermann, N.E., Elith, J., Graham, C.H., Phillips, S. & Peterson, A.T. 2007. What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? *Ecological Monographs* 77: 615–630.

Häring, T. Accepted. Facilitation of Environmental Modeling by Means of Scripting: the Combination of R and SAGA. *Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie*.

Häring, T. & Schröder, B. 2010. A review of model-error in digital soil mapping: confronting statistical soil landscape models with large-scale field validation data. *Geophysical Research Abstracts* 12: EGU2010–EGU12757.

Hastie, T.J., Tibshirani, R.J. & Friedman, J.H. 2009. *The elements of statistical learning: data mining, inference, and prediction*, 2nd ed. Springer, New York, NY, US.

Hengl, T. & Reuter, H.I. (eds.) 2009. *Geomorphometry. Concepts, software, applications*. Elsevier, Amsterdam, NL.

Hera, U., Rötzer, T., Zimmermann, L., Schulz, C., Maier, H., Weber, H. & Kölling, C. 2012. Klima en détail. Neue, hochaufgelöste Klimakarten bilden wichtige Basis zur klimatischen Regionalisierung Bayerns. *LWF Aktuell* 86: 34–37.

Hothorn, T., Bühlmann, P., Kneib, T., Schmid, M. & Hofner, B. 2010. Model-based Boosting 2.0. *Journal of Machine Learning Research* 11: 2109–2113.

Hothorn, T., Müller, J., Schröder, B., Kneib, T. & Brandl, R. 2011. Decomposing environmental, spatial, and spatio-temporal components of species distributions. *Ecological Monographs* 81: 329–347.

IUSS Working Group WRB 2007. *World Reference Base for Soil Resources 2006, first update 2007*. World Soil Resources Reports No. 103. FAO, Rome, IT.

Jasper, K., Calanca, P. & Fuhrer, J. 2006. Changes in summertime soil water patterns in complex terrain due to climatic change. *Journal of Hydrology* 327: 550–563.

Kellomäki, S. & Leinonen, S. (eds.) 2005. *Management of European forests under changing climatic conditions*. Tiedonantoja/Research Notes No. 163. University of Joensuu, Faculty of Forestry, FI.

Kneib, T., Müller, J. & Hothorn, T. 2008. Spatial smoothing techniques for the assessment of habitat suitability. *Environmental and Ecological Statistics* 15: 343–364.

Lane, S.N. 1998. Hydraulic modelling in hydrology and geomorphology: a review of high resolution approaches. *Hydrological Processes* 12: 1131–1150.

Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm. *Ecology* 74: 1659.

Legendre, P. & Legendre, L. 1998. *Numerical ecology*. 2nd ed. Elsevier, Amsterdam, NL.

Lennon, J.J. 2000. Red-shifts and red herrings in geographical ecology. *Ecography* 23: 101–113.

Lichstein, J.W., Simons, T.R., Shriner, S.A. & Franzreb, K.E. 2002. Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs* 72: 445–463.

Ludwig, R. & Mauser, W. 2000. Modelling catchment hydrology within a GIS based SVAT-model framework. *Hydrology and Earth System Sciences* 4: 239–249.

Maracchi, G., Sirotenko, O. & Bindi, M. 2005. Impacts of present and future climate variability on agriculture and forestry in the temperate regions: Europe. *Climatic Change* 70: 117–135.

McBratney, A.B., Mendonça Santos, M.L. & Minasny, B. 2003. On digital soil mapping. *Geoderma* 117: 3–52.

Meinshausen, N. & Bühlmann, P. 2010. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72: 417–473.

Miller, J.A. & Hanham, R.Q. 2011. Spatial nonstationarity and the scale of species–environment relationships in the Mojave Desert, California, USA. *International Journal of Geographic Information Science* 25: 423–438.

Miller, J.A., Franklin, J. & Aspinall, R. 2007. Incorporating spatial dependence in predictive vegetation models. *Ecological Modelling* 202: 225–242.

Niinemets, U. & Valladares, F. 2006. Tolerance to shade, drought and waterlogging of temperate, Northern hemisphere trees and shrubs. *Ecological Monographs* 76: 521–547.

Parry, M.L., Canziani, O.F., Palutikof, J.P., van der Linden, P.J. & Hanson, C.E. (eds) 2007. *Contribution of working group II to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge University Press, Cambridge, UK.

Pojar, J., Klinka, K. & Meidinger, D.V. 1987. Biogeoclimatic Ecosystem Classification in British Columbia. *Forest Ecology and Management* 22: 119–154.

Praskievicz, S. & Chang, H. 2009. A review of hydrological modelling of basin-scale climate change and urban development impacts. *Progress in Physical Geography* 33: 650–671.

Reger, B., Kölling, C. & Ewald, J. 2011. Modelling effective thermal climate for mountain forests in the Bavarian Alps: which is the best model? *Journal of Vegetation Science* 22: 677–687.

Reger, B., Schüpferling, R., Beck, J., Dietz, E., Morovitz, D., Schaller, R., Wilhelm, G. & Ewald, J. (In press). WINALP-ecobase – Ecological database of mountain forests in the Bavarian Alps. *Biodiversity and Ecology* 4.

Schaffers, A.P. & Sýkora, K.V. 2000. Reliability of Ellenberg indicator values for moisture, nitrogen and soil reaction: a comparison with field measurements. *Journal of Vegetation Science* 11: 225–244.

Schmid, M., Hothorn, T., Maloney, K.O., Weller, D.E. & Potapov, S. 2011. Geoadditive Regression Modeling of Stream Biological Condition. *Environmental and Ecological Statistics* 18: 709–733.

Schmidtlein, S. 2005. Imaging spectroscopy as a tool for mapping Ellenberg indicator values. *Journal of Applied Ecology* 42: 966–974.

Schmidtlein, S. & Sassin, J. 2004. Mapping of continuous floristic gradients in grasslands using hyperspectral imagery. *Remote Sensing of Environment* 92: 126–138.

Schröder, B., Rudner, M., Biedermann, R., Kögl, H. & Kleyer, M. 2008. A landscape model for quantifying the trade-off between conservation needs and economic constraints in

 Applied Vegetation Science
Doi: 10.1111/j.1654-109X.2012.01210.x © 2012 International Association for Vegetation Science

the management of a semi-natural grassland community. *Biological Conservation* 141: 719–732.
Schwärzel, K., Feger, K.H., Häntzschel, J., Menzer, A., Spank, U., Clausnitzer, F., Kostner, B. & Bernhofer, C. 2009. A novel approach in model-based mapping of soil water conditions at forest sites. *Forest Ecology and Management* 258: 2163–2174.
Seidl, R., Rammer, W. & Lexer, M.J. 2011. Climate change vulnerability of sustainable forest management in the Eastern Alps. *Climatic Change* 106: 225–254.
Wamelink, G.W.W., Joosten, V., van Dobben, H.F. & Berendse, F. 2002. Validity of Ellenberg indicator values judged from physico-chemical field measurements. *Journal of Vegetation Science* 13: 269–278.
Wamelink, G.W.W., Goedhart, P.W., Van Dobben, H.F. & Berendse, F. 2005. Plant species as predictors of soil pH: replacing expert judgment by measurements. *Journal of Vegetation Science* 16: 461–470.
Wilson, J.B. 2011. Cover plus: ways of measuring plant canopies and the terms used for them. *Journal of Vegetation Science* 22: 197–206.
Wilson, S.Mc.G., Pyatt, D.G., Malcolm, D.C. & Connolly, T. 2001. The use of ground vegetation and humus type as indicators of soil nutrient regime for an ecological site classification of British forests. *Forest Ecology and Management* 140: 101–116.
Wisskirchen, R. & Haeupler, H. 1998. Standardliste der Farn- und Blütenpflanzen Deutschlands. Ulmer, Stuttgart, DE.

Supporting Information

Additional supporting information may be found in the online version of this article:

Appendix S1. Spatially varying coefficients of the final model.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

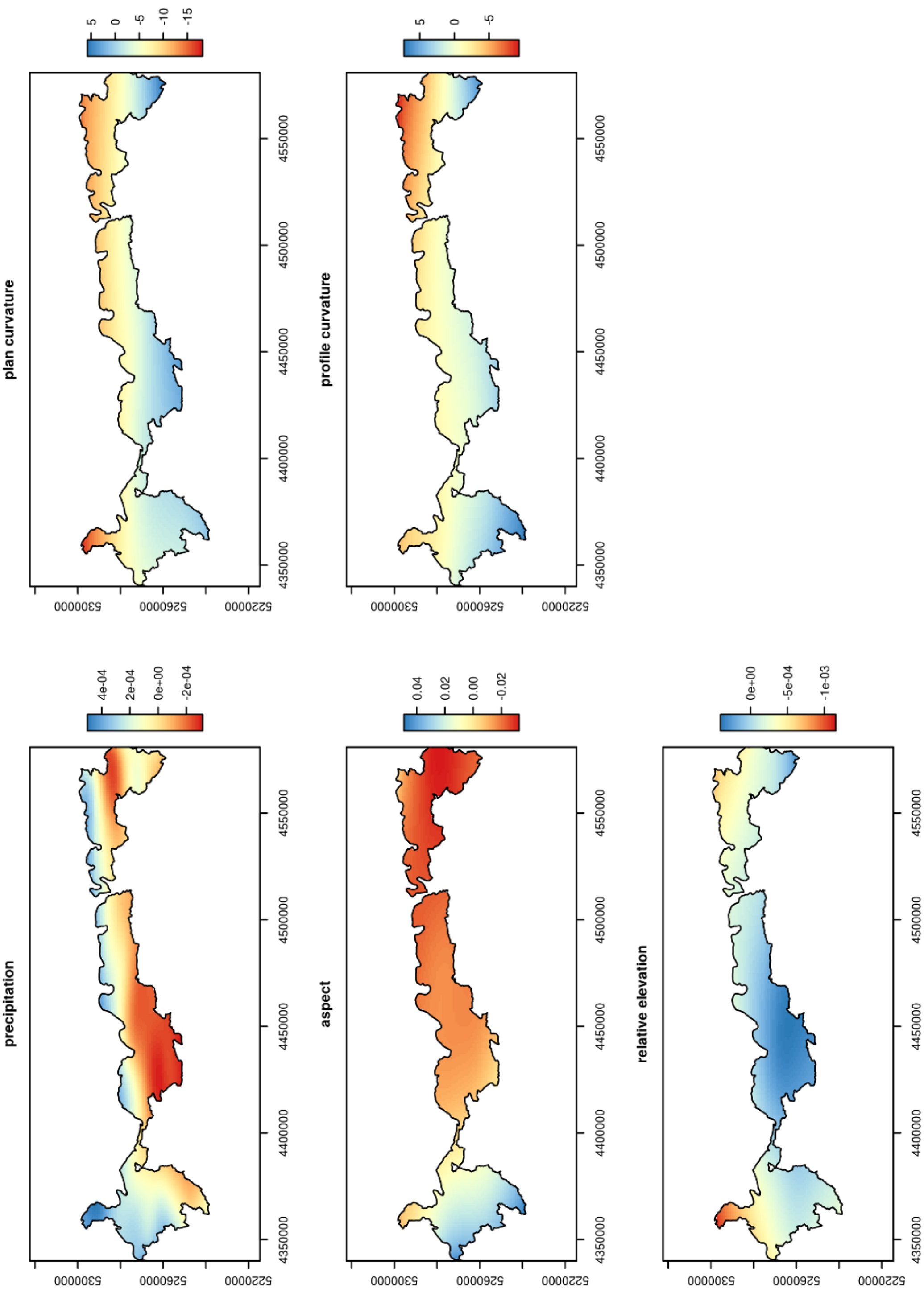
Electronic Supplement Material

Predicting Ellenberg's soil moisture indicator value in the Bavarian Alps using additive georegression

Tim Häring[*,†] Birgit Reger[§] Jörg Ewald[‡] Torsten Hothorn[§]
Boris Schröder[¶]

April 30, 2012

[*]E-mail: tim.haering@basf.com
[†]BASF SE, Environmental Fate – Modelling, Speyerer Str. 2, 67117 Limburgerhof, Germany
[‡]Faculty of Forestry, University of Applied Sciences Weihenstephan-Triesdorf, Hans-Carl von-Carlowitz-Platz 3, 85354 Freising, Germany
[§]Institut für Statistik, Ludwig-Maximilians-Universität München, Ludwigstraße 33, 80539 München, Germany
[¶]Landscape Ecology, Technische Universität München, Emil-Ramann-Strasse 6, 85350 Freising, Germany



Appendix S 1: Spatially varying coefficients of the final model.

A.3. Publication III

Häring, T., Reger, B., Ewald, J., Hothorn, T. & Schröder, B. (2014) Regionalising indicator values for soil reaction in the Bavarian Alps - from averages to multivariate spectra. *Folia Geobotanica*, doi:10.1007/s12224-013-9157-1.

Folia Geobot
DOI 10.1007/s12224-013-9157-1

Regionalizing Indicator Values for Soil Reaction in the Bavarian Alps – from Averages to Multivariate Spectra

Tim Häring · Birgit Reger · Jörg Ewald · Torsten Hothorn · Boris Schröder

© Institute of Botany, Academy of Sciences of the Czech Republic 2013

Abstract We present an approach to produce maps of Ellenberg values for soil reaction (R-value) in the Bavarian Alps. Eleven meaningful environmental predictors covering GIS-derived information on climatic, topographic and soil conditions were used to predict R-values. As dependent variables, Ellenberg indicator values for soil reaction were queried from plot records in the vegetation database WINALPecobase. We used an additive georegression model, which combines complex prediction models and the increased prediction accuracy of a boosting algorithm. In addition to environmental predictors we included spatial effects into the model to account for spatial autocorrelation. As we were particularly interested in the usefulness of averaged R-values for spatial prediction, we applied two different models: (1) a geo-additive regression model that estimates mean R-values and (2) a proportional odds model predicting the probability distribution over R-values 1 to 9. We found meaningful dependencies between the R-value and our predictors. Both models

Electronic supplementary material The online version of this article (doi:10.1007/s12224-013-9157-1) contains supplementary material, which is available to authorized users.

T. Häring
BASF SE, Environmental Fate – Modelling, Speyerer Str. 2, 67117 Limburgerhof, Germany

T. Häring (✉) · B. Schröder
Landscape Ecology, Technische Universität München, Emil-Ramann-Strasse 6,
85354 Freising, Germany
e-mail: tim.haering@basf.com

B. Reger · J. Ewald
Faculty of Forestry, University of Applied Sciences Weihenstephan-Triesdorf,
Hans-Carl von-Carlowitz-Platz 3, 85354 Freising, Germany

T. Hothorn
Institut für Sozial- und Präventivmedizin, Universität Zürich, Hirschengraben 84,
8001 Zürich, Switzerland

Published online: 03 July 2013

produced the same spatial pattern of predictions. Spatial effects had an impact only in the first model. The main drawback of mean R-values is the oversimplification of complex conditions of soil reaction, which is entailed by averaging and regression to mean values. Therefore, regionalized average indicator values provide only limited information on site-ecological characteristics. Model 1 failed to predict the range and shapes of original indicator spectra precisely. In contrast, the second model provided a more sophisticated picture of soil reaction. To make the multivariate output of model 2 comparable to that of model 1, we propose to plot the distribution in a three-dimensional color-space. In addition, comparison of both models based on a multiple linear regression model resulted in a R^2 of 0.93. The proportional odds model is a promising approach also for other indicator values and different regions as well as for other ordinal-scaled ecological parameters.

Keywords Boosting · Geo-additive regression · Proportional odds model · Spatial effect · Species distribution modeling

Introduction

Indicator values of plant species (Ellenberg et al. 2001) have widely been used in the assessment of site-ecological conditions (Hawkes et al. 1997; Diekmann 2003). They were assigned to plant species along a nine-point ordinal scale according to their estimated ecological optima on different environmental gradients (Käfer and Witte 2004). The possibility of a fast classification of a site by visual interpretation of plant species composition makes indicator values an attractive tool for ecologists as well as an easy-to-use assessment compared to time-consuming analysis of soil properties in the lab (Diekmann 2003).

By analyzing a set of different indicator values, e.g., for soil moisture, soil reaction, and temperature, one is able to classify sites into groups of similar environmental conditions (Reger et al. 2014, this issue). Using sophisticated methods from the field of species distribution modeling (cf. Elith and Leathwick 2009; Franklin 2010) maps of Ellenberg values providing detailed information on site characteristics can be produced (Reger et al. 2011; Häring et al. 2013) and these could be used for practical ecosystem management and further analysis. Reger et al. (2014, this issue) present a TRM-model (temperature-reaction-moisture) of potential natural forest vegetation in the Bavarian Alps by combining different Ellenberg indicator values.

Studies using indicator values as data input are based on vegetation plot record (relevé) databases. Plant species observations were aggregated to an average indicator value of a vegetation plot. Even though this approach has been criticized because it is mathematically incorrect for ordinal values, it is common practice in vegetation ecology to calculate average values (Diekmann 2003; Käfer and Witte 2004). Average indicator values are said to reflect the site conditions better than the indicator values of individual species, because the occurrence of a species in a relevé may deviate from its optimum due to ecological tolerance (Kowarik and Seidling 1989; Käfer and Witte 2004). The calculation of averages weighted by the logarithm of species abundance has been recommended (Böcker et al. 1983; Kowarik and Seidling 1989; Schaffers and Sýkora 2000; Käfer and Witte 2004). However, Reger et al. (2011) and Häring et al. (2013)

obtained better modeling results with unweighted mean values for predictive vegetation mapping compared to weighted averages.

Like all kinds of Gaussian regression models, calculating *average* indicator values and using statistical methods that estimate the conditional mean value, requires an approximately Gaussian distribution of indicator values. Otherwise, this approach will lead to wrong ecological estimates. However, this is not always the case (Ewald 2007). Besides unimodal Gaussian distributions, indicator spectra of the vegetation plot records from the Bavarian Alps provided by the WINALPecobase (GIVD-code EU-D-003, Dengler et al. 2011; Reger et al. 2012) are often skewed, uniform or even bimodal (cf. Fig. S1 in Electronic Supplementary Material).

Calculating mean indicator values on such data as well as modeling them based on a set of environmental predictors may yield misleading estimates of site-ecological conditions.

In this study, we propose a proportional odds model as an alternative approach to predict species indicator values, which, instead of the average indicator value, estimates the probability distribution over the range of indicator values for a given set of environmental gradients. We hypothesize that using a multivariate regression model provides a more realistic estimate of site-ecological conditions. We want to analyze whether spatial predictions of mean reaction values with a Gaussian regression model differ from predictions of actual distributions of indicator values. Using the mboost regression framework (Hothorn et al. 2011) we can estimate a mean indicator value (Gaussian regression model) as well as an ordinal value distribution (proportional odds model) with the same set of vegetation data and environmental predictors.

Materials and Methods

Study Area

The study area comprises the Bavarian Alps, a high mountain range in Southern Germany with an area of ca. 4,600 km^2 (Fig. 1). The mountain range is characterized by a long altitudinal gradient from the Saalach valley (470 m a.s.l.) to Germany's highest mountain (Zugspitze 2,962 m a.s.l.). The climate is cool and humid with mean annual temperature decreasing from 9°C at lower elevations to −4°C at the summits and annual precipitation ranging from 1,175 mm up to 2,800 mm. Due to geological diversity, the Bavarian Alps are characterized by a rich mosaic of soil types. Thus, limestones give rise to lithic as well as rendzic leptosols, histosols and terrae fuscae, whereas cambisols, stagnosols and gleyosols are widespread on sandstone, mudstone and marl (classification according to WRB 2007).

Vegetation Data

Our study was performed using vegetation plot records from the WINALPecobase (Reger et al. 2012). This database consists of 1,505 vegetation plot descriptions located in the Bavarian Alps, which were collected in 2009 (Fig. 1). Ellenberg indicator values were assigned to 904 vascular plants and 238 bryophytes. In sum, WINALPecobase consist of 57,186 individual plant records. The database is an ecological repository for forest vegetation with concomitant soil profile descriptions.

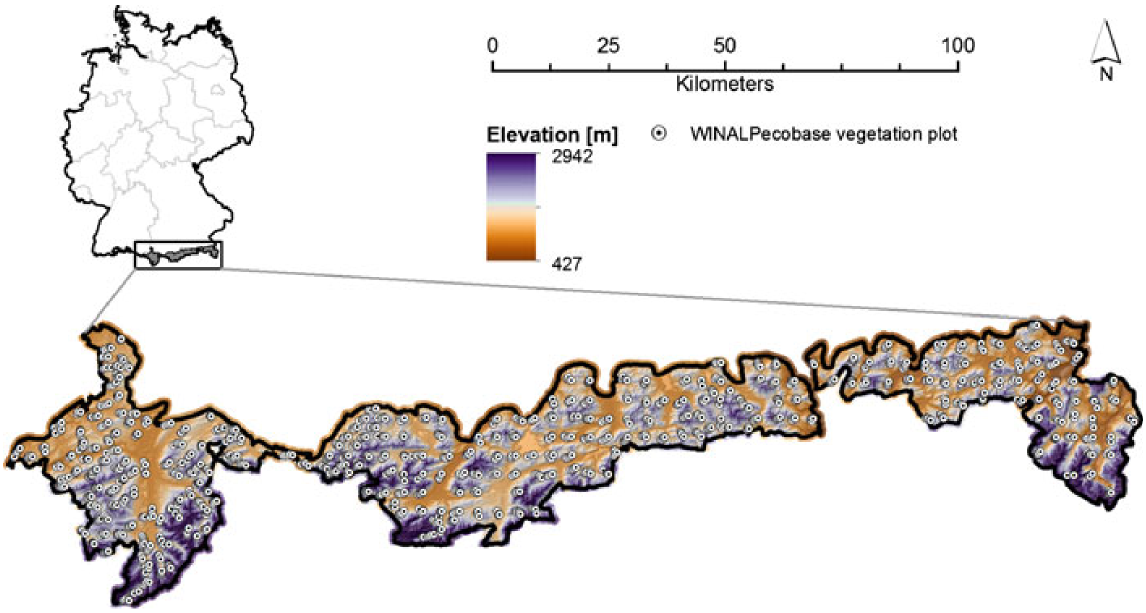


Fig. 1 Study area and location of the 1,505 vegetation plots. The sketch map indicates the location of the study area in the south of Germany

Detailed description of the database, vegetation analysis, and sampling design can be found in Reger et al. (2012).

To test our hypothesis, we focus on Ellenberg value for soil reaction (R-value), because it has a more heterogeneous distribution over its range compared to other indicators in our database. These indicator values are part of an expert system that ranks all Central European plant species according to empirical knowledge on an ordinal nine-point scale. The diverse geological environment in the study area gives rise to a broad range of R-values (1 to 9) and to corresponding heterogeneous vegetation patterns.

Following common practice (Schaffers and Sýkora 2000; Käfer and Witte 2004), Ellenberg values of all species in the plot were averaged by calculating log-abundance weighted or unweighted averages of indicator values for soil reaction based on vascular plants only or based additionally on bryophytes. The four variants of average indicator values for reaction were highly correlated (ranging between r=0.94 and r=0.98). Therefore, we restrict our analysis to the unweighted mean value calculated on vascular plants only, because this average value showed best modeling results in previous studies based on WINALPecobase (cf. Reger et al. 2011; Häring et al. 2013). Average R values range from 1.6 to 8.2 (mean = 5.91, median = 6.1, standard deviation = 0.98, cf. Figs. 1 and 10a).

Predictor Variables

We used eleven spatially explicit predictor variables reflecting meaningful ecological information to predict R-values, which are summarized in Table 1.

Most predictors were soil chemical parameters such as pH value ($CaCl_2$), cation exchange capacity, depth of decalcification, and storage of carbon and nitrogen. In addition, the fraction of sand in the top 1 m is used as a soil parameter. All soil parameters were extracted from a 1 : 25,000 soil map and a soil profile database. The soil map consists of 201 different soil map units to which representative soil profiles with analytically derived soil attributes were assigned. Therefore, the GIS layers of soil properties consist of discrete map entities and not of interpolated continuous surfaces.

Table 1 Summary statistics of predictor variables

Variable [unit]	Min	Median	Max	Standard deviation
Temperature [°C]	7.68	11.76	15.59	1.45
Slope gradient [°]	0.15	25.96	65.05	11.09
Convergence Index	48.95	99.28	165.50	20.05
Topographical Wetness Index	5.10	8.10	15.50	1.72
pH (topsoil)	2.92	6.27	7.62	1.39
pH (humus)	3.30	4.17	4.80	0.36
Depth of decalcification [cm]	0.00	21.00	250.00	47.61
Sand fraction [%]	0.00	34.60	73.17	13.40
Cation exchange capacity [kmol ha^{-1}]	81.78	477.55	2002.38	271.59
Carbon storage [t ha^{-1}]	30.00	31.96	38.87	1.52
Nitrogen storage [t ha^{-1}]	1.20	1.30	1.70	0.08

Mean temperature in the vegetation period (May to September) was used as a climatic parameter. It was derived from monthly climate maps of Bavaria with a spatial resolution of 50 m (Hera et al. 2012).

Three terrain parameters were derived from a digital elevation model (DEM) of the Bavarian Topographical Survey with a resolution of 10 m: 1) The slope summarizes the influence of gravitational processes. It was calculated according to Zevenbergen and Thorne (1987). 2) To account for interactions between soil reaction and soil moisture (Pakeman et al. 2008), we used the topographical wetness index (according to Böhner et al. 2002) as a measure for soil moisture conditions. 3) As an approximation of the small-scale variability of the terrain we use the convergence index, which gives a measure of how flow in a cell diverges or converges.

This set of predictor variables was assigned to the 1,505 vegetation plots for model calibration and validation.

Statistical Modeling

Due to the complex nature of ecological data, including nonlinear relationships, nonstationarity, temporal or spatial autocorrelation, and non-normal errors, ecologists generally apply flexible modeling approaches to address these complexities. Machine learning algorithms have become very attractive in predictive vegetation mapping (Franklin 2010), because they achieve higher prediction accuracies compared to standard regression models (Elith et al. 2008). They also incorporate methods to address overfitting and variable selection. However, their drawback is the lack of interpretability. Machine learning methods, like ensemble tree methods, are mostly black-box models, in which the predictor-response dependencies are hard to detect.

We apply a recently developed modeling technique called mboost (model-based boosting, Hothorn et al. 2010) that combines two modeling approaches: Complex prediction models for fitting additive or linear models that preserve the easy interpretability of predictor-response relationship, and also the increased prediction accuracy of a component-wise functional gradient descent boosting algorithm to estimate the model parameters (Maloney et al. 2011). The method has shown promising results for ecological modeling (Hothorn et al. 2011; Schmid et al. 2011; Häring et al. 2013).

This study aims to analyze the differences in the spatial prediction of an average R-value (model 1) to a multivariate prediction of the probability distribution (model 2). The first model follows the conventional approach of modeling indicator values. We apply a boosted geo-additive regression model given by

$$E(Y|x, s_1, s_2) = f_1(x_1) + \ldots + f_q(x_q) + f_{spatial}(s_1, s_2)$$

where Y denotes the averaged R-values on each of the 1,505 vegetation plots; $x_1,\ldots x_q$ are the explanatory variables, $f_1(x_1)+\ldots+f_q(x_q)$ are smooth functions of continuous covariates such as slope gradient or topsoil pH value, $f_{spatial}$ (s_1, s_2) is a spatial effect defined upon the x- and y-coordinates (s_1, s_2). In mboost, smooth functions were defined as P-spline functions, which can model the (possibly non-linear) response of mean R-values to environmental predictors. The model was fitted by minimizing the squared error risk that leads to the conditional mean of a numeric response. The spatial effect was added to the model to account for unobserved heterogeneity at the sampling localizations and thereby the spatial correlation in the data (Kneib et al. 2009).

Model 2 has been adopted from Schmid et al. (2011), who analyzed the influence of watershed characteristics on stream biological conditions. Therefore, they present an mboost model for regressing an ordinally scaled index of biotic integrity, IBI, on watershed-specific explanatory variables. The model is based on the proportional odds model for ordinal outcomes and is therefore appropriate for modeling ordinal species indicator values (McCullagh 1980). For each location in the study-area (i.e., every raster cell) with a specific combination of predictor variables, the model estimates a set of threshold values for the nine R-values. It is given by

$$P(Y \leq j|x, s_1, s_2) = logit^{-1}\left(\theta_j + f_1(x_1) + \ldots + f_q(x_q) + f_{spatial}(s_1, s_2)\right)$$

where Y is the Ellenberg R-value with j=1,…8 ordered categories, θ_j estimates the discrete distribution, which is adjusted on the logit scale through f.

A detailed description of the model can be found in Schmid et al. (2011). In contrast to model 1, here we do not use averaged R-values as response, but every single plant species with an assigned R-value in WINALPecobase. Thereby, the number of data points for modeling increases to 44,017.

Statistical modeling was carried out with the mboost package (version 2.1-1, Hothorn et al. 2010) for R (version 2.14.0). Spatial predictions were made with the R-package raster (1.9-55). Terrain analysis was conducted with SAGA GIS (version 2.0.8).

Model Validation

Model validation was conducted based on random sampling of the 1,505 vegetation plots. For model 1 we used two-thirds of the data as calibration data and one-third as an independent validation dataset.

Estimation of model performance of model 2 is less straightforward than the well-known R^2 value of model 1, because model 2 is more complex due to its multivariate output. We also create two subsamples of the data. Samples were chosen based on vegetation plots. For model calibration we use vegetation records belonging to two-third of the available vegetation plots. All vegetation records belonging to the remaining one-third vegetation plots were used for model validation.

Validation of model 2 is based on the model risk. The risk is calculated as the negative out-of-sample log likelihood of the proportional odds model. Typically, boosting is stopped before convergence to avoid overfitting the data and improve prediction accuracy (early stopping strategy, cf. Hastie et al. 2009). We used a ten-fold cross-validation to choose the best number of boosting iterations, i.e., the iteration with the lowest empirical risk. In mboost this is defined as the *mstop* parameter.

Comparison of Spatial Predictions

Because of the multivariate output of model 2, comparing the spatial patterns of soil reaction extracted by the two models is complicated. We propose different alternatives to solve this problem.

First, the spatial pattern of both predicted maps can be compared qualitatively and visually. However, the visual comparison of a single map of predicted mean R-values of model 1 with a set of nine different maps of predicted probabilities of model 2 is unhandy. Therefore, we aggregate the stack of predicted R-values of model 2 to a single map by using a RGB (red-green-blue) composite as used in remote sensing for multi-spectral data (cf. Hengl et al. 2007 for an example of a RGB soil map). All nine maps of predicted probabilities for R1,…,R9 were transferred into a 3D-color space to get a single RGB map that illustrates the properties of the probability distribution. Therefore, the predicted probabilities have to be aggregated into three values. To account for the left-skewed distribution of the R-value (Fig. **10**), we do not use the obvious aggregation of R = 1+2+3, G = 4+5+6 and B = 7+8+9, but in a modified fashion: R = 1+2+3+4, G = 5+6 and B = 7+8+9. Using this RGB map, a direct visual, yet still qualitative comparison with the map of predicted mean R-values of model 1 is achieved.

To quantify the accordance of both predictions, we fitted a multiple linear regression, in which the predictions of model 1 (average R-value) act as dependent, the output of model 2 as predictor variables (P(R=1), P(R=2),…,P(R=9)).

Results

Model Validation

Model validation for model 1 is based on the R^2. We used the calibration data (two-thirds of the entire dataset) to fit the regression model and calculated an R^2 between observed and predicted mean R-value of the validation data set. Figure **2** shows the predicted mean Ellenberg value for soil reaction plotted against the observed values for the calibration and validation data.

We get a R^2 of 0.42 on the independent validation data. The plot indicates promising correlation for indicator values between five and eight. However, values smaller than 5 were overestimated with the model.

Validation of model 2 is based on the negative out-of-sample log likelihood of the proportional odds model. The development of model risk over the number of boosting

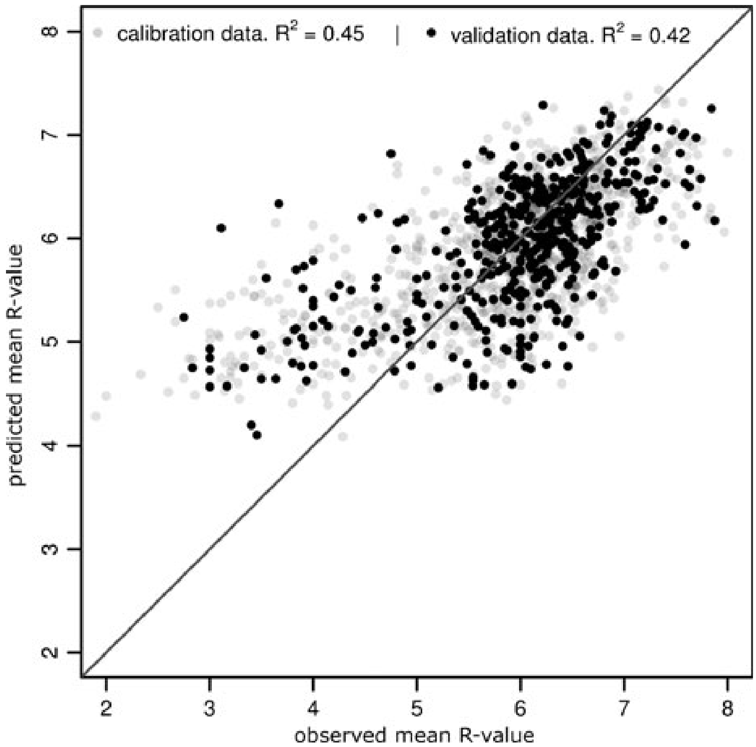


Fig. 2 Predicted *vs* observed mean Ellenberg indicator values for soil reaction with the Gaussian model (model 1). Calibration data consist of 1,019 vegetation plots of the WINALPecobase, validation data are the remaining 1/3 of all plots (471 plots)

iterations for model 2 is illustrated in Fig. S2 in Electronic Supplementary Material. The model shows the best fit with 2,500 boosting iterations. If more boosting iterations are performed, the model risk increases again, which is an indication of overfitting.

Response Curves

Partial dependency plots (Fig. 3) show the relationship between mean R-value ("Gaussian") and R-value spectrum ("PropOdds") respectively and the predictors. The response curves look quite similar for both models with only minor differences.

Spatial Effect

Spatial effects, which were included in the models in addition to the environmental covariates, are illustrated in Fig. 4. In both models the effects are significant. We can conclude that there are spatial variations in the depicted relationships that could not be explained by the predictor variables.

Prediction

Model 1 as well as model 2 were used to generate maps of predicted mean R-value (model 1) and probabilities for R-value 1 to 9 (model 2). Prediction was made on 10-m raster data that were available for the entire study area (Figs. 5 and 6).

The resulting maps show a detailed and diverse spatial pattern of soil reaction in our study area. Compared to the soil attribute maps, which serve as the only area-

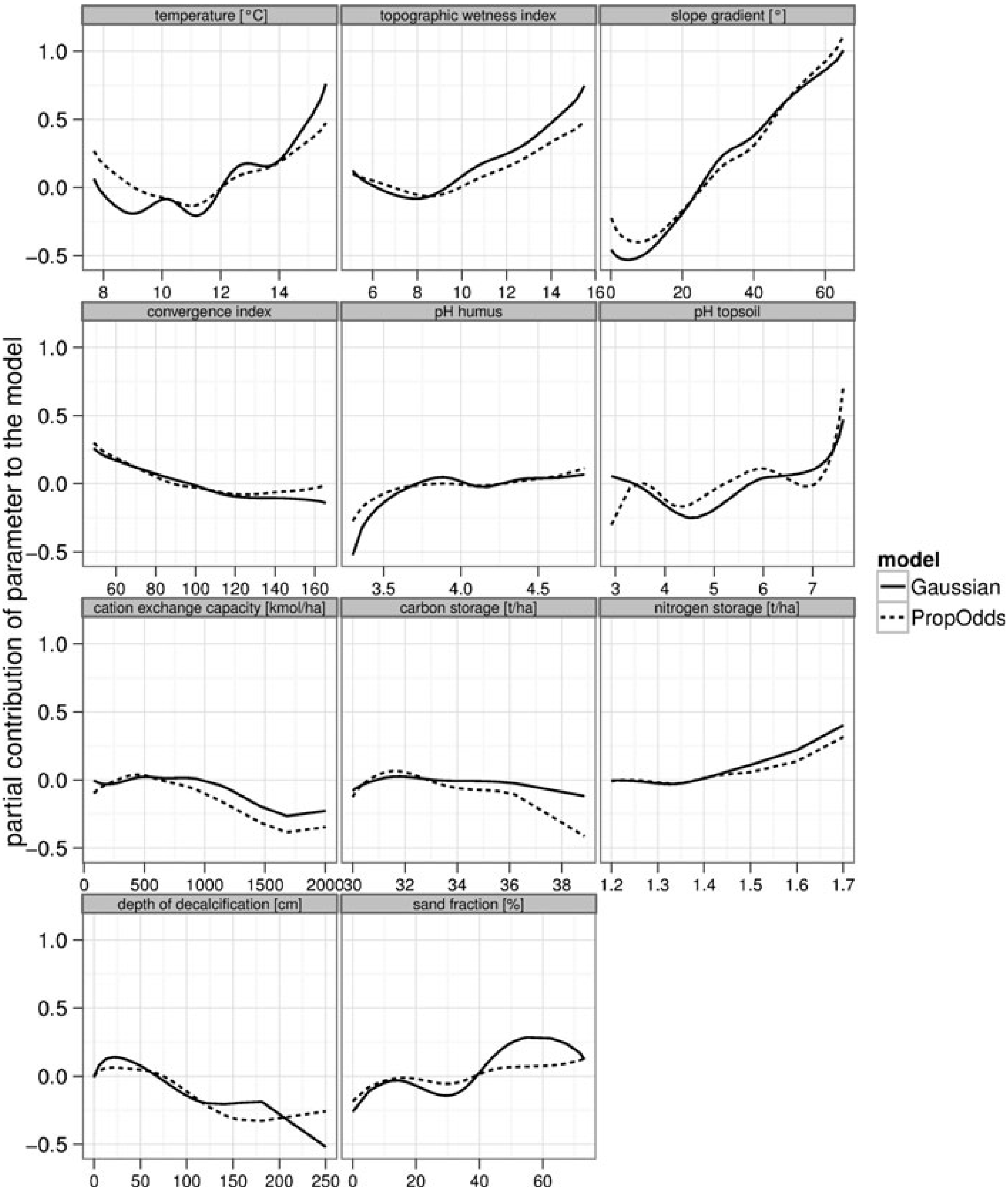


Fig. 3 Partial response curves of all environmental covariates for both models. Model 1 indicates the Gaussian regression model, and model 2 indicates the proportional odds model. The higher the partial contribution, the stronger the impact is on the predicted R-values

wide source of information on soil reaction up to now, our resulting maps show a tremendous improvement with a much higher spatial resolution.

Comparison of Models

The comparison of both models was done in different ways, both qualitative-visual as well as quantitative. The main interest is to analyze whether the spatial pattern of soil reaction in both maps differs or not, e.g., if sites with high values for mean soil reaction have high probabilities for calcareous R-values and *vice versa*.

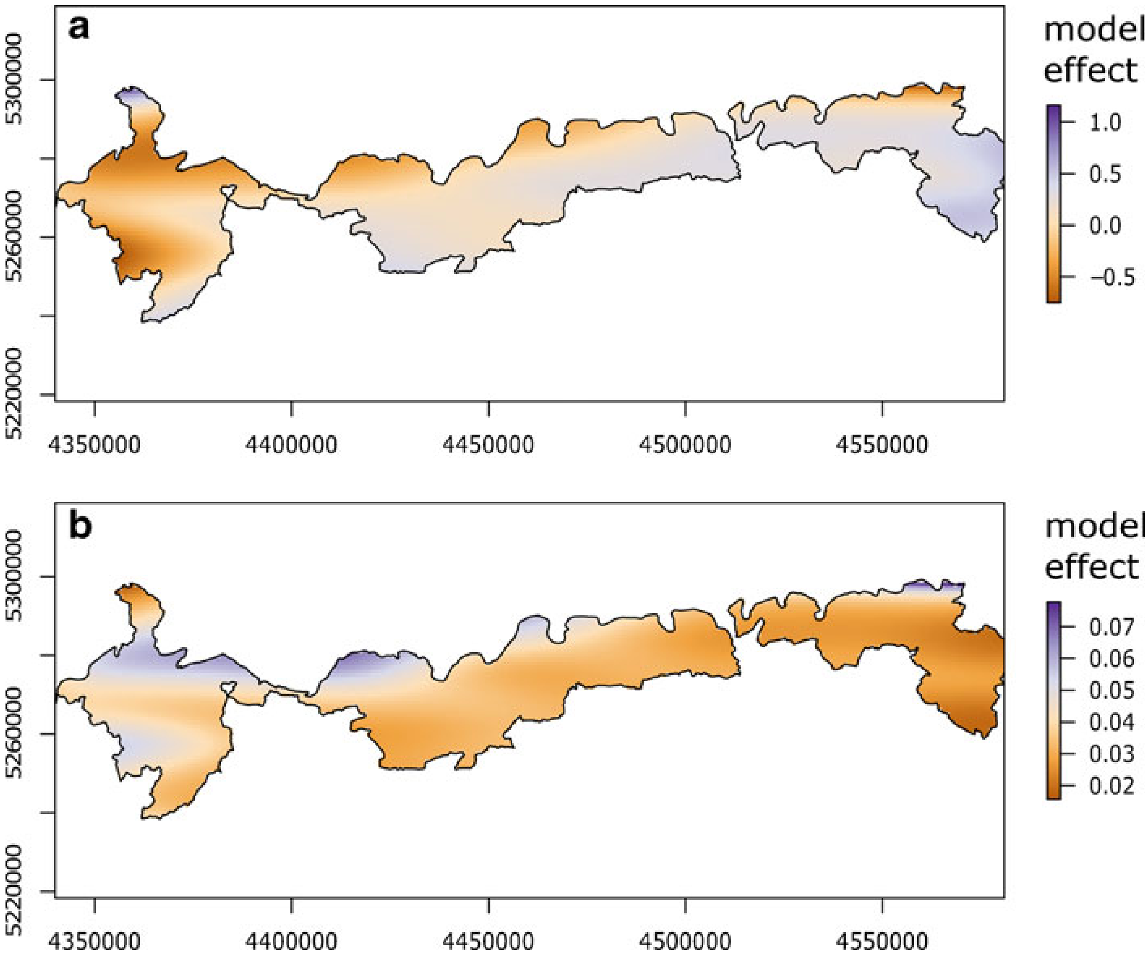


Fig. 4 Marginal spatial effect for Gaussian regression model (model 1; **a**) and proportional odds model (model 2; **b**). Color indicates the contribution of the spatial effect to the model. Positive values indicate higher R-values, negative contribution indicate lower R-values

First, we draw 1,000 randomly selected raster cells in our study area and plot the cumulative sum of the probability distribution of model 2 (Fig. 5), i.e., in each case the probability distribution of a specific raster cell in all nine maps in Fig. 6. With this curve we can visualize the shape and the range of the probability distribution. The curves are located between the two extreme values: The upper curve indicates a more right-skewed distribution and the lower curve indicates a left-skewed distribution. If the distribution tends to be uniform, the curve becomes linear.

To compare the predictions of both models, we map the predicted mean R-value at a specific raster cell with colors to the plot. The color gradient between the different lines closely follows the shape of the distribution – from light green in the upper part to blue in the lower. We can see many blue lines with steep slope from R-value 6 to 7 and even more from 7 to 8, i.e., raster cells with a left-skewed distribution, which can be found in calcareous areas, as well as many lines in light green with a steep slope from R-value 1 to 2 or 3 to 4. We can conclude that the spatial pattern of predictions of both models is very similar. This was also indicated by the same shape of the partial response curves (Fig. 3).

Nevertheless, there are indeed some deviations from this pattern. Blue lines in the upper part or light-green lines in the lower indicate differences in predicted soil reaction conditions.

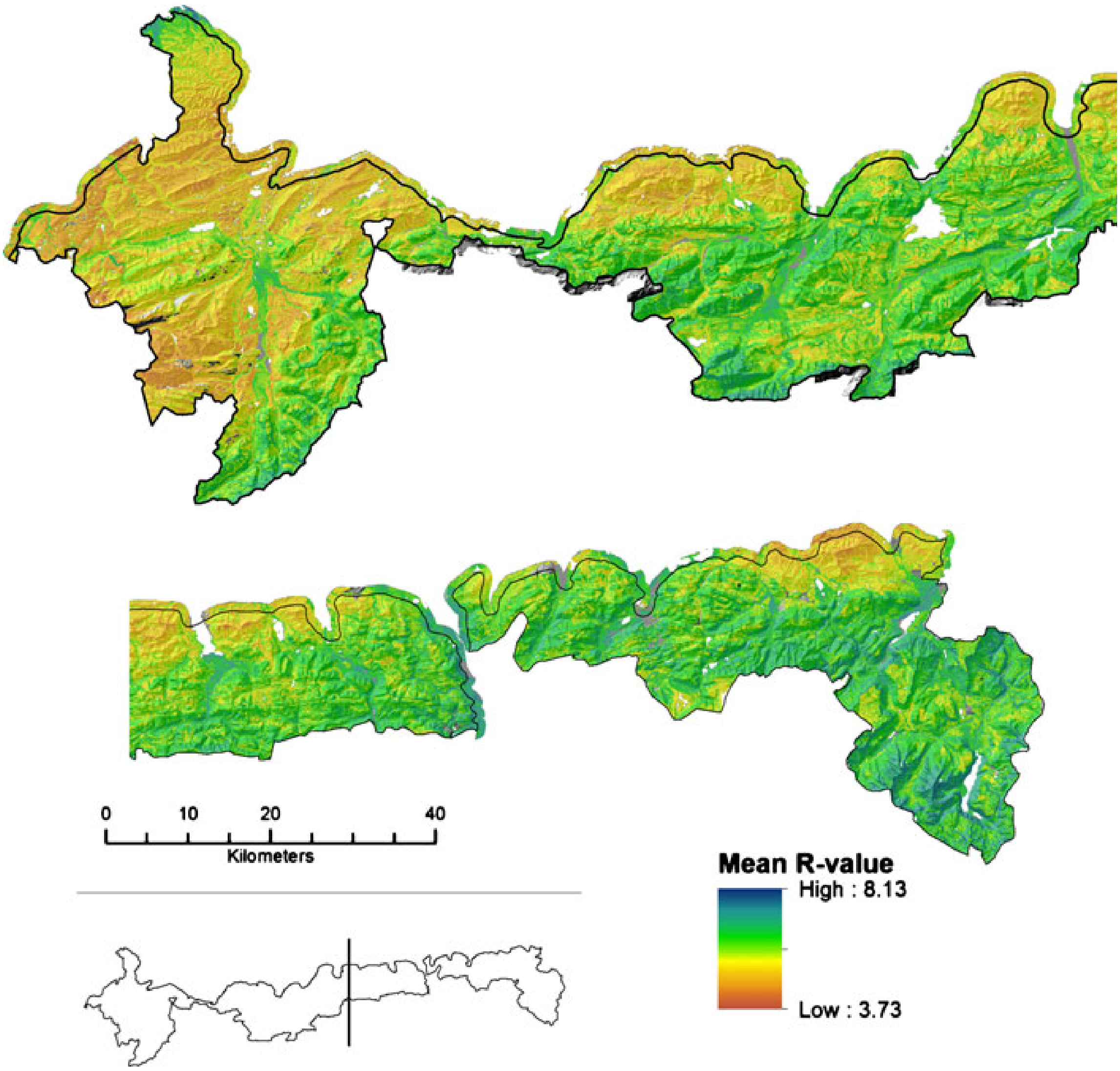


Fig. 5 Map of predicted mean R-values with Gaussian regression model (model 1) on a 10-m grid. Because the study area is a small strip in E-W direction, the map was divided in two parts (indicated with the small sketch map)

Although the spatial pattern is very similar, the predicted ranges of R- and mean R-values differ considerably, as also noted above.

For a visual comparison of the spatial predictions on both models we transferred the multivariate output of model 2 into a RGB map (Fig. 8).

As in Fig. 5 also the RGB map shows a diverse pattern of Ellenberg reaction value in the Bavarian Alps. Moreover, the pattern looks quite similar reflecting the large-scale (geological and geomorphological conditions) as well as the fine-scale spatial distribution. Therefore, as discussed above, we can conclude that the spatial pattern of predicted Ellenberg values for soil reaction of model 1 and model 2 are in high accordance. However, the big difference is the multivariate characteristic of the RGB map. Large areas in the map are displayed in a mixture of red, green and blue value (colors somewhere inside the RGB triangle), meaning that the probability distribution covers a wider range. There is a higher information depth in RGB map compared to the map of mean R-values in Fig. 5 even though the nine probabilities were aggregated to three values.

To get a quantitative estimate of the agreement of both predictions we fit a multiple linear regression model that explains the variance of average R-values with the

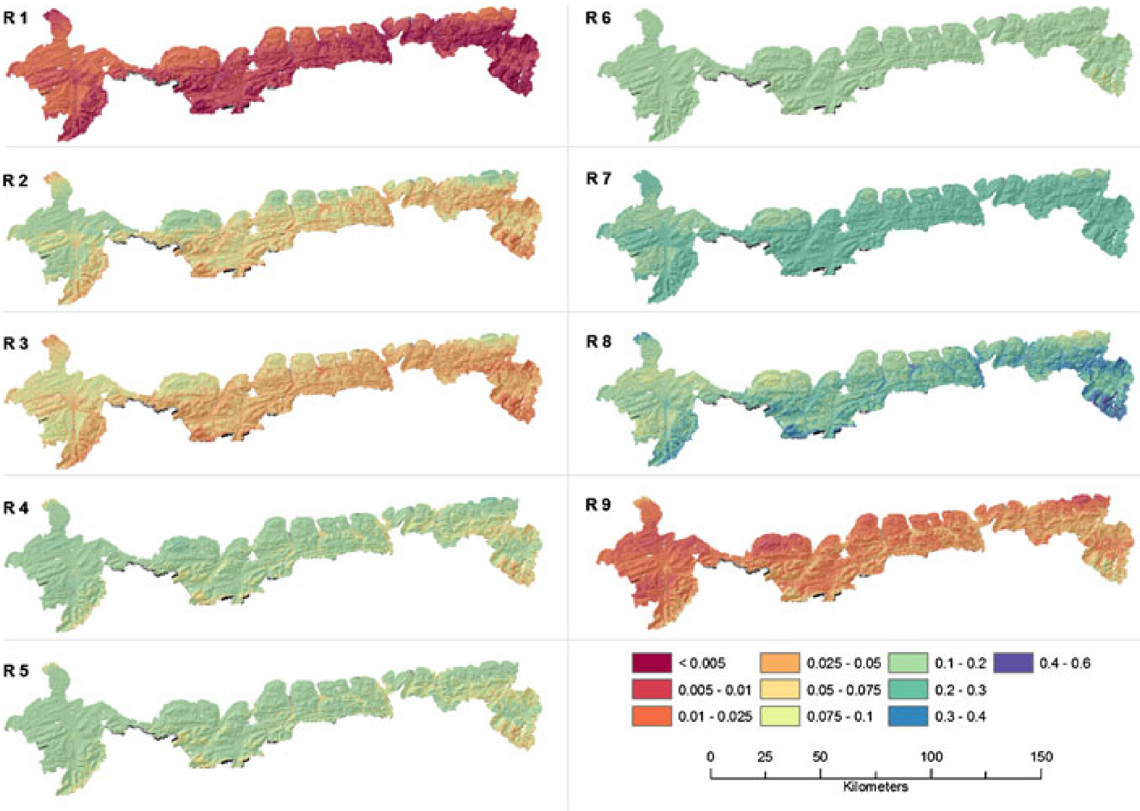


Fig. 6 Maps of predicted probabilities for R value 1 to 9. Probabilities were calculated with the proportional odds model (model 2)

predicted probability distribution of model 2. The model was estimated using every raster cell of the study area (>2.5 million) to get the most realistic estimate of correlation. The regression model shows a R^2 of 0.93.

The tight correlation between regionalized averaged R values and predicted R spectra indicates that the information of averaged R values is actually contained in the spectra. Similarly, Ewald (2007) showed that the average indicator value for nutrients (N) corresponded to the first principal component in an ordination of the spectra, whereas the second component represented the modality of the spectrum. While bimodal N value spectra represent ecosystem dynamics, bimodal R spectra are characteristic of cold temperate coniferous forests on calcareous bedrock with extremely steep vertical pH gradients and high lateral variation in pH (Ewald 1999).

While to date such complex patterns had to be inspected in individual field plots, modeling proportional odds allows projecting them into the landscape.

Discussion

We get promising correlation for model 1 with a R^2 of 0.42 on the independent validation data. However, we found the best fit only for indicator values between five and eight. However, values smaller than 5 were overestimated with the model. This may be due to a small sample of low R-values (10 %-quantile of mean R-value = 4.5 in our dataset). Because our dataset is a representative sample of the study area, areas with low R-values have a limited extent mainly in the northwestern and northern part

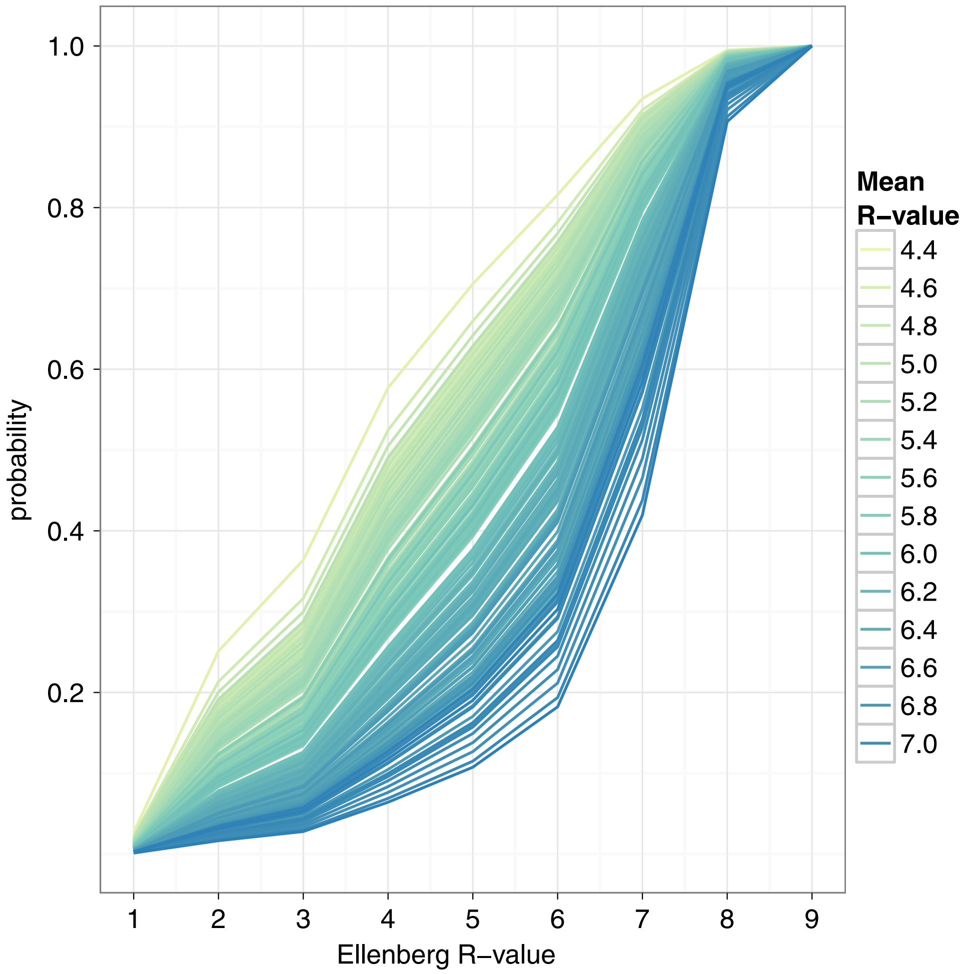


Fig. 7 Cumulative sum of the probability distribution of model 2 of 1,000 randomly selected raster cells. Coloring of the plot indicates predicted mean R-values, i.e., each line represents one raster cell. The x-axis indicates Ellenberg R-values, and y-axis indicates the cumulative probability for R-values

of the study area. Most of the study area is located in the sample range of 4.5 to 7.5 for mean R-value (cf. Fig. 2), where we can assume to get sound predictions.

The relationships between predictor variables and dependent variables depicted in partial dependency plots (Fig. 3) reflect the ecological processes that determine the spatial distribution of soil reaction. Lower temperatures in the vegetation period have a negative influence on R-values. With increasing temperature the partial effect becomes positive. In colder areas, which also have a short vegetation period, mineralization of organic matter is inhibited, thus favoring duff accumulation and leaching of topsoil. Obviously, there is a strong influence of topographical characteristics on the (mean) R-value, because the magnitude of the partial response curves for slope gradient, topographical wetness index and convergence index is higher compared to the soil parameters. Slope gradient has the highest range of partial effects. In areas with slope gradients up to 25° the effect is negative in both models. If slope gradient becomes higher, the partial effect increases linearly. There is a strong relation between soil reaction and gravitational processes like soil erosion and rock fall, especially in mountainous regions, because mineral soil acidification is counteracted by morphodynamics. Steep slopes also inhibit accumulation of thick acidic duff layers (Bochter 1984). This is also indicated by the positive response to topographical wetness index where high values have a positive effect on R-values. High values of

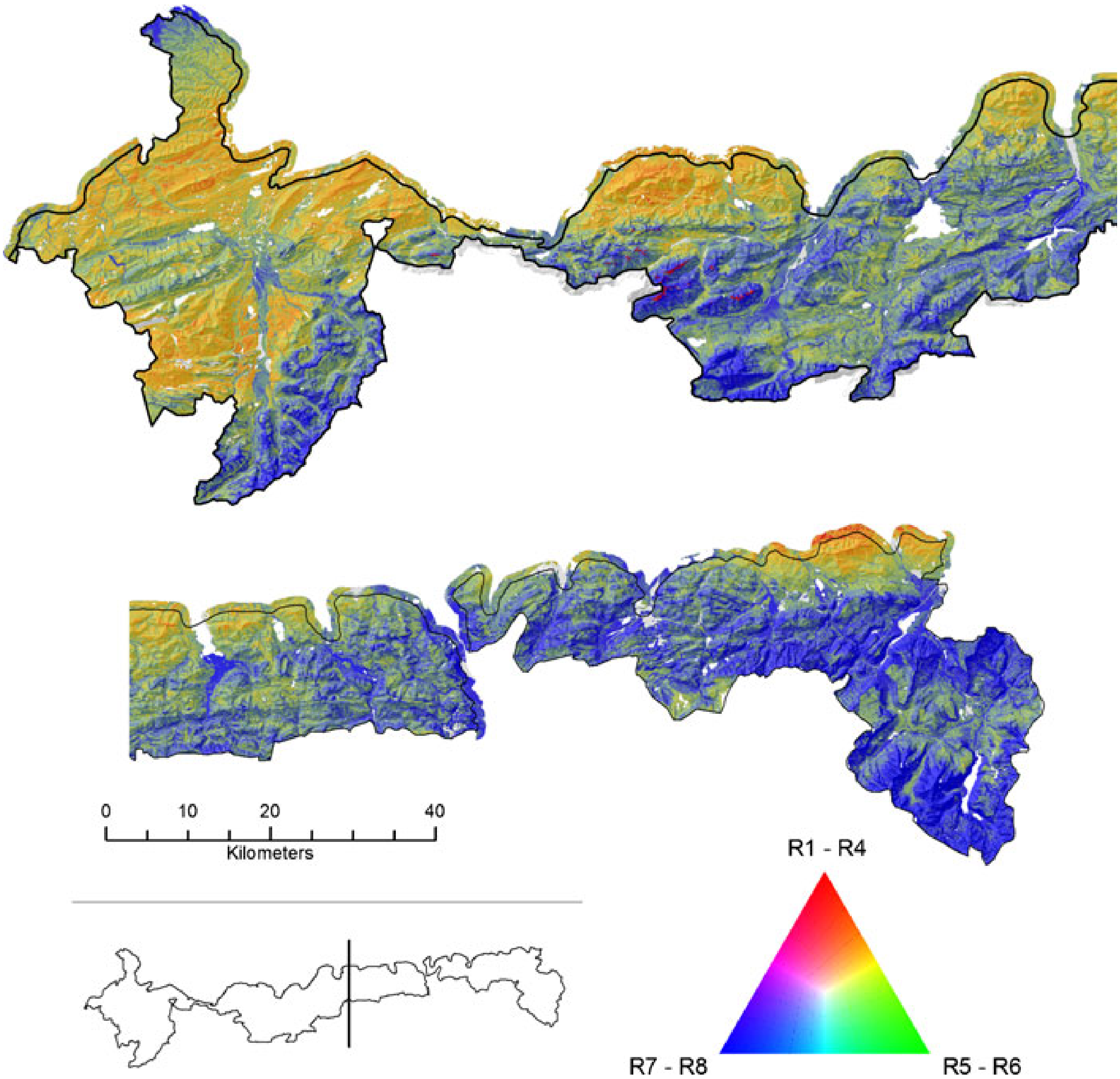


Fig. 8 Red-green-blue map of predicted probabilities for Ellenberg value R1,…,R9 calculated with model 2 (proportional odds model)

topographical wetness index indicate areas where surface water and eroded soil material accumulates. This soil material consists mainly of fine topsoil and humus which is rich in base cations. Areas with high values for convergence index, i.e., exposed landscape positions like ridges, have lower R-values whereas lower values have a positive effect, which is also an indication of mass balance.

The relationship between pH of humus layer and topsoil, and R-value is obvious because the R-value is an approximation of soil reaction: The higher the pH value the more positive its influence. Consequently, the influence of pH humus layer is only negative or zero because the value ranges only up to 4.8. However, the relationship appears to be curvilinear, which was also reported by Schaffers and Sýkora (2000). They argue that this is probably caused by the narrow pH tolerances of species with high pH. The relationship between depth of decalcification and R-value is also clear, because carbonate buffers soils at high pH. However, the partial dependence of cation exchange capacity is somewhat contradictory to soil chemical processes, because its effect becomes negative with increasing value even though cation exchange capacity and soil pH are positively correlated. This could be due to the coarse resolution of the soil map.

The marginal spatial effects for both models illustrated in Fig. 4 show similar spatial patterns with a gradient from north to south. Petrographic and geomorphological conditions in the study area can clearly be distinguished, e.g., the limestone rocks in the southern part or the Flysch facies in the north and the molasse basins in the west. However, apart from this similarity there are large differences in the maps. Bearing the high similarity of the response curves in mind (Fig. 3) this is of particular importance. The spatial effect for model 1 is positive in the central and eastern part of the study area, the northern and western part is negative. The effect for model 2 is nearly the opposite. Highest values can be found in the north and west whereas low values are localized in the south and east. However, here the effect in the entire study area is positive. Regarding the magnitude of the effect there is also a large difference. For model 1 the effect ranges from −0.75 to 1.16, but for model 2 only from 0.016 to 0.078.

The spatial effect of model 1 follows an expected spatial pattern. Mean R-values tend to be higher in areas with limestone whereas in areas with flysch (marginal area at northern boundary) or molasse (western part) sandstones R-value tends to be lower. Because of this and because of the high magnitude of the effect we conclude that the spatial effect is of high importance for model 1, which is not the case for the proportional odds model.

To have a closer look at this topic, we calculated R^2 values for model 1 with and without the spatial effect $f_{spatial}$ (s_1, s_2) (cf. section Statistical Modeling). We fit the models and calculated the R^2 on the independent validation data (one third of the entire dataset). To get a stable result we repeated this procedure 50 times. Figure 9 illustrates the importance of the spatial effect for model 1. The mean difference between R^2 of model 1 with spatial effect and model 1 without spatial effect is 0.05. There is a none-negligible unexplained deviance in model 1 when modeling mean R-value without spatial effect. Environmental variables could not explain averaged mean R-value as detailed as model 2 could for the ordinal outcome.

The influence of spatial information in species distribution modeling has been discussed widely because biogeographical patterns appear to vary as a function of space on different scales (Foody 2004; Graham et al. 2008; Hothorn et al. 2011; Miller and Hanham 2011). Untreated spatial information may not only lead to a lower model performance (Foody 2004; Hothorn et al. 2011) but also to biased estimates (Legendre and Legendre 1998; Lennon 2000; Kühn 2007). Therefore, applying approaches accounting for spatial dependencies – such as mboost – has become an important enhancement in statistical ecology (Lichstein et al. 2002; Foody 2004; Dormann et al. 2007; Miller et al. 2007; Hothorn et al. 2011; Miller and Hanham 2011).

The difference in Fig. 9 may reflect missing predictors (Schmid et al. 2011). However, because the difference is not present in the proportional odds model, it more likely reflects problems inherit in averaging indicator values. Useful information on site-ecological characteristics may get lost by averaging R-values within a vegetation plot. The shape and range of the R-values on a vegetation plot reflect a more sophisticated picture of vegetation response than an averaged value. Diekmann (2003) points out that sample plots should be as homogeneous as possible when calculating average indicator values otherwise the results may be misleading or even nonsensical. Especially for Ellenberg's R-value for soil reaction this may lead to problems in our study area. Similar to Diekmann (2003), who refers to heterogeneous

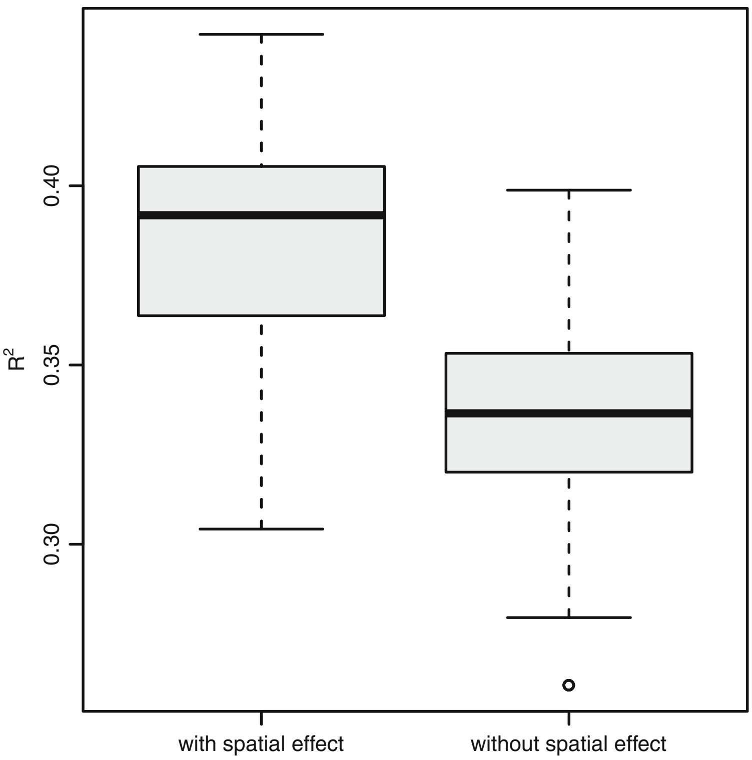


Fig. 9 R^2 values of 50 bootstrap samples of Gaussian regression model (model 1) with and without spatial effect. There is a significant contribution of the spatial effect to model 1. The box indicates the median (solid black line) and the 0.25 and 0.75 quantile, the whiskers indicates the min and max value of the distribution

indicator spectra in Fennoscandian pine forests, we study a large area with an acid humus layer on limestone bedrock where often a mixture of basiphilous and acidicolous species occur, which leads to wide and heterogeneous spectra (Ewald 1999; Ketterer and Ewald 1999).

Here, we found a clear benefit of the proportional odds model over model 1. Model 2 can exploit information about the range and shape of indicator spectra and explain these spectra with a set of environmental predictors precisely. Thereby we get a more detailed description of site-ecological conditions.

The resulting maps of predicted Ellenberg values for soil reaction in Figs. **5** and **6** show a detailed and diverse spatial pattern of soil reaction in our study area. Compared to the soil attribute maps that serve as the only area-wide source of information on soil reaction up to now, our resulting maps show a tremendous improvement with a much higher spatial resolution

Comparing the range and the shape of the predicted values with the mean R-values in the calibration dataset in Fig. **10**, differences can be observed. Compared to the vegetation plots in WINALPecobase (Fig. **10**, subplot a, range 1.6 to 8.2), the map's range of predicted mean R-values are contracted to a range from 3.7 to 8.1. Extreme values at the maximum and even more at the minimum could not be predicted well. Also the shape of the histograms reveals differences. Whereas the differences in shape results from the rich diversity of environmental conditions in our study area compared to the sample dataset, the absence of extreme values in the prediction raise concerns when using the map to draw conclusions on site-ecological characteristics. Using Gaussian regression models, which estimates the conditional mean value of the

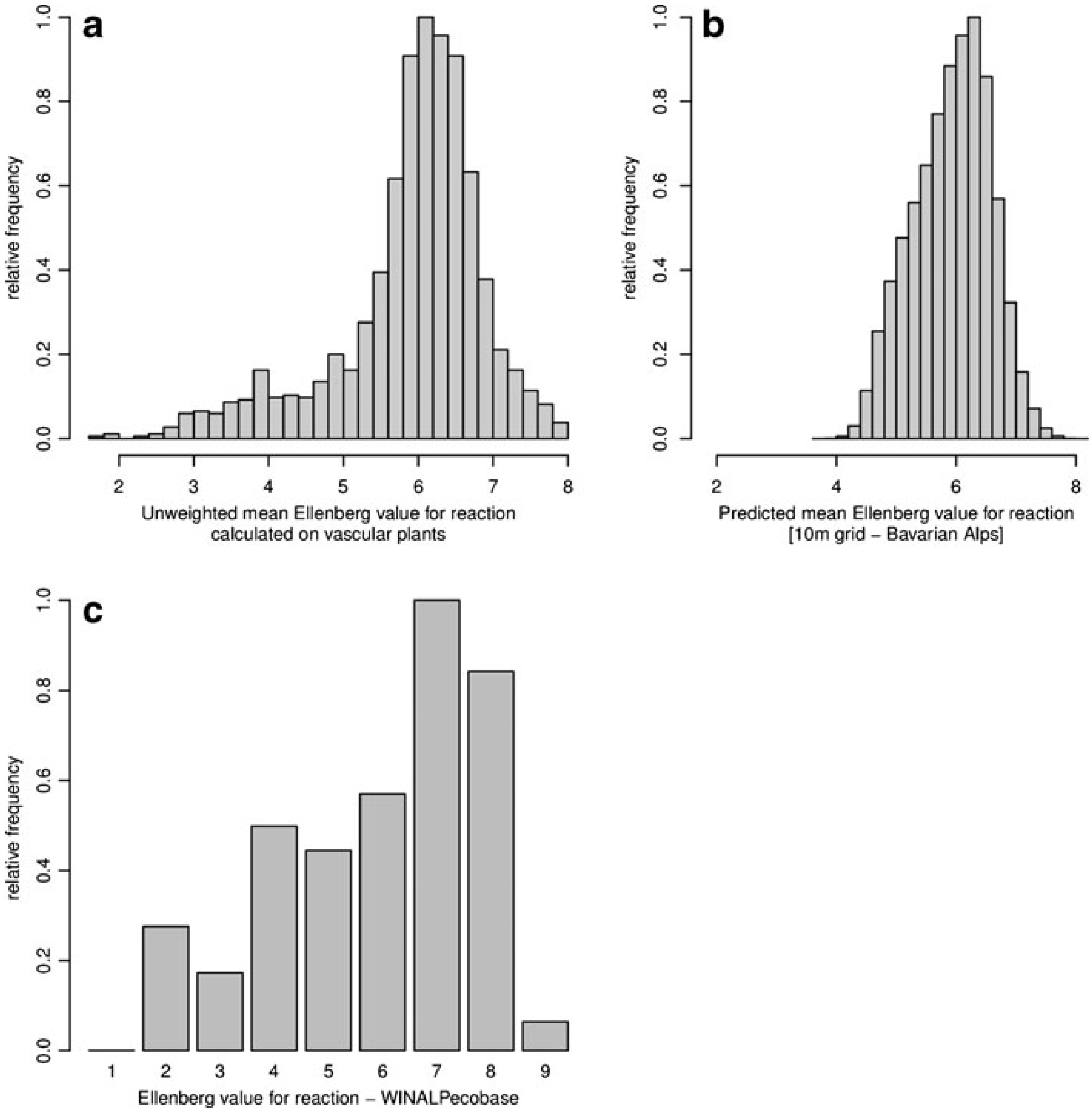


Fig. 10 Histogram of mean R-values of the 1505 vegetation plots (**a**) and the predicted mean R-value in the entire study area (**b**). Spectrum of R-values of all 44017 indicator species in the WINALPecobase (**c**) (Reger et al. 2012)

response, to estimate average indicator values leads to twofold convergence on the mean value and thereby to a reduction of the former value range – first in the calculation of the average indicator value (cf. Diekmann 2003) and second in the model itself. In Fig. **10** also the original distribution of R-values in the vegetation database (c) is illustrated (see also Ellenberg indicator value spectra of 36 randomly selected vegetation plots of WINALPecobase (Fig. S1 in Electronic Supplementary Material)). The twofold convergence in distributions when modeling average R-values can be tracked from subplot c) over a) to b).

The prediction of model 2 calculates the probability of a raster cell to contain indicator species with R-value from 1 to 9. These probabilities are plotted in Fig. **6**. Compared to a single averaged value in Fig. **10** the prediction of the proportional odds model reveals a more sophisticated picture of site-ecological characteristics because we get information on the range and shape of the probability distribution for each raster cell.

Conclusions

The preparation of spatially explicit and high-resolution information on site-ecological characteristics is crucial for sustainable management of ecosystems facing changing environmental conditions. Forest ecosystems are very likely to be influenced by climate change (Kellomäki and Leinonen 2005; Maracchi et al. 2005), especially in areas highly vulnerable to changing environmental conditions like mountainous regions (Parry et al. 2007). Ellenberg indicator values have proven to be an easy-to-use and helpful tool to achieve this purpose (Diekmann 2003). The application of statistical models to predict indicator values can be seen as a "new generation" of site classification and mapping (Reger et al. 2014, this issue). These sophisticated models can only be applied to large georeferenced vegetation databases (Ewald 2003; Guisan et al. 2007), where plots can be intersected with spatially explicit environmental data to calibrate models that then serve to produce maps based on area-wide predictors. The benefit from joining vegetation and geophysical data in one information system is mutual: vegetation can be much better modeled, whereas geophysical variables can be parameterized against a meaningful biological response. Representing the major physiological factors acting on plant growth, ecological indicator values present an obvious link between vegetation and environment.

We applied geo-additive regression models to predict Ellenberg indicator values for soil reaction on a 10 m-grid in the Bavarian Alps. The resulting maps seem very promising to support forest management in the Bavarian Alps in the future (cf. Reger et al. 2014, this issue). We analyzed, whether averaged indicator values are adequate to predict site conditions precisely. We found quite good results when predicting mean R-values. The map is a convenient tool for practical use and can easily be used for further analysis or modeling. The TRM-model (*T*emperature, *R*eaction, *M*oisture) of potential natural forest vegetation in mountain forests of Reger et al. (2014, this issue) is one example. However, users have to be careful when using predicted average R-values to draw conclusions on soil pH values, base saturation (Seidling and Rohner 1993) or calcium content (Schaffers and Sýkora 2000). We found a strong convergence to the mean value in comparison to the original range and shape of indicator spectra. Therefore, very high as well as very low mean R-values could not be predicted well. In addition, we found a strong influence of a spatial effect in our model 1. A large proportion of deviance could not be explained by environmental predictors. We conclude that averaging indicator values in areas with heterogeneous conditions regarding soil reaction like our study area could lead to biased estimates. On the vast majority of our vegetation plots we did not find bell-shapes spectra but more complex distributions. Averaging indicator values is therefore an oversimplification of the complex condition – and wrong although common practice.

To account for this issue, we present an alternative approach to predict indicator values. A proportional odds model was introduced to estimate the probabilities for all nine R-values, which means we get a multivariate prediction on every raster cell. To our knowledge this is the first approach to model and predict a spectrum of indicator values instead of an averaged value. By predicting a probability distribution over R-value 1 to 9 we get a much more sophisticated impression of soil reaction conditions. In contrast to model 1 (mean R-value) the range of the indicator spectrum was not reduced and also high and low R-values were predicted with high probabilities.

With our study we could show only one example of the application of the proportional odds model. However, a variety of different application areas in vegetation science or ecology in general are possible (e.g., Schmid et al. 2011). Thus, besides the Ellenberg value for soil reaction other indicator values exhibiting complex spectra (e.g., Ewald 2007) could be modeled more precisely with this approach. In addition, there are a lot of problems in ecology that deal with ordinal data, e.g., site and soil classification (Schaetzl et al. 2012) or measurements of biological entities in classes which are, e.g., “shorter”, “darker” or “more abundant” than others (Guisan and Harrell 2000).

However, the main drawback of the prediction of model 2 is that it is less easy to use compared to the single map of mean R-values. To prevent the use of the proportional odds model only to restricted and more sophisticated scientific applications we propose to plot the multivariate output in 3d color space by calculating a RGB composite. Similar to the predicted map of model 1, this map is a handy and helpful tool for practitioners and decision makers. In contrast to model 1, the predicted map of model 2 reflects more detail by taking the entire range and distribution of the R-values into account and is therefore an advantageous tool for forest managers as well as for predictive modelling.

Acknowledgments The research presented here forms part of the project “Forest Information System for the Northern Alps” (www.winalp.info), which is funded by the European Fund for Regional Development (EFRE) within the “INTERREG Bayern – Österreich 2007–2013” program, the Bavarian Forest Administration and the Bavaria State Forest Enterprise (BaySF). We are indebted to the Bavarian State Institute of Forestry’s “Maps for the future” crew. The digital elevation model was obtained from the Bavarian Topographical Survey (LVG), soil maps were obtained from the Bavarian State Agency for the Environment (LfU) and climate data were obtained from German Meteorological Service (DWD).

References

Bochter R (1984) *Böden naturnaher Bergwaldstandorte auf carbonatreichen Substraten.* Forschungsbericht 6, Nationalpark Berchtesgaden, Berchtesgaden

Böcker R, Kowarik I, Bornkamm R (1983) Untersuchungen zur Anwendung der Zeigerwerte nach Ellenberg. *Verh Ges Oekol* 11:35–56

Böhner J, Köthe R, Conrad O, Ringeler A, Selige T (2002) Soil regionalisation by means of terrain analysis and process parameterisation. In Micheli E, Nachtergaele F, Montanarella L (eds) *Soil classification 2001*. EUR 20398 EN, European Soil Bureau, Research Report No. 7, Luxembourg, pp 213–222

Dengler J, Jansen F, Glöckler F, Peet RK, De Cáceres M, Chytrý M, Ewald J, Oldeland J, Lopez-Gonzalez G, Finckh M, Mucina L, Rodwell JS, Schaminée JHJ, Spencer N (2011) The Global Index of Vegetation-Plot Databases (GIVD): a new resource for vegetation science. *J Veg Sci* 22:582–597

Diekmann M (2003) Species indicator values as an important tool in applied plant ecology – a review. *Basic Appl Ecol* 4:493–506

Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G, Davies RG, Hirzel A, Jetz W, Kissling D, Kühn I, Ohlemüller R, Peres-Neto PR, Reineking B, Schröder B, Schurr FM, Wilson R (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30:609–628

Elith J, Leathwick, JR (2009) Species distribution models: Ecological explanation and prediction across space and time. *Annual Rev Ecol Evol Syst* 40:677–697

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. *J Anim Ecol* 77:802–813

Ellenberg H, Weber HE, Düll R, Wirth V, Werner W (2001) *Zeigerwerte von Pflanzen in Mitteleuropa*. Verlag Erich Goltze KG, Göttingen

Ewald J (1999) Relationships between floristic and microsite variability in coniferous forests of the Bavarian Alps. *Phytocoenologia* 29:327–344

Ewald J (2003) A critique for phytosociology. *J Veg Sci* 14:291–296

Ewald J (2007) Bimodal spectra of nutrient indicators reveal abrupt eutrophication of pine forests. *Preslia* 79:391–400

Foody GM (2004) Spatial nonstationarity and scale-dependency in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. *Global Ecol Biogeogr* 13:315–320

Franklin J (2010) *Mapping species distributions: spatial inference and prediction*. Cambridge University Press, Cambridge

Graham CH, Elith J, Hijmans RJ, Guisan A, Peterson AT, Loiselle BA (2008) The influence of spatial errors in species occurrence data used in distribution models. *J Appl Ecol* 45:239–247

Guisan A, Harrell FE (2000) Ordinal response regression models in ecology. *J Veg Sci* 11:617–626

Guisan A, Spehn E, Körner C (2007) MRI Newsletter 8: Georeferenced biological databases – a tool for understanding mountain biodiversity. *Mount Res Developm* 27:80–81

Häring T, Reger B, Ewald J, Hothorn T, Schröder B (2013) Predicting Ellenberg's soil moisture indicator value in the Bavarian Alps using additive georegression. *Appl Veg Sci* 16:110–121

Hastie TJ, Tibshirani RJ, Friedman JH (2009) *The elements of statistical learning: data mining, inference, and prediction*. Ed. 2. Springer Verlag, New York

Hawkes JC, Pyatt DG, White IMS (1997) Using Ellenberg values to assess soil quality in British forests from ground vegetation: a pilot study. *J Appl Ecol* 34:375–387

Hengl T, Toomanian N, Reuter HI, Malakouti MJ (2007) Methods to interpolate soil categorical variables from profile observations: lessons from Iran. *Geoderma* 140:417–427

Hera U, Rötzer T, Zimmermann L, Schulz C, Maier H, Weber H, Kölling C (2012) Klima en détail. Neue, hochaufgelöste Klimakarten bilden wichtige Basis zur klimatischen Regionalisierung Bayerns. *LWF Aktuell* 86:34–37

Hothorn T, Bühlmann P, Kneib T, Schmid M, Hofner B (2010) Model-based boosting 2.0. *J Machine Learning Res* 11:2109–2113

Hothorn T, Müller J, Schröder B, Kneib T, Brandl R (2011) Decomposing environmental, spatial and spatio-temporal components of species distributions. *Ecol Monogr* 81:329–347

Käfer J, Witte JPM (2004) Cover-weighted averaging of indicator values in vegetation analyses. *J Veg Sci* 15:647–652

Kellomäki S, Leinonen S (eds) (2005) *Management of European forests under changing climatic conditions.* Tiedonantoja/Research Notes No. 163, University of Joensuu, Joensuu

Ketterer K, Ewald J (1999) Waldgesellschaften und Standorte auf dem Eibsee-Bergsturz bei Garmisch-Partenkirchen. *Hoppea* 60:627–690

Kneib T, Hothorn T, Tutz G (2009) Variable selection and model choice in geoadditive regression models. *Biometrics* 65:626–634

Kowarik I, Seidling W (1989) Zeigerwertberechnungen nach Ellenberg. Zu Problemen und Einschränkungen einer sinnvollen Methode. *Landschaft und Stadt* 21:132–143

Kühn I (2007) Incorporating spatial autocorrelation may invert observed patterns. *Diversity Distrib* 13:66–69

Legendre P, Legendre L (1998) *Numerical ecology.* Ed. 2. Elsevier, Amsterdam

Lennon JJ (2000) Red-shifts and red herrings in geographical ecology. *Ecography* 549:101–113

Lichstein JW, Simons TR, Shriner SA, Franzreb KE (2002) Spatial autocorrelation and autoregressive models in ecology. *Ecol Monogr* 72:445–463

Maloney KO, Schmid M, Weller DE (2011) Applying additive modelling and gradient boosting to assess the effects of watershed and reach characteristics on riverine assemblages. *Methods Ecol Evol* 3:116–128

Maracchi G, Sirotenko O, Bindi M (2005) Impacts of present and future climate variability on agriculture and forestry in the temperate regions: Europe. *Climatic Change* 70:117–135

McCullagh P (1980) Regression models for ordinal data. *J Roy Statist Soc Ser B* 42:109–142

Miller, JA, Hanham RQ (2011) Spatial nonstationarity and the scale of species-environment relationships in the Mojave Desert, California, USA. *Int J Geogr Inform Sci* 25:423–438

Miller JA, Franklin J, Aspinall R (2007) Incorporating spatial dependence in predictive vegetation models. *Ecol Modelling* 202:225–242

Pakeman RJ, Reid CL, Lennon JJ, Kent M (2008) Possible interactions between environmental factors in determining species optima. *J Veg Sci* 19:201–208

Parry ML, Canziani OF, Palutikof JP, van der Linden PJ, Hanson CE (eds) (2007) *Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge

Reger B, Kölling C, Ewald J (2011) Modelling effective thermal climate for mountain forests in the Bavarian Alps: Which is the best model? *J Veg Sci* 22:677–687

Reger B, Schüpferling R, Beck J, Dietz E, Morovitz D, Schaller R, Wilhelm G, Ewald J (2012) WINALPecobase – Ecological database of mountain forests in the Bavarian Alps. *Biodiversity Ecol* 4:167–171

Reger B, Häring T, Ewald J (2014) The TRM-model of potential natural vegetation in mountain forests. *Folia Geobot* (this issue). doi:10.1007/s12224-013-9158-0

Schaetzl RJ, Krist FJJ, Miller BA (2012) A taxonomically based ordinal estimate of soil productivity for landscape-scale analyses. *Soil Sci* 177:288–299

Schaffers AP, Sýkora KV (2000) Reliability of Ellenberg indicator values for moisture, nitrogen and soil reaction: a comparison with field measurements. *J Veg Sci* 11:225–244

Schmid M, Hothorn T, Maloney KO, Weller DE, Potapov S (2011) Geoadditive regression modeling of stream biological condition. *Environm Ecol Statist* 18:709–733

Seidling W, Rohner MS (1993) Zusammenhänge zwischen Reactions-Zeigerwerten und bodenchemischen Parametern am Beispiel von Waldbodenvegetation. *Phytocoenologia* 23:301–317

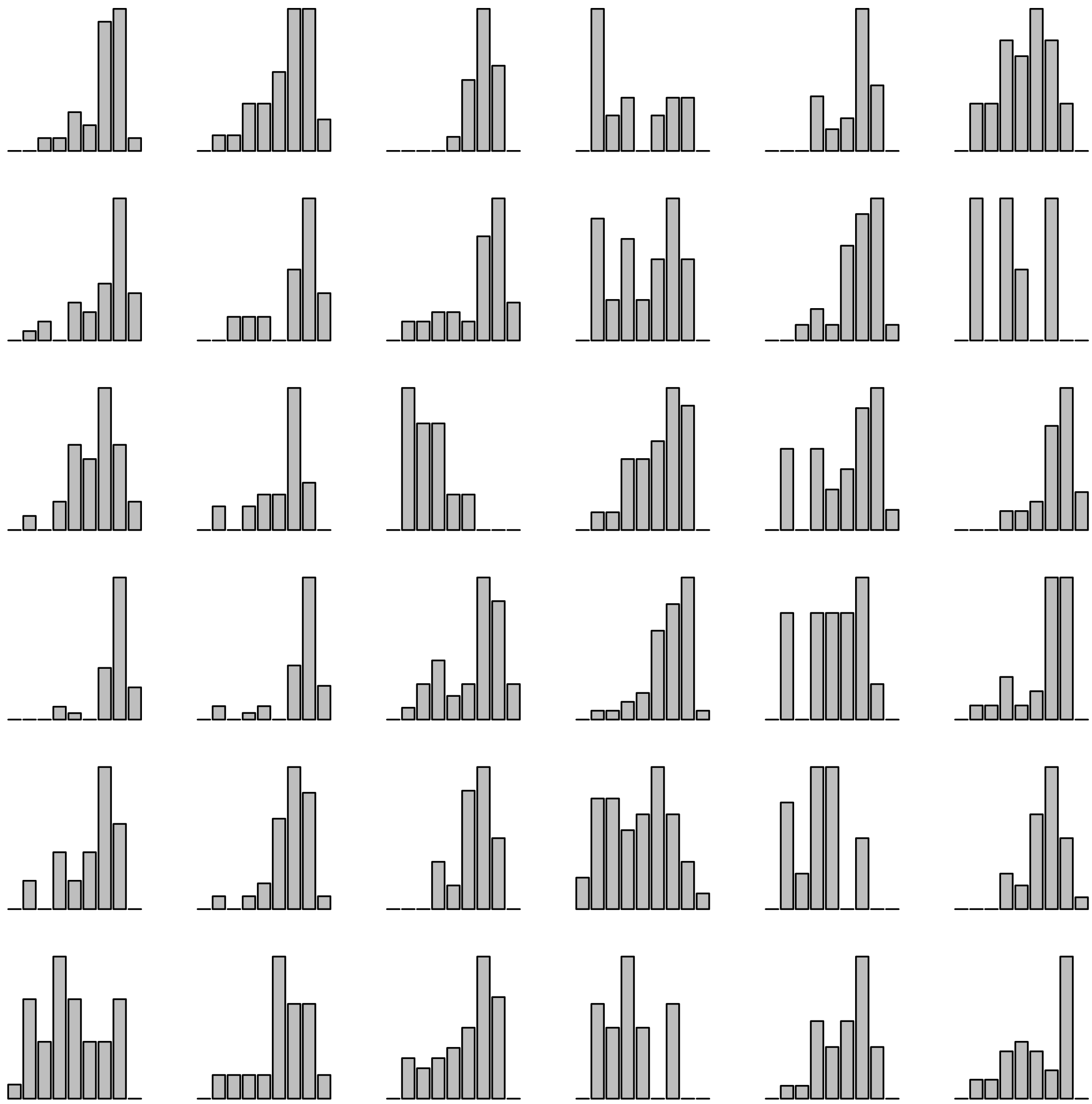
Zevenbergen LW, Thorne CR (1987) Quantitative analysis of land surface topography. *Earth Surface Processes and Landforms* 12:47–56

Received: 14 January 2012 / Revised: 24 October 2012 / Accepted: 29 October 2012

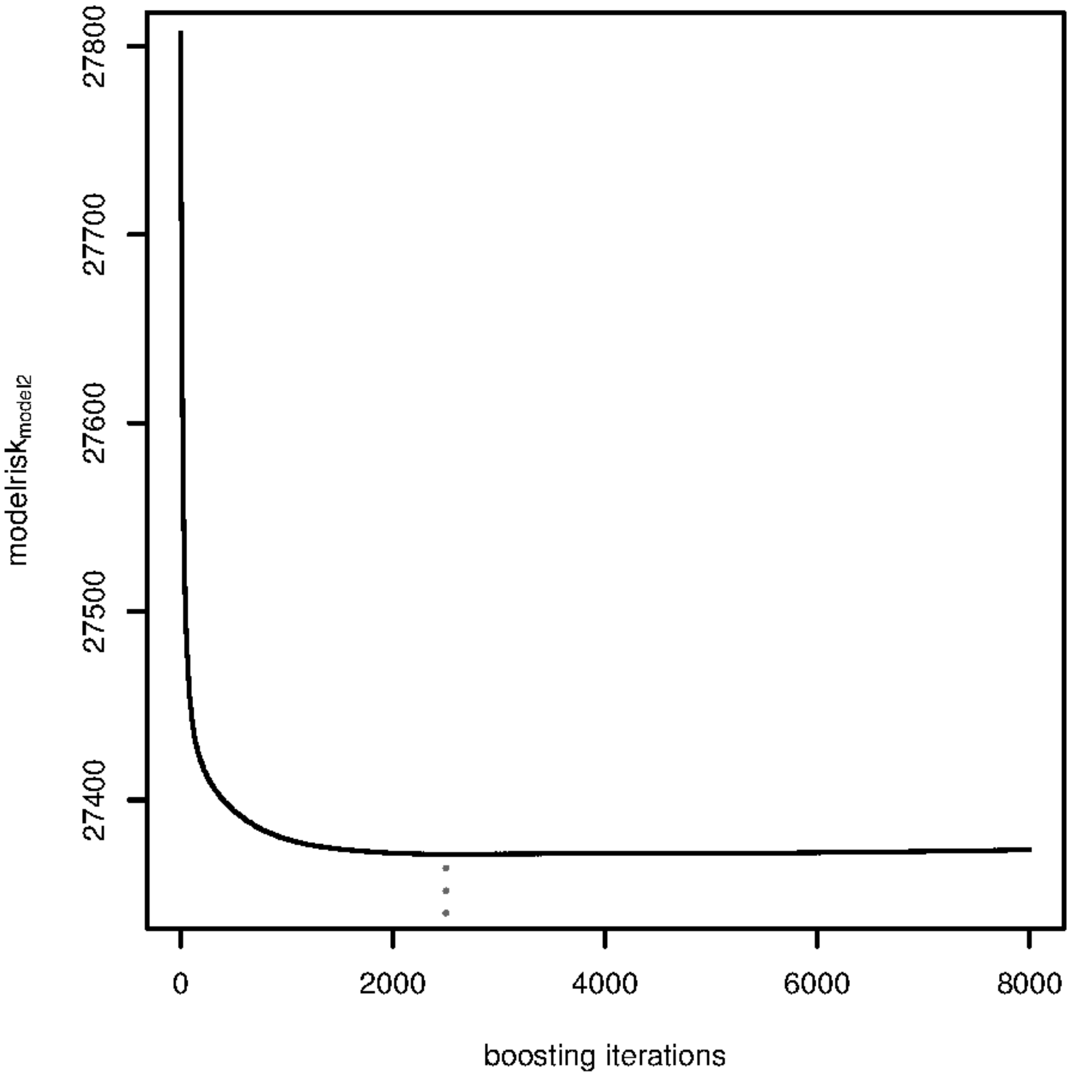
Electronic Supplementary Material

Regionalising indicator values for soil reaction in the Bavarian Alps – from averages to multivariate spectra

Tim Häring, Birgit Reger, Jörg Ewald, Torsten Hothorn and Boris Schröder



Appendix 1: Indicator Spectra of Ellenberg R-value of 36 randomly selected vegetation plots from WINALPecobase. The plots show a divers pattern of indicator spectra.



Appendix 2: Development of the negative out-of-sample log likelihood (model risk) over the number of boosting iterations for model 2. The model was restricted to 2500 boosting iterations.

Appendix B.

Legend to the map of the Bavarian growth areas and growth districts

Wuchsgebiet	Wuchsbezirk	Teilwuchsbezirk	
1 Untermainebene	1 Untermainebene		1.1
2 Spessart-Odenwald	1 Grundgebirgsspessart		2.1
	2 Buntsandsteinspessart	1 Hochspessart	2.2/1
		2 Nordspessart	2.2/2
		3 Mainspessart	2.2/3
	3 Bayerischer Odenwald		2.3
3 Rhön	1 Vorrhön		3.1
		1 Saale- und Sinn-Vorrhön	3.1/1
		2 Östliche Vorrhön	3.1/2
	2 Hohe Rhön		3.2
4 Fränkische Platte	1 Nördliche Fränkische Platte		4.1
		1 Nördliche Gipskeuperplatte und Grabfeld	4.1/1
	2 Südliche Fränkische Platte		4.2
		1 Südliche Gipskeuperplatte	4.2/1
		2 Kitzinger Sandgebiet	4.2/2
5 Fränkischer Keuper und Albvorland	1 Haßberge		5.1
	2 Steigerwald		5.2
		1 Südlicher Steigerwald	5.2/1
	3 Frankenhöhe		5.3
		1 Rangau-Hochfläche	5.3/1
	4 Itz-Baunach-Hügelland		5.4
	5 Nördliche Keuperabdachung		5.5
		1 Regnitzsenke	5.5/1
	6 Südliche Keuperabdachung		5.6
		1 Rezat-Rednitzsenke	5.6/1
	7 Nördliches Albvorland		5.7
	8 Südliches Albvorland		5.8
	9 Ries		5.9
6 Frankenalb und Oberpfälzer Jura	1 Nördliche Frankenalb und Nördlicher Oberpfälzer Jura		6.1
	2 Südliche Frankenalb und Südlicher Oberpfälzer Jura		6.2
		1 Ingolstädter Donaualb	6.2/1
	3 Schwäbische Riesalb		6.3
		1 Egaualb	6.3/1
	4 Oberfränkisches Braunjuragebiet		6.4
	5 Oberpfälzer Jurarand		6.5
7 Fränkisches Triashügelland	1 Bruchschollenland		7.1
	2 Obermainhügelland		7.2
	3 Stedtlinger Gebiet		7.3
8 Frankenwald, Fichtelgebirge und Steinwald	1 Frankenwald		8.1
	2 Münchberger Sattel		8.2
	3 Fichtelgebirge		8.3
	4 Brand-Neusorger Becken		8.4
	5 Steinwald		8.5
	6 Bayerisches Vogtland		8.6
	7 Selb-Wunsiedler Bucht		8.7
9 Oberpfälzer Becken- und Hügelland	1 Oberpfälzer Becken- und Hügelland		9.1
		1 Hessenreuther Wald	9.1/1
10 Oberpfälzer Wald	1 Mitterteicher Basaltgebiet		10.1
	2 Waldsassener Schiefergebiet und Wiesauer Senke	1 Waldsassener Schiefergebiet	10.2/1
		2 Wiesauer Senke	10.2/2
	3 Vorderer Oberpfälzer Wald		10.3
	4 Innerer Oberpfälzer Wald		10.4
	5 Cham-Further Senke		10.5
11 Bayerischer Wald	1 Westlicher Vorderer Bayerischer Wald		11.1
		1 Vorwaldrand	11.1/1
	2 Östlicher Vorderer Bayerischer Wald		11.2
		1 Lallinger Winkel	11.2/1
		2 Ilzvorland	11.2/2
	3 Innerer Bayerischer Wald		11.3
12 Tertiäres Hügelland	1 Donauried		12.1
	2 Ingolstädter Donauniederung	1 Donauau	12.2/1
		2 Donaumoos	12.2/2
	3 Ostbayerische Donauniederung	1 Donauau	12.3/1
		2 Gäuland	12.3/2
	4 Unteres Lechtal		12.4
	5 Unteres Isartal		12.5
	6 Unteres Inntal		12.6
	7 Mittelschwäbisches Schotterriedel- und Hügelland		12.7
		1 Biburger Hügelland	12.7/1
		2 Illerau	12.7/2
		3 Donauschwäbisches Hügelland	12.7/3
	8 Oberbayerisches Tertiärhügelland		12.8
		1 Aindlinger Terrassentreppe	12.8/1
	9 Niederbayerisches Tertiärhügelland	1 Westliches Niederbayerisches Tertiärhügelland	12.9/1
		2 Östliches Niederbayerisches Tertiärhügelland	12.9/2
		3 Neuburger Wald	12.9/3
13 Schwäbisch-Bayerische Schotterplatten- und Altmoränenlandschaft	1 Lechfeld		13.1
	2 Münchner Schotterebene	1 Südliche Münchner Schotterebene	13.2/1
		2 Nördliche Münchner Schotterebene	13.2/2
		3 Moose und Auen nördlich Münchens	13.2/3
	3 Mühldorfer und Öttinger Schotterfelder		13.3
	4 Vorallgäu		13.4
	5 Landsberger Altmoräne		13.5
	6 Isener Altmoräne und Hochterrasse		13.6
	7 Trostberger Altmoräne und Hochterrasse		13.7
14 Schwäbisch-Bayerische Jungmoräne und Molassevorberge	1 Bayerische Bodenseelandschaft		14.1
	2 Westallgäuer Hügelland		14.2
	3 Schwäbische Jungmoräne und Molassevorberge		14.3
	4 Oberbayerische Jungmoräne und Molassevorberge	1 Westliche kalkalpine Jungmoräne	14.4/1
		2 Inn-Jungmoräne	14.4/2
		3 Östliche kalkalpine Jungmoräne	14.4/3
15 Bayerische Alpen	1 Kürnacher Molassebergland		15.1
	2 Allgäuer Molassevoralpen		15.2
	3 Allgäuer Flysch- und Helvetikumvoralpen		15.3
	4 Oberbayerische Flysch-Voralpen	1 Ammergauer Flyschberge	15.4/1
		2 Tegernseer Flyschberge	15.4/2
		3 Teisendorfer Flyschberge	15.4/3
	5 Mittlere Bayerische Kalkalpen		15.5
	6 Chiemgauer Alpen und Saalforstamt St.Martin		15.6
	7 Allgäuer Hochalpen		15.7
	8 Karwendel und Wettersteinmassiv		15.8
	9 Berchtesgadener Hochalpen u. Saalforstamt St. Martin		15.9
		1 Leoganger Schieferberge	15.9/1

Figure B.1.: Legend to the map of the Bavarian growth areas and growth districts (Source: Walentowski et al. 2001)

Appendix C.

Permission of publishers

ELSEVIER LICENSE
TERMS AND CONDITIONS
Jul 09, 2013

This is a License Agreement between Tim Häring ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Tim Häring
License number	3184701243917
License date	Jul 09, 2013
Licensed content publisher	Elsevier
Licensed content publication	Geoderma
Licensed content title	Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils
Licensed content author	Tim Häring,Elke Dietz,Sebastian Osenstetter,Thomas Koschitzki,Boris Schröder
Licensed content date	September 2012
Licensed content volume number	185–186
Licensed content issue number	
Number of pages	11
Start Page	37
End Page	47
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Order reference number	
Title of your thesis/dissertation	Spatial Prediction Methods for the Assessment and Mapping of Forest Site Characteristics

Expected completion date	Jul 2013
Estimated size (number of pages)	150
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	0.00 USD
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

“Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER].” Also Lancet special credit - “Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier.”

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In

no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.

16. **Website**: The following terms and conditions apply to electronic reserve and author websites:
Electronic reserve: If licensed material is to be posted to website, the web site is to be password-protected and made available only to bona fide students registered on a relevant course if:
This license was made in connection with a course,
This permission is granted for 1 year only. You may obtain a license for future website posting,
All content posted to the web site must maintain the copyright information line on the bottom of each image,
A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com , and
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

17. **Author website** for journals with the following additional clauses:

All content posted to the web site must maintain the copyright information line on the bottom of each image, and the permission granted is limited to the personal version of your paper. You are not allowed to download and post the published electronic version of your article (whether PDF or HTML, proof or final version), nor may you scan the printed edition to create an electronic version. A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx . As part of our normal production process, you will receive an e-mail notice when your article appears on Elsevier's online service ScienceDirect (www.sciencedirect.com). That e-mail will include the article's Digital Object Identifier (DOI). This number provides the electronic link to the published article and should be included in the posting of your personal version. We ask that you wait until you receive this e-mail and have the DOI to do any posting.

Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

18. **Author website** for books with the following additional clauses:
Authors are permitted to place a brief summary of their work online only.
A hyper-text must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version.

Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

19. **Website** (regular and for author): A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx. or for books to the Elsevier homepage at http://www.elsevier.com

20. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

21. **Other Conditions**:

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK501061431.
Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006
For suggestions or comments regarding this order, contact RightsLink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Jul 09, 2013

This is a License Agreement between Tim Häring ("You") and John Wiley and Sons ("John Wiley and Sons") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by John Wiley and Sons, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3184721141904
License date	Jul 09, 2013
Licensed content publisher	John Wiley and Sons
Licensed content publication	Applied Vegetation Science
Licensed content title	Predicting Ellenberg's soil moisture indicator value in the Bavarian Alps using additive georegression
Licensed copyright line	© 2012 International Association for Vegetation Science
Licensed content author	Tim Häring,Birgit Reger,Jörg Ewald,Torsten Hothorn,Boris Schröder
Licensed content date	Jun 12, 2012
Start page	110
End page	121
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	No
Total	0.00 USD

Terms and Conditions

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a "Wiley Company") or a society for whom a Wiley Company has exclusive publishing rights in relation to a particular journal (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at http://myaccount.copyright.com).

Terms and Conditions

1. The materials you have requested permission to reproduce (the "Materials") are protected by copyright.

2.You are hereby granted a personal, non-exclusive, non-sublicensable, non-transferable, worldwide, limited license to reproduce the Materials for the purpose specified in the licensing process. This license is for a one-time use only with a maximum distribution equal to the number that you identified in the licensing process. Any form of republication granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before may be distributed thereafter). The Materials shall not be used in any other manner or for any other purpose. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Material. Any third party material is expressly excluded from this permission.

3. With respect to the Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Materials without the prior permission of the respective copyright owner. You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Materials, or any of the rights granted to you hereunder to any other person.

4. The Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc or one of its related companies (WILEY) or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than

the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto.

5. NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

6. WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.

7. You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.

8. IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

9. Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.

10. The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such

waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.

11. This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.

12. Any fee required for this permission shall be non-refundable after thirty (30) days from receipt

13. These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.

14. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.

15. WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

16. This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.

17. This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

Wiley Open Access Terms and Conditions

Wiley publishes Open Access articles in both its Wiley Open Access Journals program [http://www.wileyopenaccess.com/view/index.html] and as Online Open articles in its subscription journals. The majority of Wiley Open Access Journals have adopted the Creative Commons Attribution License (CC BY) which permits the unrestricted use, distribution, reproduction, adaptation and commercial exploitation of the article in any medium. No permission is required to use the article in this way provided that the article is properly cited and other license terms are observed. A small number of Wiley Open Access journals have retained the Creative Commons Attribution Non Commercial License (CC BY-NC), which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Online Open articles - Authors selecting Online Open are, unless particular exceptions apply, offered a choice of Creative Commons licenses. They may therefore select from the CC BY, the CC BY-NC and the Attribution-NoDerivatives (CC BY-NC-ND). The CC BY-NC-ND is more restrictive than the CC BY-NC as it does not permit adaptations or modifications without rights holder consent.

Wiley Open Access articles are protected by copyright and are posted to repositories and websites in accordance with the terms of the applicable Creative Commons license referenced on the article. At the time of deposit, Wiley Open Access articles include all changes made during peer review, copyediting, and publishing. Repositories and websites that host the article are responsible for incorporating any publisher-supplied amendments or retractions issued subsequently.
Wiley Open Access articles are also available without charge on Wiley's publishing platform, **Wiley Online Library** or any successor sites.

Conditions applicable to all Wiley Open Access articles:

- The authors' moral rights must not be compromised. These rights include the right of "paternity" (also known as "attribution" - the right for the author to be identified as such) and "integrity" (the right for the author not to have the work altered in such a way that the author's reputation or integrity may be damaged).

- Where content in the article is identified as belonging to a third party, it is the obligation of the user to ensure that any reuse complies with the copyright policies of the owner of that content.

- If article content is copied, downloaded or otherwise reused for research and other purposes as permitted, a link to the appropriate bibliographic citation (authors, journal, article title, volume, issue, page numbers, DOI and the link to the definitive published version on Wiley Online Library) should be maintained. Copyright notices and disclaimers must not be deleted.
    - Creative Commons licenses are copyright licenses and do not confer any other rights, including but not limited to trademark or patent rights.

- Any translations, for which a prior translation agreement with Wiley has not been agreed, must prominently display the statement: "This is an unofficial translation of an article that appeared in a Wiley publication. The publisher has not endorsed this translation."

  Conditions applicable to non-commercial licenses (CC BY-NC and CC BY-NC-ND)

For non-commercial and non-promotional purposes individual non-commercial users may access, download, copy, display and redistribute to colleagues Wiley Open Access articles. In addition, articles adopting the CC BY-NC may be adapted, translated, and text- and data-mined subject to the conditions above.

Use by commercial "for-profit" organizations

Use of non-commercial Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee. Commercial purposes include:

- Copying or downloading of articles, or linking to such articles for further redistribution, sale or licensing;
- Copying, downloading or posting by a site or service that incorporates advertising with such content;
- The inclusion or incorporation of article content in other works or services (other than normal quotations with an appropriate citation) that is then available for sale or licensing, for a fee (for example, a compilation produced for marketing purposes, inclusion in a sales pack)
- Use of article content (other than normal quotations with appropriate citation) by for-profit organizations for promotional purposes
- Linking to article content in e-mails redistributed for promotional, marketing or educational purposes;
- Use for the purposes of monetary reward by means of sale, resale, license, loan, transfer or other form of commercial exploitation such as marketing products
- Print reprints of Wiley Open Access articles can be purchased from: corporatesales@wiley.com

  The modification or adaptation for any purpose of an article referencing the CC BY-NC-ND License requires consent which can be requested from RightsLink@wiley.com .

  Other Terms and Conditions:

  BY CLICKING ON THE "I AGREE..." BOX, YOU ACKNOWLEDGE THAT YOU HAVE READ AND FULLY UNDERSTAND EACH OF THE SECTIONS OF AND PROVISIONS SET FORTH IN THIS AGREEMENT AND THAT YOU ARE IN AGREEMENT WITH AND ARE WILLING TO ACCEPT ALL OF YOUR OBLIGATIONS AS SET FORTH IN THIS

AGREEMENT.

v1.8

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK501061470.
Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

SPRINGER LICENSE
TERMS AND CONDITIONS

Jul 09, 2013

This is a License Agreement between Tim Häring ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3184690720251
License date	Jul 09, 2013
Licensed content publisher	Springer
Licensed content publication	Folia Geobotanica
Licensed content title	Regionalizing Indicator Values for Soil Reaction in the Bavarian Alps – from Averages to Multivariate Spectra
Licensed content author	Tim Häring
Licensed content date	Jan 1, 2013
Type of Use	Thesis/Dissertation
Portion	Full text
Number of copies	6
Author of this Springer article	Yes and you are a contributor of the new work
Order reference number	
Title of your thesis / dissertation	Spatial Prediction Methods for the Assessment and Mapping of Forest Site Characteristics
Expected completion date	Jul 2013
Estimated size(pages)	150
Total	0.00 USD

Terms and Conditions

Introduction

The publisher for this copyrighted material is Springer Science + Business Media. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

Limited License

With reference to your request to reprint in your thesis material on which Springer Science and Business Media control the copyright, permission is granted, free of charge, for the use indicated in your enquiry.

Licenses are for one-time use only with a maximum distribution equal to the number that you identified in the licensing process.

This License includes use in an electronic form, provided its password protected or on the university's intranet or repository, including UMI (according to the definition at the Sherpa website: http://www.sherpa.ac.uk/romeo/). For any other electronic use, please contact Springer at (permissions.dordrecht@springer.com or permissions.heidelberg@springer.com).

The material can only be used for the purpose of defending your thesis, and with a maximum of 100 extra copies in paper.

Although Springer holds copyright to the material and is entitled to negotiate on rights, this license is only valid, subject to a courtesy information to the author (address is given with the article/chapter) and provided it concerns original material which does not carry references to other sources (if material in question appears with credit to another source, authorization from that source is required as well).

Permission free of charge on this occasion does not prejudice any rights we might have to charge for reproduction of our copyrighted material in the future.

Altering/Modifying Material: Not Permitted

You may not alter or modify the material in any manner. Abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of the author(s) and/or Springer Science + Business Media. (Please contact Springer at (permissions.dordrecht@springer.com or permissions.heidelberg@springer.com)

Reservation of Rights

Springer Science + Business Media reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

Copyright Notice:Disclaimer
You must include the following copyright and permission notice in connection with any reproduction of the licensed material: "Springer and the original publisher /journal title, volume, year of publication, page, chapter/article title, name(s) of author(s), figure number(s), original copyright notice) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media"

Warranties: None

Example 1: Springer Science + Business Media makes no representations or warranties with respect to the licensed material.

Example 2: Springer Science + Business Media makes no representations or warranties with respect to the licensed material and adopts on its own behalf the limitations and disclaimers established by CCC on its behalf in its Billing and Payment terms and conditions for this licensing transaction.

Indemnity
You hereby indemnify and agree to hold harmless Springer Science + Business Media and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

No Transfer of License
This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without Springer Science + Business Media's written permission.

No Amendment Except in Writing
This license may not be amended except in a writing signed by both parties (or, in the case of Springer Science + Business Media, by CCC on Springer Science + Business Media's behalf).

Objection to Contrary Terms
Springer Science + Business Media hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and Springer Science + Business Media (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

Jurisdiction

All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in The Netherlands, in accordance with Dutch law, and to be conducted under the Rules of the 'Netherlands Arbitrage Instituut' (Netherlands Institute of Arbitration).***OR:***

All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in the Federal Republic of Germany, in accordance with German law.

Other terms and conditions:

v1.3

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK501061383.
Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

List of oral and poster presentations

Oral presentations

Häring, T. & Schröder, B. (2010) A review of model-error in digital soil mapping: Confronting statistical soil landscape models with large-scale field validation data. *Geophysical Research Abstracts*, 12, EGU2010-12757.

Häring, T. & Schröder, B. (2010) Sampling Optimization using Image Segmentation. *Proceedings of the 4^{th} International Workshop on Digital Soil Mapping*, 24.-26.05.2010, Rome.

Häring, T. (2010) Facilitation of Environmental Modeling by Means of Scripting: The Combination of R andx SAGA. *SAGA User Conference*, AGIT, 08.07.2010, Salzburg.

Poster presentations

Häring, T., Dietz, E. & Kölling, C. (2009) Zusammenhang zwischen Rastergröße und Modellgüte für die Prognose von Bodenkarten im maßstab 1 : 25.000. *Jahrestagung der Deutschen Bodenkundlichen Gesellschaft: Böden - eine endliche Ressource*, September 2009, Bonn. Berichte der DBG.

Dietz, E. Falk, W. Beck, J. **Häring, T.** & Kölling, C. (2009) Flächenhaftes Prognosemodell für Stauwasserböden unter Wald aus Bodenparametern, DGM, Klima und Vegetation. *Jahrestagung der Deutschen Bodenkundlichen Gesellschaft: Böden - eine endliche Ressource*, September 2009, Bonn. Berichte der DBG.

Häring, T., Osenstetter, S. & Dietz, E. (2011) Substratbasierte Bodenklassifikation - Wo die digitale Bodenkartierung an ihre Grenzen stößt. *1. Workshop der AG Digital Soil Mapping der DBG*, 10-11.06.2011, Hannover.

Acknowledgements

Damit ich diese Arbeit zustande bringen konnte schulde ich vielen Personen meinen Dank. Vielen Dank an...

- Boris Schröder - für die Bereitschaft, die Arbeit aus der Ferne zu betreuen, die vielen hilfreichen Gespräche und Korrekturen.
- Jörg Ewald - ohne Dich und die Möglichkeiten in WINALP würde es diese Arbeit nicht geben!
- Elke Dietz und Christan Kölling - für die Offenheit, an einer Landesanstalt eine Dissertation zu unterstützen. Ohne euch wäre es genauso wenig gegangen!
- Jörg Müller - für die Begutachtung der Arbeit.
- Johannes Kollmann - für den Vorsitz der Prüfungskommission.
- Torsten Hothorn - für die unkomplizierte und wegweisende Beratung in Sachen statistischer Modellierung.
- Birgit Reger - für die vielen hilfreichen Diskussionen, Gespräche und Duplos.
- das ganze KLIP4 Team. Wir waren einzigartig!
- Jan, Jonas, Dominik, Fred & Lenne - Ihr sei echt die coolsten Brüder!
- meine Eltern - ihr habt mich bei all meinen Entscheidungen immer unterstützt.
- Eni-Louise und Joa-Finn. Ja, jetzt endlich hab ich es geschafft!
- Steffi!!! Ich glaub, ich sehe den Mond wieder...