Nonlinear multivariate statistical analysis of long-term soil monitoring datasets of different scales

Dissertation
zur Erlangung eines Doktorgrades

eingereicht an der
Fakultät für Architektur und Bauingenieurswesen
der
Bergischen Universität Wuppertal

vorgelegt von
Carsten Schilli
aus Essen

Wuppertal 2017

Die Dissertation kann wie folgt zitiert werden:

urn:nbn:de:hbz:468-20180326-104939-5
[http://nbn-resolving.de/urn/resolver.pl?urn=urn%3Anbn%3Ade%3Ahbz%3A468-20180326-104939-5]

Table of contents

List of figures

List of tables

List of abbreviations

Abbreviation	Plaintext
bBIS	nationwide soil information system of the Federal republic of Germany
BBodSchG	Federal Soil Protection Act of the Federal Republic of Germany
BBodSchV	Federal Soil Protection and Contaminated Sites Ordinance of the Federal Republic of Germany
BImSchG	Federal Immission Control Act
BzBIG	Law on Lead in Petrol
FRG	Federal Republic of Germany
GSMP	German long-term soil monitoring program
ICP Forest	International Co-Operative Programme on Assessment and Monitoring of Air Pollution Effects on Forests
Isomap	Isometric Feature Mapping
LUCAS	Land Use/Cover Area frame Survey
PCA	Principal Component Analysis
SM	Sammon’s Mapping
SOM	Self-Organizing Map
SOM-SM	Self-Organizing Map combined with Sammons Mapping
TERENO	Terrestrial Environmental Observatories
TOC	Total organic carbon
UBA	Federal Environment Agency
WRB	World Reference Base

1. Introduction

Soils are an important compartment in the global terrestrial ecosystem. Soils are complex systems and are in contact and relating with the atmo-, hydro-, bio- and lithosphere (Tarquis et al. 2011, FAO and ITPS 2015).

The development of soils is driven by the following soil-forming factors: climate, parent material, topography, vegetation. The specific strength of the soil-forming factors varies depending on global and local geographical position. The soil-forming factors cause numerous single processes and impacts, which have an effect on the soil and its different characteristics (Scheffer & Schachtschabel 2010). The formation of soils induced by these factors is often a slow occurring process. In contrast, the changes and development in the characteristics of soils affected by anthropogenic activities can take place very quickly.

The relevance of different processes and impacts can vary considerably. Climate Change is a worldwide impact, affecting soils in different ways. Increasing temperatures can, for example, affect the soil moisture (Müller & Zhang 2016). Another overall impact on the soils is the deposition, especially of acidifying and eutrophying compounds as shown in Europe (Waldner et al. 2014). Other overall impacts can be widespread, but be limited to soils with particular land use, for example the input of trace elements via fertilizers in agricultural soils in Europe (Nziguheba & Smolders 2008). Besides the overall influence of those impacts the strength of influence can vary on different scales as shown in these studies. Further examples of scale effects are given for instance by Wang et al. (2015). They provide detailed insights into the impacts that soil texture and climate have on soil organic carbon at different scales. Biswas et al. (2013) presented scale and spatial depending effects by using soil water storage and clay content data.

In addition, impacts can be also restricted to a typical scale. An example of a large-scale (local) influence is the emission of crematories, which increase the mercury content of nearby soils significantly (Mari & Domingo 2009). As we have known for a long time, a single process can influence small subareas of a soil, as root exudates alter pH and redox potential in microsites of the rhizosphere (Fischer et al. 1989, Dorau et al. 2016). The mentioned processes and impacts reflect only a tiny part of the existing multiplicity.

Scale effects hold true for the spatial scale as well as for the time scale (Tarquis et al. 2011). Biological parameters of the soil change in the course of the year, because of changing temperature and moisture (Haag et al. 2015). Flooding events, following a more or less typical yearly pattern, can also cause influence e.g. the microbial community and soluble organic carbon content (Moche et al. 2015) as well as pH, redox potential and heavy metal content of soil solution (Schulz-Zunkel et al. 2015).

Despite the complex interplay of different spheres or the locally and globally varying characteristics of soil-forming factors, occurring impacts, and processes, it is possible to describe soils from all over the world with a single taxonomy, namely the World Reference Base (WRB) (FAO 2014). The WRB uses diagnostic horizons, properties and materials to classify and describe the soils. In addition to a precise soil classification, further knowledge of pedogenetic processes is necessary to understand the past and future development of soils (Bockheim & Gennadiyev 2000, Arrouays et al. 2014).

Soils fulfill essential functions for human life (Oliver & Gregory 2015) and for ecosystems (Brumme & Khanna 2009). In the last decades the ongoing worldwide soil degradation (FAO and ITPS 2015) has raised the awareness of soils as a non-renewable resource essential for the persistence of the ecological environment and for human society (BMU 2017, EEA 2015). Soils are used for food production, as habitat for animals and plants or as a settlement area for humans, to only name a few. The protection of soils and their function is an important part of national (BMU 2017) and international (EC 2006) political programs. The need for soil protection can be illustrated by different ways which can influence human health. Contaminated soils can harm the human health by direct human inhalation or indirectly by releasing stored contaminants in to drinking water or crops that are consumed by humans (Oliver & Gregory 2015). On the other hand, healthy soils can remove different contaminants from seepage water, e.g. biocides (Bester et al. 2011), organic pollutants (Rentz et al. 2005) or heavy metals (Yobouet et al. 2016), and therefore can be regarded as natural protectors for groundwater and surface water. To protect the important characteristics and functions of the soil requires an increasing amount of soil information (Richer de Forges & Arrouays 2010), especially with regard to processes that threaten the soils (Arrouays 2014).

In the Federal Republic of Germany the Federal (FRG) Soil Protection Act (BBodSchG – Bundes-Bodenschutzgesetz) came into force in 1998 (BBodSchG 1998). As determined in Article 1 of the BBodSchG the purpose of the Act is to protect the function of the soil and to prevent harmful changes. Soil monitoring is essential to understanding the processes of the current soil status or for changes to the soils and their characteristics (Arrouays 2009). It is also a useful tool in precautionary soil protection (Barth et al. 2001). To detect (harmful) changes in soils, the monitoring is inevitable and is in accordance to the BBodSchG. Only by understanding today's soil status as well as the processes in soils and impacts on soils, is a prognosis of future soil development possible. To understand the soil, its characteristics and the changes, a monitoring of soils was established in Germany.

Soil research and observation has a long history. At single sites, soil has been studied since the end of the 18th century (e.g. Körschens et al. 2002, Girma et al. 2007). Soil research started with the focus on agricultural aspects and the productivity of soils and suitable tillage systems (Rasmussen et al. 1998). Today the data generated at these sites is used for further investigation in to quantifying nutrient cycles, carbon sequestering, studying possible effects of climatic change and the use of pesticides (Ellmer 2012).

The long-term agronomic research sites, which can be found all over the world e.g. in the United States of America (Mitchel et al. 1991), India (Pathak et al. 2011) as well as in Europe (Körschens et al. 2013) e.g. Great Britain (Jenkinson 1991) and Germany (Körschens 2010, Körschens et al. 2014), cannot be regarded as monitoring sites in terms of systematic comprehensive soil monitoring (even if they provide valuable information), because of the clear agricultural focus of research on such sites. Comprehensive soil monitoring, as defined for the German Soil Monitoring Program, must be able to

- describe the current state of soils,
- determine changes of soil properties,
- and identify and document the causal processes

in different environmental regions as well as under different soil management and treatment conditions (Barth et al. 2001). To fulfill the aims of soil monitoring it is obvious that a network of soil monitoring sites covering the different regions and management systems is required (Morvan et al. 2008, Arrouays et al. 2012).

During the last decades of soil research, a lot of data has been collected. It is commonly felt that only a minor fraction of information has been extracted from these data so far and much more information remain to be revealed to meet the requirements of sustainable resources management. This seems to be a common feature nowadays, to cite John Naisbitt: *"We are drowning in information but starved for knowledge"* (Naisbitt 1983). The difficulty in working with large and multivariate datasets is delineating the relevant information in the dataset from the statistical noise. This problem is well known and can be overcome with the selection of adequate statistical approaches (Webster & Oliver 1990, Desaules et al. 2010, van Wesemael 2011). Nevertheless, the selection of the appropriate statistical methods is still a challenge (Arrouays 2014, Tonidanel et al. 2016).

In this thesis the datasets of the pan German long-term Soil Monitoring Program will be analyzed with innovative explorative multivariate approaches the first time. Additionally a comparable dataset from a large scale monitoring program will be analyzed with the same method. For this task the selection of appropriate statistical methods is necessary (Morvan et al. 2008, Arroays et al. 2012). The selection will be based on the demands put on the German long-term Soil Monitoring Program. Furthermore the strengths, weaknesses, and future possibilities of the German long-term Soil Monitoring Program and the belonging dataset will be evaluated.

2. Soil research and monitoring programs

To detect (harmful) changes in the soil, comprehensive soil monitoring is necessary. The objectives of a soil monitoring are defined by several authors. The authors (Arrouays et al. 1998, Barth et al. 2001, Havlikova & Klement 2002, de Gruijter et al. 2006, Desaules et al. 2010, Nicolas et al. 2014) agree in the following objectives:

- The description and evaluation of the current state of soil properties and characteristics
- The determination of long-term changes of soil properties and characteristics
- To identify reasons for changes in the soil
- To act as an early warning system for harmful changes
- To inform and advise politics and administration based on well-grounded information

To achieve the monitoring aims of determination of long-term changes of soil properties and characteristics two conditions must be met. At first the parameter spectrum of monitoring networks has to be broad to detect (unknown) common and peculiar characteristics and trends. The selection of parameters, especially when used as indicators for particular questions is difficult (Stone et al. 2016). Even if processes are known or a concrete question is defined, the choice of adequate indicators remains a difficult one (Zornoza et al. 2015, Stone et al. 2016). Therefore a broad spectrum of parameters is necessary to explore unknown interdependencies.

Second a monitoring has to be continuous. The long-term aspect of monitoring networks should not be underestimated (Prounier et al. 2015). Parameters respond at different rates to environmental influences. The required sampling frequency and time to reach first resilient results varies widely between different parameters. For example the soil respiration can clearly differ between 1 to 3 years in grassland and forest soils (Haag et al. 2015). A time span of 10 or more years can be necessary to confirm changes of nitrogen content, base cation exchange in forest soils (de Vries et al. 2009) or of organic carbon in arable soils

(Körschens 2010, Körschens et al. 2015). On the other hand changes in the environment which can influence the soil, such as changing deposition (Meesenburg et al. 2016) or climate change (Müller & Zhang 2016), extend over a long period. Even if a lot of data can be generated in a short time, e.g. by yearly investigations, the results and conclusions should be handled with care. Short-term programs, e.g. a program in New Zealand which lasts only six years (Sparling & Schipper 2004), may not detect significant changes when regarding parameters like TOC or heavy metal contents. It is possible that other programs like the Terrestrial Environmental Observatories (TERENO) project (Zacharias et al. 2011) or the program in the federal state North-Rhine Westphalia in Germany investigating soil carbon contents and stocks (Kaufmann-Boll et al. 2012) will have to deal with such problems if stopped after 15 years as planned.

In addition to the common aims, some authors include further aims, e.g.

- The evaluation and validation of analytical methods (Havlikova & Klement 2002),
- The development and validation of models for predicting changes (Arrouays et al. 1998).

All over the world numerous programs collect soil information of different kinds. The investigation areas vary in scale. A worldwide project is the development of the global soil map. To fulfill this task, data is and will be collected and merged from different programs (Arrouays et al. 2017). Further programs collect data from all over a continent, e.g. the pan-European LUCAS Program (Toth et al. 2013), with focus on monitoring land use/land cover changes and analyzing soil samples with regard to soil fertility. The International Co-Operative Programme on Assessment and Monitoring of Air Pollution Effects on Forests (ICP Forests) focus on forest health (Lorenz 2010, Sanders et al. 2016) covering Europe and North America. The parameter set by ICP Forests includes parameters of soil fertility and heavy metals (Kaufmann-Boll et al. 2012). A real global soil monitoring network does not exist today, but is needed to investigate soil quality worldwide and to check and to support outputs of models on a small as well as at a global scale (Montanarella & Vargas 2012, FAO and ITPS 2015, Arrouays et al. 2017).

In Europe, national borders outline most of the small-scale soil monitoring networks. Within Europe 30 countries maintain national soil monitoring networks of varying ranges and qualities (Morvan et al. 2008). The different networks are not concerted, especially with respect to uniformity in distribution of sites and analytical and practical methodology (Morvan et al. 2008, van Wesemael et al. 2011). Soil monitoring programs are primarily conducted in financially strong industrialized countries or are financed by international initiatives (Montanarella & Vargas 2012). Nevertheless there are further countries outside Europe maintaining national monitoring systems e.g. in Ghana and Bangladesh (Brammer & Nachtergaele 2015), Kazakhstan (Dzhalankuzov 2014) or the six-year monitoring program in New Zealand (Sparling & Schipper 2004). Teng et al. (2014) reviewed several soil environmental monitoring activities in China. Most of them are focused on soil productivity or soil pollution. The various Chinese programs are not coordinated and use different parameter sets, analytical methods, and aims of research, similar to the European national programs.

Apart from nationwide programs, there are many other programs using a larger scale. For example, there are 16 national parks in Germany, established between 1970 (Bayerischer Wald) and 2015 (Hunsrück-Hochwald). The land area within these parks varies from 4 km² (Hamburgisches Wattenmeer) to 24.2 km² (Bayerischer Wald) (BfN 2016). The national parks were and are supposed to persist permanently. In these national parks a comprehensive ecosystem research and monitoring is carried out, including the monitoring of soil parameters. In most national parks the regarded soil parameters are focused on interactions and relevance for plant growth and health and water quality (Schlumprecht et al. 2015, Kowatsch et al. 2011).

The network of TERENO will run for at least 15 years. It includes four regional investigation sites in Germany. The Terrestrial Environmental Observatories focus on the long-term impacts of climatic change on ecological, social, and economic development (Zacharias et al. 2011, Bogena et al. 2012). An example of a program at the scale level of a federal state is the monitoring of organic carbon in agricultural soils of the federal state of North-Rhine Westphalia over 15 years (Kaufmann-Boll et al. 2012). The measured parameters shall explore climate change effects and soil fertility.

At a local level, small catchments for ecological issues are investigated as the Lehstenbach and Steinkreuz Catchment (Gerstenberger et al. 2004) or the Chicken Creek (Gerwin et al. 2010). The artificial Chicken Creek catchment was established to study the initial development phase in a post-mining landscape. Often such research areas are supervised by universities. The mentioned programs can only represent a small part of the existing programs collecting soil data. Kaufmann-Boll et al. 2012 report from nine programs at the national scale in Germany alone.

Most of the European national monitoring networks focused on comprehensive soil monitoring comprise physical, chemical, and biological parameters, just as do, e.g. the network of the Czech Republic (Havlikova & Klement 2002), Germany (Barth et al. 2001) or Swiss (Schwab et al. 2006). Nevertheless, the general parameter set (pollutants, nutrients, and physical characteristics) is similar. Some parameters are determined in a few countries only, such as earthworm diversity (France, Germany, Lativa, and Netherlands) or soil water retention (Austria, Romania and Spain) (Morvan et al. 2008). Therefore, it can be stated, that there is a lack of harmonization between the existing (national) networks (FAO and ITPS 2015).

In contrast to the uniform targets of soil monitoring and similar parameter sets the programs vary in the way monitoring sites were selected. The selection and spatial distribution of the sites should be based on the characteristics of the monitoring area, but follows different rules if the several national monitoring systems are compared. While, for example Hungary, France, and Poland chose the sites by considering land use and terrain information, other countries select the sites by using a grid e.g. Wales, Denmark, Scotland, or Ireland. An uncoordinated selection of sampling sites can be found in such places as Spain (Van Camp et al. 2004, Morvan et al. 2008).

2.1 The German long-term Soil Monitoring Program (GSMP)

The German long-term soil monitoring program (GSMP) was initiated in the FRG in the 1980's. The responsibilities lie with the federal states and not with the federation. The main goals of the GSMP are to describe the current state of soils, to monitor long-term changes of soil properties, and to assess the future development (Barth et al. 2001). After the German reunification in 1990, the program was extended to the area of the former German Democratic Republic. Here, the measurements mostly started in the middle of the 1990s (Huschek et al. 2004). Today, soil monitoring is performed at about 800 sites stretched all over the FRG. The sites of GSMP mainly comprise grassland, arable land, and forest sites. Less than 7 % of the sites are fallow ground, specialized cultivation or urban soils (Figure 1).

Because there is no national law or guideline for selecting monitoring sites in each federal state different numbers of sites were selected based on varying criteria. The sites were chosen by considering the typical soils, landscapes and land use or using expert knowledge or statistical criteria. Additionally a few sites were chosen to reflect specific characteristics, such as contaminated sites or special soil forms. Another important criterion was the long-term availability of the sites, to facilitate long-term measurements. The various procedures have generated different numbers, distributions and characteristics of sites within the federal states and the FRG.

The recommendations of Barth et al. 2001 include the classification of the monitoring sites in to two types: basic monitoring sites for trait documentation and intensive monitoring sites for additional documentation of processes. At basis monitoring sites at least 70 mandatory physical, chemical and biological parameters should be measured. Soil samples should be collected as composite samples from different horizons down to 200 cm. Depending on the parameter the measurements should be repeated from once a year (mostly biological parameters) to once in five years (mostly chemical parameters) (Table 11 in Appendix I). These frequencies are typical for European soil monitoring networks in order to meet the requirements for a trend analysis, e.g. for soil contaminates (DIN ISO 16133).

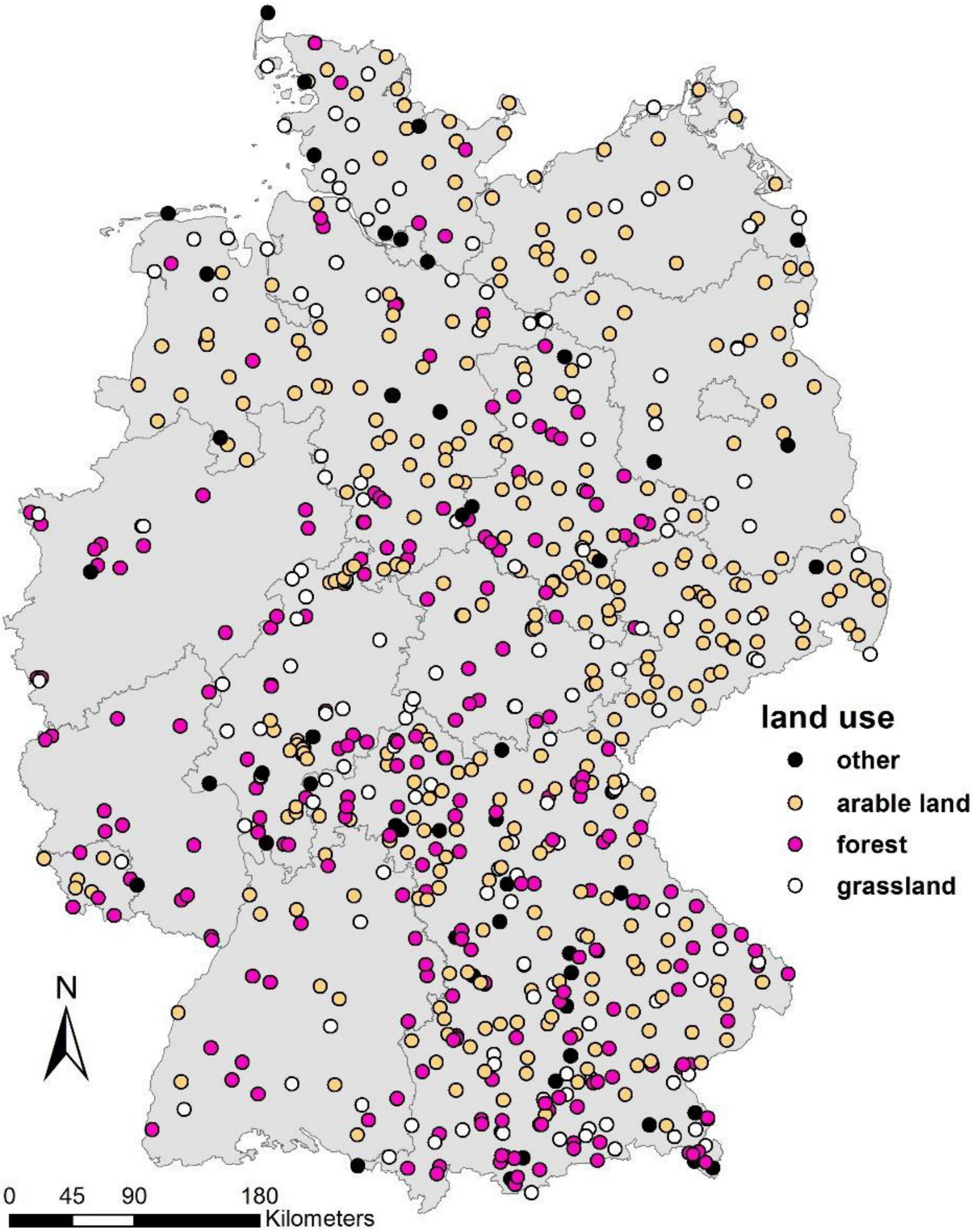

Figure 1: Monitoring sites of the GSMP. Grey lines delineate the federal states of Germany.

The Federal Environment Agency (UBA) compiled all the information and data of the GSMP on to a single database (bBIS – nationwide soil information system) to give an overview and to allow nationwide analyses (UBA 2010). The bBIS contains the descriptions of the monitoring sites and the quality proofed analytical data of measured soil matrix samples. The bBIS includes information from 11 of the 16 federal states and from more than 700 study sites in 2014. A continuous complementation is intended in the future. Up to 2016, however, this has been carried out only for individual questions and therefore parameter-specific and not comprehensively.

Despite the recommendations for temporal replications (Barth et al. 2001), the number of analytical results in the bBIS is rather low. The number of temporal measurements usually is less than the recommended frequency of replication. For example, only seven federal states provided heavy metal content data covering a period of 10 or more years till 2014. In some states this period is covered by two samples, which means one replication after 10 years (Schilli et al. 2011). Similarly other European countries monitoring programs also lack an adequate number of temporal replicates (Arrouays 2009).

At more than 90 monitoring sites of the GSMP the soil matrix and the soil solution is sampled. These sites are called intensive monitoring sites. The soil solution is collected continuously with mostly three to six suction cups or plates at each depth at biweekly to four weekly intervals. Different soil solutions of each depth are pooled. Chemical analysis is conducted on the merged samples. The analyzed parameter set from the soil solution is similar to that of extractions of the soil matrix samples (cf. Barth et al. 2001). The maintenance of the sites varies, especially in the case of the measured parameters and the use of analytical methods.

The bBIS currently does not include these data from intensive monitoring sites. In most cases the federal states maintain their own database. Some federal states, e.g. Lower Saxony, Hesse and Saxony-Anhalt, pool the data from (a part of) their intensive monitoring sites which are managed by a common institution - the Northwest German Forest Research Station. Therefore a nationwide evaluation of data from intensive monitoring is currently not possible, due to the lack of a composite dataset.

Although general guidelines are provided these days, the implementation of the program as well as of the soil analytics, slightly differ between the federal states. For example, the extraction of heavy metals in soils with aqua regia is a mandatory parameter (Table 11 in Appendix I), whereas the total content after Ruppert (1987) including the extraction with hydrofluoric acid under pressure is an optional parameter (Barth et al. 2001). In a few cases only the latter has been determined. In addition, the methods to measure the contents vary. Table 1f heavy metals and As in soils.

Table 1 shows different extraction methods used in the GSMP for the “whole” content of heavy metals and As in soils.

Table 1: Methods used to analyze the content of heavy metals and arsenic in the GSMP

Parameter	Method description
Total content of heavy metals and As	Total content (Ruppert 1987)
	Total content (modified method of Ruppert 1987)
	Total content extracted with HF/$HClO_4$ (DIN ISO 14869-1 01/2003)
	Total content extracted with HF/HNO3 under pressure in microwave (GAFA 2005)
Aqua regia extractable content of heavy metals and As	Aqua regia (DIN ISO 11446)
	Aqua regia (federal state specific method with adapted HNO_3 – HCl proportion)
	Aqua regia (federal state specific method with no further method description)

However, in most cases the mandatory chemical and physical soil characteristics as given by Barth et al. (2001) are analyzed with the equivalent or comparable analytical techniques. This means it is possible to recalculate the results of different methods to obtain comparable results (Ad-hoc Arbeitsgruppe Boden 2005b, Schilli et al. 2011) for these parameters. In spite of the federal competences, it can be stated that the GSMP reveals an appreciable degree of harmonized methods and coverage of different landscapes and ecosystems (Schröder et al. 2004).

The GSMP can be characterized as a typical (national) program that fulfills all of the criteria for a comprehensive soil monitoring. The broad spectrum of physical, biological, and chemical parameters (Barth et al. 2001) enables the program to detect changes in soil characteristics beyond current tasks. The measurement of additional climatic parameters, especially at intensive monitoring sites (Barth et al. 2001), supports analyses regarding climate change. The GSMP includes several parameters like biodiversity data (Morvan et al. 2008) and bulk density and soil depth (Brammer, H. & Nachtergaele, F. O. 2015) which are often disregarded in soil monitoring programs.

The missing harmonization of methods, with respect to analytical procedures can be moderated by the often comparable or equivalent results of the different methods (Kaufmann-Boll et al. 2011). The GSMP should provide a dataset that is typical for a comprehensive (national) soil monitoring dataset and should be suitable to fulfill the determined aims of the program.

2.2 Using soil monitoring data for research

Soil monitoring is a challenging task generating huge datasets. The datasets of soil monitoring networks often include very heterogeneous data because of such things as different site characteristics, sampling depths, land use, and various involved agencies. The GSMP has generated a very large multivariate and heterogeneous dataset, which is compiled in the bBIS dataset. The analysis of such large and heterogeneous datasets is difficult because of the complexity of the investigated system soil, especially under different conditions (e.g. parent material and land use) and its interdependencies with other compartments of the environment. Such datasets require the use of adequate statistical approaches to fulfill the monitoring aims (Arrouays 2014).

The GSMP generates two different types of datasets to meet the requirements of different aims. One type is generated at basic monitoring sites to investigate the soil matrix. According to Barth et al. (2001) the soil should be characterized by numerous biological, chemical, and some optional physical parameters like soil-water content or penetration resistance.

The second type of dataset is generated by intensive monitoring sites. It is characterized by a large number of replications due to biweekly or monthly sampling of soil solution and determination of chemical parameters. Recommendations for biological and physical parameters (Barth et al. 2001) are not available for soil solution data.

Such large multivariate datasets, such as the pan-German soil monitoring dataset, open new opportunities for improving our understanding of soils, evaluating local peculiarities, identifying the onset of trends as well as providing unexpected results without requiring any pre-defined hypotheses. Single parameters or a combination of parameters are defined for special questions like soil fertility. Different chemical, biological and physical parameters can give information about varying aspects of soil quality (Zornoza et al. 2015). In most cases impacts on processes in soil influence a number of variables. For example, the podzolation is characterized by specific dynamics of e.g. pH and the contents of organic carbon, iron, and aluminum in soil matrix and soil solution (Lundström et al. 2000). Therefore, when monitoring datasets are analyzed the statistical approaches should be multivariate and explorative to identify and describe (unknown) characteristics and trends as well as relevant processes.

To investigate relationships or patterns in data, numerous approaches of clustering, ordination or trend analyses can be used. The use of such statistical approaches like e.g. multi linear regression (Riek et al. 1995, Guckland et al. 2012, Ayoubi et al. 2014), Cluster analysis (Fabietti et al. 2010, Morrison et al. 2011) or Principal component analyses (PCA) (e.g. Abollino et al. 2002, Langer & Rinklebe 2009, Ou et al. 2013, Kelepertzis & Argyraki 2015) are well known in soil science and have often been used. Most of the used approaches are linear, in spite of the documented occurrence of nonlinear effects in the environment (Ramette 2007, Lischeid 2014), in earth (e.g. Beresnev & Wen 1996) and also in soil science (e.g. Manzoni et al. 2004, Biswas et al. 2013, Shcherbak et al. 2014). Innovative nonlinear statistical approaches are used in different empirical and ecological science to deal with this problem (e.g. Isakin & Teplykh 2011, Maassen et al. 2015). There is a need for research to identify possible benefits when using such approaches to investigate soil monitoring data (Figure 2).

In addition to the different soil monitoring objectives, monitoring can vary at different scale levels (from the continental to the local scale) as well as the impacts do. Therefore it seems to be necessary to have a look at different scales when analyzing soil monitoring data. The relevance of the different scales when monitoring soil quality and organic carbon in soils is emphasized by e.g. Wang et al. 2015 and Teng et al. 2014 who stated, that it is necessary to observe soils at national, regional and local scale.

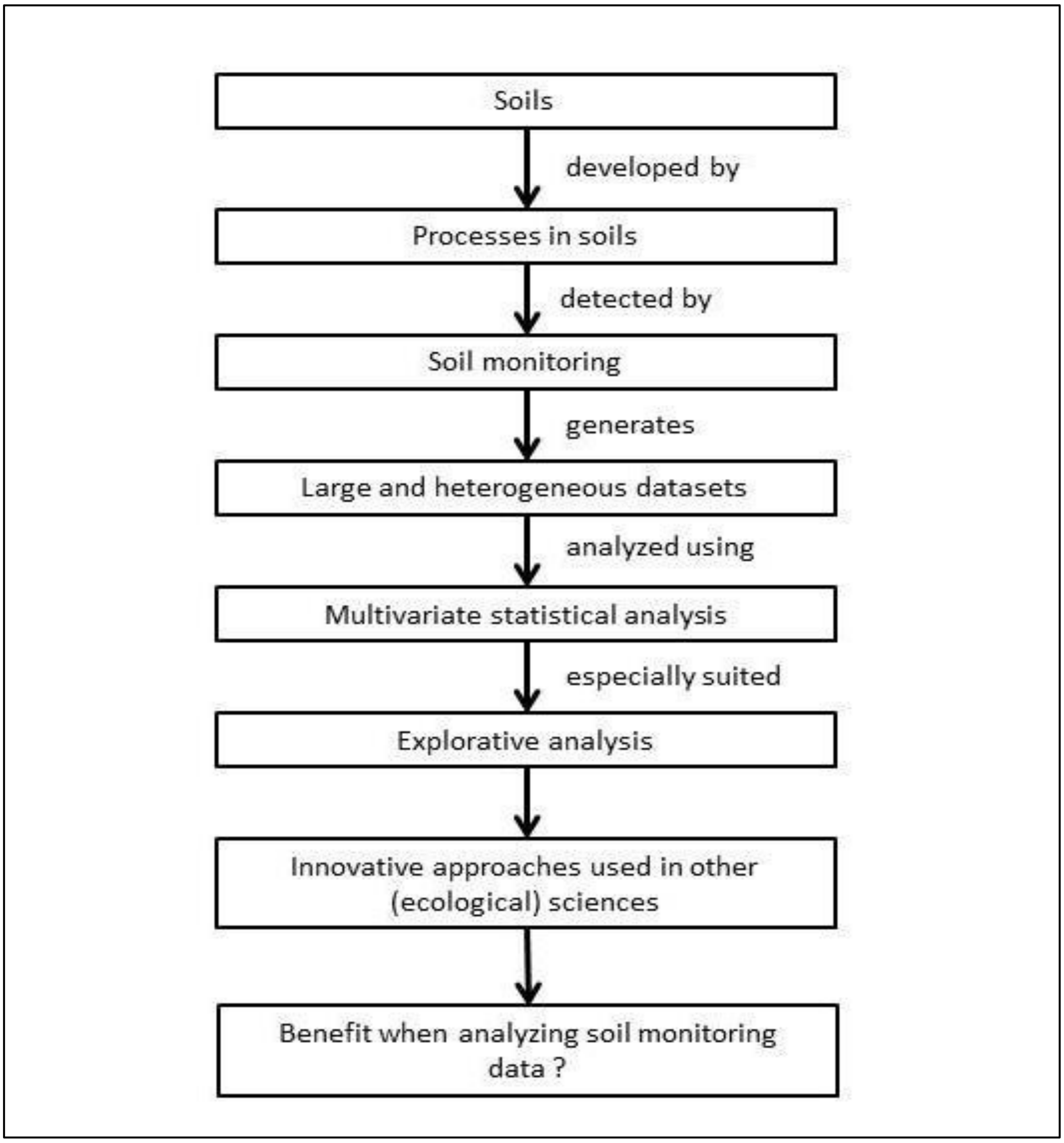

Figure 2: Derivation of the need for research

2.3 Research aims

By summing up the different tasks and utilizing the available data from soil monitoring the aims of this thesis can be defined as testing modern statistical approaches to make more efficient use of large and multivariate soil data from extensive monitoring programs in order to

- Identify the main impacts and processes on soil quality;
- Detect long-term shifts based on a large number of replicates but only a low number of temporal replications;
- Draw conclusions with respect to the suitability or need for modifications of the respective soil quality monitoring programs in order to address these questions.

To that end, advanced nonlinear methods were applied to soil quality data from various monitoring programs at different spatial scales that have rarely been applied in soil science before.

The possible relevance of scale effects requires the use of data at different scales. For this reason three different scales, with three different datasets were chosen for the analyses (Figure 3). The national scale is chosen to represent the small scale due to the available data pool of the basic monitoring sites of the GSMP (Chapter 4.1). The analysis for the medium scale deals with 5 plots from three intensive monitoring sites of the GSPM (Chapter 4.2). Four intensive monitoring sites within the Lehstenbach Catchment (Matzner 2004) will represent the large scale (Chapter 4.3).

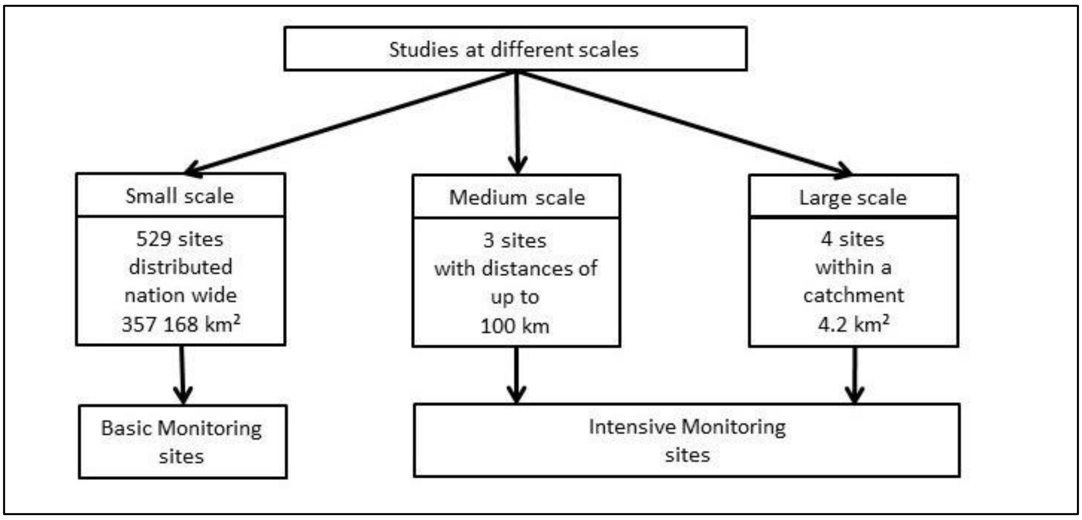

Figure 3: Study scales and data

3. Statistical methods

When analyzing complete monitoring datasets high demands are placed on the used statistical approaches. The demands on analyzing the basic and intensive monitoring datasets have some commonalties. Both large and multivariate datasets should be analyzed to uncover characteristics, processes, impacts, trends et cetera, without requiring any pre-defined hypotheses. For this reason methods of classification or pattern recognition, often used when looking for hidden characteristics in datasets, have a crucial drawback (Bahrenberg et al. 2003). They try to compare the unknown structures within the dataset with known reference patterns to assign classes or they use distinctive classifications which emphasize differences between the classes. In the latter case the variability within the group is not taken in account, which reduces the suitability.

The use of such “biased” methods impedes the exploration of new or unexpected information. Therefore, adequate explorative approaches of data mining should be used. Besides the knowledge of promising and powerful statistical approaches in fields outside the ecological research the use of the well-known and established statistical approaches, e.g. nonmetric multidimensional scaling (NMDS), PCA or correspondence analysis, are preferred by the bulk of ecological researchers until today (de Carvalho et al. 2015). Nevertheless, there is as necessity to try and prove the use of modern statistical approaches especially for analyses of extensive datasets (Lischeid et al. 2016).

As presented in chapter 2.1 the monitoring datasets include a large number of parameters. Preceding studies frequently considered only one or two of the numerous monitoring parameters to be compared with each other like in the studies in Saxony-Anhalt (Tischer 2015) and North Rhine Westphalia (Haag et al. 2015). Since soil characteristics and processes mostly are affected by or affect more than one parameter (Lundström et al. 2000, Schulz-Zunkel 2015) and one measured parameter is usually influenced by numerous different impacts and processes (Lischeid 2014). Therefore the approaches should be multivariate to regard possible interdependencies between the measured variables (Arrouays et al. 2011, Zornoza et al. 2015).

The presentation of results of high dimensional datasets can be difficult because of the various measured parameters and interactions between them. The interpretation of high multivariate datasets can be supported by dimension reduction (De Carvalho et al. 2015, Fernandez et al. 2015). Approaches of dimension reduction try to display a large amount of the information which is included in a dataset with a lower number of dimensions. A typical and well-known kind of such an approach is the PCA which tries to display a large part of the variance within a dataset by only a few components. Each component reflects a structure in the dataset which can be interpreted scientifically (Webster 2001). Further approaches try to “map” the variance of a dataset in a low dimensional space to visualize the included information. E.g. the results of the PCA or other multivariate approaches can be presented in diffusion maps (Chen et al. 2008, Fernandez et al. 2015). Other approaches, like the Self-Organizing-Map (Kohonen 2001), create their own plane of projection to display the variance of a dataset.

Common approaches like PCA account only for linear relationships. On the other hand there are often nonlinear structures in environmental dataset sets (James & McCulloch 1990, Tarquis et al. 2011, Lischeid 2014). Nonlinear structures cannot be captured using linear approaches (Kerschen & Golinval 2002) without transforming the dataset, but the transformation of data to create linearity should be avoided (Lane 2002) because of a possible loss of information. Consequently nonlinear statistical approaches often achieve better results than linear approaches when analyzing ecological datasets to identify relevant processes or impacts (Tennenbaum et al. 2000, Mahecha et al. 2007, Schilli et al. 2011, Lischeid 2014). In soil science (explorative) nonlinear statistical approaches are rarely used, whereas they are more common since several years in other empirical sciences (Tennenbaum et al. 2000, Lane 2002, Isakin & Teplykh 2011) and ecological sciences, as climate research (Gamez et al. 2004), hydrology (Böttcher et al. 2014, Maassen et al. 2015) and molecular biology (Scholz et al. 2005). Nevertheless, the use of such statistical approaches remains comparatively uncommon in soil science, although their use and necessity in soil science is known (Tarquis et al. 2011).

When analyzing datasets from soil monitoring networks there is a requirement for determination of statistical changes with adequate accuracy and power (Desaules et al. 2010, Schilli et al. 2011, van Wesemael et al. 2011).

In this thesis nonlinear explorative statistical approaches will be used. It will be examined, if promising nonlinear approaches successfully applied in other ecological sciences are provide any benefit when analyzing soil monitoring data. The approaches were evaluated and selected with regard to their applicability to the different kinds of datasets generated by basic and intensive soil monitoring. Justifications for the selection and the description of the respective statistical methods are given in the corresponding chapters.

4. Case Studies

In this chapter three case studies are presented. The studies vary in scale, from a nationwide scale (area of ca. 360.000 km²) down to the catchment scale (4.2 km²). The small scale study (Chapter 4.1) comprises the entire FRG and analyses the soil matrix data of the basic monitoring sites of the GSMP. The medium and large scale study is based on soil solution data. In chapter 4.2, the medium scale study, data from three different intensive monitoring sites of the GSMP up to 100 km apart from each other were used. Chapter 4.3 presents results based on data from intensive monitoring sites independent of the GSMP. While chapter 4.1 focuses on the visualization of the results without quantifying the impacts, the medium and large scale study compare different statistical approaches to quantify the main drivers influencing the soil solution chemistry.

4.1 Visualization of data of the basic monitoring of the GSMP

4.1.1 Introduction

There is urgent need for environmental and soil resources authorities for efficient tools to address these challenges with those large datasets (Arrouays et al. 2012). These tools are required for the initial step of the analysis of large multivariate datasets in order to structure subsequent steps of the analysis. To present data of high dimensionality or complex relationships data visualization is used since several decades. The use of visual data mining, especially if the dataset is large and the previous knowledge is low, has been proven (Keim 2002, Vesanto 2002, Chen et al. 2008). These tools should definitely not replace well-known existing statistical methods but can be used to give information to adequate further investigations.

Data visualization uses the human eye's and the human brain's excellent ability to detect patterns in visual sensations. The human brain is the most powerful tool when it comes to pattern analysis and exceeds the performance of computers by far (Chen et al. 2008). The human visual cognition is able to reduce an extremely high number of input data to low number of relevant information faster than every computer. The human eye can collect about 10^{10} bit/s using the photoreceptors of the eye. The amount of data is reduced intuitively when passing the visual nerve and frontal lobe down to 100 bits/s of essential information that are recorded cognitively (Welsch et al. 2013).

Different patterns and connections within the visually recorded information are extracted fast and automatically by the human brain (Dulclerci & Tavares 2003). Highly inhomogeneous and noisy data, which can be expected in environmental monitoring, can be handled by the flexibility and creativity of the human brain and its cognitive performance conducting visual data exploration (Keim 2002). By data visualization of large datasets the information complexity is reduced and the users knowledge acquisition is supported (Koua 2003).

Thus, data visualization is considered as the most powerful interface between large datasets, stored in a computer, and the human brain. Consequently, maps and computer screens play an outstanding role in science as well as in every day's life (Kohonen 2001, Keim 2002). The methods of visualization vary from presentation in simple graphics (e.g. diagrams) to complex approaches. Here approaches of Data Mining and Machine Learning can be taken into account to simplify mapping processes and visual representation (Dulclerci & Tavares 2003).

An explorative Data Mining tool focusing on data visualization is the Self-Organizing Map (SOM). The SOM is a kind of an artificial neural network and was introduced by Kohonen 2001. The SOM has been successfully used in different ecological (Annas et al. 2007, Olawoyin et al. 2013, Lischeid et al. 2016) and soil science studies (Merdun 2011, Voyslavov et al. 2012, Ye et al. 2015). Rivera et al. 2015 explicitly emphasize the applicability of the SOM for data exploration and recommend the SOM for analyses of complex monitoring dataset including physical, chemical and biological parameters. The performance of visual interpretation of the SOM can be enhanced by further statistical analyses (Vesanto 2002).

The SOM belongs to the field of unsupervised learning methods which makes it suitable to discover unknown patterns and interdependencies, because no predefined category information or classes are used when data training is conducted (Koua 2003). On the other hand the SOM allocates datasets to certain distinct units, which can be a drawback when looking for hidden characteristics (Bahrenberg et al. 1996, cf. chapter 3). If relations between data sets within an SOM unit should be interpreted, a further projection regarding and visualizing their similarity is necessary (Kohonen 2001). The static class boundaries should be resolved and possible transitions visualized. Kohonen (2001) suggested combining Self-Organizing Maps with Sammon's Mapping (Sammon 1969) to increase visualization quality.

The requirements given above result in the selection of the SOM as a statistical approach, supported by the Sammon's Mapping (SM) approach (Sammon 1969), as recommended by Kohonen (2001) and conducted by e.g. Lischeid et al. 2016 using hydrological data. The SOM is known for being able to handle multivariate noisy datasets and possible nonlinearity in the dataset, which cannot be excluded in soil monitoring datasets (Kohonen 2001, Rivera et al. 2015).

An additional aim of the statistical analyses was to perform multivariate trend analyses. The approach should be able to perform this even with a low number of temporal replications but a large number of spatial replicates, as described for the available dataset of basic monitoring sites of the GSMP (Schilli et al. 2011, cf. chapter 2.1). To handle this problem an innovative way to analyze the result of the combined Self-Organizing Map and Sammons Mapping (SOM-SM) with regard to time trends is presented (chapter 4.1).

Here, this approach will be applied to the data of the nation-wide German long-term Soil Monitoring program. Parts of this dataset have been used in various studies to test pre-defined hypotheses (Mindrup et al. 2011, Olbricht 2011), mostly within the borders of the different federal states and only for a single variable. In other cases numerous variables are regarded, but each one was used to explain the change of contents of a single one. Such studies using GSMP data were presented by e.g. Nerger et al. 2016 analyzing changes of TOC in a dataset of the federal state Schleswig-Holstein or Haag et al. 2015 analyzing changes of biological and chemical variables in a dataset of North-Rhine Westphalia.

The main objectives of the study are:

1) To explore the large and multivariate dataset from the GSMP and visualize the current state of soil characteristics and the relationships with other the soil properties;
2) To determine multivariate long-term trends with many spatial replicates but a limited number of temporal replicates;
3) To evaluate the suitability of the SOM-SM approach for analyzing monitoring data in soil science with respect to monitoring goals.

4.1.2 Material and methods

4.1.2.1 Dataset and data preposition

The statistical analyze is based on the bBIS dataset, provided by the UBA. The bBIS contains data from 733 monitoring sites within Germany. Data from more than 8.000 soil samples from nearly 3.800 different organic and mineral soil horizons are available. Over 330.000 analyzed soil matrix's of different physical, chemical and biological parameters are quality-assured and could be compiled in to the bBIS. This study aimed at including as many samples and as many parameters as possible. The selection of the dataset was conditioned by the selected approach of the SOM, which does not except blank values in the data (Kohonen 2001).

According to this, each parameter should be measured at each soil sample within a dataset in the ideal case. In the dataset the most frequently and nationwide measured chemical parameters are total organic carbon (TOC), total nitrogen (N), pH in $CaCl_2$ (pH) as well as the concentration of cadmium (Cd), chromium (Cr), copper (Cu), lead (Pb), nickel (Ni), and zinc (Zn) in aqua regia (Table 2) (Schilli et al. 2011).

Table 2: Regarded parameters and number of measurements

TOC	N	pH(CaCl2)	Cd	Cr	Cu	Pb	Ni	Zn
5023	5492	5627	4135	4246	4324	4357	4368	4384

The mandatory biological and physical parameters were measured in less than 1000 samples each. The exception is soil texture with >5000 measurements (Schilli et al. 2011). In this study soil texture was treated as an independent variable that could explain a part of the variance of soil quality.

In a first step of data preparation blank values in the data sets were filled. Missing values could be deduced by results of further parameters. For example TOC was calculated in 430 times by subtracting the total inorganic carbon from total carbon when these parameters were measured. Missing pH($CaCl_2$) were estimated by existing measurements of pH(KCl) and pH(H_2O) or by calculated soil texture and horizon specific means, based on the pH measurements. The uncertainty of pH estimation in $CaCl_2$ based on the measurement with other methods is accepted, because the resulting values should be close to reality (Schilli et al. 2011).

The second step replaces values below detection limit by zero. The decision against the replacement by half or quarter of the detection limit is justified by the different detection limit in the data of different federal states and the often decreasing detection limit with time. Half of the detection limit for e.g. Cd at the beginning of the monitoring period in a federal state was more than double the youngest detected value in the dataset of another federal state (Schilli et al. 2011).

In a last step samples with two or more missing values were excluded from further statistical analyses, because of the need of complete data sets. Single missing values were replaced by land use and horizon specific means of the respective parameter. This was carried out for Ni (1 case), Pb (4), Cu (23), Cr (23), pH (113), Cd (148), TOC (200) and N (204). To assign equal weight to the variables the data were z-normalized (mean = 0, standard deviation = 1) separately for each variable. In total, 3724 samples from 2343 different horizons and 529 sites met the requirements, covering a period from 1985 to 2007 (Schilli et al. 2011).

The dataset comprises samples from organic and mineral soil horizons of various land use classes and of different parent material. Consequently, the samples exhibited large differences of soil bulk density. Therefore it would have been desirable to calculate the stocks of the elements and include them into the statistical analyses. However, since those data were not available from all sites and horizons the analyses were restricted to concentration data (Schilli et al. 2011).

4.1.2.2 Statistical methods

A combination of the SOM and the SM was chosen to analyze the GSMP dataset. The SOM-SM aims at visualizing large high-dimensional datasets in a low dimensional, usually two dimensional, projection (Kohonen 2001). The SOM is a very efficient tool to discover structures and dependencies in large multivariate datasets and is recommended for Data Mining in ecological datasets (Giraudel & Lek 2001, Merdun 2011) and soil monitoring data (Rivera et al. 2015) especially if efficiency and visualization is improved by combining the SOM with further approaches such as Sammon's Mapping (Kohonen 2001, Lischeid 2014).

Self-Organizing Map

The SOM tries to display a high dimensional dataset in a 2-dimensional, as in this study, or 3-dimensional space. As an unsupervised artificial neural network the SOM clusters and projects the dataset in an iterative training procedure (Kohonen 2001). In this map patterns can be visualized which can be interpreted as typical characteristics of the dataset, respectively the analyzed soils.

In a first step an initial map is set up, which is a network with a defined number of units (in this case a hexagonal grid of 60 times 40 units). Each unit can be understood as a vector of the same dimensionality as the underlying dataset. The vectors are initialized randomly. Therefore, if the SOM generation is repeated, the results will never be exactly the same, but very similar (Kohonen 2001). In general, the SOM performs a kind of two-step cluster analysis, aiming first at assigning similar data vectors (i.e., soil samples with similar parameter values) to the same units, and second arranging units with similar data vectors close to each other. Figure 4 provides an example of a SOM-grid with 5 times 3 units and three dimensions (i.e., the number of the observed parameters).

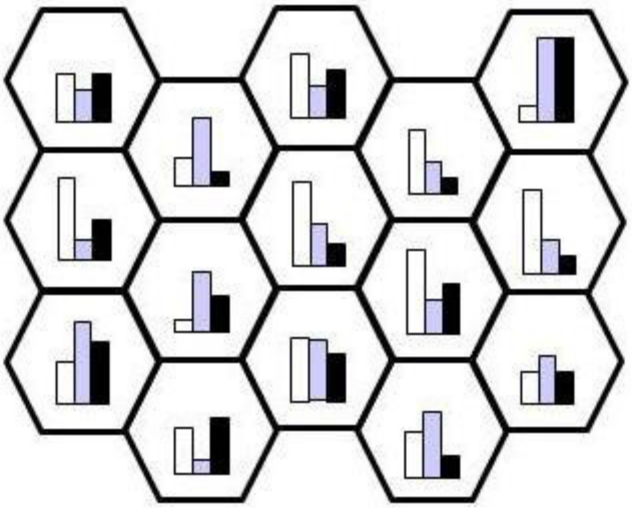

Figure 4: Scheme of an SOM grid with three variables

In an iterative “learning” procedure each sample of the dataset u_d, is presented to the network and the “winnig-unit” u_w is identified which is the most similar to the respective sample. The winning unit u_w is then updated to match the values of the sample u_d even a little bit more closely:

$$u_w(t+1) = u_w(t) + h(t) * (u_d(t) - u_w(t))$$

where t is the learning index and t_{max} is the number of learning steps. In this study 800 learning steps were used. Besides the winning unit nearby units in the network are updated as well, but to a lower degree, defined by the neighborhood function h(t). Here a Gaussian function was used:

$$h(t) = \alpha(t) * \exp\left(-\frac{\|u_w - u_{all}\|}{2\sigma^2(t)}\right)$$

The term $||u_w - u_{all}||$ is the multivariate Euclidean distance between the winning unit and each dataset unit on the map. The learning procedure is defined by the α(t) and σ(t) parameters which both approach 0 with increasing number of learning steps, where $\alpha(t)$ is the “learning-rate factor” and $\sigma(t)$ is the width of the kernel of the neighborhood function. The latter covers 2/3 of the units of the SOM in the beginning of the learning procedure.

Sammon's Mapping

The SM approach uses the results of the SOM and arranges them corresponding to the multivariate similarity of the data. Sammon's Mapping was developed to project a high dimensional dataset into a low-dimension space (Sammon 1969). Starting from a random initialization or with the scores of the first two principal components, the location of single data vectors is stepwise adjusted in an iterative procedure in order to maximize the correlation between interpoint distances in the high-dimensional data space and the low-dimensional projection. As recommended by *Kohonen* (2001) the coordinates of the 2-D projected SOM units were used for initialization instead.

The interpoint distances of the original data set (d_{ij*}) and the interpoint distances of adapted SOM coordinates (d_{ij}) are used to define an error function E

$$E = \frac{1}{\sum_{i<j}[d_{ij*}]} \sum_{i<j}^{N} \frac{[d_{ij*} - d_{ij}]}{d_{ij*}}$$

The error is stepwise minimized ($t_{max} = 100$) by adapting the SOM coordinates x_{SOM} by

$$x_{SOM}(t+1) = x_{SOM}(t) - MF * \Delta(t)$$

where $\Delta(t)$ is

$$\Delta(t) = \frac{\partial E(t)}{\partial x_{SOM}(t)} \Big/ \frac{\partial^2 E(t)}{\partial x_{SOM}(t)^2}$$

and α is the learning rate ($0 < \alpha < 1$).

It has been shown that combining SOM and SM (SOM-SM) is a very powerful tool for low-dimensional projection of large multivariate datasets (Kohonen 2001). The application of SM supports the visual pattern recognition and allows trend analyses, even with a low number of temporal replications if a larger number of spatial replications are given, which will be presented in chapter 4.1.3.

For statistical analyses and creating diagrams the R-software package (Version 2.10.0) (R Development Core Team 2006), the Kohonen extension (Version 2.0.5) for Self-Organizing Maps (Wehrens & Buydens 2007) and the MASS Extension (Version 7.3-13) for Sammons Mapping (Venables and Ripley 2002) which are available at *http://www.r-project.org* were used.

4.1.3 Results

The outcome of the SOM-SM is a single graph where every data point denotes one sample of the dataset. In contrast to more common ways of visualization of data, the location of the symbols within the graph, and their projections on the axes do not bear any information. Thus, coordinates are given only for orientation. Here, every sample is represented by a symbol. Distances between any two symbols can be interpreted to be proportional to the dissimilarities of the respective soil samples with respect to all nine parameters. Thus, the more similar any two samples are, the closer they are plotted to each other in the graph. In this study, correlation between dissimilarities of the soil samples with respect to the nine variables and distance in the SOM-SM is $r^2 = 0.94$.

The same graph will be shown with various colour codes or gray shadings for different values and characteristics of the samples. In this regard the SOM-SM can be compared with a topographic map, where elevation is colour-coded and (usually) does not increase continuously along the x- or y-axis. In this study colour codes used for concentration of different parameters (Table 2) as well as further information such as land use, soil texture etc.. Please note that only the concentration of Cd, Cr, Cu, Pb, N, Ni, TOC, Zn, and pH values have been used for setting up the SOM-SM. However, every soil sample can be associated with soil type, land use, etc. and thus, the colour coding can reflect these properties as well. Therefore, sample or site information can be considered and displayed in the graph even if not available for all samples. For example soil texture was not measured in organic soil samples e.g. peat soil or forest floor horizons. In Figure 5 to Figure 8 certain data points of the total dataset are highlighted, and the remaining data points are represented by light grey symbols to allow comparison between different figures. About 1.6 % of data points would be plotted outside the range presented in Figure 5 to Figure 8 and will not be considered in the following, allowing focusing on the majority of the samples.

In general, the more the samples from a specific subgroup spread over the graph the larger the variance of the respective parameter values. The more clear the pattern of the values of a parameter are, e.g., a clear increase of parameter values from one side to another side of the graph, the more this parameter is correlated with other parameters. Correspondingly, similar patterns for different parameters in the SOM-SM point to a strong correlation between the respective parameters. In contrast, gradients that run perpendicular to each other indicate that there is no correlation between the respective two variables.

TOC and pH

Figure 5A shows a high density of symbols in the upper range of the graph and a lower density towards the lower left corner. In addition, the cloud splits into a larger cluster to the right and a small cluster to the left, as illustrated by a solid line. Accordingly, the dataset can be divided in two groups of different chemical characteristics. Please note that this line has been added for orientation only and to support the visualization.

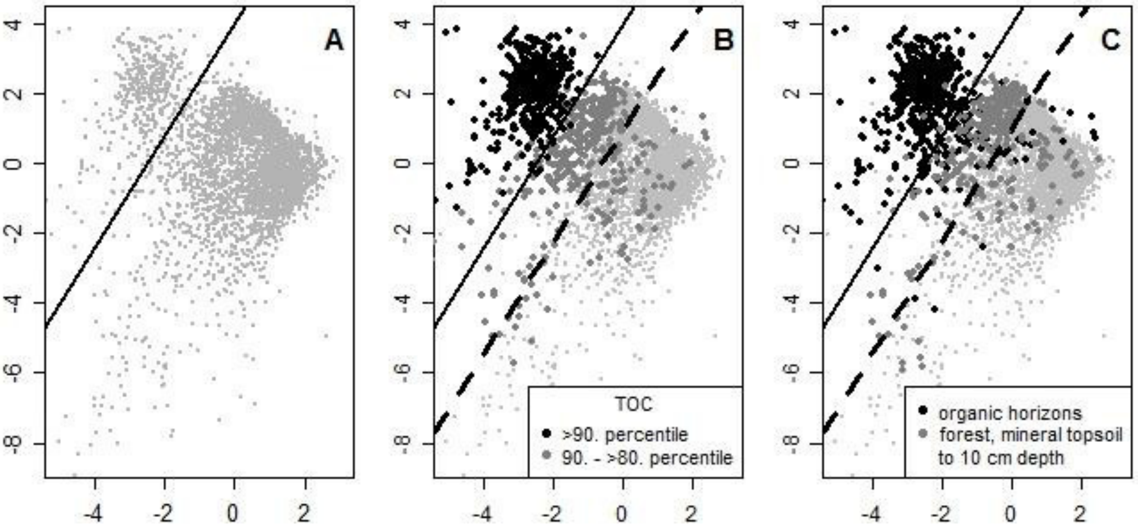

Figure 5: Total organic carbon of soil samples, and samples from organic horizons and from mineral topsoil horizons shown in the SOM-SM graph.

In Figure 5B samples with high TOC content are highlighted. The smaller cluster (left of the line) is characterized by TOC exceeding the 90. percentile (that is, > 27.76%; cf. Table 3). Samples with slightly lower TOC contents plot next to these in the bigger cluster (80.-90. percentile; Figure 5B) between the solid and the dashed line. Here, TOC contents are between 6.5% and 27.76%. The TOC indicates a tripartition of the SOM-SM in organic samples from forest floor and peat soil horizons, and mineral samples (Figure 5C) as indicated by TOC contents (Figure 5B). The solid line divides the small cluster with samples from the organic horizons (organic cluster, left of the line) from the cluster with mainly mineral samples (mineral cluster, right of the line). The mineral cluster can be differentiated by the dashed line in mineral horizons of forest soils down to 10 cm depth (mineral cluster; left of the line) and samples from soils under agricultural use or deeper mineral horizons of forest sites (mineral cluster; right of the line). The patterns of N and TOC are very similar (r = 0.75). Therefore, N will not be discussed separately.

Table 3: Statistical values of the studied parameters (n=3724), Q = Quantile

	Cd	Cr	Cu	Ni	Pb	Zn	pH	TOC
	mg kg $^{-1}$							%
Min.	b.d.l.	b.d.l.	b.d.l.	b.d.l.	b.d.l.	b.d.l.	1.90	b.d.l.
Q 10	b.d.l.	4.24	2.01	2.90	3.02	12.8	3.50	0.12
Q 20	b.d.l.	8.39	4.34	5.60	8.82	23.1	3.99	0.30
Q 30	0.03	12.8	7.00	8.70	12.60	32.9	4.39	0.49
Q 40	0.08	17.9	9.30	11.9	16.9	42.9	4.83	0.80
Q 50	0.11	24.0	11.4	16.0	21.6	52.0	5.31	1.21
Q 60	0.17	30.8	14.5	20.2	27.1	60.8	5.90	1.79
Q 70	0.28	39.7	18.1	26.5	35.3	72.0	6.35	2.97
Q 80	0.48	50.0	24.00	33.0	51.0	89.0	6.88	6.50
Q 90	0.89	67.0	38.2	45.4	92.6	128	7.30	27.7
Max.	103	361	914	544	7985	5905	8.20	81.7

The pH shows a different pattern for mineral and organic samples (Figure 6A). Interpretation of the pattern has to account for the fact that the symbols partially overlay each other. The sample points were coloured beginning with the first quintile and ending with the fifth quintile. The mineral samples reveal a very clear gradient. The pH-values exhibit a clear increase from the upper left to the lower right part of the cluster. Most of the high pH-values belong to samples from arable soils, while low values belong to upper horizons of forest soils (Figure 5C, Figure 6A). In the smaller organic cluster hardly any samples with pH-values of the fourth and fifth quintile are found. Here, the gradient is roughly perpendicular to that of the mineral cluster, increasing from the lower left to the upper right. However, the gradient is less clear compared to the mineral soil samples cluster.

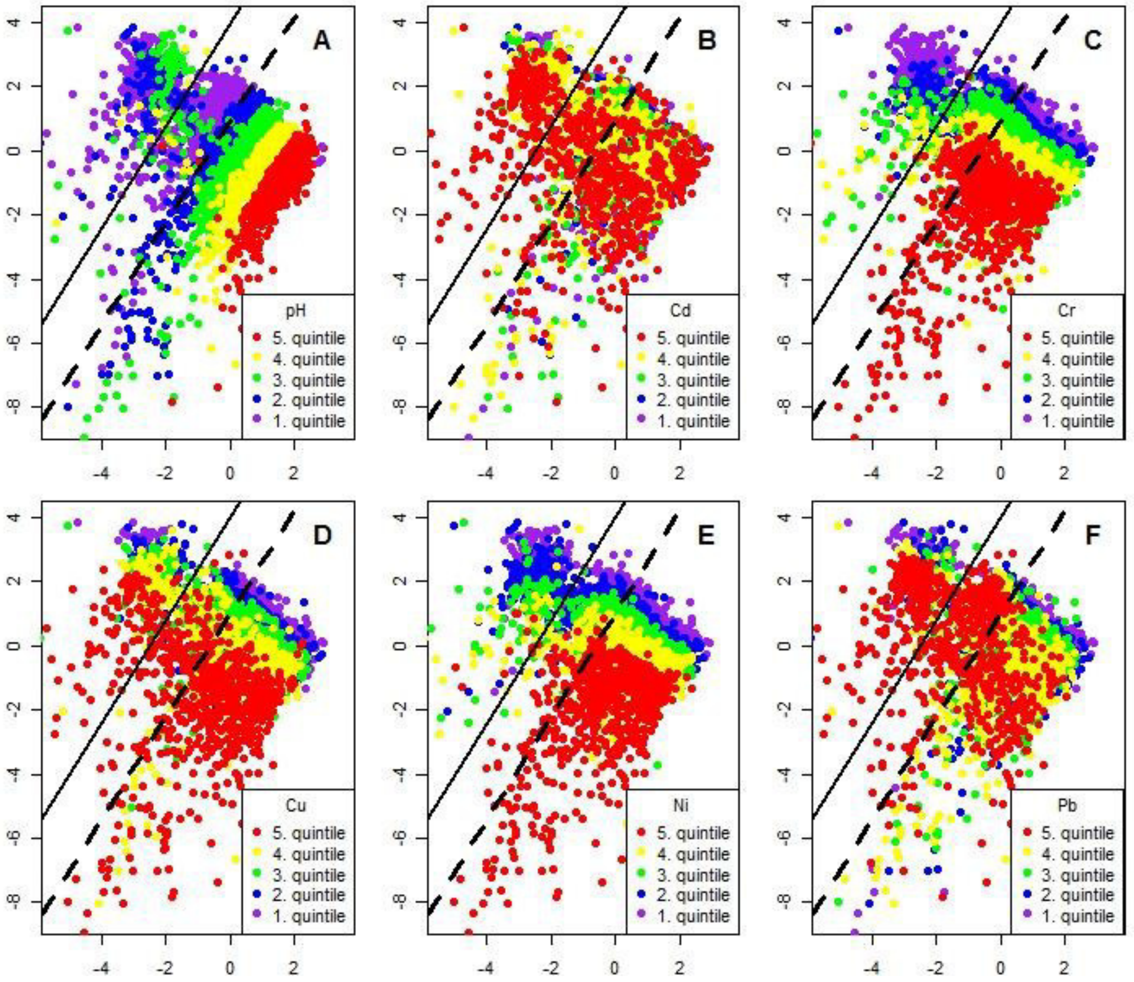

Figure 6: Heavy metal concentration and pH values of the soil samples shown in the SOM-SM.

Heavy metals

Figure 6B-F depicts the quintiles of heavy metal concentrations. In general, different elements roughly exhibit similar patterns. The results of Cu and Zn are very similar (r = 0.77). Therefore, Zn will not be discussed in detail.

In contrast to the pattern of pH values (Figure 6A) the lowest values of the heavy metals (except for Cd), were found in a small sub region along the upper edge of both clusters, and element concentration increases towards the lower left. Especially for Cr, Cu, and Ni the width of the strips for single quintiles tend to increase with increasing concentration. This seems to be due to the highly skewed distribution of the data. Concentration values in the lowest quintile are within a small range and that of the higher quintiles exhibit much larger variance

(Table 3). In contrast to the patterns of pH values and TOC content (Figure 6A, Figure 5B) the organic and mineral clusters do hardly differ with respect to the heavy metal contents. Only for Ni the cluster of organic horizons left of the solid line is predominated by the lower quintiles. The same holds true, although to a lesser extent, for the mineral topsoil samples from afforested sites located between the solid and the dashed line (Figure 6D; cf. Figure 5C). This general pattern is obvious for Ni, Cu, and less clear for Pb.

Texture and parent material

Soil texture roughly reflects the gradient of the heavy metals in the mineral cluster (Figure 7A; cf. Figure 6B-F). Sandy samples plot at the top of the mineral cluster, silty and clayey samples more towards the lower left. The latter's exhibit large variance (large scatter) and widely overlap each other. For the parent material a similar pattern can be found (Figure 7B). Samples from sand and sand-loess are located at the upper right of the cluster where low heavy metal concentrations prevail (Figure 6B-F). Limestone and marl with clayey weathering products reveal mainly mean to high contents of heavy metal elements. Samples with silty weathering products of alkaline igneous and metamorphic rock scatter to the lower left (Figure 4B) indicating high contents of heavy metals (Figure 6B-F). This relation of soil texture and geogenic content of heavy metals in German soils and the high heavy metal concentrations in silty weathering products of alkaline igneous and metamorphic rocks is in good agreement with previous results (Utermann et al. 2010).

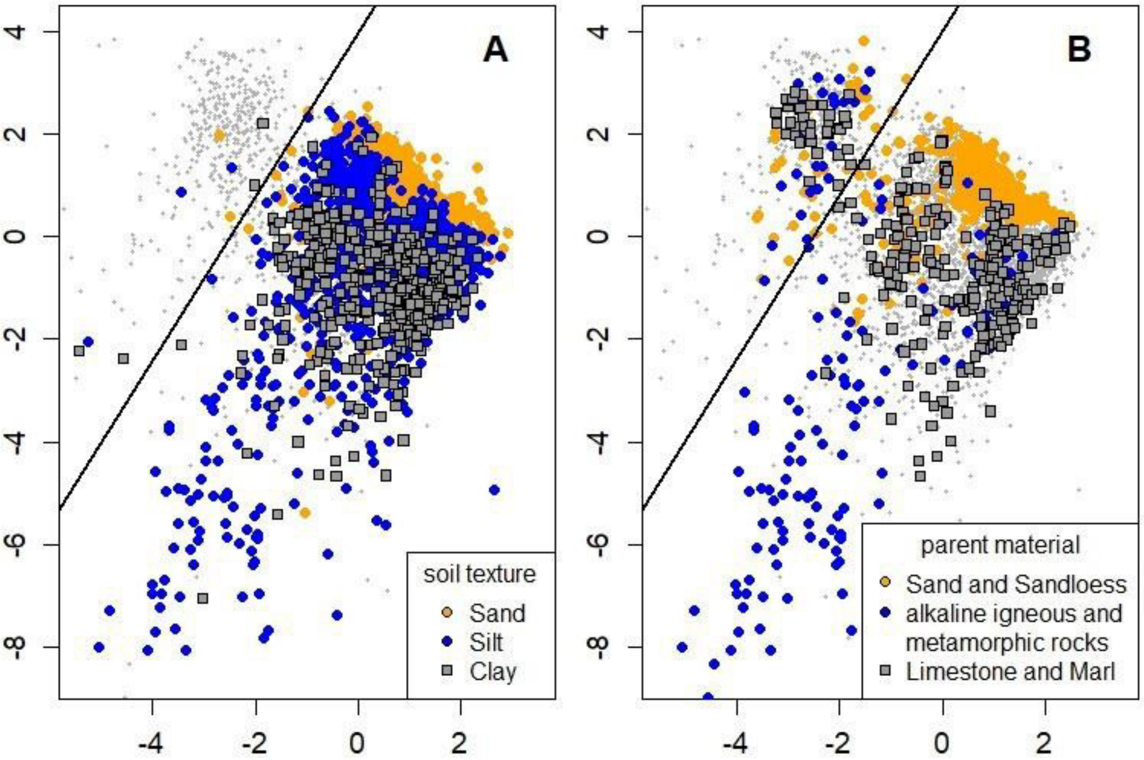

Figure 7: Soil texture and parent material of the soil samples shown in the SOM-SM graph.

In addition the data points were coloured using additional spatial information of the sampling sites, i.e., climatic region (Finke et al. 2001), type of landscape (BfN 2007), ecological classification (UBA 2001), mean temperature and mean precipitation (not shown). However, for those cases no clear patterns could be found indicating only poor relationships to the additional spatial characteristics of sampling sites.

Organic horizons

Within the organic cluster left of the solid line different soil horizons can be distinguished (Figure 8). At the top of the organic cluster (Y > 2.5) most of the samples from peat and litter horizons are plotted. The O horizons are located in the lower part of the cluster. Some samples from cultivated fens and organic horizons are located right of the continuous line, in the direction of the mineral sample cluster. Please note that the location of the samples in the SOM-SM is based only on soil chemical data without considering any information about soil type or depth. Thus, it can be concluded that these different groups of soil horizons differ systematically with respect to the soil chemical characteristics.

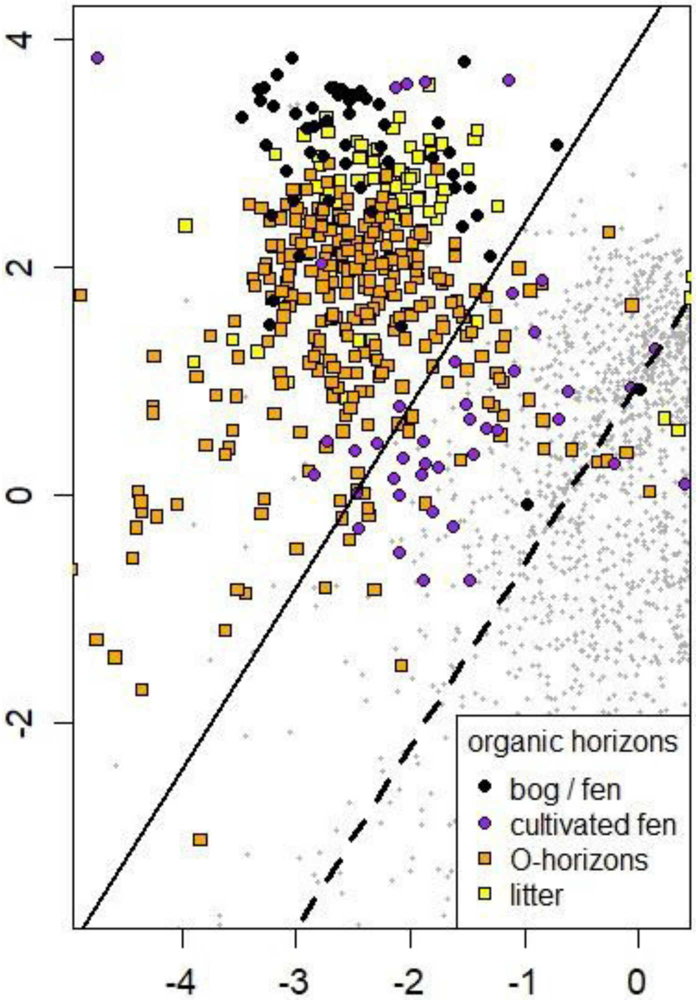

Figure 8: Samples of different organic horizons shown in the SOM-SM graph.

Multivariate Trend analysis

The SOM-SM was used for multivariate trend analysis as well. Corresponding to, e.g., comparing different groups of samples with respect to location in the graph, temporal replicates from the same sites can be compared. Multivariate trends can be identified by systematic shifts within the SOM-SM of samples from different sampling dates. Please note that any visible shift in the SOM-SM would be either due to substantial changes for single parameters, or, more likely, less substantial but correlated changes of different variables. To detect a shift the coordinates of data points in the SOM-SM of the first and last samples of the respective sites from the same horizons each were compared. Here only sites with at least ten years monitoring were considered, according to the recommendations of e.g. Bak et al. (1997) and Desaules et al. (2010).

In a first step trend analyses were conducted using samples from all depths. No clear trend could be identified (not shown). In a second step the dataset was restricted to samples from the upper mineral soil horizons (organic samples were not regarded). Differentiation between top and subsoil samples gave more clear results in a former study that analyzed heavy metal sorption in German soils with artificial neural networks (Anagu et al. 2009). Trend analyses of the organic forest floor and mineral subsoil horizons were not conducted because of the low number of available data covering ten or more years. Despite the long-time of soil monitoring the number of data is relatively low (Schilli et al. 2011).

The mean shift of paired upper mineral soil horizons samples in the SOM-SM is shown in Figure 9. In Figure 9A the calculated coordinate means of the compared land use specific youngest and oldest data points are depicted. The relative differences between paired samples are shown in Figure 9B as single points. A single point reflects the shift in the SOM-SM. The 95%-confidence interval was calculated for the shift in direction of the x- and y-axis. The ellipses in Figure 9B provide the 95%-confidence intervals of all individual time shifts of respective samples. A shift is significant, if an ellipse does not include the intersection of the axes.

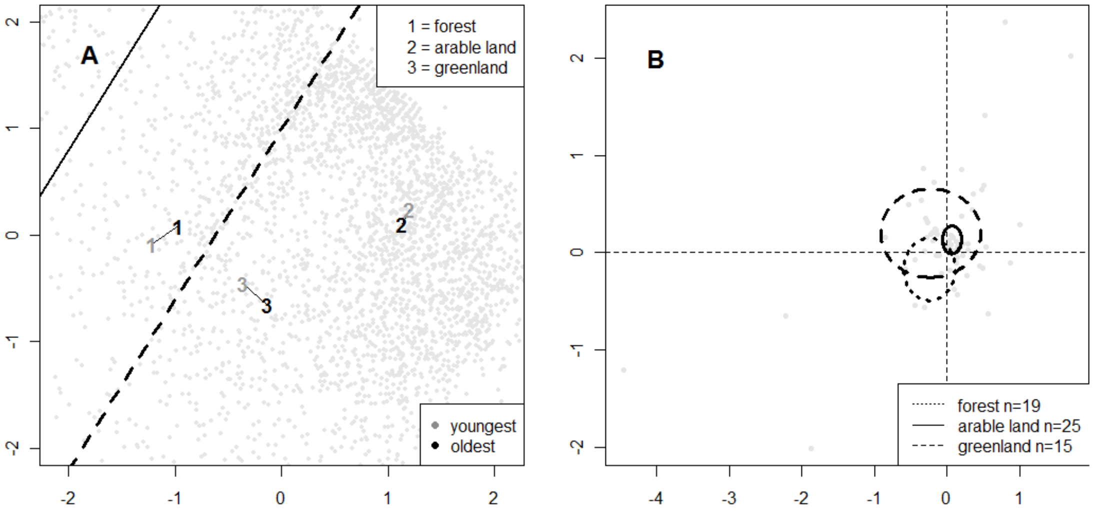

Figure 9: Trend analysis with the SOM-SM. A: Land use specific coordinate means of compared upper mineral soil samples in the SOM-SM. B: Land use specific 95% confidence interval of the shift of compared upper mineral soil samples.

One can observe a tendency for opposing trends for arable soils and forest soil, respectively, reflected by land use specific ellipses of the 95%-confidence interval. However, none of those trends is significant, indicated by the intersection of calculated ellipses and the axes. Arable soils show a shift into the direction of the upper right of the SOM-SM, while the opposite is shown for forest soils. For grassland soils, there was no clear trend (Figure 9B).

4.1.4 Discussion

TOC and pH

The pH pattern in the organic cluster (Figure 6A) might be mainly caused by decomposition of organic matter of the forest floor horizons. At the top of the cluster (Y > 2.5) most of the samples of less decomposed organic material originated from litter horizons are plotted (Figure 6A, cf. Figure 8). Litter samples reveal the similar pH than conifers or leaves. The more decomposed samples of O horizons are located in the middle and lower part of the organic cluster. The pH values are decreasing from the upper right to the lower left, indicating an acidification during decomposition (Schilli et al. 2011). This might be due to processes, such as nitrification of ammonium and the formation of sulfuric acids from sulfur oxides. The acidification is additionally supported by acidic precipitation. On the other hand the acidification of the humus layer is opposed to the decarboxylation process which consumes protons during decomposition (Barekzai & Mengel 1992). The pH gradient indicates that the consumption of protons during decomposition cannot buffer the acidic input. Some samples from pristine bogs and fens did not follow this pH gradient (upper part of the organic cluster with pH values within the first quantile). Here, the pH might reflect that of the groundwater rather than that of soil processes.

Some samples plot between the organic and mineral cluster (Figure 8, close to the solid line). Those samples originated from Oh and Ah horizons, according to the German soil classification (Wittmann et al. 1997). The location of these samples between the clusters may indicate a smooth transition between the forest floor and the underlying mineral horizons caused by biological activity mixing O and A material. On the other hand difficulties in dividing forest floor horizons from the top mineral soil during sampling (Kaste et al. 2003) might be responsible. The location of those data points indicates that the typical chemical characteristics of those samples neither correspond to typical upper mineral topsoil samples nor to the organic horizons of forest soils regarding the researched parameters. Other samples in this area can be ascribed to cultivated fens. The agricultural use of those horizons leads to higher pH (Figure 6A, cf. Figure 8). Organic horizons of cultivated fens are often mixed with sand (e.g. by plowing) and mostly drained. The agricultural use of fens influences the chemical characteristic in the direction of mineral topsoil horizons at afforested sites.

The differentiation within the mineral cluster left and right of the dashed line (Figure 5B and C) obviously is related to the accumulation of TOC and acidification of the topsoils under forest compared to agricultural soils (e.g. Andersen et al. 2002). The pH increases from the upper forest soil horizons left of the dashed line to the lower right, were arable soils dominate (Schilli et al. 2011).

A general relation of pH and TOC to soil texture was not detected (cf. Figure 5B, Figure 6A, Figure 7A). Only the top soils of arable land show a significant relation between pH and soil texture (Table 4). This is presumably due to the fact that pH was be controlled by liming or application of alkali organic fertilizers.

Table 4: Mean values and 95%-confidence interval of pH of the uppermost mineral soil horizons.

Land use	Texture	pH mean	95%-confidence interval	n
Arable land	Sand	5.76	0.15	94
	Silt	6.79	0.10	153
	Clay	7.21	0.10	48
Grassland	Sand	6.19	0.29	27
	Silt	5.58	0.43	65
	Clay	6.40	0.27	18
Forest	Sand	3.67	0.31	52
	Silt	3.98	0.18	34
	Clay	5.00	0.40	128

Chromium and Ni

In the mineral cluster all quintiles of Cr and Ni concentration are represented. Contents of Cr and Ni show clear relations with parent material (Wahsha et al. 2014, Kelepertzis & Argyraki 2015) and soil texture (Schilli et al. 2011). They generally increase from sandy to clayey samples (Figure 6C, E; Figure 7A). Several studies have previously documented that enhanced concentrations of Cr and Ni can be associated with higher clay contents (e.g. Andersen et al. 2002, Utermann et al. 2010) which is in good agreement with the presented results. The strong impact of parent material on the contents of Cr and Ni (Alloway 2013) can cover other impacts such as land use (Kelepertzis 2014, Zhou et al. 2016). Even the input of Cr into arable soils by phosphate fertilizers, which exceeds the German average of atmospheric deposition by a factor of three, is comparably small compared to the geogenic stock in soils (Nziguheba & Smolders 2008).

For Ni, less samples of the fifth quintile can be found between the lines in the topsoils of the forest sites compared to Cr (Figure 6C, E). The depletion of Ni in the topsoils of the forest sites compared to farmland is described by Anderson et al. (2002) and explained by increased weathering and leaching. The leaching of nickel within the soil profile is also described for numerous Austrian forest soils (Türtscher et al. 2017). Another possible explanation for differences between Ni and Cr could be that Ni is mainly sorbed to oxides and clay minerals (Rinklebe & Shaheen 2014) while Cr is strongly sorbed or occluded in pedogenic oxides and is sorbed by or builds up strong complexes with organic substances in soils (Alloway 2013). Generally, the mobility of Ni in soils increases faster with decreasing pH than the mobility of chromium, which should lead to a more pronounced leaching of Ni. Furthermore, Huang et al. (2011) describe an input of Cr by litterfall exceeding the sum of bulk precipitation and interception deposition in a German catchment which was not the case for Ni. A translocation of Cr from deeper soil horizons to plants and subsequent accumulation in the topsoil following litterfall is indicated (Schilli et al. 2011).

In spite of the documented translocation and input by litterfall high contents of Cr and Ni (4. and 5. quintile) were very rarely detected in the organic samples (Figure 5C; cf. Figure 6C, E). Those findings correspond very well with the observation that an important source for Cr and Ni in soils is the parent material (e.g. Andersen et al. 2002, Hernandez 2003). The input by deposition is relatively low in comparison to the stocks of the mineral soil. This is reflected by missing Ni and Cr contents of the 4. and 5. quintile in the organic samples. Even the above mentioned input by litterfall (Huang et al. 2011) is not able to compensate for those differences.

Litter decomposition and mineralization is usually connected with a mass loss of leaf material (Scheid et al. 2009). In addition there is an increase of storage capacity during decomposition due to changes of the structure of the organic material. As a consequence the concentration of heavy metals can increase (e.g. Scheid et al. 2009, Huang et al. 2011). The gradient of Ni and Cr in the organic cluster may follow the grade of decomposition.

Pristine, non-cultivated bogs and fens generally contain little decomposed organic matter. Iron and manganese oxides are often missing in bogs and fens due to anoxic conditions. The exchange capacity of the organic material of bogs and fens might be lower compared to the more decomposed O horizons. Therefore, lower concentrations of heavy metals can be detected (Figure 6C, E; cf. Figure 8) (Schilli et al. 2011).

Copper

The pattern of Cu content of the mineral soils is similar to that of Cr which indicates similar impacts (Figure 6C, D). The parent material and clay contents have been identified as important general sources of Cu by several authors (e.g. Andersen et al. 2002, Utermann et al. 2010, Mihajlovic et al. 2012, Wahsha et al. 2014) supporting the interpretation of the presented results.

For topsoils of agricultural soils (Figure 6D, mostly to the right of the dashed line) fertilizers such as manure (especially pig slurry) and sewage sludge might be relevant sources (Boysen 1992, Bak et al. 1997) as well as agrochemicals (Kelepertzis 2014, Simoncic et al. 2017). The anthropogenic impact on Cu contents should not be underestimated and can be of similar relevance than natural sources, especially in and around areas known for industrial activity (Rautengarten et al. 1995, Alloway 2013, Zhou et al. 2016). In this case study the pronounced pattern of Cu contents at the right site of the dashed line indicates that anthropogenic impacts are not able to compensate the influence of the parent material, respectively the soil texture (Schilli et al. 2011).

For forest topsoil samples (between the dashed and the solid line) the gradient is less clear. Some samples of the fifth quintile scatter to the upper right. Those samples show no similarities in soil characteristics or in spatial distribution. A possible explanation would be an influence of deposition (Schilli et al. 2011). Comparing the deposition amount and the soil stocks of Cu, the ratio is closer than for Cr and Ni, as shown by Bak et al. (1997) for Denmark. Additionally, Cu has a strong affinity to organic matter (Altaher 2001, Neupane & Roberts 2009) which is enhanced in forest topsoils compared to agricultural soils (Andersen 2002, Fabietti et al. 2010). On the other hand Neupane & Roberts (2009) did not find any enrichment of Cu in topsoils of adjacent broadleaf forest and arable soils in spite of enhanced Cu contents, presumably due to aerosol deposition. This is in accordance to the presented results, were significantly higher contents of topsoils of forest sites compared with arable topsoils of the same parent material could be not be found (Schilli et al. 2011). The less pronounced pattern could be interpreted as an evidence for the relevance of deposition as a Cu source, being consistent to previous studies (Bak et al. 1997, Zhou et al. 2016), but this cannot be confirmed here.

Copper contents of the litter and of the O horizon are within the third to fifth quintile, increasing from litter to more decomposed material of O horizons, corresponding to the results of Scheid et al. (2009). Besides deposition there could be an additional enrichment of Cu in the mineral horizons by plant uptake and subsequent litterfall (Jonczak & Parzych 2014, Türtscher et al. 2017). The observed contents in the range of the fifth and fourth quintile in forest floor samples might point to deposition and litterfall as possible sources of Cu into the soils (Schilli et al. 2011).

Lead

In contrast to Cr and Ni atmospheric deposition has been identified as an important source for Pb in soils (Zhou et al. 2016, Türtscher et al. 2017) due to the use of leaded gasoline. Lead concentration in fuel in West Germany was stepwise reduced from 0.4 g l^{-1} (1972) to 0.0 g l^{-1} (1988). This led to a substantial reduction of lead deposition since the 1980’s (Huang et al. 2011). Nevertheless, considerable amounts of the deposited lead are still stored in the forest floor (Kaste et al. 2003) and the uppermost mineral soil horizons (Hernandez et al. 2003, LUBW 2008). Corresponding to those studies high lead concentrations in the forest floor and in the upper mineral horizons of the forest sites were found (Figure 6F). Exceptions, which plot right of the dashed line, are, e.g. sites in the ore containing Harz Mountains were high concentrations of Pb are common (e.g. Bartels 1996) and floodplain soils downstream the Harz as well as exceptions like an arable soil close to a smelter which is extremely affected by deposition (Jacob 2008).

The Pb pattern in the organic cluster shows numerous samples of the highest concentration class for O horizons (Figure 6F, cf. Figure 8). Samples of fens, bogs and litter material exhibit lower Pb concentrations. During decomposition of leaf litter a significant increase of Pb was observed (Scheid et al. 2009). The high contents of Pb in O horizons give some evidence for atmospheric input and Pb accumulation in those horizons (Schilli et al. 2011, Türtscher et al. 2017). Different studies found that Pb is leached from the forest floor and accumulates in the top mineral soil (Kaste et al. 2003, Hernandez et al. 2003, LUBW 2008). In spite of the reduction of lead deposition the forest floor can release the stored Pb which is then leached to greater depth (Huang et al. 2008).

Leached Pb from the forest floor is mainly sorbed in the uppermost humic mineral soil horizons of forest sites (Kaste et al. 2003, LUBW 2008), especially in acidic soils (Degryse et al. 2009). A high affinity of Pb to organic substances is reported by, e.g. Huang et al. (2008), Neupane & Roberts (2009) and Rinklebe & Shaheen (2014). This is confirmed in this study by high Pb concentrations in the cluster of topsoil horizons of forest sites between the continuous line and the dashed line (Figure 6F). The input of anthropogenic lead might dominate over the influence of the geogenic source for mineral topsoils at the forest sites.

Compared to other heavy metals high Pb contents seem not to be closely related to finer texture (Figure 6F, Figure 7A). This is in contrast to various studies which described significant relations between total Pb and clay content (Anderson et al. 2009). Utermann et al. (2010) give background values for German subsoils for different textures, soil depths and land use. For Pb the given median of background values of silty/loamy subsoils were similar to those of the sandy topsoils regardless of land use. Corresponding to these results a clear texture gradient for Pb (Figure 6F, Figure 7A) is not recognizable. Copper and Ni reveal high concentrations in soils developed from igneous and metamorphic rock. In contrast to Cr and Ni only medium to low contents of Pb were found except for forest topsoil samples (Figure 6C,E,F, Figure 7B). This corresponds to rather low Pb contents in the bedrock (Utermann et al. 2010). Substantially higher Pb contents in the forest topsoils underline the role of deposition as an important source for Pb for these soils (Türtscher et al. 2017).

Cadmium

Among the heavy metals Cd exhibits the least clear gradient, indicating first order controls that differ from the other heavy metals. In contrast to the other heavy metals samples of the fifth quintile are located in the area of sandy and silty samples in the upper part of the mineral cluster (Figure 6B; cf. Figure 7A). Samples of the fifth and forth quintile are rarely represented in the upper part of the mineral cluster left of the dashed line, which represents acidic sandy samples of forest sites. The results of the analyses of the GSMP correspond to Andersen et al. (2002) who found slightly enhanced Cd concentration in arable topsoils compared with nearby afforested sites. As possible explanation accumulation induced by fertilizer application on arable soils (Nziguheba & Smolders 2008, Rochayati et al. 2011) and cadmium leaching because of low pH in the studied forest soils are mentioned. This is in accordance to e.g. Huang et al. 2011 describing Cadmium leaching induced by low pH for forest soils.

High concentration in arable soils right of the dashed line is presumably due to fertilizers, especially to phosphate fertilizers (Boysen 1992, Loganathan et al. 2003), which can increase the Cd content in soils (Knappe et al. 2008, Kelepertzis 2014). Cadmium contents in fertilizers differ widely (Nziguheba 2008) as well as the applied amounts of fertilizers. This might explain why Cd concentrations are not equally enhanced in all arable soils (Schilli et al. 2011).

In the SOM-SM area around X=2, Y=0 several samples with high concentrations of Cd were found (Figure 6B). Most of the samples were taken from different sites in the region of the Harz Mountain and its foreland which is known to exhibit enhanced geogenic contents of Cd from different ores (Bartels 1996). However, not all of the samples in the region of the Harz Mountains show equally high Cd concentrations, presumably due to various local effects (Schilli et al. 2011).

In the organic cluster more than 90% of the samples of the O horizons belong to the fourth and fifth quintile. However, samples from pristine bogs and fens often show low Cd concentrations (Figure 6B; cf. Figure 8). More than 30% of the samples belong to the second quintile. Like for Cr and Ni this can be mainly ascribed to the lower exchange capacity of bogs and fens (Schilli et al. 2011).

The missing accordance of the pattern of Cd concentrations with the pattern of other heavy metal Cd concentrations indicates that the Cd concentration is driven by different impacts or processes. Fabietti et al. (2010) explain the missing correlation of Cd concentration of Italian soils with other variables, such as land use, with a generally low degree of contamination. Neither land use, texture, deposition nor parent material seemed to be main drivers for the spatial differences of the Cd content in the soils. (Schilli et al. 2011). Thus the main driver for high Cd concentration remains unclear.

Main drivers

As main driver for TOC, N and pH the land use can be identified (Figure 5, Figure 6). In forest soils the distribution of organic matter is caused by natural processes. High concentration of TOC can be found typically in the forest floor, where only small amounts of mineral compounds can be found. The TOC accumulation in forest soils is mostly restricted to a few centimeters close to the surface. Exceptions are acidified soils e.g. Podzols, where TOC is transported to the sub soil and accumulate in deeper horzions (Sauer et al. 2007). The exceptions are bogs and fens, revealing high TOC and N values because of the hydrological site characteristics.

The results indicate two main drivers of soil chemistry when regarding concentrations of Cd, Cu, Cr, Pb, Ni and Zn in aqua regia extract. The impact of deposition can be visualized focusing on Pb compared with mostly geogenic Cr in the subregion of the SOM-SM between the solid and dashed line. Increased lead content in humus layers and top soil horizons of natural grown soils generally indicates anthropogenic impacts, especially deposition (Filipinski & Gruppe 1990), even if the input of Pb has decreased (Kaste et al. 2003). Comparing the concentration gradient for Cr and Pb for the mineral top soils at forest sites different directions of the gradient can be recognized.

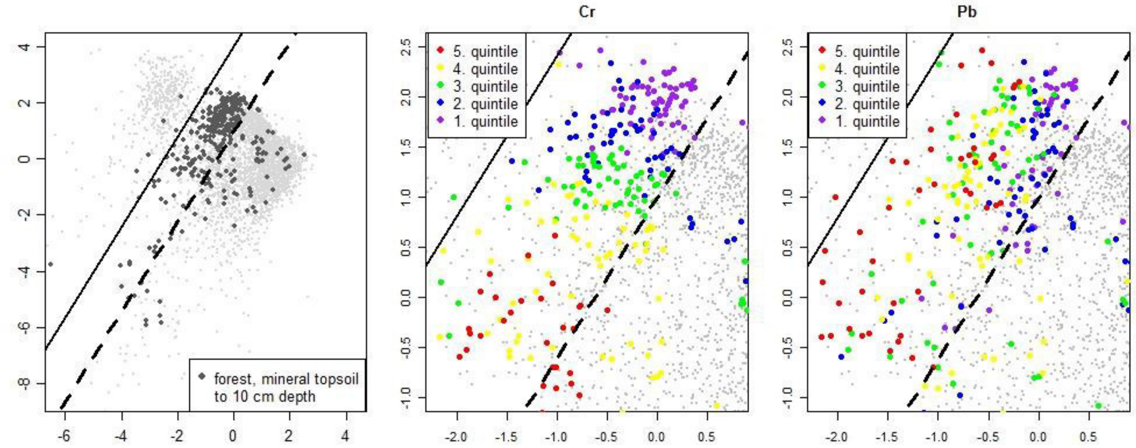

Figure 10: Comparison of Cr and Pb content for mineral top soils at forest sites

The Cr gradient follows the typical gradient from upper right to lower left reflecting the geogenic contents of the soil (Figure 10). For Pb the gradient changes the direction from more or less right to left. Contents of mineral top soil samples of forest sites from the first and second quintile can be found in the area of X-axis from -0.3 to 0.5. In the area of -1.0 to -0.3 contents of the third and fourth quintile are plotted, while high contents of the fifth quantile can be found with X-axis <1.0. The contents increase in the direction of the organic horizons (small cluster), which are mainly effected by deposition (Schilli et al. 2011), which is in accordance to higher efficiency of forests in air pollution regulation compared to other land use classes (Meesenburg et al. 2016). At forest sites the Pb from deposition is stored in the humus layer and the uppermost, TOC-rich mineral soil horizon (Huang et al. 2008). The latter often is only a few cm depth with low density.

The deposition effect is not pronounced for Pb for other mineral soil samples, like top soils of agricultural soils or sub soils at forest sites. The thickness of top soil horizons of grassland or arable land is generally quoted with about 10cm (grassland) to 25-30cm (arable land) in the underlying dataset. Besides the thickness agricultural top soils also reveal higher soil densities. The deposition input is distributed to a larger amount of soil and therefore diluted. Here, the geogenic impact can cover the deposition effect (Schilli et al. 2011).

The deposition impact is pronounced in organic horizons located at the (upper) left of the different figures, at the left of the solid line. At the right site of the dashed line the impact of parent material dominates the contents of heavy metals (Figure

6 cf. Figure 7). In general it can be stated, that the influence of deposition on the soil sample chemistry increases, the more the sample is located in the direction of the (upper) left of the figure. In the direction of the lower left the general level of heavy metal concentration increases (Schilli et al. 2011).

Multivariate Trend analysis

The multivariate trend analysis revealed no significant trend as indicated by the overlapping confidence intervals (Figure 9B). However, it gave some evidence for systematic changes of some parameters in the mineral topsoil horizons. The indicated trends of arable land and forest sites demonstrate an increase of dissimilarity of the chemical characteristics. The distance between the calculated coordinate mean of the compared data points increases (Figure 9A).

Arable soils indicate a shift towards the upper right in the SOM-SM, that is, towards lower heavy metal contents (except Cd). There may be several reasons for that. Lower heavy metal contents in atmospheric deposition (Huang et al. 2011) and fertilizers (Nziguheba & Smolders 2008), e.g. due to regulatory requirements, lead to lower input in soils. Increasing biomass production and continuous plant uptake can enhance the removal of heavy metals from arable soils by harvesting the plants. Additionally an increase of plowing depth might contribute to this phenomenon through mixing the topsoil with the subsoil material with lower heavy metal contents as described for contaminants in arable soils of the GSMP of Lower-Saxony (Höper & Meesenburg 2012).

For grassland sites the calculated confidence interval is relatively high. On the one hand similar site treatments as for arable land could be expected and thus, a similar trend could be assumed. On the other hand different grasslands sites had been used as arable land earlier as revealed by an old plow horizon. In grassland soils without plowing activity anthropogenic heavy metals and organic carbon accumulate mainly in the upper 10 cm of the soils. Thus, care should be taken not to misinterpret land use change induced changes as a trend of single parameters (Schilli et al. 2011). To prevent such misinterpretation land use and possible varying plow depth should be considered when interpreting results for single sites (Fortmann et al. 2012). Therefore a careful documentation of the site handling of agricultural soils is necessary in a long-term monitoring.

The analysis for the top mineral soils of the forest sites pointed to increasing heavy metal contents. Those results seem to be inconsistent with the decreasing deposition. Some studies described the forest floor as a current source for different heavy metals which were accumulated in former decades (e.g. Huang et al. 2008). The mineral soil might still act as a sink for heavy metals which are mobilized from the forest floor (Huang et al. 2011). This might increase the contents of heavy metals in the mineral soil in spite of decreasing deposition as described for single GSMP forest sites in the federal state Baden-Württemberg (LUBW 2008).

The size of the confidence intervals decreases with increasing number of samples. Thus, assuming that the mean location of the samples in the SOM-SM would not change when additional samples were considered, these shifts would prove to be significant. Thus, it is recommend to increase the number of sites with replications in the future (Schilli et al. 2011).

4.1.5 Conclusions

Soils are a basic resource for agricultural and forest products and provide a vast range of ecosystem services. Long-term soil quality monitoring is a prerequisite for assessing the current status, revealing long-term shifts, and identifying needs for soil protection actions. Differing from most scientific studies analysis of monitoring data by the respective authorities urgently needs tools for efficient screening of the large and multivariate datasets in order to identify and evaluate local peculiarities, to identify the onset of trends, or to yield some hints on prevailing processes without requiring any pre-defined hypotheses. Therefore, the SOM-SM approach was used and it can be stated that this is a versatile tool for visualization and first analysis of a large heterogeneous dataset like the one from the German national long-term monitoring program. General patterns and additional spatial peculiarities were identified. Dissimilarities between different elements could be easily detected. The most pronounced differentiation in the dataset was between samples from organic and mineral horizons that exhibited substantially different behavior with respect to the observed parameters. The chemism of organic horizons seemed to be mainly influenced by decomposition processes and deposition. Within the organic samples, clear differences between decomposed O horizons and less decomposed material of litter, bog and fen horizons were identified. The characteristics of mineral horizons were dominated by the parent material. For the top mineral soil horizons of forest sites the deposition seemed to be an important factor determining the contents of heavy metals.

Land use effects on heavy metal contents and TOC content were shown for surface near mineral horizons, particularly for forest sites. The cultivation of fens obviously led to a shift of chemical characteristics in direction of upper mineral soil samples of afforested sites regarding the studied nine parameters. There was no clear relationship between pH and heavy metal content in mineral soils.

Pairwise comparison of the first and last samples from sites that had been monitored for at least ten years did not yield significant differences. However, there were evidences for systematic shifts for different land use classes that deserve more attention.

The presented SOM-SM is highly recommended as a very efficient tool for a first inspection of large datasets in order to structure subsequent steps of the analysis without requiring pre-defined hypotheses. The SOM-SM makes use of the most efficient interface between a data base and the human's brain. The SOM-SM generates a single graph, that is, a low-dimensional projection of the dataset. The study shows that this graph can be used for a quick scan of the dataset to answer very different questions and giving clear hints on even unexpected features. Thus, it allows authorities and science to "play" with large datasets in very efficient ways in order to allow a more efficient use of large soil monitoring datasets.

4.2 Bedrock effects on forest soils: Commonalities and differences

4.2.1 Introduction

The soil is a complex compartment of the environment and can be influenced by the interplay and impact of atmosphere, lithosphere, hydrosphere and human activity. In spite of the complexity a worldwide valid classification for soils could be developed (FAO 2014), which is possible, amongst other reasons, because similar processes prevail in soils all over the world (Bockheim & Gennadiyev 2000). On the other hand different processes in soils can vary at smaller scales, e.g. in a landscape or within single soil profiles (Wang et al. 2013), affecting biological, physical and chemical soil characteristics.

Due to a large variety of factors and processes known to impact soil chemical properties differences between various sites usually can equally well ascribed to very different causes. This renders soil resources management very difficult and more an art rather than a science. However, as soil science now approaches the era of big data, both quality and quantity of available data as well as available methods allow addressing these issues. Thus this study aims at delineating various candidate processes with respect to their impact on soil solution quality. The dataset comprised soil solution data from three forested sites with contrasting bedrock lithology and from different depths. More than 10.000 soil solution samples, comprising a time span of 15 years, were analyzed. The size of the dataset, combined with powerful statistical approaches might enable to differentiate between commonalities and differences in spite of substantially differing sites conditions and a heterogeneous dataset, far from any orthogonal design of classical experiments. Thus the study was meant to check and to demonstrate the potential of big data approaches being applied to monitoring datasets. The performance of a linear and of a nonlinear statistical approach was compared. Using those approaches the relevant processes in the soils should be identified, quantified and distinguished in overall and site specific impacts and processes.

4.2.2 Material and Methods

4.2.2.1 Statistical methods

According to the aims in a second step the identification and a quantification of processes in soils and impacts on soil chemistry is focused. For this purpose data from a medium and small scale were chosen. In chapter 4.2 the results from analyzing soil solution of three soil profiles with distances of up to 100 km as medium scale and in chapter 4.3 the results from four soil profiles within a catchment of 4,2 km^2 as small scale are presented.

The chosen statistical approach should be able to identify the main impacts and processes on the soil characteristics. Because most processes and impacts are influenced by or influence more than one parameter multivariate approaches should be used.

The identified factors should be separated from each other and the relevance quantified. Those factors should be clearly delineated from other factors to be interpreted more easily and to allow further statistical analyses without the problem of multicollinearity. Therefore they should be stochastically independent. To reach this goal principal component analyses or factor analyses can be used and are popular (Bahrenberg et al. 2003).

In this thesis the PCA was chosen. As a first step explorative explanation tool the PCA is more suitable than the factor analyses, which is a model based approach. To regard probable hidden nonlinearity in the datasets (e.g. Manzoni et al. 2004, Biswas et al. 2013) the nonlinear Isometric Feature Mapping (Isomap) approach, which results can be interpreted similar to the principal component analysis, is used here. The similarity of the approaches minimizes a possible effect of varying algorithms influencing the results when comparing a linear and nonlinear approach. The often stated (sligth) superiority of Isomap using environmental datasets (e.g. Tarquis et al. 2011, Lischeid 2014, Maassen et al. 2015) was tested for soil science monitoring dataset in this way. Possibly the use of the more complicated and time-consuming nonlinear analyze is needless. For instance, the comparison of both approaches using a medical dataset showed the Isomap to be only slightly inferior to the PCA (Dai et al. 2015). The results will be compared to assess the use of nonlinear approaches and to give recommendations for further analyses.

In comprehensive soil monitoring often a large number of variables are measured. To analyze the relation and interdependencies of all this variables a consideration of the different variables is necessary. In single graphs only two or three variables can be presented and interpreted. Therefore a large number of figures would be necessary to regard all possible cases. Even then it is not ensured, that a display of up to three variables is enough to understand complex interdependencies.

To deal with this problem it is useful to reduce the dimensionality of the dataset. The idea of dimension reduction is to identify few representative characteristics in the underlying dataset explaining a large part of the variance of all measured variables which can be projected in graphs. The axes of the graphs can reflect the most relevant characteristics, which represent a large fraction of the explained variance of the dataset (Legendre & Legendre 2012).

For this purpose the PCA has been frequently used in different ecological studies (Selle et al. 2013, Ansari et al. 2014) as well as in soil science (Jia et. al. 2010, Kelepertzis 2014).

The PCA extracts so called principal components by conducting an eigenvalue decomposition of the covariance matrix of the measured variables. These components are able to describe circumstances, interactions, or processes which cannot be measured directly (Selle et al. 2013). To regard possible nonlinear relationships in environmental datasets (James & McCulloch 1990) the Isomap is used additionally. The Isomap approach has been developed and presented by Tenenbaum et al. (2000).

Both approaches are based on a matrix X_{nm} with n observations of m variables, which is the basis for the distance matrix $D^{(x)}$ of Euclidean distances $d_x(i,j)$ between all datapairs i,j in the space X *which* are calculated as

$$\sqrt{\sum_{n=1}^{m}(x_{in}-x_{jn})^2}$$

Both approaches are sensitive to missing values (Legendre & Legendre 2012), which is why missing values should be replenished or the whole data set cannot be regarded. To assign equal weight to possible different scaled variables the data should be z-normalized (mean = 0, standard deviation = 1) before calculating the distance matrix (Webster 2001).

Principal component analysis (PCA)

The PCA tries to project a high-dimensional dataset in a low-dimensional embedding of data points. The given amount variance in the high-dimensional data should be reflected in the low-dimensional embedding as large as possible. For this reason the PCA extracts so called principal components by conducting an eigenvalue decomposition of the covariance matrix of the measured variables. Those components are able to describe circumstances, interactions, or processes which cannot be measured directly (Selle et al. 2013).

In general the PCA is based on a covariance matrix and its eigenvalue decomposition (Legendre & Legendre 2012). In this thesis the PCA and the Isomap approach (see next section) will be compared. Therefore, the eigenvectors and eigenvalues (based on distance matrix of Euclidean distances $D^{(x)}$) are calculated as

$$\Lambda = E^T * D * E$$

The eigenvector matrix of E^T is E_{mm}. The eigenvalue matrix D is diagonal. The first principal component is the eigenvector with the highest associated eigenvalue and so on. Since E is m dimensional m principal components can be calculated (Jolliffe 2010).

It is assumed that different components can be used as quantitative measures of the effect of different processes and impacts. Ascribing components to those is based on loadings of the variables on the components, that is, their Pearson correlation coefficients with the respective component.

Isometric Feature Mapping (Isomap)

The main difference between the Isomap approach and the classical MDS and PCA, is the use of the geodesic distances when calculating interpoint distances. The Isomap tries to grasp the nonlinear structures by a stepwise linear approximation. For this purpose the Isometric Feature Mapping passes three steps (Tenenbaum et al. 2000), which are described in more detail in Appendix III.

Step 1: Set up a neighborhood graph using a chosen number of nearest neighbors k based on the distance matrix.

Step 2: Set up a distance matrix $D^{(G)}$ using geodesic distances calculated from the neighborhood graph. The geodesic distance is approximated by finding the shortest path between the data points. If k is equal to n each data point would be connected with each other and shortest path would be equal to the Euclidean distance. The calculation of the useful geodesic distance depends on the choice of the regarded nearest neighbors k. The optimal k can only be determined by trial and error.

Step 3: Based on the distance matrix $D^{(G)}$ a singular value decomposition is performed.

Following these steps a piecewise linear regression is performed in a high dimensional data space. The requirement of any pre-defined mathematical structure is not given. Accordingly high nonlinear structures can be recognized.

While the performance of the PCA can be evaluated by eigenvalues, the nonlinear approach of Isomap cannot use eigenvalues as quality criterion of the performance. For Isomap the squared Pearson correlation of $D^{(x)}$ to the distance matrix of the low-dimensional projection, consisting of the component scores, is used, to estimate the explained cumulative variances (Schilli et al. 2011). However, as relationships might be nonlinear, the rank-based Spearman ρ correlation coefficient will be used instead of the linear Pearson correlation coefficient. In addition, boxplots will support component interpretations.

Interpreting components

The Isomap and PCA components can be interpreted analogously. The scores of the different components can be understood as measures of the effect size of single processes. The components of PCA and Isomap are numbered in decreasing order of explained variance of the data set. Therefore, components with higher number mostly show lower loadings. For this reason, the detection of nonlinear relationships might be difficult. To that end, for all components (with exception of the first component) residuals of linear regression of the variables with all preceding components were used and related to the appropriate component. The linear regression was used for PCA and Isomap, because it is much more stable than various nonlinear regression approaches that had been tested. The resulting error is presumably negligible when using a nonlinear rank-based measure of correlations (when conducting the Isomap) and with nonlinear relationships close to monotonous relationships.

Boxplots will support the interpretation of the components. The boxes show the 25th and the 75th percentile, the black bar the median of the analyzed data. Whiskers denote the range as long as they do not differ more than 1.5 times from the interquartile range from the box. In that case the 1.5 times interquartile range will be displayed, and values outside the whiskers are displayed as single symbols.

4.2.2.2 Study sites

This study comprises data from three different sites located in the federal states of Lower Saxony (Lange Bramke and Solling site) and Hesse (Zierenberg site) in Germany (Figure 11). All sites are afforested and incorporated in the Intensive Forest Monitoring Program of the International Co-operative Program on Assessment and Monitoring of Air Pollution Effects on Forests (ICP Forests). The Lange Bramke site comprises three subplots at different hillslope positions, while the Zierenberg and Solling sites had only one subplot each. A summary of the main site characteristics is given in Table 5. In Table 6 Depth and site specific means of soil solution variables are given. More detailed information for Zierenberg and Solling can be found in Meesenburg & Brumme (2009) and Meesenburg et al. (2009a) and for Lange Bramke in Meesenburg et al. (2010).

Table 5: Site characteristics

	Zierenberg	Solling	Lange Bramke		
			North slope	South slope	Ridge
Location	51°22'N 09°16'E	51°45'N 09°34'E	51°51'N 10°25'E		
Elevation [m]	450	500	600	600	660
Exposure	north-east (15°)	south (0-3°)	North (5-10°)	South (10-15°)	south (0-5°)
Precipitation [mm]	754	1193	1339		
Mean temperature [°C]	6.9	6.9	5.9		
Bedrock	Basalt over limestone	Sandstone with loess cover	Quarzitic sandstone		
Texture	Clayey Silt	Loamy Silt	Loamy sandy Silt		
Soil type WRB	Hablic Cambisol	Dystric Cambisol	Dystric Cambisol		Entic Podzol
Humus type	mull	moder	mor		
Main tree species	*Fagus sylvatica*	*Fagus sylvatica*	*Picea abies*		
Stand age [years in 2013]	158	166	65		
Sampling depths [cm]	20, 60, 100	0^{*1}, 10, 20, 40, 80, 90	80, 300	0^{*1}, 80, 300	20, 80, 300

*1 transition zone of humus layer and mineral soil

Table 6: Depth and site specific means of soil solution variables (LB= Lange Bramke, b.d.l.=below detection limit)

site		number	depth	pH	Na	K	Mg	Ca	Al	Fe	Mn	Cl	NH_4-N	NO_3-N	SO_4-S	DOC
			cm	-	mg/l											
Zierenberg		2397	20	6.4	3.16	0.28	5.26	8.27	0.27	0.09	b.d.l.	4.20	b.d.l.	5.27	3.60	16.25
		1804	60	6.9	3.71	0.15	5.48	9.57	0.02	b.d.l.	b.d.l.	5.72	b.d.l.	4.46	4.20	6.08
		652	100	7.3	4.66	0.20	6.26	15.36	b.d.l.	b.d.l.	b.d.l.	5.97	b.d.l.	3.83	4.94	5.50
Solling		615	0	3.8	1.79	1.11	0.32	0.82	0.72	0.42	0.29	2.99	0.13	0.96	1.19	37.61
		387	10	4.1	1.73	0.33	0.26	0.40	1.65	0.31	0.32	3.11	0.09	0.53	1.47	16.23
		422	20	4.4	1.88	0.08	0.19	0.25	1.68	b.d.l.	0.29	3.51	b.d.l.	0.08	1.93	6.83
		461	40	4.5	1.89	0.14	0.25	0.36	1.63	b.d.l.	0.27	3.42	b.d.l.	0.64	2.38	2.33
		282	80[*1]	4.4	2.32	0.17	0.14	0.17	1.87	b.d.l.	0.20	4.38	b.d.l.	0.21	3.45	1.83
		279	90	4.5	1.55	0.36	0.21	0.20	1.05	b.d.l.	0.19	2.66	b.d.l.	0.25	2.17	1.27
LB	North	1082	80	4.4	1.16	0.31	0.15	0.29	2.09	b.d.l.	0.18	1.77	b.d.l.	1.63	2.09	2.35
		251	300	4.4	1.65	0.59	0.26	0.74	2.24	b.d.l.	0.57	2.65	b.d.l.	2.19	3.84	1.62
	South	64	0	4.2	0.80	0.83	0.10	0.61	0.08	b.d.l.	0.14	1.36	0.52	0.67	1.08	16.85
		957	80	4.3	1.93	0.26	0.25	0.31	2.70	b.d.l.	0.37	3.17	b.d.l.	b.d.l.	6.15	2.50
		224	300	4.3	2.25	1.07	1.59	3.57	2.50	b.d.l.	0.65	3.78	b.d.l.	b.d.l.	9.99	1.56
	Ridge	486	20	3.9	2.15	0.43	0.23	0.77	0.87	0.59	0.19	2.96	b.d.l.	b.d.l.	2.11	27.10
		497	80	4.2	2.39	0.16	0.19	0.26	1.51	b.d.l.	0.21	3.54	b.d.l.	b.d.l.	3.42	6.85
		155	300	4.3	1.70	0.37	0.22	0.55	1.84	b.d.l.	0.31	2.56	b.d.l.	b.d.l.	4.78	1.82

[*1] not measured within the whole study period

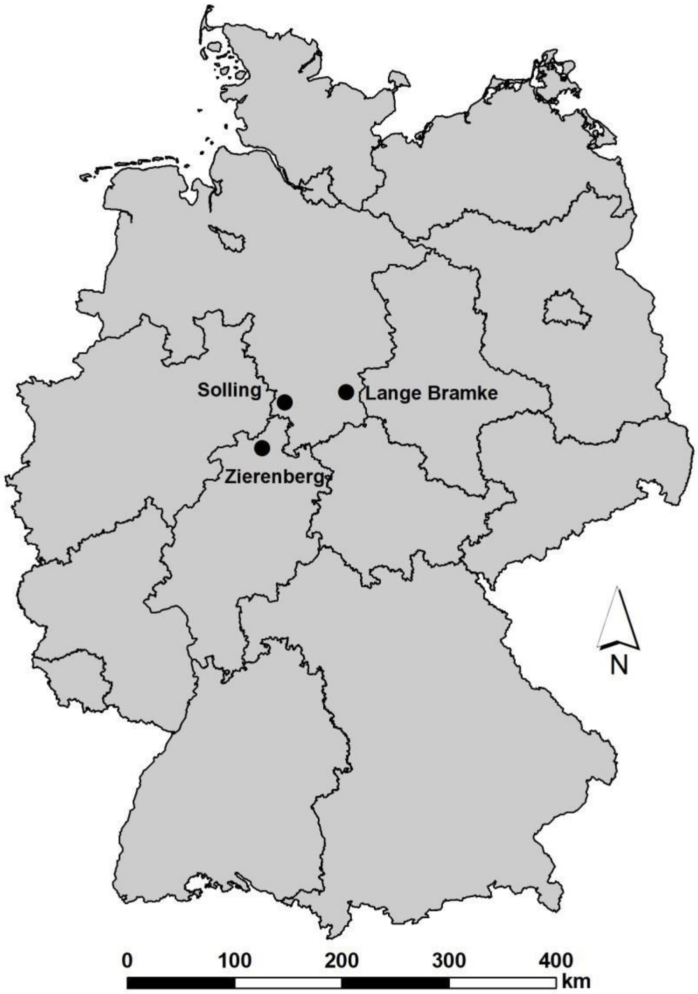

Figure 11: Location of the study sites in the Federal Republic of Germany

4.2.2.3 Dataset

For this study data from 30.000 soil solution samples from the three sites were available from the period 1993-2008. Soil solution was collected with ceramic suction cups or ceramic plates. The cups were sampled at weekly to biweekly intervals and pooled to monthly samples at Solling and Lange Bramke and at biweekly intervals, (pooled to monthly samples since 2006) at Zierenberg. In the depth of 0 cm (transition zone of humus layer and mineral soil) at Lange Bramke (south) and Solling suction plates were installed with identical measuring intervals and sample treatment. The plates at Solling were operated with continuous low suction, while the Lange Bramke (south) plate used no tension (Table 7). For further information and detailed description of soil solution sampling see Nieminen et al. (2013) and Klinck et al. (2012) for Solling and Lange Bramke and Brumme et al. (2009) for Zierenberg.

Table 7: Sampling methods at selected sites and depths

Study site	Depth [cm]	Method
LB-S	0	Zero tension sampling plates
Solling	0	Low tension sampling plates
Solling	10	Suction cups
LB-R	20	Suction cups

Values below detection limit were replaced by half the detection limit. Only variables with long and continuous observation periods were considered. Because of the necessity of complete data sets for the intended statistical analysis, the period was limited from December 1993 to September 2008 and the number of variables was reduced from more than 25 variables to a set of 14 variables. Data sets with more than two missing values were excluded. The data cleaning left 11.015 data sets for statistical analyses. Total nitrogen was excluded afterwards due to strong correlations to NO_3 (Spearman ρ 0.94), leaving 13 variables available for the subsequent statistical analyses:

Al, Ca, Cl, Fe, K, Mg, Mn, Na, NH_4, NO_3, pH, SO_4 and DOC.

Replenishing 660 missing values with the variable site and depth specific mean (pH mean was calculated with delogarithmized pH values), a complete data set was generated. Only in the case of Mn (1.80 %), NO_3 (1.03 %) and TOC (1.16 %) more than 1 % of the data were replenished. Additionally the data were z-normalized by subtracting the variable specific mean from each value and subsequent division by the standard deviation.

For statistical analyses and generating diagrams the software package R (Version 2.10.0) and Vegan extension (Version 1.15-4) for Isometric Feature Mapping available at *http:/www.r-project.org* (R Development Core Team. 2006) was used.

4.2.3 Results and discussion

The calculation of the geodesic distances for Isomap has been performed using different *k* values (1000, 3000, 5000, 7000, 7500, 8000, 8500 and 9000). The best results were achieved for *k*=8000, which were superior to the results of PCA with respect to the explained variance (Figure 12). Further figures and analyses are based on the results of Isomap with *k*=8000. A superiority of Isomap in comparison to PCA could also be approved in other studies (Tenenbaum et al. 2000, Lischeid & Bittersohl 2008, Lischeid 2014). The superiority of the nonlinear approach accounts for relationships which are not detected by the linear approach. On the other hand the superiority is not guaranteed, as presented by Dai et al. 2015.

More than 95 % of the variance of the dataset can be explained by the first four components, and more than half (54.4 %) by the first Isomap component (Figure 12). Considerable negative correlations (Spearman $\rho < -0.5$) for this component were found for Al, Mn and NH_4 (Figure 13). Comparable positive correlations were calculated for pH, Na, Mg, Ca, Cl and NO_3. Positive correlations for Fe and DOC in contrast to negative loadings of Al and SO_4 were calculated for the second component explaining 31.5 % of the dataset variance. The third component shows positive correlations throughout, except of pH and accounts for 7.5 % of the dataset variance. Additional 2.2 % of the variance of the dataset can be ascribed to the fourth component. This component is characterized by negative loadings of DOC and Fe and positive loadings of NH_4 and K especially.

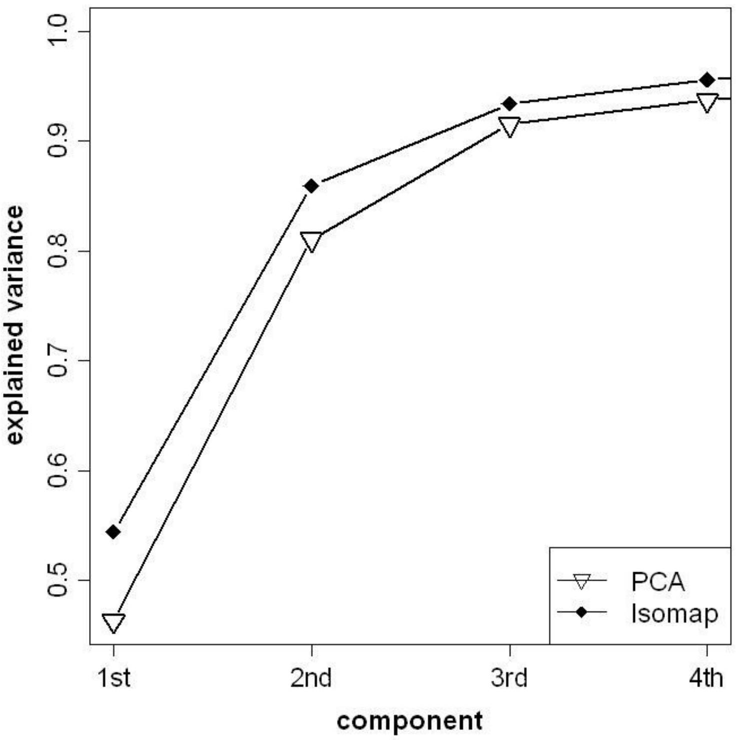

Figure 12: Explained variance by PCA and Isomap

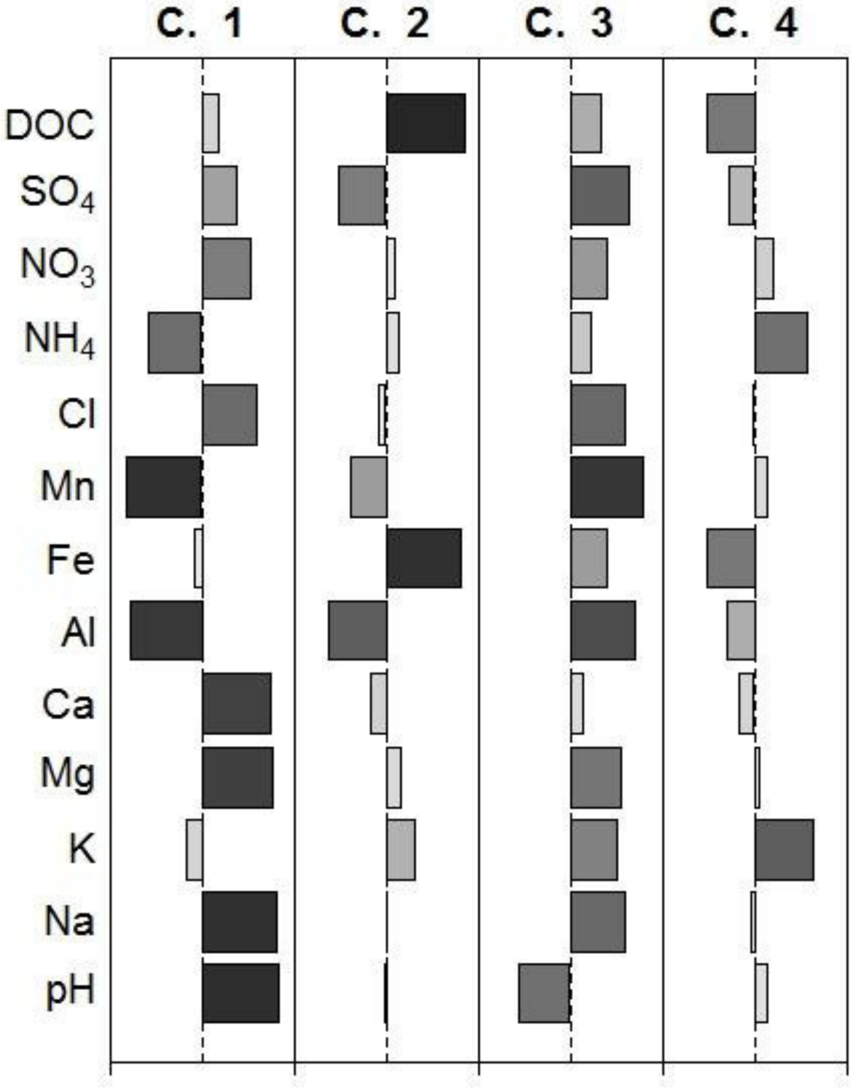

Figure 13: Parameter loadings determined as Spearman ρ for residuals of regression of single variables with the scores of the preceding Isomap dimensions. The X-axis is scaled between [−1;1] for every component. Grey shading is proportional to the absolute values of the Spearman ρ coefficient.

4.2.3.1 Component 1: Bedrock

The first component indicates the chemical peculiarities of the more acidic sites developed on sandstone (Solling and Lange Bramke) and the less acidic site on basalt over limestone (Zierenberg). Negative component values are mostly calculated at Lange Bramke and Solling, while positive component values exclusively occur at Zierenberg (Figure 14 - left). It is the only component out of the first four components studied in more detail that clearly differentiates between the sites.

In comparison to acidic sites, the soil on substrate rich in bases mostly shows a higher pH and higher base cation availability, which is indicated by high positive loadings of Ca, Mg, Na and pH (cf. Figure 13, Figure 14). In contrast, release of Al and Mn at the more acidic sites is reflected by negative loadings of Al and Mn. The buffering of acids by clay minerals and Al-(hydr-)oxides (pH < 5) leads to increased Al contents in soil solution, while the dissolution of Mn-(hydr-)oxides releases Mn (Ulrich 1986). With sustained dissolution of Al- and Mn-(hydr-)oxides, increasing Al and Mn concentrations in soil solution and resorption at the soil matrix the concentration of base cations decreases in soil solution. Besides Al and Mn, a negative loading was calculated for NH_4. Ammonium can be oxidized to NO_3 by microbial activity, removed from soil solution by root uptake, fixated in interlayers of clay minerals or immobilized in organic layers close to the surface (Brumme et al. 2009, Nieder et al. 2011). The differences in sampled depths may explain the negative loading of NH_4. Surface near soil solution samples (< 20 cm), where ammonium can be detected most frequently, are missing at Zierenberg (see Table 6).

At the Solling and the Zierenberg site a depth gradient can be recognized. The component values tend to increase from the surface to deeper horizons (Figure 14 - right). At the Solling site to a maximum of 80 cm depth. Lange Bramke reveals a depth gradient at the south slope only, which is poorly developed (not shown). Soil formation causes depth gradients, affecting the upper horizons to a larger degree. Such depth gradients are consistent with previous findings (Marschner et al. 1998, Meesenburg et al. 2016). Obviously, soil formation and atmospheric input lead to a development of soil solution composition ranging from the state in the subsoil at Zierenberg towards the characteristics of the upper horizons at the Solling site. Although the soil formation processes play an important role, the influence of the bedrock dominates and distinguished the sites developed on sandstone from the Zierenberg with basalt over limestone.

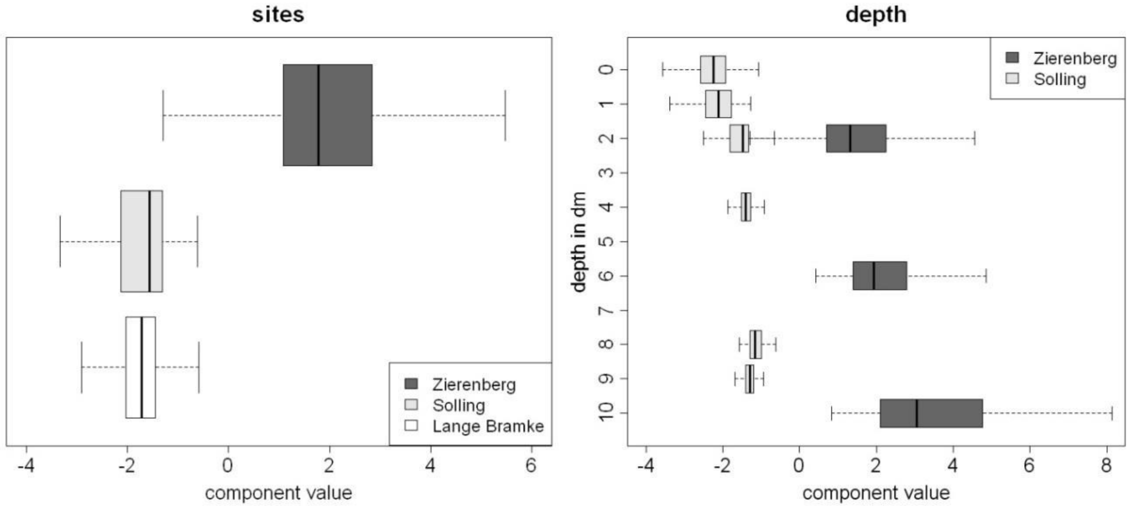

Figure 14: Scores of the first component at different sites and site-specific depths

4.2.3.2 Component 2: Acid induced soil formation effects

The second component relates to different soil formation effects associated with the buffering of acidic inputs. In contrast to the first component, the second component - explaining 31.5 % of the variance - does not reveal distinct site differences. The second component is dominated by depth gradients. These gradients are similar at each site. Positive component values occur in the upper soil horizons (Figure 15). The component reveals positive loadings of DOC and Fe and negative loadings of Al, SO_4 and Mn.

Negative component values reflect reactions in the exchanger buffer range, typical for pH 4.2 - 5.0. The buffering of acidic inputs by Mn-oxides and clay minerals releases Al and Mn to soil solution (Ulrich 1986).

The buffering of acids by iron minerals, e.g. $Fe(OH)_3$, associated with the formation of organic Fe complexes, is indicated by positive values. The latter is typical for mineral soils with pH values < 3.8 in the aluminium /iron buffer range (Ulrich 1986). Single pH measurements down to < 3,7 can be found at Lange Bramke till 80cm and Solling till 10cm depth (Table 6). Lowest pH median (3.9) and highest average component values can be found in 20 cm depth at Lange Bramke (ridge) supporting the interpretation.

Suction cups and plates revealing positive component values are installed in upper soil horizons with elevated contents of humic substances, like the forest floor or A horizon, respectively E horizon. Acidic atmospheric input is often linked with podzolization processes and the formation of iron-humate complexes in the upper soil. The latter are translocated in deeper soil horizons. (Riise et al. 2000). At Zierenberg a median of pH 6.4 could be measured in 20 cm depth which is untypical for podzolization processes and is unexpected with regard to the positive component values found (Figure 15). On the other hand the formation and dislocation of iron-humate complexes and development of podzolic horizons could be shown at different sites with pH > 7 (Protz et al. 1984). The necessary organic acids may originate from plants or litter decomposition. Additionally, the exudation of organic acids by roots, fungi and microorganisms is possible (Lundström et al. 2000).

Typically, podzolization processes also lead to a formation of Al-organic complexes, similar to Fe, which is contrary to the negative loading of Al in these results. For the Solling site the decoupling of Fe and Al dynamic because of the formation of Al-(Hydroxy-)Sulfates is described in the exchanger buffer range (Ulrich 1986), which support opposing loadings of Al and SO_4 to Fe. Al-(Hydroxy-)Sulfates are formed by buffering processes induced by sulfide entries at pH 4,2 - 5.0 values. In the range of pH 4.2 and lower, the solution of aluminum sulfates begins an important role in the buffering of acid contents (Ulrich 1986, Brümmer 2010). Accordingly, this process will play only subordinate role at the Solling and Lange Bramke site, where pH values in this range where measured. For this reason the loading of SO_4 may be lower than the ones of Al and Mn.

According with the component characteristic Al concentration is usually lower close to the surface were higher DOC concentrations and detectable Fe concentration were measured (Table 6). Another explanation may be the competition of Fe and Al for binding sites on dissolved organic matter (DOM). At pH values similar to Solling and Lange Bramke, Fe shows a stronger tendency for soluble complexation with DOM than Al (Jansen et al. 2002).

In summary, this component combines numerous processes (e.g. formation and dissolution of minerals and organic substances, vertical translocation of substances), all of which can be related to the buffering of acid inputs.

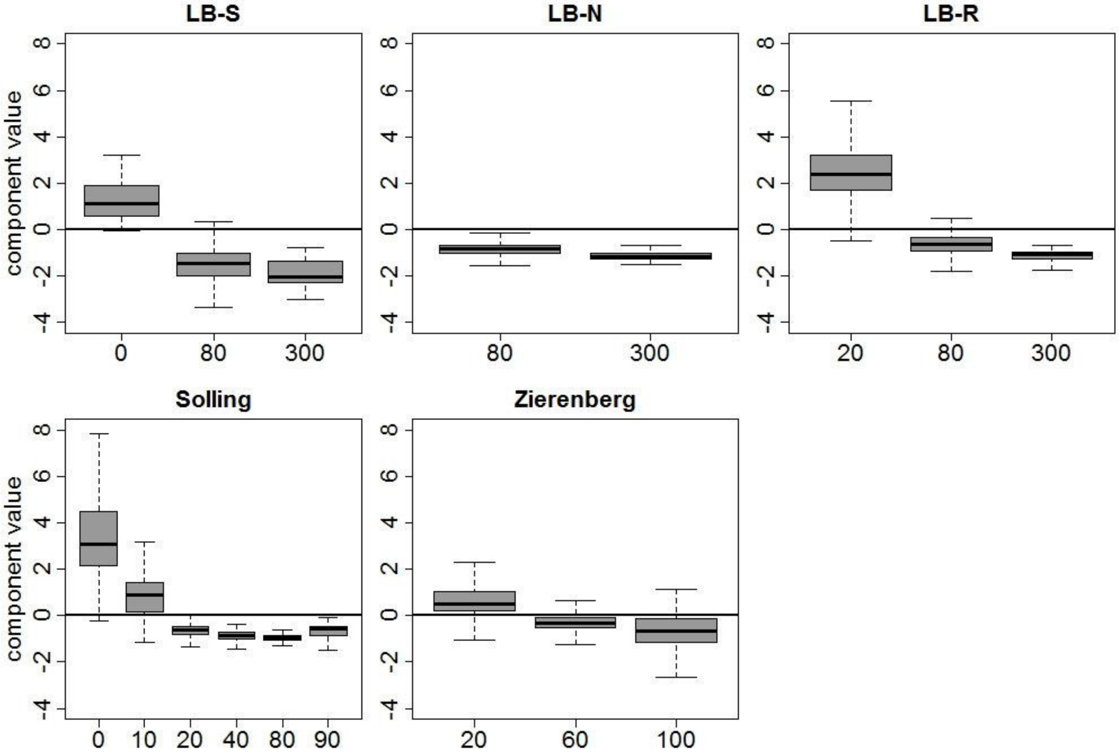

Figure 15: Scores of the second component at site specific depths, LB=Lange Bramke, N=North, S=South, R=Ridge

4.2.3.3 Component 3: Concentration through evapotranspiration

Explaining 7.5 % of the dataset variance, the third component reveals positive loadings for all studied solutes with exception of pH (Figure 13). On the other hand, the negative loading of pH can be regarded as a positive correlation with H^+ ions. Accordingly, this component indicates a process, which influences all parameters in a similar way. Solutes derived from mineral weathering such as Mn or Al reveal highest loadings. Comparably low positive loadings are calculated for Ca, nitrogenous compounds and DOC.

The most pronounced pattern for this component is revealed when regarding the yearly median at depth of 80cm at Lange Bramke, 60cm at Zierenberg and 90cm at Solling site (Figure 16). The depth of 90cm at the Solling site was chosen since the depth of 80cm was not active during the entire study period. Similar trends can be recognized for Lange Bramke Ridge and Solling ($r^2 = 0,9$), Lange Bramke Ridge and Zierenberg and Zierenberg and Solling (both $r^2 = 0,64$). The similarity indicates an overall impact on the soils. It is also striking that the level of the medians of the yearly component values at the Lange Bramke site can be distinguished by the surface morphology slope. The highest values can be found at the south slope, while the lowest ones are calculated for the north slope. The correlations between the yearly medians of the component values and the precipitation do not exceed an r^2 of 0.05. The opposite is the case for yearly temperature medians (Table 8). With r^2 of 0.63 to 0.75 a relation between both parameters is indicated. An exception is the Lange Bramke (north) with a lower r^2 of 0.4.

Table 8: r^2 of yearly means of temperature and yearly medians of the third component differentiated according to study sites

site		r^2
Solling		0.63
Zierenberg		0.75
Lange Bramke	South	0.65
	Ridge	0.72
	North	0.40

Regarding the monthly medians of the component values, the values tend to be higher in summer months and lower in spring for Zierenberg, Solling and Lange Bramke Ridge. This trend is particulary evident on the southern slope of Lange Bramke. At Lange Bramke north it is hardly pronounced (Figure 17). The observed overall seasonal pattern may be attributed to evapotranspiration. Sutmöller et al. (2007) estimated by means of model calculations for the Lange Bramke sites the highest evapotranspiration rates at the south facing slope and the lowest rates at the north facing slope, supporting the results shown in Figure 17 with higher component values at the south and lower ones at Lange Bramke (north).

The relevance of evapotranspiration short-term effects on soil solution composition has been considered in several studies identifying long-term trends. A relative increase of solutes in soil solution because of evapotranspiration is well known. The effect of evapotransipiration and root water uptake on solutes with high loadings as Cl and SO_4 is described (Wesselink et al. 1995, Brumme et al. 2009). Evapotranspiration directly influences the soil moisture. In turn soil moisture is known to effect the ionic strength of soil solution (e.g. van Hess et al. 2000; Rennert & Rinklebe, 2010), which is consistent with generally higher loadings in times of plant growth and lower ones in spring. The annual differences in the component values can be attributed to the temperature fluctuations and the expected higher evapotranspiration respectively lower soil water content.

However, the evapotranspiration and the root uptake are not known to increase concentrations of dissolved substances of mostly geogenic origin such as Al and Mn. Especially manganese as a nutritional element is more likely to be taken by plants (Türtscher et al. 2017), than it should come to a concentration. The same is true for cations K, Ca, Mg and Na, especially in the tree growing season (Nietfeld et al. 2017). The extent to which such a nutrient uptake by water removal can be compensated is not known.

On the other hand an increased water uptake will enhance the residence time of the remaining soil water. The longer residence time will increase the effect of weathering. The release of nutrients and typical geogenic solutes by weathering processes may exceed the plant uptake (Calvaruso et al. 2014) and can explain the higher loadings Mg, Na and K as well as Al and Mn. Furthermore all sites can be assigned to the Exchanger buffer range. Here mainly Al and Mn are released during buffer processes (Ulrich 1986), which supports their high positive component loadings.

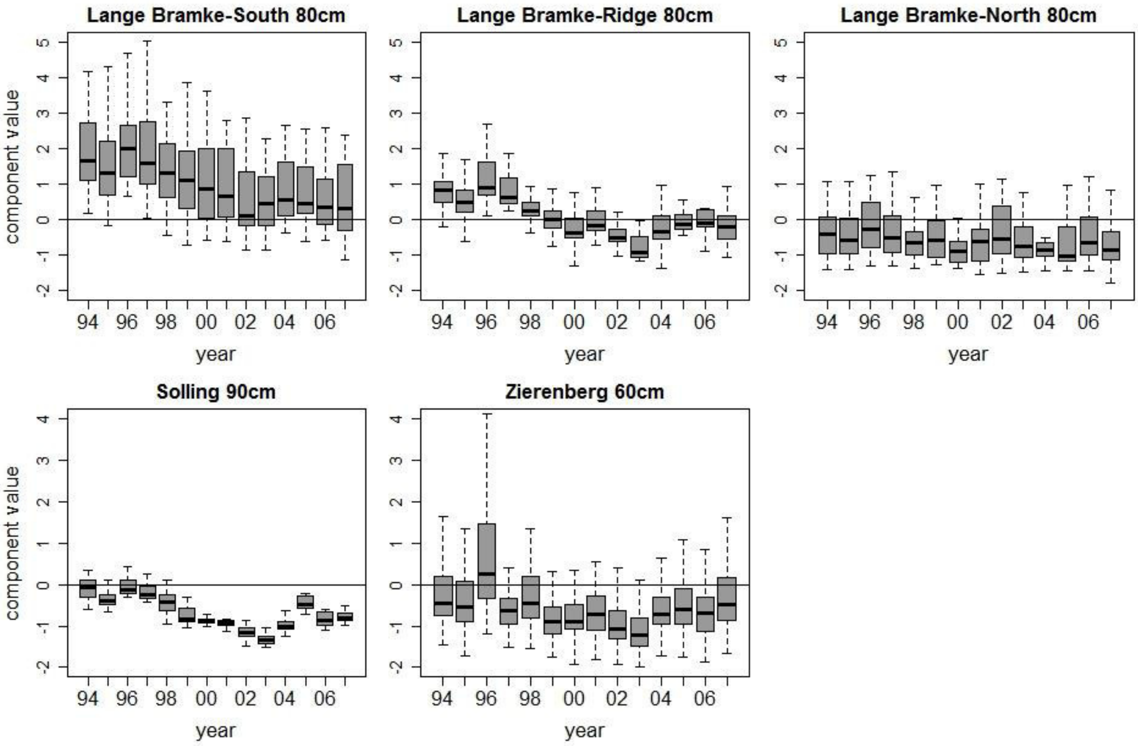

Figure 16: Site specific scores of the third component

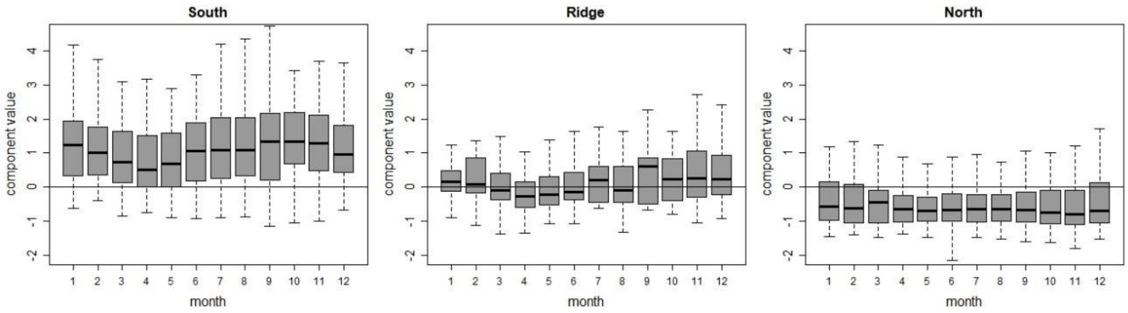

Figure 17: Component scores divided by months for different sites of Lange Bramke

4.2.3.4 Component 4: surface processes

The process indicated by the fourth component is most pronounced in horizons close to the surface. In depths of > 20 cm the component value median is close to 0 within narrow limits (mostly between 1 and 1) at all sites (Figure 18). Therefore, this component is not discussed for Lange Bramke (north) with suction cups in 80 cm and 300 cm. At Lange Bramke (south) clearly positive component values are calculated for the 0 cm depth. Further positive component values exceeding the typical range [-1; 1] can be found at the Solling site in 0 cm only. Values below -1 are found for the depth of 0 cm at the Solling site, the ridge of Lange Bramke in 20 cm depth and sometimes in 10 cm at the Solling site. Accordingly the processes reflected by this component are mainly restricted to the humus layer and the top soil, especially pronounced at the forest floor of Lange Bramke (south).

The decomposition and mineralization of litter is a main source for nutrients in soils. While several other nutrients can be of geogenic origin (Johnson-Maynard et al. 2005) the decomposition is a main source of nitrogen compounds and potassium in soils corresponding to the positive loadings of the component (Brumme et al. 2009). The decomposition is driven by temperature effects leading to higher decomposition rates in the summer and lower ones in winter (Bisht et al. 2014). A resulting seasonality is shown in Figure 19. Induced by the decomposition process Vestin et al. (2008) reported a release of DOC, SO_4, Al and Ca caused by the pedogenetic processes of litter degradation and weathering because of acidic input for top soils. Similar negative loadings of Fe and DOC may reflect the presence of Fe-organic complexes often developed when organic acids are released in the topsoil, which would be typical for the podzolic soils (Lundström et al. 2000) in the study. The observations of Vestin et al. (2008) and Lundström et al. (2000) as well as the quick immobilization of K and NH_4 in interlayers of clay minerals (Kretzschmar 2010) correspond with the negative loadings of the component in the mineral soil close to the surface. Therefore, positive values of this component are most pronounced in the depth of 0cm, while the negative ones are dominating in the mineral soil close to the surface.

Furthermore, deposition (precipitation and throughfall) is described as a major source of K and nitrogen compounds in soils (Meesenburg et al. 1995, Hojjati et al. 2009), corresponding to the positive loadings. Enhanced concentrations of K, nitrogen compounds (especially NH_4) in throughfall compared to soil solution are documented for Zierenberg and Solling (Meesenburg et al. 2009b). For Mn in soils plant cycling can be a relevant process (Türtscher et al. 2017) including litterfall and leaching. At the Solling site in 0 cm depth the highest component value medians are calculated for (early) summer (Figure 19), which is period of intense growth of *Fagus Sylvatica*. In the depth of 0 cm of Lange Bramke (south) a similar seasonality is indicated, but not shown because of the uncertainty because of the low number of data (n=64; cf. Table 6).

During this period the leaching of e.g. K from recently developed leaves or needles can be intensified (Chen et al. 2010). In contrast to the indicated seasonality (Figure 19) Chen et al. (2010) reported a decrease of NH_4 in throughfall by canopy effects during the growing season. On the other hand, the contents of NH_4 in throughfall in comparison to soil solution, especially of the mineral soil, remain high (Ulrich 1986, Meesenburg et al. 2009b). Nevertheless, the detected seasonality indicates a decisive influence of the decomposition. In this component the effects of such pedogenetic processes induced by acidic organic inputs of the decomposition process (negative loadings) are opposed to the initial nutrient input by decomposition and throughfall effects (positive loadings).

The obvious differences of the 0 cm depth at Solling and Lange Bramke (south) can not be explained definitely. Different explanations are possible, always taking into account the small number of measurements on the Lange Bramke site compared to the Solling site. A first one is based on natural processes. The more pronounced characteristic of the component at Lange Bramke may be caused by the tree species. Hojjati et al. (2009) showed, that K and nitrogen leaching is higher at spruce sites compared to beech sites, possibly leading to higher component scores at 0 cm at Lange Bramke (south) compared to Solling. Additionally, the site specific tree species, humus type and microbial activity may regulate soil solution composition (Titeux & Delvaux 2009). However, no significant differences for microbial biomass and microbial activity were found between Solling and Lange Bramke (Mintrup & Höper 2012). On the other hand, the high values at Lange Bramke are measured on the southern slope where a higher influence of the sun radiation can be expected which may increase the decomposition (Bisht et al. 2014).

Another explanation of technical nature are possible differences during installation of the suction plates without suction at Lange Bramke and the installation of low suction plates at the Solling site. The suction plates were installed at the transition zone from organic layer to mineral soil. Nieminen et al. (2013) describe the difficult installation of (zero tension) plate lysimeters and refers to disturbance of the humus layer. In Figure 20 the component values of the four suction plates of the Solling site are presented separately. The plates 2 and 4 show highest and mostly clear positive values with a large span from April to August. Plate 4 reveals very high component values which are similar to the ones measured and the Lange Bramke site. The plate 3 of the Solling site shows negative values similar to the depth of 10cm in most cases. Plate 1 fluctuates weakly by the value zero. Such differences can not be recognized for the first and second component and is only slightly pronounced for the third component. As described above this component is strongly influenced by processes taking place above the mineral soil which is why possible weak deviations of the installation locations (transition zone of humus layer and mineral soil) may have a correspondingly strong effect. At the Solling site the vacuum applied may additionally induce a negative hydraulic gradient from the subjacent mineral soil to the lysimeter plates if installed closer to or within the mineral soil supporting the occurrence of negative component values.

The pattern of the fourth component is similar to effects of different methods. It is known that the sampling of soil solutions with different methods leads to varying results (Marques et al. 1996, Nieminen et al. 2013). A comparison of the chemistry of soil solution collected with zero-tension lysimeters and suction lysimeters differs considerable. When comparing the methods higher concentrations of Ca and K for zero tension lysimeters could be proven, while concentrations of e.g. Al, DOC, Cl, Mg and the nitrogen compounds were enhanced in soil solution from suction lysimeters (Marques et al. 1996). The described results do not correspond completely to the signature of the fourth component (cf. Figure 13), but reveal similarities for e.g. K, Al or DOC.

Such differences are explainable by difference residence time of the collected soil solution in the soil. Zero suction lysimeters rather collect soil solution with lower residence time like fast draining water in macro pores (Nieminen et al. 2013), which can be close to the chemical characteristics of rainfall or throughfall. However suction cups collect soil solution which closer to the equilibrium between soil matrix and soil solution (Marques et al. 1996), which is mostly from meso- and fine pores. These differences should be regarded, especially when dealing with parameters with fast kinetic like NH_4 and K.

A possible main influence of the sampling method can be excluded, because of a missing effect in depths below 10cm. However, an intensified pattern of the component is possible due to different methods.

Summarizing the results it can be stated, that this component reflects a variety of possible and hardly distinguishable influences. By this reason, the various impacts and processes are summarized as surface processes.

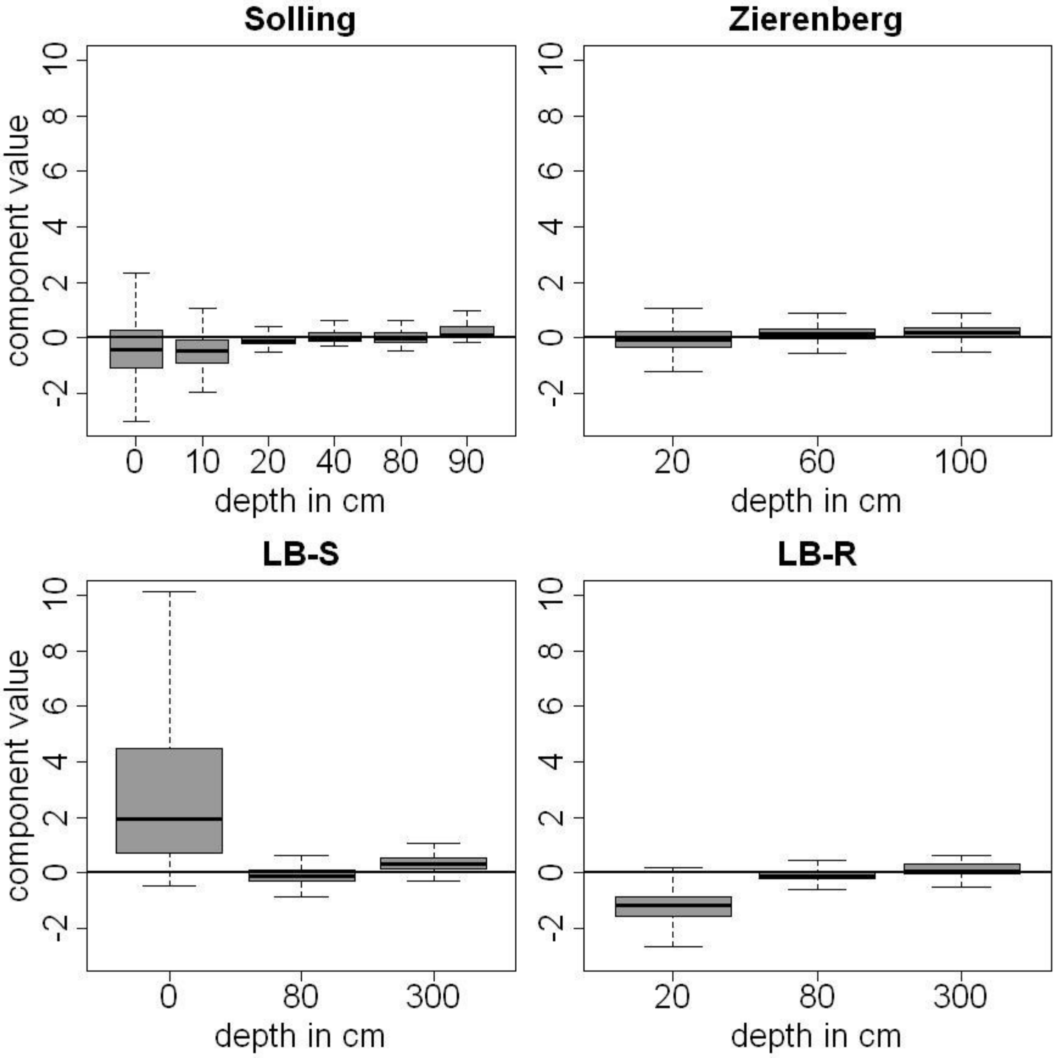

Figure 18: Scores of the fourth component at different sites and depths

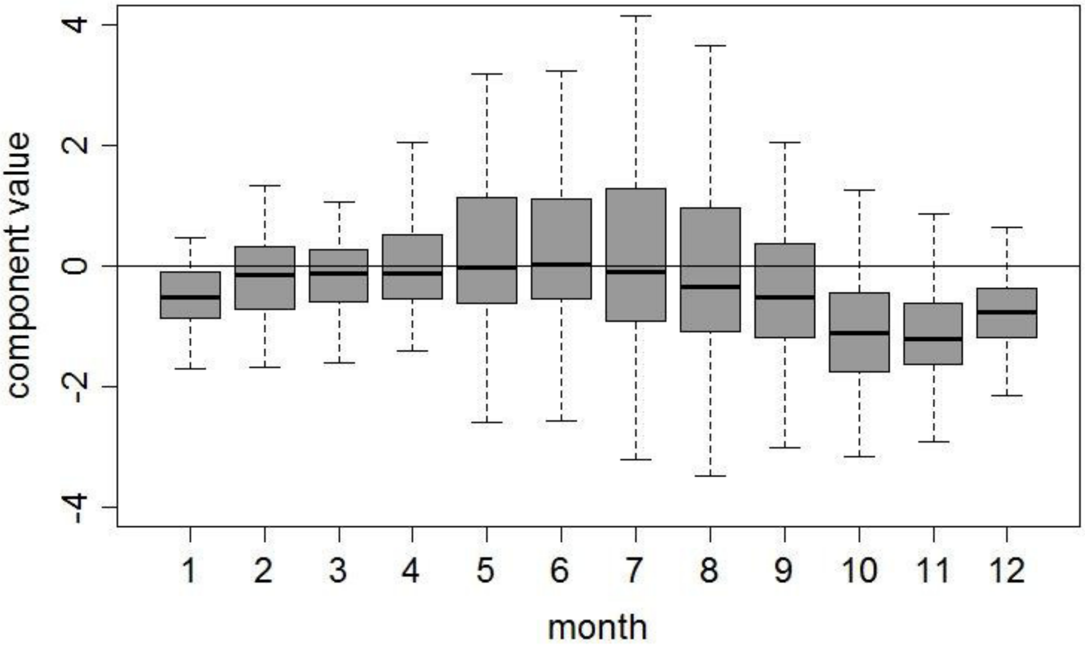

Figure 19: Scores of the fourth component at the Solling site at 0 cm depth

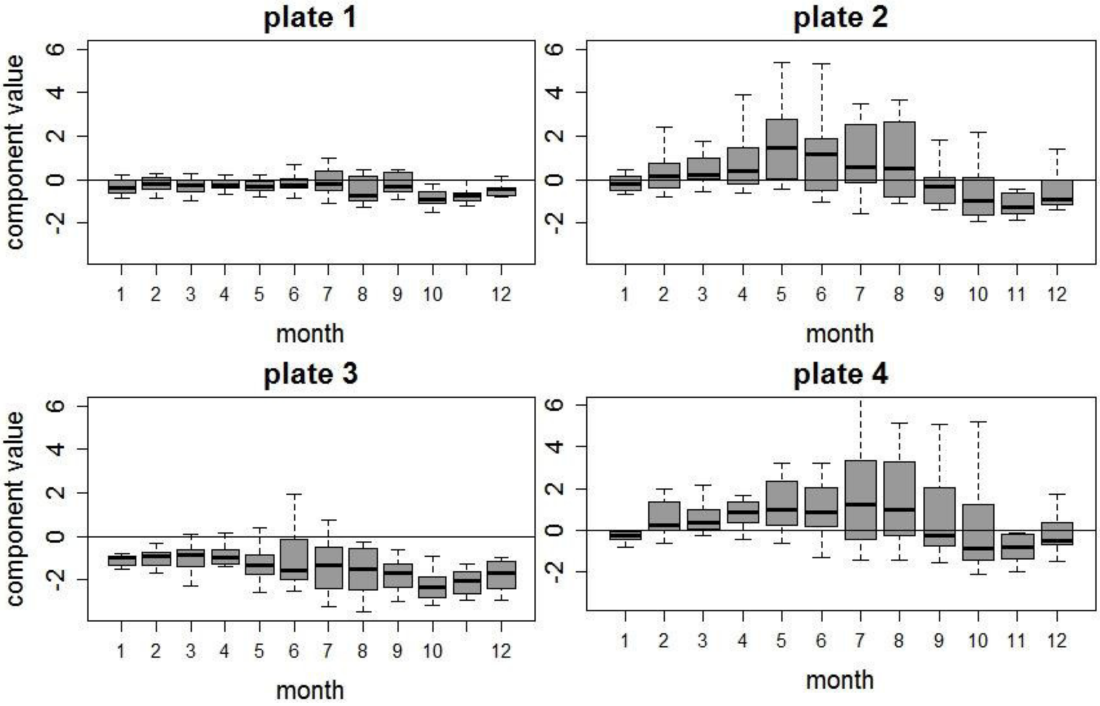

Figure 20: Scores of the fourth component at the Solling site at 0 cm depth for each suction plate

4.2.3.5 Single solutes

Besides the identification of processes using several variables, the Isomap gives information about the interplay between single solutes and the different processes. In Figure 13 it is shown, which variable is influenced by the different processes. Impacts of single processes now could be assessed with regard to their impact on single solute concentration. For example the concentration of Al in soil solution is influenced clearly by the first three components, while Ca concentration is dominated by the first component. Therefore, a number of relevant variables can be chosen for the setup of further research when analyzing single processes. The presented results show for example, that Ca should be regarded, when comparing the soil development on the three different sites (first component), but can be dismissed when dealing with impacts connected with the third component, like evapotranspiration.

4.2.4 Conclusions

Taking advantage of the available large datasets as well as of a modern nonlinear dimensionality reduction approach this study aimed at disentangling different effects that were supposed to act at different intensities at the three different sites. By outperforming the PCA slightly the Isomap revealed nonlinear relationships in the dataset, which can be expected in soil monitoring datasets.

That approach obviously was successful in extracting the effects of site-specific different bedrock lithology (first component) from soil formation effects (second component) and general climatic effects (third component) that were immanent at all three sites at about the same magnitude (Table 9). The forth component emphasized the noticeable peculiarity of a specific depth. Since the depth of 0 cm was regarded at two sites only, an overall impact cannot be proven, but expected.

In addition the results allowed to weight the different processes according to the respective fraction of variance explained (Table 9). In total, four processes with different spatial and temporal patterns could be identified, explaining nearly 96 % of the variance. In spite of clearly differing site characteristics, overall processes and impacts such as buffering of acidic inputs, evapotranspiration are responsible for 39 % of the variance in soil solution composition. Component specific spatial and time trends could be extracted.

Table 9: Similarities and differences of the researched sites

Component	Name	Explained variance	Impact
1	Bedrock	54.4 %	site specific
2	Acid induced soil formation effects	31.5 %	overall
3	Concentration through evapotranspiration	7.5 %	overall
4	Surface processes	2.2 %	site specific because of sampling depths; expected overall

The results of Isomap are additionally valuable for further applications. Different solutes of the dataset can be rated with respect to their relevance for advanced investigations.

4.3 Further intensive monitoring sites

In this chapter a study dealing with soil solution data from outside the GSMP are presented. The results were published in the scientific journal GEODERMA in 2010 (Appendix III). The data of soil solution analyses were collected in the Lehstenbach catchment, which is located in the \`Fichtelgebirge\` in South Germany. In contrast to the GSMP the data were not collected within the focus on soil science. Nevertheless, the underlying dataset of the presented study is a large, high dimensional and very heterogeneous (Appendix III) and against this background comparable to the complex datasets generated by the GSMP (chapter 2.1).

The intensive investigation of a small stream catchment area (4.2 km^2) aimed at the exploration of ecosystem processes from e.g. hydrology, soil science, climatology etc. and their interdependencies (Matzner 2004). However, the techniques of sampling using suction cups (cf. Appendix III and chapter 4.2.2.3) and measurement technique are similar. In the Lehstenbach catchment the study sites are close to each other, providing the opportunity to have a look at large scale processes compared to chapter 4.1 and 4.2. In the GSMP the distances between different monitoring sites are far larger (Figure 1) in most cases. Exceptions were only made at a few single sites with enclosed subplots, as presented for the Lange Bramke site in Table 5. Regarding the large scale in this chapter all scales of the research concept (chapter 2.2) are considered.

The data were provided by Prof. E. Matzner from the University of Bayreuth and by the former Federal Bavarian Water Resources Agency. The analyses of soil solution and field work, including maintaining the study sites and collecting the samples, were conducted by the former Bayreuth Institute of Terrestrial Ecosystem Research. Research idea and concept of interpreting the data were generated by Prof. Gunnar Lischeid and me. Prof. Gunnar Lischeid and Prof. Jörg Rinklebe supported the published paper (Appendix III) by helpful discussions and correcting the manuscript before submission.

4.3.1 Abstract

The large scale study comprises more than 4000 soil solution data analyses of 16 parameters from several depths and three soil types which could be analyzed. The used nonlinear isomap approach slightly outperformed the linear PCA, indicating nonlinear relationships in the dataset (Appendix III).

As dominating and most important impact on soil solution chemistry the deposition could be identified. Nearly 60% of the dataset variance can be explained by the first component “Deposition” and the third component “Changing deposition chemistry”, clearly reflecting the importance of deposition for soil solution chemistry. The second component “Acidification”, explaining 28%, is mainly initiated by deposition (Meesenburg et al. 2016). Further impacts, like matrix-solution interactions (fourth component – 5.8% of explained variance) and decomposition of organic matter (fifth component 1.6% of explained variance) can only explain a small part of the variance in soil solution chemistry within the Lehstenbach catchment (Appendix III).

4.3.2 Conclusions

As well as in the medium and small scale spatial and temporal patterns could be shown. Spatial pattern and/or temporal patterns (seasonal and long-term) could be clearly assigned to the different components. The higher efficiency of Isomap indicates that nonlinear relationships can be found in the dataset, which is in accordance with other studies (Ter Braak 1998, Mahecha et al. 2007, Lischeid 2014, chapter 4.2). The analyses show, that often used linear approaches in soil science, like the PCA (e.g. Astel. et al. 2011, Henriksson et al. 2013, Zhou et al. 2016), may not be the best choice. Therefore, the Isomap approach is more suitable to reach the aim of intensive soil monitoring, the identification and documentation of processes influencing the soil system by measuring the soil solution, to quantify the main drivers on soil solution chemistry and to denominate the most relevant variables, than the PCA (Appendix III). Thus, the Isomap approach can be recommended when analyzing large heterogeneous datasets in soil science.

The characteristics of the impacts and processes could be identified. More than 94% of the variance of the high dimensional dataset can be explained by only five components. As the main driver of soil solution chemistry in the soils of the Lehstenbach catchment the deposition can be identified (Appendix III).

Even if not separated as an own component the impact of the soil form can still be recognized in the first component. The relevant impact of parent material, as presented in chapter 4.1.3 (Figure 7) and chapter 4.2.3.1 (Figure 14), here is represented e.g. by the histosol and its redox-characteristics (Appendix III). The redox processes modify the effects of deposition input (Appendix III). Groundwater influenced soils, like e.g. Gleysols and different histosols mostly reveal anoxic conditions and low redox potentials (Scheffer & Schachtschabel 2010). Such anoxic conditions, which are proven for the histosols in the Lehstenbach Catchment (Alewell et al. 2006, Lischeid et al. 2007), can reduce anions like NO3 and SO_4 to gaseous forms and completely remove them from the soil solution (Kretzschmer 2010, Fritzsche et al. 2016). The decreasing amount of cations in the soil solution in consequence of the loss of anions can cover the effect of deposition input, as indicated by the first component (Appendix III). Similarly Weyer et al. 2014 identified redox processes as the most relevant impact when regarding soil -, ground- and stream water of the Lehstenbach catchment. The use of groundwater samples which exhibited more strong redox characteristics (Lischeid & Bittersohl 2008) may explain the higher impact of redox processes in the mentioned study.

In opposite to the soil type the impact of soil acidification is separated from deposition by an own component. Nevertheless it should be kept in mind that the acidification of soils is mainly caused by the input of acidic ions like SO_4 with deposition, even though the input has been reduced considerably in the last decades (Brumme et al. 2009, Waldner et al. 2014, Meesenburg et al. 2016) which is well documented for the Lehstenbach Catchment (Matzner et al. 2004) as well. The impacts of acidification could equally well be identified in a dataset including soil solution, groundwater and stream water samples explaining about 30% of the variance supporting the findings (Weyer et al. 2014) confirming the results and supporting the demarcation as a separate component.

Further processes like evapotranspiration, mineral weathering, plant uptake and decomposition, cause only small differences in the soil solution chemistry. Less than 8% of the dataset variance could be ascribed to the components associated with those processes (Appendix III).

It is exciting how well the processes identified by soil chemical information can also be reflected in other data. The results using the chemical analysis of the soil solution are very consistent with the results using the soil matrix potential. Lischeid et al. 2017 calculated the relevance of evapotraspiration for hydrological signals in the top soil with 5,8% using data from 2002-2009 of two sites in the Lehstenbach catchment, which is exactly the same explained variance as calculated by Schilli et al. 2010 (Appendix III).

5. Overall Discussion

In this thesis the three studies were presented, analyzing soil monitoring data at varying scales, with different parameters and of different origin. Innovative statistical approaches for analyzing soil monitoring datasets were used. The results of the different studies will be compared and interpreted in an overall discussion. The findings are compared in three sections under different aspects. The first chapter focuses on the soil processes, soil chemistry and complexity of soils in the different studies (chapter 5.1). The second chapter will evaluate the use of the presented statistical approaches when analyzing soil monitoring data (chapter 5.2). Finally recommendations for the GSMP will be given to improve its quality (chapter 5.3).

5.1 Soil chemistry and complexity of soils

Three studies were conducted to identify the characteristics of soil chemistry and the responsible processes and impacts. In this chapter the focus lies on the question, in how far the recognized characteristics and influences reflect themselves in the different scales.

In a first step (chapter 5.1.1), the results of the medium and large scale study are be compared, since they have used the same statistical approach, soil medium and a similar parameter set. The discussion of the identified and quantified processes in the comparison of the two studies is in the foreground. In a second step (chapter 5.1.2), the results are compared with those of the third study, using different scale, statistical method, parameter set and soil samples from sites with varying land use. It will be shown how the results are different due to differences in data records and methods, or are similar despite the differences. In chapter 5.1.3 conclusion of the results of the previous two chapters are given.

5.1.1 Process identification at the large and medium scale

The comparison of the identified components, the explained variance and the related impacts of the large and medium scale study (Table 10) reveals clear similarities. In both studies the parent material as well as its the specific reaction on deposition explains about half of the variance in soil solution chemistry. The calculated explained variance, 54.4% at the medium scale and 58.6% (45.7% first component plus 12.9% third component), are very close to each other. The high relevance of deposition for these components is reflected by the importance of NO_3, SO_4 and Cl which are mostly of atmospheric deposition origin. In the second study the different bedrock (acidic sites with sandstone vs. alkaline basalt over limestone) cause the differentiation, while at the large scale the redox-conditions of the Fibric Histosol are responsible for the site differentiation. The high relevance of the parent material in both studies supports the necessity to use the soil form instead of the soil type when characterizing sites.

The results of both studies are very similar regarding the second component. In the medium scale the consequences of acid induced soil formation effects explains 31.5% of dataset variance (Chapter 4.2.3.2). The second component of the large scale “Acidification”, responsible for 28% of dataset variance, shows the consequences of acidic input by deposition, which is indicated by the described depth gradient (Appendix III). The second component is marked by positive loadings of DOC and Fe opposed to SO_4, Mn and Al and most pronounced in the upper soil. Horizons close to the surface are separated from deeper horizons. When measured all sites show positive mean component values in depths down to 35cm, with exception of the Solling site, revealing positive means only down to 10cm. Despite the significant differences of the soils in both studies the similarities are very clear.

Table 10: Comparison of the different components of the medium and large scale study

C.	Medium scale study			Small scale study		
	E [%]	Component-name	Impacts and processes	E [%]	Component-name	Impacts and processes
1	54.4	Bedrock	site characteristic (bedrock / soil form), deposition, soil formation	45.7	Deposition	Deposition, site characteristic (parent material and redox conditions / soil form)
2	31.5	Acid induced soil formation effects	formation and dissolution of minerals and organic substances by different processes, podzolization	27.6	Acidification	Buffering acidic input, podzolization
3	7.5	Concentration through evapotranspiration	Evapotranspiration, root uptake, weathering	12.9	Changing deposition chemistry	Deposition chemistry, exchange processes
4	2.2	Surface processes	decomosition, mineralization, throughfall,	5.8	Matrix–solution interactions	Residence time, weathering, evapotranspiration
5	-	-		1.7	Decomposition	Decomposition, mineralization

C. = Component; E [%] = Explained variance in %

The second component in both studies should be distinguished from the deposition dominated components because of different reasons. The component characteristics, respectively the characterizing ions, are typical for the pedogenetic processes of podzolization and the buffering acidic inputs and not for deposition chemistry. Furthermore, deposition is not the only source for acidic substances in the soil. They may also derive from e.g. litter decomposition, root exudation or microbial activity (Lundström et al. 2000). However, the deposition can be named as a primary source of acidity (Brumme et. al 2009, Meesenburg et al. 2016). The significant reduction of acidic components with precipitation in Germany (UBA 2013) does not eliminate its relevance for soil acidification, because the emission reduction holds true for cationic inputs too (e.g. Waldner et al. 2015, Meesenburg et al. 2016).

The evapotranspiration is most relevant impact of the third component in the medium scale and also important for the fourth component in the small scale. Again the explained variance is quite similar with 7.5% and 5.8% (Table 10). In both studies, a separation of the different sites and the identification of seasonal effects were possible. Besides the evapotranspiration, other impacts affecting the soil water balance (e.g. plants, soil texture) as well as the residence time of water (influenced by evapotranspiration) show relevance for the chemical signature of the components.

Comparable percentage of dataset variance is also explained by the fourth component in the medium scale and the fifth component in the large scale. Both components having positive loadings of NO_3, NH_4 and K opposed to Fe, Al, SO_4 and DOC pronounced in surface near horizons. The positive loadings reflect the release of ions from organic material by e.g. decomposition and mineralization (c.f. chapter 5.5 Appendix III, chapter 4.2.3.4). The peculiarity of the surface process effects in the medium scale study may be explained by the sampling depth of 0 cm with low / no suction plates, which was not regarded in the Lehstenbach Catchment. In spite of the low explained variance and the uncertainty in interpretation due to possible natural effects or different sampling methods a similarity of the different studies is indicated.

The main difference between the large and medium scale is the separation of deposition influence in two components for the large scale. The changing deposition chemistry can be identified as a distinguishable impact from the first component dealing with deposition too. The distinction in to two components may be explained by the particular situation at the study site. The region of the Lehstenbach catchment, as well as large areas close to the border of the Czech Republic, are well known for being heavily polluted with acidic, sulphur-containing deposition loaded with heavy metals in the 90's and former decades (Rautengarten et al. 1995). Also a drastic decrease of acidic input for this area during the following years (Matzner et al. 2004) is common knowledge and presented in Figure 8 in Appendix III for the study period. A decreasing acidic deposition is known for the sites at the medium scale too (Meesenburg et al.

2016). On the other hand the deposition at the nearby sites of the small Lehstenbach catchment will be very similar compared to the sites far apart from each other of the second study. This may support the lacking differentiation in two components in the second study.

It can be stated, that the soils in both studies are influenced in a similar way by the same processes and impacts. Despite the differences in scale and soil forms regarded in the studies, the relevance and characteristic of the processes and impacts is comparable, reflected by the components and the explained variance. Even at different scales and varying substrates the complex system of soil seems to be influenced in similar ways in most instances. The main impacts given in Table 10 can explain about 90% of the variance in soil solution chemistry. The depth gradients given in some components indicate that the relevance of single impacts will change or may be not detected when the regarded depths would vary.

5.1.2 Comparison to the results of the small scale

In opposite to the other studies the contents of different heavy metals, TOC and N as well as pH of the soil matrix with different land use were evaluated in the small scale study. Nevertheless the results show clear similarities to the results of the two studies dealing with soil solution and a different parameter set.

At the small scale a general differentiation of organic horizons, mostly humus layers of forest sites, and mineral soil samples is visible in the SOM-SM. The cluster of the organic horizons and mineral horizons show clear differences in the chemical characteristics by the location in the SOM-SM (Figure 5C). Even if not pronounced in a similar way than for the large and medium scale the differentiation between forest floor horizons (respectively horizons close to the surface) and mineral horizons can be recognized in the small scale study too. The peculiar chemical characteristic of forest floor horizons is also indicated by the fourth component in the medium scale study (Chapter 4.2.3.4) and the fifth component in the Lehstenbach Catchment study (Appendix III). The more pronounced differentiation in the small scale study can be explained by a more direct consideration of the organic horizons. Here the material of the organic horizons was sampled and analyzed. In the medium scale only two sites were equipped with suction plates in contact with the organic material of the forest floor. Therefore a considerable impact of the forest floor on the soil solution chemistry could not be expected. Nevertheless it was revealed by the mentioned components.

The strong impact of deposition, especially wet deposition, could be expected when regarding the soil solution. However, the relevance of deposition is also pronounced when regarding the chemistry of the soil matrix (Chapter 4.1). Here the impact can be visualized and evaluated for each single parameter (Figure 10) in opposite to the multivariate quantification of these parameters from the second and third study. The input of further substances with deposition is indicated by Pb (Figure 10). Highest concentrations can be found in the humus layer and top soils of forest sites (Figure 6F) supporting the interpretation.

The depth gradient, which can be found for soil solution chemistry (e.g. Figure 14) is also indicated in Figure 5C and Figure 6A. The acidification can be visualized by lower pH values in top soils of forest sites compared to sub soils.

Besides the deposition the parent material of the soil is of vital importance for soil chemistry. The geogenic content of heavy metals and the texture of the weathered parent material is the most important impact for the investigated heavy metal concentrations (except Pb in forest top soils and Cd) in the mineral soil. The mineral soil samples in the small scale study clearly reflect the relevance of the parent material for the soil chemistry (Figure 6 and Figure 7), and this is also given in the medium (Chapter 4.2.3.1) and large scale (Appendix III) as relevant driver. In Figure 14 the difference between sites with alkaline and more acidic material can be shown for soil solution regarding nutrients, reflecting the relevance of the parent material.

The clear differentiation of organic horizons from bogs, fens, and mineral horizons could be found in both studies dealing with such soil horizons (large and small scale study). The first component of the large scale study divides between mineral horizons and organic horizons of the histosol (Appendix III). A clear differentiation between such horizons is also given in the SOM-SM (Figure 8). Soil water (Appendix III - Figure 5) and soil matrix (chapter 4.1.4) are characterized by lower concentrations of most of the investigated parameter, which is explainable by e.g. the low redox potential. Low redox potential can reduce the quality of organic matter and exchange capacity (Weyer *et al.* 2014). In soil solution the amount of anions (e.g. nitrate and sulphate) is reduced under such conditions (Fritzsche *et al.* 2016) leading to lower ionic strength of the solution. The differentiation supports the relevance of the soil form on chemistry of soil matrix and soil solution. In the small scale study the active bogs and fens can be distinguished (Figure 8) indicating different chemical characteristics compared to other samples, especially to mineral soil samples. Instead of the consumption of NO_3 or SO_4 (not measured in the small scale study) the low exchange capacity of the soil matrix of bogs and fens is crucial for the typical characteristics. This is caused by hardly decomposed organic matter and low the amount of reactive oxides cause of microbial consumption, both typical for anoxic conditions in pristine bogs and fens.

Other known influences on soil chemistry, as e.g. land use on heavy metal contents (Anderson et al. 2002, Schilli & Hütter 2005), are often covered by the impact of the parent material or deposition (Neupane & Roberts 2008, Zhou et al.

2016), especially on a small scale (Anderson et al. 2000, Desaules et al. 2010). Exceptions can be found for single parameter and single land use like Cu in vineyard soils (Fabietti et al. 2010, Simoncic et al. 2017).

In the presented case studies the expected complex interplay of atmosphere (e.g. deposition), hydrosphere (e.g. groundwater in bogs and fens) and biosphere (e.g. evapotranspiration and microbial activity) can be identified as reason for the characteristics of soil chemistry. Nevertheless it can be stated, that despite the complexity of the system soil similar results are achieved in the different studies. Even if regarding varying parameter sets, land use, soil horizons and scales the main impacts and processes stay the same.

5.1.3 Conclusion

To describe the soil and its characteristics different taxonomies are in use, e.g. the World Reference Base (FAO 2014) or the German soil Taxonomy (Ad-hoc Arbeitsgruppe Boden 2005). In such soil taxonomies soils were classified by diagnostic horizons, their properties, pedogenetic processes and/or the soil material. Those characteristics are summed up in the soil typ. The type of soil is a result of the different impacts and processes on the parent material with time. The results of the three studies clearly indicate, that the soil type is not sufficient to define the chemical characteristics of a soil. For example the two sites Zierenberg and Solling are classified as Cambisols, revealing a similar soil type (Table 5). Nevertheless the chemical properties are substantially different (Chapter 4.2.3.1, cf. Table 9). The relevance of the parent material for e.g. heavy metals in the soil matrix is pronounced in the first study (Chapter 4.1.3). The physical characteristic of the soil texture, which is important for e.g. heavy metal contents, is covered by the impact of the parent material for soils developed on alkaline igneous and metamorphic rocks (Figure 7A; cf. Figure 7B). Therefore the use of soil forms, as combination of pedogenetic characteristics and characteristics of the soil material, is strongly recommended.

It can be summarized, that the soil is a complex system in contact with different spheres of the environment. Numerous processes and impacts influencing the soil and soil solution chemistry are proven. From overall impacts and processes, like weathering (Johnson-Maynard et al. 2005) or deposition (Marschner et al. 1998, Waldner et al. 2015), to local ones, as are root exudates (Haichar et al. 2014), the soil-influencing effects are well known. Additionally only small changes in one of the numerous environmental conditions, like e.g. temperature or soil moisture, can influence the sensitive system soil considerable (Szukics et al. 2010, FAO 2015). However, there are a few main impacts and processes influencing soils of different pedogenesis and / or land use. The same is true for soil horizons in different depths. Besides the complex soil system and numerous known soil processes and impacts on soil chemistry different soils seem to have more commonalities than expected.

The main difference seems to be the intensity of the impacts or processes, which depends among other things on the recorded parameter set and the soils included in the underlying dataset. For example, the relevance of the special characteristic of bogs and fens cannot be found in the medium scale, because no samples from such soils were included. The same is true for, e.g. effects in the forest floor in the large scale study.

Main findings:

- The complex soil system is dominated by the influence of the same few impacts and processes at different sites.
- Despite the different scales and parameter sets, the identified relevant impacts and processes are comparable.
- The most relevant impacts from national to local scale are deposition of acidifying compounds, the parent material and the acidification. More than 80% of the dataset variance at the medium and large scale can be explained by these three main impacts and processes
- The quantified relevance (conducted at medium and large scale only) of the dominating impacts and processes can be regarded as similar despite the different scale levels.
- Soils should be described by the soil form, as the soil type of different soil taxonomies excludes the relevant parent material in most cases.

5.2 Statistical methods for analysis of soil monitoring data

The analysis of long-term soil monitoring data is a challenging task. Different common challenges were investigated in the presented studies. The first is the problem of the dataset characteristics. On the one hand soil matrix datasets comprise a large number variables and mostly a small number of replications (Morvan et al. 2008, Arrouays et al. 2012). The statistical method of the SOM-SM, which was used to analyze the dataset, is discussed with regard to the use of data visualization and the problem of identifying the main characteristics and impacts on soil chemistry and to perform trend analyses if replications are low. On the other hand, there are soil solution datasets with a high number of replications but a lower one of study sites (Chapter 2.1). Furthermore, several studies find that nonlinearity cannot be excluded when dealing with environmental monitoring datasets (Scholz et al. 2005, Lischeid and Bittersohl 2008, Tarquis et al. 2011) and can be expected in soil science (Biswas et al. 2013, Shcherbak et al. 2014), which is why the results of the comparison of the PCA and the Isomap receive special attention.

The use and potential of the used innovative nonlinear approaches will be evaluated in the separate chapter 5.2.2. As a first step the different datasets are evaluated with regard to their evaluability in the next subsection.

5.2.1 Soil monitoring datasets and requirements on statistical approaches

The decentralized management of the various monitoring areas is a problem that leads to a non-uniform dataset due to the different methods of the responsible ones (Chapter 2.1). Nevertheless, the existing guidelines for the GSMP (Barth et al. 2001) lead to comparable results, even if the used procedures are not identical (Chapter 4.1, Schröder et al. 2004, Schilli et al. 2011). The results of the three case studies show that analyses of the used datasets can fulfill the aims of monitoring, with the restriction that a trend analysis using of the basic data at the present time is difficult with standard methods due to the small number of replications. The questions which can be answered by analyzing such huge soil monitoring datasets with an extensive parameter sets exceed soil science. For example different questions dealing with climatic change (Körschen et al. 2014), human health (Zornoza et al. 2015), environmental management (Meesenburg 2013) or the success of political actions (Chapter 0 and 4.3, LUBW 2008). The latter is also the case regarding the third component of the large scale study (changing deposition chemistry) as a consequence of the implementation of the BImSchG (2013) coming into effect in 1974.

In contrast to other environmental monitoring programs, like water or air quality monitoring, a common problem of soil monitoring programs is the mentioned low number of temporal replicates when monitoring the soil matrix because of the low frequency of sampling, e.g. 5-10 years. Most of the analyzed parameters of the soil matrix in the first study (Chapter 4.1), like the contents of heavy metals (Desaules et al. 2010) and carbon (Körschens et al. 2014), react very slowly to environmental influences in most cases. Therefore, a higher frequency is not reasonable for numerous parameters (Desaules et al. 2010) and would not increase the quality of the results. Accordingly, sites in many national long-term monitoring networks have been sampled and analyzed only once to three times until today (Arrouays 2009, Arrouays et al. 2012, Marx et al. 2015). On the other hand, the soil of the monitoring sites is consumed by sampling (in particular forest soil), which means that a higher sampling frequency is not useful. Therefore the frequency of soil sampling should be defined carefully to conserve the soil monitoring sites for long-term monitoring.

In most cases the soil solution reacts more sensitively to environmental impacts, like e.g. precipitation, than the soil matrix. Therefore, the measurement of the chemical composition of the soil solution enables an early recognition of changes in the soil system (Barth et al. 2001). The parameters measured in soil solution are subjected to different kinetics. While parameters, like NH_4, K and PO_4, are known to react quickly, parameters mainly originating from soil minerals, like Si have slower kinetics in the solution. Additionally the soil solution can be and is sampled at a higher frequency than the soil matrix without consuming the soil of the monitoring site. Because of this reasons annual (e.g. Figure 4 - Appendix III) and interannual trends (e.g. Figure 9 - Appendix III; Figure 16) could be identified by the studies using soil solution, while there were only overall indications for a trend in the first study. However, it must be borne in mind that the parameter set of intensive monitoring sites is significantly smaller due to the lack of biological and physical parameters. Especially the biological parameters are of great interest because of their potential power of expression for e.g. climate change (Giardina et al. 2014, Haag et al. 2015) issues which is why the basic monitoring set is of high value.

5.2.2 Use and potential of innovative approaches

The results presented in chapter 4 reveal the high potential of the used statistical approaches, which are innovative in soil science, when analyzing soil monitoring data. The use of the Isomap approach identified three (chapter 4.2), respectively four components (Appendix III), each explaining more than 5 % of the dataset variance. In both studies the approach was able to explain about 93% of the variance with only few components. The SOM-SM was able to visualize 94% of the variance within the dataset, mainly influenced by two different drivers without quantifying them. Both approaches are able to deal with such large and heterogeneous multivariate datasets which are typical for long-term soil monitoring and explain most of the variance of the dataset. Besides the results and the proven use of both approaches the advantages and disadvantages of SOM-SM and Isomap are discussed with respect to the monitoring datasets and the aims of soil monitoring in the following.

As presented in chapter 4.1 the SOM-SM was able to identify the state of the soil. The identification and documentation of causal processes is possible. Very promising is the use of the SOM-SM to perform multivariate trend analyses. Assuming a continued sampling frequency of 5-10 years there will be not enough results in the close future for traditional trend analyses, like regression analyses or Mann-Kendall-Test. The presented approach of SOM-SM offers a alternative for multivariate trend analyses, if there are numerous sites with a low number of replicates. In chapter 4.1.3 it was shown, that there are hints for a multivariate trend, based on the data of few sites from the GSMP with low number of replications. The number of sites with available data from replicates will increase with time when monitoring continues. Accordingly it can be assumed, that the indicated trends (Figure 9) can be verified by decreasing confidence intervals because of the increasing number of data, even if only a period of 10 years is covered.

Advantage and disadvantage of the SOM-SM is the output for interpreting the results. The output of the SOM-SM is only one graph and has no output function as known from other artificial neural networks. Nevertheless, the entire explained variance is visualized in this single graph, and different effects and processes can be derived by the colouring of the graph in different ways, depending on location and pattern characteristics. The results of PCA and Isomap have to be presented in numerous figures and tables with statistical values (e.g. graphs of parameter loadings or component values) depending on the number of relevant components. Compared to the quantitative results of PCA and Isomap, in which different effects and processes can be quantified and distinguished, the visualization of the SOM-SM is, however, qualitative.

The interpretation of the results of SOM-SM and Isomap differs clearly. The SOM-SM approach aims at arranging the observations by their similarity. The scientific interpretation of the results is supported by the power of the human brain. Different colour codes used for the SOM-SM can be interpreted by eye-minded pattern recognition. The human brain is able to recognize and analyze patterns and anomalies in patterns in presented in figures in very short time, independent from the shown scientific contents. Therefore the results may be easier to understand even for a non-statistically skilled audience. Additionally the ability of the human brain to abstract can recognize hazy patterns (Keim 2002) which may not be detected when using other statistical approaches. Therefore, the SOM-SM can be used as a first step analyzing tool, which can identify pattern fast and efficient.

The interpretation of the PCA and Isomap results is a task for experts. The attribution of a component to a process, impact or a combination of processes or impacts is complicated and requires a deeper knowledge of the researched topic. Moreover, it is possible that the results obtained can be caused for various impacts which can not be unambiguously identified and distinguished from one another, as shown in chapter 4.2.3.4. The presentation of the results to layperson is very difficult. The comparison of the linear PCA and the nonlinear Isomap approach shows in both comparative studies that the nonlinear Isomap approach outperformed the linear PCA, even if only slightly. Correspondingly it can be stated that there are nonlinear interactions influencing soil chemistry and these are not detected by the linear approach of the PCA. This supports the findings of different other studies dealing with ecological data (Mahecha et al. 2007; Ramette 2007, Weyer et al. 2014). With regard to the monitoring aims of process identification the superiority of the Isomap for this purpose could be proven in chapter 4.2 and 4.3. Based on the results the use nonlinear statistical approaches are recommended when analyzing soil monitoring data, especially when nonlinear relationships cannot be excluded.

The need of complete data sets is one disadvantage both approaches have in common. In all studies the necessity of complete data sets reduced the size of the useable dataset considerable. The problem of incomplete data sets will be distinct when dealing with data from diverse sources, like different institutions, federal states or states, which use different parameter sets or methods in their monitoring programs (Schilli et al. 2011). Another problem will be a change the analysis methods and/or parameter sets. The information of not continuously (comparable) measured parameter cannot be regarded in the analyses.

The results of the three studies show, that the datasets and used statistical approaches are suitable tools when analyzing soil monitoring datasets to achieve the aims of soil monitoring. Both approaches seem to be promising tools for future analyses.

Main findings:

- The GSMP can provide information to fulfill all national and international demands and aims of a comprehensive monitoring.
- Intensive monitoring sites of the GSMP can provided large datasets with high temporal resolution suitable for trend analyses and to identify inner annual characteristics and impacts.
- The frequency of the soil matrix sampling of the monitoring areas provides a dataset which is currently only conditionally suitable for trend analyzes.
- The SOM-SM approach is a powerful tool to give insight in information hidden in soil monitoring datasets as a qualitative first step analyze.
- The successful trend analysis with the SOM-SM can deal with the typical problem of the soil matrix monitoring datasets – the numerous sites with a low number of measurements in time.
- The nonlinear isomap approach (slightly) outperformed the similar linear PCA, thus demonstrating nonlinearity in soil datasets.
- Both methods can
    - deal with soil monitoring datasets and perform a clear dimensional reduction without a substantial loss of information (explained variance >93%).
    - analyze the soil monitoring datasets and answer the questions addressed to the soil monitoring
- The analysis of soil monitoring data can be used to answer questions that go beyond the field of soil science.

5.3 Recommendations for monitoring programs

When researching complex system as the pedosphere which is in permanent contact and interrelation with other systems each measured parameter possibly gives specific information. Therefore a wide spectrum of regarded parameters is useful. To ensure the reliability and comparability of the data from different stakeholders each regarded parameter or characteristic should be measured or described identically in the ideal case, or at least in a comparable way. On the other hand it is known, that measurements in different laboratories using standardized analytical methods can provide differences in the results, as shown for the European forest monitoring network (Cools et al. 2004), revealing the necessity of high quality standards and management. Often it is necessary to exclude possible valuable parameters from analyses because of different reasons, like a large number of comparable or missing values because of different measuring methods (table 11 Appendix I). The recommendations given in Barth et al. 2001 aim at avoiding such problems. However, when analyzing the dataset of the GSMP only a few of more than 70 mandatory parameters could be used (Chapter 4.1.2). The recommendations given in Barth et al. (2001) have not been implemented in the different federal countries in a similar way. Even if the degree of harmonization of methods of the GSMP can already be characterized as high for large parts (Schröder et al. 2004), it should be improved for different parameters further to increase the quality of the data. A comprehensive harmonization of sampling and analysis methodology is essential for soil monitoring (Nieminen et al. 2013).

Analyzing for time trends in soil monitoring data is a challenging task, especially when regarding time spans of more than a decade with data provided by different stakeholders. The first and most important thing for analyses in all cases and fields is a reliable dataset. The necessity of comparable methods was already mentioned. Besides the harmonization of working procedures the technical advances support and increase the data quality by e.g. more accurate analyzing tools. On the other hand the increased accuracy of measurements can complicate the interpretation of the data. Within the time span of more than 20 year of the GSMP the technical standards had changed and make it necessary to have a closer look at the data. Here, the measured contents of different parameters often were close to the detection limit, for example for Cd, Hg and organic pollutants. Generally the detection limits are decreasing with time. The detection limits for Cd are given in the GSMP dataset vary in some federal states from 0.5 mg/kg (1985) to 0.01 mg/kg (2007). Additionally, the detection limits vary between the different federal states, adding a spatial uncertainty in federal wide analyses for such parameters.

The demand on soil monitoring data to answer questions of different topics changes with time. For example 2009 in France the most inquired topics were contamination, erosion and soil organic carbon decline, while soil acidification - one of the top topics of the 80's – takes a backseat (Richer de Forges & Arrouays 2010). The decreasing relevance of acidification may be explained by first sign revealing a recovery of soils from acidification (Graf Pannatier et al. 2005, Meesenburg et al. 2016), even if the effects are still recognizable in the soils (chapter 4.2.3.2 and 4.3.1). Against the background of the increasing number of environmental quality standards in the European Union and expansion of the parameter set seems to be necessary. Until today the amendment of the BBodSchV (BMUB 2015) is in discussion, even if the first step towards implementation was carried out in May 2017 (BMUB 2017). Threshold values for so far not regarded elements (e.g. thallium, vanadium) will be defined. At least the new parameters of a revision of the BBodSchV should be added to the mandatory parameter set. With modern analytical methods the measurement of additional parameters is cost-efficient compared to the maintenance of the sampling sites and sampling procedure. A consideration of additional parameters which can be measured by modern analytical apparatus easily and at low costs seems to be efficient and forward-looking, but depends on the used apparatus, supporting further method harmonization. Reducing the list of parameters should carefully be thought about because the prospective relevance cannot be clearly evaluated. As an example the measurement of different radioactive isotopes at different GSMP sites in Bavaria (Spörlein & Wölfel 2011) or Austria (Smidt et al. 2012) can be mentioned, induced by the reactor catastrophe in Tschernobyl in 1986. The general recording of radioactive isotope receives new relevance with the consequences of Fukoshima accident caused by a Tsunami. Therefore, the elimination of the parameters should be avoided unless absolutely necessary.

It can be stated, that the creation of a high quality soil monitoring dataset is a complex task, which has to face several problems. To deal with such problems, as changing analytical standards and parameter sets or potential measurement uncertainty, an adequate quality management system seems to be necessary. A useful quality management system based on the measurement of retain samples, especially for soil monitoring, was presented by Meuli et al. (2013). Meuli et al. (2013) advice a parallel measurement of a retain samples when analyzing new samples. The new analyses can be referred to the retain samples, using the difference of measurements (∆1 in Figure 21 A) as correction value for the values of the new measurement. As shown in Figure 21 the indicated time trend by the non-referenced measurement cannot be found for the referenced time line, revealing the relevance of such quality assurance.

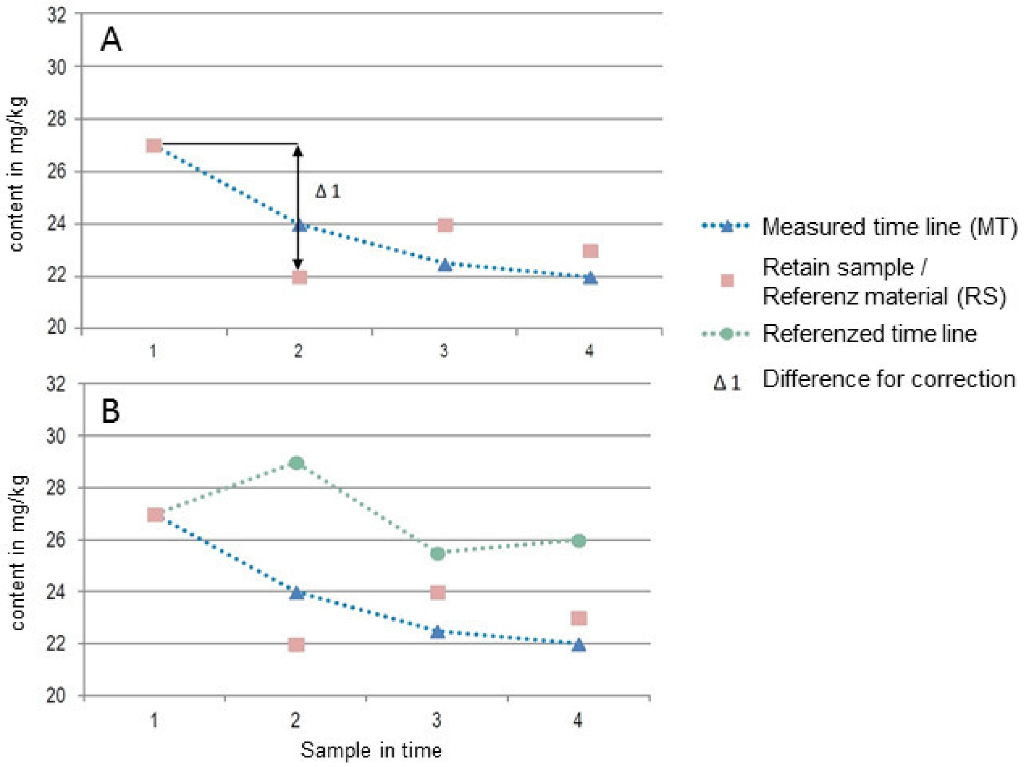

Figure 21: Part of the Swiss soil monitoring reference system (adopted from Meuli et al. 2013)

Such a reference system would provide several advantages to monitoring datasets.

- Missing values of parameters (never measured or measured with different non-comparable methods) can ascertained when analyzing the retain samples.
- The reference system also minimizes the laboratory bias caused by e.g. changing analyzing laboratories / apparatus or different employees.
- Results below the detection limit may change to analyzed low concentration because of advances in analytical techniques.

By using such quality assurance the quality of monitoring data and the number of measured parameters can be increased, especially if samples have been analyzed for a long time span and the parameter spectrum has changed in between.

The use of the presented quality assurance requires an adequate amount of retain samples. This can be a problem for long-term soil monitoring systems regarding the soil matrix because more soil material would be sampled and stored. The collection of additional soil material is especially problematic at forest sites with a humus layer. Here, it is necessary to collect material from a large area because of the low raw density and possible low thickness of humus layers. Therefore a careful setup of the sampling and analytical design with a long-term strategy is necessary. Such problems generally do not occur, or occur with smaller relevance, when monitoring air or water in similar frequencies and can be regarded as an additionally challenge for soil scientists. The increasing amount of necessary retain samples increases the cost for storing the samples over a long period. Different parameters make different demands on sample stocking (Pezzolesi et al. 2000). While air-dried samples are appropriate for e.g. heavy metal contents frozen retain samples are necessary for different biological parameters or organic contaminates. The storage and related preparation of the latter is more complicated and expensive. Accordingly the mentioned system of quality assurance is limited to available adequate retain samples which are often missing.

Besides the varying used methods the documentation of those is not always sufficient, as presented in Table 1 and Table 12. To avoid insufficient documentation the UBA presented a method code developed for the GSMP dataset (Kaufmann-Boll et al. 2011).

In conclusion, it should be noted that a continuation of the monitoring is urgently recommended. Ideally, it is continued in at least the existing or an adapted and expanded frame, as mentioned above. The continuation of soil monitoring will increase the quality of the dataset considerably for two reasons:

- The number of measurements will increase. High population supports statistical analyses.
- The investigated time span increase, and this is very important for research regarding parameters with slow kinetic (Desaules et al. 2010, Körschens 2010).

The continuously growing number of data sets will increase the efficiency by far when identifying time trends and secure them statistically. Nonetheless, it can be stated that even if the number of replicates at present is low, multivariate trend analyses of soil monitoring data are possible if adequate statistical approaches were used.

Main findings:

- Maintenance of basic monitoring is necessary, as the value of the data will increase considerable in close future, when more quality assured replications are available for trend analyses.
- The operational procedure like soil sampling, soil analyses and data management should be harmonized at the national level to ensure the comparability and a joint analyzability of the data.
- Gaps in the documentation should be filled if possible.
- The broad parameter set should be kept or even expanded to be able to answer prospective questions because of changing environmental quality standards.
- The implementation of a quality management system to reduce laboratory bias and time effects is recommended and should be connected with an adapted management of retain samples.
- Measurement of retain samples is recommended
    - as part of a quality assurance as presented above,
    - to complete data sets,
    - to analyze parameters of a possibly adapted parameter set,
    - to analyze parameters which were not detected in former times because of high detection limits in past decades
- The federal fragmentation within the GSMP lead to difficulties when joining data for nationwide analysis.

6. Synthesis

A comprehensive soil monitoring is a challenging task generating heterogeneous and multivariate datasets. To analyze these datasets adequate statistical approaches are necessary. Statistical approach should be selected in dependence of the aim of analyses and the underlying dataset. Here two different approaches were used which are innovative in soil science.

A tool of visual Data Mining was chosen for the first step analyze of the basic monitoring data, which is promising when the dataset is large and the previous knowledge is low. The use of the SOM-SM approach as a tool of data visualization is well known. The data visualization and interpretation by the human brain showed numerous well–defined pattern, which indicate the most relevant impacts and processes. Difficult-to-recognize patterns show questions, such as for Cd in chapter 4.1.3, which should be investigated more closely in a next step. Additionally the SOM-SM can provide a way to successfully perform trend analyses for datasets with low number of replications at numerous sites.

The aim of the analyses of intensive monitoring sites was the identification and a quantification of the most relevant processes in soils and trends in soil chemistry. For this purpose approaches of dimension reduction were used. Since nonlinear relationships in ecological science are known a similar linear (PCA) and nonlinear (Isomap) approach was used, to identify a possible benefit of nonlinear analyses. The slight advantage of the results of the Isomap approach, compared to the PCA, had revealed that nonlinear interactions in soils can be found and should be expected and regarded in the future. Therefore the use of nonlinear approaches can be recommended, as long as nonlinearity cannot be excluded. Further research concepts can be designed more precisely using the achieved results with higher accuracy from nonlinear approaches. Additionally the dimension reduction done by the Isomap approach can be used to set up models with lower dimensions resulting in a reduced model uncertainty.

The three presented studies are based on datasets varying in different characteristics, e.g. soil medium (soil matrix and soil solution), regarded land use, observed parameters and sampling frequency. Despite of all these differences all the studies show similar main drivers of the chemical characteristics of the soil. The changes in the chemical characteristics of the complex soil system can be attributed to and explained by only a small number of driving impacts and processes. In the presented studies the main driver's deposition and parent material, as well as acidification, could be identified in all studies, regardless of the study scale. Despite the complexity of the system soil, the different soils seem to have more commonalities than expected. The identified impacts and processes can be distinguished from each other and quantified as shown for the different components in chapter 4.2.3 and Appendix III. The determination of the main drivers on soil chemistry increases the understanding of the complex soil system. The identified processes can explain the past and can give hints to the future development of soils. The effects of political actions like the reduction of emissions of e.g. SO_4 by the BImSchG (chapter 4.2 and 4.3) or Pb by the BzBIG (LUBW 2008) on soils can be identified and evaluated.

The studies could clearly reveal the high relevance of the parent material of soils for the characteristics of the soil. To describe soils and the characteristics and to compare soils always the soil form, and not only the soil type, should be used.

Deficits within the GSMP were identified. The federal fragmentation of the GSMP can lead to difficulties in nationwide analysis. Besides the non-uniform parameter set in different federal states, the main problem is the diversity of used analytical methods, although recommendations for the parameter set and standard methods are available. Besides these shortcomings the GSMP can fulfill the national and international criteria of a comprehensive soil monitoring. Recommendations could be given to solve parts of the problems and to increase the quality of the dataset by analyzing retain samples.

7. Outlook and research requirements

The used statistical approaches reveal high potential for analyzing soil monitoring data. Nevertheless there is a need for further research. The advantage of the nonlinear Isomap was only slight, which is why it should be used on further monitoring datasets and compared again to PCA to confirm the superiority. Furthermore, monitoring data from soils should be used which are influenced from more dynamic or additional processes such as floodplain soil. Possibly the advantage of Isomap will be more pronounced at such sites. The results achieved with the Isomap approach (and PCA) are still interpretations and could be complemented by an isotope technique in order to, e.g. a clear determination of the sources of different parameters (Houle et al. 2014).

Both approaches should not be limited to a first step analyze of soil monitoring datasets as presented in this thesis. A combination with each other or further statistical result seems to be promising as presented by several authors of ecological research (e.g. Fujino & Yoshida 2006, Annas et al. 2007, Lischeid 2014). For example the approaches of SOM-SM and Isomap could be combined when using soil monitoring data, by e.g. colouring the data points of SOM-SM according to calculated component values of Isomap. A simple visualization of complex, quantitative results of the soil monitoring should be possible.

The SOM-SM and Isomap results can be used for further investigation. The reduction of dimensions by calculating components is promising for the use in models. The results of Isomap could be combined with further statistical approaches. The presented approaches are only a part of the numerous possibilities in the area big data analyses. Big data provide the diverse requirements and possibilities for evaluation methods.

Besides the used statistical approaches there are further promising statistical approaches when analyzing large datasets, for example approaches of Machine Learning (Qiu et al. 2016). Methods that are worth testing are numerous and in most cases free of charge (Tonidandel et al. 2016).

One of the most important advantages of comprehensive soil monitoring, which collects data and analyses parameter not only of the soil, is the large potential to answer unexpected questions, even beyond soil science. Besides soil focused research further questions should be taken into account when dealing with soil monitoring data. Actually the topic “climate change” and the consequences are discussed on political and public level. Studies in soil science can support the estimation of possible effects of climate change (Giardina et al. 2014, Marx et al. 2015). Furthermore high quality monitoring data and associated analyses can act as decision support for forest management and environmental management (Meesenburg 2013, BMU 2017) and may provide useful information for future actions, as measured contaminants may serve as a basis to develop benchmarks (Nicolas et al. 2014).

Against the background of changing statutory framework in the European Union and the FRG the parameter set should be reconsidered. In the course of time new parameters get relevant. For example different medical products in soils, as antibiotics (Aga et al. 2016), are of interest for human and ecological health as well as nanoparticels in soils (Anjum et al. 2013). Further the intended amendment of the German soil protection act will set up threshold values for to date unstated parameters, e.g. thallium, vanadium or antimony (BMUB 2015). Therefore an adaption of the parameter set of the GSMP, as recommended in chapter 5.3 and different studies (e.g. Marx et al. 2015, Nerger et al. 2016) seems to be necessary. The adaption of the parameter set should be combined with a harmonization of methods.

8. References

Abollino, O.; Aceto, M.; Malandrino, M.; Mentasti, E.; Sarzanini, C. & Petrella, F. (2002): Heavy metals in agricultaral soils from Piedmont Italy. Distribution, speciation and chemometric data treatment, Chemosphere 49: 545-557.

Ad-hoc Arbeitsgruppe Boden (2005). Bodenkundliche Kartieranleitung, E. Schweitzer-bart`sche Verlagsbuchhandlung, Stuttgart.

Ad-hoc Arbeitsgruppe Boden (2005b). Leitfaden zum Umrechnungsprogramm – Methodenvergleich Gesamtgehalte Haupt- und Spurenelemente, Hannover, *http://www.bgr.bund.de/DE/Themen/Boden/Netzwerke/Adhocag/Downloads/methodenvergleich_ad_hoc_ag.html* (Download 27.01.2015)

Aga, D.S.; Lenczewski, M.; Snow, D. Muurinen, J. Sallach, B.J. & Wallace, J.S. (2016): Challenges in the Measurement of Antibiotics and in Evaluating their Impacts in Agroecosystems: A Critical Review Journal of Environmental Quality 45: 407-419.

Alewell, C.; Paul, S.; Lischeid, G.; Küsel, K. & Gehre, M. (2006): Characterizing the redox status in three different forested wetlands with geochemical data. Environmental Science and Technology 40: 7609-7615.

Alloway, B. J. (2013): Heavy metals in soils – Trace Metals and Metalloids in Soils and their Bioavailability. Springer, Dordrecht.

Altaher, H. M. (2001): Factors affecting Mobility of copper in soil-water matrices Dissertation. Submitted to the Faculty of the Virginia Polytechnic Institute and State University. Blacksburg.

Anagu, I.; Ingwersen, J.; Utermann, J. & Streck, T. (2009): Estimation of heavy metal sorption in German soils using artifical neural networks. Geoderma 152: 104-112.

Anjum, N.A.; Gill, S.S.; Duarte, A.C.; Pereira, E. & Ahmad, I. (2013): Silver nanoparticels in soil-plant system, Journal of Nanoparticle Research 15: 1896-1923.

Annas, S.; Kanai, T. & Koyama, S. (2007): Principal Component Analysis and Self-Organizing Map for Visualizing and Classifying Fire Risks in Forest Regions. Agricultural Information Research 16: 44-51.

Andersen, M.K.; Raulund-Rasmussen, K.; Hanse, H.C.B. & Strobel, B.W. (2002): Distribution and fractionation of heavy metals in pairs of arable and afforested soils in Denmark. European Journal of Soil Science 202: 491-502.

Anderson, J. P. E. & Domsch, K. H. (1978): A physiological method for the quantitative measurement of microbial biomass in soils. Soil Biology and Biochemistry 10: 215-221.

Anderson, J. P. E. & Domsch, K. H. (1990): Application of ecophysiological quotients (qCO2 and qD) on microbial biomasses from soils of different cropping histories. Soil Biology and Biochemistry 22: 251-255.

Ansari, K.G.T.M.; Manokaran, S.; Raja, S.; Lyla, P.S. & Khan, S.A. (2014): Interaction of free-living marine nematodes in the artificial mangrove environment (southeast coast of india). Environmental Monitoring and Assessment 186: 293-305.

Arrouays, D.; Vogel, H., Eckelmann, W.; Armstrong-Brown, S.; Loveland, P. & Coulter, B. (1998): Soil monitoring networks in Europe, 16th World Congress of Soil Science. Montpellier, France.

Arrouays, D. (2009): Soil inventory and monitoring. Current issues and gaps, European Journal of Soil Science 60: 721–722.

Arrouays, D.; Saby, N.P.A.; Thioulouse, J.; Jolivet, C.; Boulonne, L. & Ratie, C. (2011): Large trends in French topsoil characteristics are revealed by spatially constrained multivariate analysis, Geoderma 161: 107-114

Arrouays, D. ; Marchant, B.P.; Saby, N.P.A.; Meersmans, J.; Orton, T.G.; Martin, M.P.; Bellamy, P.H.; Lark, R.M. & Kibblewhite, M. (2012): Generic Issues on Broad-Scale Soil Monitoring Schemes: A Review. Pedosphere 22: 456–469.

Arrouays, D. (2014): Current issues and applications of soil monitoring, European Journal of Soil Science 65: 777–778.

Arrouays, D.; Leenaars, J. G. B.; Richer de Forges, A.; Adhikari, K. Ballabio, C.; Greve, M.; Grundy, M., Guerrero, E.; Hempel, J.; Hengl, T. et al. (2017): Soil legacy data rescue via GlobalSoilMap and other international and national initiatives. GeoResJ 14: 1-19

Astel, AM.; Chepanova, L.; & Simeonov, V. (2011): Soil Contamination Interpretation by the Use of Monitoring Data Analysis. Water Air and Soil Pollution 216: 375-390.

Ayoubi, S.; Mehnatkesh, A.; Jalalian, A.; Sahrawat, K.L.; Gheysari, M. (2014): Relationships between grain protein, Zn, Cu, Fe and Mn contents in wheat and soil and topographic attributes. Archives of Agronomy and Soil Science 60: 625-638

Bahrenberg, G.; Giese, E. & Nipper, J. (2003): Statistische Methoden in der Geographie 2 – Multivariate Statistik, Gebrüder Borntraeger Verlagsbuchhandlung, Berlin

Bak, J.; Jensen, J.; Larsen, M.M.; Pritzl, G. & Scott-Fordsmand, J. (1997): A heavy metal monitoring-programme in Denmark. The Science of the Total Environment 207: 179-186.

Barekzai, A. & Mengel, K. (1992): Effect of microbial decomposition of mature leaves on soil pH. Journal of Plant Nutrition and Soil Science 156: 93–94.

Bartels, C. (1996) Mining in the Middle Ages and Modern Era in the Harz Mountains and its impact on the environment. Naturwissenschaften 83: 483-491.

Barth, N., Brandtner, W., Cordsen, E., Dann, T., Emmerich, K.-H., Feldhaus, D.,Kleefisch, B., Schilling, B. & Utermann, J. (2001): Boden-Dauerbeobachtung. Einrichtung und Betrieb von Boden-Dauerbeobachtungsflächen. In: Rosenkranz, D., Bachmann, G., König, W., Einsele, G. (Hrsg.): Bodenschutz – Ergänzbares Handbuch der Maßnahmen und Empfehlungen für Schutz, Pflege und Sanierungvon Böden, Landschaft und Grundwasser, Loseblattsammlung.

Bauchhenß, J. (1981): Artenspektrum, Biomasse, Diversität und Umsatzleistung von Lumbriciden auf unterschiedlich bewirtschafteten Grünlandflächen verschiedener Standorte Bayerns. - Bayerisches Landwirtschaftliches Jahrbuch 59: 119-124.

Bauchhenß, J. (1997): Regenwurm-Fauna. - In: Boden-Dauerbeobachtungs-Flächen (BDF): Bericht nach 10 jähriger Laufzeit 1985-1995, Teil III Boden: Gefüge, Organische Substanz, Bodenorganismen, Vegetation. - Schriftenreihe der Bayer. Landesanstalt für Bodenkultur und Pflanzenbau 6: 219-234.

BBodSchG (1998): Federal Soil Protection Act of 17 March 1998, Federal Law Gazette I p. 502.

BBodSchV (1999): Federal Soil Protection and Contaminated Sites Ordinance of 12 July 1999, Federal Law Gazette I p. 1554 (altered by Article 2 of an Ordinance from 23 December 2004, Federal Law Gazette I p. 3758).

Beresnev, I. A. & Wen, K-L. (1996): Nonlinear Soil Response – A Reality?, Bulletin of the Seismological Society of America 86: 1964-1978

Bester, K.; Banzhaf, S.; Burkhardt, M.; Janzen, N.; Niederstrasser, B. & Scheytt, T. (2011): Activated soil filters for removal of biocides from contaminated run-off and waste-waters. Chemosphere 85: 1233-1240.

BfN - Federal Agency for Nature Conservation (2007): Karte der naturräumlichen Gliederung der Bundesrepublik Deutschland 1:200.000. Bonn.

BfN – Federal Agency for Nature Conservation (2016): *https://www.bfn.de/0308_nlp.html* (Download 13.10.2016)

BImSchG (2013): Federal Immision Control Act of 17 May 2013, Federal Law Gazette I p. 1274 (altered by Article 1 of an Ordinance from 20 November 2014, Federal Law Gazette I p. 1740).

Bisht, V. K.; Nautiyal, B. P. ; Kuniyal, C. P.; Prasad, P. & Sundriyal, R. C. (2014): Litter Production, Decomposition, and Nutrient Release in Subalpine Forest Communities of the Northwest Himalaya, Journal of Ecosystems - *https://www.hindawi.com/archive/2014/294867/* (Download 21.11.2016)

Biswas, A.; Cresswell, H.P., Viscarra Rossel, R. A. & Si, B. C. (2013): Characterizing scale- and location-specific variation in non-linear soil systems using the wavelet transform. European Journal of Soil Science 64: 706-715.

BMU - Federal Ministry for the Environment, Nature Conservation and Nuclear Safty (2017): Vierter Bodenschutzbericht der Bundesregierung - Beschluss des Bundeskabinetts vom 27. September 2017, Bonn. http://www.bmub.bund.de/fileadmin/Daten_BMU/Download_PDF/Bodenschutz/vierter_bodenschutzbericht_bf.pdf (Download 29.09.2017)

BMUB - Federal Ministry for the Environment, Nature Conservation, Construction and Nuclear Safty (2015): Mantelverordnung zur Festlegung von Anforderungen für das Einbringen und das Einleiten von Stoffen in das Grundwasser, an den Einbau von Ersatzbaustoffen und für die Verwendung von Boden und bodenähnlichem Material. 3. Arbeitsentwurf 23.07.2015, *http://www.bmub.bund.de/fileadmin/Daten_BMU/Download_PDF/Bodenschutz/mantelv_entwurf_3_bf.pdf* (Download 05.05.2015)

BMUB - Federal Ministry for the Environment, Nature Conservation, Construction and Nuclear Safty (2017): Neue Regeln für mineralische Abfälle und Bodenschutz, *http://www.bmub.bund.de/pressemitteilung/ hendricks-tragfaehige-loesung-fuer-umgang-mit-bauabfaellen/* (Download 18.08.2017)

Bockheim, J.G. & Gennadiyev, A.N. (2000): The role of soil-forming processes in the definition of taxa in Soil Taxonomy and the World Soil Reference Base, Geoderma 95: 53-72.

Böttcher, S.; Merz, C., Lischeid, G. & Dannowski, R. (2014): Using Isomap to differentiate between anthropogenic and natural effects on groundwater dynamics in a complex geological setting. Journal of Hydrology 519: 1634–1641

Bogena, H.; Kunkel, R.; Puetz, T.; Vereecken, H.; Krueger, E.; Zacharias, S.; Dietrich, P.; Wollschlaeger, U.; Kunstmann, H.; Papen, H.; Schmid, H. P.; Munch, J.C.; Priesack, E.; Schwank, M.; Bens, O.; Brauer, A.; Borg, E.& Hajnsek, I. (2012): TERENO - Long-term monitoring network for terrestrial environmental research. Hydrologie und Wasserbewirtschaftung 7: 138-143.

Boysen, P. (1992): Schwermetalle und andere Schadstoffe in Düngemitteln. UBA-Texte Nr. 89/2011, Dessau-Roßlau.

Brammer, H. & Nachtergaele, F. O. (2015): Implications of soil complexity for environmental monitoring. International Journal of Environmental Studies 72: 56-73.

Brümmer, G.W (2010): Hauptnährelemente,- In: Blume, H.-P.; Brümmer, G.W.; Horn, R.; Kandeler, E.; Kogel-Knabner, I.; Kretzschmar, R.; Stahr, K. & Wilke, B.-M.: Lehrbuch der Bodenkunde. Spektrum. P. 401-429.

Brumme. R. & Khanna. P.K. (2009): Functioning and Management of European Beech Ecosystems. Ecological Studies 208, Springer. Berlin

Brumme, R.; Meesenburg, H.; Bredemeier, M.; Jacobsen, C.; Schönfelder, E.; Meiwes, K.J. & Eichhorn, J. (2009): Changes in Soil Solution Chemistry, Seepage Losses, and Input–Output Budgets at Three Beech Forests in Response to Atmospheric Depositions - in: Brumme. R., Khanna. P.K. (Eds.):

Functioning and Management of European Beech Ecosystems. Ecological Studies 208, Springer. Berlin, pp. 303-336.

Calvaruso, C.; Collignon, C.; Kies, A. & Turpault, M.-P. (2014): Seasonal Evolution of the Rhizosphere Effect on Major and Trace Elements in Soil Solutions of Norway Spruce (Picea abies Karst) and Beech (Fagus sylvatica) in an Acidic Forest Soil. Open Journal of Soil Science, 4, 323-336.

Chen, C-H, Härdle, W. & Unwin, A. (2008): Handbook of Data Visualization, Springer, Berlin.

Chen, J.; Li, W.; & Gao, F. (2010): Biogeochemical effects of forest vegetation on acid precipitation-related water chemistry: A case study in southwest China. Journal of Environmental Monitoring 12: 1799-1806.

Cools, N.; Delanote, V.; Scheldeman, X.; Quataert, P.; De Vos, B. & Roskams, P. (2004): Quality assurance and quality control in forest soil analyses: a comparison between European soil laboratories. Accreditation and Quality Assurance 9: 688-694

Dai, P.; Gwadry-Sridhar, F.; Bauer, M.A. & Borrie, M. (2015): A hybrid manifold learning algorithm for the diagnosis and prognostication of Alzheimer’s disease. AMIA Annual Symposium Proceedings Archive: 475-483 *http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4765614/* (Download 20.4.2016)

De Carvalho, G. C.; De Sa-Neto, R. J. & Barros, F. (2015): Thirty years of use of multivariate quantitative methods in benthic community ecology of marine and coastal habitats: looking to the past to planning the future. Scientometrics 105 :593–610

De Gruijter, J.; Brus, D.; Bierkens, D.J. & Knotters, M.F.P. (2006): Sampling for natural Resource Monitoring. Springer. Heidelberg.

De Vries, W., Reinds, G.J. & Bierkens, M.F.P. (2009): Assessment of the optimal time interval for repeated soil surveys at intensively monitored forest plots. Journal of Environmental Monitoring 11: 2009-2021.

Degryse, F.; Smolders, E. & Parker, D.R. (2009): Partitioning of metals (Cd, Co, Cu, Ni, Pb, Zn) in soils: concepts, methodologies, prediction and applications – a review, European Journal of Soil Science 60: 590-612.

Desaules, A., Ammann, S. & Schwab, P. (2010): Advances in long-term soil-pollution monitoring of Switzerland, Journal of Plant Nutrition and Soil Science 173, 525–535.

DIN ISO 16133 (2006): Soil quality – Guidance on the establishment and maintenance of monitoring programmes (ISO 16133:2004)

Dorau, K.; Eickmeier, M. & Mansfeld, T. (2016): Comparison of Manganese and Iron Oxide-Coated Redox Bars for Characterization of the Redox Status in Wetland Soils, Wetlands 36: 133-141.

Dulclerci, S. A. & Tavares, J. M. R. S. (2003): Introduction of Human Perception in Visualization, Imaging and Robotics 4: 60-70.

Dunger, W. & Fiedler, H. J. (1998): Methoden der Bodenbiologie. Fischer, Jena.

Dzhalankuzov, T. D. (2014): Concept and Results of Soil Monitoring in North Kazakhstan - in: Mueller, L.; Saparov, A. & Lischeid, G. (Eds.): Novel Measurement and Assessment Tools for Monitoring and Management of Land and Water Resources in Agricultural Landscapes of Central Asia. Environmental Science and Engineering, Springer. Cham, pp. 653-666.

EC – European Community (2006). Proposal for a Directive of the European Parliament and of the Council establishing a Framework for the Protection of Soil and Amending Directive 2004/35/EC, Brussels.

EEA – European Environmental Agency (2015): Land and soil losing ground to human activities, Copenhagen.

Ellmer, F. (2012): Landwirtschaftliche Dauerfeldversuche,- in: Kaufmann-Boll, C.; Tischler, B. & Siebigs, A. (Eds.): Bodendaten in Deutschland - Übersicht über die wichtigsten Mess- und Erhebungsaktivitäten für Böden, Dessau-Roßlau, pp. 33-38.

FAO – Food and Agricultural Organization of the United Nations (2014): World reference base for soil resources. World Soil Resources Rep. 106, Rome.

FAO and ITPS – Food and Agricultural Organization of the United Nations and Intergovernmental Technical Panel on Soils (2015): Status of the World's Soil Resources. Rome.

Fabietti, G.; Biasioli, M.; Barberis, R. & Ajmone-Marsan, F. (2010): Soil contamination by organic and inorganic pollutants at the regional scale: the case of Piedmont, Italy. Journal of Soils and Sediments 10: 290-300.

Fernandez, A.; Gonzales, A. M.; Diaz, J.; & Dorronsoro, J. (2015): Diffusion Maps for dimensionality reduction and visualization of meteorological data. Neurocomputing 163: 25-37.

Filipinski, M. & Gruppe, M. (1990): Verteilungsmuster lithogener, pedogener und anthropogener Schwermetalle in Böden. Journal of Plant Nutrition and Soil Science 153: 69-73.

Finke, P.; Hartwich, R.; Dudal, R.; Ibanez, J.; Jamagne, M.; King, D.; Montanarella, L. & Yassoglou, N. (2001): Georeferenced soil database for Europe. Manual of procedures (ver. 1.1). Report EUR 18092, European Soil Bureau, Joint Research Centre, Ispra.

Fischer, W.R.; Flessa, H. & Schaller, G. (1989): pH values and redox potential in microsites oft he rhizospere. Journal of Plant Nutrition and Soil Science 152: 191-195.

Fortmann, H.; Rademacher, P.; Groh, H. & Höper, H. (2012): Stoffgehalte und -vorräte im Boden und deren Veränderungen – in: Höper, H. & Meesenburg, H. (Eds.): 20 Jahre Bodendauerbeobachtung in Niedersachsen, GeoBerichte vol. 23, Hannover, pp. 48-69.

Fritzsche, A.; Pagels, B. & Totsche, K. U. (2016):The composition of mobile matter in a floodplain topsoil: A comparative study with soil columns and field lysimeters. Journal of Plant Nutrition and Soil Science 179: 18-28.

Fujino, M. & Yoshida, M. (2006): Development and validation of a method of forestry region classification using PCA and cluster analysis together with SOM algorithm. Journal of the Japanese Forest Society 88: 221-230.

GAFA - Gutachterausschuss Forstliche Analytik (2005): Handbook of forestal analytik. Loose-leaf booklet of analytical methods in forestry.

Gamez, A.J.; Zhou, C.S.; Timmermann, A. & Kurths, J. (2004): Nonlinear dimensionality reduction in climate data. Nonlinear Processes in Geophysics 11: 393-398.

Gerstberger, P.; Foken, T. & Kalbitz, K. (2004): The Lehstenbach and Steinkreuz catchments in NE Bavaria, Germany. In: Matzner, E. (Editor), Biogeochemistry of forested catchments in a changing environment - A German case study. Ecological Studies, 172, Springer, Heidelberg, pp. 15-40.

Gerwin, W.; Schaaf, W.; Biemelt, D.; Elmer, M.; Maurer, T. & Schneider, A. (2010): The artificial catchment 'Hühnerwasser' (Chicken Creek): Construction and initial properties. Ecosystem Development Band 1, FZLB, Cottbus, 56pp

Giardina, C.P.; Litton, C. M.; Crow, S. E. & Asner (2014): Warming-related increases in soil CO2 efflux are explained by increasing below-ground carbon flux. Nature Climate Change 4: 822-827

Giraudel, J. L. & Lek, S. (2001): A comparison of self-organized map algorithm and some conventional statistical methods for ecological community ordination. Ecological Modelling 146: 329-339.

Girma, K., Starr, L.H., Daryl, B.A., Brenda, S.T. & William, R.R. (2007): The Magruder Plots: Untangling the puzzle. Agronomy Journal 99: 1191-1198

Graefe, U. (1991): Ein Enchyträentest zur Bestimmung der Säure- und Metalltoxizität im Boden. - Mitteilungen der Deutschen Bodenkundlichen Gesellschaft 66: 487-490.

Graf Pannatier, E.; Luster, J.; Zimmermann, S. & Blaser, P. (2005): Acidification of Soil Solution in a Chestnut Forest Stand in Southern Switzerland: Are There Signs of Recovery? Environmental Science & Technology 39: 7761-7767.

Guckland, A.; Ahrends, B.; Paar, U.; Dammann, I.; Evers, J.; Meiwes, K.J.; Schonfelder, E.; Ullrich, T.; Mintrup, M.; König, N. & Eichhorn, J. (2012): Predicting depth translocation of base cations after forest liming: results from long-term experiments. European Journal of Forest Research 131: 1869-1887.

Haag, R.; Stempelmann, I. & Haider, J. (2015): Bodenbiologische Untersuchungen auf Bodendauerbeobachtungsflächen in Nordrhein-Westfalen im Zeitraum 1995 – 2007. Corrected version from 28.7.2015 of the 2009 published report of the Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen (LANUV NRW). Essen.

Haichar, F. Z.; Santaella, C.; Heulin, T. & Achouak, W. (2014): Root exudates mediated interactions belowground. Soil Biology and Biochemistry 77: 69-80.

Havlikova, S. & Klement, V. (2002): Monitoring of Agricultural Soils and Inputs into Soils in The Czech Republic, Archives of Agronomy and Soil Science 48:161-169.

Henriksson, S.; Hagberg, J.; Backstrom, M.; Persson, I. & Lindstrom, G. (2013): Assessment of PCDD/Fs levels in soil at a contaminated sawmill site in Sweden - A GIS and PCA approach to interpret the contamination pattern and distribution. Environmental Pollution 180: 19-26.

Hernandez, L. (2003): Heavy metal distribution in some French forest soils: evidence for atmospheric contamination. The Science of the Total Environment 312: 195-219.

Heinemeyer, O., Insam, H., Kaiser, E. A. & Walenzik, G. (1989): Soil microbial biomass and respiration measurements: An automated technique based on infra-red gas analysis. Plant and Soil 116: 191-195.

Höper, H. & Meesenburg, H. (Eds.) (2012): 20 Jahre Bodendauerbeobachtung in Niedersachsen, GeoBerichte vol. 23, Hannover.

Hojjati, S.M.; Hagen-Thorn, A. & Lamersdorf. N.P. (2009): Canopy composition as a measure to identify patterns of nutrient input in a mixed European beech and Norway spruce forest in central Europe. European Journal of Forest Research 128: 13-25.

Houle, D.; Marty, C.; Duchesne, L. & Gagnon, C. (2014): Humus layer is the main locus of secondary SO_4 production in boreal forests. Geochimica et Cosmochimica Acta 126: 18-29.

Huang, J-H; Kalbitz, K. & Matzner, E. (2008): Mobility of trimethyllead and total lead in the forest floor. Soil Science Society of America Journal 72: 978-984.

Huang, J-H; Ilgen, G. & Matzner, E. (2011): Fluxes and budgets of Cd, Zn, Cu, Cr and Ni in a remote forested catchment in Germany. Biogeochemistry 103: 59-70.

Huschek, G., Krengel, D., Kayser, M., Bauriegel, A. & Burger, H. (2004): Länderübergreifende Auswertung von Daten der Boden-Dauerbeobachtung der Länder. UBA-Text 50/2004, Dessau.

Isakin, M. & Teplykh, G.V. (2011): Research of higher engineering education quality on the base of students Interviewing data by nonlinear principal components analysis (NLPCA), Applied Ecometrics 21: 70-96

Jacob, R. (2008): Untersuchung von Schadstoff-Depositionen verschiedener Sammelmethoden auf den Bodendauerbeobachtungsflächen des Sächischen Landesamtes für Umwelt und Geologie. Diploma-Thesis. Technical University of Dresden.

James, F. & McCulloch, C.M. (1990): Multivariate analysis in ecology and systematics: Panacea or Pandora’s Box. Annual Review of Ecology and Systematics, 21: 129-166.

Jansen, B.; Nierop, K.G.J. & Verstraten, J.M. (2002): Influence of pH and metal/carbon ratios on soluble organic complexation of Fe(II). Fe(III) and Al(III) in soil solutions determined by diffusive gradients in thin films. Analytica Chemica Acta 454: 259-270.

Jenkinson D.S. (1991): The Rothamsted long-term experiments – Are they still of use. Agronomy Journal 83: 2-10.

Jia, L.; Wang, W. Li, Y. & Yang, L. (2010):Heavy Metals in Soil and Crops of an Intensively Farmed Area: A Case Study in Yucheng City, Shandong Province, China, Journal of Environmental Research and Public Health 7: 395-412

Johnson-Maynard, J. L.; Graham, R. C.; Shouse, P. J. & Quideau S. A. (2005): Base cation and silicon biogeochemistry under pine and scrub oak monocultures: implications for weathering rates. Geoderma 126: 353-365.

Jolliffe, I. T. (2010): Principal Component Analysis, Springer, New York.

Jonczak, J. & Parzych, A. (2014): The Content of Heavy Metals in the Soil and Litterfall an a Beech-Pine-Spruce Stand in Northern Poland. Archives of Environmental Protection 40:67-77

Kaste, J. M.; Friedland, A. J. & Stürup, S. (2003): Using stable and radioactive isotopes to trace atmospherically deposited Pb in montane forest soils. Environmental Science and Technology 37: 3560-3567.

Kaufmann-Boll, C.; Lazar, S.; Schilli, C. & Rinklebe, J. (2011): Auswertung der Veränderung des Bodenzustands für Boden-Dauerbeobachtungsflächen (BDF) und Validierung räumlicher Trends unter Einbeziehung anderer Messnetze - Teil A: Methodencode und Umgang mit Verfahrenswechseln, UBA-Texte Nr. 89/2011, Dessau-Roßlau.

Kaufmann-Boll, C.; Tischler, B. & Siebigs, A. (2012): Bodendaten in Deutschland - Übersicht über die wichtigsten Mess- und Erhebungsaktivitäten für Böden, Dessau-Roßlau. *http://www.umweltbundesamt.de/sites/default/files/medien/publikation/long/4291.pdf* (Download: 21.04.2014)

Keim, D.A. (2002): Information Visualization and Visual Data Mining. IEEE Transactions on visualization and computer graphics 7: 100-107.

Kelepertzis, E. (2014): Accumulation of heavy metals in agricultural soils of Mediterranean: Insights from Argolida basin, Peloponnese, Greece. Geoderma 221-222:82-90.

Kelepertzis, E. & Argyraki, A. (2015): Geochemical associations for evaluating the availability of potentially harmful elements in urban soils: Lessons learnt from Athens, Greece. Applied Geochemistry 59: 63-73.

Kerschen, G. & Golinval, J.-C. (2002): Non-linear generalization of Principal Component Analysis: From a global to a local approach. Journal of Sound and Vibration 254: 867-876.

Klinck, U.; Rademacher, P.; Meesenburg, H.; Fier, A.; Höper, H.& Schäfer, W. (2012): Stoffkonzentrationen im Sickerwasser - in Höper, H. & Meesenburg, H. (Eds.): Tagungsband 20 Jahre Boden-Dauerbeobachtung, GeoBerichte vol. 23, Hannover, pp. 133-151.

Knappe, F.; Möhler, S.; Ostermayer, A.; Lazar, S. & Kaufmann, C. (2008): Vergleichende Auswertung von Schadstoffeinträgen in Böden über verschiedene Eintragspfade, UBA-Texte Nr. 36/2008, Dessau-Roßlau.

Kohonen, T. (1982): Analysis of a simple self-organizing process. Biological Cybernetics 44: 135–140.

Kohonen, T. (2001): Self-organizing maps, Springer, Berlin.

Körschens, M.; Merbach, I. & Schulz, E. (2002): 100 Jahre Statischer Düngungsversuch Bad Lauchstädt. (Report), Bad Lauchstädt.

Körschens, M. (2010): Der organische Kohlenstoff im Boden (Corg) – Bedeutung, Bestimmung, Bewertung. Archives of Agronomy and Soil Science 56: 375-392.

Körschens, M.; Albert, E.; Armbruster, M.; Barkusky, M.; Baumecker, M.; Behle-Schalk, L.; Bischoff. R.; Cergan. Z.; Ellmer, F.; Herbst, F.; Hoffmann, S.;

Hofmann, B.; Kismanyoky, T.; Kubat, J.; Kunzova, E.; Lopez-Fando, C.; Merbach, I.; Merbach, W.; Teresa Pardo, M.; Rogasik, J.; Rühlmann, J.; Spiegel, H.; Schulz, E.; Tajnsek, A.; Toth, Z.; Wegener, H. & Zorn W. (2013): Effect of different mineral and organic fertilization on yield, N-uptake, C- and N-balance, as well as C-content and C-dynamics in the soil, derived from the results of 21 long-term field experiments in the 21th century. Archives of Agronomy and Soil Science 59: 1017-1040.

Körschens, M.; Albert, E.; Baumecker, M.; Ellmer, f.; Grunert, M.; Hoffmann, S.; Kismanyoky, T.; Kubat, J.; Kunzova, E.; Marx, M.; Rogasik, J.; Rinklebe, J.; Rühlmann, J.; Schilli, C.; Schröter, H.; Schroetter, S.; Schweizer, K.; Toth, Z.; Zimmer, J. & Zorn, W. (2014): Humus und Klimaänderung - Ergebnisse aus 15 langjährigen Dauerfeldversuchen. Archives of Agronomy and Soil Science 60: 1485-1517.

Koua, E.L. (2003): Using Self-Organizing Maps for information visualization and knowledge discovery in complex geospatial datasets. Proceedings of the 21st International Cartographic Conference (ICC): 1694-1701.

Kowatsch, A., Hampicke, U., Kruse-Graumann, L. & Plachter, H. (2011): Indikatoren für ein integratives Monitoring in deutschen Großschutzgebieten. BfN-Skripten 202, Bonn-Bad Godesberg.

Kretzschmer, R. (2010): Chemische Eigenschaften und Prozesse,- In: Blume, H.-P.; Brümmer, G.W.; Horn, R.; Kandeler, E.; Kogel-Knabner, I.; Kretzschmar, R.; Stahr, K. & Wilke, B.-M.: Lehrbuch der Bodenkunde. Spektrum. P. 121-170.

Lane, P.W. (2002): Generalized linear models in soil science, European Journal of Soil Science 53: 241-251

Langer, U. & Rinklebe, J. (2009): Lipid biomarkers for assessment of microbial communities in floodplain soils of the Elbe River (Germany). Wetlands 29: 353-362.

Legendre, P. & Legendre, L. (2012): Numerical Ecology. Elsevier, Amsterdam.

Lischeid, G.; Kolb, A.; Alewell, C. & Paul, S. (2007): Impact of redox and transport processes in a riparian wetland on stream water quality in the Fichtelgebirge region, southern Germany Hydrological Processes 21: 123–132.

Lischeid, G. & Bittersohl, J. (2008): Tracing biogeochemical processes in stream water and groundwater using nonlinear statistics. Journal of Hydrology 357: 11-28.

Lischeid, G. (2014): Non-Linear Approaches to Assess Water and Soil Quality - in: Mueller, L.; Saparov, A. & Lischeid, G. (Eds.): Novel Measurement and Assessment Tools for Monitoring and Management of Land and Water Resources in Agricultural Landscapes of Central Asia. Environmental Science and Engineering, Springer. Cham, pp. 653-666.

Lischeid, G.; Kalettka, T.; Merz, C. & Steidel, J. (2016): Monitoring the phase space of ecosystems: Concept and examples from the Quillow catchment, Uckermark. Ecological Indicators 65: 55-65.

Lischeid, G.; Frei, S.; Huwe, B.; Bogner, C.; Lüers; Babel, W. & Foken, T. (2017): Catchment Evapotranspiration and Runoff. In: Foken, T. (Editor), Energy and Matter Fluxes of a Spruce Forest Ecosystem, Springer, Berlin, pp. 355-375.

Loganathan, P.; Hedley, M. J.; Grace, N. D.; Lee, J.; Cronin, S. J.; Bolan, N. S. & Zanders, J. M. (2003): Fertiliser contaminants in New Zealand grazed pasture with special reference to cadmium and fluorine — a review. Australian Journal of Soil Research 41: 501 – 532.

Lorenz, M. (2010): Objectives, Strategy and Implementation of ICP Forests. Manual Part I - in: Institute for World Forestry (Eds.): Manual on methods and criteria for harmonized sampling, assessment, monitoring and analysis of the effects of air pollution on forests. UNECE. Hamburg. pp.1-21.

LUBW - Landesamt für Umwelt, Messungen und Naturschutz Baden-Württemberg (2008): 20 Jahre Bodendauerbeobachtung in Baden-Württemberg, Karlsruhe. *http://www.fachdokumente.lubw.baden-wuerttemberg.de/servlet/is/91065/20_jahre_bodendauerbeobachtung.pdf?command=downloadContent&filename=20_jahre_bodendauerbeobachtung.pdf&FIS=199* (Download: 13.03.2015)

Lundström, U.S.; van Breemen, N. & Bain, D. (2000): The podzolization process. A review. Geoderma 94: 91-107.

Maassen, S.; Balla, D. & Dannowski, R. (2015): Long-term behavior of groundwater chemistry in a periodically rewetted fen area covered with macrophytes. Journal of Environmental Management 150: 412-419

Mahecha, M.D.; Martinez, A.; Lischeid, G. & Beck, E. (2007): Nonlinear dimensionality reduction: Alternative ordination approaches for extraction and visualizing biodiversity patterns in tropical montane forest vegetation data. Ecological informatics 2: 138-149.

Manzoni, S.; Porporato, A.; D'Odorico, P.; Laio, F. & Rodriguez-Iturbe, I. (2004): Soil nutrient cycles as a nonlinear dynamical system. Nonlinear Processes in Geophysics 11: 589-598.

Mari, M. & Domingo, J. L. (2009): Toxic emissions from crematories. Environment International 36:131-137.

Marques, R.; Ranger, D.; Gelhaye, D.; Pollier, B.; Ponette, Q. & Goedert, O. (1996): Comparison of chemical composition of solutions collected by zero-tension plate lysimeters with those from ceramic-cup lysimeters in a forest soil. European journal of soil science 47: 407-417.

Marschner, B.; Gensior, A. & Fischer, U. (1998): Response of soil solution chemistry to recent declines in atmospheric deposition in two forest ecosystems in Berlin, Germany. Geoderma 83: 83-101.

Marx, M.; Schilli, C.; Kastler, M.; Molt, C.; Kaufmann-Boll, C.; Körschens, M.; Lischeid, G.; Rinklebe, J. (2015): Development of technical, legitimate and organizational backgrounds for adaptation to climate change from the perspective of soil protection - Project 03 - Determination of the changes of organic matter content and the causes in agricultural soils in Germany (FKZ 3711 71 213/03) - final report (in german). Wuppertal

Matzner, E. (Ed.) (2004): Biogeochemistry of forested catchments 685 in a changing environment - A German case study. Ecological Studies 172, Springer, Heidelberg

Matzner, E.; Zuber, T.; Alewell, C.; Lischeid, G. & Moritz, K. (2004): Trends in Deposition and Canopy Leaching of Mineral Elements as Indicated by Bulk Deposition and Throughfall Measurements. In: Matzner, E. (Editor), Biogeochemistry of forested catchments in a changing environment - A German case study. Ecological Studies, 172, Springer, Heidelberg, pp. 233-250.

Meesenburg, H.; Meiwes K. J. & Rademacher P. (1995): Long term trends in atmospheric deposition and seepage output in northwest German forest ecosystems. Water, Air & Soil Pollution 85: 611-616.

Meesenburg, H. & Brumme, R. (2009): General Description of Study Sites – in: Brumme. R.. Khanna. P.K. (Eds.): Functioning and Management of European Beech Ecosystems. Ecological Studies 208, Springer. Berlin, pp. 7-11

Meesenburg, H.; Brumme, R.; Jacobsen, C.; Meiwes, K.J. & Eichhorn. J. (2009a): Soil Properties – in: Brumme. R., Khanna. P.K. (Eds.): Functioning and Management of European Beech Ecosystems. Ecological Studies 208, Springer. Berlin, pp.33-46

Meesenburg, H.; Eichhorn, J. & Meiwes. K.J. (2009b): Atmospheric Deposition and Canopy Interactions – in: Brumme, R. & Khanna. P.K. (Eds.) 2009: Functioning and Management of European Beech Ecosystems. Ecological Studies 208, Springer. Berlin, pp. 265-302.

Meesenburg, H.; Sutmöller, J. & Hentschel, S. (2010): Retrospective and prospective evaluation of water budgets at Lange Bramke. Harz Mountains. Germany: Effects of plant cover and climate change. IAHS-AISH Publication 336: 239-244

Meesenburg, H. (2013): Forest monitoring in Germany - decision support for forest management and environmental management – Proceedings of International Symposium for Health Oktober 2013 Seoul, Korea

Meesenburg, H.; Ahrends, B.; Fleck, S.; Wagner, M.; Fortmann, H.; Scheler, B.; Klinck, U.; Dammann, I.; Eichhorn, J; Mindrup, M. & Meives, K.J. (2016): Long-term changes of ecosystem services at Solling, Germany: Recovery from acidification, but increasing nitrogen saturation? Ecological Indicators 65: 103-112.

Merdun, H. (2011): Self-organizing map artificial neural network application in multidimensional soil data analysis. Neural Computing and Applications 20: 1295-1303.

Meuli, R. G.; Schwab, P.; Wächter, D. & Ammann, S. (2013): Ergebnisse der Nationalen Bodenbeobachtung (NABO) - Zustand und Veränderungen der anorganischen Schadstoffe und Bodenbegleitparameter über 15 Jahre, Eigenverlag des Bundesamt für Umwelt der Schweiz, Bern.

Mihajlovic, D.; Antic-Mladenovic, S.; Radanovic, D.; Predic, T.; Babic, M.; Markovic, S. & Malicevic, Z. (2012): Contents of nickel, zinc, copper and lead in agricultural soils of the plains in the northwestern part of the Republic of Srpska. Agroknowledge Journal 13: 123-134.

Mindrup, M.; Dammann, I.; Talkner, U. & Meiwes, K.J. (2011): Zeitliche Entwicklung und Bewertung der Baumernährung – in: Höper, H. & Meesenburg, H. (Eds.): 20 Jahre Bodendauerbeobachtung in Niedersachsen, GeoBerichte vol. 23, Hannover, pp. 121-132.

Mindrup, M. & Höper, H. (2012): Bodenbiologische Untersuchungen – in: Höper, H. & Meesenburg, H. (Eds.): 20 Jahre Bodendauerbeobachtung in Niedersachsen, GeoBerichte vol. 23, Hannover, pp. 79-94.

Mitchel, C.C.; Westermann, R.L.; Brown, J.R.; Peck, T.R. (1991): Overview of long-term agronomic research. Agronomy Journal 83: 24-29.

Moche, M.; Gutknecht, J.; Schulz, E.; Langer, U.; Rinklebe, J. (2015): Monthly dynamics of microbial community structure and their controlling factors in three floodplain soils. Soil Biology and Biochemistry 90. 169-178.

Montanarella, L. & Vargas, R., 2012: Global governance of soil resources as a necessary condition for sustainable development. Current Opinion in Environmental Sustainability 4, 559–564.

Morvan, X.P.P.; Saby, N.P.A.; Arrouays, D.; Le Bas, C.; Jones, R.J.A., Verheijen, F.G.A., Bellamy, P.H., Stephens, M. & Kibblewhite, M.G. (2008): Soil monitoring in Europe: a review of existing systems and requirements for harmonization. Science of the Total Environment 391: 1–12.

Morrison, J.M.; Goldhaber, M.B.; Ellefsen, K.J. & Mills, C.T. (2011): Cluster analysis of a regional-scale soil geochemical dataset in northern California. Applied Geochemisty 26 Supplement: 105-107.

Müller, B. & Zhang, X. (2016): Causes of drying trends in northern hemispheric land areas in reconstructed soil moisture data. Climatic Change: 255-267.

Naisbitt, J. (1983): Megatrends: Ten New Directions Transforming Our Lives. Warner Books, New York.

Nerger, R.; Beylich, A.; & Fohrer, N. (2016): Long-term monitoring of soil quality changes in Northern Germany, Geoderma Regional 7:239-249.

Neupane, G. & Roberts, S.J. (2009): Quantitative Comparison of Heavy Metals and As Accumulation in Agricultural and Forest Soils near Bowling Green, Ohio. Water air and soil pollution 197: 289-301.

Nicolas, M.; Jolivet, C. & Jonrad, M. (2014): How monitoring networks contribute to the understanding and to the management of soil and forest ecosystems?, Revue forestière française 4: 95-103.

Nieder, R.; Benbi, D.K. & Scherer, H.W. (2011): Fixation and defixation of ammonium in soils: A review. Biology and Fertility of Soils 47: 1-14.

Nieminen, T.M.; Derome, K.; Meesenburg, H. & de Vos, B. (2013): Soil Solution: Sampling and Chemical Analyses - in: Ferretti, M. & Fischer, R. (Eds.): Forest Monitoring: Methods for terrestrial investigations in Europe with an overview of North America and Asia, Developments in Environmental Science vol. 12, Elsevier, Amsterdam, pp. 301-315.

Nietfeld, H.; Prenzel, J.; Helmisaari, H.-S.; Polle, A. & Beese, F. (2017): Modeling of mineral nutrient uptake of spruce tree roots as affected by the ion dynamics in the rhizosphere: Upscaling of model results to field plot scale, Ecological Modelling 345: 150-164.

Nziguheba, G. & Smolders, E. (2008): Inputs of trace elements in agricultural soils via phosphate fertilizers in European countries. The Science of the total environment 390: 53-57.

Olawoyin, R.; Nieto A.; Grayson, R.L.; Hardisty, F. & Oyewole, S. (2013): Application of artificial neural network (ANN)–self-organizing map (SOM) for the categorization of water, soil and sediment quality in petrochemical regions. Expert Systems with Applications 40: 3634–3648

Olbricht, W. (2011): Bleifrei? – Die Bleibelastung der Böden Bayerns vor 25 Jahren und heute (LfU)- In: Bayerisches Landesamt für Umwelt (Eds): Den Boden fest im Blick – 25 Jahre Bodendauerbeobachtung in Bayern. Augsburg, pp. 41-59.

Oliver, M.A. & Gregory, P.J. (2015): Soil, food security and human health: A review. European Journal of Soil Science 66:257-276.

Ou, H.-S.; Wei, C.-H.; Deng, Y. & Gao, N.-Y. (2013): Principal component analysis to assess the composition and fate of impurities in a large river embedded reservoir: Quigcaosha Reservoir. Environmental Science: Process & Impacts 15: 1613-1621.

Pathak, H.; Byjesh, K.; CharkrabartI, B. & Aggarwal, P. K. (2011): Potential and cost of carbon sequestration in Indian agriculture: Estimates from long-term field experiments. Field Crop Research 120: 102-111.

Pezzolesi, T.P.; Zartman, R.E. & Hickey, M.G. (2000): Effects of storage methods on chemical values of waterlogged soils, Wetlands 20: 189-193.

Protz, R.; Martini, P.; Ross, G.J. & Terasmae, J. (1984): Rate of podzolic soil formation near Hudson Bay, Ontario, Canadian Journal of Soil Science 64: 31-49.

Prounier, J.; Chabaux, F.; Stille, P.; Gangloff, S.; Pierret, M.C., Viville, D. & Aubert, D. (2015): Geochemical and isotopic (Sr, U) monitoring of soil solutions from the Strengbach catchment (Vosges mountains, France): Evidence for recent weathering evolution. Chemical Geology 417: 289-305.

Qiu, J.; Wu, Q.; Ding, G.; Xu, Y. & Feng, S. (2016): A survey of machine learning for big data processing. Journal on Advances in Signal Processing 67: 1-16.

R Development Core Team (2006): R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, *http://www.R-project.org.*

Ramette, A. (2007): Multivariate analyses in microbial ecology. FEMS Microbiology Ecology 62: 142-160.

Rasmussen, P. E.; Goulding, K. W. T.; Brown, J. R.; Grace, P. R.; Janzen, H. H.& Körschens, M. (1998): Long-term agroecosystem experiments: Assessing agricultural sustainability and global change. Science 282: 893–896.

Rautengarten, A. M.; Schnoor, J. L.; Anderberg, S.; Olendrzynski, K. & Stigliani, W. M. (1995): Soil sensitivity due to acid and heavy metal deposition in east central Europe. Waterm, Air and Soil Pollution 85: 737-742.

Rennert, T. & Rinklebe, J. (2010): Release of Ni and Zn from Contaminated Floodplain Soils under Saturated Flow Conditions. Water, Air & Soil Pollution 205: 93-105.

Rentz, J.A.; Alvarez, P.J.J. & Schnoor (2005): Benzo[a]pyrene co-metabolism in the presence of plant root extracts and exudates: Implications for phytoremediation. Environmental Pollution 136:477-484.

Richer De Forges, A.C. & Arrouays, D. (2010): Analysis of requests for information and data from a national soil data centre in France. Soil Use and Management 26:399-411.

Riek, W.; Wessolek, G.; Renger, M. and Vetterlein, E. (1995): Air capacity, plant available water and field-capacity of horizon substrate clusters – a statistical-analysis of soil survey laboratory data. Zeitschrift für Pflanzenernährung und Bodenkunde 158: 485-491.

Rinklebe, J. & Shaheen, S.M. (2014): Assessing the mobilization of cadmium, lead, and nickel using a seven-step sequential extraction technique in contaminated floodplain soil profiles along the central Elbe River, Germany. Water, Air, & Soil Pollution. 225:2039-2048.

Riise, G.; van Hees, P.; Lundström, U. & Strand, L.T. (2000): Mobility of different size fractions of organic carbon. Al. Fe. Mn and Si in podzols. Geoderma 94: 237-247.

Rivera, D.; Sandoval, M. & Godoy, A. (2015): Exploring soil databases: a self-organizing map approach. Soil Use and Management 31: 121-131.

Rochayati, S.; Du Laing, G.; Rinklebe, J.; Meissner, R. & Verloo, M. (2011): Use of reactive phosphate rocks as fertilizer on acid upland soils in Indonesia: accumulation of cadmium and zinc in soils and shoots of maize plants. Journal of Plant Nutrition and Soil Science 174: 186-194.

Ruppert, H. (1987): Bestimmung von Schwermetallen im Boden sowie die ihr Verhalten beeinflussenden Bodeneigenschaften. In: Ruppert, H. (1987): Natürliche Grundgehalte und anthropogene Anreicherungen von Schwermetallen in Böden Bayerns,- GLA-Fachberichte 2, München.

SAG - Sonderarbeitsgruppe Informationsgrundlagen Bodenschutz der Umweltminister-Konferenz (1991): Bericht der UAG Boden-Dauerbeobachtungsflächen. Arbeitshefte Bodenschutz 1, Bayerisches Staatsministerium für Landesentwicklung und Umweltfragen, München.

Sammon, J.W. (1969): A Nonlinear Mapping for Data Structure Analysis. IEEE Transaction on computers 18: 401-409.

Sanders, T.G.M; Michel, A.K. & Ferretti, M. (2016) 30 years of monitoring the effects of long-range transboundary air pollution on forests in Europe and beyond, 30 Years Anniversary Report, UNECE/ICP Forests, Eberswalde.

Sauer, D.; Sponagel, H.; Sommer, M.; Giani, L.; Jahn, R. & Stahr, K. (2007): Podzol: Soil of the Year 2007. A review on its genesis, occurrence, and functions. Journal of Plant Nutrition and Soil Science 170: 581-597.

Scheid, S., Günthardt-Goerg, M.S., Schulin, R. & Nowack, B. (2009): Accumulation and solubility of metals during leaf litter decomposition in non-polluted and polluted soil. European Journal of Soil Science 60: 613–621.

Scheffer, E. & Schachtschabel, P. (2010): Lehrbuch der Bodenkunde. 16 Edition, Spektrum, Heidelberg.

Schilli, C. & Hütter, M. (2005): Vorratsbasierte digitale Bodenbelastungskarte am Beispiel der Stadt Essen. Bodenschutz 1, Seite 4-10.

Schilli, C.; Lischeid, G. & Rinklebe, J. (2010): Which processes prevail? Analyzing long-term soil solution monitoring data using nonlinear statistics. Geoderma 158: 412-420

Schilli, C.; Rinklebe, J.; Lischeid, G.; Kaufmann-Boll, C. & Lazar, S. (2011): Auswertung der Veränderung des Bodenzustands für Boden-Dauerbeobachtungsflächen (BDF) und Validierung räumlicher Trends unter Einbeziehung anderer Messnetze - Teil B: Datenauswertung und Weiterentwicklung des Monitorings, UBA-Texte Nr. 90/2011, Dessau-Roßlau

Schinner, F., Öhlinger, R., Kandeler, E. & Margesin, R. (1993): Bodenbiologische Arbeitsmethoden, Springer, Heidelberg.

Scholz, M.; Kaplan, F.; Guy, C. L.; Kopka, J. & Selbig, J. (2005): Non-linear PCA: a missing data approach, Bioinformatics 20: 3887-3895.

Schröder, W.; Pesch, R. & Schmidt, G. (2004): Soil Monitoring in Germany - Spatial Representatively and Methodical Comparability. Journal of Soils and Sediments 4: 49- 58.

Schlumprecht, H.; Knuff, A. & Scherfose, V. (2015): Vorschläge zur Gliederung und zu den Inhalten von Nationalpark-Plänen. BfN-Skripten 425, Bad Godesberg.

Schulz-Zunkel, C.; Krueger, F.; Rupp, H.; Meissner, R.; Gruber, B.; Gerisch, M. & Bork, H.-R. (2013): Spatial and seasonal distribution of trace metals in floodplain soils. A case study with the Middle Elbe River, Germany. Geoderma 211: 128-137.

Schwab, P.; Weisskopf, P.; Oberholzer, H.; Scheid, S. & Berli, M. (2006): Langzeitbeobachtung von physikalischen und biologischen Bodeneigenschaften - Pilotprojekt LAZBO - Teil 1 Einleitung und Grundlagen. Agroscope FAL Reckenholz, Eidg. Forschungsanstalt für Agrarökologie und Landbau, Zürich.

Selle, B.; Schwientek, M. & Lischeid, G. (2013): Understanding processes govering water quality in catchments using principal component scores. Journal of Hydrology 486: 31-38.

Shcherbak, I.; Millar, N. & Robertson, G. P. (2014): Global metaanalysis of the nonlinear response of soil nitrous oxide (N2O) emissions to fertilizer nitrogen. Proceedings of the National Academy of Sciences of the United States of America 111: 9199-9204.

Simoncic, A.; Susin, J.; Sinkovec, M.; Leskovsek, R.; Cus, F.; Pongrac, V.Z. & Cesnik, H.B. (2017): Twelve-year investigation of copper soil concentrations shows that vineyards are at risk. Acta Agriculturae Scandinavica - Section B 67: 381-394.

Smidt, S.; Jandl, R.; Bauer, H.; Fürst, A.; Mutsch, F.; Zechmeister, H. & Seidel, C. (2012): Trace metals and radionuclides in Austrian forest ecosystems. In: Ishwaran, N. (Editor): The Biosphere. InTech, Europe, pp. 94-118.

Sparling, G. & Schipper, L. (2004): Soil quality monitoring in New Zealand: trends and issues arising from a broad-scale survey. Agriculture, Ecosystems and Environment 104: 545–552.

Spörlein, P. & Wölfl, P. (2011): Ergebnisse aus zwei Jahrzehnten Radioaktivitätsbodendauerbeobachtung. – in: LfU - Bayerisches Landesamt für Umwelt (Eds.): Den Boden fest im Blick – 25 Jahre Bodendauerbeobachtung in Bayern. pp. 20-30 *http://www.bestellen.bayern.de/application/stmug_app000009?SID=1473864933&ACTIONxSESSxSHOWPIC(BILDxKEY:lfu_bod_00083,BILDxCLASS:Artikel,BILDxTYPE:PDF)* (Download: 11.02.2015)

Stone, D.; Ritz, K.; Griffiths, B.G.; Orgiazzi, A. & Creamer, R.E. (2016): Selection of biological indicators appropriate for European soil monitoring. Applied Soil Ecology 97: 12–22.

Sutmöller, J.; Hentschel, S.; Meesenburg, H. & Spellmann. H. (2007): Auswirkungen forstlicher Maßnahmen auf den Wasserhaushalt in bewaldeten Einzugsgebieten. - in: Miegel, K.; Trübner, E.-R. & Kleeberg. H.-B. (Eds.): Einfluss von Bewirtschaftung und Klima auf Wasser- und Stoffhaushalt von Gewässern. Forum für Hydrologie und Wasserbewirtschaftung 20.07 in Rostock, pp. 235-246.

Szukics, U.; Abell, G. C. J.; Hödl, V.; Mitter, B.; Sessitsch, A.; Hackl, E. & Zechmeister-Boltenstern S. (2010): Nitrifiers and denitrifers respond rapidly to changed moisture and increasing temperature in a pristine forest soil. Microbiology Ecology 72: 395-406.

Tarquis, A. M.; de Lima, J. L. M. P., Krajewski, W. F.; Cheng, Q. & Gaonac'h, H. (2011): Nonlinear and Scaling Processes in Hydrology and Soil Science. Nonlinear processes in geophysics 18: 899–902.

Tenenbaum, J. B.; de Silva, V. & Langford, J. C. (2000): A global geometric framework for nonlinear dimensionality reduction. Science 290: 2319-2323.

Teng, Y.; Wu, J.; Lu, S.; Wang, Y.; Jiao, X. & Song, L. (2014): Soil and soil environmental quality monitoring in China: A review. Environment International 69: 177–199

Ter Braak, C.J.F. (1988): A Theory of Gradient Analysis. Advances in Ecological Research 18: 271-317.

Tischer, S. (2015): Microbial biomass and enzyme activities on soil monitoring sites in Saxony-Anhalt, Germany. Archives of Agronomy and Soil Science 51: 673-685.

Titeux, H. & Delvaux. B. (2009): Experimental study of DOC. nutrients and metals release from forest floors developed under beech (Fagus sylvatica L.) on a Cambisol and a Podzol. Geoderma 148: 291-298.

Tonidandel, S.; King, E.B. & Cortina, J. M. (2016): Big Data Methods: Leveraging Modern Data Analytic Techniques to Build Organizational Science. Organizational Research Methods. 1-23 *http://journals.sagepub.com/doi/pdf/10.1177/1094428116677299* (Download: 07.01.2017)

Toth, G; Jones, A. & Montanarella, L. (2013): LUCAS Topsoil Survey - methodology, data and results. JRC Technical Reports, Luxembourg.

Türtscher, S.; Berger, P.; Lindebner, L. & Berger, T.W. (2017): Declining atmospheric deposition of heavy metals over the last three decades is reflected in soil and foliage of 97 beech (Fagus sylvatica) stands in the Vienna Woods, Environmental Pollution 230, 561-573

UBA - Federal Environment Agency (2014): *http://www. umweltbundesamt.de/altlast/web1/berichte/pdf/bbodschv-engl.pdf* (Download: 12.01.2014)

UBA - Federal Environment Agency (2013): Das Luftmessnetz des Umweltbundesamtes - Langzeitmessungen, Prozessverständnis und Wirkungen ferntransportierter Luftverunreinigungen, Dessau-Roßlau

UBA - Federal Environment Agency (2010): *https://www.umweltbundesamt.de/sites/default/files/medien/pdfs/beschreibung_xml_schema_uba.pdf* (Download: 31.08.2016)

UBA - Federal Environment Agency (2001): Konkretisierung des Umweltbeobachtungsprogrammes im Rahmen eines Stufenkonzeptes der Umweltbeobachtung des Bundes und der Länder; Umweltprobenbank einschließlich Human- und Biomonitoring. Berlin.

Ulrich, B. (1986): Natural and anthrophogenic components of soil acidification, Journal of Plant Nutrition and Soil Science 149: 702-717.

Utermann, J.; Fuchs, M. & Düwel, O. (2010): Hintergrundwerte für Spurenelemente in Böden mit Blick auf Bodenarten-Hauptgruppen (Background values of trace metals in soils regarding the soil texture). Bodenschutz 4: 103-108.

Vance, E.D., Brookes, P.C. & Jenkinson, D.S. (1987). An extraction method for measuring soil microbial biomass. - Soil Biology and Biochemistry 19: 703-707.

van Hess, P.A.W.; Lundström, U.S.; Starr, M. & Giesler, R. (2000): Factors influencing aluminium concentrations in soil solution from podzols. Geoderma 94: 289-310.

van Wesemael, B.; Paustian, K.; Andrén, O; Cerri, C. E. P.; Dodd, M.; Etchevers, J.; Goidts, E.; Grace, P.; Kätterer, T.; McConkey, B. G.; Ogle, S.; Pan, G. & Siebner, C. (2011): How can soil monitoring networks be used to improve predictions of organic carbon pool dynamics and CO2 fluxes in agricultural soils?. Soil and Plant 338: 247-259.

Venables W.N. & Ripley B.D. (2002): Modern applied statistics with S, Springer, New York

Vesanto, J. (2002): Data exploration process based on the Self-Organizing Map. PhD Thesis at Helsinki University of Technology. Helsinki

Vestin, J.L.K.; Norström, S.H.; Bylund, D.; Mellander, P.-E. & Lundström, U.S. (2008): Soil Solution and stream water chemistry in a forested catchment I: Dynamics. Geoderma 144: 256-270.

Voyslavov, T.; Tsakovski, S. & Simeonov, V. (2012): Soil contamination interpretation using self-organizing maps. Global Nest Journal 14: 3-9.

Waldner, P.; Marchetto, A.; Thimonier,A.; Schmitt, M.; Rogara, M.; Granke, O.; Mues, V.; Hansen,K. Karlson,G.P.; Zlindra, D. et al. (2014): Detection of temporal trends in atmospheric deposition of inorganic nitrogen and sulphate to forests in Europe. Atmospheric Environment 95: 363-374.

Wang, S.; Cao, Z.; Li, X.; Liao, Z.; Hu, B.; Ni, J. & Ruan, H. (2013): Spatial-seasonal variation of soil denitrification under three riparian vegetation types around the Dianchi Lake in Yunnan, China. Environmental Science: Process & Impacts 15: 963-971

Wang, M.; Liu, D.; Jia, J. & Zhang, X. (2015): Global trends in soil monitoring research from 1999–2013: a bibliometric analysis. Acta Agriculturae Scandinavica, Section B — Soil & Plant Science 65: 483-495.

Wahsha, M. Bini, C., Zilioli, D.; Spiandorello, M. & Gallo, M. (2014): Potentially harmful elements in terraced agroecosystems of NE Italy: Geogenic vs anthropogenic enrichment. Journal of Geochemical Exploration 144 B: 355-362.

Webster, R. & Oliver, M.A. (1990): Statistical Methods in Soil and Land Resource Survey. Oxford University Press, Oxford.

Webster, R. (2001): Statistics to support soil research and their presentation, European Journal of Soil Science 5: 331-340.

Wehrens, R. & Buydens, L.M.C. (2007): Self- and Super-organizing Maps in R: The kohonen Package. Journal of Statistical Software 21: 1-19.

Welsch, N.; Schwab, J. & Liebmann, C. Chr. (2013): Materie. Springer, Berlin Heidelberg.

Wesselink, L.G.; Meiwes, K.J.; Matzner, E. & Stein, A. (1995): Long-term changes in water and soil chemistry in spruce and beech forests, Solling, Germany. Environmental Science &. Technology 29: 51–58.

Weyer, C.; Pfeiffer, S.; Schulze, K.; Borken,W. & Lischeid, G. (2014):Catchments as heterogeneous and multi-species reactors: An integral approach for identifying biogeochemical hot-spots at the catchment scale. Journal of Hydrology 519B: 1560-1571.

Wittmann, O.; Blume, H.-P.; Filipinski, M. & Witt, S. (1997): Soil classification of the federal republic of Germany, Mitteilungen der Deutschen Bodenkundlichen Gesellschaft 84: 253-275.

Ye, C.; Li, S.; Yang, Y.; Shu, X.; Zhang, J.; & Zhang, Q. (2015): Advancing Analysis of Spatio-Temporal Variations of Soil Nutrients in the Water Level Fluctuation Zone of China’s Three Gorges Reservoir Using Self-Organizing Map. PLOS ONE – Peer reviewed Online Ressource, *http://dx.doi.org/10.1371/journal.pone.0121210* (Download: 20.03.2016)

Yobouet, Y.A.; Adouby, K. & Drogui, P. (2016): Experimental Methodology to Assess Retention of Heavy Metals Using Soils from Municipal Waste Landfills. Water, Air, & Soil Pollution 227: 36-44

Zacharias, S.; Bogena, H.; Samaniego, L.; Mauder, M.; Fuß, R.; Pütz, T; Frenzel, M.; Schwank, M; Baessler, C. & Butterbach-Bahl, K. et al. (2011): A Network of Terrestrial Environmental Observatories in Germany. Vadose Zone Journal 10. 955-973

Zhou, J.; Feng, K.; Pei, Zongping; Meng, F. & Sun, J. (2016): Multivariate analysis combined with GIS to source identification of heavy metals in soils around an abandoned industrial area, Eastern China. Ecotoxicology 25: 380-388

Zornoza, R.; Acosta, J.A.; Bastida, F.; Dominguez, S.G.; Toledo, D.M. & Faz, A. (2015): Identification of sensitive indicators to assess the interrelationship between soil quality, management practices and human health. Soil 1: 173-185.

Appendix I

Table 11: Mandatory parameters for the soil solid phase (altered from Barth et al 2001)

Parameter	Method	Frequency
Soil physic		
Grain size	DIN 19683-2	once
Raw density, dry	DIN 19683-12, ISO 11 272	once
Solid substance density	DIN 19683-11	once
Pore size distribution	DIN 19683-5	once
Saturated water conductivity (kf)	DIN 19683-9	once
Soil chemistry		
pH value	DIN 19684-1, DIN ISO 10390-7	> 5 years
$C_{total.}$, C_{org}	DIN 19684-2, DIN ISO 10694	> 5 years
N_{total}	DIN 19684-4, DIN ISO 11261-8	> 5 years
Carbonate content	DIN 19684-5, DIN ISO 10693	> 5 years
CEC_{pot} (Ca, Mg, Na, K, H-Wert)	DIN 19684-8, DIN ISO 13536	> 5 years
CEC_{eff} (Ca, Mg, Na, K, Al, H, Mn, Fe)	BZE (1994); bei pH($CaCl_2$) <6,5	> 5 years
Extractable aqua regia contents: Ca, Mg, Fe, K, Mn, P, S	DIN 38414-7, DIN ISO 11446	> 5 years
Extractable aqua regia contents: Cd, Cr, Cu, Hg, Ni, Pb, Zn, As, Al	DIN 38414-7, DIN ISO 11446	> 5 years or event related
Long-lived radionuclides: ^{137}Cs, ^{134}Cs	BMU (1997)	> 5 years or event related
Polychlorinated biphenyls: PCB 28, 52, 101, 138, 153, 180	DIN 38407-2 F2	> 5 years or event related
Chlorine pesticides: HCB, α-, β-, γ-, δ-HCH, DDD, DDT, DDE	DIN 38407-2 F2	> 5 years or event related
Polycyclic aromatic hydrocarbons: 16 PACs (EPA)	DIN ISO 13877	> 5 years or event related
Soil microbiology		
Microbial biomass	Substrate-induced respiration (according to Anderson & Domsch 1978 and Heinemeyer et al. 1989), DIN ISO 14240-1	≥1 year
Microbial biomass	Fumigation-extraction method (according to Vance et al. 1987), DIN ISO 14240-2	≥1 year
Microbial Basal respiration	Continuous flow method Heinemeyer et al. (1989) or determination of O_2 uptake (Schinner et al. (1993), DIN 19737	≥1 year
Metabolic quotient	Anderson & Domsch 1990)	≥1 year
Soil zoology		
Lumbricids	Hand picked (Graefe 1991, Bauchhenß 1997)	> 5 years
Lumbricids	Formalin expulsion (Graefe 1991, Bauchhenß 1981)	> 5 years
Small annelids	Wet extraction according to Graefe 1991 or Graefe in Dunger & Fiedler 1998	> 5 years

Table 12: Different methods when analyzing the parameter Cd listed in the database of the German long-term soil monitoring program (altered from Schilli et al. 2011)

Parameter	Method
Cd	Aqua regia extractable (DIN 38414-7)
Cd	Aqua regia extractable (DIN ISO 11466)
Cd	Aqua regia extractable (laboratory specific method)
Cd	Effective Cation Exchange Capacity (laboratory specific method)
Cd	Potential Cation Exchange Capacity (laboratory specific method)
Cd	Ammonium nitrate extractable (DIN EN 19730)
Cd	EDTA-extractable (DIN 68406 E29)
Cd	DTPA-extractable
Cd	Oxalate extractable (laboratory specific method)
Cd	HNO_3-extractable
Cd	Total contents (federal state specific method)
Cd	Total contents (GAFA 2005 – chapter 3.3.2)
Cd	Total contents (laboratory specific method)
Cd	Total contents (Ruppert 1987)
Cd	X-ray fluorescence analysis
Cd	Method not defined

Appendix II

Curriculum Vitae

Der Lebenslauf ist in der Online-Version aus Gründen des Datenschutzes nicht enthalten.

The CV is not included in the online version for privacy reasons.

Der Lebenslauf ist in der Online-Version aus Gründen des Datenschutzes nicht enthalten.

The CV is not included in the online version for privacy reasons.

Appendix III

I thank ELSEVIER for granting permission to use this publication in my dissertation.

Reprinted from

Schilli, C.; Lischeid, G. & Rinklebe, J. (2010): Which processes prevail? Analyzing long-term soil solution monitoring data using nonlinear statistics. Geoderma 158: 412-420

with permission from ELSEVIER.

Geoderma 158 (2010) 412–420

Contents lists available at ScienceDirect

Geoderma

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

Which processes prevail?
Analyzing long-term soil solution monitoring data using nonlinear statistics

Carsten Schilli [a,*], Gunnar Lischeid [b], Jörg Rinklebe [a]

[a] *Soil- and Groundwater-Management, Institute for Soil Engineering, Water- and Waste-Management, Department D, University of Wuppertal, Pauluskirchstraße 7, 42285 Wuppertal, Germany*
[b] *Institute of Landscape Hydrology, Leibniz Centre for Agricultural Landscape Research (ZALF), Eberswalder Straße 84, 15374 Müncheberg, Germany*

ARTICLE INFO

Article history:
Received 3 February 2010
Received in revised form 4 June 2010
Accepted 21 June 2010
Available online 14 July 2010

Keywords:
Nonlinear statistics
Soil solution
Soil monitoring
Process identification
Trend
Deposition

ABSTRACT

Soil monitoring yields large and heterogeneous data sets. To identify the prevailing processes as well as identifying spatial patterns or temporal trends, mostly linear approaches are used. Here, a nonlinear approach, Isometric Feature Mapping (Isomap), was applied and compared to the established linear Principal Component Analysis (PCA) to a data set from a long-term monitoring program in the forested Lehstenbach catchment (Fichtelgebirge, Germany). The data set comprised more than 4000 soil solution samples from different periods, soil types and varying depths, where 16 solutes were determined.

The nonlinear Isomap approach achieved slightly better results than the linear procedure. More than 94% of the variance of the given data set was explained by the first five components. About 46% of the variance was ascribed to the impact of long-term atmospheric deposition. Soil acidification may explain the characteristics of the second component and another 28% of the data set's variance. The third component indicated a long-term shift of deposition chemistry that accounted for nearly 13% of the variance. Matrix–solution interactions and decomposition of organic matter were ascribed to the fourth and fifth component, explaining another 5.8% and 1.6% of the variance of the data set. Thus, long-term deposition could be interpreted as the most important factor influencing soil solution chemistry in different ways. Based on the Isomap results spatial and temporal patterns were investigated. Different redox conditions and depth of sampling accounted for much of the spatial variance. The identified components differed substantially with respect to seasonal patterns or long-term trends. The nonlinear Isomap approach revealed applicability and further potential for analyzing comprehensive data sets in soil science.

1. Introduction

Long-term soil monitoring programs have generated large multivariate data sets all over the world. To understand the observed temporal changes identification of the relevant processes is necessary. Besides, the latter is necessary for a better understanding of spatial heterogeneities. In most cases, various processes have an effect on single solutes. As soon as the prevailing processes are identified low-dimensional models can be set up with reduced model uncertainty compared to high dimensional models.

In soil science as well as in other ecological sciences different multivariate methods are used to analyze monitoring data. For the purpose of dimensionality reduction and process identification methods like multivariate linear regression (e.g. Graf Pannatier et al., 2005) and the common Principal Component Analysis (PCA) (e.g. Gupta et al., 2006; Zhang et al., 2007; Weyer et al., 2008; Langer and Rinklebe, 2009) are frequently used. However, many of those methods require linear data sets, but linearity often is more an exception rather than the rule in ecological data sets (Mahecha et al., 2007; Ramette, 2007). Mathematical methods based on linearity assumptions can cause difficulties in interpretation (James and Culloch, 1990). Nonlinear analyses of ecological data have revealed higher efficiency (e.g. Tenenbaum et al., 2000; Mahecha et al., 2007) analyzing different ecological data sets. However, nonlinear approaches in soil science are still very rarely used today.

In this study the linear PCA and the nonlinear Isometric Feature Mapping (Isomap) (Tenenbaum et al., 2000) were applied to analyze long-term monitoring soil solution data. The use of PCA for investigating ecological datasets, even in soil science is well established. Isomap was successfully used in different ecological sciences like botany (Mahecha et al., 2007), hydrology (Lischeid and Bittersohl, 2008) and climate research (Gámez et al., 2004). The data set was collected from different sites of an intensively studied forested catchment in south Germany (e.g. Moritz et al., 1994; Lischeid et al., 2002; Matzner et al., 2004a; Weyer et al., 2008). The sites differ with respect to, e.g., soil type and vegetation and with respect to the number of collected samples resulting in a very heterogeneous data set.

---

* Corresponding author. Fax: +49 202 4394196.
E-mail addresses: schilli@uni-wuppertal.de (C. Schilli), lischeid@zalf.de (G. Lischeid), rinklebe@uni-wuppertal.de (J. Rinklebe).

0016-7061/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.geoderma.2010.06.014

The aim of this study is i) to identify the dominating impacts and soil-chemical processes based on soil solution chemistry and ii) to study possible long-term shifts of the prevailing processes using and comparing a linear and a nonlinear approach. The results will be interpreted with regard to their relevance when analyzing long-term monitoring data.

2. Sites

The Lehstenbach catchment (4.2 km²) is located in the mountainous region "Fichtelgebirge" in South Germany, close to the border to the Czech Republic (Fig. 1). Here, numerous studies have been performed (e.g. Alewell et al., 2000; Matzner et al., 2004a,b; Bogner et al., 2008; Lischeid and Bittersohl, 2008). The altitude varies between 695 and 877 m above sea level. The mean annual temperature is 5.3 °C (1971–2000), the mean annual precipitation is approximately 985 mm (1988–1999) (Gerstberger et al., 2004).

The geological parent material is Variscan granite, intensively weathered during tertiary. Thickness of the regolith varies between 0 m and more than 30 m. Fens and bogs have developed in topographic depressions. Acidic soils, like Podzols or Dystric Areno- and Cambisols (according to IUSS/ISRIC/FAO, 2006), prevail in more than two thirds of the catchment area. Wetland soils cover about 30% of the area (Gerstberger et al., 2004).

3. Data set

Soil solution data from the Lehstenbach catchment measured from 1992 to 2000 have been used. This data pool includes 4061 samples of four different sampling sites within the catchment (Table 1). Vegetation is Norway spruce (*Picea abies*) at all sites.

At the Coulissenhieb site the soil is classified as a Haplic Podzol (according to IUSS/ISRIC/FAO, 2006) with an incumbent humus form of mor type (Gerstberger et al., 2004). Soil solution concentration data was available from four sub-plots here, covering the period from 1992 to 2000. Data were provided by the former Bayreuth Institute of Terrestrial Ecosystem Research (BITÖK) at the University of Bayreuth. The data of the four different nearby sub-plots are merged and treated as single site. The samples were taken at 20 and 90 cm depths. In addition, samples from 35 cm were taken 1992–1995 only. Four ceramic suction cups with mean pore diameter of 0.45 µm and constant suction of 240–320 hPa were installed per depth (Manderscheid and Matzner, 1995). Samples were collected at biweekly intervals and analyzed separately.

Table 1
Site characteristics (Moritz et al., 1994; Gerstberger et al., 2004).

Study site	Coulissenhieb	Köhlerloh	Gemös	Weidenbrunnen
Depths [cm]	20, 35, 90	50, 100	50, 100, 150, 200	50, 100, 150, 200
Size [ha]	2.50	~0.10	0.11	~0.15
Depth to ground water [m]	About 4.00	0.77–1.90	2.52–4.53	3.00–7.00
Stand age 2009 [year]	136	73	124	53
Mean height of tree 1992 [m]	26.70	19.93	26.95	14.90
Texture	Loam to sandy loam	Silty to sandy loam	Silty sand to silty loam	Silty to loamy grus
Exposition/form	West/slope	East/ depression	East/slope	Southwest/slope
Soil type-FAO-classification	Haplic Podzol	Fibric Histosol	Haplic Podzol	Dystric Arenosol
Number of samples	2562	338	565	596

In addition, soil solution data were provided by the Bavarian Environmental Agency (LfU; former Federal Bavarian Water Resources Agency) from another three sites in the catchment (Moritz et al., 1994). The soils at these sites were classified as Haplic Podzol at Gemös, Dystric Arenosol at Weidenbrunnen, both covered with a thick raw humus, and as Sapric Histosol (according to IUSS/ISRIC/FAO, 2006) at the Köhlerloh site (Moritz et al., 1994). Sampling depth was 50, 100, 150 and 200 cm at Gemös and Weidenbrunnen, and 50 and 100 cm only at Köhlerloh, due to the high groundwater table at this site (Table 1).

Samples were collected via ceramic aluminium-oxide sintercups with mean pore diameter of 0.5 µm and periodical suction (4–5 days) of 300 hPa (Moritz et al., 1994). Sub-samples of four different cups per depth were pooled to one sample before analysis. Monthly data are available for those sites from 1996 through 2000.

In total, 39 variables (e.g. heavy metals, nutrients, temperature and conductivity) were determined with different frequencies and periods. Variables measured in less than 3000 observations were not included in the analysis. Nitrogen and sulfur were excluded due to strong correlations to NO_3 and SO_4 (Spearmans ρ exceeding 0.9). As a result

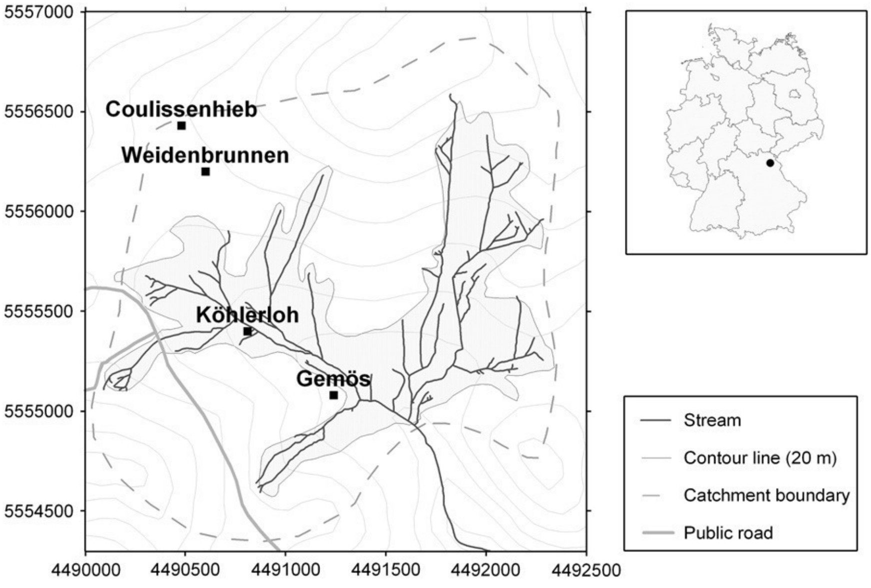

Fig. 1. Map of the study sites (left) and location of the Lehstenbach catchment in Germany (upper right).

16 variables were available for the subsequent statistical analyses: Al, Ba, Ca, Cl, DOC, Fe, K, Mg, Mn, Na, NH_4, NO_3, pH, PO_4, Si and SO_4.

4. Methods

Data values below detection limit were replaced by half the detection limit. Samples with more than two missing values were disregarded. Single missing values were replaced by the variable site and depth specific mean. This was necessary for 8% of all samples regarding a single and 9% regarding two variables. We replenished more than 1% of data for PO_4 (1.03%), SO_4 (1.3%), Si (2.5%) and DOC (11.4%). In total, data from 4061 samples could be used. In a next step the data were z-normalized (mean = 0, standard deviation = 1) to assign equal weight to the different variables.

Data pre-processing was performed using MS Access 2007. Statistical analyses and diagrams were generated with the R-software package (Version 2.6.1) and the Vegan package (Version 1.13-0) for Isometric Feature Mapping available at http://www.r-project.org (R Development Core Team, 2006). The data set was analyzed by applying PCA and the Isomap.

The basic principle for either approach can be characterized as follows: The data set forms a matrix X_{nm} with n observations of m variables. Thus, every instant is located in an m dimensional space. A distance matrix $D^{(x)}$ is build up by the Euclidean distances $d_x(i,j)$ between all datapairs i,j in the space X, following Eq. (1)

$$d_{ij} = ||x_i - x_j|| \qquad (1)$$

Both approaches aim at preserving as much of the variance in the m dimensional data space as possible in a low-dimensional projection. The projection is purely linear in case of the principal component analysis, whereas Isomap performs a piecewise linear approximation of a nonlinear manifold. In fact, principal component analysis can be regarded as a specific form of Isomap, and both approaches yield exactly the same results for a specific parameterization of the Isomap approach. The axes of the coordinate system of the projection are assumed to represent independent processes. Mapping single data points on these axes gives a quantitative assessment of the effect of that process.

4.1. Principal component analysis (PCA)

The PCA is a common mathematical approach for dimensionality reduction aiming to characterize a high dimensional dataset by few components (see, e.g. Legendre and Legendre, 1998) which are independent from each other. In a mathematical sense, PCA is an eigenvalue decomposition of the covariance matrix. Interpretation of components is based on loadings of the single variables on the components, that is, correlations with the respective component.

4.2. Isometric feature mapping (Isomap)

The Isomap approach, presented by Tenenbaum et al. (2000), can be regarded as a variant of the classical multidimensional scaling of the Euclidean distance matrix of the data. The Isomap approach consists of three steps (Tenenbaum et al., 2000):

(1) Constructing a neighborhood graph for a defined number of nearest neighbors (k) in $D^{(x)}$ using Euclidean distances;

(2) Setting up a distance matrix $D^{(G)}$ by calculating the shortest path between the pairs of data points using the sum of the smallest interpoint distances defined in Eq. (1). The smallest interpoint distances can be calculated using Eq. (2)

$$g_{ij} = \sum_{k=1}^{m} \frac{\partial X_k}{\partial \theta_i} \frac{\partial X_k}{\partial \theta_j} \qquad (2)$$

Θ is part of X defined by the chosen k. If the chosen k is equal to $n-1$, than $gij = dij$ and Eqs. (1) and (2) would lead to the same results. Calculating geodesic distances following Eq. (2) in Θ the distance matrix $D^{(G)}$ can be constructed.

(3) Performing a singular value decomposition of the distance matrix.

The Isomap approach performs a piecewise linear regression in a high dimensional data space without requiring any pre-defined mathematical structure. Thus, the Isomap approach enables to map even high nonlinear structures in the data set.

The parameter k has to be optimized in an iterative way. High k values will overestimate the low-dimensional manifold, whereas too low values will lead to a loss of useful information (Gámez et al., 2004). In contrast, nonlinear relationships can be mapped by using rather low k values. Only then the nonlinear Isomap approach will be clearly superior to linear methods (Tenenbaum et al., 2000; Gámez et al., 2004; Lischeid and Bittersohl, 2008).

In contrast to PCA, performance of the Isomap approach cannot be assessed using eigenvalues. Instead, a more universal measure is used and applied to PCA as well. It is equal to the squared Pearson correlation of the distance matrix of the 16-dimensional data set compared to the distance matrix of the low-dimensional projection, e.g., scores of the first three Isomap components. This measure will be referred to as "explained variance" in the following. Please note that it is related to but not equal to the eigenvalue based assessment commonly used for PCA.

The Isomap components can be interpreted analogously to the components of the PCA. However, as relationships might be nonlinear, the rank-based Spearman ρ correlation coefficient will be used instead of the linear Pearson correlation coefficient. In addition, boxplots will support component interpretations. The boxes show the 25- and 75-percentile of the analyzed data. Whiskers denote the range, as long as it is within the 1.5 interquartile range from the box, or the 1.5 interquartile range otherwise. The component scores are interpreted as measures of the effects of single processes.

4.3. Loadings of residuals

Both for PCA and Isomap the components are numbered in decreasing order according to their contribution to explaining the total variance of the data. Consequently nonlinear relationships might be very difficult to detect for components of higher number that exhibit only weak correlations with single variables. Thus, in this study loadings are analyzed based on using residuals of the respective variables. For all components except for the first one, residuals of linear regression of the variables with all preceding components are related to the respective component. Linear regression does not grasp nonlinear relationships that the Isomap approach could have identified. However, linear regression is much more stable compared to various nonlinear regression approaches that had been tested. As long as the nonlinear relationships are close to monotonic relationships, and a nonlinear rank-based measure of correlations like Spearman ρ is used, the resulting error is assumed to be negligible.

5. Results and discussion

Table 2 shows the site and depth specific means. Most variables were negatively skewed except Cl and pH, whereas normality was rarely found. Site and depth specific differences are given in Table 2. E.g., the Köhlerloh site exhibits very low NO_3, SO_4 and Al concentrations, attended by enhanced Na concentration and pH in comparison to the other sites. Highest DOC concentrations were measured at the uppermost depth at Coulissenhieb and at Köhlerloh. Trends in soil solution for different variables are known (e.g. Matzner et al., 2004a) such as a decrease of ionic strength in soil solution

Table 2
Depth and site specific means of variables with less than 25% of measurements below detection limit. C = Coulissenhieb, K = Köhlerloh, G = Gemös, W = Weidenbrunnen, b.d.l. = below detection limit.

Site	Depth	pH	Al	Ba	Ca	K	Mg	Mn	Na	Si	Cl	NO$_3$	SO$_4$	DOC
	[cm]	CaCl$_2$	[mg/l]											
C	20	3.63	3.10	0.04	2.53	1.52	0.47	0.09	1.62	5.97	1.99	18.9	15.0	24.5
	35	3.88	5.43	0.05	2.60	1.12	0.58	0.11	1.79	7.45	2.25	21.6	19.9	19.5
	90	4.13	5.55	0.05	1.96	1.13	0.48	0.15	1.89	5.45	2.11	19.3	23.3	4.8
K	50	4.93	0.39	b.d.l.	0.70	0.48	0.11	b.d.l.	3.36	7.73	3.05	b.d.l.	1.4	14.7
	100	5.25	0.16	b.d.l.	1.05	0.49	0.27	b.d.l.	3.82	8.19	2.34	b.d.l.	2.2	6.7
G	50	4.25	5.26	0.02	0.47	0.98	0.32	0.05	1.61	6.03	1.93	1.9	29.2	6.1
	100	4.22	6.44	0.02	0.55	1.58	0.31	0.15	1.55	3.93	2.68	6.5	32.3	4.5
	150	4.28	5.00	0.03	0.37	1.10	0.21	0.12	1.65	2.96	1.68	2.6	28.7	2.4
	200	4.22	5.02	0.04	0.57	1.01	0.34	0.11	1.93	4.32	1.87	1.6	30.3	4.6
W	50	4.13	10.27	0.14	1.15	3.82	0.53	0.24	1.12	4.73	2.08	60.9	14.8	4.8
	100	4.12	9.66	0.08	1.26	1.78	0.54	0.34	1.51	3.47	2.41	52.8	16.8	3.5
	150	4.05	11.08	0.06	0.99	1.41	0.47	0.24	1.36	3.41	2.88	53.3	20.0	3.5
	200	4.07	9.95	0.06	0.97	1.57	0.42	0.20	1.44	2.55	3.29	49.3	18.9	5.9

with time. Based on this heterogenic dataset PCA and Isomap were performed to identify the processes causing or influencing these differences.

The performance of Isomap depends on the connectivity of the underlying graph, which is equal to the number of links to nearest neighbor's k. Isomap has been run with different k values (500, 1000, 2000, 2500, 3000, 3500, 4000) and yielded best results for $k = 3000$. Fig. 2 shows the slight superiority of Isomap ($k = 3000$) in comparison to PCA and Isomap ($k = 500$). The first five components explained more than 94% of the total variance.

Isomap was slightly superior to the linear PCA with respect to explained variance. The Isomap dimensions were very similar to the PCA components. In fact, the Isomap components are easier to identify, to interpret and to assign to processes in comparison to the linear PCA components. Thus we have focused our discussion on the Isomap results. Further figures and analyses are based on the Isomap ($k = 3000$) variant.

Similarly, Lischeid and Bittersohl (2008) found that Isomap was slightly superior to PCA when investigating stream and groundwater quality data from the Lehstenbach catchment. In other studies differing results were found. Isomap clearly outperformed linear approaches in various studies (Tenenbaum et al., 2000; Gámez et al., 2004; Mahecha et al., 2007). These differences can be explained by varying degrees of nonlinearity in the respective data sets. Determining the different components via Isomap was the first step. In the second step these components were assigned to biogeochemical processes.

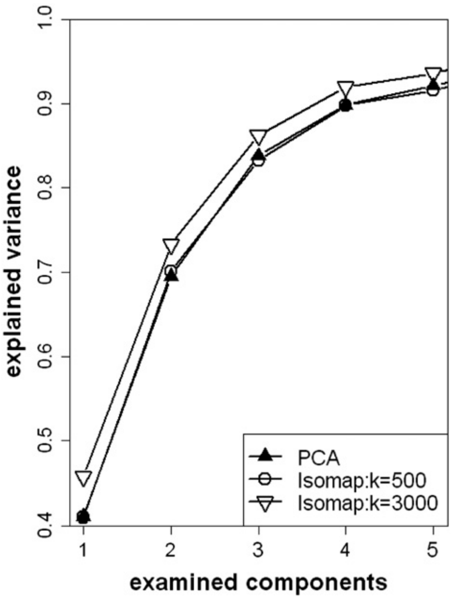

Fig. 2. Comparison of the explained variance by PCA and Isomap.

5.1. Component 1: Deposition

The first component explains almost 46% of the variance of the data set (Fig. 2). Very high positive loadings (>0.7) were found for Al, Ba, Mg, Mn and NO$_3$. Loadings between 0.4 and 0.7 were calculated for Ca, K, Cl and SO$_4$ (Fig. 3).

Component means decreased in the long-term at Coulissenhieb (Fig. 4) (1993–2000) and at Gemös (1996–2000) at all depths but to different degrees. In contrast, there was no clear trend at Weidenbrunnen and Köhlerloh. The results of Coulissenhieb and Gemös correspond to the well documented decrease of atmospheric deposition (Marschner et al., 1998; Oulehle et al., 2006; Schmid, 2008; Wu et al., 2010). Nitrate, SO$_4$ and Cl are mostly due to atmospheric deposition. Furthermore most of the Mg found in the topsoil originates from deposition (Weyer et al., 2008). Moreover, Al and Mn concentration in susceptible soils are known to be enhanced by atmospheric deposition. An obvious decrease of K, Mg, Ca, SO$_4$ and H$^+$ concentration was observed in the Lehstenbach catchment in bulk deposition and throughfall from 1987 to 2001 (Matzner et al., 2004b). A corresponding decline of ionic strength in acidic forest mineral soil horizon solution is consistent with previous findings (Marschner et al., 1998; Graf Pannatier et al., 2005; Wu et al., 2010). Thus, the first component seems to reflect the reaction of soil solution chemistry to reduced atmospheric deposition of the last decades.

Dissolved organic carbon is not correlated with the first component. This is contrary to the assumption that the decreased deposition of sulfur would increase DOC mobility because of increasing charge on hydrophobic organic colloids. On the other hand there is an enhanced competition between DOC and SO$_4$ for adsorption sites with decreasing sulfur deposition (Wu et al., 2010).

Beside temporal trends the decrease of deposition input is reflected by a depth gradient. The uppermost soil layers sampled at Coulissenhieb and Gemös, that is, 20 cm at Coulissenhieb and 50 cm at Gemös, exhibit lower component scores compared to the deeper soil layers (Fig. 5). Similar depth gradients were described for the terrestrial soils of the Lehstenbach catchment by Matzner et al. (2001). Obviously, the effect of decreasing deposition is the most pronounced in the upper soil layers. Such a depth gradient as reaction of changing deposition was described for different forest soils (e.g. Marschner et al., 1998; Oulehle et al., 2006).

At Weidenbrunnen no corresponding depth gradient was found. This could be due to reduced input of deposition at Weidenbrunnen in

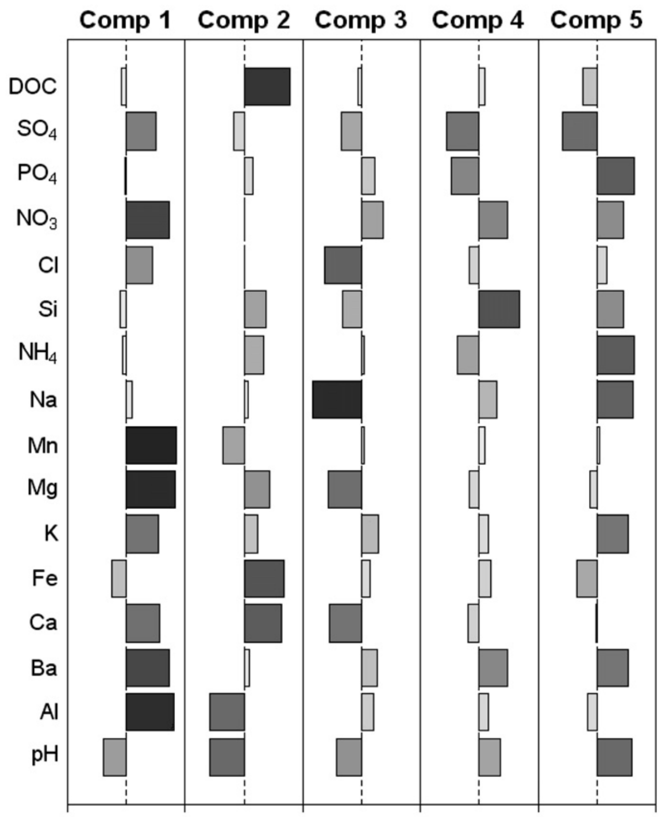

Fig. 3. Parameter loadings (Spearman ρ). Correlations of components 2–5 are calculated for residuals. *X*-axis is scaled between [−1;1] for every component. Grey shading is proportional to Spearman ρ.

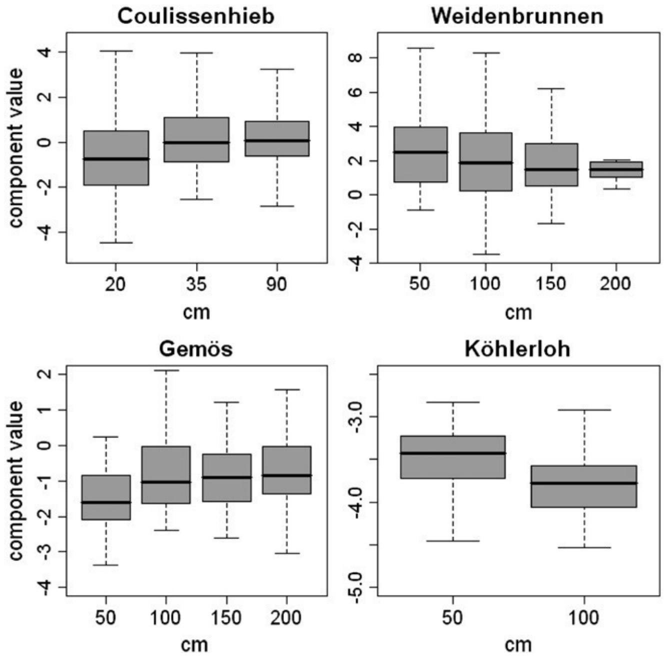

Fig. 5. Scores of the first component for different sites.

comparison to Coulissenhieb and Gemös because of lower stand age (Table 1). Tree height at Weidenbrunnen is lower compared to the surrounding stands supporting lower deposition influence.

The Histosol at the Köhlerloh site exhibits very low component values and even lower component scores at greater depth (Fig. 5). Nitrate and SO_4 concentrations at Köhlerloh are very low in comparison to the other sites (Table 2). Here, NO_3 does not reach the detection limit in nearly 90% of the samples, whereas Cl exhibits similar values compared to the other sites. Due to high groundwater level at the Köhlerloh site, anoxic conditions prevail (Lischeid et al., 2007). The low concentrations of NO_3 and SO_4 are ascribed to denitrification and desulfurication in anaerobic conditions in the Histosol (Matzner et al., 2001; Küsel and Alewell, 2004; Lischeid et al., 2007) causing considerably lower ionic strength and rather high pH values caused by proton consumption (Yu et al., 2007) and the release of Fe^{2+} by reduction of soil immanent iron oxides (Küsel and Alewell, 2004). Correspondingly, pH and Fe exhibit slight negative loadings.

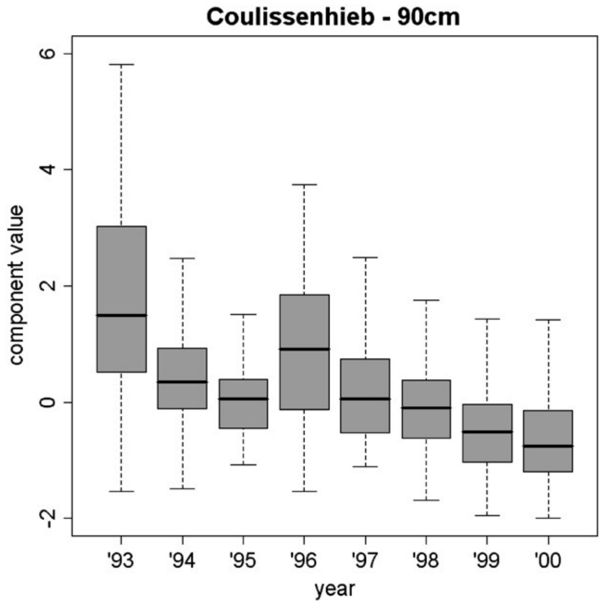

Fig. 4. Scores of the first component at Coulissenhieb. The black bar represents the mean. Upper and lower whiskers extend to the most extreme data points within the 1.5 interquartile range from the mean.

The differences of soil solution ionic strength between the terrestrial soils might be additionally conditional on general varieties in soil properties, soil moisture, variable amounts of litterfall and uptake by roots. The impact of soil moisture on ionic strength in soils was emphasized in numerous studies (e.g. van Hess et al., 2000; Dyer et al., 2008; Rennert and Rinklebe, 2010). Manderscheid and Matzner (1995) ascribed the spatial variation of soil solution chemistry at a small scale mainly due to patterns in throughfall chemistry depending on distance from the stems at the Coulissenhieb site. Larger scale spatial heterogeneities in throughfall fluxes in the Lehstenbach catchment caused by stand age, exposition and elevation are documented but low (Matzner et al., 2001). Differences in root uptake or throughfall may support the span of component values and the variable specific variance.

5.2. Component 2: Acidification

The second component explains 28% of the variance. The first two components explain more than two third of the variance of the data set. The second component is dominated by positive loadings of Fe and DOC and a negative loading of pH (Fig. 3). The combination of Fe and DOC contrary to pH is a typical signature for podzolization processes (Lundström et al., 2000a). Likewise Al is pronounced with a correlation of −0.5 and Ca with 0.64. Highest component scores were found at 20 cm depth, and slightly lower scores at 35 cm at Coulissenhieb. In contrast, component scores are much lower and usually below 0 at greater depth at all sites (Fig. 6). We allocate this second component to soil acidification and associated impacts.

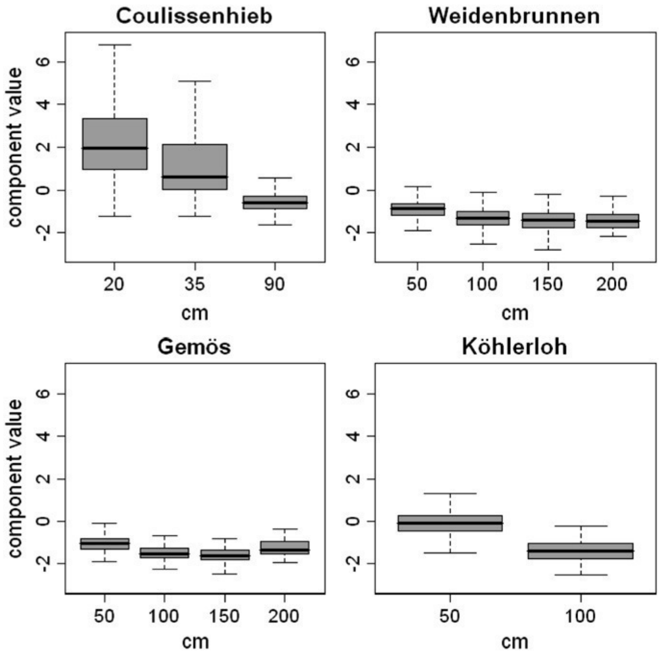

Fig. 6. Scores of the second component for different sites.

Acidification of sandy soils under forest is frequently linked with podzolation. The suction cups at 20 cm and 35 cm depth are installed in the Bs-horizon and in the transition zone between the Bs- and Bw-horizon of the Haplic Podzol at Coulissenhieb. The location in the accumulation horizons of the Podzol is reflected by the strong positive correlations with DOC and Fe opposed to pH. Thus, positive component scores are mainly represented in data of the upper soil horizons at Coulissenhieb with high DOC concentrations and low pH (Table 2), supporting the interpretation as acid induced podzolation.

The slightly positive correlation of Si with the second component scores may be partly ascribed to enhanced mineral weathering because of e.g. higher mycorrhizal activity and lower pH in upper mineral soil horizons (Lundström et al., 2000b). Differences of base cation correlations could be explained by their varying association to organic material. Contrary to Na, the cations Ca, K and Mg are primarily bound to humic compounds in soil solution (Gustafsson et al., 2000) explaining the different strength of correlation. Certainly the negative correlation of Al (in opposite to Fe and DOC) is contrary to expectations during podzolization. On the other hand, Al is rather ubiquitous in these acidic soils. Thus no clear loading emerged for this component.

5.3. Component 3: Changing deposition chemistry

Nearly 13% of variance of the data set is explained by this component. Calcium, Mg, Na, Cl, SO_4 and pH correlates negatively with the third component in opposite to the positive loading for NO_3 (Fig. 3). Component scores show similar long-term trends and depth gradients compared to the first component (Fig. 7).

A substantial long-term decrease of sulfur and corresponding cation (e.g. Ca and Na) deposition but ongoing high nitrogen deposition have been observed at many sites in Europe (e.g. Marschner et al., 1998; Alewell et al., 2000; Oulehle et al., 2006; Schmid, 2008). The same holds true for the Lehstenbach catchment. During the observation period we can constitute a decrease in deposition of Ca and especially SO_4, while NO_3 and K do not clearly decrease in the Lehstenbach catchment (Fig. 8). Additionally a decline for Cl and Mg was observed (Matzner et al., 2004b), which corresponds to negative loadings of the third component.

Thus, the third component seems to reflect the effects of a shift from sulfur dominated to a more nitrogen dominated deposition on soil solution.

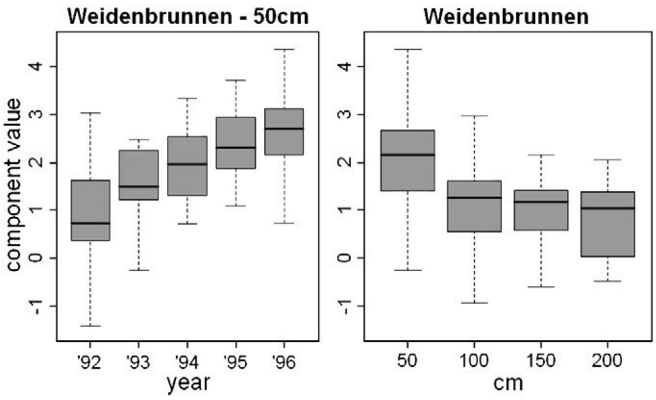

Fig. 7. Scores of the third component.

Decrease of SO_4 concentration in soil solution could enhance remobilization of adsorbed ions like NO_3 (Matzner et al., 2001). Graf Pannatier et al. (2005) describe a relative enrichment of Al in soil solution because of decreasing base cation input according to our findings, showing slight positive Al values. In addition, negative loading of pH and positive loading of Fe might be related to decreasing base cation input via deposition (Meesenburg et al., 1995).

The long-term increase of component values is in line with the changing deposition chemistry that was observed at 50 cm depth at Weidenbrunnen, indicating decreasing sulfate dominance in soil solution. Correspondingly, the observed depth gradients are consistent with a long-term shift in deposition chemistry.

5.4. Component 4: Matrix–solution interactions

The fourth component explains almost 6% of the variance of the data set and revealed positive correlations with pH, Ba, Na, NO_3, Si, and negative correlations with NH_4, SO_4 and PO_4 (Fig. 3). Loadings of Si and Na give hints on feldspar weathering (e.g. Lischeid and Bittersohl, 2008). The kinetics of feldspar weathering is very slow. Thus soil water residence time is likely to have an effect on its imprint on solute concentration. Soil water content and thus residence time depends on different impacts, e.g. soil matrix effects, plant growth period, precipitation and temperature (Robinson et al., 2008).

The positive loadings of mostly mineral-borne solutes like Si, Ba, Al and Fe can be interpreted as the slow occurring processes of weathering indicating longer residence time in soil (Velbel, 1993; Lischeid et al., 2002; Lischeid and Bittersohl, 2008; Vestin et al., 2008). The local bedrock, mainly feldspar, is naturally rich in Na and poor in Mg, which can cause the positive Na and opposing Mg loading. Weathering could also explain the positive loading of pH. Infiltrating rain water decreases concentrations of different weathering products like Si because of dilution effects (Vestin et al., 2008). For this reason

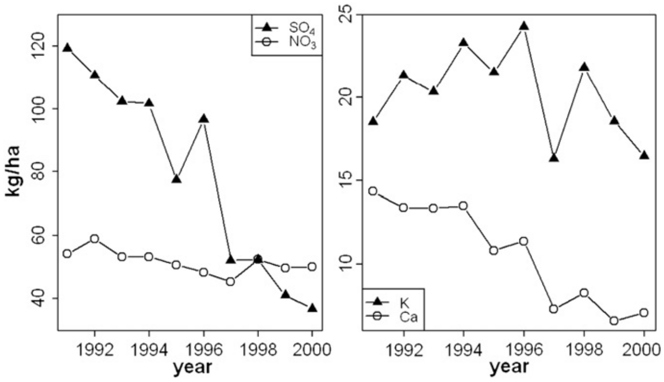

Fig. 8. Deposition in Lehstenbach catchment during the observation period.

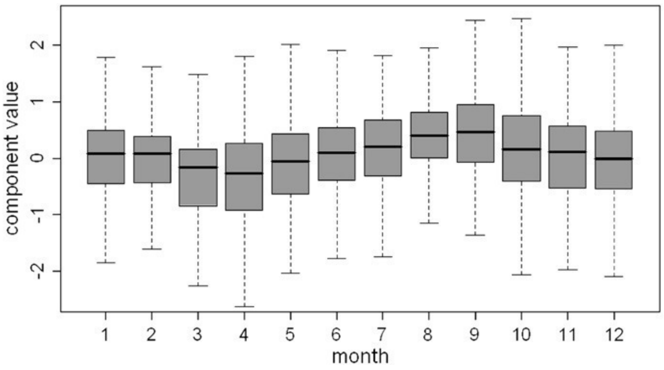

Fig. 9. Scores of the fourth component (all sites and depths).

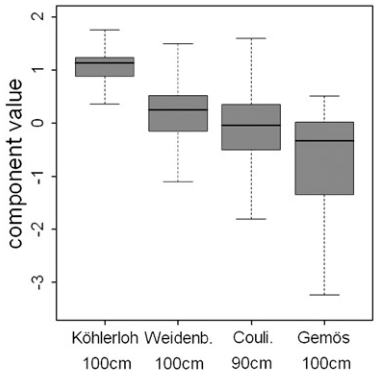

Fig. 10. Scores of the fourth component for different sites.

lower component values would be explainable by shorter water residence time. Most of the variables with negative loadings (e.g. NH_4, PO_4 and SO_4) are known to be quickly adsorbed to the soil matrix. Thus enhanced concentrations are only observed if water of lower residence time was sampled.

Fig. 9 shows the seasonal characteristic of the fourth component. The seasonality is comparable to that of air temperature, presumably caused by lower soil moisture during periods of higher evapotranspiration. The variance of component values in single months is wide, which possibly reflects different number and intensity of rain events in different years. Samples at 20 and 35 cm depth exhibit the strongest variance, probably as a result of stronger effects of temperature and rainfall in comparison to deeper soil horizons. Low component scores in deeper horizons may be explained by low residence time in consequence of fast seepage, which was described by Bogner et al. (2008) for the Coulissenhieb site where precipitation mainly follows preferential flow pathways in the upper soil horizons down to 40 cm depth, whereas matrix flow dominates below.

The site differences correspond to those of the first component for the terrestrial soils. The highest component means of terrestrial soils were found at Weidenbrunnen (Fig. 10). An explanation might be different minerals of Gemös and the Weidenbrunnen and Coulissenhieb sites. The Köhlerloh site exhibits the highest component scores. Suction cups in 50 and 100 cm depth actually sample laterally flowing groundwater with long residence time, suiting to higher Si concentrations (Table 2). The depth gradient may be related to the dilution effect of precipitation or lower residence time in the upper horizons, leading to lower component scores.

5.5. Component 5: Decomposition

The fifth component is positively correlated with pH, Ba, K, Na, NH_4, PO_4, Si and NO_3, and negatively with SO_4 and Fe (Fig. 3). This component explains about 2% of the dataset's variance.

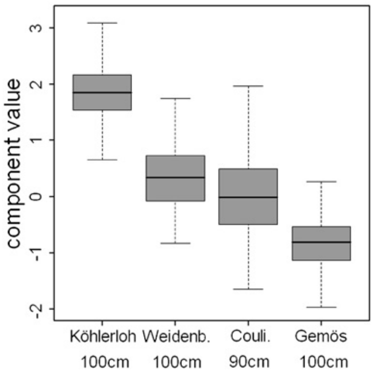

Fig. 11. Scores of the fifth component for different sites.

We ascribe positive correlations of Na, K, PO_4 and nitrogen compounds to solute release during decomposition of soil organic matter. Low or negative correlations point to nutrient consumption, e.g., by root uptake. The fifth component is interpreted to reflect the effect of net decomposition, that is, solute release by decomposition minus nutrient uptake by plants and microorganisms. Additionally sorption of decomposition products has to be considered.

The component shows substantial differences between the sites (Fig. 11). Outliers with very low values nearly exclusively occur at Gemös and Coulissenhieb (not shown). While the mature stands at Coulissenhieb and Gemös show negative component means, Weidenbrunnen and Köhlerloh have positive component means. Influence of tree age and crown density on decomposition (Moore, 1986) may explain the deviating characteristic of the fifth component. Cerli et al. (2006) have reported of accelerated mineralization of organic matter in younger stands. The higher the plant age the more carbon and nutrient accumulation increases and the more organic acids can be mobilized, resulting in decreasing pH values. That corresponds to our findings of positive pH and negative DOC loading. To mobilize nutrients in the soil, plants release different organic acids, and decrease the pH (Vestin et al., 2008; Du Laing et al., 2009), documented by opposing correlation of DOC to pH and base cations. Those organic acids may build up complexes with Fe, resulting in similar negative loadings of DOC and Fe.

Different nutrients like K, Mg and P are preferentially taken up by plant roots while others, like Al and Fe are excluded (Lundström et al., 2000b). The latters show negative loadings according to plant uptake effects. Such an influence of uptake on base cations and other nutrients on concentrations in soil solution is well documented (e.g. Johnson-Maynard et al., 2005; Vestin et al., 2008). Vestin et al. (2008) have reported increasing concentrations of DOC, SO_4 and Al in soil solution during the growing season, corresponding to the characteristics of our fifth component.

Callesen et al. (2007) measured high amounts of leachable NO_3 following soil-freezing events. They suppose, that an increased mortality of roots and organisms and thus reduced nutrient uptake supports the leaching. Lower component values dominate from December to February, and higher scores from March to August at Gemös and Weidenbrunnen (not shown). Beneath freezing/thawing cycles drying/wetting phases can have similar effects (Borken and Matzner, 2009). Those phases do not occur every year, which might explain why this component does not exhibit any clear seasonality like the fourth component. Furthermore, the available data set did not include humus layer or E-horizon samples where these patterns presumably are much more pronounced.

5.6. Multivariate process-solutes relationships

Variable loadings shown in Fig. 3 offer the possibility to identify processes which dominate the concentration dynamics of single variables. Few variables show high loadings in only one component.

Thereby this component is a crucial one for the variable concentration dynamics in the dataset. Other variables are obviously prone to a variety of different processes. For example acidic deposition seems not to be the only process influencing the SO_4 concentrations and pH, as may be expected. Instead, both variables are affected by additional processes, e.g. matrix–solution interactions (loadings of third component: SO_4 0.55; pH −0.36). For Mn the amount of deposition is the most crucial one. For the DOC-dynamics acidification turned out as the most important process, actually more important than decomposition. The nonlinear Isomap approach gave insight into the interplay between different processes and solutes. Thus, the effects of different processes on single solutes could be differentiated.

6. Conclusions

The results achieved with Isomap have slightly outperformed the results gained with PCA as a linear approach. The higher efficiency of Isomap which was detected in other ecological studies could be confirmed. Moreover, Isomap is able to discriminate between differences caused by temporally and spatially varying impact factors. Thus, the nonlinear Isomap approach can be deemed to be applicable to analyze large heterogeneous data sets as generated e.g. by long-term soil monitoring. Isomap is a promising alternative to linear approaches if explorative statistical analyses are accomplished and nonlinearity cannot be excluded. Therefore the higher amount of work is justified. In our study, five components explained more than 94% of the variance of the data set. Deposition is the main driver influencing the soil solution chemistry at the studied sites. With 46% of the first and 13% of the third component the impact of changing deposition explains almost 60% of the data set's variance. Factors, like redox processes at Köhlerloh, modify the strength of the influence of deposition. Other processes like mineral weathering and decomposition of organic matter could be identified and assigned to different components. Isotope techniques should allow discriminating between the origin of variables such as deposition, geogenic origin, and decomposition. Moreover these techniques might be a promising way to verify the interpretations in future.

Using the Isomap approach offers insights in the prevailing processes; additionally the relevant variables can be identified. Those advantages of Isomap which were also detected in other ecological studies could be confirmed. Isomap offers a high potential in the future and should be used in soil science when investigating large datasets.

Acknowledgments

The non-linear statistical analyses were conducted during a project funded by the German Federal Environmental Agency (Umweltbundesamt, UBA) (grant no. FKZ 3707 71 203). Part of the data was provided by the former Federal Bavarian Water Resources Agency which is highly appreciated. Field work and chemical analyses of the former Bayreuth Institute of Terrestrial Ecosystem Research (BITÖK) were funded by the Federal Ministry for Education and Research, or Federal Ministry for Education, Science, Research and Technology, respectively, (grants no. PT BEO 51-0339476 A-C). The authors thank Prof. E. Matzner (University of Bayreuth, Germany) for providing the dataset.

References

Alewell, C., Manderscheid, B., Gerstberger, P., Matzner, E., 2000. Effects of reduced atmospheric deposition on soil solution chemistry and elemental contents of spruce needles in NE-Bavaria, Germany. Journal of Plant Nutrition and Soil Science 163, 509–516.

Bogner, C., Wolf, B., Schlather, M., Huwe, B., 2008. Analysing flow patterns from dye tracer experiments in a forest soil using extreme value statistics. European Journal of Soil Science 59, 103–113.

Borken, W., Matzner, E., 2009. Reapprasial of drying and wetting effects on C and N mineralization and fluxes in soils. Global Change Biology 15, 808–824.

Callesen, I., Borken, W., Kalbitz, K., Matzner, E., 2007. Long-term development of nitrogen fluxes in a coniferous ecosystem: does soil freezing trigger nitrate leaching? Journal of Plant Nutrition and Soil Science 170, 189–196.

Cerli, C., Celi, L., Johansson, M.-B., Kögel-Knabner, I., Rosenquvist, L., Zanini, E., 2006. Soil organic matter changes in a spruce chronosequence on Swedish former agricultural soil: I. Carbon and Lignin Dynamics. Soil Science 171, 837–849.

Du Laing, G., Rinklebe, J., Vandecasteele, B., Meers, E., Tack, F.M.G., 2009. Trace metal behaviour in estuarine and riverine floodplain soils and sediments: a review. Science of the Total Environment 407, 3972–3985.

Dyer, C.L., Kopittke, P.M., Sheldon, A.R., Menzies, N.W., 2008. Influence of soil moisture content on soil solution composition. Soil Science Society of America Journal 72, 355–361.

Gámez, A.J., Zhou, C.S., Timmermann, A., Kurths, J., 2004. Nonlinear dimensionality reduction in climate data. Nonlinear Processes in Geophysics 11, 393–398.

Gerstberger, P., Foken, T., Kalbitz, K., 2004. The Lehstenbach and Steinkreuz catchments in NE Bavaria, Germany. Biogeochemistry of forested catchments in a changing environment — a German case study. In: Matzner, E. (Ed.), Ecological Studies, 172. Springer, Heidelberg, pp. 15–40.

Graf Pannatier, E., Luster, J., Zimmermann, S., Blaser, P., 2005. Acidification of soil solution in a chestnut forest stand in southern Switzerland: are there signs of recovery? Environmental Science & Technology 39, 7761–7767.

Gupta, A.K., Sinha, S., Basant, A., Singh, K.P., 2006. Multivariate analysis of selected metals in agricultural soil receiving UASB treated tannery effluent at Jajmau, Kanpur (India). Bulletin of Environmental Contamination and Toxicology 79, 577–582.

Gustafsson, J.P., van Hees, P., Starr, M., Karltun, E., Lundström, U.S., 2000. Partitioning of base cations and sulphate between solid and dissolved phases in three podzolised forest soils. Geoderma 94, 311–333.

IUSS/ISRIC/FAO, 2006. World Reference Base for Soil Resources 2006. FAO, Rome.

James, F., Culloch, C.M., 1990. Multivariate analysis in ecology and systematics: Panacea or Pandora's Box. Annual Review of Ecology and Systematics 21, 129–166.

Johnson-Maynard, J.L., Graham, R.C., Shouse, P.J., Quideau, S.A., 2005. Base cation and silicon biogeochemistry under pine and scrub oak monocultures: implications for weathering rates. Geoderma 126, 353–365.

Küsel, K., Alewell, C., 2004. Riparian zones in a forested catchment: hot spots for microbial reductive process. Biogeochemistry of forested catchments in a changing environment — a German case study. In: Matzner, E. (Ed.), Ecological Studies, 172. Springer, Heidelberg, pp. 377–398.

Langer, U., Rinklebe, J., 2009. Lipid biomarkers for assessment of microbial communities in floodplain soils of the Elbe River (Germany). Wetlands 29, 353–362.

Legendre, P., Legendre, L., 1998. Numerical Ecology. Elsevier, Amsterdam.

Lischeid, G., Bittersohl, J., 2008. Tracing biogeochemical processes in stream water and groundwater using nonlinear statistics. Journal of Hydrology 357, 11–28.

Lischeid, G., Kolb, A., Alewell, C., 2002. Apparent translatory flow in groundwater recharge and runoff generation. Journal of Hydrology 265, 195–211.

Lischeid, G., Kolb, A., Alewell, C., Paul, S., 2007. Impact of redox and transport processes in a riparian wetland on stream water quality in the Fichtelgebirge region, southern Germany. Hydrological Processes 21, 123–132.

Lundström, U.S., van Breemen, N., Bain, D., 2000a. The podzolization process. A review. Geoderma 94, 91–107.

Lundström, U.S., van Breemen, N., Bain, D., van Hess, P.A.W., Giesler, R., Gustafsson, J.P., Ilvesniemi, H., Karltun, E., Melkerud, P.-A., Olsson, M., Riise, G., Wahlberg, O., Bergelin, A., Bishop, K., Finlay, R., Jongmans, A.G., Magnusson, T., Mannerkoski, H., Nordgren, A., Nyberg, L., Starr, M., Tau Strand, L., 2000b. Advances in understanding the podzolization process resulting from a multidisciplinary study of three coniferous forest soils in the Nordic Countries. Geoderma 94, 335–353.

Mahecha, M.D., Martinez, A., Lischeid, G., Beck, E., 2007. Nonlinear dimensionality reduction: alternative ordination approaches for extraction and visualizing biodiversity patterns in tropical montane forest vegetation data. Ecological informatics 2, 138–149.

Manderscheid, M., Matzner, E., 1995. Spatial and temporal variation of soil solution chemistry and ion fluxes through the soil in a mature Norway Spruce (*Picea abies* (L.)Karst.) stand. Biogeochemistry 30, 99–114.

Marschner, B., Gensior, A., Fischer, U., 1998. Response of soil solution chemistry to recent declines in atmospheric deposition in two forest ecosystems in Berlin, Germany. Geoderma 83, 83–101.

Matzner, E., Alewell, C., Bittersohl, J., Lischeid, G., Kammerer, G., 2001. Biogeochemistry of a spruce forest catchment of the Fichtelgebirge in response to changing atmospheric deposition. In: Tenhunen, J.D. (Ed.), Ecosystem Approaches to Landscape Management in Central Europe. Springer, Berlin, pp. 463–503.

Matzner, E., Zuber, T., Lischeid, G., 2004a. Response of soil solution chemistry and solute fluxes to changing deposition rates. Biogeochemistry of forested catchments in a changing environment — a German case study. In: Matzner, E. (Ed.), Ecological Studies, 172. Springer, Heidelberg, pp. 339–360.

Matzner, E., Zuber, T., Alewell, C., Lischeid, G., Moritz, K., 2004b. Trends in deposition and canopy leaching of mineral elements as indicated by bulk deposition and throughfall measurements. Biogeochemistry of forested catchments in a changing environment — a German case study. : In: Matzner, E. (Ed.), Ecological Studies, 172. Springer, Heidelberg, pp. 233–250.

Meesenburg, H., Meiwes, K.J., Rademacher, P., 1995. Long term trends in atmospheric deposition and seepage output in northwest German forest ecosystems. Water, Air & Soil Pollution 85, 611–616.

Moore, A.M., 1986. Temperature and moisture dependence of decomposition rates of hardwood and coniferous leaf litter. Soil Biology & Biochemistry 18, 427–435.

Moritz, K., Bittersohl, J., Müller, F.-X., Krebs, M., 1994. Auswirkungen des sauren Regens und des Waldsterbens auf das Grundwasser. Bayerisches Landesamt für Wasserwirtschaft (Editor), München.

Oulehle, F., Hofmeister, J., Cudlin, P., Hruska, J., 2006. The effect of reduced atmospheric deposition on soil and soil solution chemistry at a site subjected to long-term acidification, Nacetin, Czech Republic. The Science of the Total Environment 370, 532–544.

R Development Core Team, 2006. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org.

Ramette, A., 2007. Multivariate analyses in microbial ecology. FEMS Microbiology Ecology 62, 142–160.

Rennert, T., Rinklebe, J., 2010. Release of Ni and Zn from contaminated floodplain soils under saturated flow conditions. Water, Air & Soil Pollution 205, 93–105.

Robinson, D.A., Campbell, C.S., Hopmans, J.W., Hornbuckle, B.K., Jones, S.B., Knight, R., Ogden, F., Selker, J., Wendroth, O., 2008. Soil moisture measurement for ecological and hydrological watershed-scale observatories: a review. Vadose Zone Journal 7, 358–389.

Schmid, H., 2008. Bodenversauerung in Bayern. Bayerisches Landesamt für Umwelt (Editor), Augsburg.

Tenenbaum, J.B., de Silva, V., Langford, J.C., 2000. A global geometric framework for nonlinear dimensionality reduction. Science 290, 2319–2323.

Van Hess, P.A.W., Lundström, U.S., Starr, M., Giesler, R., 2000. Factors influencing aluminium concentrations in soil solution from podzols. Geoderma 94, 289–310.

Velbel, M.A., 1993. Constancy of silicate-mineral weathering-rate ratios between natural and experimental weathering: implications for hydrologic control of differences in absolute rates. Chemical Geology 105, 89–99.

Vestin, J.L.K., Norström, S.H., Bylund, D., Mellander, P.-E., Lundström, U.S., 2008. Soil solution and stream water chemistry in a forested catchment I: dynamics. Geoderma 144, 256–270.

Weyer, C., Lischeid, G., Aquilina, L., Pierson-Wickmann, A.-C., Martin, C., 2008. Mineralogical sources of the buffer capacity in a granite catchment determined by strontium isotopes. Applied Geochemistry 23, 2888–2905.

Wu, Y., Clarke, N., Mulder, J., 2010. Dissolved organic carbon concentrations in throughfall and soil waters at level II monitoring plots in Norway: short- and long-term variations. Water, Air & Soil Pollution 205, 273–288.

Yu, K., Böhme, F., Rinklebe, J., Neue, H.-U., DeLaune, R.D., 2007. Major biogeochemical processes in soils — a microcosm incubation from reducing to oxidizing conditions. Soil Science Society of America Journal 71, 1406–1417.

Zhang, H.B., Luo, Y.M., Wong, M.H., Zhao, Q.G., Zhang, G.L., 2007. Concentrations and possible sources of polychlorinated biphenyls in the soils of Hong Kong. Geoderma 138, 244–251.