



HELSINKI UNIVERSITY OF TECHNOLOGY
Department of Engineering
Physics and Mathematics

Mikael Sulkava

Identifying spatial and temporal profiles from forest nutrition data

Master's thesis submitted in partial fulfillment of the requirements for the degree
of Master of Science in Technology

Espoo, 23rd May 2003

Supervisor: Professor Jaakko Hollmén

Instructor: Professor Jaakko Hollmén

Tekijä:	Mikael Sulkava
Osasto:	Teknillisen fysiikan ja matematiikan osasto
Pääaine:	Informaatiotekniikka
Sivuaine:	Lääketieteellinen tekniikka
Työn nimi:	Spatiaalisten ja ajallisten profiilien tunnistaminen metsän ravinnedatasta
Title in English:	Identifying spatial and temporal profiles from forest nutrition data
Professuurin koodi ja nimi:	T-122 Informaatiotekniikka
Työn valvoja:	Ma. Prof. Jaakko Hollmén
Työn ohjaaja:	Ma. Prof. Jaakko Hollmén
Tiivistelmä:	<p>Tämä diplomityö on tehty Teknillisen korkeakoulun Informaatiotekniikan laboratorion Intelligent Data Engineering -tutkimusryhmässä yhteistyössä Metsäntutkimuslaitoksen kanssa.</p> <p>Työssä tutkittiin erilaisten data-analyysimenetelmien avulla männyn- ja kuusenneulasten ravinnepitoisuuksia Suomessa ja Itävallassa vuosina 1987–2000. Tavoitteena oli analysoida ravinteiden maantieteellistä ja ajallista jakaumaa sekä yleisesti selvittää minkälainen mittausdatan sisäinen rakenne on. Kyseessä oli siis varsin selkeästi data-analyysiongelma. Käytettyjä analyysimenetelmiä olivat spatiaaliset tilastomenetelmät, itseorganisoivan kartan klusterointi sekä aikasarjamallitus.</p> <p>Työssä selvisi, että spatiaalinen tilastollinen tunnusluku semivarianssi on kohtalaisen käyttökelpoinen, muttei kuitenkaan ällistytävän hyvä mitta paikallisten ravinnepitoisuuksiin vaikuttavien tekijöiden tutkimiseen. Lisäksi havaittiin, että mittaustuloksia interpoloimalla voidaan tuottaa miellyttävän näköisiä sekä datan maantieteellisen rakenteen hahmottamista helpottavia kuvia.</p> <p>Itseorganisoivan kartan käyttöön perustuva klusterointimenetelmä tuotti uutta tietoa ravinnepitoisuuksien suhteista eri vuosina ja eri paikoissa. Sen avulla mittaustulokset voitiin jakaa kuuteen ryhmään, joissa kussakin neulasten kasvu ja ravinteiden määrät ovat omanlaisiansa. Eri ryhmät kuvasivat siis erilaisia metsän kasvuolosuhteita. Menetelmän antamien tulosten perusteella metsäasiantuntijat pystyivät rakentamaan Suomen metsien kehitystä kuvaavan mallin.</p> <p>Aikasarja-analyysissä käytetyn kätketyn Markov-mallin avulla saatu kuvaus mittausten ajallisesta rakenteesta ei ollut yhtä informatiivinen kuin muiden tutkittujen menetelmien tulokset. Mallin avulla saatu datan luokittelu kahteen ryhmään ei antanut erityisen kiinnostavaa tietoa eri mittausten välisistä yhteyksistä.</p>
Sivumäärä:	76
Avainsanat:	Itseorganisoiva kartta, klusterointi, kätketty Markov-malli, spatiaaliset tilastomenetelmät, metsä
Täytetään osastolla	
Hyväksytty:	Kirjasto:

Author:	Mikael Sulkava
Department:	Department of engineering physics and mathematics
Major subject:	Computer and information science
Minor subject:	Biomedical engineering
Title:	Identifying spatial and temporal profiles from forest nutrition data
Title in Finnish:	Spatiaalisten ja ajallisten profiilien tunnistaminen metsän ravinnedatasta
Chair:	T-122 Computer and information science
Supervisor:	Prof. (pro tem) Jaakko Hollmén
Instructor:	Prof. (pro tem) Jaakko Hollmén
Abstract:	<p>This Master's thesis has been done with the Intelligent Data Engineering group at the Laboratory of Computer and Information Science of Helsinki University of Technology in cooperation with the Finnish Forest Research Institute.</p> <p>In this work, the nutrient concentrations of pine and spruce needles in Finland and Austria between 1987–2000 were studied using different data analysis methods. The aim was to analyze the spatial and temporal distribution of the nutrients and generally find out what kind of internal structure exists in the data. Thus, the work was rather clearly a data analysis problem. The analysis methods used in the work were spatial statistics, clustering of the self-organizing map and time series modeling.</p> <p>It was found that a spatial statistic, semivariance, is a reasonably usable, but not exceptionally good, measure for analyzing the local factors affecting the nutrient concentrations. It was also noted that by interpolating the measurements, figures can be drawn that are visually appealing and also make it easier to understand the geographical structure of the data.</p> <p>The clustering method of the self-organizing map provided new information about the relations of the nutrients between different years and locations. With the clustering method, we were able to divide the measurements into six groups. In each group the growth of the needles and the amounts of the nutrients were different and thus, different groups represented different kinds of growing conditions. Using the result of the clustering method, forest experts were able to construct a model that characterizes the development of the forests of Finland.</p> <p>The hidden Markov model used in the time series analysis did not yield much information about the temporal structure of the data compared to the other methods. The classification of the data into two groups did not give any especially interesting information about the connections between the measurements.</p>
Number of pages:	76
Keywords:	Self-organizing map, clustering, hidden Markov model, spatial statistics, forest
Department fills	
Approved:	Library code:

Preface

This work was carried out in the Intelligent Data Engineering group at the Laboratory of Computer and Information Science of Helsinki University of Technology. The thesis was done as a part of a cooperation project with the Finnish Forest Research Institute.

First of all, I want to thank my supervisor and instructor Jaakko Hollmén for his endless support and valuable advice on the thesis.

I wish to thank Sebastiaan Luyssaert from the Finnish Forest Research Institute for making it possible to work on this topic, the fruitful cooperation and good comments on the thesis as well as the interpretation of the results from a biological point of view.

I also wish to thank Olli Simula for making the necessary financial arrangements as well as commenting and proofreading the text.

I want to thank Johan Himberg and Pasi Lehtimäki for their valuable comments and advice on the text.

I would also like to thank my former instructor Juha Vesanto for his expertise and cooperation with the methodology of the thesis.

Last but not least, I want to thank my relatives and friends and all the people who have had an encouraging attitude towards my work and studying.

Espoo, 23rd May 2003

Mika Sulkava

Contents

1	Introduction	1
2	Foliar nutrient concentration data and weather measurements	3
2.1	Finland	3
2.2	Austria	5
3	Analysis and modeling of the data	7
3.1	Spatial statistics	8
3.1.1	Semivariogram	8
3.2	Clustering using the self-organizing map	10
3.2.1	The self-organizing map	10
3.2.2	The VS clustering algorithm	12
3.2.3	Model validation	16
3.3	Time series modeling	16
3.3.1	Hidden Markov model	17
3.3.2	Learning with the EM-algorithm	19
3.3.3	Analysis of the model	21
3.3.4	From the cluster model to a transition graph	24
4	Results of the analyses of nutrient concentration data	25
4.1	Foliage of Finland	25
4.1.1	Spatial statistics	25
4.1.2	Finding the nutrition profiles	29
4.1.3	Temporal modeling of foliar nutrient concentrations	38
4.2	Foliage of Austria	41

4.2.1	Spatial statistics	41
4.2.2	Finding the nutrition profiles	42
4.2.3	Temporal modeling of foliar nutrient concentrations	48
4.3	Discussion	50
4.3.1	Spatial statistics	50
4.3.2	Clustering using the self-organizing map	51
4.3.3	Time series modeling	53
5	Summary and conclusions	56
5.1	Future work	57
A	Tables	59

Chapter 1

Introduction

Living plants are capable of taking up substances from the environment and using them for the synthesis of their cellular components. These nutrients play an important role in the physiological and biochemical processes of forest ecosystems. Owing to the relations between the environment and the foliar mineral composition, chemical foliar analysis is a useful diagnostic and monitoring tool in forestry and environmental studies [1].

In this study, different data analysis methods including spatial statistics, clustering of the self-organizing map and time series analysis with the hidden Markov model are used to analyze the relations between environmental changes and mineral composition of tree foliage. The spatial and temporal distributions of different nutrients are studied in order to find the possible structures in the data. Measurements from pine and spruce forests in Finland and Austria between 1987 and 2000 are used as the data sets for the methods.

It is found that clustering of the self-organizing map is a useful tool in forest nutrition analysis. The clustering method is able to represent the structure of the relations of nutrient concentrations in a new informative way. The methods of spatial statistics are able to enlighten the geographical structure of the data. The interpretation of the results of the hidden Markov model does not provide much insight into the temporal relations of the data.

The structure of the thesis is as follows. The nutrient data used in the experiments is described in Chapter 2. The data analysis methods: spatial statistics, clustering and time series modeling are introduced in Chapter 3. The results of

the different methods with discussion are explained in Chapter 4. Some results in table format are in Appendix A. Summary and conclusions of the methods and results are in Chapter 5.

Chapter 2

Foliar nutrient concentration data and weather measurements

The data used in the experiments consisted of two parts: one data set from Finland and one from Austria. The measurements were made from needle samples collected from the conifer trees in forests of Finland and Austria. The measured variables were different nutrient concentrations and the weight of the needles. The measurements and analyses of the data described in this chapter were mostly carried out by the personnel of the Finnish Forest Research Institute.

Every year between October and November, the same personnel collected the needle samples from the trees on the stands. In each stand, a few trees were selected and current (C) and previous (C+1) years' twigs were pruned from the top third of the crown. Twigs with the same needle-year class were treated as a separate sample. The needles were dried and the needle mass (NM) of 1000 needles was weighted. To determine the foliar element concentrations, the dried needles were ground. For each stand, the elemental composition was determined from the needle powder using various chemical analysis methods [1].

2.1 Finland

In the data from Finland, there were 12 concentration measurements (see Table 2.1). In addition, the needle mass was reported as the weight of 1000 needles ($g/1000$). The measurements were made in 36 measurement stands scattered

around Finland in the years 1987–2000. There was, however, some missing data: altogether 29% of the measurements from C needles were missing (Table 2.1). In 16 stands, the main tree species was Norway spruce (*Picea abies*) and in the rest Scots pine (*Pinus sylvestris*).

The quality of the measurements was analyzed in inter-laboratory tests. It was found that the relative quality of the laboratories was good. The relative standard deviation (rsd), based on repeated measurements of 11 different reference samples ranged in the sampling period 1987–2000 for N between 0.7–1.8%, for S 1.5–5.1% and for P 1.5–4.1% [1].

In addition to the concentrations and needle mass, the geographical positions of the stands were known and there were some weather measurements available. There were numerous weather stations from which measurements were available. For each station, the average temperature of January, July and the whole year as well as the precipitation sum of the year were known. In addition to these, the normality of temperature and precipitation of the year in each station were calculated. The weather data of the nearest weather station was used for the measurement stands.

For each weather station i , the average temperature and precipitation value O were available on a monthly base ($m = 1, \dots, 12$) for the years $t = 1970, \dots, 2000$. The average monthly observation (temperature or precipitation) for station i in year t for month m can be written as $O_{i,t,m}$. The observations were normally distributed for all stations and all months. This was confirmed with the Kolmogorov-Smirnov test. The long-term (1970–2000) average $\mu(O_{i,m})$ and standard deviation $\sigma(O_{i,m})$ were known for all stations and months. The probability $P(O_{i,t,m})$ was calculated that a normally distributed random variable with mean and standard deviation equal to $\mu(O_{i,m})$ and $\sigma(O_{i,m})$ will get a value less than the observation $O_{i,t,m}$.

$$P(O_{i,t,m}) = P(O_{i,t,m} > x | X \sim N(\mu(O_{i,m}), \sigma(O_{i,m}))) \quad (2.1)$$

The variables describing the annual temperature and precipitation were calculated by averaging these monthly probabilities:

$$P_{i,t}(O) = \frac{1}{12} \left[P(O_{i,t-1,12}) + \sum_{m=1}^{11} P(O_{i,t,m}) \right] \quad (2.2)$$

Table 2.1: *The number of missing values of the measured variables in both C and C+1 needles from Finland and Austria. An “X” means that the measurement was not made at all.*

Measurement	Finland		Austria	
	C	C+1	C	C+1
Nitrogen N (<i>mg/g</i>)	137	137	11	629
Sulfur S (<i>mg/g</i>)	137	137	11	71
Phosphorus P (<i>mg/g</i>)	137	137	11	629
Calcium Ca (<i>mg/g</i>)	137	137	11	629
Magnesium Mg (<i>mg/g</i>)	137	137	221	629
Potassium K (<i>mg/g</i>)	137	137	11	629
Zinc Zn ($\mu\text{g/g}$)	137	137	221	629
Manganese Mn ($\mu\text{g/g}$)	137	137	11	629
Iron Fe ($\mu\text{g/g}$)	139	137	221	629
Copper Cu ($\mu\text{g/g}$)	138	141	X	X
Aluminum Al ($\mu\text{g/g}$)	137	137	X	X
Boron B ($\mu\text{g/g}$)	162	163	X	X
Needle mass NM (<i>g/1000</i>)	216	216	653	776
Temperature		0		X
Precipitation		0		X
Altitude (<i>m</i>)		X		0

2.2 Austria

There were needle mass and 9 concentration measurements available from Austria (Table 2.1). The number of measurement stands was 71, and the measurements were made for 11 years: 1989–1999. In total, 18% of the measurements from C needles were missing. The amount of missing data from the C+1 needles was higher (Table 2.1). The geographical positions and the altitudes of the stands were known. There was no weather data available from Austria. In 66 stands the main tree species was Norway spruce and in five stands Scots pine.

The quality of the measurements was analyzed between 1995–1999 with repeated measurements of 18 different reference samples. The relative standard

deviation ranged between 1.1–4.2% for N, 1.0–3.2% for S and 1.5–4.1% for P. The relative quality, compared with other laboratories was thus good. As a consequence, the measurement methods can contribute only a minor part of the total spatial and temporal variation of the data [1].

Chapter 3

Analysis and modeling of the data

The mineral composition of the forest nutrition data was analyzed with three different types of data analysis methods: spatial statistics, clustering and time series modeling.

The spatial distribution of the measurement values was analyzed with spatial statistics methods. Interpolation of the nutrient concentration values was used to get an overview of the distribution of the data. In addition, semivariograms of the measurement values were computed to further analyze the spatial correlations between in the data.

A clustering algorithm was used to divide the measurements of different stands in different years into clusters such that similar measurements belong to the same cluster. The clusters and their meaningfulness was studied and a temporal model was constructed that describes the switching of the stands between different clusters.

A time series model was trained for the measurements in order to better understand the possible changes in the tree stands in different years. The model assumes that each year, a stand belongs to one of a few discrete states. Each state describes different growing conditions. By analyzing the state sequences, information concerning the development of the forests can be obtained.

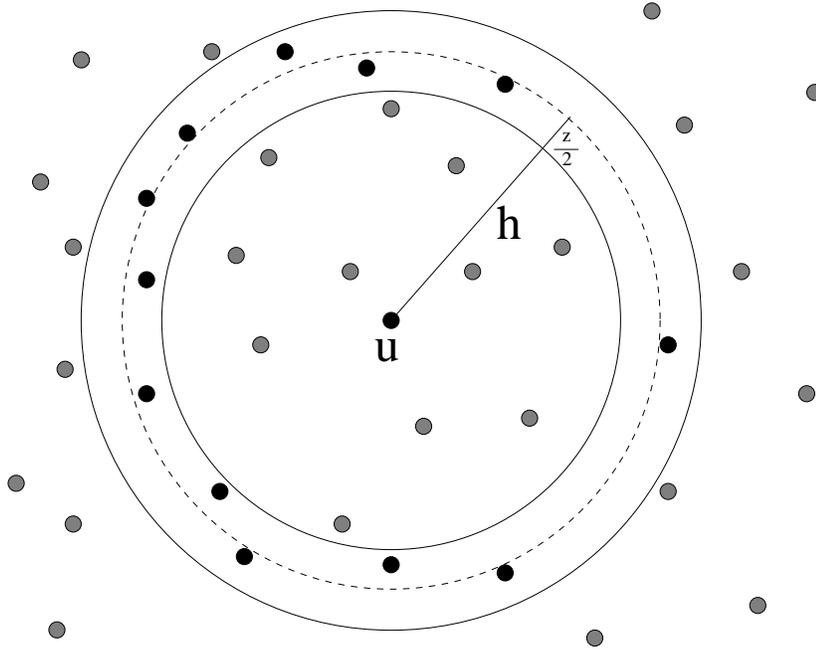


Figure 3.1: *The semivariance is calculated as the average squared difference between all the data pairs at a given lag distance interval h from each other. The black dots between the two solid circles are the stands whose measurement values are compared to the stand u_α in the center of the circles.*

3.1 Spatial statistics

Spatial statistics is statistical study of spatial patterns and processes. Spatial (point) patterns are simply patterns formed by some points, for example measurement stands, on the map. In this work, we are more interested in finding out how the nutrients are distributed. That way we could gain some insight on the spatial processes affecting the nutrition status of the forests. Different methods and models have been discussed in the literature for example in [2, 3, 4, 5].

3.1.1 Semivariogram

The experimental semivariogram $\gamma(h)$ measures the average dissimilarity between data separated by a lag distance h . It is computed as half the average squared

difference between the components of every data pair [6]:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [x(u_\alpha) - x(u_\alpha + h)]^2, \quad (3.1)$$

where $x(u_\alpha)$ is the measurement value at location u_α and $N(h)$ is the number of data pairs at a distance h from each other (see Figure 3.1). The semivariogram value at a given lag h is called the semivariance. The semivariogram is a measure of spatial variability over the full range of attribute values. Due to limited number of measurements, the distances at which the semivariance is calculated for the semivariogram are usually of form:

$$h = \left(n - \frac{1}{2}\right) z, \quad n \in \mathbb{N}, z > 0, \quad (3.2)$$

where z defines the width of the lag intervals and consequently the resolution of the semivariogram. Thus, instead of a single lag distance h , we actually use a lag distance interval $h \pm \frac{z}{2}$. When choosing the suitable value for z , one has to compromise between the resolution and noise of the semivariogram.

Indicator semivariogram can be used to characterize the spatial distribution of an indicator function $i(u_\alpha)$ that is defined to get the value 1, if a certain condition holds and 0 otherwise. In case of classified data, the indicator function is defined as follows:

$$i(u_\alpha, C) = \begin{cases} 1 & \text{if } u_\alpha \in C \\ 0 & \text{otherwise} \end{cases} \quad (3.3)$$

The indicator functions simply tells if stand u_α belongs to class C . Now, the indicator semivariogram is computed as

$$\gamma_I(h, C) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [i(u_\alpha, C) - i(u_\alpha + h, C)]^2. \quad (3.4)$$

The indicator variogram value $2\gamma_I(h, C)$ measures how often exactly one of two stands separated by lag h belongs to class C [6]. In other words, $2\gamma_I(h, C)$ estimates the correlation between belonging to class C and not belonging to it as a function of h .

3.2 Clustering using the self-organizing map

Clustering algorithms can be divided into two main categories: partitive and hierarchical algorithms [7]. Partitive algorithms divide the data set into non-overlapping partitions whereas hierarchical algorithms construct a hierarchy tree of the clusters. The structure of the hierarchy is such that all the data belongs to the top level (root) cluster and at the bottom level, each data vector forms a separate cluster.

Many kinds of clustering algorithms have been developed for different problems. Recently, new algorithms have emerged, for example, for biotechnology [8, 9], very large databases [10] and probabilistic models [11].

A clustering algorithm based on the self-organizing map (SOM) [12, 13] was chosen for this problem because of its good visualization properties. The SOM preserves neighborhood relations and presents high-dimensional datasets on a 2-dimensional grid and is therefore an important tool in data mining with its main applications in visualization and clustering. In forest research, the self-organizing map has been used for example in tree mortality prediction [14].

3.2.1 The self-organizing map

The self-organizing map consists of a low-dimensional (usually 2-D) regular grid of map units that are connected to adjacent ones by a neighborhood relation [15]. The grid can be effectively used to visualize and explore properties of the data [16]. Each map unit i is represented by a prototype vector, $\mathbf{m}_i = [m_{i1}, \dots, m_{id}]^T$, where d is the dimension of the input vectors \mathbf{x} . The training of the map consists of the following phases:

1. Initialization. The initial values of the prototype vectors \mathbf{m}_i are chosen either completely randomly, selecting input vectors \mathbf{x} randomly as prototype vectors or initializing the prototype vectors linearly along the greatest eigenvectors of the data.
2. Sampling. Select an input vector \mathbf{x} with a given probability.

3. Similarity testing. Find the winning unit $i(\mathbf{x})$ with the smallest distance:

$$i(\mathbf{x}) = \arg \min_j \|\mathbf{x} - \mathbf{m}_j\|, \quad j = 1, \dots, l \quad (3.5)$$

4. Updating. The prototype vectors are adjusted using the following equation:

$$\mathbf{m}_j(t+1) = \mathbf{m}_j(t) + \eta(t)h_{j,i(\mathbf{x})}(t)(\mathbf{x} - \mathbf{m}_j(t)) \quad (3.6)$$

where $\eta(t)$ is a parameter of learning rate and $h_{j,i(\mathbf{x})}(t)$ is a neighborhood function surrounding the winning unit $i(\mathbf{x})$. Both parameters decrease monotonically during the training. The update rule moves the prototype vectors of the winning unit and its neighbors towards the input vector.

5. Return to phase 2 if there are any significant changes in the map.

The neighborhood function must be symmetrical around its maximum on the winning unit. To ensure convergence, it has to decrease monotonically to zero as the distance from the map unit increases. A commonly used function is the Gaussian neighborhood function.

$$h_{j,i(\mathbf{x})} = e^{-\frac{d_{ji}^2}{2\sigma^2}}, \quad (3.7)$$

where d_{ji} is the Euclidean distance from the winning unit on the map grid. The width of the neighborhood function should decrease during training. Usually, parameter σ decreases linearly to 1. The learning parameter η is a decreasing function. An exponentially decreasing function is usually used.

$$\eta(t) = \eta_0 e^{-\frac{t}{\tau}}, \quad t = 0, 1, 2, \dots, \quad (3.8)$$

where τ is a time constant.

The training of the map is done in two phases. In the rough training phase, the prototype vectors are topologically ordered. The neighborhood function should be rather wide in the beginning of the training and get slowly narrower. In the fine tuning phase, the prototype vectors are slightly adjusted to find a more accurate representation of the input space. The width of the neighborhood function is small.

After the training, the prototype vectors define a tessellation of the input space into a set of Voronoi sets

$$V_i = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{m}_i\| < \|\mathbf{x} - \mathbf{m}_j\| \forall j \neq i\}, \quad (3.9)$$

where \mathbf{x} are the data vectors and $\|\cdot\|$ is the Euclidean norm. In effect, each data vector belongs to the Voronoi set of the prototype to which it is closest. The map unit with the closest prototype vector is called the best matching unit (BMU). In other words, SOM quantizes the training data set with a representative set of prototype vectors. The quantization process is regularized by the neighborhood relation such that topology of the data set is preserved. Thus, the algorithm can be thought as a kind of nonlinear regression. Without the neighborhood relation, the SOM algorithm reduces to the k-means algorithm [15].

There are numerous ways to measure the quality of the SOM [17, 18, 19, 20, 21, 22]. In this study, average quantization error and topographic error were used for the task because they are both simple and intuitive measures for assessing the quality of the SOM. The quality of quantization of the SOM can be measured with the average quantization error, which is simply the average distance from each data vector to the best matching unit. The preservation of the topology of the maps can be measured with the topographic error. It is the percentage of data vectors for which the best matching unit and the second best matching unit are not neighboring map units [23].

3.2.2 The VS clustering algorithm

The VS clustering algorithm used in this study has been co-developed by the author [24]. It is a two-level approach: first the SOM is trained and the data is partitioned into a large number of Voronoi sets, each corresponding to one map unit. Subsequently, the map units are clustered. All data vectors in a Voronoi set belong to the same cluster as the corresponding map unit. An advantage over traditional methods, like k-means, is that the result can be effectively visualized on the 2-dimensional grid. Another advantage is that by clustering the SOM rather than the data directly, significant gains in the speed of clustering can be obtained [16].

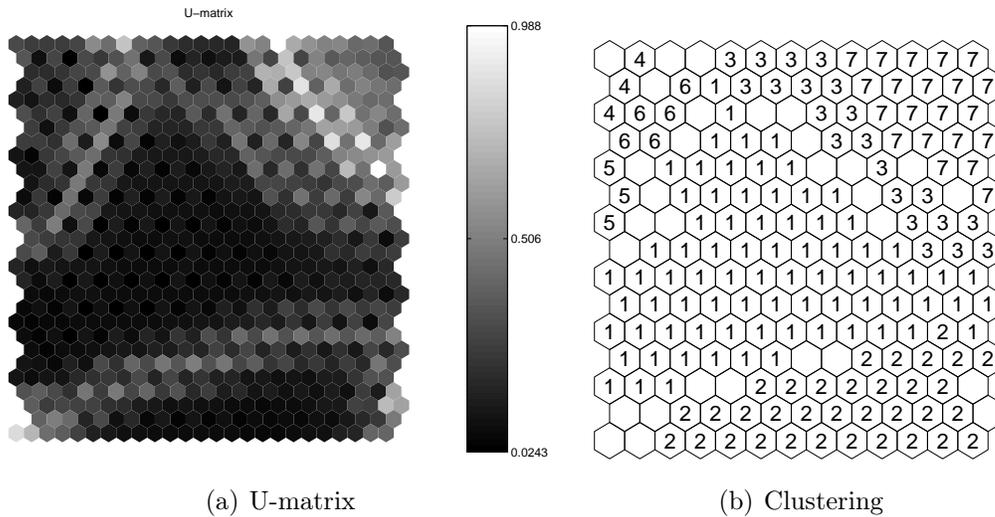


Figure 3.2: (a) An example of U-matrix showing the distances between neighboring map units of SOM. (b) The corresponding clustering result with 7 clusters. The numbers (1–7) indicate which cluster the map unit belongs to.

U-matrix is a commonly used tool to cluster the SOM visually [25]. It visualizes distances between each map unit and its neighbors. Unfortunately, when clusters are identified visually, the results obtained by different people are not necessarily the same. Therefore, an automated clustering algorithm that follows the results of the U-matrix (Figure 3.2a) was used. The details of the algorithm are presented in [24]. The basic idea of the algorithm is as follows:

1. The data is quantized with SOM and a distance matrix, showing the median distances between neighboring map units is calculated.
2. The map is divided into a set of base clusters. This is done using region growing with local minima of the distance matrix as seed points.
3. A cluster hierarchy is constructed from the base clusters using an agglomerative algorithm [26] and a pruning procedure.
4. The final partitioning with suitable number of clusters is obtained from the hierarchy (Figure 3.2b).

Region growing

The region-growing starts with setting the local minima of the distance matrix as seed points. These are the map units whose median distance to neighboring units is smaller than the median distance of any of the neighboring units to their neighbors. Next, the unassigned map unit with smallest distance to a cluster is found and assigned to the corresponding cluster. Here, a continuity constraint is used to ensure that the clusters form continuous areas on the map. Only those unassigned map units are considered for merging which are neighbors of map units that already belong to a cluster. Assigning the map units to clusters is continued until all map units belong to some cluster.

This procedure provides a partitioning of the map into a set of base clusters, number of which is equal to the number of local minima on the distance matrix. A problem is that the distance matrix may have some local minima which are products of random variations in the data rather than real local maxima of the probability density function of the data. Such base clusters are pruned out of the clustering in a hierarchical fashion.

Cluster hierarchy

In cluster analysis, constructing a cluster hierarchy is often beneficial [27, 28]. Apart from the need for pruning explained above, a cluster hierarchy may represent the true structure of the data better than a single-level partitioning. Some clusters can be considered super-clusters, consisting of several sub-clusters, which allows the data to be investigated at several levels of detail.

An agglomerative clustering algorithm was used to construct the cluster hierarchy from the base clusters. This, however, produces a binary tree which may not be representative of the true structure of the data. If in reality, a super-cluster consists of three (or more) sub-clusters, the binary tree will have one (or more) extra intermediate clusters. These are pruned out as follows:

1. Start from root (top level) cluster.
2. For the cluster c under investigation, generate different sub-cluster sets. A sub-cluster set may contain either sub-clusters of cluster c or sub-clusters of c 's sub-clusters (sub-sub-clusters).

3. Each sub-cluster set defines a partitioning of the data in the investigated cluster. Investigate each partitioning using a clustering validity measure, the gap-index I_{gap} .
4. Select the sub-cluster set with minimum I_{gap} , and prune the corresponding intermediate clusters.
5. Select an uninvestigated cluster (if any), and continue from step 2.

The clustering validity measure used in the pruning was a measure of the gap between the two clusters. It compares the sizes of the clusters with the distance between them.

$$I_{gap} = \frac{1}{C} \sum_{i=1}^C \max_j \left\{ \frac{S_i + S_j}{d_{ij}} \right\}, \quad \text{where} \quad (3.10)$$

$$S_i = E\{\|\mathbf{m}_k - \mathbf{m}_l\| \mid \mathbf{m}_k, \mathbf{m}_l \in C_i, k \in N_l, V_k, V_l \neq \emptyset\}, \quad (3.11)$$

$$d_{ij} = E\{a\|\mathbf{m}_k - \mathbf{m}_l\| \mid \mathbf{m}_k \in C_i, \mathbf{m}_l \in C_j, k \in N_l\} \quad (3.12)$$

Above, C is the number of clusters, E is the average, C_i is the set of prototype vectors which belong to cluster i , N_l is the set of neighboring map units of the map unit l and

$$a = \begin{cases} 2 & \text{iff } V_k = \emptyset \vee V_l = \emptyset \\ 1 & \text{otherwise.} \end{cases} \quad (3.13)$$

The coefficient a is used to reflect the fact that an empty map unit between two clusters does not really belong to either and thus, the distance between the clusters is approximately twice the distance from either cluster to the empty map unit.

This procedure gives a pruned cluster hierarchy together with validity measures of the clustering quality of the sub-cluster sets of each node in the tree. A particular partitioning is obtained from this tree by starting from the top with all data in a single cluster, and traversing the tree downwards by always splitting the intermediate node with best clustering validity index, until a predetermined number of clusters has been obtained. The selection criteria for the number of final clusters were the following: the sub-clusters were combined into a super-cluster if the difference between the means of the sub-clusters was smaller than the standard deviation of the whole data or if one of the sub-clusters consisted of very few (<5) map units.

3.2.3 Model validation

When analyzing and interpreting a model, many different hypotheses concerning the model are often proposed. These hypotheses should, however, be tested in order to be able to draw conclusions about the model. In case of measurement data and a theoretical model, empirical hypothesis testing can be done with, for example, permutation tests [29].

Permutation tests are useful tools in hypothesis testing. In case of a clustering problem, the possible correlation between the clusters and any measurement data can be analyzed with permutation tests. The principle is to compare a cluster to the rest of the data, and see if there is any difference when the cluster is replaced by a set of randomly chosen data points. That is, the test answers to the question: “Is there a certain difference between the data in the cluster and the rest of the data?” In this study, the test gives an estimate of the probability

$$P_{pt}(C) = P(E\{x_C\} - E\{x_{D \setminus C}\} < E\{x_R\} - E\{x_{D \setminus R}\}), \quad (3.14)$$

where E is the expectation value, x_C are some measurement values (for example NM) of vectors that belong to cluster C , D is the whole data set and R is a set of randomly selected points. In statistical significance testing, it is important to keep in mind what the test actually tells about the data and what it does not [30]. For example the above test (Equation 3.14) does not tell anything about inequality to the other direction.

3.3 Time series modeling

Time series are ordered measurements, i.e. the measurements are obtained at consecutive time steps. The data used in this study is time series data, but the time dimension is completely lost in the clustering method. Therefore, a time series model that preserves the temporal relations of the measurements was used to analyze the data.

Regime switching models are useful tools in time series analysis [31]. The main structure of the models is the following: the system generating the observations is at all times in some discrete state and the output of the system depends on the

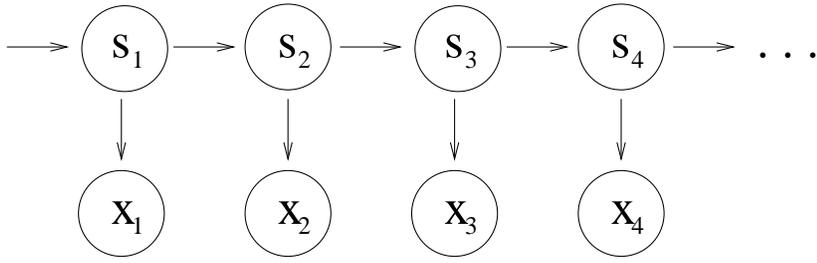


Figure 3.3: *The hidden Markov model. The hidden variable s_t changes its state in time with $P(s_t|s_{t-1})$. The observations depend on the hidden variable as $P(x_t|s_t)$.*

state. The early studies of switching regression are [32, 33] and more advanced models have been studied in [34, 35, 36, 37, 38, 39]. The models have been successfully used in, for example, economics [40, 41, 42, 43], but they have also been criticized [44]. An introduction for nonlinear time series modeling has been published in [45].

3.3.1 Hidden Markov model

The hidden Markov model (HMM) is a discrete-time model that belongs to the general framework of probabilistic independent networks [46]. The properties that make it suitable for our purpose are that it can handle missing data as well as time series of different lengths. The model assumes sequential data and has been successfully used, for example, in sequence processing [47], speech recognition [48, 49, 50] and fault detection [51]. The hidden Markov model assumes a hidden sequence of discrete states $S = \{s_0, \dots, s_T\}$ and an observation sequence $X = \{x_0, \dots, x_T\}$. At each time step t :

1. The process is assumed to be in some unobserved state s_t .
2. An observation x_t depending on the state is acquired.

The probability of being in a particular state at time t depends only on the state at previous time step $t - 1$. This is called the Markov property, i.e. given the previous state, the current state is conditionally independent of the whole history.

$$P(s_t|s_{t-1}, s_{t-2}, \dots, s_1) = P(s_t|s_{t-1}) \quad (3.15)$$

Also, given the current state, the current observation is conditionally independent of the whole history [52] (see Figure 3.3).

$$P(x_t | s_t, s_{t-1}, \dots, s_1) = P(x_t | s_t) \quad (3.16)$$

The state changes in time stochastically according to a $N \times N$ transition matrix, where N is the number of states

$$A = (a_{ij}) = P(s_t = j | s_{t-1} = i), \quad i, j = 1, \dots, N \quad (3.17)$$

with the following constraints:

$$a_{ij} \geq 0, \quad i, j = 1, \dots, N \quad (3.18)$$

$$\sum_{j=1}^N a_{ij} = 1, \quad i = 1, \dots, N \quad (3.19)$$

The observed variables of the data sequence can be either continuous or discrete. In our experiments, the observed values were continuous. The probability density of the observed variable depends on the state through an emission probability density function $P(x_t | s_t = j; \theta)$, parameterized by θ . The joint probability density parameterized by A and θ can be expressed as follows [47, 52]

$$P(X, S) = P(x_0, s_0) \prod_{t=1}^T P(s_t | s_{t-1}; A) \prod_{t=1}^T P(x_t | s_t; \theta), \quad (3.20)$$

where $P(x_0, s_0)$ is the prior of the state and data at $t = 0$.

In this work, the hidden state is the unknown combined effect of the environment (for example soil condition, weather, pollution, fungi, insect pests and tree history) on the foliar nutrient concentrations and needle mass. The observation sequence of each measurement stand i contains the observed nutrient concentrations of N, S and P and the needle mass for each year.

$$\mathbf{X}_i = \begin{pmatrix} c_{i,1}(N) & c_{i,2}(N) & \dots & c_{i,t}(N) & \dots & c_{i,T}(N) \\ c_{i,1}(S) & c_{i,2}(S) & \dots & c_{i,t}(S) & \dots & c_{i,T}(S) \\ c_{i,1}(P) & c_{i,2}(P) & \dots & c_{i,t}(P) & \dots & c_{i,T}(P) \\ NM_{i,1} & NM_{i,2} & \dots & NM_{i,t} & \dots & NM_{i,T} \end{pmatrix} \quad (3.21)$$

There are many variants of the traditional hidden Markov model. For example, a constrained HMM assumes a transition matrix constrained by neighborhood relations of the states [53], an idea rather closely related to the self-organizing map. For modeling data that involves different types of data, an HMM for metric and event-based data [54] can be used. For a comprehensive tutorial on hidden Markov models, see [55, 56] and for a review, see [50].

3.3.2 Learning with the EM-algorithm

There are three basic problems for hidden Markov models that must be solved for the model to be useful in real-world applications [48]:

1. Given the observation sequence X and the model θ , how to efficiently compute the probability of the observation sequence given the model $P(X|\theta)$?
2. Given the observation sequence X and the model θ , how to choose a state sequence S which is optimal in some meaningful sense?
3. How to adjust the model parameters to maximize $P(X, \theta)$?

The problem is to train the model, to learn the maximum likelihood estimates (MLE) for the parameters of the model. The MLE's are the parameters that maximize the probability of the observed data. Because in the HMM, the state sequence is not directly observed, the problem of obtaining the MLE's is not completely straightforward. The EM-algorithm (Expectation-Maximization) [47, 57] can be used to estimate the parameters of the hidden Markov model from a data set available.

The EM-algorithm is an iterative algorithm that finds the MLE's for the parameters in incomplete data problems. Here, incomplete data means that there is a many-one mapping from the unobserved state to the observed measurements, i.e. the same observation can be measured in many different states [57]. Because the density functions of the states are overlapping each other, it is only possible to tell the probability of a set of measurements (here the concentrations and needle mass) to belong to a certain state. The expected log-likelihood of the complete data is

$$Q(\theta|\theta^p) = E(\log P(X, S|\theta)|X, \theta^p), \quad (3.22)$$

where the log-likelihood of the complete data is parameterized by θ and the expectation is taken with respect to the second distribution parameterized by the current parameters θ^p . The algorithm consists of two steps: the E-step (expectation) and the M-step (maximization). The value of Q is calculated in the E-step. In the M-step, the parameter values θ^{p+1} are updated such that the value of $Q(\theta|\theta^p)$ is maximized. The EM-algorithm converges monotonically to a local maximum of the likelihood function of the data [57, 58].

Inference is the procedure for obtaining the probabilities $P(s_t|X)$ for $t = 1, \dots, T$ given the observation sequence X . To be able to calculate the maximum likelihood estimates of the parameters, the state probability sequence must be inferred for the time series. The inference is done in two parts: filtering and smoothing [31]. The filtering starts from the beginning of the time series and proceeds forward in time. At each time step t , the predicted state s_t is calculated given the observation sequence up to $t - 1$:

$$X_{t-1} = \{x_1, \dots, x_{t-1}\} \quad (3.23)$$

and the probability density of s_{t-1} .

$$P(s_t = j|X_{t-1}) = \sum_{i=1}^N a_{ij}P(s_{t-1} = i|X_{t-1}) \quad (3.24)$$

After the filtered probabilities $P(s_t = j|X_{t-1})$ have been estimated for $t = 1, \dots, T$, the smoothed probabilities can be computed. Smoothing is done backwards in time for $t = T - 1, \dots, 1$. The procedure consists of two equations that are evaluated alternately.

$$P(s_t = j, s_{t+1} = i|X_T) = P(s_t = j|X_t) \frac{P(s_{t+1} = i|X_T)}{P(s_{t+1} = i|X_t)} a_{ji} \quad (3.25)$$

$$P(s_t = j|X_T) = \sum_{i=1}^N P(s_t = j, s_{t+1} = i|X_T), \quad (3.26)$$

where

$$X_T = \{x_1, \dots, x_t, \dots, x_T\} \quad (3.27)$$

The filtering is done starting from the beginning of the sequence and smoothing backwards in time. Therefore, the inference algorithm is sometimes referred to as the “forward-backward” algorithm [59].

The probability density function of the observed variable is assumed to be a Gaussian probability density function with mean and covariance depending on the state. The parameters of the model for each state are the following: the mean values μ of the variables, the variances σ^2 of the variables and the probabilities to switch to another state a_{ij} , $i, j = 1, \dots, N$, i.e. the transition matrix. The variance of all the variables is assumed to be the same to decrease the number of free parameters. By decreasing the number of parameters, learning from limited amount of data becomes more stable and less dependent on random variations of the data.

The model parameters are updated using the following system of equations [31]. For a more detailed description of the EM-algorithm, see [60].

$$\hat{\mu}_i = \frac{\sum_{k=1}^S \sum_{t=1}^T x_{t,k} P(s_{t,k} = i | X_{T,k})}{\sum_{k=1}^S \sum_{t=1}^T P(s_{t,k} = i | X_{T,k})}, \quad i = 1, 2, \dots, N \quad (3.28)$$

$$\hat{\sigma}_i^2 = \frac{\sum_{k=1}^S \sum_{t=1}^T (x_{t,k} - \hat{\mu}_i)^2 P(s_{t,k} = i | X_{T,k})}{\sum_{k=1}^S \sum_{t=1}^T P(s_{t,k} = i | X_{T,k})}, \quad i = 1, 2, \dots, N \quad (3.29)$$

$$\hat{a}_{ij} = \frac{\sum_{k=1}^S \sum_{t=2}^T P(s_{t,k} = j, s_{t-1,k} = i | X_{T,k})}{\sum_{k=1}^S \sum_{t=2}^T P(s_{t-1,k} = i | X_{T,k})}, \quad i, j = 1, 2, \dots, N \quad (3.30)$$

In above, $\hat{\mu}_i$ and $\hat{\sigma}_i^2$ are the estimates of the mean and variance of state i . The Equations 3.29–3.30 are used in the case that there are more than one time series that are assumed to be generated by the same process (Finland 36, Austria 71). The model is trained as follows: First, one filtering and smoothing step is carried out for each of the time series $k = 1, \dots, S$. Then, the parameters are updated using the inferred state probabilities. Next, the filtering and smoothing steps are carried out again separately for all the time series and the parameters are updated. These inference and updating steps are continued until the parameters have reached some stationary values. Learning the model parameters this way assumes independence between the time series.

3.3.3 Analysis of the model

The quality of the model

The quality of learning can be tested with cross-validation [61]. The data is divided randomly into two equally sized parts: a training set, and a test set.

The training set is used to train the model, and the log-likelihood values Q of both sets are computed. Then, the sets are switched such that the original test set is used as the training set and the original training set as the test set. The model is trained with the new training set and the log-likelihood values of both data sets are computed. This is repeated several times in order to reduce the effects of a specific division. This kind of procedure is called two-fold cross-validation because the data set is divided into two parts. There are also other possibilities. For example, in ten-fold cross-validation, the data set is divided into ten parts, nine of which are used as the training set at a time and one as the test set and each choice of nine sets is tried. For a review and comparison of the accuracy estimation methods cross-validation and bootstrap, see [62]. The difference between the likelihoods of the test and training sets reflects the quality of the model. The smaller the difference, the better the model fits to the real distribution of the data, i.e. corresponds to the true structure of the data. Big difference in the likelihoods can be a sign of overfitting, i.e. the model being unable to generalize.

Comparison of the state sequences

Using the HMM, we can calculate a probability sequence for each stand that indicates the probability of belonging to a state. Using these sequences, different measures between the stands can be calculated. The Kullback-Leibler divergence (KL), Hellinger (H) and Pearson's ϕ^2 (PHI) dissimilarity measures and L_1 distance are defined, respectively, as

$$KL(\varphi, \psi) = \sum_{t=1}^T \varphi(t) \log \frac{\varphi(t)}{\psi(t)} + \psi(t) - \varphi(t) \quad (3.31)$$

$$H(\varphi, \psi) = \sum_{t=1}^T \left| \sqrt{\varphi(t)} - \sqrt{\psi(t)} \right|^2 \quad (3.32)$$

$$PHI(\varphi, \psi) = \sum_{t=1}^T \frac{|\varphi(t) - \psi(t)|^2}{\psi(t)} \quad (3.33)$$

$$L_1(\varphi, \psi) = \sum_{t=1}^T |\varphi(t) - \psi(t)|, \quad (3.34)$$

where φ and ψ denote the state probabilities of the two time series [63] that are compared to each other. From the four measures, only H and L_1 are symmetric. One drawback of the L_1 distance is that it largely ignores the differences in the tails of the probability density functions. In case of missing measurements, the measures can be scaled with a parameter $\frac{T}{m_{\varphi,\psi}}$, where $m_{\varphi,\psi}$ is the number of time steps, in which there were measurements from both time series.

In addition to these, a distance measure for Markov-sources was used [64]:

$$J(\theta_0, \theta) = \frac{1}{2}[D(\theta_0, \theta) + D(\theta, \theta_0)], \quad (3.35)$$

where

$$D(\theta_0, \theta) = \lim_{T \rightarrow \infty} \frac{1}{T} [\log q(X|\theta_0) - \log q(X|\theta)] \quad (3.36)$$

and $q(X|\theta)$ is the likelihood value of observation sequence X of length T with model parameters θ . The observation sequences can be randomly generated using parameters θ and the inferred state probabilities. The distance between two state probability sequences can be calculated by generating many (in the experiments 20) observation sequences from the inferred state probabilities and taking the average over their J -distances.

Using these measures, it is possible to compare the state probability sequences of the time series with each other instead of only single probabilities. These measures can also be compared with the geographical distances. The possible correlations between the geographical and the above mentioned measures can be investigated using permutation testing similarly as explained in Section 3.2.3. The tested hypothesis was that stands that are close to each other also have smaller values of the above mentioned distance (and other) measures (Equations 3.31–3.35). The stand pairs that are closer to each other than some threshold distance are compared to the rest of the stand pairs. The result of the test is the probability that the difference between the mean dissimilarity (KL, H, PHI, L_1, J) of the stands that are close to each other and the mean dissimilarity of the rest of the stands is smaller than the difference between the mean dissimilarities of two randomly generated groups of stands.

3.3.4 From the cluster model to a transition graph

There are some significant differences between the solution of the clustering and the hidden Markov model:

1. The probability densities of the HMM states overlap each other and therefore, only the probability can be calculated for a measurement to belong to a certain state. The clusters do not overlap each other. A measurement can belong only to a single cluster.
2. In the training, time affects the probabilities of the HMM states. The probability of the state depends on the state at the previous time step through the transition matrix. The clusters do not have any temporal relations.
3. The shape of the distributions of the HMM states is always Gaussian. The shape of the clusters is not restricted.

Even though the time dimension is not taken into account when training the cluster model, it can be used in the analysis of the model by, for example constructing transition graphs that show the probabilities of switching from one cluster to another. The transition graphs can simply be constructed as follows. First, the number of data vectors belonging to a cluster Z_i is calculated for each of the clusters $i = 1, \dots, C$. The number of measurement pairs that are consecutive in time and the former measurement belongs to cluster i and the latter to cluster j is denoted by Z_{ij} . Then, the maximum likelihood estimates of the transition probabilities for the whole data set can be calculated as follows.

$$\hat{a}_{ij} = \frac{Z_{ij}}{Z_i}, \quad i, j = 1, \dots, C \quad (3.37)$$

Using the acquired transition matrix, it is easy to construct visualizations that show the typical cluster switches.

Chapter 4

Results of the analyses of nutrient concentration data

4.1 Foliage of Finland

4.1.1 Spatial statistics

Interpolation of the measurement values was done using a triangle based cubic method. It is based on a Delaunay triangulation of the data [2]. The interpolated smooth surface always goes through the data points and has no discontinuities in the first derivative.

The results of interpolation are shown in Figure 4.1. It can be seen that the values of N and S are quite similar every year. In south-eastern Lapland there is a constant deficiency of N and S. Otherwise, the spatial variations of the two nutrients are quite small. In the years 1995, 1996 and 1999, the N concentrations were low in most of the country and in 2000 they were high. The S concentrations have decreased over the years. In the 1980's there was clearly more S than in the 90's, the concentrations were particularly low in 1999. The values of P are less stable. The two clear properties of P concentration are that it is high in southern Finland and low in middle Finland. In eastern Finland, there is a turnover from low to high concentrations between years 1995–1997. The P concentrations were higher in 1987 and 1988 and slightly lower in 1992 and 1999. The needle masses of spruce change rather much from one year to another. The needles were heavy

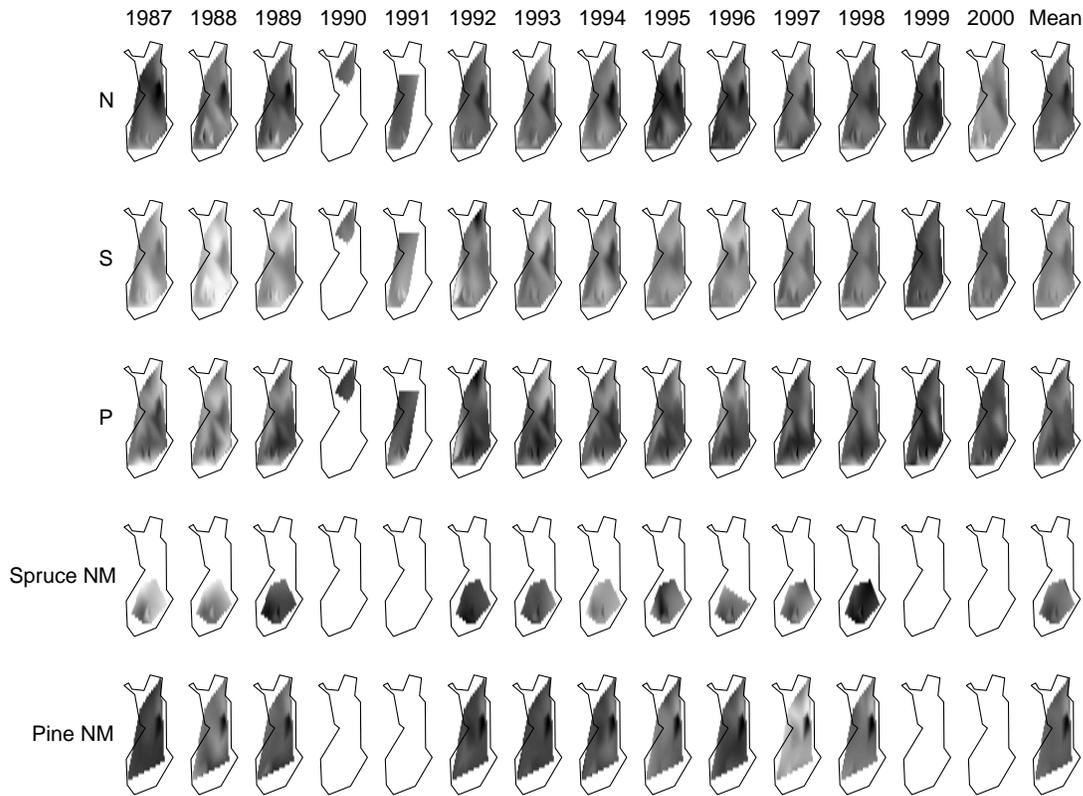


Figure 4.1: *The interpolated values of the nutrients N, S and P and needle mass in Finland for all years. In the two bottom rows are the needle masses of spruce and pine separately. Darker color indicates smaller value. The mean values over all years are in the rightmost column.*

in the years 1987, 1988 and 1994 and light in 1989, 1992 and 1998. The needle mass of pine is usually (not surprisingly) low in northern and high in southern Finland. In 1988 and 1997, the pine needles were quite heavy whereas in 1987 and 1992, they were light. In these years, the temperatures of June were lower than normally. After the latter year the weight of the needles have had an increasing trend.

The semivariograms of measurements from Finland were calculated for N, S, P and NM measurements. They are shown in Figure 4.2. It can be seen that there is quite a lot variation in the shapes of the semivariograms between different years. This means that the source and processes which affect the nutrient

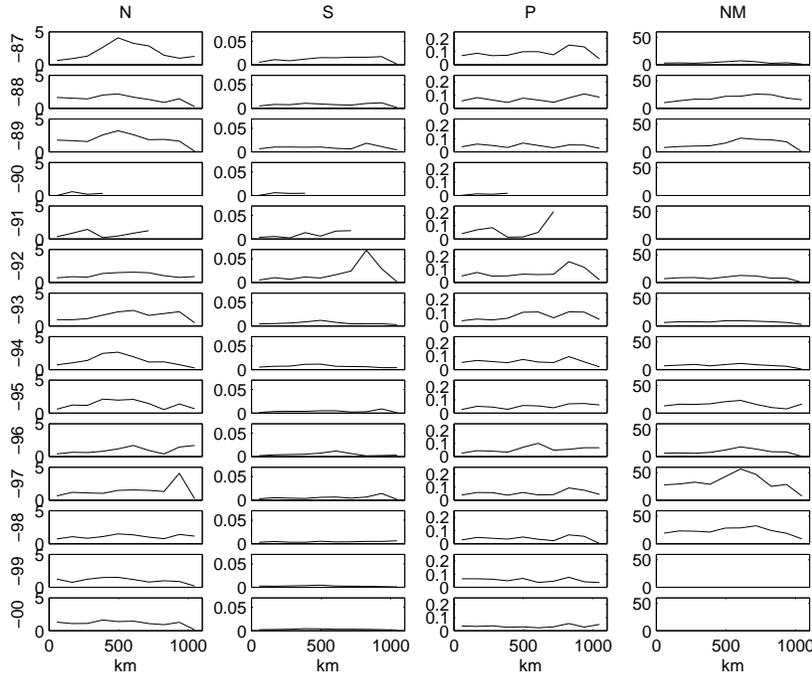


Figure 4.2: *The semivariograms of the N, S, P and NM measurements from Finland based on the annual data. The rows correspond to the years 1987–2000.*

concentrations change from one year to another. The least amount of variation is in the semivariograms of P. This suggests that the processes controlling the concentration of P change slower.

The mean semivariogram of all years were calculated with different lag distances (110–550km). The mean semivariograms were also calculated for the tree species separately. The results are combined in Figure 4.3. The mean semivariograms of N and NM have peaks at distances around 600–700km and around 900km with S and P. Thus, there seems to be some spatial correlation in the data. The results, however, get much clearer when the tree species are handled separately.

When considering only spruce stands, the values of N and NM have rather clearly such structure that stands that are close to each other have smaller mutual distances. This can also be seen in S but less clearly. Similar structure is also visible in pine stands with N and S, but for pine, there is no structure in NM. The reason for the NM semivariogram of the complete data getting lower values with

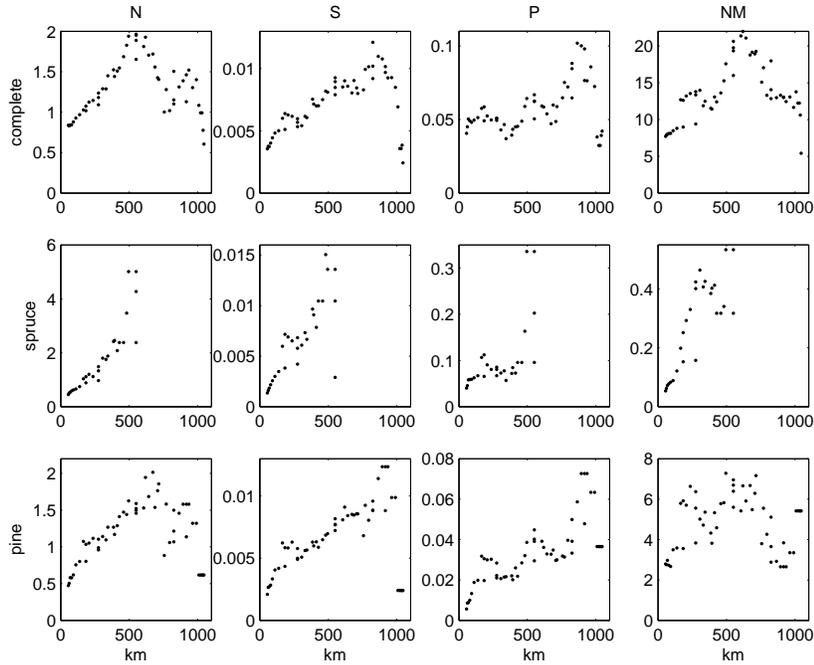


Figure 4.3: *The mean semivariograms of N, S, P and NM measurements from Finland with lag distances 110–550km for all years. In the first row are the semivariograms for the complete data, in the second row for spruce and in the third row for pine.*

lag distances longer than 600km is most likely that 600km is the maximum distance between spruce stands. There are spruce stands only in southern Finland. With longer lag distances, only pine stands can be compared to each other and the differences in NM between the same species are smaller than the differences between different species. The results of the mean semivariograms suggest that similar N, S and NM measurements are often made at stands that are close to each other. Therefore, there could be some kind of a connection between the clusters and their locations. This can be studied with the indicator semivariogram (see Section 4.1.2).

4.1.2 Finding the nutrition profiles

Because there was only a little a priori information about the structure of the data, a clustering method was used to analyze the relationships between the nutrients at the stand level. The nitrogen, sulphur and phosphorus concentrations (c) in the needles and needle mass (NM) of the current year’s needles were used with the clustering method. These elements were chosen because of their importance to the growth of the tree, their dynamic temporal behavior and because the number of measurements was not particularly high and therefore, we wanted to keep the dimensionality of the data limited. Needle mass was used because it enabled us to interpret the clusters based on nutrient concentration as well as nutrient content. The 4-dimensional data vector \mathbf{x} used in the clustering method for measurement stand i and year t was:

$$\mathbf{x}_{i,t} = [c_{i,t}(N), c_{i,t}(S), c_{i,t}(P), NM_{i,t}]^T. \quad (4.1)$$

Before the actual clustering, the data was normalized so that the mean of each variable was 0 and variance 1 to ensure that all variables have equal weights in the training.

The data was clustered using the algorithm explained in Section 3.2. The linearly initialized map was trained with SOM Toolbox¹ [65] using the batch algorithm [15] in two rough training epochs and five fine tuning epochs. The final neighborhood width was 1 in order to ensure good quality of quantization. For the Finnish data, a map consisting of a regular hexagonal grid with 6×9 map units was used.

The average quantization error of the map was 0.82. The error was calculated using the normalized variable values. Increasing the map size could have decreased the error but could also have led to overfitting. The topographic error was 5.5%. The topographic error was typical for this data; map size didn’t affect it much. According to the topographic error, the topology of the data set was preserved rather well in the quantization process. Therefore, clustering the SOM can be assumed to give reliable results.

The U-matrix of the SOM of Finnish data is shown in Figure 4.4a and the final clustering result in Figure 4.4b. The clustering result is similar to the U-

¹<http://www.cis.hut.fi/projects/somtoolbox>

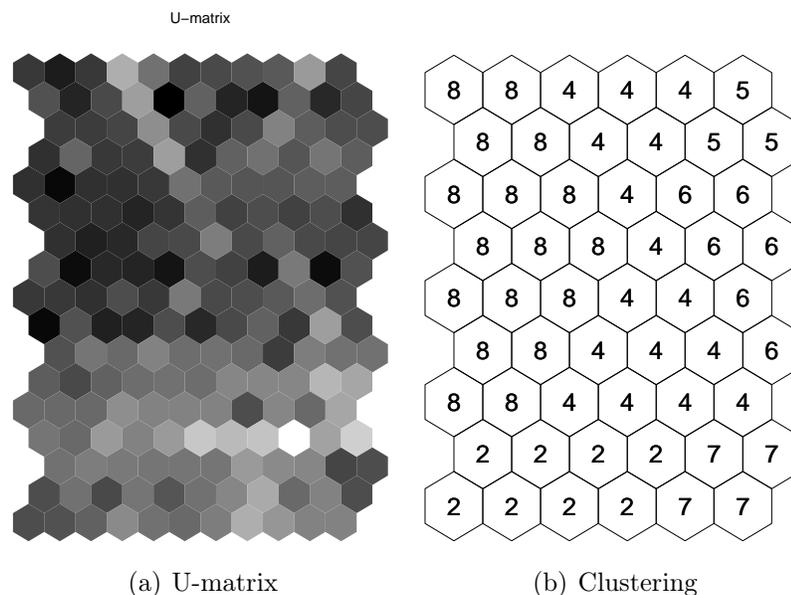


Figure 4.4: (a) *U*-matrix for the SOM of Finnish data. Darker color indicates smaller value. (b) The clustering result. The numbers (2–8) indicate which cluster the map unit belongs to. Some numbers are not present, because those clusters were not chosen from the hierarchy to the final clustering.

matrix. The hierarchical structure of the clustering can be seen in Figure 4.5. One cluster (9) was pruned out of the hierarchy and bottom level clusters 1 and 3 were combined into cluster 8 in the final clustering. Clusters 2 and 7 are the most different from all other clusters. These two clusters have the highest mean values of the concentration variables.

The mean values and standard deviations of all the measurements for the clusters are shown in Table A.2. The component planes of the SOM of Finnish data are shown in Figure 4.6. Using the component planes and the mean values, the six clusters of the data from Finland can be described qualitatively as:

- 2: High N and S, average P, high NM.
- 4: Average N and S, low P and NM.
- 5: Low N, S, P, and NM.
- 6: Average N, S and P, low NM.

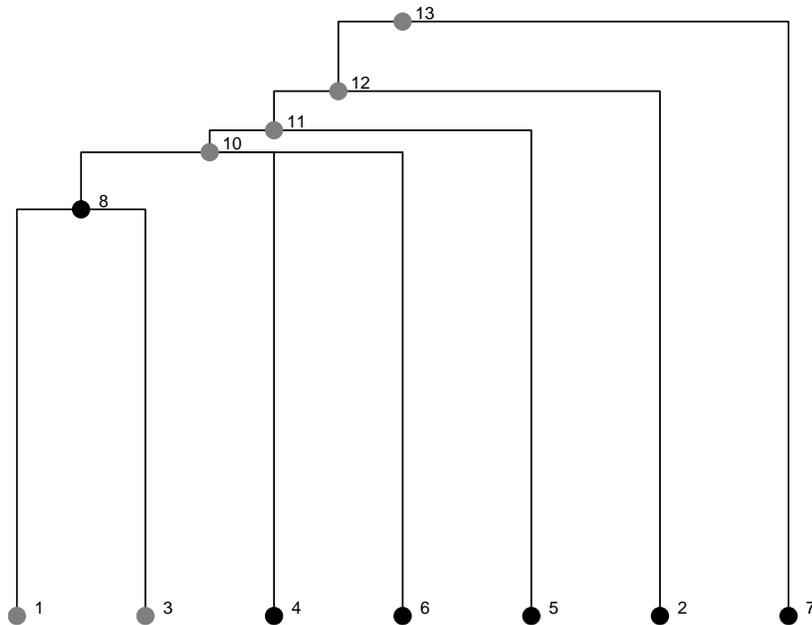


Figure 4.5: *Cluster hierarchy of the SOM of Finnish data. Black circles are the final clusters.*

7: High N, S and P, low NM.

8: Average N, low S and P, high NM.

Needle mass is the variable that had the most missing measurements in Finland. Despite of that, needle mass contributed quite much to the clustering result. It can be seen that clusters 4–7 have rather low and clusters 2 and 8 rather high mean needle mass values. These two groups of clusters do not overlap each other much with respect to needle mass, i.e. their standard deviations are fairly small. The reason for this is that according to histograms (not shown), needle mass has a clearly bimodal probability density function, whereas all the concentration variables have more or less unimodal probability density functions. The bimodality of needle mass is caused by the two different tree species having needles of different size. Now, the two cluster groups approximately correspond to the two different peaks of the probability density function of needle mass.

The clusters have slightly different meanings for different tree species. The nutrition profiles, i.e. the mean values and standard deviations of the clusters for

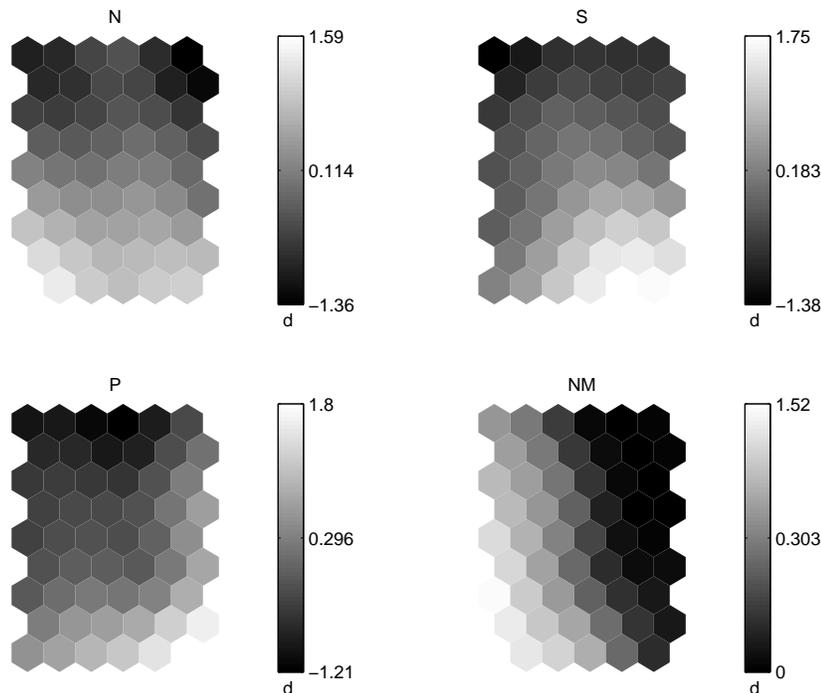


Figure 4.6: *The component planes of the SOM of Finnish data. The values of the component planes correspond to the values of the normalized data.*

the two species are in Tables A.3 and A.4. For pine, the clusters can be explained as follows [66]. Cluster 5 represents trees with multiple-nutrient deficiency. All the concentrations and needle mass are low. Clusters 4 and 6 represent a sub-optimal nutrient status. Cluster 4 is characterized by a deficiency of P and cluster 6 may have P-excess. Clusters 2 and 7 have high S and P concentrations. Both clusters have excess of these nutrients but only cluster 2 has high needle mass. Cluster 8, which is the most common one, has favorable S and P concentrations but N is probably a limiting factor of the growth.

When considering spruce, there were no expert's interpretations available for the clusters. According to Tables A.3 and A.4, there are no significant differences in the concentrations of N, S and P, but the NM of the clusters are differently related to each other. Clusters 4, 5, 6 and 7 have very similar NM and clusters 2 and 8 only contain measurements without NM information. Because the clusters have so similar needle masses, it is hard to say, which clusters, if any, could

Table 4.1: *The cluster switch probability matrix of Finland. The rows show the conditional probabilities of switching from a certain cluster to another.*

Previous\Current	2	4	5	6	7	8
2	0.40	0.15	0.00	0.00	0.03	0.42
4	0.09	0.49	0.06	0.12	0.05	0.19
5	0.04	0.32	0.52	0.04	0.00	0.08
6	0.05	0.16	0.05	0.55	0.11	0.09
7	0.23	0.08	0.00	0.23	0.38	0.08
8	0.10	0.09	0.00	0.01	0.00	0.79

possibly represent the optimal environment for spruce growth.

The switching of the cluster of a stand was analyzed to find out if there is any structure in the development of nutrient concentrations. The most common switches were 8–2, 8–4, 2–8, 4–8 and 6–4. When considering two consecutive switches, the most common series were 8–2–8, 2–8–2, 8–4–8, 4–5–4, 4–6–4 and 4–8–2. These results are not very surprising since the most common clusters are 8, 4 and 2. Usually the switches happen between the most common clusters.

Because the absolute numbers of switches did not reveal much information about the data, the conditional probabilities of switching the cluster were calculated. They are shown in Table 4.1. The high probabilities tell something about the most typical switches. The process seems to have some tendency to converge to clusters 4 and especially 8. Also, switches 7–2 and 7–6 have above average probabilities.

More information about the switches can be extracted by analyzing the tree species separately. Graphs showing the structure of switching the cluster for both species separately are shown in Figure 4.7. Spruce stands usually belong to clusters 4–7, cluster 4 being the most common one and 6 the second most common. Also, switches that happen with high probability are 5–4 and 8–4. In pine stands, the process is most of the time in cluster 8. As time goes on, more stands switch to cluster 8 than from cluster 8. Clusters 5, 6 and 7 are less common than 2 and 4. Switches that happen with high probability are 2–8, 4–8, 6–8 and 7–2.

In the years 1987, 1988, 1991 and 1993, the high needle mass cluster 8 was less

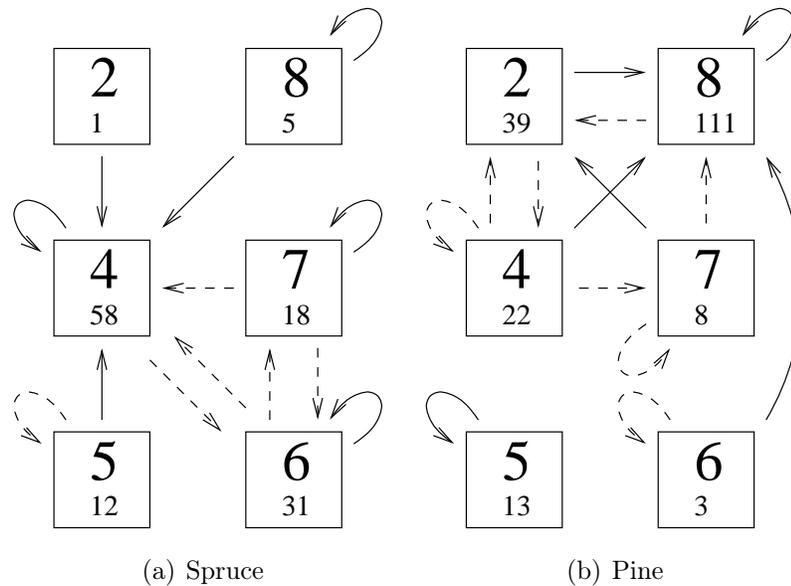


Figure 4.7: *Graphs showing the typical cluster switches in Finland. Solid line denotes a probability higher than 0.4 and dashed line a probability between 0.1 and 0.4. The smaller number under the cluster number is the number of years a stand has belonged to that cluster. Switch probabilities for spruce (a) and pine (b) are shown separately.*

usual than otherwise. In 1987 and 1988, the high sulfur concentration clusters 2 and 7 were more usual than normally. The reason for high number of low needle mass clusters in 1991 might be that there were a lot of measurements missing from everywhere else but southern Finland, where the low needle mass clusters are normally more probable than elsewhere. This is caused by the fact that there are no spruce stands in northern Finland. What reduces the significance of this result is that in 1991, there were needle mass measurements only from two stands. In 1993, the low needle mass cluster 4 was the most common one. Starting from 1995, the number of stands in cluster 4 has decreased.

The indicator semivariograms (Equation 3.4) for the clusters of Finland are shown in Figure 4.8. They suggest that there may be some kind of structure in the geographical locations of clusters 2, 5 and 7. For cluster 5 this seems reasonable, because according to Figure 4.9 it is found almost only in the N and S poor area in south-eastern Lapland. The problem is that the graphs have so much noise

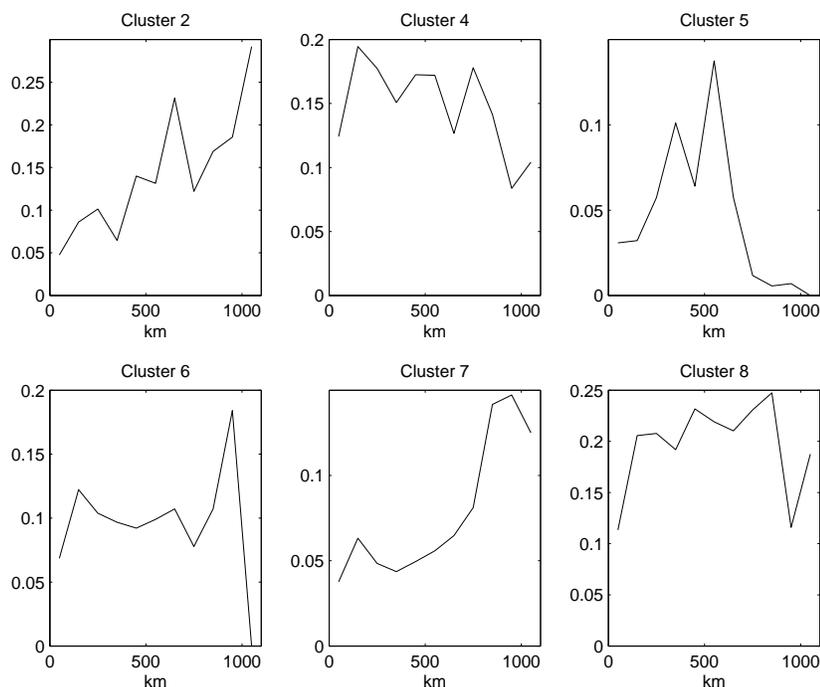


Figure 4.8: *The mean indicator semivariograms of the clusters of Finland for all years.*

that it is hard to say whether this impression reflects any true properties of the data or not. However, it seems clear that there is not much, if any, correlation between the locations of clusters 4, 6 and 8.

The clustering result of Finland on a geographical map for each year can be seen in Figure 4.9. The probability of a stand to be in cluster 8 does not seem to be very much connected to the geographical position of the stand. In southern Finland, other clusters are a little more common than cluster 8. Stands in the other clusters are spread more unevenly on the map. A stand in cluster 2 is most likely in northern or south-western Finland. Clusters 4 and 7 can usually be found in southern Finland and cluster 4 also in middle Finland. Cluster 5 exists most often in south-eastern Lapland and cluster 6 in southern and western Finland.

The possible connections between weather and the clustering were analyzed with permutation testing. The tested measurements were six weather variables:

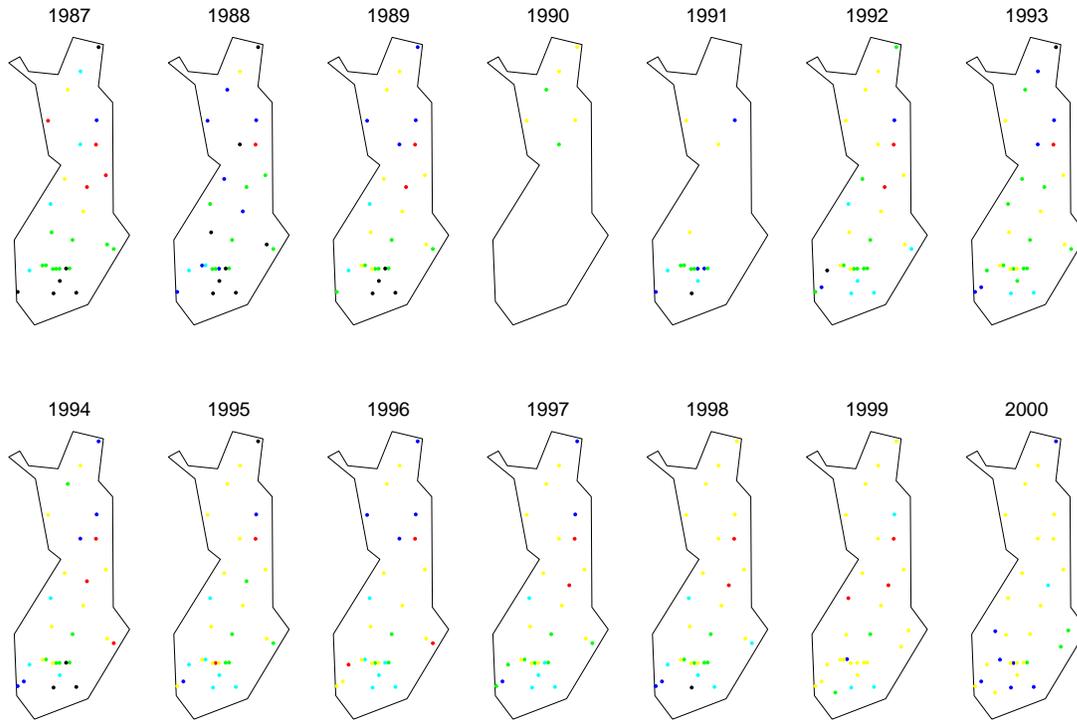


Figure 4.9: *Clustering of the measurement stands of Finland for each year. Color coding: dark blue = cluster 2, green = cluster 4, red = cluster 5, light blue = cluster 6, black = cluster 7, yellow = cluster 8.*

the average temperature of the year, temperatures in January, and July, precipitation and the probabilities that a random observation gets a lower value than the observation of temperature and precipitation for each year. The results with 10000 repetitions are shown in Table A.5.

It can be seen that clusters 2 and 8 have very small probabilities that the difference between the mean probability of temperature in the cluster and outside the cluster is smaller than the difference between the mean probability of temperature in a random cluster and outside the random cluster. The opposite holds for clusters 4, 5, 6, and 7. Thus, the temperature in those clusters is more likely low than in clusters 2 and 8. The permutation test probabilities for precipitation are low for clusters 7 and 8 and high for clusters 4, 5 and 6. Low precipitation is less likely in clusters 7 and 8 than in clusters 4, 5 and 6.

Since there seems to be a connection between weather and the clustering,

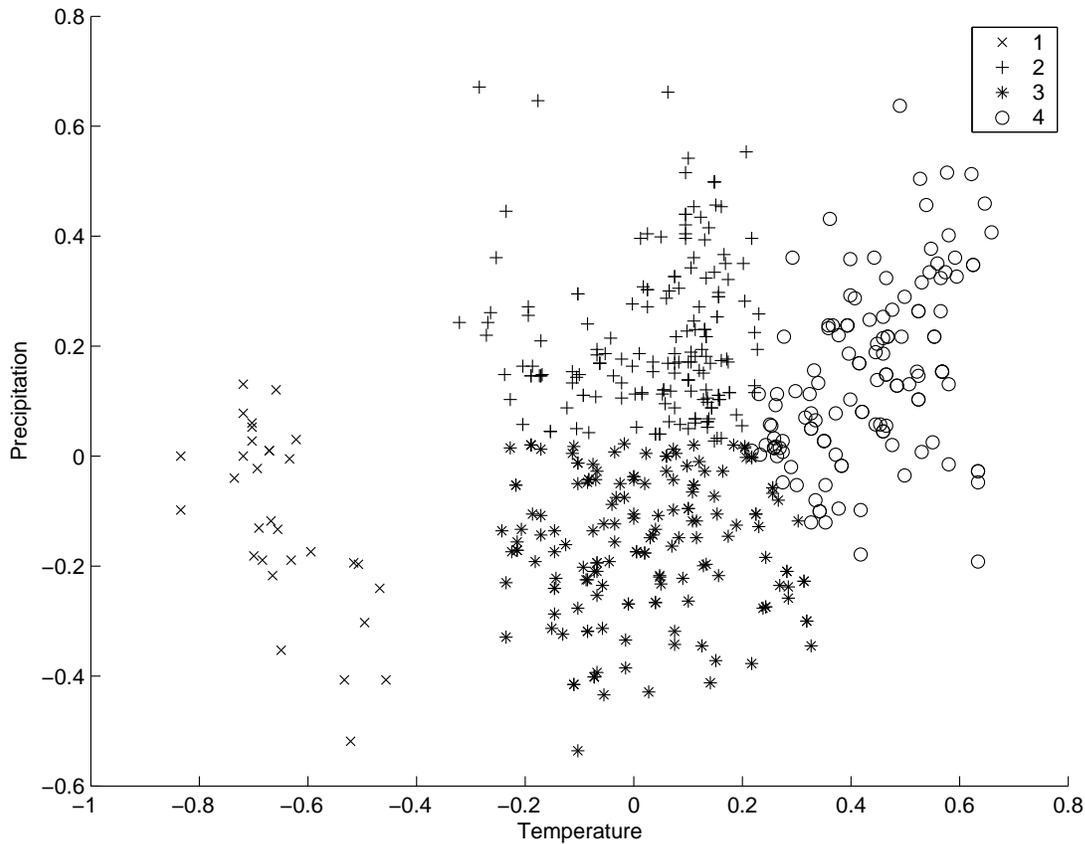


Figure 4.10: *The weather data. The values of the axes are the normalized temperature and precipitation values.*

the effect of the weather on the clustering was analyzed further. The weather probabilities were mapped back to the input space with the inverse of the normal cumulative distribution function with $\mu = 0$ and $\sigma = 1$. The obtained normalized weather data was divided into 4 clusters using the k-means algorithm. The structure and the clusters of the weather data are shown in Figure 4.10. For clarity, the weather clusters are from now on called weather states. Weather state one corresponds to cold temperature and slightly lower precipitation. States two and three correspond to normal temperature, state two has higher precipitation than state three. State four corresponds to high temperature and slightly above average precipitation.

A transition matrix (as in Table 4.1) for the clusters can be constructed for

each weather state separately. The transition matrices were calculated and it was found that they are different from each other. It seems that in weather state two, there is a higher probability to switch to cluster 7 and a lower probability to switch to clusters 8 and 4. Also, the probability of not switching the cluster at all is higher. In weather state three, the switch does not often happen to cluster 7, but from 7 to 2 or 4. In weather state 4, there is a high tendency to switch from all the clusters to cluster 8. Thus, warm temperature seems to be connected to cluster 8. The transition matrix for weather state one could not be calculated because all the measurements in that state were from the year 1987 and there was no information about the clusters in the previous year. As can be seen in Figure 4.10, the temperatures in 1987 were highly exceptional. In state one, the number of stands in clusters 4, 5 and 7 was higher than otherwise and lower in clusters 2 and 8. It seems reasonable that when the temperature is low, the needles grow slower and the number of stands in clusters with high needle mass decreases.

4.1.3 Temporal modeling of foliar nutrient concentrations

In the experiments, the N, S and P concentrations of the needles and needle mass were used in the time series. The data was normalized so that the mean of each variable was 0 and variance 1 to ensure that all variables have equal weights in the training. So, the data consisted of 36 four-dimensional time series with length of 14 years.

The parameters of the model were trained using the EM-algorithm with 150 iterations. The initial values of the parameters were the following: the mean values were spread around 0 with a gaps of size 0.2 between them, all the variances were 0.5 and the probability of staying in each state was 2/3. The number of states varied in the experiments from 2 to 5.

The convergence of the mean values μ during training is shown in Figure 4.11 for two and three state models. The models with more states behave identically to the three state model. With any number of states, one state converges to high mean values and the rest to low values. All the parameters of the low mean value states are exactly the same including the means, variances and transition

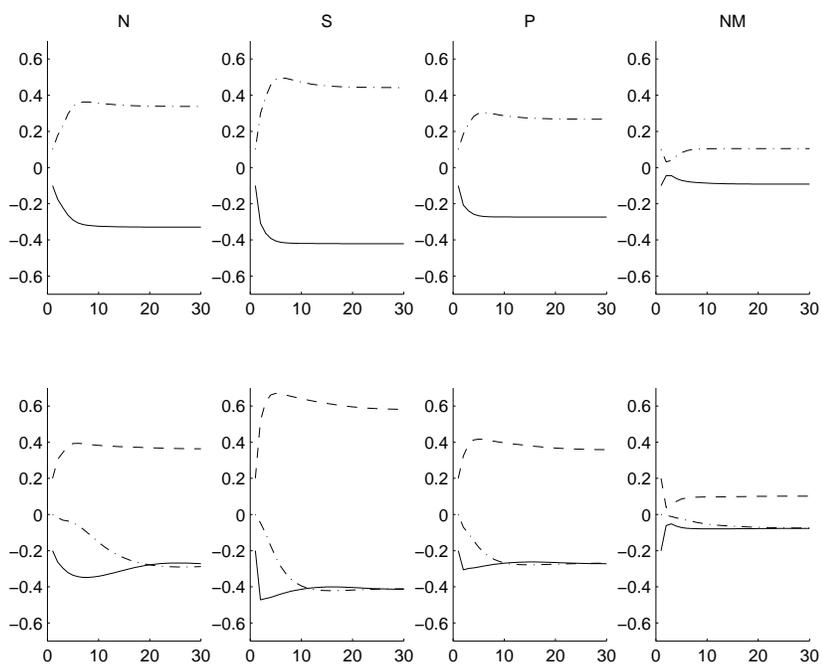


Figure 4.11: *The convergence of the states' mean values in Finland during training. In the top row is the two-state model and in the bottom row the three-state model. The scale on y-axis is normalized.*

probabilities. Apparently, there are no more than two HMM-states to be found in the data. The probability of staying in the high concentration state is about 0.5 and the probability of switching from any of the low concentration states to the high concentration state is 0.2–0.4, the more states the lower the probability. This is interesting, because the low concentration states are essentially the same. The mean values of the high concentration state, however, are higher in the multi-state models.

The mean values and standard deviations of the states for the two-state models of both countries are shown in Table 4.3. It can be seen that the states have quite a lot of overlap, especially in P concentration and needle mass. The states are so similar, that accounting for the precision of the chemical analysis methods it would be impossible to distinguish them.

Two-fold cross-validation was used to test the quality of the learning, the results were as follows. With 20 iterations, the mean Q -value for the training set

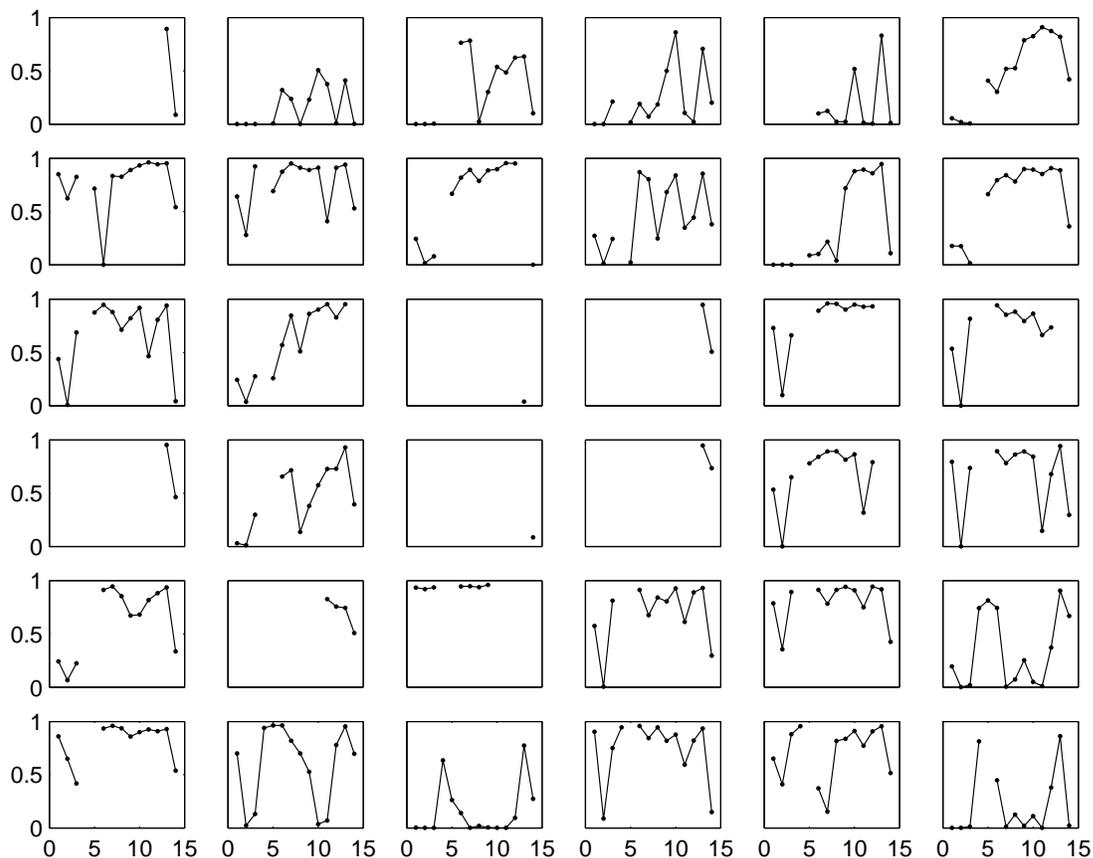


Figure 4.12: *The state probabilities for the 36 measurement stands of Finland during the 14 year sampling period. The curves illustrate the probability of the low concentration state 1.*

was -280 and the standard deviation 30 . For the test set, the values were -330 and 50 . The quality of the model can be considered good, because the difference between the log-likelihood of the training and test set is rather small. A few times the log-likelihood of the test set was even higher than the log-likelihood of the training set. The model seems to fit well to the data.

The probability sequences of the low concentration state 1 are shown for all stands in Figure 4.12. In many stands, the probability of state 1 is low in 1988, high in about 1993–1999 and drops significantly in 2000. Both in 1988 and 2000, in many stands the concentration of N was higher than normally.

Using the measures explained in Section 3.3.3, similarity of the state prob-

Table 4.2: *The permutation test results for the dissimilarity measures using different threshold distances for Finland. The values are the probabilities that the difference between the mean dissimilarity (KL, H, PHI, L_1, J) of the stands that are closer to each other than a threshold distance and the mean dissimilarity of the rest of the stands is smaller than the difference between the mean dissimilarities of two randomly generated groups of stands.*

Distance (km)	100	200	300	400	500	600	700	800	900	1000
KL	0.06	0.95	1.00	1.00	1.00	1.00	0.90	0.94	0.60	0.15
H	0.05	0.94	1.00	1.00	1.00	0.98	0.82	0.94	0.35	0.11
PHI	0.20	0.43	0.82	0.88	0.98	0.80	0.74	0.88	0.93	0.34
L_1	0.08	0.92	1.00	1.00	1.00	0.98	0.77	0.94	0.29	0.07
J	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.92	0.91

ability sequences of the measurement stands can be studied. All the different measures between stands were compared to the geographical distance, but there did not seem to be any strong correlation between the geographical position of the stand and its state probability sequence. Only in the case of J -distance, there seemed to be some faint structure.

The results of the permutation test are shown in Table 4.2. They are related to the P -values of the test as $1 - P$. These results also suggest that only J -distance has a connection with the geographical distance. With the other measures, there seems to be a correlation with the geographical distance at distances around $500km$, but not, for example, under $100km$, which suggests that geographical distances are not very closely connected to the probability sequences.

4.2 Foliage of Austria

4.2.1 Spatial statistics

The results of the triangle based cubic interpolation of the nutrient concentrations and needle mass are shown in Figure 4.13. The distribution of the nutrients on the map is more random than in Finland. There are some small areas that

Table 4.3: *The means and standard deviations of both countries' two-state models.*

	Finland				Austria			
	State 1		State 2		State 1		State 2	
	μ	σ	μ	σ	μ	σ	μ	σ
N (<i>mg/g</i>)	11.7	1.0	12.6	1.4	12.8	1.1	13.1	1.4
S (<i>mg/g</i>)	0.89	0.09	0.99	0.12	0.97	0.10	0.99	0.12
P (<i>mg/g</i>)	1.48	0.19	1.62	0.27	1.48	0.29	1.50	0.35
NM (<i>g/1000</i>)	7.7	3.0	8.5	4.1	4.80	0.89	4.87	1.06

constantly have less N than others, but they are spread quite randomly on the map. The average N concentration was particularly low in the years 1991, 1992 and 1994. The concentration of S decreases with time and towards west. In 1994 and from 1997 on, the concentrations were somewhat lower than otherwise. The distribution of P is quite uneven: some stands have higher and some lower concentrations. On average, the values are lower in northern Austria. The trend is that P concentrations decrease in time. Not much can be said about the distribution of NM of spruce, because there were measurements only from two years and they are completely different from each other. In 1996, the needles were significantly heavier than in 1995.

The semivariograms of measurements of Austria (Figure 4.14) show that there are less differences between different year's measurements than in Finland. The mean semivariogram (Figure 4.15) of NM has a peak around $400km$, whereas the concentration measurements do not have any significant peaks. The trend of NM cannot really be interpreted much due to the limited amount of data and the fact that the semivariograms of the years 1995 and 1996 look completely different. If the trend is real, it means that there is structure in the short distance, but it is lost in the long distances. There is no clear spatial structure to be found in the semivariograms of the measurements of Austria.

4.2.2 Finding the nutrition profiles

The preprocessing of the data, training of the SOM and clustering were done similarly as with the data of Finland (see Section 4.1.2). For the Austrian data,

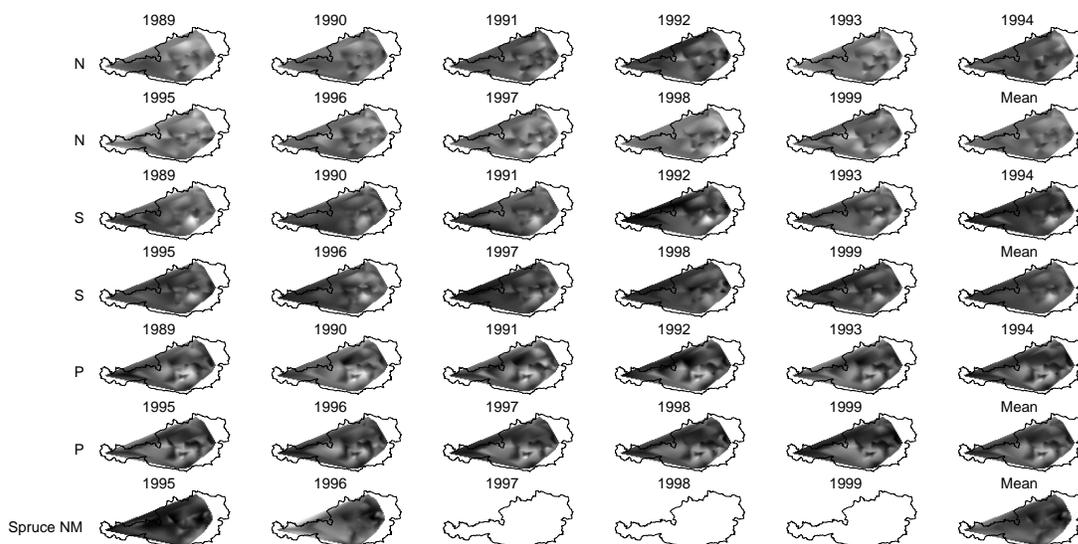


Figure 4.13: *The interpolated values of the nutrients N, S and P and needle mass in Austria for all years. In the top row are the N values for years 1989–1994, in the second row the N values for years 1995–1999 and the mean value over all years, etc. Darker color indicates smaller value. Needle masses are only from spruce.*

a larger map of size 8×12 was used due to the larger size of the data set.

The average quantization error of the map was 0.54. The lower average quantization error for Austrian data is a result of bigger map size. As with the map of Finland, the topographic error was not alarmingly high: 4.3%.

The U-matrix and clustering result for the Austrian data are shown in Figure 4.16 and the hierarchical structure of the clustering in Figure 4.17. Three clusters (11, 12 and 16) were pruned out of the clustering and bottom level clusters 2, 3, 5, 6 and 8 were combined into cluster 14. In the hierarchy, cluster 9 is the most dissimilar to other clusters. Its mean sulfur and phosphorus concentration values are the highest.

The mean values and standard deviations of all the measurements of the clusters (nutrition profiles) are shown in Table A.1. Using it and the component planes of the SOM (Figure 4.18) the six clusters of Austria can be qualitatively described as follows:

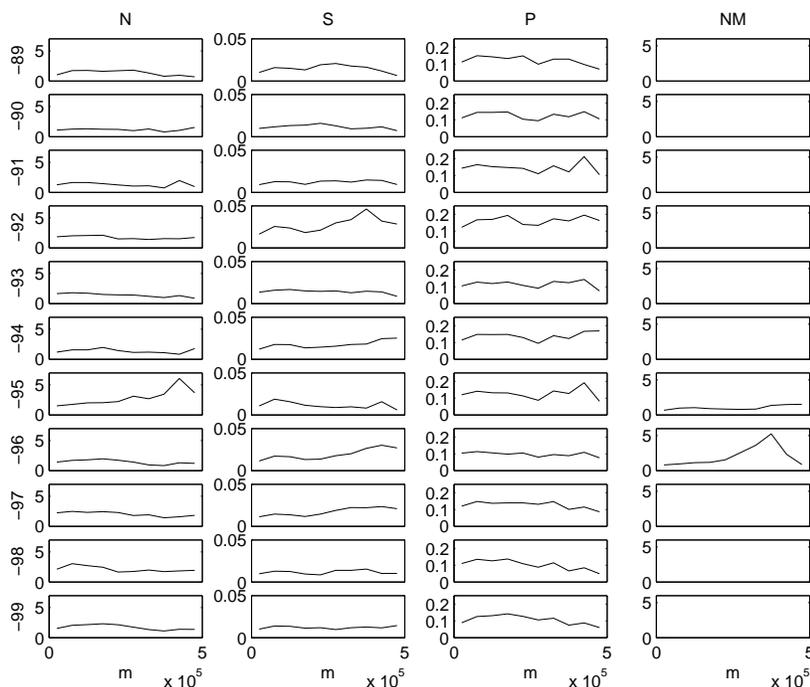


Figure 4.14: *The semivariograms of N, S, P and NM measurements from Austria based on the annual data. The rows correspond to the years 1989–1999.*

- 1: High N, average S and P, low NM.
- 4: Low N, S, P and NM.
- 7: Low N, S and P, average NM.
- 9: Average N, high S and P, low NM.
- 10: Low N, average S, high P, low NM.
- 14: Average N and S, low P, average NM.

In Austria, cluster 7 is the one with the highest needle mass. The concentrations of N, S and P are low. In the other clusters, the needle masses are lower. However, due to the high number of missing needle mass measurements, this should not be emphasized too much. In cluster 4, all the concentrations are very low. Cluster 14 is the average cluster with respect to the nutrient concentrations.

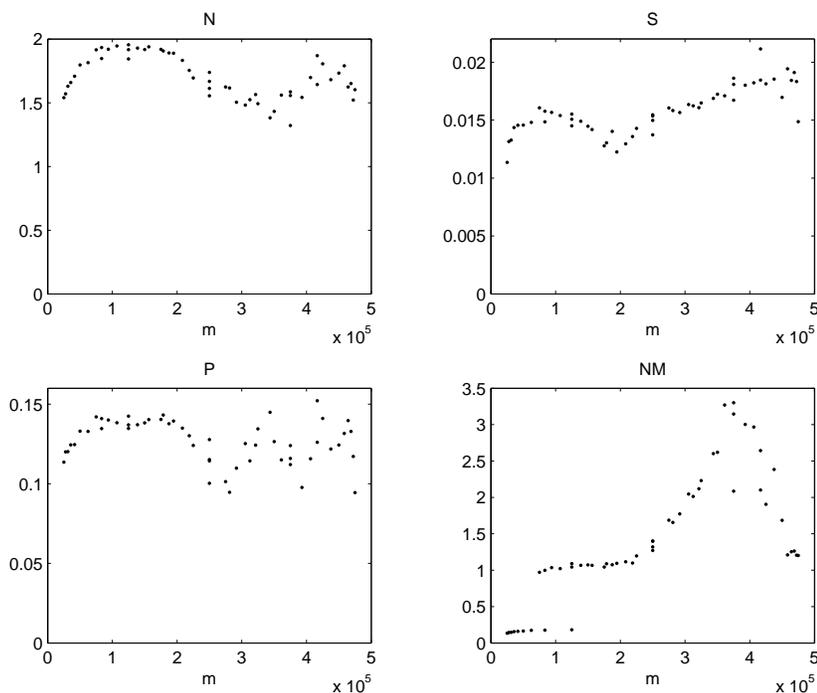


Figure 4.15: *The mean semivariograms of N , S , P and NM measurements from Austria with lag distances 50–250km for all years.*

The main difference between clusters 14 and 10 is that in cluster 10 the concentration of P is much higher. In cluster 9, the concentrations of S and P are the highest and in cluster 1 the concentration of N is the highest.

There was very much missing needle mass data from Austria; there were measurements only from the years 1995 and 1996, i.e. the concentration variables contributed more to the clustering result than needle mass. Therefore, needle mass was not taken into account when selecting the number of clusters. Unlike in Finland, the probability density function of needle mass is not bimodal and the clustering result of Austria was not divided with the tree species as in Finland. This is caused by the fact that there were no needle mass measurements from pine trees. With respect to the other measurements, pine and spruce are a lot harder to distinguish. Due to the very small weight of needle mass in the learning process, one should be careful when drawing conclusions regarding the clusters' mean needle masses.

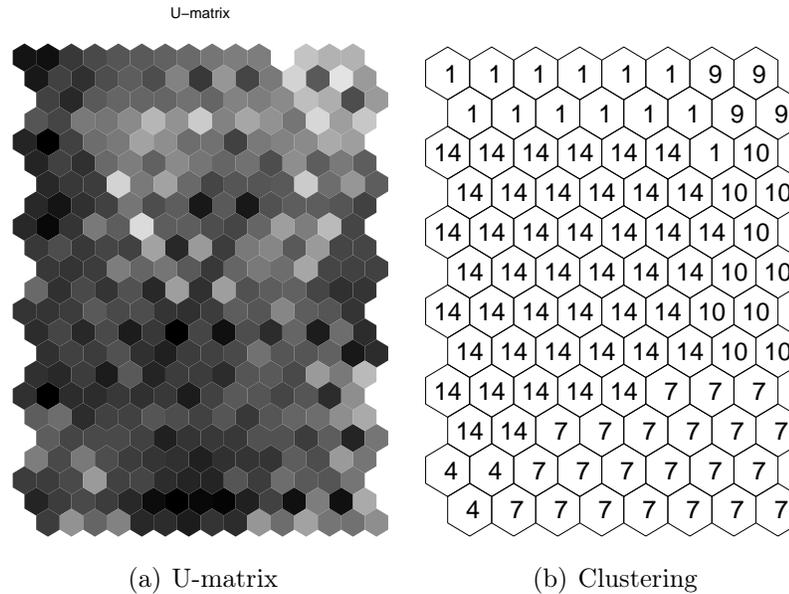


Figure 4.16: (a) *U-matrix for the SOM of Austrian data. Darker color indicates smaller value.* (b) *The clustering result. The numbers indicate which cluster the map unit belongs to. Some numbers are not present, because those clusters were not chosen from the hierarchy to the final clustering.*

The switching of the clusters was also analyzed. The most common switches were 14–7, 7–14, 1–14 and 14–1. When considering two consecutive switches, the most usual combinations are 14–7–14, 7–14–7, 14–1–14, 1–14–1. It can be concluded that most switches happen between the biggest clusters with the exception of cluster 10. The switching probabilities of the clusters were also computed. The transition matrix is shown in Table 4.4. The system is shown graphically in Figure 4.19. Because only a small minority of the stands were pine stands, there was no need to separate the tree species here.

According to Figure 4.19, there seems to be a flow from clusters 1, 4, 9 and 10 to clusters 7 and 14. It should be kept in mind, however, that the values in Table 4.4 are probabilities and they do not necessarily tell the whole truth about the transitions. For example the probability of switching from 4 to 14 is 0.25 and from 14 to 4 0.04, but still the number of switches from 14 to 4 is higher than from 4 to 14, because the total number of years a stand belonged to cluster 14 is so much higher. There were only 5 pine stands in the data set, but a difference can

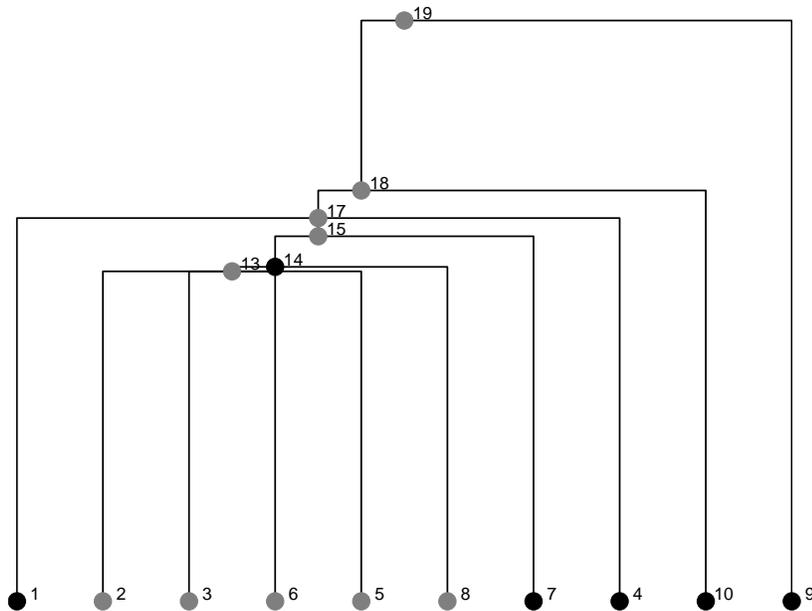


Figure 4.17: *Cluster hierarchy of the SOM of Austrian data. Black circles are the final clusters.*

be seen between the different species. Pine stands belong less often to clusters 4, 7, 9 and 10 than spruce stands.

In the years 1994 and 1996–1999, more stands were in cluster 7 than usually. The amount of S in the needles was lower than normally in those years. Between 1993 and 1994, the amount of N and S decreased quite much and many stands switched from cluster 14 to 7. In 1993 and from year 1996 on, there were fewer stands in clusters 10 than earlier. In 1989, 1995 and 1998, the number of stands in cluster 1 was slightly higher. In these years, the amount of N was above average. In 1992, there was a peak in the number of stands in cluster 4 and a drop in the amount of N.

The mean indicator semivariogram of the Austrian clusters for all years is shown in Figure 4.20. There could be some weak spatial correlation between the stands in cluster 7 but no clear structure can be seen in the locations of the clusters.

The clustering result of Austria on a geographical map for each year can be seen in Figure 4.21. It tells a lot more about the spatial distribution of the clusters

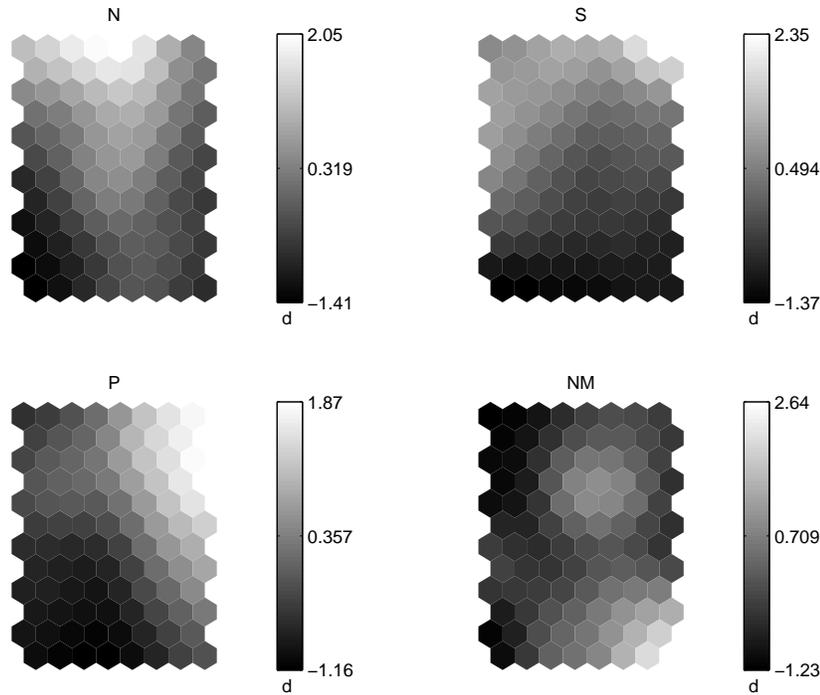


Figure 4.18: *The component planes of the SOM of Austrian data. The values of the component planes correspond to the values of the normalized data.*

than the indicator semivariogram. Stands in the most common clusters 14 and 7 are spread quite evenly on the map. High phosphorus and sulfur concentration cluster 9 is most likely to be found in a few stands little south of the middle of Austria. Stands in cluster 10 are usually in southern Austria. In that area, there are mountains and that explains the clearly higher mean altitude value of that cluster. Cluster 1 exists most often in the plains of north-eastern Austria and cluster 4 in eastern and middle Austria.

4.2.3 Temporal modeling of foliar nutrient concentrations

The convergence of the mean values during training is shown in Figure 4.22 for two and three state models. The convergence of the parameters is similar to the model of Finland. One state has high mean value of all concentrations and all the rest have the exact same low values as well as the same variances and transition probabilities. The probability of switching to the high concentration

Table 4.4: *The cluster switch probability matrix of Austria. The rows show the conditional probabilities of switching from a certain cluster to another.*

Previous\Current	1	4	7	9	10	14
1	0.38	0.00	0.12	0.00	0.06	0.44
4	0.00	0.22	0.50	0.00	0.03	0.25
7	0.06	0.11	0.42	0.00	0.04	0.38
9	0.07	0.00	0.00	0.66	0.17	0.10
10	0.04	0.01	0.12	0.02	0.57	0.23
14	0.10	0.04	0.19	0.01	0.06	0.60

state is approximately 0.3–0.5 from any other state. As in Finland, high number of states decreases the probability of switching to the high concentration state. It should be noted that the number of needle mass measurements is much lower than the number of the other measurements and thus, its weight in the learning was rather small.

As can be seen from Table 4.3, the states of the model of Austria overlap each other even more than the states of the model of Finland. Due to the accuracy of the chemical analysis methods, the states are practically indistinguishable.

The cross-validation results for Austria were as good as for Finland. The mean and standard deviation of Q of the training set were -560 and 70 . For the test set the values were -610 and 60 . Again, a few times the log-likelihood of the test set was higher than the log-likelihood of the training set. It seems that the model fits well to the data, but not that well that it could be considered overfitting.

The state probability sequences of the stands for low mean concentration state 1 are shown in Figure 4.23. The dissimilarities between the sequences using different measures (see Section 3.3.3) suggest that there is no correlation between the geographical position of the stand and its state probability sequence.

The different measures of the state sequences were compared with the geographical distances using permutation testing. The results are in Table 4.5. They are related to the P -values of the test as $1 - P$. According to the results, there is no clear connection between the the geographical distance and the state sequences. Only J -distance may have some relation with the geographical dis-

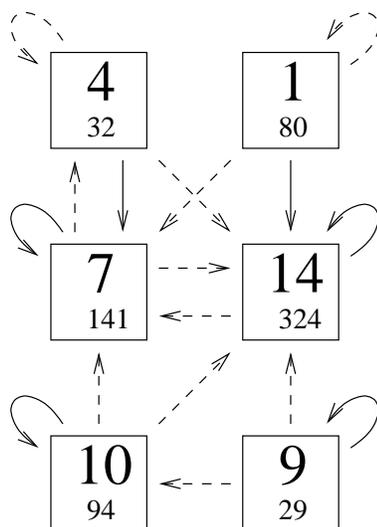


Figure 4.19: *Graphs showing the typical cluster switches in Austria. Solid line denotes a probability higher than 0.4 and dashed line a probability between 0.1 and 0.4. The smaller number under the cluster number is the number of years a stand has belonged to that cluster.*

tances, but the probabilities are significant only with threshold distances 50 and 200km.

4.3 Discussion

4.3.1 Spatial statistics

The analysis of the measurements using methods of spatial statistics was found to be useful. In Finland, according to the interpolated measurement values, there is some spatial structure in the data. Also, the semivariograms suggest that the measurements are spatially correlated. The result that the shape of the semivariograms of P does not change much supports the idea that P comes to the tree mainly from the soil. The other nutrients may be affected more by a combination of local and transboundary sources.

In Austria, there is no clear spatial structure in the data. The existence of the short-distance structure in NM is dubious. Some stands have constantly smaller

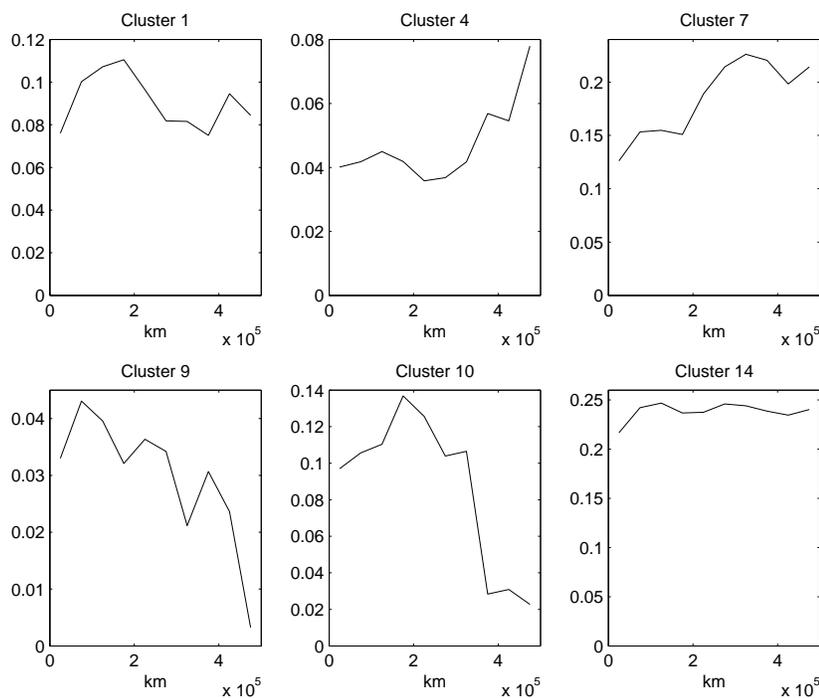


Figure 4.20: *The mean indicator semivariograms of the clusters of Austria for all years.*

and some higher concentrations, but they seem to be spread rather randomly throughout the country.

4.3.2 Clustering using the self-organizing map

Clustering the data with the SOM-based algorithm revealed some rather interesting properties of the data sets, especially from Finland.

In Finland, a quite clear structure of the data was found. The data was divided into two groups of clusters: one for the high and one for the low needle masses. This way, the algorithm approximately divided the tree species to their own clusters. Experts were able to give all the clusters some meaningful interpretations about the state of the forest and thus, the clustering method can be considered successful. In addition, a connection between the clusters and weather was found.

The clustering result of the Austrian data was less clear than for the data of

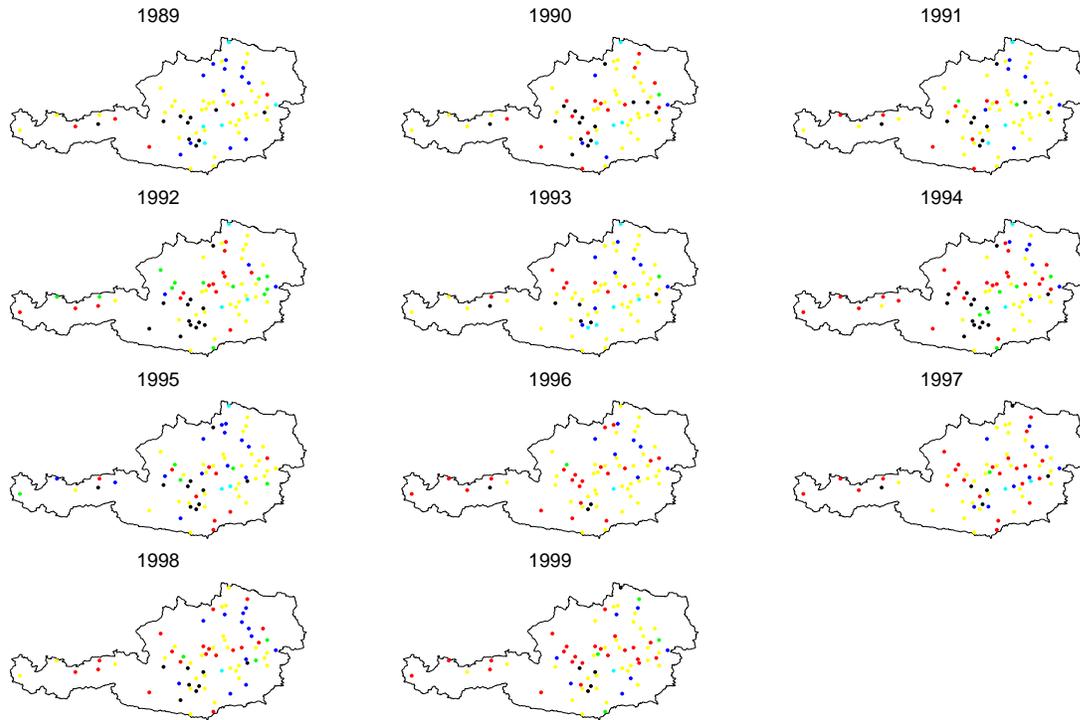


Figure 4.21: *Clustering of the measurement stands of Austria for each year. Color coding: dark blue = cluster 1, green = cluster 4, red = cluster 7, light blue = cluster 9, black = cluster 10, yellow = cluster 14.*

Finland. Similar clusters to those in Finland were found, but the very limited amount of needle mass measurements made interpretation of the results much more difficult. Also, the lack of weather data reduced the possibilities of validation of the result. There are, however, no reasons to doubt the quality of the clustering other than the lack of needle mass measurements. The connection between the clusters and the altitude of the measurement stands suggests that the clustering result of Austria is meaningful, as well as that the same method gave meaningful results for similar data from Finland.

The results of the clustering method were validated by forest experts and they were found to be meaningful. Also, the switching of the cluster of the stands seemed reasonable. The results of this method gave the reason to introduce the concept of nutrition profile in forest research. The concept is a powerful method for describing the elemental variation of the data and a significant improvement

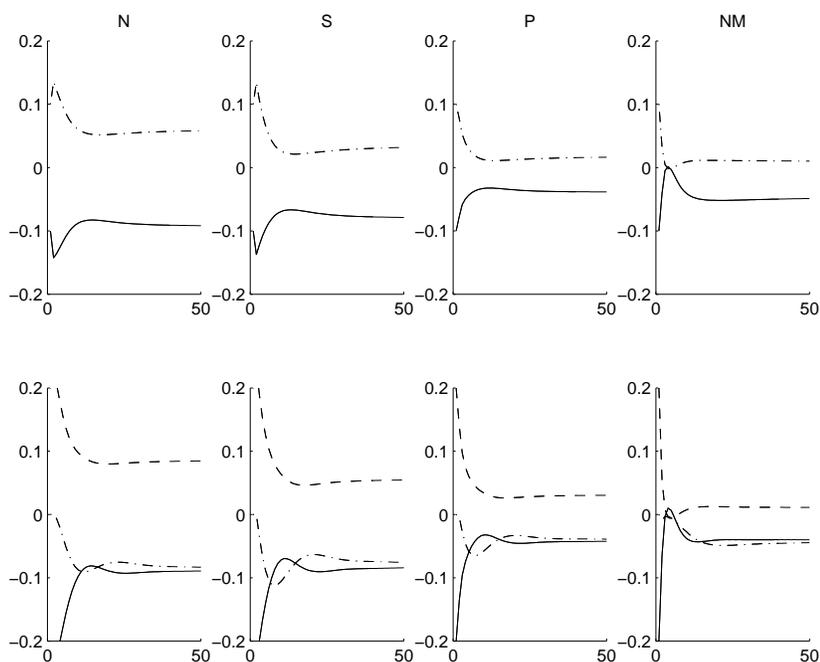


Figure 4.22: *The convergence of the states' mean values in Austria during training. The rows correspond to the number of states of the model: top 2, bottom 3. The scale of y-axis is normalized.*

compared to the traditional use of single nutrient concentrations or ratios between two concentrations to describe the state of a tree [66].

4.3.3 Time series modeling

The results of the Hidden Markov model contained less information than the results of the clustering. In both countries, the number of states was two. The states correspond simply to high and low measurement values. The main problem with the states is that the differences between the mean values of the states are usually smaller than the accuracy of the chemical analysis methods. Thus, it is impossible to actually distinguish the states from each other. In addition, there is plenty of data that does not fit too well into either state. This means that the generalization property of the model is not satisfactory. Also, the inability of the model to find more than two states from the data reduced the usability of the model.

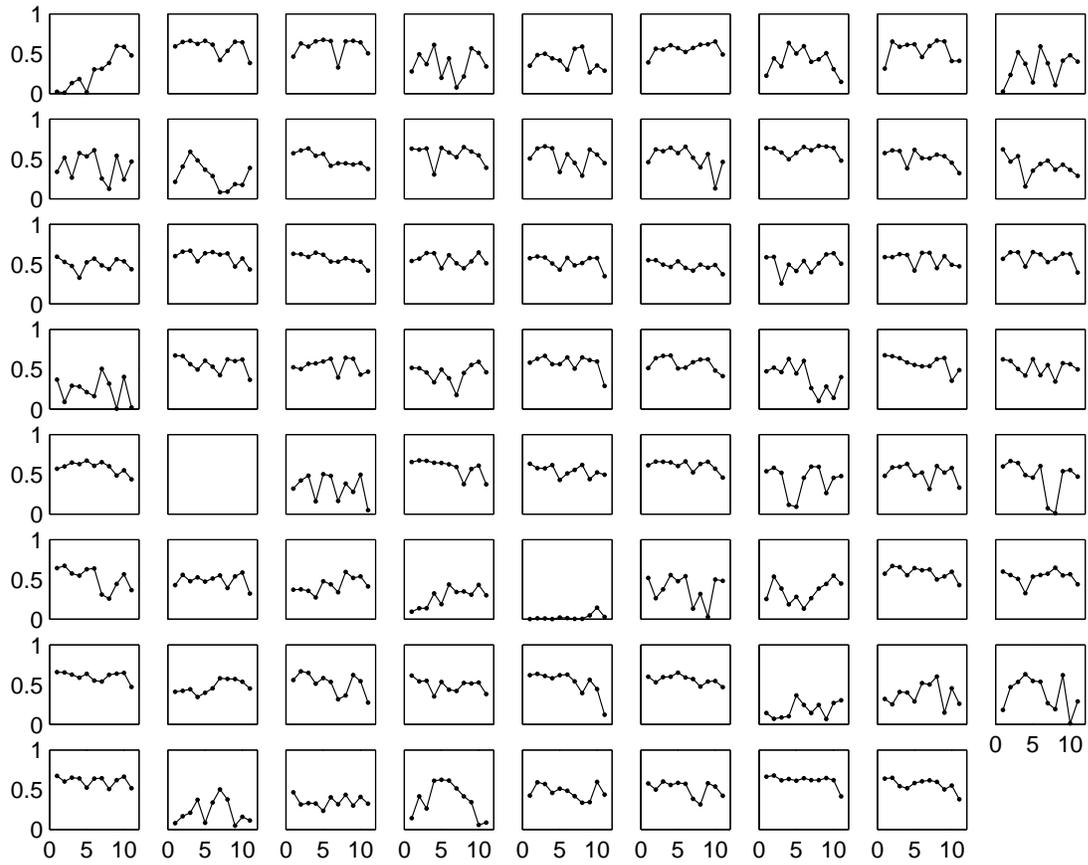


Figure 4.23: *The state probabilities for the 71 measurement stands of Austria during the 11 year sampling period. The curves denote the probability of state 1.*

Table 4.5: *The permutation test results for the dissimilarity measures using different threshold distances for Austria. The values are the probabilities that the difference between the mean dissimilarity (KL, H, PHI, L_1, J) of the stands that are closer to each other than a threshold distance and the mean dissimilarity of the rest of the stands is smaller than the difference between the mean dissimilarities of two randomly generated groups of stands.*

Distance (km)	50	100	150	200	250	300	350	400	450
KL	0.84	0.01	0.00	0.04	0.04	0.11	0.12	0.04	0.00
H	0.91	0.03	0.02	0.14	0.12	0.09	0.10	0.04	0.01
PHI	0.59	0.00	0.00	0.07	0.03	0.18	0.24	0.27	0.00
L_1	0.88	0.00	0.01	0.01	0.01	0.00	0.01	0.03	0.01
J	0.99	0.67	0.81	0.99	0.91	0.69	0.57	0.05	0.00

Chapter 5

Summary and conclusions

In this study, the nutrient concentrations of pine and spruce needles were analyzed using different data analysis methods. The data was collected from Finland and Austria in 1987–2000. The tested analysis methods included a few spatial statistics methods: semivariograms and interpolation, clustering of the self-organizing map with some simple temporal analysis of the clusters and actual time series modeling with the hidden Markov model. The aim of the study was to analyze the spatial and temporal distribution of the nutrient concentrations and simply try to find out what kind of internal structure there is in the data and how the different data analysis methods perform with this kind of data.

It was found that semivariance, a spatial statistic, is a reasonably usable measure for analyzing the factors that affect the nutrient concentrations on a local scale. With the data used in this study, the semivariograms showed some trends in the similarity of nearby stands. The problem with the graphs was that they were rather noisy and therefore not very easy to interpret. It was also noted that interpolation of the measurements makes it possible to draw figures that are both visually appealing and help understand the geographical structure of the data.

The hidden Markov model used in the time series analysis did not yield much information about the temporal structure of the data. When compared to the clustering method, the results of time series modeling were clearly inferior. The classification of the data into two groups did not give any especially interesting information about the connections between the measurements. The two states

were simply so similar to each other that no remarkable conclusions could be drawn considering the possible source that could have generated the measurements. Apparently, the basic hidden Markov model was not the optimal time series model to be used with this kind of data.

The VS clustering algorithm based on the self-organizing map provided new information about the relations of the nutrients between different years and locations. With the clustering method, it was possible to divide the measurements into six groups. In each group, the growth of the needles and the amounts of the nutrients were different, i.e. different groups represented different types of growing conditions. Forest experts were able to construct a model that characterizes the development of the condition of forests in Finland using the result of the clustering method.

5.1 Future work

The results of the tested methods, especially the clustering method were promising, but not that excellent that there would not be any need for improvements. There are some adjustments that might be worth trying to the models in the future.

First, a model should be constructed that effectively uses both the spatial and temporal dimensions of the data. One possibility to achieve this would be to construct different time series models for different parts of the country. It would probably also be worth trying to use separate models for different tree species and perhaps even different weather conditions. The weather data could be included as a more internal part to the models. In this study, only the current year's needles were analyzed. Using both the current and previous year's needles, more information about the growth and development of the needles could be extracted.

In the future, the clustering model could be enhanced by using probability distributions instead of the crisp clusters. This way, only the probability could be given that a measurement belongs to a certain cluster. This kind of fuzzy clustering can be obtained by using for example Gaussian mixture models or building the Gaussian distributions on top of the self-organizing map as in [67]. In addition to the SOM, simple and perhaps useful visualizations of the data

could be achieved using other projection methods like Sammon's mapping and principal component analysis.

Also, the use of more complex time series models could be beneficial. So far, we have tested some autoregressive switching models [31], but due to numerical problems in training, no results can be shown here yet. The complexity of the models should not, however, be too high, because of the limited amount of measurements. In time series analysis, instead of the actual measurement data, the change of the measurements between two consecutive time steps could be used to train the model as for example in [40].

Appendix A

Tables

Table A.1: *The means and standard deviations of the concentration and needle mass measurements, weather probabilities and the geographical positions of the clusters in Austria.*

Cluster	1	4	7	9	10	14
N (<i>mg/g</i>)	15.5±1.0	10.8±0.6	12.2±0.6	13.4±1.1	12.4±0.7	13.1±1.2
S (<i>mg/g</i>)	1.10±0.08	0.80±0.06	0.86±0.05	1.30±0.14	0.98±0.06	1.01±0.08
P (<i>mg/g</i>)	1.57±0.33	1.16±0.17	1.27±0.24	2.12±0.23	2.02±0.26	1.42±0.26
Ca (<i>mg/g</i>)	4.05±1.29	5.84±1.81	4.62±1.72	4.87±1.29	3.12±1.10	4.44±1.50
Mg (<i>mg/g</i>)	1.24±0.26	1.26±0.42	1.26±0.36	1.43±0.31	1.41±0.25	1.26±0.29
K (<i>mg/g</i>)	6.06±1.15	5.84±1.50	5.65±1.18	7.83±1.57	5.70±1.60	6.24±1.32
Zn (<i>μg/g</i>)	36±10	42±12	36±10	62±77	30±9	36±11
Mn (<i>μg/g</i>)	932±581	365±462	395±426	765±343	727±399	573±509
Fe (<i>μg/g</i>)	61±31	38±15	37±18	57±15	39±14	46±21
NM (<i>g/1000</i>)	4.16±0.76	3.32±0.38	5.87±1.09	4.53±0.47	4.18±0.39	4.89±1.00
Latitude (°)	47.9±0.6	47.6±0.4	47.6±0.5	47.7±0.7	47.4±0.4	47.6±0.5
Longitude (°)	15.1±0.9	14.7±1.5	14.0±1.6	15.0±0.4	14.2±0.9	14.8±1.2
Altitude (<i>m</i>)	721±348	828±286	963±349	964±311	1328±285	907±353

Table A.2: *The means and standard deviations of the concentration and needle mass measurements, geographical positions of the clusters in Finland and the weather variables.*

Cluster	2	4	5	6	7	8
N (<i>mg/g</i>)	13.7±1.1	12.3±0.9	9.9±0.6	11.6±0.7	13.6±0.9	11.8±1.0
S (<i>mg/g</i>)	1.04±0.09	0.96±0.11	0.85±0.08	0.91±0.07	1.14±0.10	0.88±0.08
P (<i>mg/g</i>)	1.76±0.16	1.38±0.20	1.54±0.17	1.72±0.17	2.05±0.21	1.43±0.13
Ca (<i>mg/g</i>)	2.63±0.93	3.95±1.54	3.12±0.96	4.77±1.48	4.24±1.60	2.25±1.08
Mg (<i>mg/g</i>)	1.14±0.15	1.14±0.15	1.12±0.12	1.21±0.16	1.27±0.14	1.07±0.13
K (<i>mg/g</i>)	5.77±0.65	6.03±0.97	6.40±0.76	6.37±1.11	6.30±0.98	5.24±0.72
Zn (<i>μg/g</i>)	45±8	34±8	33±8	37±9	40±11	40±6
Mn (<i>μg/g</i>)	472±190	621±241	572±176	788±335	769±354	445±170
Fe (<i>μg/g</i>)	40±11	34±14	26±11	30±8	43±13	31±8
Cu (<i>μg/g</i>)	3.14±0.69	2.28±0.61	1.71±0.43	2.26±0.61	2.97±1.02	2.72±0.60
Al (<i>μg/g</i>)	313±128	94±99	68±71	61±56	146±143	218±90
B (<i>μg/g</i>)	12±5	13±5	12±3	12±5	14±6	12±4
NM (<i>g/1000</i>)	12.6±2.8	5.4±2.3	4.9±1.3	4.6±1.2	6.1±2.7	11.1±2.6
Latitude (°)	64.6±3.2	62.7±1.7	65.0±1.5	62.4±1.8	62.9±3.1	64.4±2.6
Longitude (°)	25.5±2.8	25.8±2.0	27.8±1.8	25.2±1.9	25.9±1.9	26.0±2.1
Avg. temp. (°C)	2.4±2.8	3.0±1.6	1.0±1.7	3.2±1.7	2.6±2.6	2.2±2.1
Jan. temp. (°C)	-7.6±5.2	-7.8±5.8	-11.6±5.0	-7.1±5.1	-8.9±6.8	-8.7±4.3
July temp. (°C)	15.3±2.0	15.8±1.8	14.6±1.8	15.6±1.7	16.3±2.5	15.3±1.7
Prob. of temp.	0.56±0.09	0.51±0.13	0.50±0.13	0.51±0.11	0.48±0.15	0.56±0.10
Precip. (<i>mm</i>)	591±120	641±102	601±131	634±111	611±120	613±112
Prob. of precip.	0.53±0.09	0.51±0.08	0.50±0.10	0.50±0.08	0.55±0.08	0.54±0.09

Table A.3: *The means and standard deviations of the concentration and needle mass measurements, geographical positions of the clusters in Finland and the weather variables for spruce.*

Cluster	2	4	5	6	7	8
N (<i>mg/g</i>)	14.9±1.8	12.2±0.9	10.1±0.5	11.6±0.7	13.6±0.9	12.4±1.6
S (<i>mg/g</i>)	0.95±0.07	0.95±0.12	0.84±0.07	0.91±0.07	1.11±0.10	0.82±0.06
P (<i>mg/g</i>)	1.79±0.12	1.32±0.19	1.50±0.21	1.71±0.17	2.05±0.24	1.36±0.17
Ca (<i>mg/g</i>)	4.84±1.31	4.61±1.26	3.29±0.90	4.98±1.30	5.07±1.16	5.23±1.43
Mg (<i>mg/g</i>)	1.15±0.16	1.16±0.16	1.12±0.08	1.22±0.16	1.28±0.12	1.16±0.19
K (<i>mg/g</i>)	7.12±0.53	6.38±0.86	6.34±0.91	6.42±1.10	6.49±1.10	6.82±0.51
Zn (<i>μg/g</i>)	31±12	31±6	31±7	36±9	37±11	36±8
Mn (<i>μg/g</i>)	811±400	687±245	656±172	821±319	932±300	734±290
Fe (<i>μg/g</i>)	30±5	29±9	23±4	29±6	39±10	23±3
Cu (<i>μg/g</i>)	2.06±0.34	2.03±0.42	1.72±0.29	2.20±0.59	2.51±0.55	2.08±0.33
Al (<i>μg/g</i>)	48±22	41±21	33±16	47±17	62±29	34±13
B (<i>μg/g</i>)	17±9	14±5	11±3	12±5	13±6	12±6
NM (<i>g/1000</i>)	-	4.2±0.8	4.4±0.9	4.3±0.8	4.5±0.7	-
Latitude (°)	61.5±0.5	62.3±0.8	63.8±1.2	62.0±1.2	61.3±0.4	62.0±0.8
Longitude (°)	25.3±1.0	25.9±1.9	27.2±2.2	25.0±1.8	25.3±1.1	25.0±1.9
Avg. temp. (°C)	4.9±1.1	3.1±1.2	2.3±1.3	3.6±1.1	3.8±1.6	4.4±0.8
Jan. temp. (°C)	-5.8±1.9	-7.0±4.8	-9.8±4.7	-6.1±3.6	-7.5±7.4	-7.2±2.5
July temp. (°C)	16.2±0.3	16.0±1.7	15.5±1.6	15.8±1.6	16.9±2.0	16.3±0.6
Prob. of temp.	0.63±0.03	0.52±0.12	0.51±0.12	0.52±0.10	0.50±0.16	0.64±0.04
Precip. (<i>mm</i>)	663±43	653±94	577±102	646±100	639±87	661±63
Prob. of precip.	0.54±0.08	0.50±0.08	0.49±0.07	0.52±0.08	0.55±0.06	0.52±0.08

Table A.4: *The means and standard deviations of the concentration and needle mass measurements, geographical positions of the clusters in Finland and the weather variables for pine.*

Cluster	2	4	5	6	7	8
N (<i>mg/g</i>)	13.6±0.9	12.5±0.7	9.7±0.7	11.4±1.1	13.6±0.7	11.8±1.0
S (<i>mg/g</i>)	1.05±0.09	0.97±0.08	0.86±0.09	0.94±0.10	1.18±0.10	0.89±0.08
P (<i>mg/g</i>)	1.76±0.16	1.51±0.14	1.58±0.12	1.75±0.17	2.04±0.13	1.43±0.12
Ca (<i>mg/g</i>)	2.38±0.40	2.20±0.40	2.96±1.02	1.80±0.31	2.36±0.30	1.97±0.40
Mg (<i>mg/g</i>)	1.14±0.15	1.06±0.09	1.12±0.15	1.05±0.15	1.24±0.19	1.06±0.13
K (<i>mg/g</i>)	5.61±0.45	5.10±0.56	6.45±0.62	5.58±1.02	5.87±0.45	5.09±0.53
Zn (<i>μg/g</i>)	47±6	43±5	36±8	42±4	48±5	40±6
Mn (<i>μg/g</i>)	432±100	447±104	495±147	310±112	403±84	417±124
Fe (<i>μg/g</i>)	41±11	45±20	28±14	43±21	50±16	32±8
Cu (<i>μg/g</i>)	3.27±0.61	2.92±0.57	1.70±0.54	3.11±0.15	4.02±1.07	2.78±0.58
Al (<i>μg/g</i>)	345±95	235±84	101±86	258±45	333±114	235±72
B (<i>μg/g</i>)	12±4	13±3	12±2	10±5	15±4	12±3
NM (<i>g/1000</i>)	12.6±2.8	8.9±1.5	5.2±1.4	8.6±1.7	9.6±1.7	11.1±2.6
Latitude (°)	65.0±3.2	63.6±2.7	66.2±0.4	67.3±1.3	66.4±3.7	64.7±2.6
Longitude (°)	25.5±2.9	25.4±2.4	28.2±1.3	27.6±0.8	27.2±2.8	26.1±2.1
Avg. temp. (°C)	2.1±2.8	2.5±2.4	-0.3±0.9	-1.7±1.4	0.0±2.5	2.0±2.1
Jan. temp. (°C)	-7.8±5.4	-10.0±7.5	-13.3±4.8	-21.0±2.7	-12.0±4.0	-8.9±4.4
July temp. (°C)	15.2±2.1	15.5±2.1	13.8±1.6	13.0±1.5	15.0±3.2	15.2±1.7
Prob. of temp.	0.55±0.09	0.51±0.15	0.48±0.13	0.40±0.14	0.46±0.13	0.55±0.11
Precip. (<i>mm</i>)	582±123	607±116	622±153	460±133	548±163	608±114
Prob. of precip.	0.53±0.09	0.52±0.08	0.51±0.12	0.41±0.09	0.52±0.11	0.54±0.10

Table A.5: *The permutation test results P_{pt} for the clusters of Finland using temperature and precipitation measurements and their probabilities as the tested measurements.*

	Cluster					
	2	4	5	6	7	8
Complete data						
Average temperature	0.96	0.05	1.00	0	0.67	1.00
Temperature in January	0.08	0.16	1.00	0	0.92	0.91
Temperature in July	0.89	0.02	1.00	0.35	0	0.98
Precipitation	1.00	0.03	0.98	0.12	0.85	0.86
Prob. of temperature	0	0.98	1.00	0.98	1.00	0
Prob. of precipitation	0.23	0.96	0.95	1.00	0.01	0.01
Spruce						
Average temperature	0	0.01	0.12	0	0.05	0
Temperature in January	0.97	0.33	0.39	0.86	0.11	0.86
Temperature in July	1.00	1.00	0.97	0.84	1.00	0.99
Prob. of temperature	0.17	0.02	0.80	0.75	0.29	0.20
Precipitation	0.03	0.69	0	0.98	0.53	0
Prob. of precipitation	0	0.50	0.03	0	0.06	0.16
Pine						
Average temperature	0.56	0.02	0.44	0.12	0.96	0.61
Temperature in January	0.09	0.93	0.11	0.98	0.51	0.77
Temperature in July	1.00	1.00	1.00	1.00	0.18	0.93
Prob. of temperature	1.00	1.00	1.00	1.00	1.00	1.00
Precipitation	1.00	1.00	0.80	1.00	1.00	0.62
Prob. of precipitation	0.72	0.34	0.64	0.05	0.47	0.05

Bibliography

- [1] S. Luyssaert, H. Raitio, and A. Fürst. Forest nutrition at the Finnish and Austrian level I plots in 1987-2000. Technical report, The Finnish Forest Research Institute and Austrian Federal Office Research Centre for Forests, 2003. To appear.
- [2] Brian D. Ripley. *Spatial Statistics*. Wiley Series in Probability And Mathematical Statistics. John Wiley & Sons, 1981.
- [3] Gary L. Gaile and Cort J. Willmott, editors. *Spatial Statistics And Models*, volume 40. D. Reidel Publishing Company, 1984.
- [4] Graham Upton and Bernard Fingleton. *Spatial Data Analysis by Example*, volume 1: Point Pattern and Quantitative Data of *Wiley Series in probability and statistics*. John Wiley & Sons, 1985.
- [5] Brian D. Ripley. *Statistical inference for spatial processes*. Cambridge University Press, 1988.
- [6] Pierre Goovaerts. *Geostatistics for Natural Resources Evaluation*. Applied Geostatistics. Oxford University Press, 1997.
- [7] Sergios Theodoridis and Konstantinos Koutroumbas. *Pattern Recognition*. Academic Press, 1999.
- [8] Martin Bogdan, Alexei Babanine, Jörg Kaniecki, and Wolfgang Rosenstiel. Nerve signal processing using artificial neural nets. In *International Workshop on Information Processing in Cells and Tissues (IPCAT 95)*, pages 55–68, 1995.

- [9] Martin Bogdan and Wolfgang Rosenstiel. Detection of cluster in self-organizing maps for controlling a prostheses using nerve signals. In *Proceedings of the 9th European Symposium on Artificial Neural Networks (ESANN 2001)*, pages 131–136, April 2001.
- [10] Aoying Zhou, Weining Qian, Hailei Qian, Jin Wen, Shuigeng Zhou, and Ye Fan. A hybrid approach to clustering in very large databases. In David Wai-Lok Cheung, Graham J. Williams, and Qing Li, editors, *Advances in Knowledge Discovery and Data Mining. 5th Pacific-Asia Conference, PAKDD 2001*, volume 2035 of *Lecture Notes in Computer Science*, pages 519–524. Springer, April 2001.
- [11] Jaakko Hollmén, Volker Tresp, and Olli Simula. A self-organizing map algorithm for clustering probabilistic models. In *Proceedings of the Ninth International Conference on Artificial Neural Networks (ICANN'99)*, volume 2, pages 946–951. IEE, 1999.
- [12] Teuvo Kohonen. Automatic formation of topological maps of patterns in a self-organizing system. In Erkki Oja and Olli Simula, editors, *Proceedings of The Second Scandinavian Conference on Image Analysis*, pages 214–220, 1981.
- [13] Teuvo Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990.
- [14] Hubert Hasenauer and Dieter Merkl. Forest tree mortality simulation in uneven-aged stands using connectionist networks. In *Proceedings of the International Conference on Engineering Applications of Neural Networks (EANN'97)*, pages 341–348, Stockholm, Sweden, June 1997.
- [15] Teuvo Kohonen. *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer, Berlin, third edition, 2001.
- [16] Juha Vesanto and Esa Alhoniemi. Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*, 11(3):586–600, May 2000.

- [17] M. Cottrell, E. de Bodt, and M. Verleysen. A statistical tool to assess the reliability of self-organizing maps. In Nigel Allinson, Hujun Yin, Lesley Allinson, and Jon Slack, editors, *Advances in Self-Organizing Maps, Proceedings of the Workshop on Self-Organizing Maps 2001*, pages 7–14. Springer, June 2001.
- [18] A. Hämmäläinen. A measure of disorder for the self-organizing map. In *Proceedings of International Conference on Neural Networks (ICNN) 94*, pages 659–664, 1994.
- [19] Jarkko Venna and Samuel Kaski. Neighborhood preservation in nonlinear projection methods: an experimental study. In *Proceedings of International Conference on Artificial Neural Networks (ICANN) 2001*, 2001.
- [20] T. Villmann, R. Der, M. Herrmann, and T.M. Martinetz. Topology preservation in self-organizing feature maps: Exact definition and measurement. *IEEE Transactions on Neural Networks*, 8(2):256–266, 1997.
- [21] Th. Villmann, R. Der, and Th. Martinetz. A new quantitative measure of topology preservation in Kohonen’s feature maps. In *Proc of International Conference on Neural Networks (ICNN) 94*, pages 645–648, 1994.
- [22] S. Kaski and K. Lagus. Comparing self-organizing maps. In *Proceedings of International Conference on Artificial Neural Networks (ICANN) 96*, pages 809 – 814, 1996.
- [23] Kimmo Kiviluoto. Topology preservation in self-organizing maps. In *Proceedings of the International Conference on Neural Networks (ICNN’96)*, volume 1, pages 294–299, Piscataway, New Jersey, USA, June 1996. IEEE Neural Networks Council.
- [24] Juha Vesanto and Mika Sulkava. Distance matrix based clustering of the self-organizing map. In *Proceedings of the International Conference on Artificial Neural Networks - ICANN 2002*, Lecture Notes in Computer Science, No. 2415, pages 951–956. Springer-Verlag, 2002.

- [25] A. Ultsch and H. P. Siemon. Kohonen’s Self Organizing Feature Maps for Exploratory Data Analysis. In *Proceedings of International Neural Network Conference (INNC’90)*, pages 305–308, Dordrecht, Netherlands, 1990. Kluwer.
- [26] Jiawei Han and Micheline Kamber. *Data Mining: Concepts and Techniques*. The Morgan Kaufmann Series in Data Mining Systems. Morgan Kaufmann, 2001.
- [27] George Karypis, Eui-Hong (Sam) Han, and Vipin Kumar. Chameleon: Hierarchical clustering using dynamic modeling. *Computer*, 32(8):68–75, August 1999.
- [28] Manoranjan Dash and Huan Liu. Efficient hierarchical clustering algorithms using partially overlapping partitions. In David Wai-Lok Cheung, Graham J. Williams, and Qing Li, editors, *Advances in Knowledge Discovery and Data Mining. 5th Pacific-Asia Conference, PAKDD 2001*, volume 2035 of *Lecture Notes in Computer Science*, pages 495–506. Springer, April 2001.
- [29] Phillip Good. *Permutation Tests — A Practical Guide to Resampling Methods for Testing Hypotheses*. Springer Series in Statistics. Springer Verlag, second edition, 2000.
- [30] David F. Parkhurst. Statistical significance tests: Equivalence and reverse tests should reduce misinterpretation. *BioScience*, 51(12):1051–1057, December 2001.
- [31] James D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.
- [32] Richard E. Quandt. The estimation of parameters of linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, 53:873–880, 1958.
- [33] Richard E. Quandt. A new approach to estimating switching regressions. *Journal of the American Statistical Association*, 67(338):306–310, 1972.
- [34] R. H. Shumway and D. S. Stoffer. Dynamic linear models with switching. *Journal of the American Statistical Association*, 86(415):763–769, 1991.

- [35] Chang-Jin Kim. Dynamical linear models with Markov-switching. *Journal of Econometrics*, 60:1–22, 1994.
- [36] V. Kadiramanathan. Recursive nonlinear identification using multiple model algorithm. In *Proceedings of the Fifth IEEE Workshop on Neural Networks for Signal Processing (NNSP'95)*, pages 171–180, 1995.
- [37] Andreas S. Weigend. Nonlinear gated experts for time series: Discovering regimes and avoiding overfitting. *International Journal of Neural Systems*, 6(4):373–399, 1995.
- [38] Merrilee Hurn, Ana Justel, and Christian P. Robert. Estimating mixtures of regressions. *Journal of Computational and Graphical Statistics*, 12:55–79, 2003.
- [39] Hans-Martin Krolzig. Predicting Markov-switching vector autoregressive processes. Technical report, Department of Economics and Nuffield College, Oxford, 2000.
- [40] James D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384, 1989.
- [41] James D. Hamilton. Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45:39–70, 1990.
- [42] Michael P. Clements and Hans-Martin Krolzig. A comparison of the forecast performance of Markov-switching and threshold autoregressive models of us gnp. *The Econometrics Journal*, 1:C47–C75, 1998.
- [43] Colin Lizieri, Steven Satchell, Elaine Worzala, and Robert Dacco. Real interest regimes and real estate performance: A comparison of U.K. and U.S. markets. *The Journal of Real Estate Research*, 16(3):339–356, 1998.
- [44] Robert Dacco and Steve Satchell. Why do regime-switching models forecast so badly? *Journal of Forecasting*, 18(1):1–16, 1999.

- [45] Simon M. Potter. Nonlinear time series modelling: An introduction. Technical report, Federal Reserve Bank of New York, 1999.
- [46] Padhraic Smyth, David Heckerman, and Michael I. Jordan. Probabilistic independence networks for hidden Markov probability models. *Neural Computation*, 9(2):227–269, 1997.
- [47] Leonard E. Baum. An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. *Inequalities*, 3:1–8, 1972.
- [48] Lawrence R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.
- [49] B.H. Juang and L.R. Rabiner. Hidden Markov models for speech recognition. *Technometrics*, 33(3):251–272, 1991.
- [50] Yoshua Bengio. Markovian models for sequential data. *Neural Computing Surveys*, 2:129–162, 1999.
- [51] Padhraic Smyth. Hidden Markov models for fault detection in dynamic systems. *Pattern Recognition*, 27(1):149–164, 1994.
- [52] Jaakko Hollmén. *User Profiling and Classification for Fraud Detection in Mobile Communications Networks*. PhD thesis, Helsinki University of Technology, 2000.
- [53] Sam Roweis. Constrained hidden Markov models. In *Advances in Neural Information Processing Systems: Proceedings of the 1999 Conference (NIPS'12)*, pages 782–788. MIT Press, 1999.
- [54] Jaakko Hollmén and Volker Tresp. Hidden Markov model for metric and event-based data. In *Proceedings of EUSIPCO 2000 — X European Signal Processing Conference*, volume II, pages 737–740, 2000.
- [55] Alan B. Poritz. Hidden Markov models: A guided tour. In *Proceedings of the IEEE International conference of Acoustics, Speech and Signal Processing (ICASSP'88)*, pages 7–13, 1988.

- [56] Arthur Nádas and Robert L. Mercer. Hidden Markov models and some connections with artificial neural nets. In P. Smolensky, M. C. Mozer, and D. E. Rumelhart, editors, *Mathematical Perspectives on Neural Networks*, chapter 17, pages 603–650. Lawrence Erlbaum, 1996.
- [57] A. P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39:1–38, 1977.
- [58] C.F. Jeff Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11(1):95–103, 1983.
- [59] S.E. Levinson, L.R. Rabiner, and M.M. Sondhi. An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition. *The Bell System Technical Journal*, 62(4):1035–1074, 1983.
- [60] Geoffrey J. McLachlan. *The EM Algorithm and Extensions*. Wiley & Sons, 1996.
- [61] M. Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society, Series B (Methodological)*, 36(2):111–147, 1974.
- [62] Ron Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, IJCAI 95*, volume 2, pages 1137–1145, Montréal, Canada, August 1995. Morgan Kaufmann.
- [63] P. P. B. Eggermont and V. N. LaRiggia. *Maximum Penalized Likelihood Estimation*, volume I: Density Estimation of *Springer Series in Statistics*. Springer Verlag, 2001.
- [64] B.-H. Juang and L.R. Rabiner. A probabilistic distance measure for hidden Markov models. *AT&T Technical Journal*, 64(2):391–408, February 1985.

- [65] Juha Vesanto, Johan Himberg, Esa Alhoniemi, and Juha Parhankangas. SOM Toolbox for Matlab 5. Report A57, Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo, Finland, 2000.
- [66] Sebastiaan Luyssaert, Mika Sulkava, Jaakko Hollmén, and Hannu Raitio. Nutrition profiles and their usage to evaluate forest nutrition based on large-scale foliar surveys. Manuscript submitted to *Tree Physiology*, 2003.
- [67] Esa Alhoniemi, Johan Himberg, and Juha Vesanto. Probabilistic measures for responses of self-organizing map units. In *Proceedings of the International ICSC Congress on Computational Intelligence Methods and Applications (CIMA '99)*, pages 286–290. ICSC Academic Press, 1999.