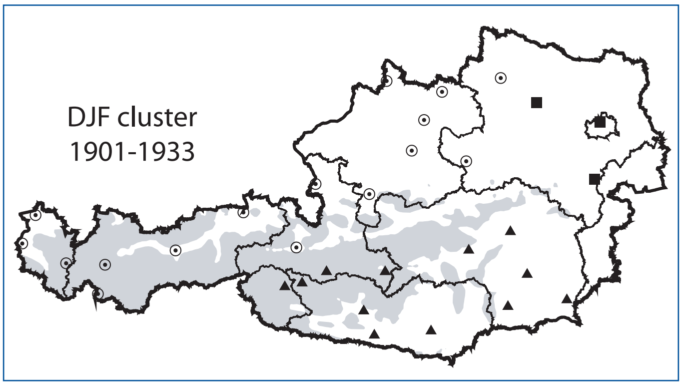
GKSS FORSCHUNGSZENTRUM in der HELMHOLTZ GEMEINSCHAFT

Detection of homogeneous precipitation regions in Austria during the 20th century



Autoren:

C. Matulla
E. K. Penlap
P. Haas
H. Formayer

ISSN 0344-9629

GKSS 2003/8

GKSS 2003/8

Detection of homogeneous precipitation regions in Austria during the 20th century

Authors:

C. Matulla
(Institute of Meteorology and Physics, University of Natural Resources and Applied Life Sciences, Vienna, Austria, and GKSS Research Centre, Institute for Coastal Research, Geesthacht, Germany)

E. K. Penlap
(Atmospheric Sciences Lab., Dep. of Physics, Faculty of Sciences, University of Yaounde I, Cameroon, and GKSS Research Centre, Institute for Coastal Research, Geesthacht, Germany)

P. Haas
H. Formayer
(Institute of Meteorology and Physics, University of Natural Resources and Applied Life Sciences, Vienna, Austria)

GKSS-Forschungszentrum Geesthacht GmbH • Geesthacht • 2003

Corresponding Author:
Dr. Christoph Matulla
current email: matulla@gkss.de

Die Berichte der GKSS werden kostenlos abgegeben.
The delivery of the GKSS reports is free of charge.

Anforderungen/Requests:

GKSS-Forschungszentrum Geesthacht GmbH
Bibliothek/Library
Postfach 11 60
D-21494 Geesthacht
Germany
Fax.: (49) 04152/871717

ISSN 0344-9629

GKSS-Forschungszentrum Geesthacht GmbH · Telefon (04152)87-0
Max-Planck-Straße · D-21502 Geesthacht / Postfach 11 60 · D-21494 Geesthacht

GKSS 2003/8

Detection of homogeneous precipitation regions in Austria during the 20th century

Christoph Matulla, Edouard K. Penlap, Patrick Haas, Herbert Formayer

27 pages with 8 figures and 3 tables

Abstract

Recently a discussion about the benefit of using homogeneous precipitation regions in downscaling and impact studies was stimulated. Up to now it is believed that the application of statistical methods to homogeneous regions generally improves the performance of the downscaling. However, assessing homogeneous precipitation regions is also useful for climatological and forecasting models, detection of climatic fluctuations and station network design. In this study we investigate the usefulness of three different statistical methods to detect homogeneous precipitation regions in Austria, during three different periods covering the 20th century. Hence, the study contributes to a better understanding of Austria's precipitation climate during the 20th century and shows the usefulness of three different statistical methods (Rotated Empirical Orthogonalfunctions, Cluster Analysis and Artificial Neural Nets) in detecting homogeneous regions.

Homogene Niederschlagsregionen in Österreich während des 20. Jahrhunderts

Zusammenfassung

Die Leistungssteigerung von Downscaling durch das Vorschalten eines Verfahrens, das in den Beobachtungsdaten homogene Regionen detektiert, ist in Diskussion. Generell kann man wohl davon ausgehen, dass es zu einer Verbesserung der Ergebnisse führt. Neben dieser Anwendung ist das Auffinden von homogenen Regionen auch für andere Zwecke nützlich, beispielsweise zum Testen von Klima- und Vorhersagemodellen, zur Detektion von Klimaschwankungen oder Klimawandel und zum Design von Stationsnetzen. Hier suchen wir mit Hilfe von drei statistischen Verfahren (Rotated Empirical Orthogonal Functions, Cluster Analysis und Artificial Neural Nets) im 20. Jahrhundert homogene Niederschlagsregionen in Österreich auszugrenzen und hoffen so, zu einem besseren Verständnis des Niederschlagsklimas in Österreich beitragen zu können.

Manuscript received / Manuskripteingang in TDB: 21. März 2003

1 Introduction

Precipitation is one of the main climatic elements and its distribution can be, particularly over complex terrain, highly variable in space and time. Long term fluctuations of precipitation affect the composition of vegetation (Lexer et al. 2002) directly. Hence, recording of past, and simulation of future variations and changes of homogeneous precipitation regions are principal tasks of climate research.

This work pursues two aims. First to illustrate objective methods to detect homogeneous precipitation regions and thereby to contribute to a better understanding of the spatial and seasonal variability of precipitation climate in Austria. Second, to study precipitation behavior during the 20th century. These goals appear to be of value since Austrias precipitation climate is marked by complicated patterns of spatial and seasonal variability (Auer 1993).

Austria's landscape is complex. It ranges from basins with small altitudes of only a few hundred metres to high mountains from 2000 up to more than 3500 metres and deep valleys along the Alpine chain.

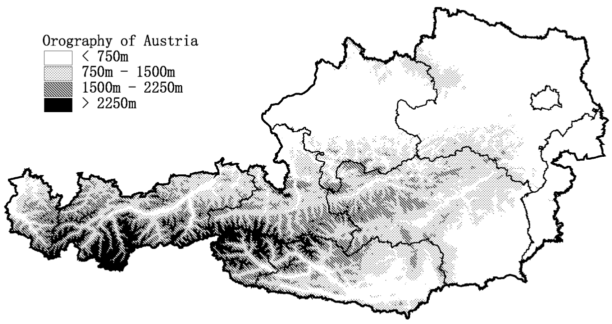


Figure 1: Orography of Austria

The weather in Austria is dominated by three main airflows originating from the Atlantic, from the Mediterranean and from Eastern Europe, respectively. The Alpine crest seperates Austria into several climatic provinces. The northern and central Alps form a barrier for the northwestern airflow. From Vorarlberg across the Bavarian Alps to the Salzburger Alps, mean annual precipitation totals reach values of more than 3000mm (Schwab et al. 2001) having its maximum, as in all parts of Austria, during summer. The precipitation totals decrease eastwards, and inner alpine valleys like the Inntal and the Ötztal, located at the lee side, are not so wet. The eastern parts of Austria including the Mühl-, Wald- and Weinviertel, as well as the Burgenland and areas in the Steiermark, are relatively dry, with mean annual totals from less than 500 to 700mm. The climatic impact in this region is conditioned by the continent and shows large differences between summer and winter season. Areas south of the main alpine peaks such as the Klagenfurter valley and the east-styrian uplands receive higher amounts of precipitation than the eastern region. This is mainly due to the advection of humid airmasses from the Adriatic sea during the summer months. In this part of the country the frequency of occurrence of thunderstorms and hail is higher than in any other part of Austria.

To identify different homogeneous regions, precipitation distributions are regionalized using different methods. The application of methods is repeated for disjoint periods during the 20th century. Three samples of equal length were allocated (1901-33, 1934-66, 1967-99). This was done to fulfill the WMO recommendation of using at least 30 year periods within climate analysis (WMO 1992). Moreover this arbitrary choice matches typical research conditions.

Some practical applications of homogeneous precipitation regions are: (i) application in Downscaling and Impact studies (ii) use for climatological and forecasting models (iii) comparison of different statistical methods (iv) detection of climatic fluctuations and (v) station network design.

The manuscript is organized in the following way: The high quality homogenized long term climate dataset called ALOCLIM (Auer et al. 2001), which serves as a basis of this study is introduced in section 2. In section 3 the three different methods used (rotated EOFs, cluster analysis and a ANN technique) as well as the achieved results are presented. Section 4 contains a comparison and an interpretation of the achieved results.

2 Data

The Austrian LOng term CLIMate ALOCLIM (Auer et al. 2001) is one of the first datasets that meets all necessary requirements to describe climate and its variability. Such requirement are: (i) high density and long term station records, (ii) multiple-element datasets and (iii) high data quality in terms of non-climatic inhomogeneities.

All time series entering the ALOCLIM dataset have been investigated for breaks and systematic biases in the long-term station records. This is achieved by a homogenizing procedure (Böhm et al. 2001; Auer et al. 2001) using meta data information and mathematical homogeneity tests. Due to a collaboration of the Austrian weather service with national weather services of most neighboring countries the quality of the ALOCLIM data is homogeneous in space (i.e. not limited by the borders of the Austrian territory). ALOCLIM contains precipitation, temperature, pressure and sunshine duration on a monthly basis.

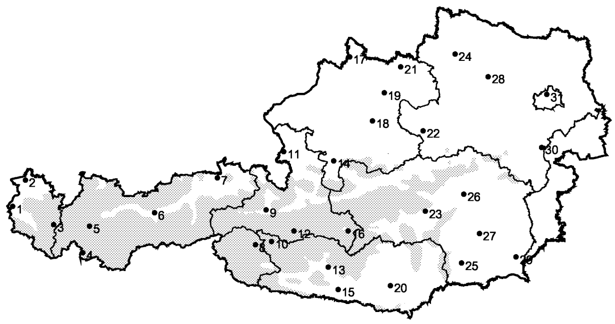


Figure 2: Distribution of the stations used in this study. Shading indicates area above 1200m.

Monthly precipitation totals from 31 stations (see Table 1 and Figure 2 for their spatial distribution), spread over Austria, were extracted for the 20th century. Only 0.6% of the start–data are missing. To simplify matters, they are replaced by their hundred-year means.

The missing data are not uniformly distributed over the stations, e.g. Kals (8) shows a total of about 10% missing data (eight years of missing values in the nineteen-thirties account for the main part of it). Within the Kufstein record (7) four years of data at the beginning of the 20th century are missing. Millstatt (13) on the other hand shows about 2% missing data spread over the whole century. All other twentynine stations have almost no data missing.

Table 1: Stations utilized in this study. Numbers in the first column are used to identify the stations within this work.

Nr.	Name of station	lon.[^{o}]	lat.[^{o}]	alt.[m]
01.	FELDKIRCH	9.60	47.27	439
02.	BREGENZ	9.75	47.50	436
03.	LANGEN	10.12	47.13	1270
04.	NAUDERS	10.50	46.90	1360
05.	LANDECK	10.57	47.13	818
06.	INNSBRUCK-UNIVERSITY	11.38	47.26	577
07.	KUFSTEIN	12.16	47.57	492
08.	KALS	12.65	47.00	1347
09.	ZELL-AM-SEE	12.78	47.30	753
10.	HEILIGENBLUT	12.85	47.03	1242
11.	SALZBURG-AIRPORT	13.00	47.80	430
12.	BAD-GASTEIN	13.13	47.12	1100
13.	MILLSTATT	13.56	46.81	792
14.	BAD-ISCHL	13.63	47.72	469
15.	BAD-BLEIBERG	13.68	46.62	904
16.	TAMSWEG	13.81	47.12	1012
17.	KOLLERSCHLAG	13.84	48.61	725
18.	KREMSMUENSTER	14.13	48.06	383
19.	LINZ	14.19	48.24	298
20.	KLAGENFURT	14.33	46.65	447
21.	FREISTADT	14.50	48.52	548
22.	WAIDHOFEN/YBBS	14.77	47.97	365
23.	SECKAU	14.78	47.28	874
24.	STIFT-ZWETTL	15.20	48.62	506
25.	DEUTSCHLANDSBERG	15.22	46.83	410
26.	BRUCK-MUR	15.27	47.42	489
27.	GRAZ-UNIVERSITY	15.45	47.08	366
28.	KREMS	15.60	48.42	223
29.	BAD-GLEICHENBERG	15.90	46.87	303
30.	WIENER-NEUSTADT	16.27	47.80	270
31.	WIEN-HOHE-WARTE	16.36	48.25	203

3 Methods and Results

Three techniques are used in order to group the stations into homogeneous precipitation regions. They are outlined in this section. The mathematics of the methods are not introduced, as they are given in various statistical publications (c.f. von Storch and Zwiers 1999; Haykin 1994). However, the main ideas are presented.

The utilized methods are: (i) a Principal Component Analysis (PCA) performed on seasonal correlation matrices of precipitation data followed by a varimax rotation (REOF), (ii) a cluster analysis (CLRA) and (iii) 'competitive learning', which is an unsupervised learning procedure named of artificial neural networks (ANN).

3.1 Rotated Empirical Orthogonal Functions (REOFs)

Rotation was introduced to meteorology by Richman (1986) and its goal is to derive simple but meaningful patterns. Ehrendorfer (1987) used REOFs to identify homogeneous regions for summer and winter half-years from 1951 to 1980 in Austria. He utilized networks with somewhat less than 30 stations and found three precipitation regions of Austria for both, winter and summer half-year. Widmann (1996) regionalized Swiss precipitation from 1961-1990 and Alpine precipitation from 1978-1991 using REOFs.

Principal component analysis (PCA) is used to identify a low dimensional subspace of the original data-space that contains most of its variability. This subspace is spanned by the leading eigenvectors, named Empirical Orthogonal Functions (EOFs). The rotation procedure usually follows a PCA in order to separate the noise from the signal.

The desired 'simple' patterns, are obtained by applying an orthonormal transformation to the EOFs. The required transformation solves a variational problem, which minimizes a cost function (von Storch and Zwiers 1999). The form of the cost function characterizes the shape of the REOFs.

Simplicity can be achieved for the REOFs or their time coefficients but not for both at the same time, hence, the REOFs can be orthogonal or the coefficients can be uncorrelated. Within this study the so-called 'varimax' method (Richman 1986) is used to determine the form of the cost function. Figure 3 shows the flow chart of the applied approach.

$\mathbf{X}$ is the random variable (the seasonal totals at the stations). $(\mathbf{X} - \mu)/\sigma$ its standardized form, where μ and σ are the arithmetic mean and standard deviation, respectively.

The first step is to perform a PCA on the standardized random variable i.e. to diagonalize the correlation matrix. The corresponding eigenvectors (EOFs) form an orthogonal basis and their time coefficients are uncorrelated. The resulting eigenvalues are utilized, via the so called 'logarithm of eigenvalue plots' (Preisendorfer 1988), to determine the di-

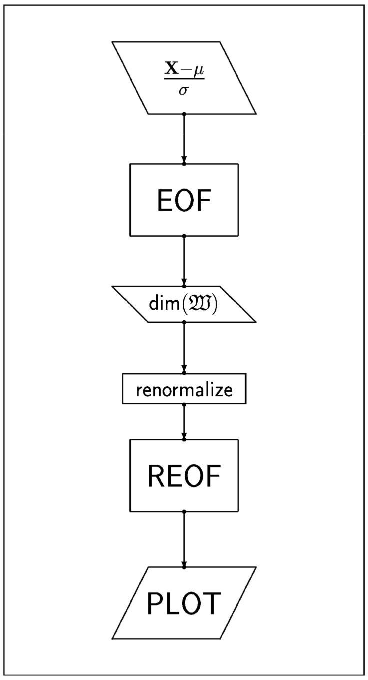


Figure 3: Flow chart of the REOF technique.

mension of the subspace containing the main fraction of variance. This step is symbolized by '$\mathsf{dim}(\mathfrak{W})$' in Figure 3.
In case of DJF the subspaces spanned by the first three (four) EOFs contain more than 75(80)% of the intra-annual variance. During the summer season (JJA) the first four (five) EOFs explain more than 65(70)%, cf. Table 2.
However, the quality of the EOFs declines with increasing index and hence, including more and more EOFs will not solve the problem. von Storch and Hannoschöck (1985) showed that the variance of the eigenvalue estimates is large and biased. In general large eigenvalues are overestimated and small ones are underestimated. These errors become considerably large if the degree of freedom exceeds the sample size. Hence it was decided to take the first three DJF-EOFs and the first four JJA-EOFs. Ehrendorfer (1987), who investigated the period 1951-1980, took three EOFs for both, the winter- and summer-half-year.
Subsequently the varimax-rotation that provides the REOFs is applied. With respect to subsequent applications (e.g. Downscaling) it was decided to keep the time coefficients of the rotated patterns uncorrelated, which implies that the REOFs are not orthogonal. Thus the EOFs and their coefficents have to be renormalized prior to rotation (von Storch and Zwiers 1999).

Stations which share the highest value at the same REOF are combined into one group.

Table 2: Fraction of variance explained by the leading DJF-(JJA)-EOFs. xplvr abbr. explained variance.

xplvr %	season			
period	DJF		JJA	
	3EOFs	4EOFs	4EOFs	5EOFs
1901-1933	77.4	81.9	70.8	74.8
1934-1966	81.1	85.6	74.4	79.0
1967-1999	76.2	81.9	67.4	72.3

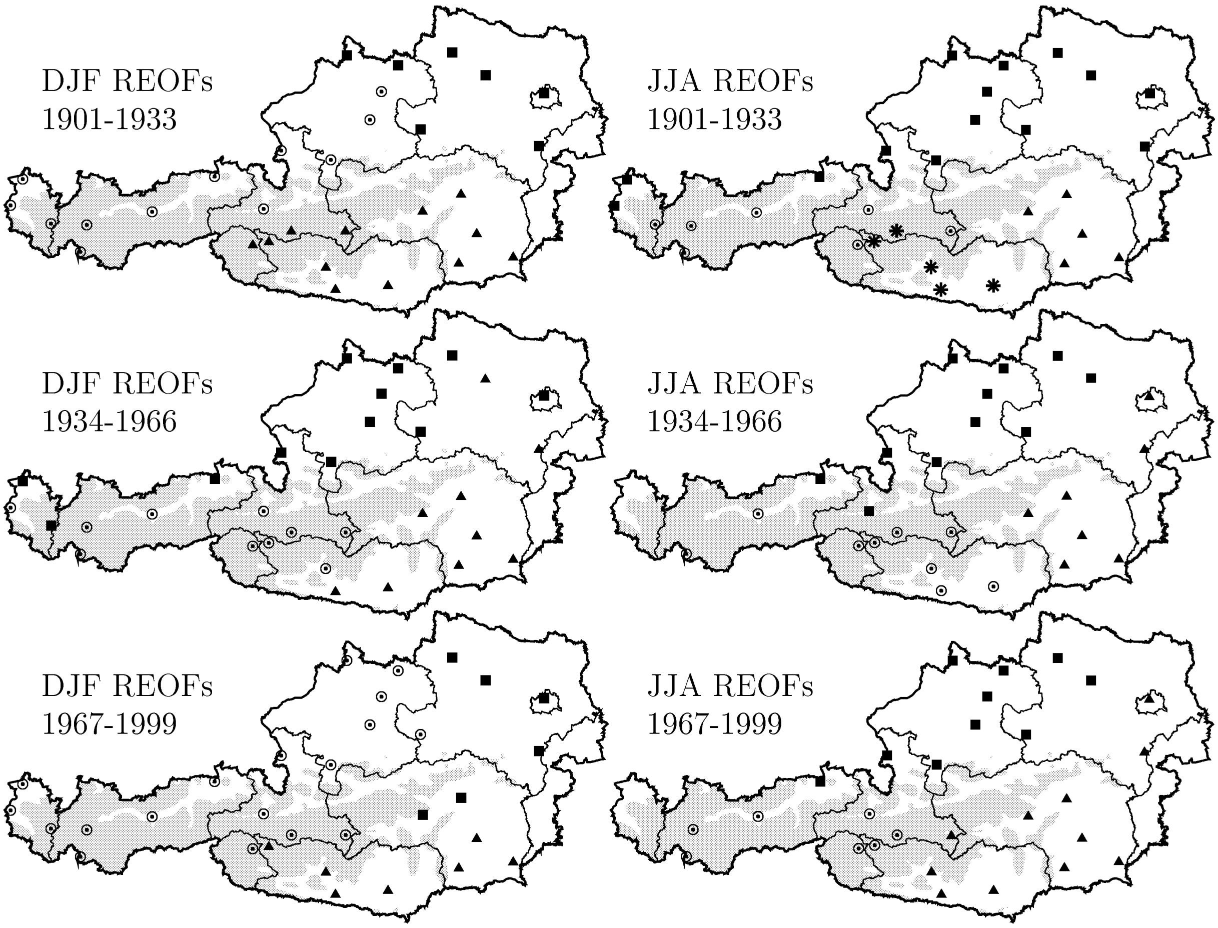


Figure 4: Left (right) hand side: DJF (JJA)-groups formed by three (four) REOFs. Shading indicates area above 1200m.

Hence the maximal number of regions is given by the dimension of the subspace retained, i.e. three (four) for the winter (summer) season. Figure 4 displays the regions found during the winter (left hand side) and the summer seasons (right hand side).

The results are briefly described:

DJF 1901-33: Region (□) lies in the nord-east of Austria. It extends from the Mühl- and the Weinviertel to the Vienna basin. A second region (◯) covers the territory in the north of the main alpine peaks from Vorarlberg to the Innviertel. The third region (△) runs along the alpine chain and includes the stations in its south.

DJF 1934-66: This period (□) combines most stations in the northern plateaus (Wein-, Wald- and Mühlviertel), three stations (7, 11, 14) in Salzburg as well as two in Vorarlberg (2, 3). Another unit (△) extends the southern basins (Grazer and Klagenfurter basin) across station 30 to station 28. The remaining stations (◯) cover large parts of East Tirol and Tirol.

DJF 1967-99: The situation is related to the first period. Only stations along the '1901-33-borders' are different. The most noticeable differences occour along the top line of the alpine chain.

In order to achieve approximately the same fraction of explained variance in summer as in winter it was necessary to take a four dimensional subspace into account. Hence, the results contain one more group.

JJA 1901-33: Region (□) is clearly larger than the others. It includes the northern plateaus, three stations in Salzburg (14, 11, 7) and the two most western stations (1, 2). Region (△) covers the Grazer basin, (∗) the Klagenfurter basin plus two stations (10, 12) and region (◯) the rest of Austria.

JJA 1934-66: The stations in the most western part of Austria form a stand alone unit (+). Stations (□) in and around Salzburg are linked with stations in the Mühl- and the Waldviertel. Stations in the eastern and south-eastern areas, like the Grazer basin, form a group (△).

JJA 1967-99: The distribution is similar to the 1934-66 period. Changes appear along the borders. The region covering the east and south-east lowlands extends now towards the western parts. However stations in Vorarlberg form an own unit (□).

3.2 Cluster Analysis (CLRA)

Woth (2001) used a hierarchial cluster analysis to find homogeneous regions of winter (DJF) precipitation on the Iberian Peninsula and in the south of France. Ramos (2001) investigated autumn and spring precipitation distribution patterns in north-eastern Spain for the period 1889-1999 by applying hierarchial and non hierarchial CLRA techniques. Jackson and Weinand (1994) classified tropical precipitation stations distributed over the globe, using PCA and a clustering procedure. They utilized daily data records of different lengths, ranging from five to more than fifty years.

The purpose of a cluster analysis is to sort objects into clusters according to different aspects. Clusters are units containing objects. Similar to Woth (2001), a hierarchial type of cluster analysis is used to identify different homogeneous precipitation regions.

A hierarchial cluster analysis is situated between two extreme states. One at its beginning, when every object forms a cluster on its own, and another at its end, when all objects are joined into one cluster. Between these extremes the objects are onwardly aggregated into clusters. In this study the objects under investigation are normalized anomalies of seasonal precipitation totals at Austrian stations. Cluster Analysis offers many possibilities of grouping objects. This wide choice corresponds to the ways of answering the following questions:

1. 'how can similarity between the objects/clusters be quantified?'

2. 'when should two objects/clusters be joined into one?'.

Both questions deal with fuzziness - different objects or clusters become indistinguishable with increasing fuzziness. In this work the correlation coefficient was selected to quantify fuzziness and thereby to answer the questions above. In reference to question 1 – a high correlation coefficient between two objects accounts for much similarity between them. In reference to question 2 – two clusters are not joined together until the correlation coefficient between their most dissimilar objects is lower than the considered fuzziness. This intersection-technique is called the 'complete linkage method'. It does not combine different clusters until the Euclidean distance between the most dissimilar objects underruns a given value. The utilization of the correlation coefficient ρ advises the use of normalized anomalies $(\mathbf{X}, \mathbf{Y})$ as objects, mainly because it permits the formulation of the correlation coefficient as a simple function of the distance:

$$\rho\,(\mathbf{X}, \mathbf{Y}) = 1 - \frac{1}{2}\|\mathbf{X} - \mathbf{Y}\|^2$$

An effective cluster analysis should provide clusters inheriting a high degree of inner homogeneity and outer separation, i.e. the correlation coefficients between the objects inside the clusters should be high and between different clusters the corresponding correlation coefficient should be low. Figure 5 displays the strategy of our approach. At the top, the

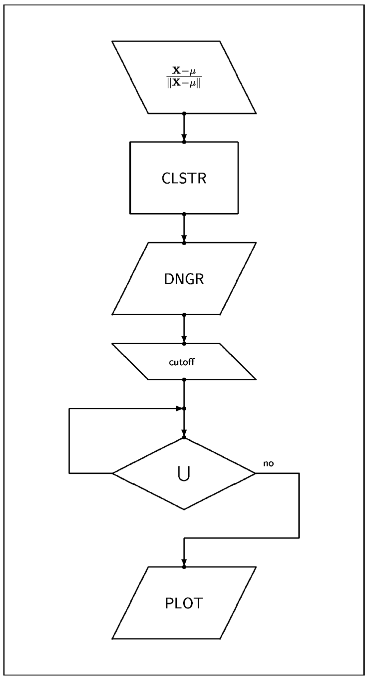


Figure 5: Flow chard of the REOF technique.

normalized time series at the stations enter the cluster analysis. The result of the cluster analysis is diagrammed in a so-called dendogram.

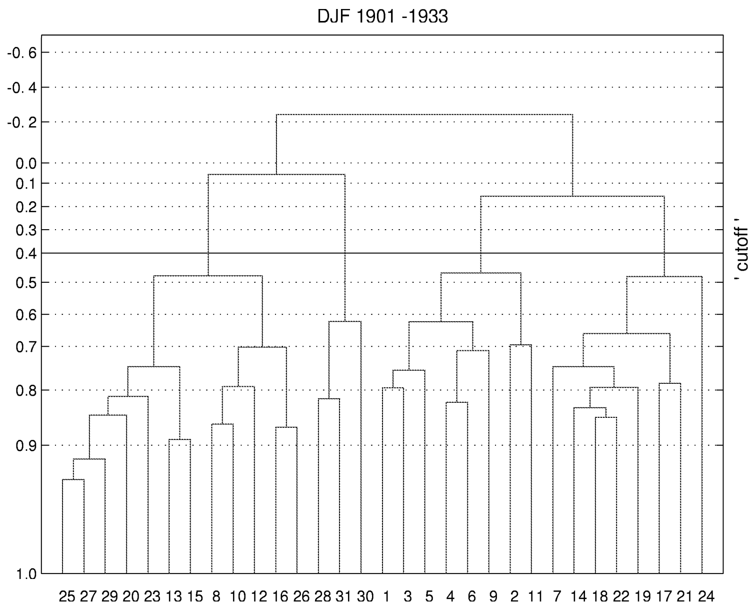


Figure 6: DJF-dendogram for for the period 1901-1933.

Table 3: Assessment of CLRA results for the winter-season. Fischer- and Pearson-correlationcoef. measure the homogeneity inside the clusters. The correlationmatix indicates the similarity between the clusters.

Period	Maß	clstr.1	clstr.2	clstr.3	noclstr
1901-1933	Fischer	1.01	1.01	0.75	0.48
	Pearson	0.77	0.76	0.64	0.45
	Corrmat.	1.00	0.42	0.24	
		0.42	1.00	0.44	
		0.24	0.44	1.00	
1934-1966	Fischer	0.90	0.92	0.91	0.59
	Pearson	0.72	0.72	0.72	0.53
	Corrmat.	1.00	0.58	0.41	
		0.58	1.00	0.71	
		0.41	0.71	1.00	
1967-1999	Fischer	1.12	0.77	0.82	0.48
	Pearson	0.81	0.65	0.68	0.45
	Corrmat.	1.00	0.57	-0.01	
		0.57	1.00	0.50	
		-0.01	0.50	1.00	

An example of a dendogram is given in Figure 6. The abscissa displays the stations and the ordinate the correlation coefficient. At the beginning the correlation coefficient is equal to 1 (i.e. uniqueness). Thus two objects form a cluster only if they are identical. Along the y-axis fuzziness (the correlation coefficient) is increasing (decreasing). Hence the number of clusters is decreasing. At an arbitrary value of the correlation coefficient the aggregation is truncated ('cutoff' in Figure 5) and the effectiveness of the analysis (inner homogeneity and outer separation) is evaluated (e.g. Table 3). If the result is not acceptable a further union of clusters or a cutoff at a higher correlation coefficient is necessary. This iterative process is prompted by the if-condition in Figure 5. Table 3 displays the inner homogeneity and outer distinctiveness (i.e. the quality-measures of the analysis). At the end of the procedure the precipitation regions are plotted (see Figure 7). A drawback of the method is that it critically depends on the initial state. Hence objects which are actually related, however situated in different clusters (at the beginning), may not be combined until the end of the cluster analysis. Even the high diversity of grouping the objects implies a (a priori) nonexistent assumption: knowledge about the data under investigation.

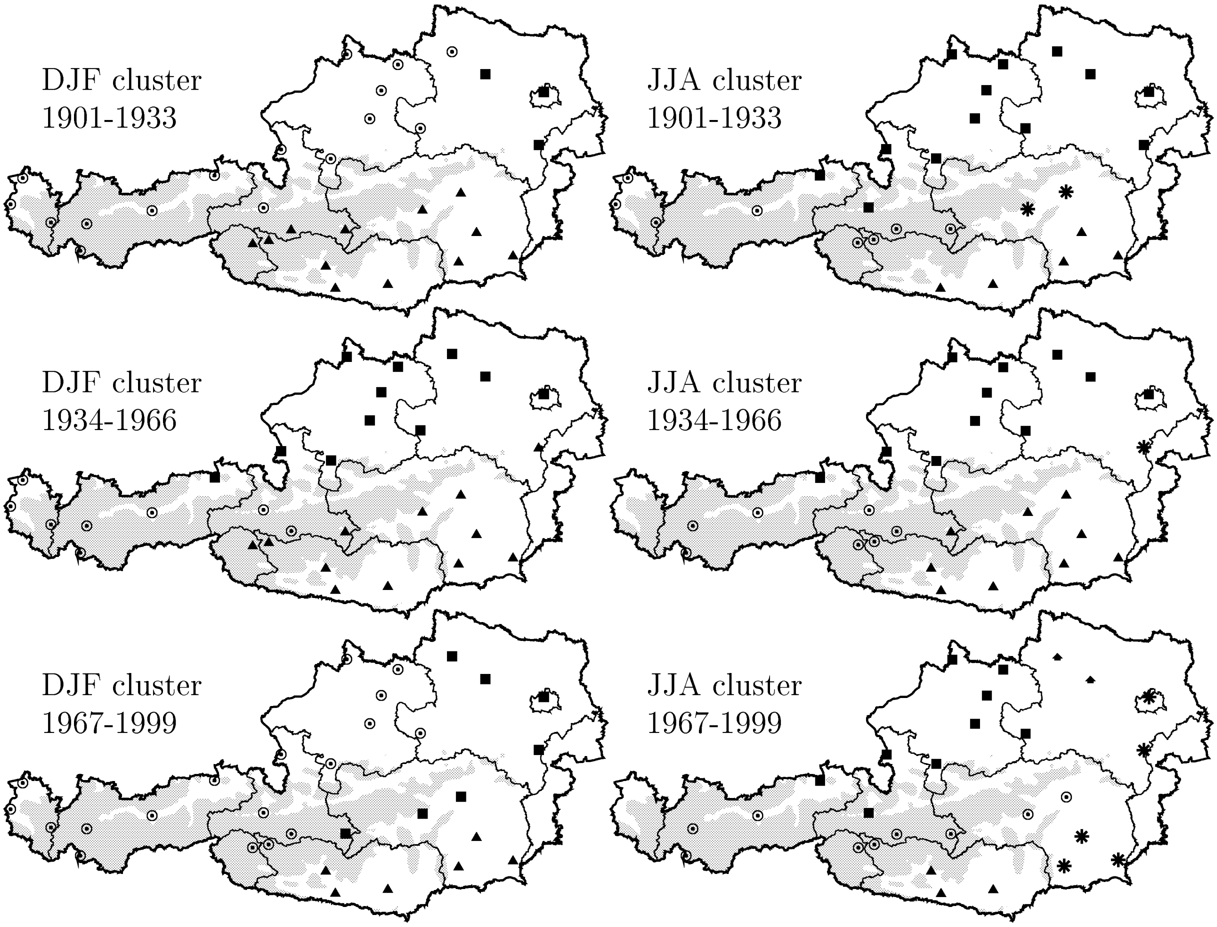


Figure 7: left (right) hand side: DJF (JJA)-precipitation-clusters. derived by hierarchial cluster analysis. Reference periods are dislpayed in the panels. Shading indicates area above 1200m.

As in case of REOF the results are briefly described:
DJF 1901-33: The cluster in the north-eastern part of Austria (□) contains three stations (28, 30, 31). Another cluster (△) stretches from the alpine crest southwards and links Eastern Tirol with the Klagenfurter and Grazer basin. The third cluster (◯) covers the territory in the north of the main alpine peaks from Vorarlberg to the western edge of the Waldviertel.
DJF 1934-66: During this period stations in the northern plateaus (Wein-, Wald- and Mühlviertel) and some stations in Salzburg (7, 11, 14) form an unit (□). Stations at the top line of the alpine chain and in its south (△) are linked together (except for 12 but with 30). The remaining stations in the western part of Austria (◯) were unified with station 9 and 12.
DJF 1967-99: This period reminds of the first period. The western region (◯) is in good

agreement with that of 1901-33. Stations along the alpine crest, which belonged to the southern cluster are now divided between the western (◯) and eastern (□) group. Thus cluster (△) is now restricted to stations in the southern basins.

During the summer-season the situation is more complicated. The measure between inner homogeinity and outer dissimilarity is more unfavorable than in winter, which results in a higher number of clusters here.

JJA 1901-33: The largest cluster is formed by stations in the northern plateaus (□) and stations in and around Salzburg. In the west of Austria one cluster dominates (◯), however station 4 and 5 (+) built an unit on their own. Station 26 and 23 (∗) form a 'mini-cluster' and Millstatt (13) is a stand-alone unit. Stations (△) along the southern border are linked together.

JJA 1934-66: Again, two clusters split the western part of Austria: (◯) and (+). The northern plateaus and parts of Salzburg are united (□) as well as the southern basins with stations along the central alpine peaks (△). Station 30 stands alone.

JJA 1967-99: Clusters containing two or three stations are: (⇑) in the Waldviertel, (△) around the Klagenfurter basin and the stations in Vorarlberg (+). Stations along the main peaks of the alpine crest and some in Tirol (4, 5, 6) are marked by (◯). Cluster (□) contains stations in Salzburg and the Mühlviertel and cluster (∗) those in the north-east to south-east lowlands.

3.3 Self-Organizing networks

Artificial Neural Network (ANN) have been used for a wide range of applications. Foody (1999) for example used ANNs to classify vegetation data in environmental sciences, Michaelides et al. (2001) and Hewitson and Crane (2002) applied them to the climatological domain. Michaelides et al. (2001) used ANNs to classify rainfall variability in Cyprus. He found the ANNs performing more realistic than a cluster analysis.

Self-organising networks are a subgroup of artificial neural networks intended to perform a mapping of arbitrary distributed data (input patterns) on a low dimensional space. More precise, it is an 'unsupervised learning' procedure. There are many types of self-organizing networks applicable to a wide area of problems. Hewitson and Crane (2002) recently used 'self-organizing feature maps' to describe changes of synoptic circulation and discuss in detail its performance and its application within the climatology domain. Here another technique 'competitive learning' (Rumelhart and Zipser 1985), which is realted to 'self-organizing feature maps' is used. This kind of self-organizing networks divides a set of input patterns into groups. Hewitson and Crane (2002) recently used 'self-organizing feature maps', which are related to 'competitive learning', to describe changes of synoptic circulation and discuss in detail its performance and its application within the climatology domain.

The architecture of the neural network used in this paper consists of an input layer containing n nodes and an output layer with n_0 nodes. n corresponds to the dimension of the input vector, which equals in our case to the length of the time series, and n_0 to the desired number of clusters. Each node of the output layer is connected to all nodes of the input layer through the *connection weights*. The input data are generally arranged as a matrix of $m \times n$ dimension where m and n denote the number of observations and variables, respectively. These data are, through an iterative process, assigned to the output nodes. An iteration consists of selecting an observation (input vector) at random, finding its "best matching" output node (the one having the smallest Euclidean distance with the input vector) and updating the *connection weights*. The updating formula is a function of one *learning rate* (η). In total there are two learning rates η and η' ($\eta \gg \eta'$), which are fixed during the whole process. η is used to update the weights of the "best matching" node and η' for all others. η' is used to prevent the situation of only one node being the "best matching" node. If further iterations cause no alteration of the compositions of each node, the learning process is complete. All stations mapped onto the same node form a group.

As input data we used standardized seasonal anomalies of monthly precipitation at the considered stations. They are arranged as a $m \times n$ matrix as previously mentioned.

The connection weights are randomly initialized between -1 and 1 and the learning rate parameters η and η' are fixed to 0.08 and 0.001, respectively. The desired number of clusters are chosen as 3 and 4 for winter and summer, respectively. During the winter seasons convergence is reached after 2000 iterations. In case of the summer seasons 5000 iterations are necessary. Figure 8 shows the regions found by the ANN technique used.

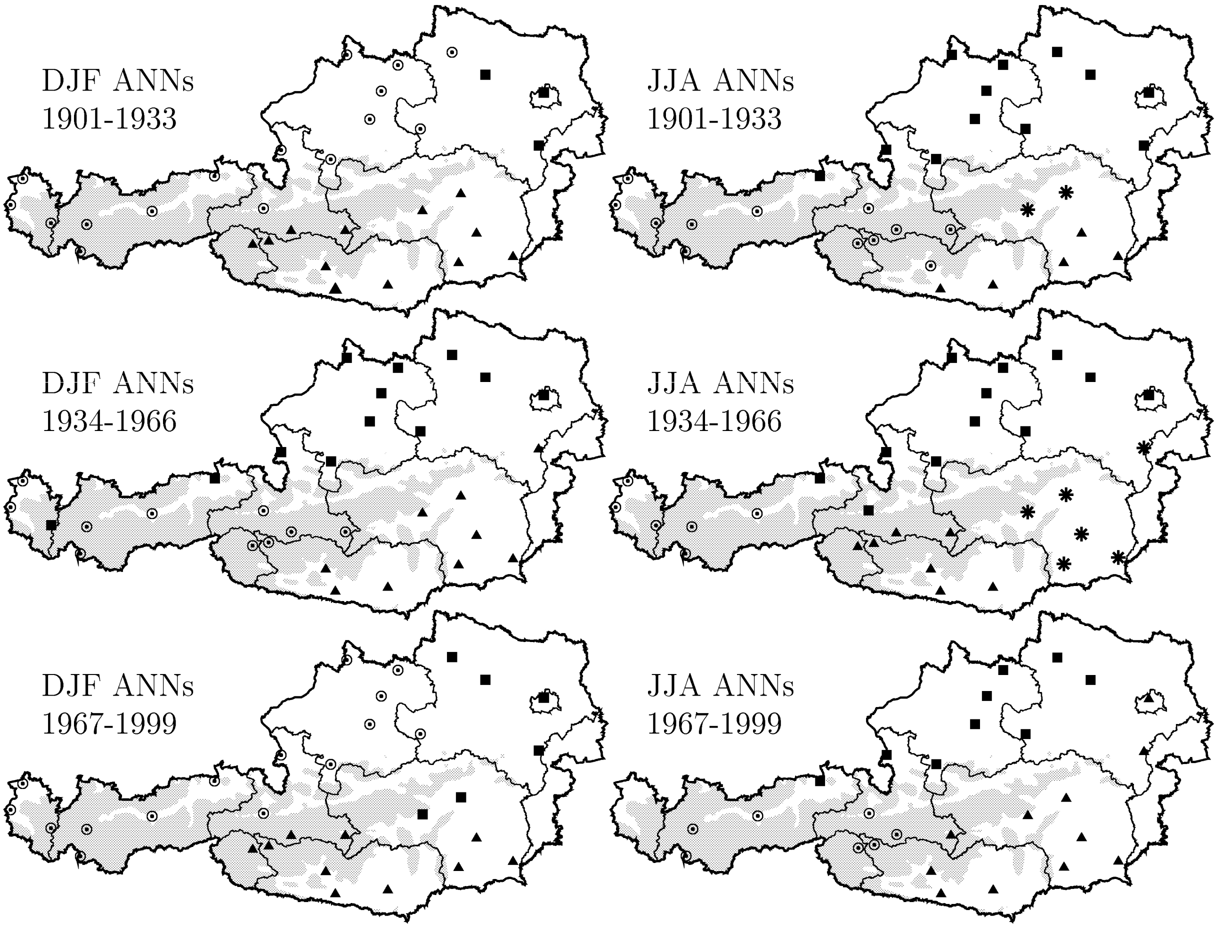


Figure 8: Left hand side DJF clusters, right hand side JJA clusters, found by ANNs. Shading indicates area above 1200m.

DJF 1901-33: During this period, results obtained using the ANN technique are identical to those derived by the cluster analysis. Three regions are found: one in the north-eastern part of Austria (□), one along and southwards from the alpine crest (△) and one runs from Vorarlberg (◯) across Tirol and Salzburg to the western edge of the Waldviertel.

DJF 1934-66: The achieved results are related to both cluster analysis and REOFs. The region, which contains the northern plateaus (Wein-, Wald- and Mühlviertel), some stations in and around Salzburg as well as Langen (station 3) (□) might be a little closer to the one found by cluster analysis. The south-eastern group (△), as well as the one running from Salburg across Tirol to Vorarlberg (◯) are similar to the REOF findings.

DJF 1967-99: Again, the ANN results are alike to the results of the other methods. The group labeled by (□) is identical to the corresponding in the REOF case. The western region (◯) differs only in parts of the main alpine peaks from the others (stations 8, 10, 12) and hence the southern group is in good accordance to the cluster analysis and REOFs.

As previously listed, one more group than in winter is desired. The higher number of iterations needed to achieve convergence during summer might indicate a more complicated situation.

JJA 1901-33: (□) covering the northern plateaus and stations in and around Salzburg is dominating. Regions indicated by ($*$) and (△) located around the Styrian Alps and the southern basins, respectively are identical to the cluster analysis-groups. The west of Austria is joined into one unit (◯).

JJA 1934-66: (□) is almost unaltered and close to the REOF result. Except of Wien Hohe Warte (station 31) the group labeled by ($*$) equals to the corresponding REOF group. The two areas are remaining: one in Kärnten and around its borders (△), the other one in Tirol and Vorarlberg (◯).

JJA 1967-99: The situation is similar to the period from 1934 to 1966 and ist exept from station 17 (Kollerschlag) congruent to the REOF case.

4 DJF synopsis

1901-1933 Findings of cluster analysis and ANNs are identical. Differences to the RE-OFs appear only along the edge between region (□), lying in the the north-east of Austria, and region (◯), covering the territory in the north of the main alpine peaks from Vorarlberg to the Mühlviertel. The third region (△) runs along the alpine chain and includes all stations to its south.

1934-1966 During this period, results of REOFs and ANNs are strongly linked more so than to the results of cluster analysis. All techniques combine stations in the northern plateaus (Wein-, Wald- and Mühlviertel) with stations in and around Salzburg (□). Additionally, ANNs and REOFs show a connection to stations in Vorarlberg. Unit (△) extends from stations in the vincinity of the southern basins (Grazer and Klagenfurter basin) to the north-east. The remaining stations (◯) form a group, which runs from Vorarlberg across Tirol to the southern edge of Salzburg. Noticeable differences between the techniques appear around the Hohe Tauern.

1967-1999 The situation is related to the first period. During this period, results of REOF and cluster analysis are sightly closer to each other than to findings of ANNs. However, differences occur only in the vincinity of the Hohe Tauern. One feature appears worth mentioning – the extension of the north-eastern group (□) into Steiermark, which is found by all methods.

During all investigated episodes three regions can be distinguished. A rough classification might be: one region covers Austrias south, another comprises the stations in its north-east and the third contains the western parts of Austria. The agreement among the methods is satisfying. Two features appear worth mentioning: The large region covering the northern plateaus (□) during the second period (1934-1966) and the combination of the north-eastern part with stations in the Steiermark (1967-1999) are both detected by all methods. The patterns found in the first (1901-1933) and third (1967-1999) episodes share more similarities than to those found in period 1934-1966.

JJA synopsis

1901-1933 REOFs and ANNs detect four groups, cluster analysis six. Similar regions are: (□), which coveres large parts the northern plateaus and extend westwards, (△) combining stations along the southern border of Austria and (◯), running along and north of the alpine crest from Salzburg to Vorarlberg. ANNs and cluster analysis find groups, labeled (△), (∗) and (□), which are (almost) identical.

1934-1966 The northern region (□), as well as the cluster in Austria's south-east, labeled (△) in case of REOFs and cluster analysis or (∗) in the ANNs case, are found by all methods. Moreover REOFs and cluster analysis rearrange stations in Vorarlberg to a unit (+).

1967-1999 REOF findings are almost in perfect congruence with those derived by ANNs. Both methods show four regions: one in Vorarlberg (+), another (◯) stretches from Tirol to Salzburg, the third (△) follows the alpine crest eastwards and covers basins in its souths and the Vienna basin and the last one (□) combines large parts of the northern plateaus. The picture drawn by cluster analysis differs.

The situation during summer is more complex than during winter and the methods show differences. However, similarities outweigh the differences. Three regions are almost always detected: one covering the northern plateaus, one combining areas in Austria's south-eastern part and one forming a unit in Vorarlberg. In detail several groups of station can be found, which are always joined together.

5 Discussion and Conclusions

From a climatological point of view the derived results are reliable and hence the techniques appear suitable for these kinds of problems. However, the fact that the results are reliable also reflects a high data quality.

Findings of all three techniques show the main precipitation modulating effects in Austria. Accordingly, the Alps being a barrier for the main airflows, concerning north and southward advection. The continental influence in the east is captured owing to a quasi-permanent easterly high. During the warm season convective processes dominate the precipitation regime. Independently from the considered periods, all techniques seperate Austria into three regions during winter. Differences emerging between the periods are reproduced by all techniques in a very similar way.

During the first period the winter season shows a 'classical' pattern. The first region covers alpine areas in Vorarlberg and runs eastwards to the western edge of Niederösterreich, the second one combines the alpine chain with basins in its south, and the third region lies in Austrias north-east. In this succession they reflect the Atlantic, Mediterranean and continental influences, respectively. During the second period large fractions of the 1901-1933 western and eastern groups are joined into one. This might be the most striking result of the whole analysis. In order to find the SLP pattern, which is associated with precipitaion in this region during the 20th century, the geographical sector from 50^oW to 30^o and 35^oN to 65^oN is extracted from the northern hemisphere SLP dataset (Trenberth and Paolino Jr. 1980) and investigated by means of PCA and MLR. Demonstrative features found are: (i) a negative anomaly over large parts of Europe having its minimum over Skandinavia, (ii) almost zero values south of Island and a poor developed positive anomaly in the west of Spain. This points to air mass advection from north-west leading to precipitation at the northern border of the entire Alps. However, another region, more clearly influenced by the Atlantic, can be found in the central Alps. The third period only shows a small region influenced by the Mediterranean and a larger one under continental influence. During this period the region influenced by the Atlantic is similar to the first period. Ehrendorfer (1987), uses other data, another period (1951-1980) and even investigates half-years and his result resembles well to our findings.

During the summer season Austria's precipitation pattern is more patchy. Not all regions can be easily related to different air flows or precipitation regimes. Contrary to winter, only three regions are almost similar concerning all techniques. The first one is located in the Central Alps (Tyrolian Alps and Hohe Tauern), an area which is known for low convective precipitation compared to other parts in Austria. The second one covers areas from Salzburg to the Waldviertel in the north of the Alps and the third region is situated in the south-east of Austria. The differences between the periods are of the same magnitude as the differences produced by different techniques within the periods. Thus, a comprehensive meteorological interpretation is not possible.

Taking into account the different rain producing mechanisms it is not surprising, that precipitation during summer, is highly mutable in comparison to winter-precipitation. Precipitation events during winter are mainly triggered by large scale advective processes, with sizable temporal extent. During summer, precipitation is dominated by local scale short-time convective processes. These differences are responsible for the precipitation distribution depending on the seasonal cycle.

The techniques are able to reveal the seasonal dependence of Austria's precipitation patterns. Differences among the methods are larger during summer than during winter. In the case of cluster analysis the differences between summer and winter can directly be seen in the dendrograms. During summer, it is difficult to distinguish clearly between different groups and this leads to a reduced outer separation and inner homogeneity of the final regions. In the case of REOFs, the dimension of the subspace, required to comprise approximately the same fraction of variance as in winter, is higher. Hence, the corresponding patterns are likely to be more complicated than in winter. In the case of ANNs, during winter only one choice of parameters was sufficient for all episodes. During summer these parameters had to be adjusted for each period separately. Hence, differences between the methods are more pronounced in summer than in winter.

To summarize, the results found by the different techniques are in accordance. This is particularly true for the winter season but even during summer, common features can be detected . This fact inspires confidence in the usefulness of the methods.

Cluster Analysis, REOFs and ANNs depend on the objective of the research: CLA offers many possibilities of quantifying similarity, i.e. to measure how two stations are alike and how they differ; REOFs depend on the number of EOFs retained and the rotation method, and in ANNs several parameters (e.g. learning rates, distance weighting, etc.) can be adjusted.

6 Outlook

Results of this work are now being used in downscaling studies. For every region a single downscaling model is established (Woth 2001), to infer information from larger to smaller scales. Further goals might be the application of the presented techniques to: (i) larger amounts of station data, which generally comprise only a short period in comparison to the total period available in this work, but also (ii) to larger regions, as for example the whole alpine area and (iii) detection of homogeneous regions with respect to time.

Acknowledgements
This study was conducted within the research project: 'Usability of different downscaling methods in complex terrain' funded by the Austrian Federal-Ministry of education, science and culture. The Zentralanstalt für Meteorologie und Geodynamik provided the ALOCLIM dataset. We would like to thank H. Matulla, S. Wagner, D. Bray and B. Hünicke for fruitful discussions and for helping us with the manuscript.

References

Auer, I., 1993: *Niederschlagsschwankungen in Österreich.* Number 7 in Österreichische Beiträge zu Meteorologie und Geophysik. Zentralanstalt für Meteorologie und Geodynamik.

Auer, I., R. Böhm, and W. Schöner, 2001: *Austrian long-term climate 1767-2000*, Volume 25 of *Österreichische Beiträge zu Meteorologie und Geophysik.* Central Institute for Meteorology and Geodynamics.

Böhm, R., I. Auer, M. Brunetti, M. Maugeri, T. Nanni, and W. Schöner, 2001: Regional temperature variability in the European Alps: 1760-1998 from homogenized instrumental time series. *Int. J. Climatol.*, **21**, 1779–1801.

Ehrendorfer, M., 1987: A regionalization of Austria's precipitation climate using principal component analysis. *J. Climatology*, **7**, 71–89.

Foody, G.M., 1999: Applications of the self-organising feature map neural network in community data analysis. *Ecological Modelling*, **120**, 97–107.

Haykin, S., 1994: *Neural Networks: A Comprehensive Foundation.* Macmillan College Publishing Company.

Hewitson, B.C., and R.G. Crane, 2002: Self-organizing mags: applications to synoptic climatology. *Clim. Res.*, **22**, 13–26.

Jackson, I.J., and H. Weinand, 1994: Towards a classification of tropical rainfall stations. *Int. J. Climatol.*, **14**, 263–286.

Lexer, M., K. Hönninger, H. Scheifinger, C. Matulla, N. Groll, H. Kromp-Kolb, K. Schaudauer, F. Starlinger, and M. Englisch, 2002: The sensitivity of Austrian forests to scenarios of climate change: a large-scale risk assesment based on a modified gap model and forest inventory data. *Forest Ecology and Management*, **162**, 53–72.

Michaelides, S.C., C.S. Pattichis, and G. Kleovoulou, 2001: Classification of rainfall variability by using Artificial neural networks. *Int. J. Climatol.*, **21**, 1401–1414.

Preisendorfer, R.W., 1988: *Principal Component Analysis in Meteorology and Oceanography.* Seattle: Elsevier, 425pp.

Ramos, M.C., 2001: Divisive and hierarchial clustering techniques to analyse variability of rainfall distribution patterns in a Mediterranean region. *Atmos. Res.*, **57**, 123–138.

Richman, M.B., 1986: Rotation of principal components. *Int. J. Climatol.*, **6**, 293–335.

Rumelhart, D.E., and D. Zipser, 1985: Feature discovery by competitive learning. *Cognitive Science*, **9**, 75–112.

Schwab, M., C. Daly, C. Frei, and C. Schär, 2001: *Hydrologischer Atlas der Schweiz.* Bundesamt für Landestopographie, Wabern - Bern.

Trenberth, K.E., and D.A. Paolino Jr., 1980: The Northern Hemisphere sea-level pressure data set: Trends, errors and discontinuities. *Mon. Wea. Rev.*, **108**, 855–872.

von Storch, H., and G. Hannoschöck, 1985: Statistical aspects of estimated principal vectors (EOFs) based on small sample sizes. *J. Clim. Appl. Meteor*, **24**, 716–724.

von Storch, H., and F. Zwiers, 1999: *Statistical Analysis in Climate Research.* Cambridge University Press, 528 pp.

Widmann, M.L., 1996: *Mesoscale variability and long-term trends of Alpine precipitation and their relation to the synoptic-scale flow.* Ph. D. thesis, ETH Zürich.

WMO (Ed.), 1992: *International meteorological vocabulary*, Volume WMO/OMN/BMO-No.182. Secretariat of the World Meteorological Organization.

Woth, K., 2001: Abschätzung einer zukünftigen Niederschlagsentwicklung mit statistischen Methoden unter Einbezug räumlicher Differenzierungsverfahren am Beispiel des südwesteuropäischen Raums. Master's thesis, GKSS-Report 2001/28 - University of Trier, 101 pp.