Relating the uncertainty of short-term wind speed predictions to meteorological situations with methods from synoptic climatology

Matthias Lange, Detlev Heinemann
Energy Meteorology, Institute of Physics, Carl von Ossietzky University Oldenburg,
26111 Oldenburg, Germany, Tel: ++49-441-798-3577
email: m.lange@mail.uni-oldenburg.de, www.energiemeteorologie.de

Abstract

In order to provide an uncertainty estimate for short-term wind power predictions the accuracy of the underlying wind speed prediction is assessed quantitatively for different meteorological situations. With methods from synoptic climatology an automatic classification scheme is implemented using measurements of wind speed, wind direction and pressure at mean sea level to characterize the local weather conditions at a site. The classification procedure involves principal component analysis to efficiently reduce the data to the most relevant patterns. Cluster analysis is used to group days with similar meteorological conditions into common classes. A comparison of these clusters with weather maps shows that typical weather patterns are successfully captured by the classification scheme. The mean forecast error of the wind speed prediction of the German Weather Service is calculated for each of the clusters. It is found that different meteorological situations have indeed significant differences in the prediction error where the highest rmse can be by a factor 1.3 to 1.6 larger than the smallest rmse. Typically, high uncertainties in the forecast have to be expected in situations where low pressure systems quickly pass north of the site while stationary high pressure situations have smaller forecast errors.

Keywords: short-term prediction, wind power forecast, uncertainty, meteorology

1. Introduction

Wind speed from numerical weather prediction models is the main input into many existing wind power forecasting systems. A valuable information in addition to the power prediction itself is the expected uncertainty of a specific forecast as it allows users to evaluate the risk of an erroneous prediction. To provide this uncertainty estimate the accuracy of the underlying wind speed prediction has to be assessed quantitatively

The meteorological situation over Northern Europe can be very dynamic if a series of low pressure systems is coming in from the Atlantic ocean having frontal zones with high wind speeds. On the other hand if a high pressure system covers large parts of Europe for several days the situation is rather stationary and low wind speeds prevail. However, the performance of numerical weather prediction systems is not equally well for all the different weather situations, e.g. pointed out by Palmer [1]. In particular, in dynamic cases the real situation can evolve quite differently from the one that was numerically predicted. Hence, the challenge is to know in advance how predictable the current meteorological situation is.

The aim of this work is to find a quantitative relation between typical weather situations and the corresponding wind speed prediction error. In contrast to previous works on estimating the uncertainty of wind power prediction [2, 3] this investigation involves more meteorological variables that are closely related to the wind field. Moreover, their diurnal variations are considered to account for changing weather conditions. For this purpose synoptic climatology provides a variety of powerful methods [4, 5]. The technique used here is based on an automatic classification scheme that sorts the prevailing meteorological situation into different classes according to meteorological measurements. The classification scheme uses principal component analysis for efficient data reduction. Groups of days with similar meteorological conditions are then automatically be created by cluster analysis. Finally, for each of these groups the forecast error is determined separately and the differences between the average error values in each class are tested for significance.

2. Principal component analysis of meteorological data

Principal component analysis (PCA) in connection with cluster analysis is a well established method to produce synoptic indices to classify the current weather situation [4, 5]. The idea is to use historical measurement data over at least one year to define a number of weather classes based on the characteristic properties inherently given by the data. PCA is used to decompose the time series of meteorological variables into several eigenmodes that can be ordered according to their relevance for the time series.

The variables used in this investigation are the horizontal wind vector $\vec{u} = (u, v)$ at 30 m height and the air pressure reduced to mean sea level (pmsl). The wind speed data is taken from measurements of the German WMEP programme and the surface pressure data is provided by synoptic stations of the German weather service.

The temperature that is often included to determine the type of air mass [6, 4] is not considered here. This was done because temperature tends to be a dominating variable that requires the data sets to be split into winter and summer part. But doing so would leave only half the data points for the analysis which makes it difficult to have significant results in the end. Moreover, in contrast to the wind vector and the atmospheric pressure the absolute temperature is not directly associated with the wind field dynamics. However, it is desirable for future investigations to include temperature differences at different heights as indicator of the thermal stratification. The importance of considering atmospheric stability for prediction purposes has recently been underlined by Focken et al. [7].

To account for diurnal variations the variables were taken at 0, 6, 12, 18 and 24 hours each day. The data is normalized and written into a matrix M where the columns contain the different variables and the rows the points of time (table 1).

variables →

time ↓

day1	u_0	u_6	u_{12}	u_{18}	u_{24}	v_0	v_6	v_{12}	v_{18}	v_{24}	$pmsl_0$	$pmsl_6$	$pmsl_{12}$	$pmsl_{18}$	$pmsl_{24}$
day2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Table 1 Contruction of data matrix M

If the matrix M is constructed with time series like this the procedure is called p-mode PCA [4]. Note that M is a 365x15 matrix due to 15 variables for each of the 365 days of one year.

The PCA of M is carried out with standard numerical tools resulting in 15 principal components $\vec{p}_i$ with corresponding eigenvalues l_i. The eigenvalue spectrum for a site (Hilkenbrook) in Northern Germany located in flat terrain with rather homogeneous roughness is shown in figure 1. The eigenvalues are normalized to the sum of the eigenvalues and therefore describe the degree of variance the corresponding principal components (PC) contribute to the signal.

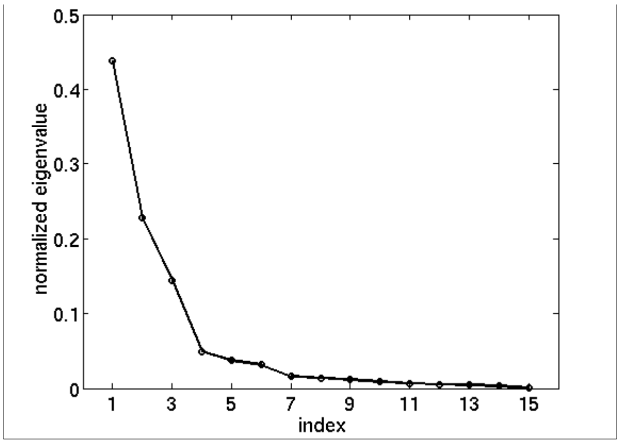

Figure 1 Spectrum of eigenvalues of the PCA of M normalized with sum of eigenvalues. Each eigenvalue describes the degree of variance the corresponding principal component contributes to the signal.

The first 6 PCs explain approx. 90% of the variance and are illustrated for Hilkenbrook in figures 2 (wind components) and figure 3 (pmsl). The first PC describes wind directions from north-east that do not vary much over the day (fig. 2) with corresponding PC of pressure being stationary at a high level (fig. 3). The second PC accounts for wind from north west also with only little changes in strength and direction over the day and still rather high pressure which is slightly rising. The third PC describes wind contributions from the south with slightly more variations in wind speed and decreasing pmsl from high to medium level. However, the fourth PC is the first dynamic one as it describes an approx. 270 degree turn of the wind vector from north to west with strongly changing wind speeds. The corresponding pressure decreases dramatically from high to very low pmsl. The fifth and sixth PC represent 180 degree changes in wind direction with falling and, respectively, rising pressure.

Though the local wind conditions at different sites in Northern Germany vary rather strongly their eigenvalue spectrum and the relevant principal components are very similar. This might be due to the fact that the basic modes that constitute the weather patterns are quite universal for the investigated area. Hence, the decomposition of the meteorological time series of $\vec{u}$ and pmsl in terms of the PCA seems to extract typical patterns of the climatology over Northern Germany.

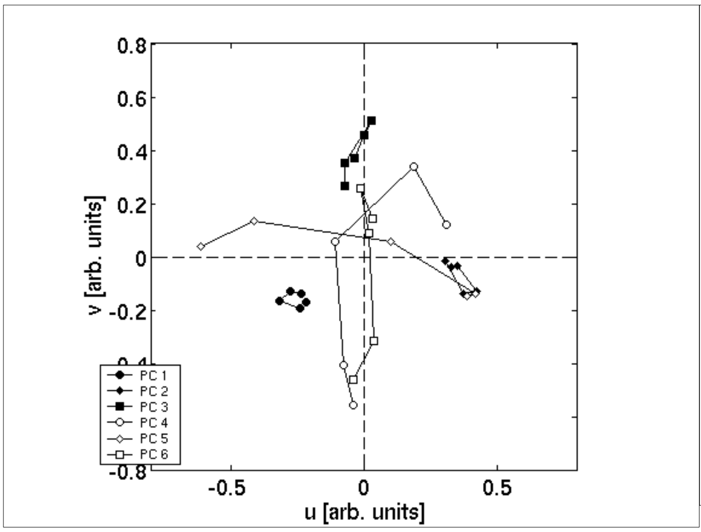

Figure 2 First six principal components (PC) of the wind vector (u,v) for the site Hilkenbrook. While PC 1 to2 are rather static the higher PC represent changes in wind speed and direction over one day.

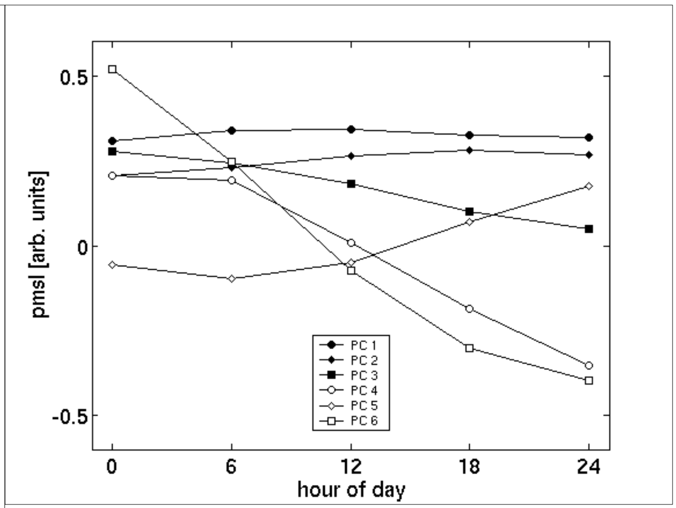

Figure 3 First six PCs of mean sea level pressure pmsl for Hilkenbrook. PC 1 and 2 represent high pmsl with minor changes while PC 3 describes an increase from low to medium pressure.

Note that the PCs do not necessarily correspond to actual meteorological conditions at the location. They constitute the basic set of vectors whose linear combination can reconstruct the actual situations. The inspection of the eigenvalue spectrum in figure 1 and the modes in figures 2 and 3 leads to the conclusion that the first six modes account for most of the relevant structures in the data. Hence, these first 6 out of the 15 PCs are used as a new vector basis that spans a reduced state space of the data. Let $\vec{v}_i$,i=1, …,6 denote this new basis then any state of the original data $\vec{w}$ can be approximately written as

$$\vec{w} \approx \sum_{i=1}^{6} a_i \vec{v}_i \tag{1}$$

where a_i are real numbers. To determine the 6 $a(j)_i$ -values on day *j* the daily values contained in the rows of matrix M (table 1) have to be projected on the basis $\vec{v}_i$:

$$a(j)_i = \vec{w}(j) \cdot \vec{v}_i \,, \qquad i=1, \ldots,6,\ j= 1, \ldots, 365 \tag{2}$$

where $\vec{w}(j)$ is the j-th row of M.

Thus, using PCA the data has been decomposed into six relevant PCs. For each day the component loading being required to reconstruct the (u,v,pmsl) time series of that day is given by the $a(j)_i$. The $a(j)_i$ are called dayscores.

5.2 Cluster analysis to classify data

So far PCA was used as an efficient way to code the rather complex meteorological information of the temporal development of three variables over 24 hours. Now the data has to be classified into different situations. This is done by clustering days with similar meteorological conditions into one class. For each day the vector of dayscores $\vec{a}(j) = (a_1, a_2, a_3, a_4, a_5, a_6)$ is one point in a six dimensional phase space. Cluster analysis simply forms different groups in this phase space that contain $\vec{a}$ which a are close together in terms of the given metric. As there are various ways how these clusters can be constructed [8, 9] a detailed discussion is beyond the scope of this work. The procedure chosen here to establish relations within and between the clusters is called complete linkage technique which produces the best results in the context of this investigation. In addition, a second linkage technique, called Ward's linkage, also provides useful results but with less pronounced distinctions between error values in the end. The procedure works iteratively in that it starts with the 365 vectors of the days and joins the two most adjacent objects into one common cluster where the last cluster naturally contains all other clusters. Hence, the difficult part is to interrupt this process at the right time. The aim is to have a rather small number of clusters that contain days with similar meteorological conditions which are different from those days in each of the other clusters. It turns out that a good choice for meteorological applications is between 5 and 9 clusters.

Each cluster is supposed to represent a typical synoptic situation. For each of the 5 clusters found at Hilkenbrook with complete linkage technique the mean values of wind speed and pmsl are shown in figures 4 and 5. Cluster 1 has strong winds from the south and decreasing pressure connected to weather situations where a low pressure area is approaching from the west or north west. Cluster 2 is a situation where a low centred in the north of the site passes (see weather map in fig. 6) with pressure first dropping and then increasing from a rather low level. The wind direction changes from south west to west and again back to SW. In contrast to this, cluster 3 is characterized by a high pressure system that is rather stationary in the north east of the site having moderate wind speeds from the east and very high pmsl (fig. 7). Cluster 4 has also relatively high pressure related to a ridge with wind from south west. Finally, cluster 5 refers to a situation with a blocking high pressure system east of the site and low pressure systems far in the west of the site. Hence, the isobars are from north west to south east with small pressure gradients and low wind speeds.

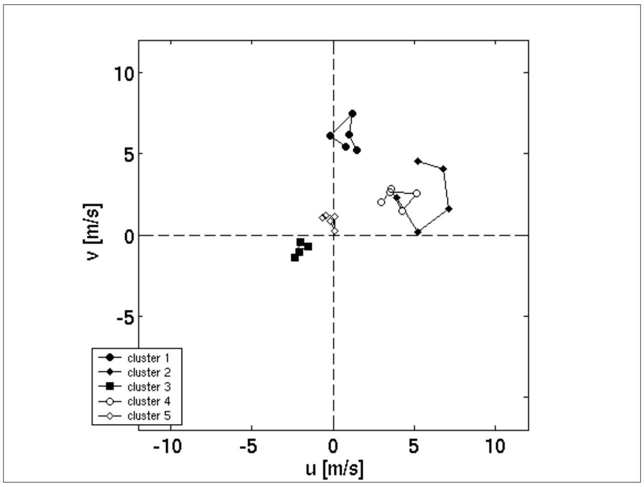

Figure 4 Mean values of $\vec{u}$ for each of the five cluster at Hilkenbrook.

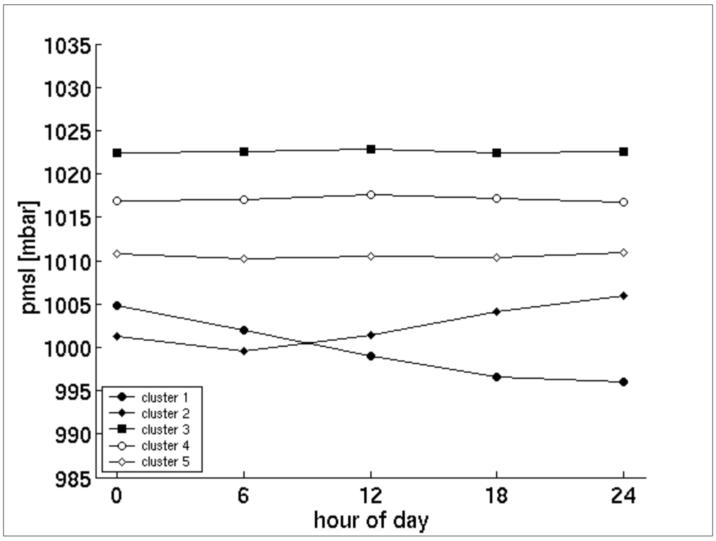

Figure 5 Mean values of pmsl for each of the five clusters at Hilkenbrook.

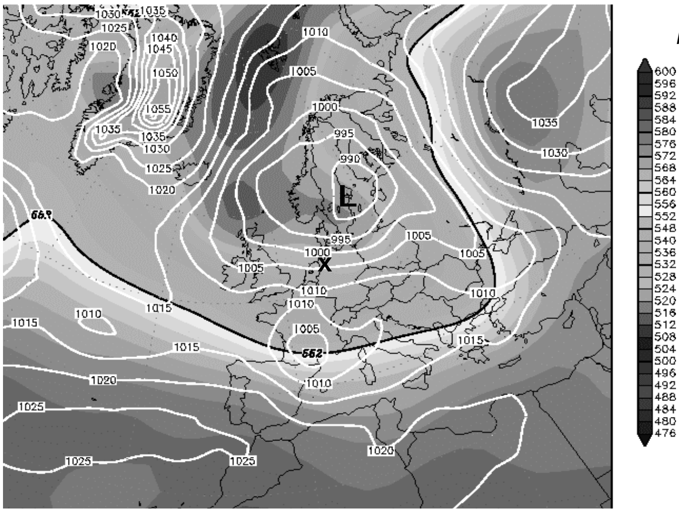

Figure 6 Typical day of cluster 2 for Hilkenbrook (marked by **x**). The white lines are isobars of the pmsl in mbar, the shaded areas show geopotential height of 500 mbar surface. A low pressure system is passing with its centre north of Germany. NCEP reanalyis plot from www.wetterzentrale.de.

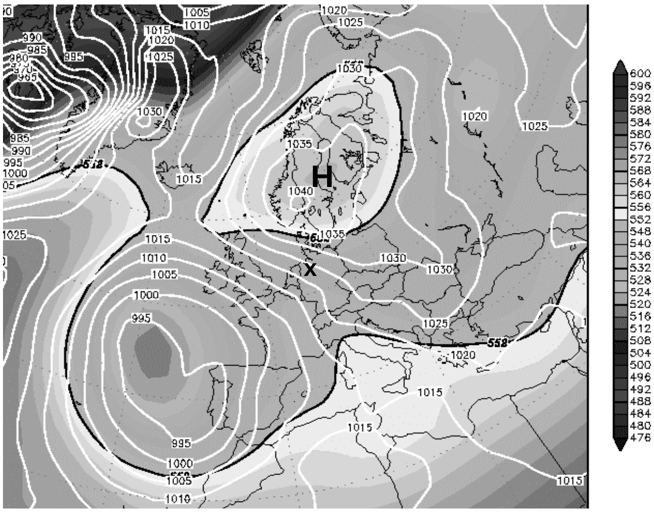

Figure 7 Typical situation for cluster 3 for Hilkenbrook (marked by **x**) with stationary high pressure over Scandinavia and easterly winds over northern Germany. NCEP reanalyis plot from www.wetterzentrale.de.

5.3 Relation to forecast error

The next step is to relate the different meteorological conditions defined by the 5 clusters to the forecast error of the wind speed. To evaluate the performance of the prediction system over one day the rmse is defined based on the 4 daily values of the wind speed where a prediction is available. Hence,

$$rmse = \sqrt{\frac{1}{4}\sum_{t}(u_{pred,t} - u_{meas,t})^2} \quad \text{with t = 6, 12, 18, 24 h} \qquad (3)$$

where $u_{pred,t}$ is the predicted wind speed at time *t* and $u_{meas,t}$ the measured wind speed. The daily error assigns one single value to each day in a cluster. The characteristic property of this definition is that the point of time at which a deviation between prediction and measurement occurs is not relevant. Coherent structures such as wrongly predicted fronts that cause errors at succeeding points of time are captured by definition (3). The numerical prediction used here is the 00 UTC run of the 10m wind speed provided by the German weather service (DWD). It is compared to measurements at the same height.

The means of the daily rmse values per cluster are shown in figure 8. Cluster 2 has the biggest rmse. It is by a factor 1.5 larger than the smallest rmse occurring for cluster 3. The 95%-confidence intervals illustrate the statistical significance of the results. A statistical test (F-test at confidence level 0.05) confirms that indeed cluster 2 has a statistically higher mean rmse than clusters 3,4 and 5. While cluster 3 has only a statistically significant difference to cluster 2.

In terms of weather situations this means that situations in cluster 2 (like the one shown in fig. 6) where a fast moving low pressure system determines the wind field typically have a larger rmse and, hence, a higher uncertainty in the wind speed prediction than the high pressure situation in cluster 3 (fig. 7). Similar results are found for other sites in Germany. In particular, sites in flat terrain with homogenous local conditions show a comparable behaviour. However, care has to be taken if local effects, e.g. orography, become important. Then there is an additional source of situation dependent error that overlays the universal effect of the numerical prediction.

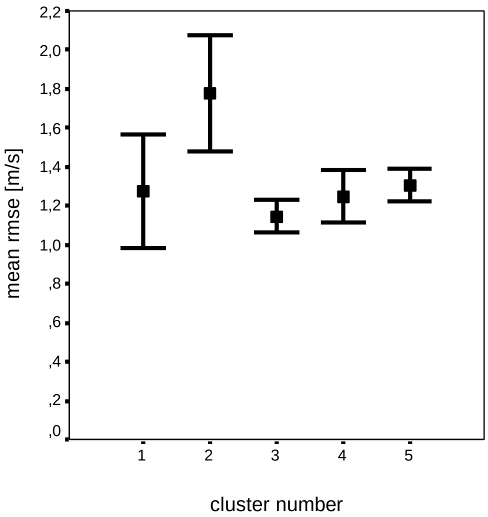

Figure 8 Mean rmse with 95%-confidence intervals for each of the 5 clusters for site Hilkenbrook (see figs. 4 and 5). The largest rmse refers to weather situations where a low pressure system passes north of the site (fig. 6). The smallest rmse occurs for high pressure situations as in fig. 7.

Conclusion

The investigation shows that the uncertainty of the wind speed forecast depends on the prevailing meteorological conditions. Typically, dynamic situations are harder to predict than stationary ones. In addition, the method used here allows for a quantitative assessment of this behaviour and, therefore, enables developers of prediction systems to include this information into the uncertainty estimate of their prediction.

Methods from synoptic climatology can successfully be applied to classify meteorological situations in different classes based on historical wind speed, wind direction and pressure measurements over one year. Comparing typical days from the clusters with large scale weather maps reveals that although the classification uses local information the clusters can be associated with the overall weather situation in most cases. For different sites the typical structures being extracted by PCA are very similar suggesting that the basic climatological modes in the time series are rather universal. Using one error value per day the average prediction uncertainty for each of the clusters is calculated and shows significant differences between different meteorological situation. The ratio between the largest and the smallest rmse is in the range of 1.3 to 1.6 for the investigated sites which is quite a profound difference.

To exploit the results of this work for an estimation of the uncertainty of the wind power prediction two steps are necessary. First, the predictability of the weather classes has to be confirmed, i.e. the cluster type has to be predicted in terms of the forecasts of wind speed, wind direction and pressure. As the prediction accuracy of direction and pressure is considerably better than wind speed the prospects are very good to accurately determine the type of meteorological situation in advance. Then the typical uncertainty of the wind speed forecast for this situation can be assigned. The second step is to calculate the uncertainty of the power prediction. An earlier investigation [2] showed that the power error can be described rather well by using the product of wind speed uncertainty and the derivative of the power curve at the predicted wind speed. In contrast to [2] the wind speed accuracy is now no longer constant but dependent on the meteorological condition.

Acknowledgements

We like to thank ISET and DWD for providing the data. This work was funded by the German Federal Environment Foundation (DBU) and the European Commission (ANEMOS project).

Literature

[1] Palmer TN. Predicting the uncertainty in forecasts of weather and climate. *Rep. Proc. Phys*. 2000, **63**:71-116

[2] Lange M, Heinemann D. Accuracy of Short-Term Wind Power Predictions Depending on Meteorological Conditions, *Proc. Global WindPower Conference*, Paris, 2002

[3] Luig A, Bofinger S, Beyer HG. Analysis of Confidence Intervals for the Prediction of the Regional Power Output. *Proc. European Wind Energy Conference EWEC* , Copenhagen, 2001, 725-728

[4] Yarnal B. *Synoptic climatology in environmental analysis – A primer*; Belhaven Press London and Florida, 1993

[5] Barry RG, Perry AH. *Synoptic Climatology*. Methuen & Co Ltd, London, 1973

[6] Shagedanova M, Burt TP, Davies TD. Synoptic Climatology of Air Pollution in Moscow. *Theoretical and Applied Climatology* 1998; **61**:85-102.

[7] Focken U, Heinemann D. Influence of Thermal Stratification on Wind Profiles for Heights up to 130m. *Proc. European Wind Energy Conference EWEC*, Madrid, 2003

[8] Everitt B. *Cluster Analysis*. Heinemann Educational Books, London, 1974

[9] Kalkstein LS. An Evaluation of Three Clustering Procedures for Use in Synoptic Climatological Classification. *Journal Climate and Meteorology* 1987; **26**:717-730.