A Modern Heuristic Method for Diagnosing Regional Near-Surface Wind Patterns

K. Kamber*, H. Wanner

Climatology and Meteorology, Institute of Geography, University of Berne, Hallerstr. 12, CH - 3012 Berne, Switzerland

**Corresponding author. Tel.: +41-79-640-7003*
E-mail address: kurt.j.kamber@opener.ch

Abstract

A four-month intensive measuring campaign with highly resolved spatial wind measurements was carried out in order to investigate regional near-surface airflow patterns in the meso-gamma to meso-beta scale in the region of Berne, situated in the Swiss Middleland between the Alps and the Jura mountains. Based on these wind measurements and on operational forecast data, a wind field information system was established as an important module for calculating the dispersion of air pollutants in complex terrain. In terms of broad operational use, the real-time diagnosis of wind patterns must be carried out with a minimum of wind information transmitted via online connections. Therefore, a modern heuristic method hereafter referred to as subspace based cluster analysis was developed. The investigations reveal that diagnoses of the typical near-surface wind patterns in the region of Berne can be established with multiple combinations of only three automated stations. A notable characteristic of the inductive method is the possibility to achieve high site-specific accuracy of the evaluated and pre-processed typical wind fields. Validation of the results demonstrates that the regional wind patterns can be diagnosed with good accuracy by using short fitting periods of 800 hours and 35 cluster representatives, which generate 15 synoptic wind field classes. The results of the study are verified by applying the method to the data set of the one-year intensive measuring campaign in the region of Basle, Switzerland (REKLIP / MISTRAL project). Diagnoses based on the typical regional wind fields correspond to the result of a time-lagged persistence of two hours. Furthermore, it is shown that the prognostic wind patterns (14 km grid size) are related to the observed near-surface wind patterns, permitting their forecast, though with distinctly lower accuracy compared to the diagnosis.

Keywords: near-surface wind; pattern recognition; air pollution; emergency response; operational use; statistical methodology

1. Introduction

Regional near-surface airflow in densely populated and heavily industrialised regions is often not investigated and even if a database exists the airflow is not known at the moment, although it is of major importance for assessing air pollutant dispersion. Especially in complex terrain reliable, fast and accurate wind information should be available in order to manage industrial hazards by calculating the time dependent regional near-surface distribution of air pollutants. Being aware that atmospheric conditions such as diffusion, mixing height, plume chemistry, and deposition have to be considered, the focus here is on wind field diagnosis. In the literature related to the meso-gamma scale (a few km to a few tens of km), widely differing approaches to this problem can be found, the majority of which attempt to provide diagnoses as well as prognoses.

Eckmann (1988) points out that wind field techniques for emergency response systems, that provide real-time dispersion estimates, should require limited field measurements. Davakis et al. (1998) thus use experimental data and account for terrain influences in order to calculate mass-consistent wind fields with one automated station and synoptic information from weather charts.

The study of Carter et al. (2000) uses the information of a network of automated weather stations in order to work out forecasts for a few hours by analog forecasting, i.e. historical pattern matching. A short-term predictive tool based on persistence and a mass-consistent wind field model presented by Cox et al. (1998) requires a highly resolved meteorological network. In this context the comparative investigations of Gross (1996) demonstrated that the calculation of wind fields in complex terrain with mass-consistent models only leads to realistic results if wind observations of a very dense network are available for initialisation.

Davis et al. (1999) anticipate that the applied operational forecast system with multiple-nest capability (MM5) can be used as a driver for other models so as to predict the dispersion of air pollutants in complex terrain. Finally, Dabberdt and Miller (2000) analyse an accidental release of a hazardous chemical under stationary conditions by applying a diagnostic mass-consistent wind field and furthermore investigate ensemble simulations in order to quantify the uncertainties in the dispersion model simulations.

The approach presented in this paper is based on the identification of the current wind pattern using as little wind information as possible and the direct access to preprocessed typical regional wind fields evaluated on the basis of an intensive measuring campaign. This approach draws on the MISTRAL project (Gassmann et al., 1996), a subproject of the REKLIP (1999) climate research project. Within the framework of the MISTRAL project a two-year intensive measuring campaign was conducted with the aim of examining near-surface regional wind patterns in the region of Basle, in the North-West of Switzerland.

Based on the findings of this project and at the request of the Swiss Federal Office of Energy, the WINDBANK project (Graber and Tinguely, 1999; Graber and Gassmann, 2000) was commissioned to inventory and examine near-surface wind systems in regions where Swiss nuclear power plants are situated. On the basis of hourly-averaged wind patterns, methods were developed to enable the diagnosis of the patterns with the aid of current meteorological information. An attempt was also made to forecast wind field evolution using information provided by the operational forecast model of the Swiss Weather Service.

The project presented in this paper uses the WINDBANK data set of a four-month (July – October 1997) intensive measuring campaign in the hilly region of Berne. There, wind conditions are characterized by strong topographic forcing (Wanner and Furger 1990) and thermally driven circulation systems.

The project adopts new approaches with the aim to develop a method for diagnosing typical wind patterns which requires a minimum of online information. This would allow regions with less facilities to build and run operational systems more cost-effectively. The results provide the basis for operational wind field information systems. Such systems can be used not only for the evaluation of current airflow conditions but also for calculating the dispersion of air pollutants using models which take account of the topography as well as observed typical wind patterns in complex terrain.

First, the measuring network and the basic data are described. The following chapter introduces the modern heuristic method and explains the difference between the inductive and the deductive approach. The chapter on results focuses on the process of station selection for the wind field diagnosis, the appropriate number of wind field classes, and the required duration of the intensive measuring campaign. Finally, the results are compared to those of a similar project and evaluated with regard to further investigations and operational use.

2. Measuring Network and Data

The measuring network in the Berne region (see Fig. 1) consisted of 22 temporarily installed wind measuring stations, 20 permanent automated stations (ANETZ) operated by the Swiss Weather Service (MeteoSwiss), and two temporarily installed SODAR systems. The SODAR systems recorded local wind conditions at 50m and 150m above ground level as well as in a device at 250m above ground level. The careful evaluation of locations for the temporary stations ensured a high-level spatial representation and maximum free flow, i.e. flow conditions which are not disturbed by obstacles leading to local perturbations. With the exception of a few tower stations, the measurements were taken at a height of 10m above ground level. Missing hourly average values (4.6%) were reconstructed by linear regression based on observed zonal and meridional wind components. A blanket view of wind conditions at a somewhat greater height above ground level was also desirable. Normally the relevant measuring or reanalysis data are not available, necessitating the use of prognostic data. This has the advantage of allowing prognostic data to be linked to observations using downscaling techniques.

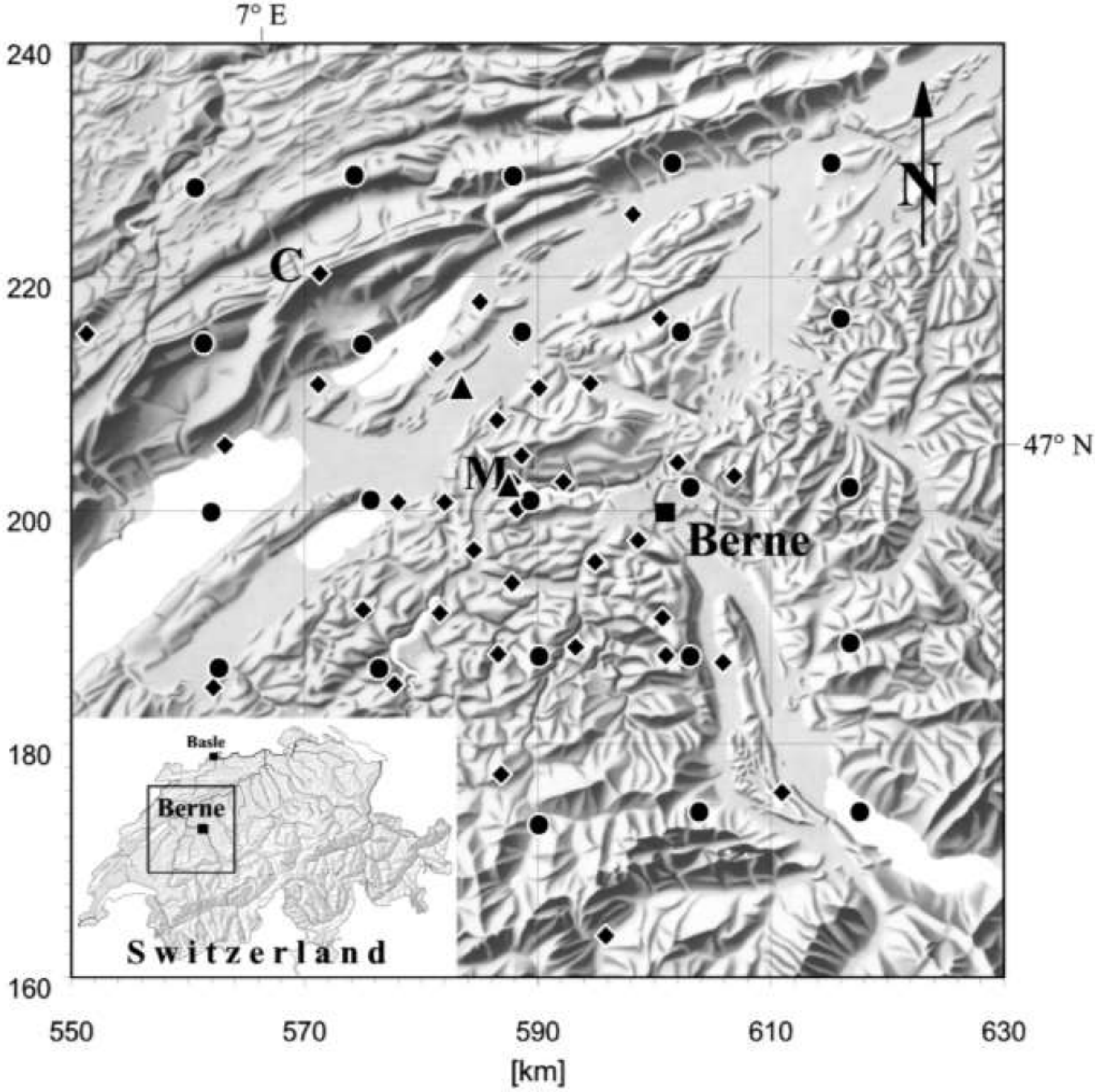

Fig. 1. Wind information network of the WINDBANK project. The topography represents the Swiss Middleland (430m – 700m asl) between the Jura Mountains in the North-West (1000m – 1500m asl) and the prealpine part of the Swiss Middleland in the South-East (1100m – 2000m asl).

The diamonds (♦) symbolize temporary and permanent wind measuring stations. The triangles (▲) designate SODAR system sites and the circles (●) near-surface grid points of the forecast model.

„Map data PK500, reproduced with approval of the Swiss Federal Office of Topography (BA002455)“

The measuring data are therefore supplemented by wind information at 25 grid points from the Swiss Model, the operational forecast model of the Swiss Weather Service (Majewski, 1991). This is a non-hydrostatic mesoscale model with a horizontal grid size of about 14 km, which was operational up to the beginning of 2001. The selected grid points are on average about 300 meters above the effective topography. The corresponding forecasts are all initialised at 12 GMT. Note that the observation sites and the grid points of the forecast model are referred to hereafter as stations.

3. Methods

The procedure for the classification of the hourly flow patterns consists of four main steps:

(a) selection of few diagnostic stations; the wind direction variables of these stations delineate the diagnostic space, which is a subspace of the state space; (b) clustering within the diagnostic space by allocation of the observations to a number of N cluster representatives; (c) calculation of the N wind field classes in the state space; and (d) assessing the quality of the classification.

The aim of the strategy is to identify a cluster structure within a diagnostic space with the attribute that the corresponding clusters in the state space produce a good value for the quality measurement.

The evaluation of the typical wind patterns is based solely on wind direction data in order to avoid the formation of wind field classes with no differences in wind direction but only in wind speed. Another reason for this approach is the fact that the regional dispersion of air pollutants depends to a large extent on the wind direction at the emission site. As will be discussed below, this aspect – with regard to specific stationary facilities – can be specially addressed by the method discussed here.

Therefore, a single flow pattern, called an event, is defined in this study by the wind direction at all stations at a fixed time. The observed events form single points in the state space which is delineated by the wind direction variables of all stations. The diagnostic space is defined as a subspace of the state space with a generally very low dimensionality.

(a) Combinations of diagnostic stations can be selected by a local search procedure or, for a very small number of diagnostic stations, by an exhaustive search procedure (Michalewicz and Fogel, 2000). In the case of the former, for a given combination of stations an additional station is sought, leading to a very good result. In the case of the second procedure, which necessitates a major calculation effort, the diagnostic stations are selected systematically. Findings show that the local search procedure very quickly produces good results.

(b) Various cluster techniques lend themselves to the formation of a specific number of N clusters (Kaufman and Rousseeuw, 1990; Arabie et al., 1996; Gordon, 1999). The procedure used in this study defines the projections of N events into the diagnostic space as cluster representatives, which systematically subdivide the diagnostic space on the basis of the Euclidean distance – in parallel with Thiessen polygons in a plane. This cluster procedure is extremely fast but it needs to screen many combinations of N stochastically selected events in order to obtain a very good solution. The described process groups all events in N clusters or classes, whereby every hour of the study period can be assigned to a particular class.

(c) Whereas the evaluation of the typical wind patterns is performed using wind direction data only, their calculation is based on wind direction as well as on wind speed. For every wind field class and every station, the wind vectors are calculated by vectorial averaging of the normalised wind vectors and by scalar averaging of the wind speeds.

(d) The quality of a classification or a diagnosis is characterized by the Root Mean Square Error of calculated wind directions, RMSE(α), which is a continuous quality measurement appropriate for regional comparisons. In the case of model fitting, RMSE(α) corresponds to the standard deviation and is calculated on the basis of the deviations between observations and the class-specific mean values assigned to them. In the case of model validation, which is carried out in an independent period, RMSE(α) is calculated on the basis of the deviations between observations and the diagnosed values. Note that due to the circular data structure, the intermediate angle α can only take on values between -180° and +180°. To exclude any influence by weak-wind observations, which is of secondary importance for regional transport, observations of wind speeds below a specific threshold β e.g. $0.5 ms^{-1}$ can be disregarded when calculating RMSE(α).

The diagnosis is carried out in the diagnostic space, which is partitioned by the cluster representatives. The observed wind directions of the diagnostic stations can be assigned to the partitions, which themselves are linked to the wind field classes. An important element of the diagnosis is the consideration of the current mean regional wind speed, which is taken into account by its reconstruction with the aid of the wind speeds of the diagnostic stations.

The method described in this paper is termed subspace based cluster analysis, in which the diagnostic space by definition is delineated by variables, i.e. the wind directions from the diagnostic stations. This is a modern heuristic method, which converts heuristic strategies into algorithmic methods for computers (Michalewicz and Fogel, 2000) and, as most heuristic methods, does not guarantee global optima. The statistical procedure is suitable for evaluating data with a linear or circular structure, using a minimum of information to achieve maximum accuracy in the diagnosis or prognosis of system states. It is particularly useful if system variables are characterised primarily by non-linear relationships.

In the procedure described in this paper, which represents the inductive solution approach, flow patterns are clustered on the basis of fewer selected station data. Once a very good solution has been found, the diagnosis is made based on the stations which were used for classification. As will be demonstrated in detail below, the validation of the results constitutes a key component of the methodical procedure. The deductive procedure calculates the wind field classes on the basis of the wind information at all stations. The optimal combination of diagnostic stations for diagnosing these classes is determined subsequently.

4. Results

The quality of classifications with respect to the number of classes and the number of diagnostic stations is shown in Fig. 2. From this, one can see that a good-quality classification of the regional pattern of wind directions can be made using only three stations. The best solutions found for combinations of four or more diagnostic stations result in minor improvements only. In around 100 classes the curves calculated using information from three or more stations show a transition to a relatively linear progress. The initially sharp drop in the curve is primarily attributable to the improvement in the representation of the system states due to the formed clusters. By contrast, the relatively linear drop of RMSE(α) is primarily attributable to the increase in the number of classes, combined with an over-fitting of the statistical model. The next step will therefore involve further investigations of 100 classes.

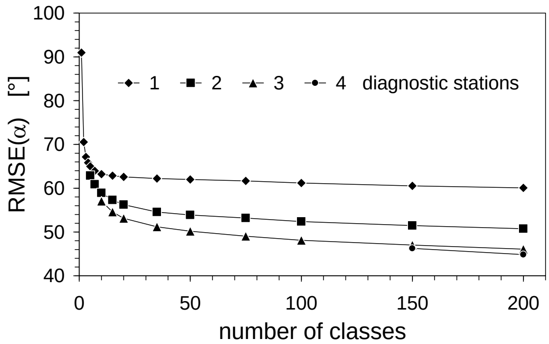

Fig. 2. Quality measurement RMSE(α) relative to the number of diagnostic stations and the number of wind field classes. RMSE(α) is the Root Mean Square Error of the intermediate angles (α) between observed and calculated wind directions. The diagram shows the results of four particularly appropriate combinations of diagnostic stations. Threshold β = $0.0 ms^{-1}$ i.e. all observations, including weak winds, were used for the calculation of RMSE(α).

Validation of the statistical model is performed periodically (Fig. 3). Model fitting is performed exemplarily in the months of July/August (Period I) and validation in the months of September/October (Period II). The results of the 59'640 station combinations are represented, resulting from the selection of three diagnostic stations out of 72 stations. The diagnoses made in the independent period result in values for RMSE(α), which are approximately 5° higher than during the fitting period. The strong correlation with a coefficient of determination (r^2) of 0.87 indicates that the method applied to the intensive measuring campaign data set exhibits robust attributes.

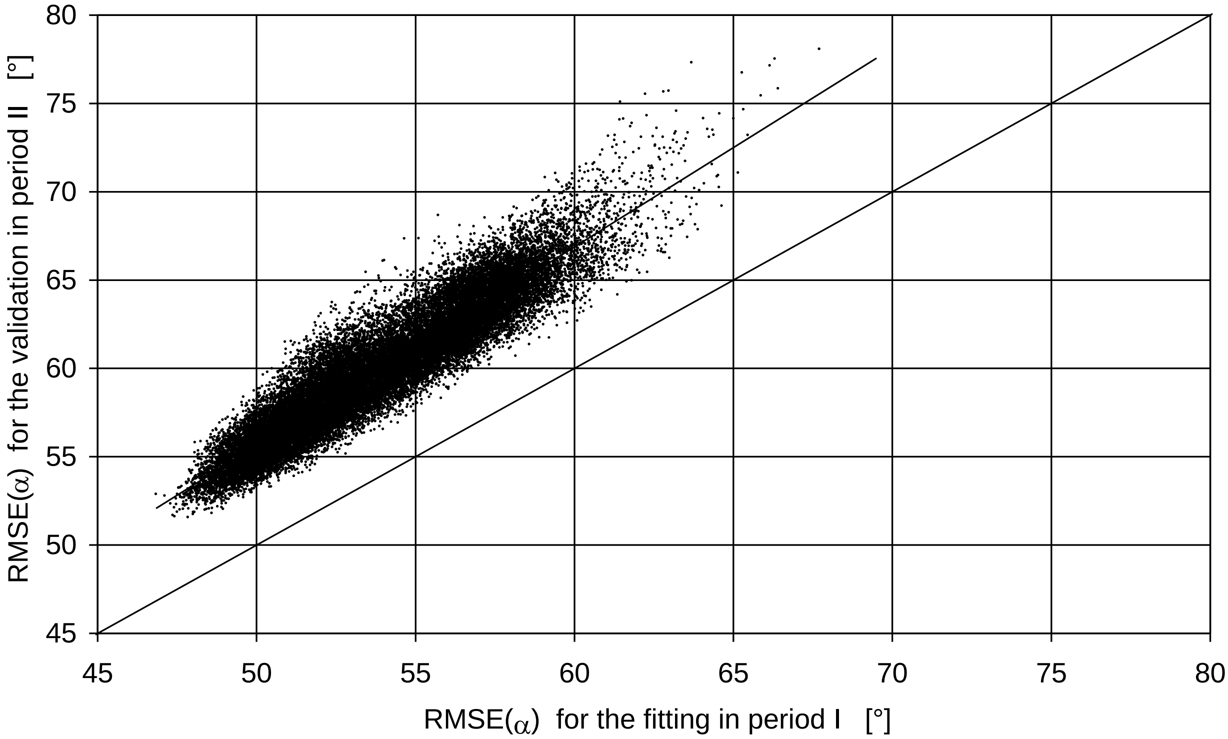

Fig. 3. Quality measurement RMSE(α) for model fitting in Period I (July, August) and model validation in Period II (September, October) for all combinations resulting from the selection of three diagnostic stations out of a total of 72 stations. Thus, each point corresponds to a specific wind field classification. 100 wind field classes, $\beta = 0.0 ms^{-1}$.

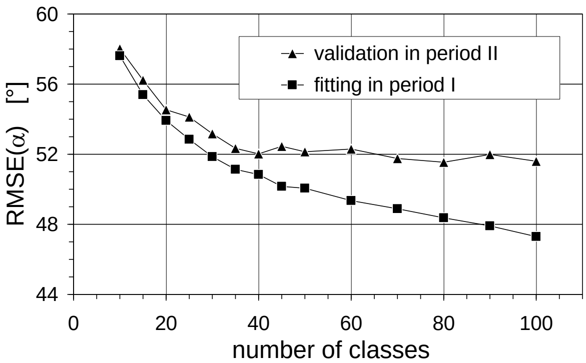

Fig. 4. Model fitting in Period I (hours 1 – 1476) and model validation in Period II (hours 1477 – 2952) dependent on the number of wind field classes. Best combination found for three diagnostic stations, $\beta = 0.0 ms^{-1}$.

The solutions that produce relatively poor results primarily involve valley stations, which are influenced by diurnally forced mountain-valley circulations.

If slight reductions in the quality of the diagnoses are acceptable, a much larger number of station combinations can be used for the diagnosis. With a view to setting up an operational system, these can then be subjected to an additional selection process based on operational, economic or other criteria.

Fig. 4 shows RMSE(α) for the fitting in Period I (July, August) and the validation in Period II (September, October) as a function of the number of wind field classes, obtained by means of the best classification found based on three diagnostic stations. It is apparent that no significant improvement in validation can be achieved in case of more than 35 classes.

Once the appropriate number of diagnostic stations, an appropriate station combination and the number of classes practical for diagnostic purposes have been determined, the influence of the duration of the fitting period is examined in terms of the performance of the statistical model in the validation period. For this purpose, model fitting is conducted in a variable time frame (hours 1 – t) while validation is performed in Period II (hours

1477 – 2952). Fig. 5 shows that only minor improvements in the validation period are achieved once the model fitting has been performed over a time frame of more than 400 hours. To obtain a more generalized verification of this result, a model fitting can be conducted over any period of 401 hours (fitting window, (t-200^h) – (t+200^h)), with model validation performed over the 2551 remaining hours of the study period.

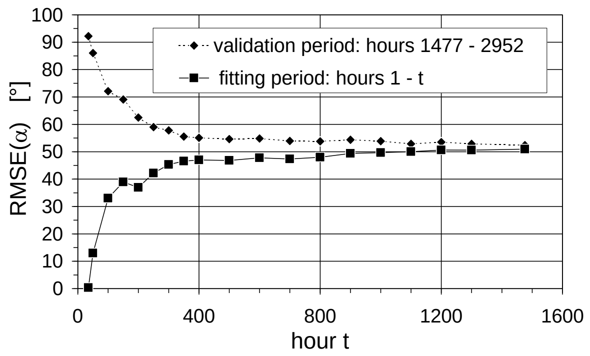

Fig. 5. Validation of the statistical model as a function of fitting periods of varying lengths (from hour 1 till hour t). Best combination found for three diagnostic stations and 35 wind field classes, β = 0.0ms^{-1}.

If this fitting window is continually shifted from the start to the end of the study period, quite homogenous results are achieved (not shown) which confirm the observation that good diagnoses can be worked out even on the basis of a very short measuring campaign.

To verify this result, the data set from the intensive measuring campaign conducted in the region of Basle was considered (Kaufmann and Weber, 1996; REKLIP, 1999), though the data set does not contain wind information from the operational forecast model of the Swiss Weather Service. Examination of a one-year data set shows that a 801-hour period (Fig. 6) is recommended for model fitting, which is related to the possibility of long-lasting large-scale weather conditions. This is demonstrated by a model fitting performed over 401 hours (Fig. 7) at a time. The spike in the validation curve is related to marked anticyclonic weather situation with inversion layer at times with fog, which prevailed from 25 November to 15 December 1991.

In the following the statistical and synoptical features of an appropriate wind field classification in the region of Berne are described. The classification is referred to as solution S_1 and is defined by three automated diagnostic stations, 35 classes, and a wind speed threshold β of 0.5ms^{-1}.

Two diagnostic stations are situated in the centre of the study region at 10m and 110m above ground level, respectively (Fig. 1, site M, Mühleberg nuclear power plant, 483m asl). The third diagnostic station is situated on the main Jura ridge at 45m above ground (Fig. 1, site C, Chasseral, 1599m asl).

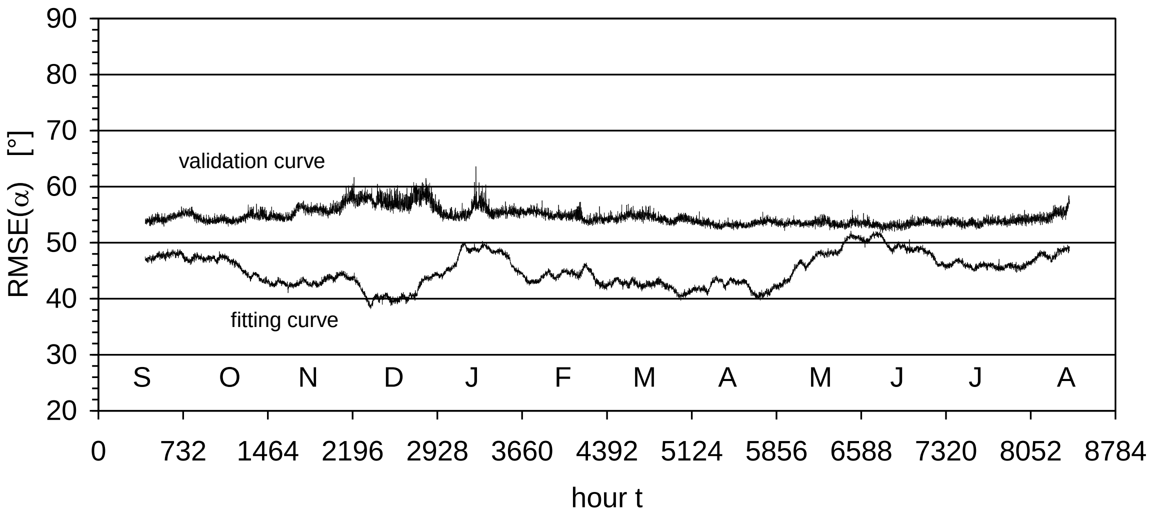

Fig. 6. Model fitting during 801-hour periods ((t-400^h) – (t+400^h)) and model validation in the 7983 remaining hours of the MISTRAL intensive measuring campaign (September 1991 – August 1992). Best combination found for three diagnostic stations and 35 wind field classes, threshold β = 0.5ms^{-1} i.e. weak wind observations below 0.5ms^{-1} were disregarded for the calculation of RMSE(α).

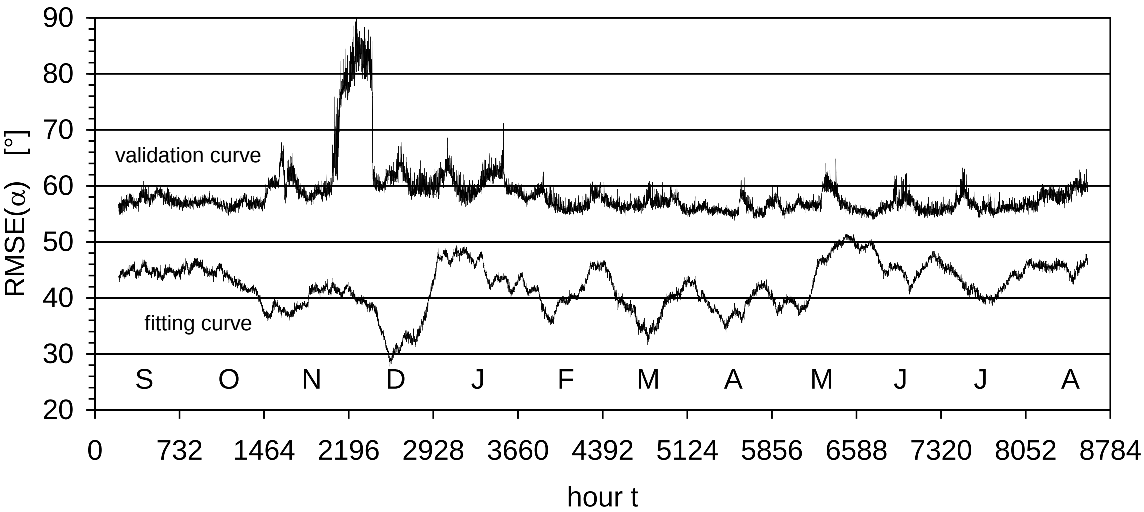

Fig 7. Similar Fig. 6 but with model fitting periods of only 401 hours.

As described in the section on methods the algorithm for the formation of the clusters screens many combinations of N cluster representatives in order to obtain a very good result. The solution S_1 is characterized by a RMSE(α) of 51.3°, which is the least value from a well shaped distribution with positive skewness and a maximal frequency at 52.8°.

Compared to the von Mises distribution, which is a highly typical distribution for circular data (Fisher, 1993), the symmetrical distribution of the intermediate angles α for solution S_1 shows a stronger centric distribution whereby around 80% of the calculated wind directions for this number of classes are located within the range of ± RMSE(α).

A special property of the inductive diagnostic method, though one which can only be observed in a very small number of diagnostic stations, is a high local accuracy of wind direction at the diagnostic stations as well as at the locations which are in close processual relation to them.

With respect to the three diagnostic stations, this phenomenon, which is inherent to the methodical procedure, leads to remarkably good values for RMSE(α) in the 18.8° – 25.7° range (solution S_1). This property allows a wind field information system to be adapted easily to operational requirements in terms of spatial diagnostic accuracy. Of particular interest is the selection of diagnostic stations at potential emission sites.

Note that the diagnosis described in the section on methods must be conducted based on information provided by the 35 cluster representatives as well as the associated wind field classes. When the 35 wind field classes defined by wind direction as well as wind speed at the stations are hierarchically clustered, 15 classes exhibit a leap in the curve of RMSE(α), which indicates that the 35 wind field classes can be grouped into 15 synoptic wind field classes. This is associated with a slight rise of RMSE(α) from 51.3° to 53.1° (solution S_1) as well as a generally complex partitioning of the diagnostic space.

The number of 15 synoptic wind field classes thus obtained is of a similar order of magnitude as the 12 wind field classes in the Basle region in North-Western Switzerland (Kaufmann and Weber, 1996) and the 12 identified basic wind patterns in Eastern Idaho (Carter et al., 2000).

5. Discussion

In this paragraph the results of the study are compared with two projects, which use the deductive methodical approach. They investigate wind fields, which are normalized in respect to wind velocity, with the intention of preventing the formation of wind field classes with no difference in flow patterns but only in wind speed. The classification as well as the diagnosis are based on the same dataset and the evaluation of the quality of the wind field diagnosis depends on the rate of hitting the correct wind field class.

Table 1
Performance for the deductive and inductive diagnostic method. Root Mean Square Error of the calculated wind direction, RMSE(α), Root Mean Square Error of the calculated wind speed RMSE(u). Region of Berne, July – October 1997

Diagnostic data	Online data and forecast data [f]				Online data	
Diagnostic method	I Deductive [a]		II Inductive [b]		III Inductive [c]	
	RMSE(α) (°)	RMSE(u) [d] (ms^{-1})	RMSE(α) (°)	RMSE(u) [d] (ms^{-1})	RMSE(α) (°)	RMSE(u) [d] (ms^{-1})
Observation sites [e]	54.7	1.18	49.1	1.22	49.1	1.24
Model grid points [g]	32.7	1.50	35.5	1.72	53.7	2.48

[a] Diagnostic data from 20 automated stations and 25 model grid points (own calculations, based on the classification of Graber and Tinguely (1999), see text, β = 0.5ms^{-1}).
[b] Diagnostic data from 2 automated stations and 1 model grid point (as solution S_1, see text, site C is replaced by the model grid point in the south-east).
[c] Diagnostic data from 3 automated stations (solution S_1, see text).
[d] Mean wind speed at the observation sites and at the model grid points, for the interpretation of RMSE(u): 2.13ms^{-1} and 5.25ms^{-1}, respectively.
[e] Results for the temporary stations and the automated stations.
[f] Forecast data used in the same way as reanalysis data.
[g] Results for the 25 model grid points of the weather prediction model.

The results of the first project with the name MISTRAL illustrate that diagnosis requires information from many stations. Thus, measurements from six automated stations are required to achieve a hit rate of 81%. If 12 automated stations are used and optimum data availability is assured, correct allocation to one of the defined 12 wind field classes can be achieved in 91% of the cases (REKLIP, 1999).

The second project, a follow-up study by Graber and Tinguely (1999) – on the same data basis as this paper – therefore uses the information from 20 automated stations and 25 model grid points of the forecast model to diagnose 12 wind field classes. It remains to be stated that the operation of such dense automated measuring networks entails high costs and calls for a high level of operating security for stations and data lines.

Table 1 compares the results of the project just reviewed with those of the inductive diagnostic method discussed here. The first two solutions (Tab. 1, column I and II) use online data as well as forecast data for the diagnosis and are of roughly comparable quality. However, they differ clearly in terms of the number of data sources required to perform the diagnosis (see footnote [a] and [b]). Hence the inductive diagnostic method turns out to be eminently suitable for obtaining good diagnostic results based on very little information. The third solution (Tab. 1, column III, solution S_1) was worked out with the aid of three automated stations. This, understandably, is the reason why the model grid points exhibit larger values for RMSE(α). Mention has already been made of the greater local accuracy of wind direction produced by the inductive solution.

The wind directions of the forecast data used agree quite well with the SODAR observations at the higher level as well as the observations at the few peak stations. Except for the strong advective weather conditions, this is not the case with the near-surface stations. However, it appears that the use of forecast data as predictors allows regional near-surface wind patterns to be forecast, albeit less accurately than by diagnosis. For the observation sites, the validated solution of a very good forecast produces a RMSE(α) of 68.4° (β = 0.5ms^{-1}), which is significantly higher than that obtained by diagnosis (49.1°, cf. Tab. 1). It remains to be pointed out that the advantages of the inductive method presented here are not realised in applying it for forecasting purposes, because the forecast model produces a great deal of wind information. Since no results at higher local accuracy could be obtained for the near-surface area using forecast data, and due to reasons of redundancy, operation of an inductive diagnostic system, which requires only online data, is recommended along with a separately running forecast system.

Assuming that a flow pattern remains persistent for a specific time, a diagnosis may be worked out based on flow patterns which were observed one, two or more hours ago. In this way, for the observation sites and for the first four hours, a diagnosis based on a time-lagged persistence reveals values for RMSE(α) of 34.4°, 49.5°, 59.1°, and 66.7° ($\beta = 0.5 ms^{-1}$). Hence the performance of the diagnosis of typical regional near-surface wind fields in the Bern region corresponds to the diagnosis, which can be achieved based on a time-lagged persistence of two hours. Very good diagnostic results could be achieved on the basis of a time-lagged persistence of one hour, but this would require an extremely dense automated network as well as continual calculation of the current wind field, which would need to be performed within the hour.

The investigations indicate that only part of the processes and the transitions, which determine flow patterns, can be diagnosed on the basis of the methodical procedure. Special studies show that particularly circulation systems in valley locations are often defined on a lower scale and are therefore decoupled from regional processes. Thus, we hypothesize that the relatively major residual variations are above all attributable to the influence of micro-scale processes, the manifold transitions between the typical flow patterns as well as to the incorporation of events with weak mean flow.

Similar to the studies of Davis et al. (1999), the part determined by recurring meso-gamma scale circulation can be examined by forming hourly wind field classes for the 24 hours of the day. Taking these circulation systems into account and with respect to observation sites, a RMSE(α) of 78.5° ($\beta = 0.5 ms^{-1}$) can be computed. Given a corresponding overall variability of 95.3°, which can be calculated on the basis of a single class, and a variability of 49.1° (cf. Tab. 1) which cannot be further reduced, only one part of the residual variations of wind direction can be explained by diurnally forced circulation systems. In this context it remains to be stated that the resultant wind directions of the 24 wind field classes are in fact defined by diurnally forced circulation systems. This can be verified simply by excluding observations of higher wind speeds when calculating the wind field classes.

Finally, it remains to be stated that only a part of the observed variability in regional near-surface wind patterns can be diagnosed by typical wind fields. This points to the possibility of synthetic modelling (Sattler et al., 1999) of the few typical wind conditions, which would greatly simplify the set-up of operational systems, since cost-intensive measuring campaigns would not need to be conducted. However, wind field information obtained through an intensive measuring campaign provides a very good basis for work in the fields of air pollution prevention and settlement planning.

The great advantage of any diagnosis of typical regional wind fields is the fact that these can be pre-processed, allowing sufficient time and effort for their once-only calculation, e.g. taking into account the impact of terrain features and energy fluxes. As comparable studies by Gross (1996) demonstrate, the calculation of wind fields using prognostic rather than diagnostic models is desirable. In particular, such approaches allow thermally driven flows to be modelled in valleys where no observations are available. Promising comparisons between observed and modelled regional wind fields are presented by Kastendeuch et al. (2000).

It can be assumed that wind field information systems will soon become a standard tool for emergency management, particularly for densely populated settlement zones. These information systems are of interest not only in terms of emergencies but also in terms of smog situations where, for example, major emission sources can be better controlled.

6. Conclusion

The subspace based cluster analysis method presented here is a modern heuristic approach to evaluate data with a linear or circular structure. It aims at diagnosing or forecasting system states based on a minimum of information – particularly where system variables are characterised primarily by non-linear relationships. With respect to the diagnosis of regional near-surface wind fields, it demonstrates how these can be developed and optimised on the basis of short-term intensive measuring campaigns and a low volume of current station data.

Validation of the results shows that diagnoses conducted in the region of Berne can be worked out based on three automated stations and a large number of related station combinations, respectively. The study shows that the diagnosis of typical wind patterns can be achieved on the basis of 35 wind field classes and a remarkably short measuring campaign of only 800 hours. A higher number of classes or a longer fitting period only leads to minor improvements in the validation period. The suitability of a measuring campaign of 800 hours was verified in particular by the data set of a one-year intensive measuring campaign conducted in the region of Basle (REKLIP / MISTRAL project), where it was observed that the impact of different seasons was very minor.

It has been shown that flow patterns in the operational forecast have a deterministic relationship with the observed near-surface patterns. Incorporation of forecast data from the Swiss Weather Service therefore enables a forecast of regional airflow, although the results for near-surface stations are not as good as those of diagnoses based on online wind information. Diagnoses with greater local accuracy, which are often of major interest for operational reasons, can be obtained only by the inductive method.

Acknowledgements

Special thanks are due to the institutions which supported and funded this research: The Fund for Projects and Studies of the Swiss Electric Utilities, the Swiss Federal Nuclear Safety Inspectorate of the Swiss Federal Office of Energy, the Werner Steiger Foundation and the Cantons of Basle Country and Basle City. We also like to thank the Swiss Weather Service (MeteoSwiss) and W. Graber for the data set of the intensive measuring campaign, and C. Schmutz, S. Brönnimann, M. Ambühl and D. Assoulin for many stimulating discussions.

References

Arabie, P., Hubert, L.J., De Soete, G., (Eds), 1996. Clustering and classification. World Scientific Publications. Singapore, New Jersey.

Cox, R.M., Sontowski, J., Fry, R.N., Dougherty, C.M., Smith, T.J., 1998. Wind and Diffusion Modeling for Complex Terrain. Journal of Applied Meteorology 37. 996–1009.

Carter, R.G., Keislar, C. and R.E., 2000. Emergency Response Transport Forecasting Using Historical Wind Field Pattern Matching. Journal of Applied Meteorology 39, 446–462.

Dabberdt, W.F., Miller, E., 2000. Uncertainty, ensembles and air quality dispersion modelling: applications and challenges. Atmospheric Environment 34, 4667–4673.

Davakis, E., Varvayanni, M., Deligiannis, P., Catsaros, N., 1998. Diagnosis of wind flow and dispersion over complex terrain based on limited meteorological data. Environmental Pollution 103, 333–343.

Davis, C., Warner, T., Astling, E., Bowers, J., 1999. Development and application of an operational, relocatable mesogamma-scale weather analysis and forecasting system. Tellus, 51A, 710–727.

Eckmann, R.M., 1988. The suitability of different wind-field techniques for an emergency-response dispersion model. In: Proceedings of the ANS Topical Meeting on Emergency Response-Planning, Technologies and Implementation, 4–4, pp. 1–6.

Fisher, N.I., 1993. Statistical Analysis of Circular Data. pp. 39–57. Cambridge University Press.

Gassmann, F., Feller, W., Schaub, O., Kamber, K., Moussiopoulos, N., Megariti, V., 1996. Results of wind field project MISTRAL and application for planning and emergency response. In: Caussade, B., Power, H., Brebbia, C.A. (Eds.), Air Pollution, vol. IV: monitoring, simulation and control, Computational Mechanics Publications, 495–506.

Gordon, A.D., 1999. Classification. 2nd ed., Monographs on statistics and applied probability. Boca Raton Chapman & Hall. 256 pp.

Graber, W.K., Tinguely M., 1999. Projekt „Windbank oberes Aaretal“. Klassifizierung, Diagnose und Prognose von Windfeldern in der Region des Kernkraftwerkes Mühleberg. Paul Scherrer Institut, Abteilung Luftfremdstoffe. PSI Bericht Nr. 99–09, Paul Scherrer Institut, Villigen, Schweiz. ISSN 1019-0643, 56 pp. (in German)

Graber, W.K., Gassmann, F., 2000. Real time modelling as an emergency decision support system for accidental release of air pollutants. Mathematics and Computers in Simulation 52, 413–426.

Gross, G., 1996. On the applicability of numerical mass-consistent wind field models. Boundary-Layer Meteorology 77, 378–394.

Kastendeuch, P.P., Lacarrere, P., Najjar, G., Noilhan, J., Gassmann, F., Paul, P., 2000. Mesoscale Simulations of Thermodynamic Fluxes over Complex Terrain. International Journal of Climatology 20. 1249–1264.

Kaufman, L., Rousseeuw, P.J., 1990. Finding groups in data: an introduction to cluster analysis. New York: Wiley.

Kaufmann, P., Weber, R. O., 1996. Classification of Mesoscale Wind Fields in the MISTRAL Field Experiment. Journal of Applied Meteorology 35, 1963–1979.

Majewski, D., 1991. The Europa-Modell of the Deutscher Wetterdienst. ECMWF Seminar Proceedings, Numerical Methods in Atmospheric Models, Reading, UK, vol. 2, 147–191.

Michalewicz, Z., Fogel, D.B. 2000. How to Solve It: Modern Heuristics. Springer-Verlag Berlin, Heidelberg.

REKLIP, 1999. Luftqualität und Regionalklima. pp. 50–56. Schlussbericht, 3. Gassmann, F., Ahrens, D., Vogel, B. (Eds). Editions Coprur, Strasbourg, France. ISBN 2–84208–035–1. (in German)

Sattler, K., Traup, S., Kruse, B., 1999. A Cluster Space Calibration Scheme for Synthetic Wind Statistics Calculated with a Mesoscale Model and its Application to Regional Wind Climate Studies. Meteorology and Atmospheric Physics 71, 189–203

Wanner, H., Furger, M., 1990. The Bise – Climatology of a Regional Wind North of the Alps. Meteorology and Atmospheric Physics 43, 105–115.