Cluster analysis of wind fields in complex terrain
P. Kaufmann, R.O. Weber
Paul Scherrer Institute, CH-5232 Villigen PSI, Switzerland

ABSTRACT

A new method for automated classification of regional scale wind fields is proposed. The classification scheme makes use of observed data only and does not require interpolation to a regular grid. It is applied to data from the REKLIP / MISTRAL project. The obtained grouping of the wind fields from a one year period reflects diurnal and seasonal variations of air flow in complex terrain.

INTRODUCTION

Automated classification schemes for wind fields on a synoptic scale have been developed for air pollution studies, e.g. by Fernau and Samson [1] and by Dorling et al. [2]. A similar method for the regional scale has been presented in Gassmann et al. [3]. All of these schemes work along the following lines: A wind field is estimated on a regular grid, as in Gassmann et al. [3] by a diagnostic wind field model (Moussiopoulos et al. [4]) which interpolates the wind fields from the irregularly distributed observations to a regular grid. In a second step, for each wind field trajectories from predefined starting points are determined. A distance measure between two wind fields is defined by means of the trajectories. Making use of the distances obtained in this way, a cluster analysis is performed over the set of all wind fields. Each of these steps has its own pitfalls which may influence the outcome of the classification procedure.

Instead of performing all the intermediate steps mentioned above, the authors decided to carry out a cluster analysis which is solely based on the observed wind at each station. A wind field in this context is a set of observations at measurement sites at a specific time rather than a set of wind vectors on a regular grid. Greene [5] used wind vector data of 50 stations for a complex principal component analysis (PCA) and performed a classification using the empirical orthogonal eigenfunctions. PCA requires complete datasets, therefore a missing value at a single station causes the exclusion of the wind field from the analysis. In our case, 50 stations with an average data loss of approximately 10% lead to a total data loss of 99.6%, which matches the expected
value for independent variables (1-0.95^2=99.5%). Furthermore, the PCA em-
phasizes variables with unique temporal behavior and thus might put too much
weight on stations with local influences, in which we are not interested. In con-
trast, the classification method presented here uses the observed data only and
does not require any preprocessing.

METHOD OF CLASSIFICATION

We decided to apply a hierarchical clustering method, which does not require
any a priori knowledge of the data structure. Like any clustering method
(Anderberg [6]), it needs a measure of dissimilarity, called distance, between
two objects. The objects of the analysis are normalized hourly mean wind
fields. The distance measure between two wind fields is defined as the mean
absolute value of the vector differences at all stations. In order to determine a
suitable clustering method, five standard methods were tested with simplified,
was used. This method assumes that the used distances are Euclidean, which is
not the case here. The complete linkage method (Anderberg [6]) turned out to
be the most robust for our purpose. It tends to build groups of similar size like
Ward's method does, but is not sensitive to details of the distance measure,
because only the ranking of the distances is important.

THE PROJECT REKLIP / MISTRAL

In the region of Basel (Switzerland)-Freiburg (Germany)-Strasbourg (France),
the so called Regio Basiliensis, an international climatological project called
REKLIP (Regio Klima Projekt) is taking place (Parlow [7]). The aim of this
project is to investigate the regional climatology of the upper Rhine valley.
The universities of Basel, Freiburg, Karlsruhe and Strasbourg are partitioning.
In addition, several public and private institutions contribute to the project. For
a better understanding of the wind climate in the region of Basel, the Paul
Scherrer Institute and the two cantons of Basel initiated a subproject of
REKLIP, called MISTRAL (Gassmann [3]).

MISTRAL (Modell für Immissions-Schutz bei Transport und Ausbreitung von
Luftfremdstoffen) is designed to measure and interpret near-surface air flow in
complex terrain. Between June 1991 and December 1992, ground based
meteorological measurements were made in the Basel area (Figure 1). The
observation network consisted of 49 stations measuring wind speed and
direction, standard deviation of the wind components, temperature and relative
humidity (Kamber and Kaufmann [8]). The vertical range of the measurement
sites extended from 230 m above mean sea level (MSL) up to 1170 m MSL.
The sensor height was approximately 10 m above ground level (AGL) at field
stations and reached a maximum of 70 m AGL on exposed high buildings in
urban areas. An additional station of the Swiss weather service, placed on top
Figure 1: Observation area ($50 \times 55$ km$^2$) around the city of Basel with the 50 measurement sites (triangles). The grey scale indicates the height above mean sea level ranging from below 300 m (white) to above 1100 m (black). White triangles are stations above 800 m MSL.

of a TV tower 262 m AGL, was included in the analysis. The observed area ($50 \times 55$ km$^2$) is shown in Figure 1 and is described in more detail in Gassmann [3]. The project MISTRAL is mainly based on the following idea: The dense network of observations should make it possible to classify the wind fields in order to get a catalogue of the typical regional flow patterns. It is planned to reinstall a certain number of stations and to run them in an on-line mode, thus providing continuous wind information. The actual wind observa-
Figure 2: Cluster membership of the hourly averaged wind fields, one day per week, from 5 September 1991 to 27 August 1992. The vertical axis indicates the day, the horizontal axis the hour of the day. The 15 clusters are distinguished by grey scales.

The catalogue of typical regional flows can also be used for air pollution management. In addition, such a dense network of stations should help to further elucidate the physical mechanisms of flow generation in inhomogeneous terrain.
RESULTS

The classification method described above was applied to a one-year period from 1 September 1991 to 31 August 1992 (366 days) of wind data from the REKLIP / MISTRAL project. An observed wind field consists of all hourly wind vectors at each station at a specific hour. One-hour averages are believed to reduce the influence of local effects, but still allow to resolve diurnal variations. The objects in this analysis are the normalized hourly mean wind fields of these 8784 hours. The result of a hierarchical cluster analysis is a hierarchical tree of all wind fields. One is in principle free to choose the number of groups. Usually a discontinuity in the sequence of the levels of similarity, at which two objects or groups merge to a new cluster, is taken as an indicator for a meaningful number of clusters. In our analysis, a number of 7 or 15 clusters seems appropriate. As we do not want to loose too much information about the fine structure of the wind fields, our wind fields are grouped into 15 clusters. An extract of the result, covering every seventh day, is given as an example in Figure 2. On most days, a clear diurnal variation of the wind patterns can be seen. Regarding the full one-year period, seasonal variations can also be observed. This result agrees well with the physical mechanism of mountain-valley wind systems. For the one year period from September 1991 to August 1992, the frequencies of the groups are listed in Table 1.

Table 1: Summary of the 15 clusters obtained by the complete linkage clustering method.

<table>
<thead>
<tr>
<th>Cluster number</th>
<th>Count</th>
<th>Relative frequency</th>
<th>Mean wind speed [ms⁻¹]</th>
<th>Std. Dev. [ms⁻¹]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>127</td>
<td>1.4%</td>
<td>1.7</td>
<td>0.60</td>
</tr>
<tr>
<td>2</td>
<td>166</td>
<td>1.9%</td>
<td>1.5</td>
<td>0.41</td>
</tr>
<tr>
<td>3</td>
<td>419</td>
<td>4.8%</td>
<td>2.1</td>
<td>0.71</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>0.3%</td>
<td>2.0</td>
<td>0.74</td>
</tr>
<tr>
<td>5</td>
<td>1692</td>
<td>19.3%</td>
<td>3.8</td>
<td>1.76</td>
</tr>
<tr>
<td>6</td>
<td>751</td>
<td>8.5%</td>
<td>2.8</td>
<td>1.19</td>
</tr>
<tr>
<td>7</td>
<td>637</td>
<td>7.3%</td>
<td>2.3</td>
<td>0.82</td>
</tr>
<tr>
<td>8</td>
<td>1643</td>
<td>18.7%</td>
<td>2.5</td>
<td>0.83</td>
</tr>
<tr>
<td>9</td>
<td>376</td>
<td>4.3%</td>
<td>1.9</td>
<td>0.61</td>
</tr>
<tr>
<td>10</td>
<td>580</td>
<td>6.6%</td>
<td>3.2</td>
<td>2.10</td>
</tr>
<tr>
<td>11</td>
<td>176</td>
<td>2.0%</td>
<td>1.3</td>
<td>0.27</td>
</tr>
<tr>
<td>12</td>
<td>407</td>
<td>4.6%</td>
<td>2.3</td>
<td>1.41</td>
</tr>
<tr>
<td>13</td>
<td>159</td>
<td>1.8%</td>
<td>1.4</td>
<td>0.39</td>
</tr>
<tr>
<td>14</td>
<td>1050</td>
<td>12.0%</td>
<td>1.9</td>
<td>0.59</td>
</tr>
<tr>
<td>15</td>
<td>571</td>
<td>6.5%</td>
<td>2.0</td>
<td>0.67</td>
</tr>
<tr>
<td>Total</td>
<td>8784</td>
<td>100.0%</td>
<td>2.6</td>
<td>1.39</td>
</tr>
</tbody>
</table>
Figure 3: Cluster averages of normalized wind vectors at all measurement sites for a) cluster 5 and b) cluster 14. 'C' labels the station on the TV tower at St. Chrischona.

Figure 4: Wind roses at thirty stations for a) cluster 5 and b) cluster 14. The wind roses are divided in 10° sectors, whose length are proportional to the frequency. The circle around each measurement site indicates the 5% frequency level. Only the lower portion (50 x 40 km²) of the observation area is shown. 'C' labels the station on the TV tower at St. Chrischona.

Two of the most frequent wind fields are presented in Figure 3. The vectors show the cluster average of the normalized wind measurements at the individual observation sites. Figure 3a (cluster 5) describes the rather simple situation with clear westerly winds over the whole investigation area, even in the valleys. Only the stations deep in the Ergolz valley show some anomalies. However, the situation in Figure 3b (cluster 14) is more complicated. While
stations at higher locations, including the Chrischona tower north-east of Basel, and the stations at the west rim of the investigation area show westerly or south-westerly winds, the measurements in moderate heights show southerly winds, and down at the valley floors, especially in the east part of the Rhine valley, an easterly wind can be observed. This flow pattern mostly occurs during nighttime and shows the local influence of cold air drainage.

The variability between all wind fields belonging to the same cluster is qualitatively described by the wind rose at each station. The wind roses show the frequencies of all observed wind directions within the cluster. For each cluster a map of the wind roses is drawn. In Figure 4a (cluster 5), all stations show a well defined prevailing westerly wind within the cluster, except one station south of the Ergolz valley in a narrow side valley. This is the class with the highest mean wind speed (3.8 ms⁻¹ vector mean, Table 1). In other classes, there is a minority of the stations not showing a distinct preference for a wind direction (e.g. cluster 14, Figure 4b). No preferred direction can e.g. be seen at a station in the lower Birs valley, at the two southernmost stations and at some of the eastern stations. The particular flow pattern is not well defined at these stations, and reversely these stations are not good representatives for the particular class. A more detailed classification would be needed if the flow at these stations should be uniquely resolved. From the inspection of the wind roses we can thus draw conclusions on the representativity of the wind measurement sites.

CONCLUSIONS

A classification of typical regional wind field patterns based on hourly wind measurements was successfully performed for the REKLIP / MISTRAL area. The result can be interpreted in terms of typical processes in complex terrain like mountain-valley winds and cold air drainage. However, the cluster analysis of the wind fields alone results in a grouping of the observed flow patterns, but does not give the meteorological interpretation. Additional meteorological variables will be considered in subsequent work to explain the physical meaning of the groups. The variability of the wind observations within the classes has been visualised by wind roses, which show the relation between the station measurements and the classes obtained for the wind fields and indicate how well the wind observations are represented by the classes. While the classification with 15 clusters seems to be appropriate for a general understanding of the dynamics, the modelling of transport of air pollutants may require more detailed information and therefore a higher number of classes, which can be derived from the same hierarchical clustering tree.
ACKNOWLEDGEMENTS

The project REKLIP / MISTRAL is partly funded by the two cantons of Basel. It would not have been realized without the continuous support of its initiator, Dr. F. Gassmann. Additional data were kindly provided by the Swiss Meteorological Institute, Zürich and the Geographical Institute of the University of Basel.

REFERENCES


