Pure appl. geophys. 164 (2007) 1383–1396
0033–4553/07/071383–14
DOI 10.1007/s00024-007-0224-5

Pure and Applied Geophysics

Climatological Tools for Low Visibility Forecasting

Otto Hyvärinen,[1] Jukka Julkunen,[2] and Vesa Nietosvaara[3]

Abstract—Forecasters need climatological forecasting tools because of limitations of numerical weather prediction models. In this article, using Finnish SYNOP observations and ERA-40 model reanalysis data, low visibility cases are studied using subjective and objective analysis techniques. For the objective analysis, we used an AutoClass clustering algorithm, concentrating on three Finnish airports, namely, the Rovaniemi in northern Finland, Kauhava in western Finland, and Maarianhamina in southwest Finland. These airports represent different climatological conditions. Results suggested that combining of subjective analysis with an objective analysis, e.g., clustering algorithms such as the AutoClass method, can be used to construct climatological guides for forecasters. Some higher level subjective "meta-clustering" was used to make the results physically more reasonable and easier to interpret by the forecasters.

Key words: Low visibility, fog, clustering, forecast model reanalysis.

1. Introduction

Low visibility negatively affects flight operations. Especially small airports and airplanes without expensive approach equipment are greatly affected. In a case of low visibility, planes have to use an alternate airport which causes extra consumption of fuel. The threshold of poor visibility is set by convention to 1 kilometer. Visibility can decrease to less than 1 kilometer for a number of different physical reasons, e.g., by fog or by rain and snow showers. An experienced forecaster realizes which weather types are the most commonly associated with these events, and also becomes familiar with the climatological peculiarities of the areas within which he mostly has to work. This kind of climatological background is important because numerical weather prediction models are not always as useful as forecasters would like them to be. However, if a forecaster has 10 or even 20 airports under his responsibility, he cannot rely on personal experience alone. At the minimum, the forecaster needs

[1] Finnish Meteorological Institute, Remote Sensing for Weather Applications, Helsinki, Finland. E-mail: otto.hyvarinen@fmi.fi
[2] Finnish Meteorological Institute, Aviation and Military Weather Service, Rovaniemi, Finland. E-mail: jukka.julkunen@fmi.fi
[3] Finnish Meteorological Institute, Remote Sensing for Weather Applications, Turku, Finland. E-mail: vesa.nietosvaara@fmi.fi

distributions of some relevant parameters such as monthly distribution of wind direction and fog. Unfortunately, these classical distributions give only an approximation of low visibility weather parameters. More advanced statistical methods should be investigated.

Often only SYNOP observations (WMO, 1995) are available for low visibility forecasting. SYNOP observations have a rather coarse time scale (3 hours), which is not optimal for nowcasters. But sheer numbers of observations (about 80,000 observations from one airport for the period of 30 years) can be used to calculate distributions from these data. On the other hand, SYNOP observations are basically point measurements. To be able to visualize the synoptic situation of low visibility cases, some kind of spatial analysis is needed. Recently, various reanalyses of past observations with modern methods have been published. For example, TARDIF (2004) used NCEP-NCAR Reanalysis Project data to characterize the synoptic weather patterns associated with fog. In the present study, the ERA-40 reanalysis data (UPPALA *et al.*, 2005) generated by the European Centre for Medium-Range Weather Forecasts is used.

An objective method that shows the typical situations of low visibility can help forecasters to learn the climatology of new areas they are assigned to. Clustering algorithms are examples of such objective methods. In this method, a set of observations is divided into groups, so that members of the same group are more alike than the members of different groups (RIPLEY, 1996). Clustering methods can be divided into three main approaches: 1) Hierarchical methods, 2) partitioning methods, and 3) methods based on mixtures of parameterized densities (e.g., probability distribution). The first two approaches are commonly used in meteorological literature. Hierarchical methods result in trees or dendrograms, which can be easily visualized. A recent article for meteorological applications on this method was given by CHENG (2004). Partitioning methods include the well-known K-means algorithm. In K-means, samples are divided to a pre-assigned number of groups. Then, the group centers and memberships are chosen so that the sum of squared distances from each sample to the group center is minimized. However, this algorithm has some shortcomings. For example, the number of clusters has to be decided by the user, and discrete data cannot be used in this method. A recent article using partitioning method was by (GUTIÉRREZ *et al.*, 2004). The third approach of clustering assumes the samples come from the same mixture of sources and they can be parameterized. These sources are usually probability distributions but fuzzy logic technique can also be used. In this study, the AutoClass classifier (CHEESEMAN *et al.*, 1988) is used, handling both continuous and discrete observations, and automatically selecting the best number of clusters. AutoClass fits probability distributions to the data and searches for the clusters with the highest probability. Probability distributions can be either discrete nominal or continuous (real-valued). Distributions for continuous data are assumed to be either single normal or multinormal. If continuous data have a clear lower bound, then the log-normal distribution can be used for the analysis.

2. Observations and Methods

Present data consist of SYNOP observations obtained from the sixteen Finnish weather stations and main airports (Fig. 1) at every three hours for about 30 years from 1970 to 2000. Minimally, the SYNOP observations consist of various observations, including pressure, clouds, temperature at 2 meters height, relative humidity, wind, and weather situations. The geopotential and temperature information from the ERA-40 reanalysis data were analyzed at the pressure levels used in opearational forecasting such as 1000 hPa, 850 hPa, and 500 hPa. It is worth noting that the ERA-40 reanalysis data are available every six hours, not every three hours

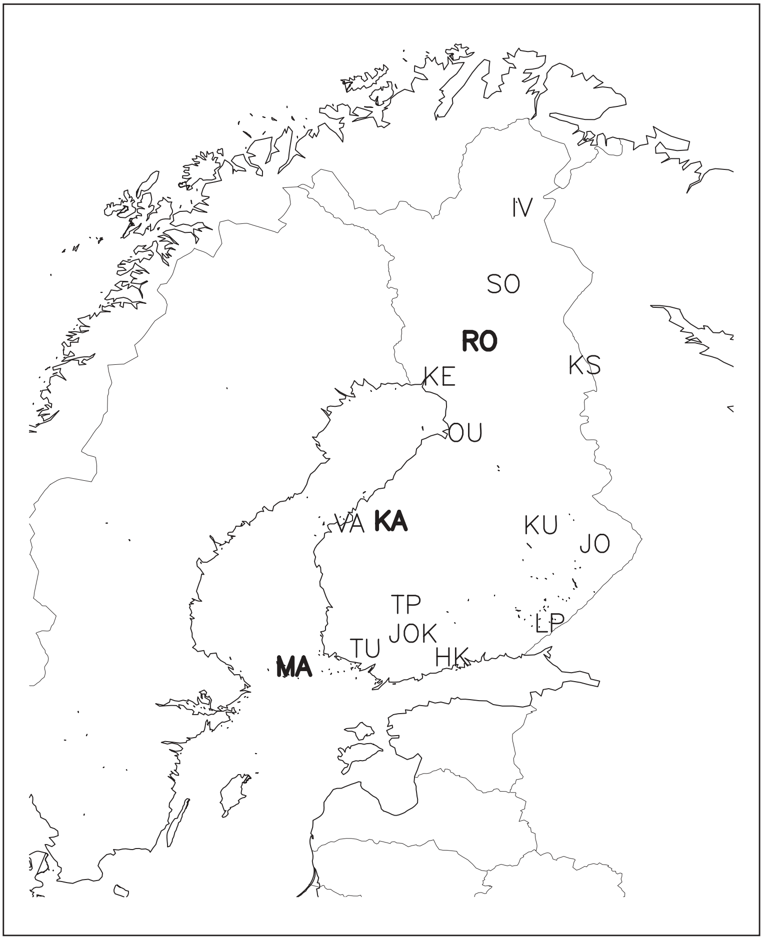

Figure 1
Finnish weather stations used in the study. Rovaniemi (RO), Maarianhamina (MA) and Kauhava (KA) were investigated in more detail.

as for the SYNOP observations. The analysis of data is performed using the R software (R Development Core Team, 2005).

2.1. Subjective Approach for Fog Climatology

In the analysis, in order to get classical distributions of visibility, wind speed and direction, and other related parameters, three airports are chosen to apply a subjective analysis of the synoptical observations. The airports, representing three different climatological regions, are: 1) Rovaniemi (RO) airport from northern Finland, 2) Maarianhamina (MA) airport from southwest Finland, and 3) Kauhava (KA) airport from western Finland (Fig. 1). The Rovaniemi airport is an elevated airport, and its altitude is approximately 195 m above sea level (ASL). It is well-known for its heavy fog periods, which often cause difficulties to aviation. The Kauhava airport represents a small airport in western Finland which is located at low altitude (42 m ASL) just at an intermediate zone between the lowlands of the coastal western Finland and slightly more elevated terrain in the east. The Maarianhamina airport is located in the inner parts of a relatively large island on the Northern Baltic Sea and represents a marine station. Its height is 2 m ASL and it is in the lower part of the island.

Weather forecasting in the above-mentioned airports is quite difficult. After a subjective analysis, the results are compared to the results of an objective analysis, and then an evaluation is made.

2.2. Objective Approach with AutoClass

The AutoClass uses a Bayesian approach, and its goal is to find the most probable set of class descriptions for a given data set and prior expectations. The use of priors automatically includes the Occam's razor to the model and prevents overfitting of the models. Other methods based on the maximum likelihood (e.g., K-means) determine classification that best predicts the data. The skill of these models to classify new data is often poor because they try to fit the learning set exactly. Full Bayesian solution is in practice too costly, and maximum *a posteriori* (MAP) parameter values are searched in the parameter space. The search has two levels. The model level search (the first level), is for the best number of clusters and the cluster probabilities of an individual case. In the second level, the MAP parameter values for probability distributions for each cluster are computed. Models are first rated by their best posterior probabilities, and then the best models and their MAP parameter values are reported. The AutoClass does not assign cases to clusters, but for each case it provides the probabilities of belonging to a cluster. These probabilities are not manipulated further and they are assigned to the clusters of highest probability.

In the present study, the AutoClass for 500 cycles took about one hour of the CPU time for each station and this 500 cycles is considered to be enough for the best classification was found much earlier.

Clustering was performed for all sixteen stations. Then, only low visibility observations were performed, i.e., when the visibility was less than 1 km. This means that we did not try to predict fog at this stage but we only wanted to see the existence of the fog type or other low visibility situations.

Before running AutoClass, the features for clustering and their probability distributions are chosen. The selection of features has a significant impact on the clustering, so great care is needed. For the present study, the amounts of low and total cloudiness, 2 m temperature, relative humidity, wind speed and direction, and the codes for present and past weather are chosen. This was based on empirical climatological knowledge of Nordic fog events and their dependence on weather. Temperatures near zero degrees Celsius usually favor certain fog types e.g., freezing fog. Cloudiness and wind were also expected to be good classifiers, since fog events on rather calm and clear conditions typically represent radiation fog events, while advective fog types are often observed in relation to certain large-scale weather patters. The probability models for these features are rather straightforward: Present and past weather have discrete nominal models and others have continuous models. In these calculations some conversions of data were needed. For example, present weather values from 00 to 99 are considered too detailed, so only the first digit is used. The wind u and v components with multinormal distributions are then used. This is a reasonable choice under the constraints of the AutoClass because richer selection of distributions is not available. Cloudiness and relative humidity in our study have a clear upper bound and are they are inverted so that their distribution is considered as log-normal. Note that, in the more general case, cloudiness and relative humidity have both lower and upper bounds; therefore, a simple log-normal or normal distribution can be quite inappropriate for many applications.

The rounding of the SYNOP messages can result in some uncertainties. This can cause some problems because any clustering method can find spurious clusters caused by rounding. This can be especially problematic with wind observations where limited resolution of wind direction might suggest false finger-like distributions of observations. We tried to remedy this by adding the Gaussian noise with standard deviation of half the bin to the wind observations, as suggested in AutoClass documentation.

After the clustering, ERA-40 data were used to visualize the average synoptic situations of all clusters for selected pressure and temperature levels. Then, the mean and standard deviation were computed for the ERA levels nearest in time to each occurrence. From the images of average conditions, an indication of the synoptic conditions associated with each cluster is then obtained.

The significance of the patterns that can be seen in the average synoptic images for the ERA data set is also explored. The 850 hPa temperature field in a geographic location for any single cluster is compared to the mean and standard deviation of the same field for the entire ERA data set. For the tests of statistical significance, a simple bootstrap system (WILKS, 1995) is used. As a metric for the distance between different distribution, the Bhattacharyya distance (BHATTACHARYYA, 1943) is used, because it uses information from both the mean and the standard deviation of any parameter. Then, for each grid point of every cluster, the following procedure was performed: First, the Bhattacharyya distance between the entire ERA data set and the cluster under investigation was calculated. Second, a distribution of the Bhattacharyya distances between the entire ERA data set and random data sets of the cluster under investigation was constructed. Third, the Bhattacharyya distance of original cluster was compared to the constructed distribution. If the Bhattacharyya distance of original cluster was larger than a certain percentage of the values in the distribution (depending on the rejection level), it is concluded what is seen in the average synoptic images is not just random fluctuations of data.

3. Results and Discussion

3.1. Subjective Approach

Results showed that the occurrence of low visibility values at Rovaniemi airport was very high. On the other hand, low visibility observations associated with calm conditions are rare. That indicates that radiative fog at the Rovaniemi airport was not very common and the fog occurrence is likely due to weather patterns and advection. The reason for this is likely that Rovaniemi airport is located on top of a hill therefore, orographical effects are likely to influence the air flow resulting in foggy weather. Most foggy weather events which occur at this airport are during southwesterly moist air flow. Additionally, because of its northern location, some of the very low visibility observations may be caused by snowfall.

At the Kauhava airport, observations suggest that there are many low visibility observations during calm conditions. Particularly in autumn morning hours low visibilities may occur due to fog occurring as a result of radiative cooling. Because most low visibility observations are during almost windless situations, it is suggested that advection fog type occurrence is quite rare. It is obvious that Kauhava airport lies in the area where radiative cooling is intense during autumn.

The Maarianhamina airport has a quite different seasonal distribution of very low visibilities compared to the other airports. The main difference is that most of the low visibility observations are in springtime, and they likely occur because of radiative cooling and advection.

3.2. Objective Approach

The number of clusters given by the AutoClass for each station varied from 12 to 23 clusters, while most of the stations had less than 16 clusters. This variation seems quite reasonable. The station that had most of the clusters (the Kauhava airport) has the reputation of being difficult to forecast for low visibilities. In all cases, the smallest clusters were substantially smaller than the largest ones. These clusters are ignored in many cases, because it is quite likely that these clusters are artifacts as a result of the result of the inability of AutoClass to model all the details of data with its inherent probability distributions. Thus small clusters probably do not give new insights to the climatology of low visibility.

There are areas of statistical significance in ERA-40 images (not shown), except for the very small clusters that are suspicious anyway. In some cases, these areas encompassed the whole area from which ERA-data were used; in most cases they covered most of the area. Therefore, weather patterns seen in ERA-40 images have statistical significance, and cannot be dismissed as random fluctuations.

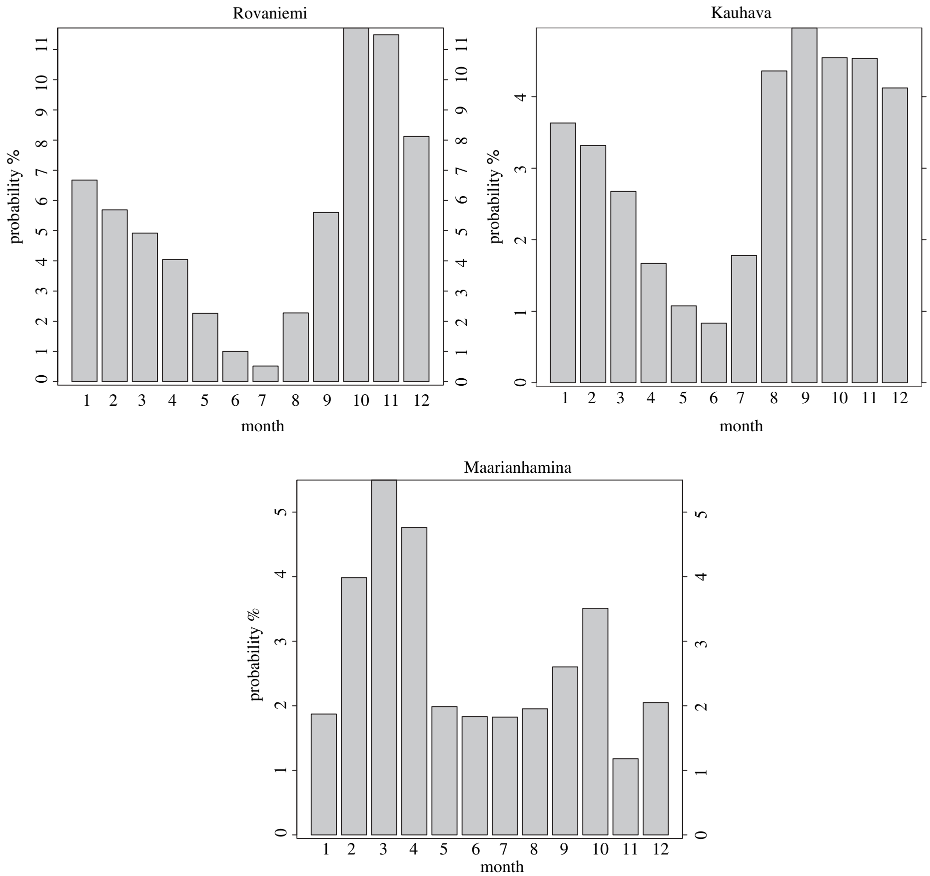

Figure 2
Monthly frequency of low visibility at the Rovaniemi, Kauhava and Maarianhamina airports.

3.3. Interpretation of the Results

Both the subjective and objective methods used in this study help us to interpret the nature of fog events at the three airports.

The simple monthly probability diagrams for low visibility already show some basic characteristics of fog occurrence at each individual station (Fig. 2). Both at Rovaniemi and Kauhava, fog occurrence is high in autumn (from October to November) and also in winter (from December to March), but at the Kauhava airport, fog is likely to occur in late summer (August). At the Maarianhamina airport, fog occurrence is at its peak in February, March and April.

A more detailed subjective analysis can be made with the help of low visibility frequency charts (Fig. 3). These charts reveal not only the monthly distribution already shown in Figure 2, but also the preferred time for fog occurrence. For the Rovaniemi airport, the autumn maximum for fog shows low visibilities occurring at any time of day, though the maximum occurrence does not change significantly during the morning hours. On the contrary, during spring months, fog normally occurs only at night and in the morning. At the Kauhava airport, the figure shows other characteristics of the fog: morning fog likely occurs from late summer to early

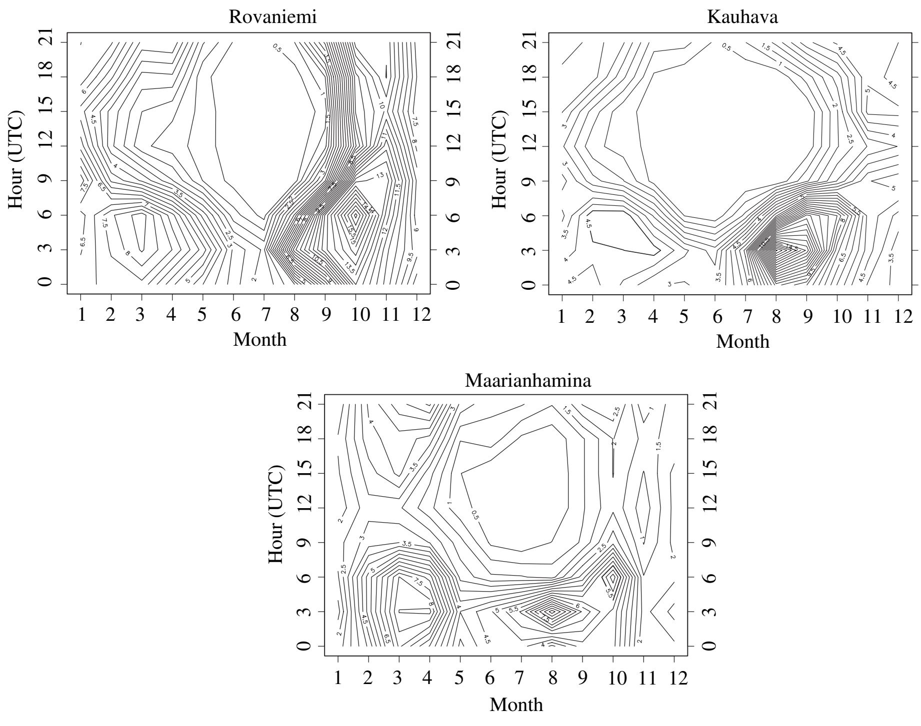

Figure 3
Annual frequency of low visibility at the Rovaniemi, Kauhava and Maarianhamina airports.

autumn. A secondary maximum can also be seen from February to April. At the Maarianhamina airport, the low visibilities are highly concentrated for late night and early morning hours throughout the year. Day time fog occurrence for this location is remarkably low.

The above results suggest some speculations on how many of the fog occurrences are explained by classifying them as radiative or advective fogs, and how much of the variability is due to the different topography and environmental factors. It is also important to know if these features can be determined by using the objective classification methods, e.g., clustering techniques.

The objective interpretation by ‘‘eyeballing'' cluster characteristics is used in the analysis. We especially considered wind, cloudiness, and SYNOP code distributions, temperature, and temporal distribution of occurrence. As it turned out, it was reasonably easy to classify clusters to three ‘‘meta-clusters'': advection fog, radiation fog and non-fog cases. The distinction between advection and radiation fog clusters was in fact surprisingly clear: while radiation fog cases typically reflect calm or weak wind conditions at surface with mostly clear skies, the advective fog cases show an association with moderate winds (4–7 m/s or even more) with overcast skies and a tendency to be independent of the time of day. Non-fog cases were typically differentiated very well by the clustering: the non-fog cases (such as low visibilities caused by snowfall) were concentrated in the clusters representing only a small percentage of all cases. An example from the Maarianhamina airport is shown in Figure 4. As with the choice of features for the clustering, we constructed the meta-clusters reflecting our knowledge of Nordic fog occurrences. It is conceivable that some of the details would change if this procedure was adopted in a vastly different climate, but essential parts of the procedure would remain the same.

Clearly, at these three stations, radiation fog dominates during summer, while the advection fog is more dominant during winter. Radiation fog is also highly dependent on the time of day, being most likely to occur at dawn, while advection fog has not such a strong dependency on the time of day. The Kauhava airport shows a remarkable exception in this respect: Purely advective fogs decrease and a vast majority of cases represent radiation fog.

The ERA-40 reanalysis data proved to be extremely useful in meta-clustering: The mean synoptic conditions for each cluster helped us to decide whether the cluster belonged to an advection or radiation fog meta-cluster. As an example of mean synoptic conditions with ERA-40 reanalysis data, images for an advection fog cluster representing the Maarianhamina airport are shown in Figure 5. This example illustrates typical conditions for the occurrence of advection fog/stratus. The following characteristic properties can be found:

— A relatively warm airmass at the lower troposphere, which can be seen as a local ridge of temperature at 850 hPa over the area of interest.
— Upper level ridge to the east of the area, indicating warm air advection over the area.

For each station then an objective evaluation is done. Of particular interest were the clusters with some systematic fog occurrence patterns for individual stations. These systematic patterns could be either a specific fog type bound to a particular

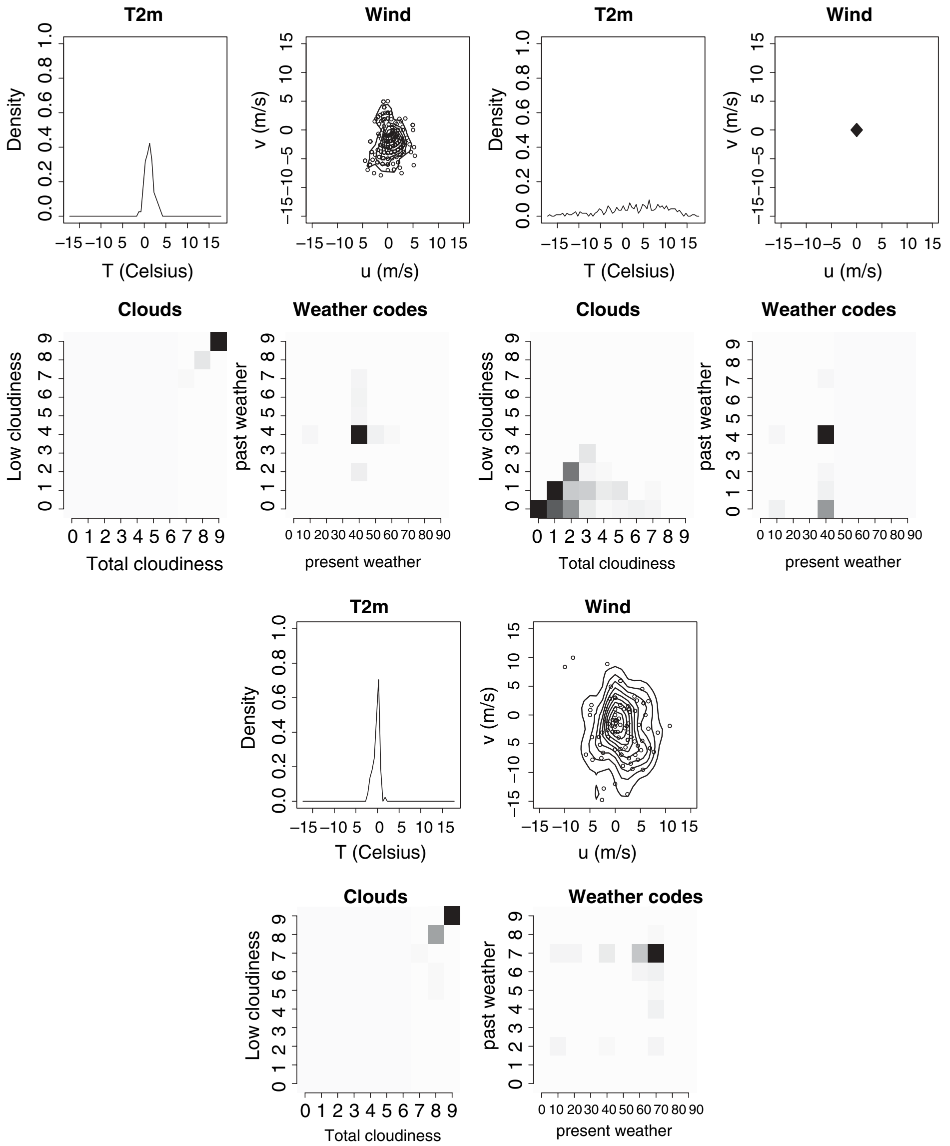

Figure 4
Three typical representatives of ‘‘advection fog’’, ‘‘radiation fog’’ and ‘‘non-fog’’ (snowfall) clusters for the Maarianhamina airport. Roughly, SYNOP codes of present weather are for fog from 40 to 49, and for snowfall are from 70 to 79. The SYNOP code of past weather for fog is 4 and for snow 7. For SYNOP codes, dark color indicates high density, light color low density.

season of the year or a fog-type associated with specific synoptic situations (as shown by ERA-40 data).

At the Rovaniemi airport, three major fog types are found: (1) Advective late autumn-wintertime fog (represents roughly 60% of all fog occurrences), (2) an early autumn radiation fog (20% of the total number of fog occurrences) and (3) a wintertime rime fog associated with weak flow and high pressure ridge or high pressure center in or to the south of northern Finland (also 20% of the total number of fog occurrences). These fog types truly represent the fog "archetypes" for this particular airport. An example of surface pressure pattern for type 3 fog is shown in Figure 6a.

Another example for the radiation fog at the Kauhava airport is considered as a sub-type for a radiation fog, which is the most typical fog type at Kauhava (FRISK, pers. Comm., 2005). This fog type can be identified as "a radiation fog associated with a secondary low formation in the Bay of Bothnia." This meta-cluster consists, in fact, of three clusters, and for each of them a distinct flow pattern is clearly seen in ERA-40 mean synoptic charts. The flow pattern suggests that when there is a less intense low pressure centered over the Bay of Bothnia (Fig. 6b), morning radiation fog is often observed. This pattern and its connection to the morning fog events has also been verified and described by the operational forecasters making daily forecasts for the Kauhava airport.

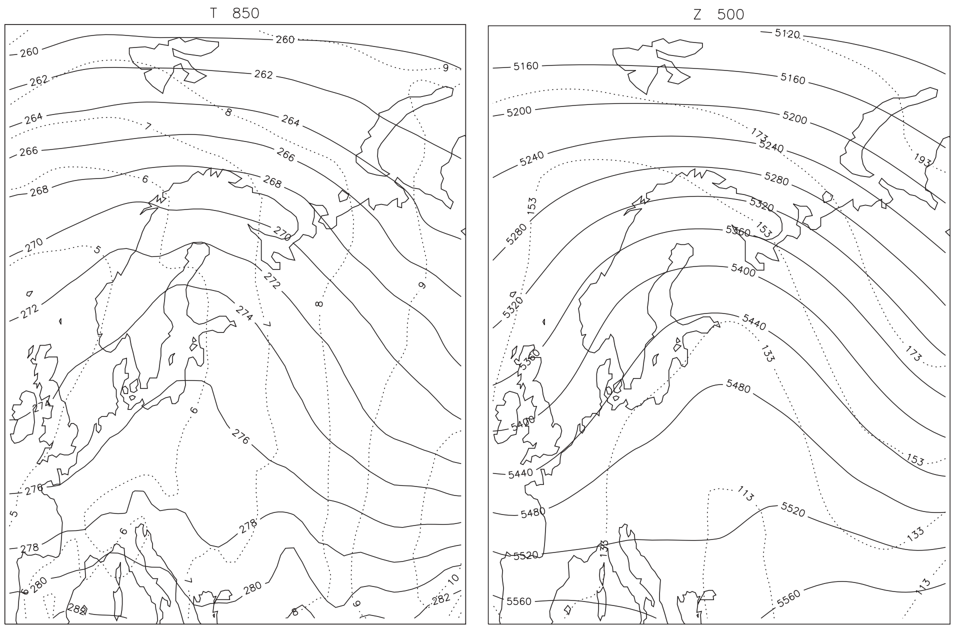

Figure 5
Synoptic conditions for one arbitrary advection fog cluster of the Maarianhamina airport. Solid lines indicate mean values, dotted lines standard deviation.

The clustering results can be combined with the other diagrams for improved results. An example is shown in Figure 7, where we have combined clustering results with the monthly probability diagram. The resulting image shows when the low visibilities likely occur, and what kind of low visibility cases exist.

4. *Summary and Conclusion*

Aviation forecasters need to know the low visibility and fog climatologies of each airport for safe aircraft landing and take-off. The purpose of this study was to investigate possible new tools that use clustering methods and ERA-40 reanalysis fields. In this study, both subjective and objective methods are used, helping us to interpret the nature of fog events at several airports. Three Finnish airports are used in the analysis, representing different climatological regions.

The results suggest that the combination of the subjective analysis and a clustering algorithm (e.g., a AutoClass method) could be used to construct climatological guidelines for forecasters. It is also found that results were in

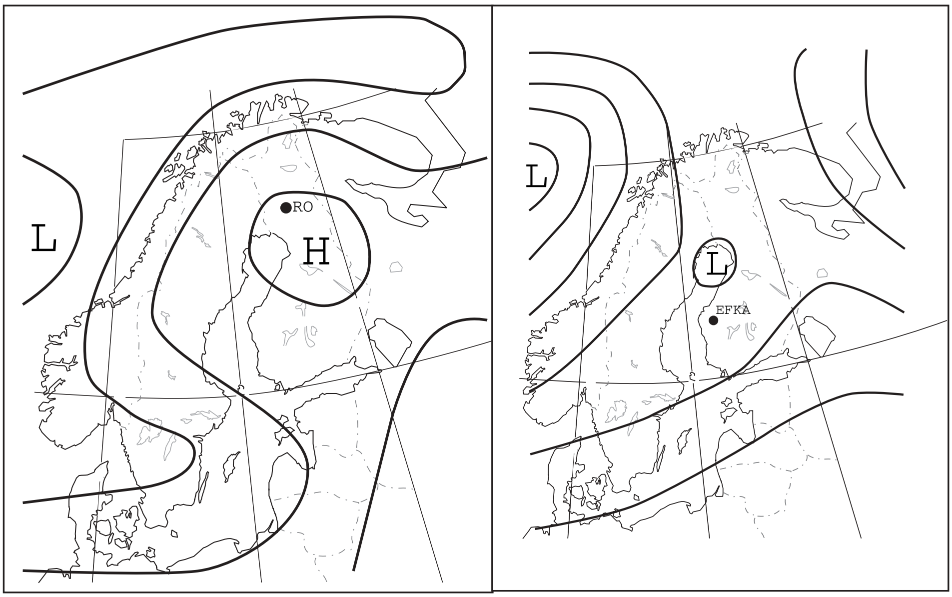

Figure 6
Schematic images showing the mean surface pressure conditions in Scandinavia for two fog types: a) Fog type at the Rovaniemi airport "A wintertime rime fog within a high pressure ridge". b) Fog type at the Kauhava airport "An autumn morning radiation fog associated with a secondary low pressure on the Bay of Bothnia". Isobars are in solid black contours. Locations of the Rovaniemi (RO) and Kauhava (KA) airports are shown with a black dot.

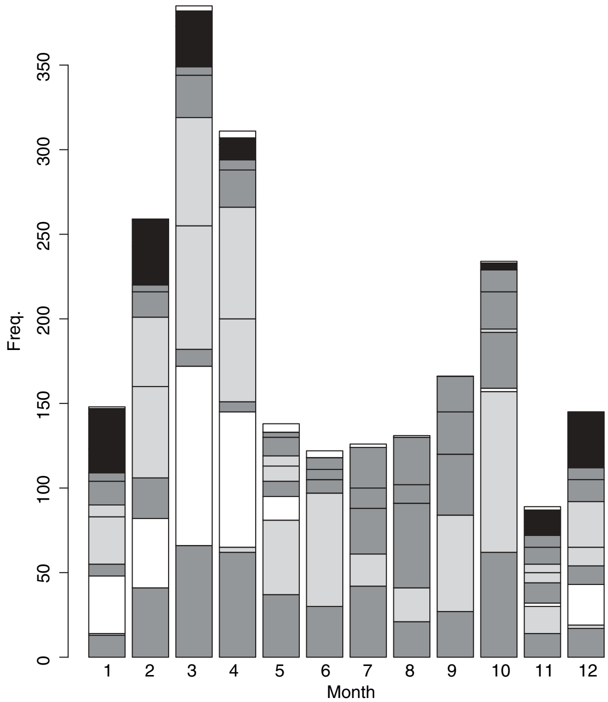

Figure 7
Monthly frequency of low visibility at the Maarianhamina airport augmented with the clustering information. Every bar segment represents a cluster. Clusters have been subjectively classified to ''meta-clusters'' and colored accordingly: light gray clusters represent occurrences of advection fog, dark gray clusters represent occurrences of radiative fog, black clusters represent non-fog events and white clusters are left unclassified.

agreement with forecasters' decisions. This supports our idea to use the integrated (combined) methods in training the forecasters.

It must be emphasized that the clustering results depend on the method used. AutoClass has a rich family of distributions, as mentioned above, but real distributions of some parameters can still be quite different. In the present work, some higher level subjective ''meta-clustering'' is performed in order to make the results more physically reasonable and easier to interpret by the forecaster.

Our future plans can be divided into three parts: 1) The method refinement, 2) broadening the scope, and 3) applying the method to the operational environments. Selection of features for clustering is quite *ad hoc* and should be revisited for future applications. For example, pressure information in the current clustering analysis is not used. It will also be interesting to widen the scope of the present investigations and apply the new methods to other airports in Finland and use more detailed observations (e.g., METAR) over the orographic environments. The quantitative way of justifying the choice of our clustering method also needs to be researched, and a detailed comparison methodology with different clustering techniques (e.g., K-means) needs to be developed.

References

BHATTACHARYA, A. (1943), *On a measure of divergence between two statistical populations defined by their probability distributions*, Bull. Calcutta Math. Soc. *35*, 99–110.

CHEESEMAN, P., KELLY, J., SELF, M., STUTZ, J., TAYLOR, W., and FREEMAN, D. (1988), *AutoClass: A Bayesian Classification System*. In Proc. Fifth Internat. Conf. on *Machine Learning*, Ann Arbor, MI. June 12–14 1988. Morgan Kaufmann Publishers, San Francisco, pp. 54–64.

CHENG, C.S., AULD, H., LI, G., KLAASSEN, J., TUGWOOD, B., and LI, Q. (2004), *An automated synoptic typing procedure to predict freezing rain: An application to Ottawa, Ontario, Canada*, Weather and Forecasting *19*, 4, 751–768.

GUTIÉRREZ, J. M., COFIÑO, A. S., and CANO, R. (2004), *Clustering methods for statistical downscaling in short-range weather forecasts*, Mon. Wea. Rev. *132*, 9, 2169–2183.

R. DEVELOPMENT CORE TEAM (2005), *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

RIPLEY, B.D. *Pattern Recognition and Neural Networks* (Cambridge University Press, Cambridge 1996).

TARDIF, R. (2004), *Characterizing fog occurrences in the Northeastern United States using historical data*, 11th Conf. Aviation, Range and Aerospace Meteorology, Hyannis Port.

UPPALA, S.M., KÅLLBERG, P.W., SIMMONS, A.J., ANDRAE, U., DA COSTA BECHTOLD, V., FIORINO, M., GIBSON, J.K., HASELER, J., HERNANDEZ, A., KELLY, G.A., LI, X., ONOGI, K., SAARINEN, S., SOKKA, N., ALLAN, R.P., ANDERSSON, E., ARPE, K., BALMASEDA, M.A., BELJAARS, A.C.M., VAN DE BERG, L., BIDLOT, J., BORMANN, N., CAIRES, S., CHEVALLIER, F., DETHOF, A., DRAGOSAVAC, M., FISHER, M., FUENTES, M., HAGEMANN, S., HÓLM, E., HOSKINS, B.J., ISAKSEN, L., JANSSEN, P.A.E.M., JENNE, R., MCNALLY, A.P., MAHFOUF, J.-F., MORCRETTE, J.-J., RAYNER, N.A., SAUNDERS, R.W., SIMON, P., STERL, A., TRENBERTH, K.E., UNTCH, A., VASILJEVIC, D., VITERBO, P., and WOOLLEN, J. (2005), The ERA-40 re-analysis., Quart. J. Roy. Meteorol. Soc., *131*, 2961–3012.

WILKS, D.S. *Statistical Methods in the Atmospheric Sciences: an Introduction* (Academic Press, San Diego, 1995).

WORLD METEOROLOGICAL ORGANIZATION (1995), *WMO-No 306, Manual on Codes, International Codes*, Volume I.1, Geneva.

(Received April 27, 2006, accepted November 1, 2006)
Published Online First: May 16, 2007