Tellus (1981), *33*, 113–122

Objective analysis of humidity by the optimum interpolation method

By JAN VAN MAANEN,[1] *Royal Netherlands Meteorological Institute (KNMI), Postbus 201, 3730 AE De Bilt, The Netherlands*

(Manuscript received February 29; in final form September 4, 1980)

ABSTRACT

Covariances and correlations of the humidity mixing ratio are calculated which are necessary for the construction of an objective humidity analysis scheme with the optimum interpolation method. The statistics are derived for winter and summer and for three levels, 850, 700 and 500 mbar. It appears that the best description of the humidity correlation function is a function of the form $\rho_0 \exp(-bs)$, s being the distance between two locations. The constants ρ_0 and b are dependent on the season, level, and guess field used in the analysis. A representative value for b^{-1} is 400 km. From the determination of ρ_0, which varies between about 0.5 and 0.9, one obtains a quantitative impression of the observation error.

The results obtained are used in a humidity analysis system and examples of analyses of the absolute humidity are given.

1. Introduction

In this paper we present some results on the determination of the humidity correlation function, and apply these results to the objective analysis of the humidity field.

Several papers have dealt with the problem of humidity analysis. Atkins (1974) described a successive correction method with two scans. The weight function she used depends not only on the distance between the grid point being analysed and the observation point, but also on the magnitude and direction of the gradient of the background humidity field. This serves the purpose of keeping details in the analysis, especially near fronts with a strong humidity gradient.

Kästner (1974) also has described an analysis made with the correction method. Not only direct humidity observations by means of radiosondes are used, but also the surface observations and calculations of various flow parameters like vertical velocity are taken into account in order to provide an estimate of the upper-air humidity.

In general it can be stated that the humidity, as an atmospheric parameter, has special properties. This should be kept in mind while constructing a humidity analysis scheme. The radiosonde instruments measuring humidity are not very accurate. This hampers the analysis, while attempts to estimate the humidity from other observed or calculated quantities—such as cloudiness or vertical motion—also encounter difficulties, because no truly reliable observations exist from which to construct a humidity value. Also, the humidity field has a large variability on short spatial scales. An observation of, say, 500 mbar height is much more representative for its surroundings than a humidity observation. This is because the humidity field is influenced more strongly by meso-scale weather systems than the mass field.

If one wants to analyse the humidity field, the first question is which quantity is to be analysed? There are several candidates, such as relative humidity, mixing ratio, logarithm of mixing ratio, dewpoint, dewpoint depression, and so on. As these variables have strongly non-linear relationships, the resulting analysis depends on the choice made. The principal step of optimum interpolation consists of minimizing a sum of squares, and it seems

[1] Present affiliation: European Centre for Medium Range Weather Forecasts, Reading, United Kingdom.

 0040-2826/81/020113-10$02.50/0

preferable that the differences between the observations and the guess-field have a normal distribution. Also it is advantageous for the humidity variable to be the same as the one used in the model, as this simplifies the necessary checking for oversaturation in the analysis, initialization and first time step of the model.

Ideally the humidity variable should be independent of temperature and also have a normal distribution. None of the variables mentioned above fulfil both these requirements even in an approximate way. It was decided to use the mixing ratio, for the following reasons: we found that the deviation of its distribution from normality is not large, and it is the variable normally used in forecasting models. (Strictly speaking, in most forecasting models specific humidity is used instead of mixing ratio. The difference between these two variables is negligible, even more so if the accuracy with which the humidity can be analysed is taken into account.) The mixing ratio is correlated with the temperature, however. We turned this correlation to our advantage as this enables us to derive the guess-field for the analysis from the temperature field (see Section 3).

In the next section some characteristics of the observations are discussed, especially the observation error variances. In Section 3 the correlation function is determined, whereas in Section 4 two examples of analyses are given. Section 5 consists of a summary and conclusions.

2. The observation error

The standard deviation of the observation error of the radiosonde humidity sensor is difficult to assess in a direct way, therefore the random error was estimated from the correlation between observations of nearby stations.

In this section the observation error for the Dutch station de Bilt (52°N, 5°E) is determined and it is assumed that the result also applies for other stations in middle and high latitudes.

All statistical data of this paper were calculated separately for the winter and summer season. The data used for the winter period are all radiosonde humidity observations of December 1976 (apart from the first 16 days), January, February, December 1977, January and February 1978. For the summer period the observations of the 6 months June, July and August of 1977 and 1978 are used.

For every radiosonde station in the vicinity of de Bilt we determined the correlation between its observations and the observations of de Bilt. In Fig. 1 a plot of the correlation of 850 mbar versus distance is shown for the winter period, in Fig. 2 the same is plotted for the summer period. The dots were fitted with the function $\rho(s) = \rho_0 \exp(-bs)$ where s is the distance from de Bilt, ρ the correlation with the observation of de Bilt, and ρ_0 and b are fitting constants. The constants ρ_0 and b were determined in such a way that the sum of squares of the vertical distances between the function $\rho(s)$ and the dots of Figs. 1 and 2 was minimized. In the next section the correlation function will be discussed in some greater detail.

Because of the existence of observation errors, the coefficient ρ_0 is not equal to one and can be shown to be

$$\rho_0 = \frac{\text{var}(r)}{\text{var}(r) + \text{var}(\varepsilon_r)}$$

where r is the true value of the mixing ratio and ε_r is the corresponding observation error. The error variance var (ε_r) has a contribution from the true instrumental error and a contribution from variability of the field on scales too small to be resolved by the observational network, as far as this variability differs significantly from the one implied by the correlation model.

One should bear in mind that with respect to the small-scale error the situation is somewhat complicated, however. For explanatory purposes, suppose the observations were without instrumental error, and the correlation model $\exp(-bs)$ were exactly valid for the true field and for all distances. The fitting procedure mentioned above would find, independent of the mutual distances of the stations used, that the constant ρ_0 is exactly one, i.e. the correlation is one at distance zero. This is the case although clearly there *is* variance at scales too small to be determined by the observational network. The existence of errors of the measuring instrument makes the value of ρ_0 less than one, but still the observed correlations would exactly fit the function $\rho_0 \exp(-bs)$. This reasoning implies that the deviation of ρ_0 from one is *not* caused by the effect of instrumental errors and errors with scales too small to be resolved, but by the combined effect of instrumental errors and the *deviations* of the

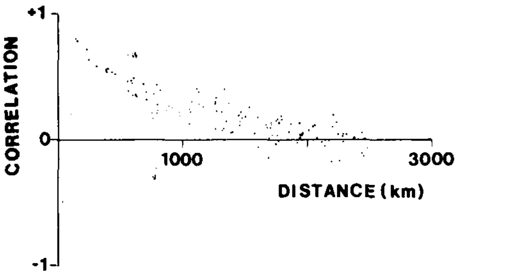

Fig. 1. Graph of the correlation of radiosonde observations of absolute humidity at 850 mbar vs distance from de Bilt, for the winter period.

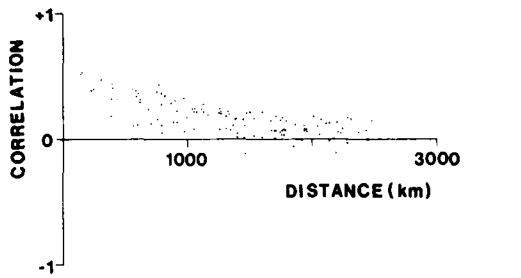

Fig. 2. As Fig. 1, for the summer period.

correlation behaviour at small distance between the true atmospheric field and the correlation model.

Note that once a correlation model (like ρ_0 exp $(-bs)$) is chosen, the value of ρ_0 to be used in the analysis is exactly the one found by the fitting procedure. The interpretation of the deviation of ρ_0 from one as being caused by the effect of "measuring and small-scale errors" is not quite correct, however, but for ease of formulation it is retained in the rest of this paper.

The variance of the humidity observation is, in fact, var (r) + var (ε_r). From the determination of the correlation function constant ρ_0 and the observation variance, the variance of the "true" humidity field (without small-scale effects) and the variance of the observation error (including small scale effects) can be determined. It is impossible, and also unnecessary, to separate the effects of instrumental errors from effects of small-scale errors, as the distance between de Bilt and nearby observations is of the same order as the distance between grid points in a numerical model of the atmosphere used for forecasting the large-scale mass and motion fields. The results for summer and winter are shown in Table 1.

These results are, strictly speaking, valid only for de Bilt. An independent check was provided by computing the same figures for the station Topeka in the U.S.A. (39° N, 96° W), and this check resulted in figures of the same order of magnitude.

The observation error at 850 mbar is much larger in summer than in winter. It is very unlikely that this is an instrumental effect caused, for instance, by higher summer temperatures.

The low correlations in summer at 850 mbar were studied more closely to determine the causes of this phenomenon. This was done by considering the surface observations. These observations have a high accuracy as compared with the radiosondes, and the network is denser. The surface

Table 1. *Values of ρ_0 and* b *obtained by fitting the function ρ_0 exp* (−bs) *to the correlations, in case climatology is used to obtain the guess value. Also given is the observation variance, and its splitting into the variance of the true humidity value and the variance of the error component (dimensionless)*

Season	Level	ρ_0	$1/b$(m)	Variance of observation	Variance of true humidity	Variance of error
Winter	850	0.95	730.10^3	$1.81.10^{-6}$	$1.73.10^{-6}$	$0.085.10^{-6}$
	700	0.74	550.10^3	$0.85.10^{-6}$	$0.63.10^{-6}$	$0.22.10^{-6}$
	500	0.93	480.10^3	$0.11.10^{-6}$	$0.10.10^{-6}$	$0.008.10^{-6}$
Summer	850	0.54	990.10^3	$2.95.10^{-6}$	$1.60.10^{-6}$	$1.35.10^{-6}$
	700	0.86	440.10^3	$1.87.10^{-6}$	$1.61.10^{-6}$	$0.26.10^{-6}$
	500	0.78	330.10^3	$0.24.10^{-6}$	$0.19.10^{-6}$	$0.054.10^{-6}$

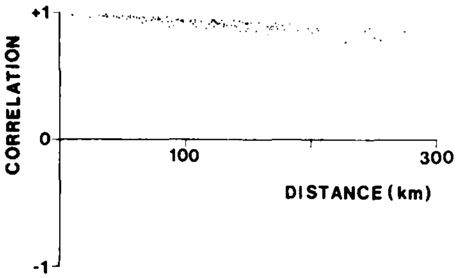

Fig. 3. Graph of correlation versus distance for observed surface mixing ratio for the Netherlands, for the winter period.

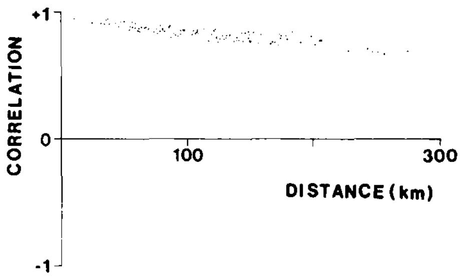

Fig. 4. As Fig. 3, for the summer period.

observations of the Netherlands from the same period were considered, and for every possible station pair the correlation of r was calculated. The plots of correlation vs distance are shown in Figs. 3 and 4. It appears that the surface correlations are somewhat higher than the 850 mbar correlations, reflecting the higher accuracy of the measuring instruments used for the determination of the surface humidity. In summer, the correlation at a distance of 300 km at the surface is about 0.65, and because of the high accuracy of the surface observations (which is reflected in the very high correlation at short distances) it is certain that the low correlation of 0.65 is a true physical effect, and is not caused by instrumental errors. If it is assumed that the correlation structure at 850 mbar is not very different from the one at the surface, we can now be sure that a substantial fraction of the high observation error in summer is caused by features of the humidity field too small to be resolved by the radiosonde observation network. These features range from individual clouds to meso-scale systems with a scale of about 100 or 200 km.

3. The correlation function

First, we will give a short derivation of the equations of optimum interpolation equations. A more detailed discussion of these equations can be found in Gandin (1963) or in Rutherford (1972).

Let us introduce the following quantities:

A_i^o, $i = 1, \ldots, n$ is the observed value of r at station i.

A_i^p, $i = 1, \ldots, n$ is the predicted or guess-field value of r at station i.

A_k^p is the predicted or guess-field value of r at the grid point being analysed.

A_k^i is the analysed value of r at the grid point being analysed.

a_i^o, $i = 1, \ldots, n$ is the observed value of r at station i minus its true value (i.e. the observation error).

a_i^p, $i = 1, \ldots, n$ is the error of the guess-field at station i.

a_k^p is the error of the guess-field at the grid point being analysed.

a_k^i is the error in the analysis.

n is the number of stations that is selected for use for the analysis of grid point k.

The analysis equation is

$$A_k^i = A_k^p + \sum_{i=1}^{n} w_{ik}(A_i^o - A_i^p) \tag{1}$$

which is equivalent to

$$a_k^i = a_k^p + \sum_{i=1}^{n} w_{ik}(a_i^o - a_i^p) \tag{2}$$

The weights w_{ik} are determined by requiring that the statistical expectation of a_k^{i2} be a minimum. If by $\langle a \rangle$ we denote the seasonal average of a the weights are the solution of the following set of linear equations:

$$\sum_{j=i}^{n} (\langle a_i^o a_j^o \rangle + \langle a_i^p a_j^p \rangle) w_{jk} = \langle a_i^p a_k^p \rangle, \quad i = 1, \ldots, n \tag{3}$$

The covariances $\langle a_i^p a_j^p \rangle$ and $\langle a_i^p a_k^p \rangle$ are functions of the observed and analysed variable, and the location of i, j and k. The covariance of the humidity field will be strongly dependent on latitude, but it will be assumed that the correlation function

$$\rho(s) = \frac{\langle a_i^p a_j^p \rangle}{[\langle a_i^{p2} \rangle \langle a_j^{p2} \rangle]^{\frac{1}{2}}} \tag{4}$$

is only a function of the distance s between the locations i and j, and not a function of the place on the earth or the direction of the line between i and j. The correlation function need not be constant with height, however.

The correlation function is dependent on the first-guess used, of course. In the previous section the correlation function was determined on the basis of climatology as a first-guess. In general climatology will not be very suitable for use as a first-guess, because the correlation distance for humidity is low and in many oceanographic areas the first-guess remains virtually unchanged. This leads to systematically high relative humidities in cold troughs and low ones in warm ridges. It is therefore more logical to use either the prediction from a numerical model or information from the temperature field. Taking a numerical forecast as guess-field can have the advantage that the guess-field is accurate, but the calculated correlation functions are model-dependent and need to be redetermined every time the model is substantially changed. Therefore as a starting point the temperature field was used to construct a background field for the humidity analysis, instead of a numerical forecast.

The guess value of the mixing ratio was derived as follows. All observations from a certain level and season were taken into consideration, and divided into classes depending on the observed temperature. The class width was 1 K. For all observations the mixing ratio was calculated, and averaged over the observations in the same class. In this way a function was derived that provides an estimate of the mixing ratio from the temperature. An advantage of deriving the guess-field from the temperature in this way is that the difference between the humidity observation and the humidity guess-field is uncorrelated with the temperature. The guess-value of the humidity as function of temperature is shown in Fig. 5. Expressed in relative humidity, the standard deviation about the average humidity in every 1 K class varied from 20% at low temperatures to 25% at high temperatures.

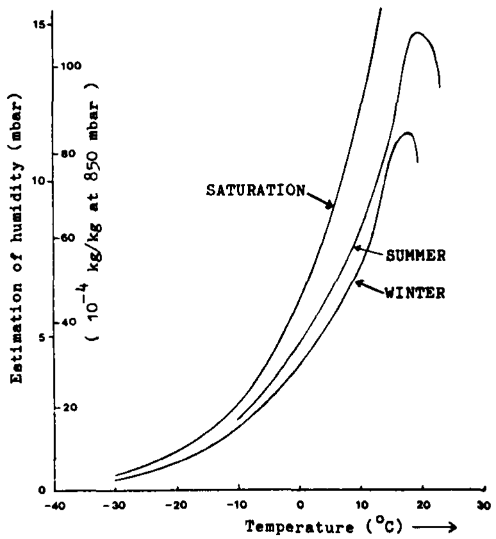

Fig. 5. The functions used for estimating the first-guess humidity at 850 mbar from the 850 mbar temperature. The graphs labelled winter and summer are used in the corresponding season. For comparison, the saturation curve is also shown. The vertical axis is labelled in vapour pressure (left) and mixing ratio at 850 mbar (right).

The next step is the determination of the correlation function for the case the guess-field is determined with the above mentioned method. The correlations of the difference between the observation and the guess-field of every station near de Bilt with de Bilt itself are shown in Figs. 6 and 7, and these correlations are fitted again with a function of the form $\rho(s) = \rho_0 \exp(-bs)$ as in Section 2.

We will mention one theoretical inconsistency of this function. Strictly taken, s is a function of four variables, the coordinates (x_i, y_i) and (x_j, y_j) of the locations of i and j, with

$$s = [(x_i - x_j)^2 + (y_i - y_j)^2]^{\frac{1}{2}}. \tag{5}$$

The covariance function used is equal to

$$\rho(s) = \rho(x_i, y_i; x_j, y_j) = \rho_o \exp[-b\{(x_i - x_j)^2 + (y_i - y_j)^2\}^{1/2}]. \tag{6}$$

Now, if the humidity field is differentiable with respect to x, so should be its covariance. The inconsistency is that the covariance $\rho(s)$ as a function of x_i is not differentiable to x_i at all points. For instance, $\partial\rho(s)/\partial x_i$ is not defined for $i = j$. This

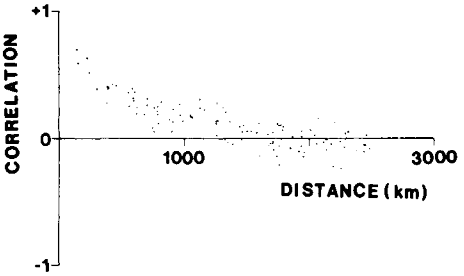

Fig. 6. Correlation vs distance from de Bilt when temperature is used for the determination of a guess value of the mixing ratio, for the winter period.

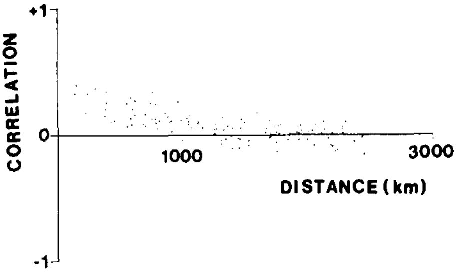

Fig. 7. As Fig. 6, for the summer period.

is more or less a problem in analyses of the height field, because observations of the wind are interpreted as observations of the height gradient and directly used for estimating the height, and vice versa. Because in the humidity analysis the humidity gradients are not involved in any way, the exact behaviour or the differentiability of the correlation function near $s = 0$ is not important, and considerations of fit with the observed correlations should be given preference when choosing a function.

The function $\rho_0 \exp(-bs)$ was used because it fits the data better than the function $\rho_0 \exp(-bs^2)$. There are no indications in the correlation graph for the correlation to drop below zero at large distances as is sometimes found for correlations of 500 mbar heights. The function $\exp(-bs)$ is positive definite, because the two-dimensional spectrum is positive definite. This spectrum $S(k)$ of $\exp(-bs)$ is given by (see Gandin, 1963; Gradshteyn, 1965)

$$S(k) = \int_0^\infty e^{-bs} J_0(ks) s\, ds = b(b^2 + k^2)^{-3/2} \tag{7}$$

A careful inspection of charts on which the correlation with de Bilt was plotted for every station on the location of this station revealed only small deviations from isotropy. Therefore it was assumed that the correlation function is isotropic. Also it is assumed that the correlation function is homogeneous, so the correlation between two locations is only a function of the distance between the two locations.

In Section 2 the observation error variance due to small-scale features of the humidity field was considered. When the guess-field is derived from the temperature field, a part of this variance will be explained by the guess-field. Thus it is to be expected that the total error variance will be lower, and it can also be expected that the variance of the deviations from the guess-field is lower. This is confirmed by the data in Table 2. For the analysis the ratio of the two variances is important and this information is contained in the value of the fitting constant ρ_o, also included in Table 2.

Table 2. *Values of* ρ_0 *and* b *obtained by fitting the function* ρ_0 *exp* (−bs) *through the correlations, in case the guess-field is derived from the temperature field. Also included is the variance of the difference between the observation and the guess-field (dimensionless)*

Season	Level	ρ_0	$1/b$(m)	Variance of difference between observation and guess field
Winter	850	0.87	530.10^3	$1.44.10^{-6}$
	700	0.69	470.10^3	$0.70.10^{-6}$
	500	0.78	370.10^3	$0.082.10^{-6}$
Summer	850	0.48	560.10^3	$2.88.10^{-6}$
	700	0.91	380.10^3	$1.93.10^{-6}$
	500	0.79	300.10^3	$0.22.10^{-6}$

The value of ρ_0 depends critically on the correlation of a few stations with a small mutual distance, and hence the accuracy of the determination of ρ_0 depends on the observation density near de Bilt. But even if the accuracy of ρ_0 is not high, it is the best estimate of ρ_0 one can get.

4. Examples of analyses

The results of the preceding sections were used to analyse the mixing ratio at 850 mbar. Here we

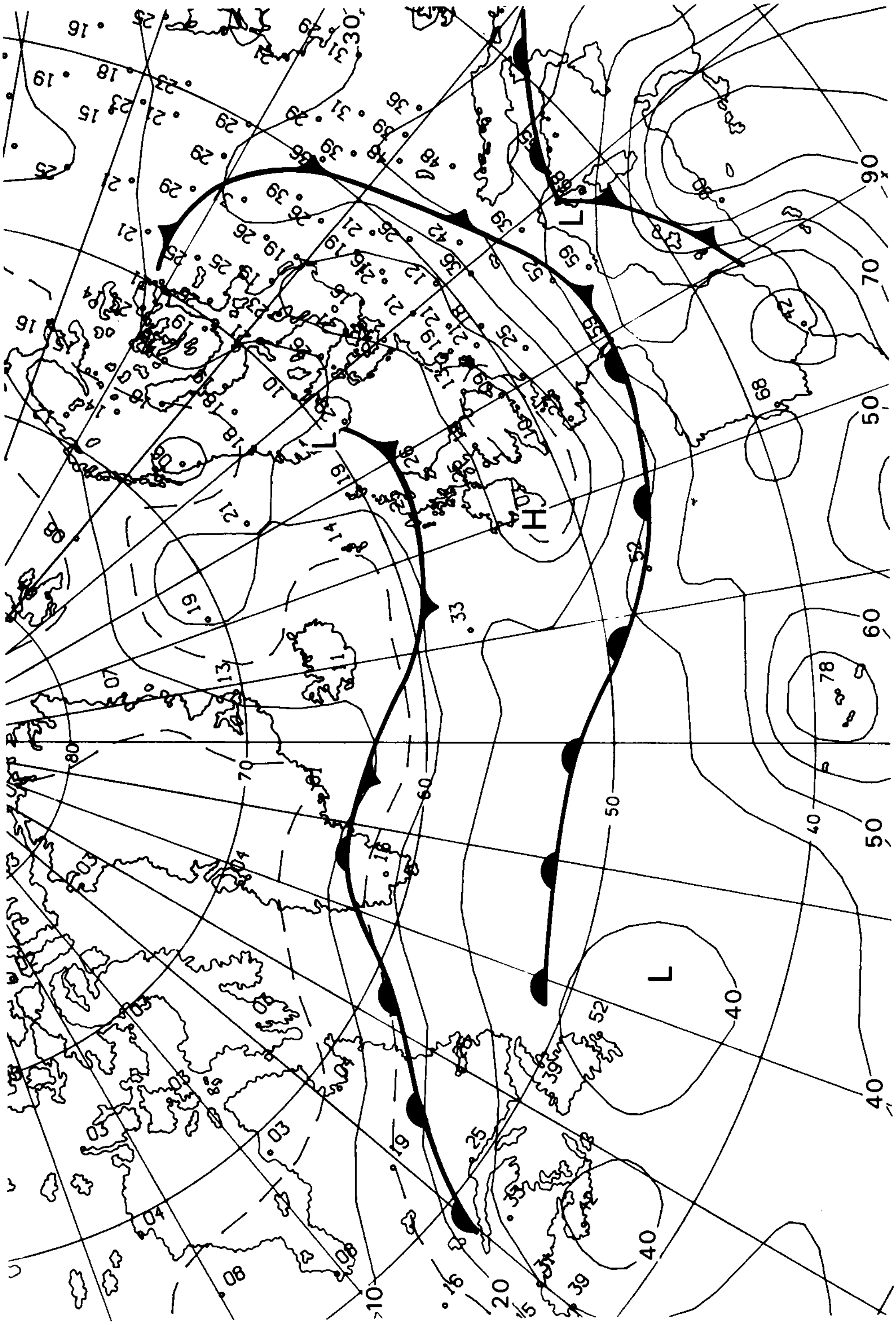

Fig. 8. Humidity analysis at 850 mbar of 1979 February 3, 00 g.m.t. Surface fronts and centres of high and low pressure are also indicated. Units are 10^{-4} kg/kg.

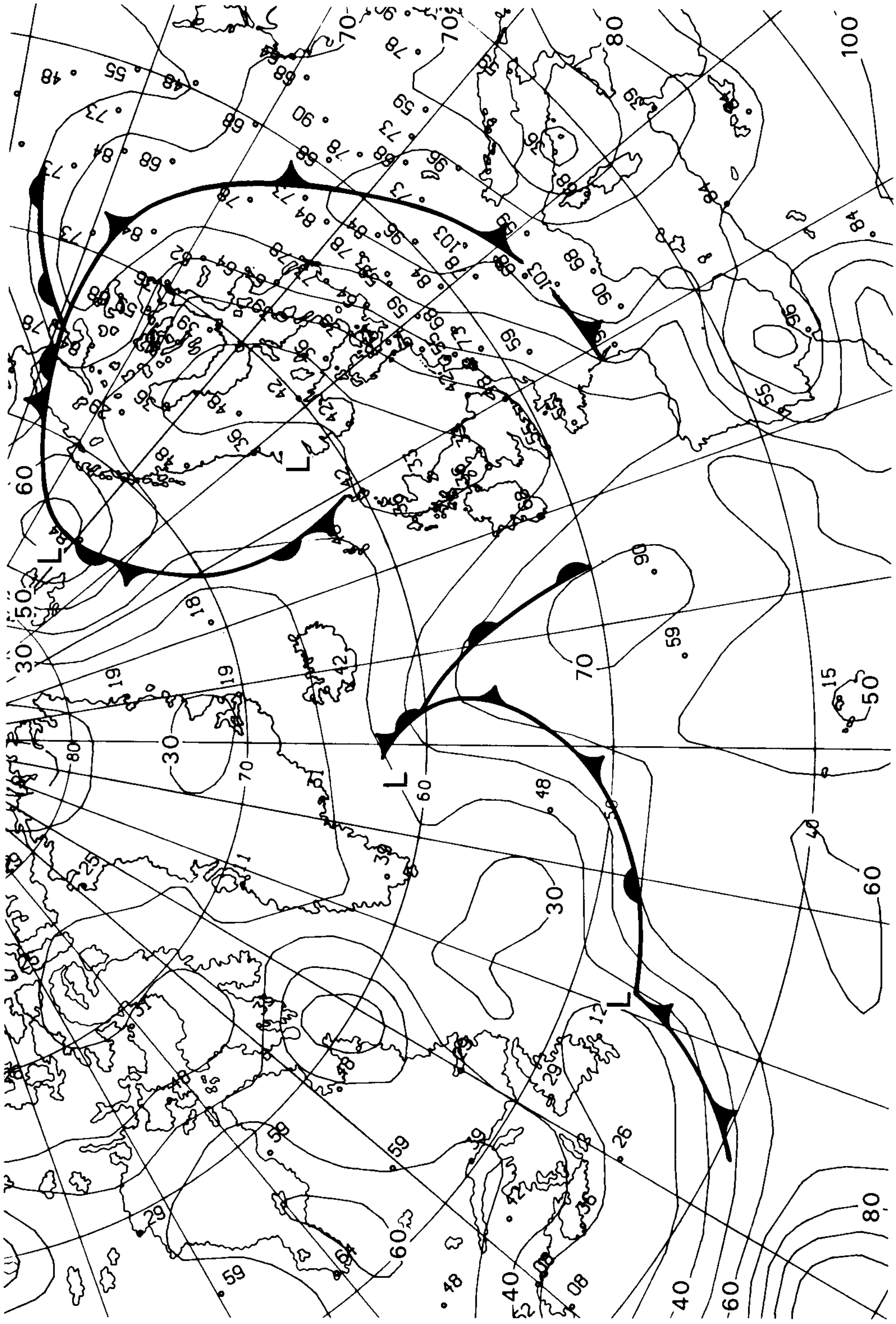

Fig. 9. As Fig. 8, for 1979 June 27, 00 g.m.t. The depression near 75° N, 20° E was very weak and disappeared within 12 h.

include two examples, one for the winter and one for the summer. The assumed correlation function of the error of the background humidity field was exp $(-bs)$, with b read from Table 2, while the ratio between the observation error and the prediction error was put equal to 0.15 and 1.08, respectively, in agreement with the data of Table 2. Note that the value of the correlation distance $1/b$ is smaller than the corresponding value for geopotential height fields (see e.g. Schlatter, 1975; Schlatter et al., 1976).

On a stereographic projection a grid was used with a grid size of 375 km. At every grid point a guess value for the mixing ratio was calculated by the method outlined in Section 3. All observations within a distance of 1500 km from a grid point were used, except when more than 16 observations were available within this range. Then the observations were sorted in order of increasing distance and the 16 nearest were used. No checks were applied to the observations.

In Fig. 8 the analysed field of February 3, 1979 is displayed, together with the observations used for the analysis. The surface fronts and pressure centres suggested by the subjective analysis of the "Deutscher Wetterdienst" are also drawn. A weak area of high pressure filled with cold and very dry air is present over Ireland. A cold front stretches from a depression over Finland to Southern France. On the warm side the air is more humid than on the cold side. The double depression near the Greenwich meridian is marked by a high absolute humidity. Over Iceland the humidity is relatively low in a cold northerly airstream.

The surface weather chart of June 27, 1979, shows a front stretching from Finland to southern France. The front is marked by a band of high humidity. In the cold air north of the front the humidity decreases with increasing distance from the front (see Fig. 9 for the humidity analysis at 850 mbar). The guess-field (not shown) indicates humid air north of Norway, and this is confirmed by the observation near 74° N, 18° E, and the analysis. The warm sector of a depression with its centre south west of Iceland is clearly marked by humid air near the 25° W meridian.

In general there seems to be a good agreement between features of the humidity field and the synoptic field. The examples shown were selected from a series of 10 days, and are representative for the quality of each one of the humidity analyses in the series. However, the synoptic features are not revealed as clearly in the observations, and hence in the analysis, in all cases as in the examples shown here.

5. Summary and conclusions

The observation error of humidity observations and the correlation function of the humidity field were determined for the winter and summer seasons from observed correlations between de Bilt and neighbouring stations, and observed variances. The observation error was determined by extrapolating the correlation function to zero distance, so it includes the variance of scales that are too small to be resolved by the observational network. It appears that due to this small-scale variation the observational error at the lower levels is much larger in summer than in winter. The correlation length does not change much from winter to summer.

The correlation function used is not differentiable at the origin. As the derivatives of the function are not involved in any way, this is not a serious drawback.

The results that are obtained are thought useful for the middle- and high-latitude regions. In principle the same methods can be used for the design of an analysis scheme for the tropics, but it is not probable that the same correlations and regressions are suitable for use in those regions. In Sections 3 and 4 the guess-field used was derived from the temperature, but one could also use a numerical forecast. The disadvantage of using a forecast is that the correlations and the prediction errors are model-dependent, hence if is necessary to recalculate the various statistics if the model is changed substantially.

Examples are given of analyses made with the optimum interpolation method. The quality of the analysis can be determined by considering the agreement between synoptic features of the flow (especially fronts) and the humidity field. Another way of verifying might be to make a forecast from the analysis, and to verify quantitatively the resulting forecasts. Forecasts of rain and cloud amounts, as well as aspects of the energy balance of the atmosphere, like latent energy release, can be verified.

The observation errors and correlation functions for absolute humidity are strongly dependent on the season. Therefore, any humidity analysis scheme should take this effect into account. In this paper the correlation functions are derived for the winter and summer seasons. For operational purposes one also needs functions for spring and autumn. These can be derived in the same way as the winter and summer functions, but a simpler solution is to interpolate in time between the winter and summer functions.

6. Acknowledgements

The author is grateful to Dr A. P. M. Baede and Mr A. Lorenc for discussions and helpful comments during the course of this study, and to Dr A. E. P. M. Abels-van Maanen for carefully reading the manuscript.

REFERENCES

Atkins, M. J. 1974. The objective analysis of relative humidity. *Tellus 26*, 663–671.

Gandin, L. S. 1963. Objective analysis of meteorological fields. Jerusalem, Israel program for Scientific Translations, 1965.

Gradshteyn, I. S. and Ryzhik, I. M. 1965. Table of integrals, series and products. New York: Academic Press.

Kästner, A. 1974. Ein Verfahren zur numerischen Analyse der Relativen Feuchte. *Arch. met. geoph. biokl., ser. A, 23*, 137–148.

Rutherford, I. D. 1972. Data assimilation by statistical interpolation of forecast error fields. *J. Atmos. Sci. 29*, 809–815.

Schlatter, T. W. 1975. Some experiments with a multivariate statistical objective analysis scheme. *Mon. Wea. Rev. 103*, 246–257.

Schlatter, T. W., Branstator, G. W. and Thiel, L. G. 1976. Testing a global multivariate statistical objective analysis scheme with observed data. *Mon. Wea. Rev. 104*, 765–783.

ОБЪЕКТИВНЫЙ АНАЛИЗ ВЛАЖНОСТИ МЕТОДОМ ОПТИМАЛЬНОЙ ИНТЕРПОЛЯЦИИ

Вычисляются различные статистические характеристики, необходимые для построения схемы объективного анализа влажности методом оптимальной интерполяции. Эти характеристики получены для зимы и лета для трех уровней 850,700 и 500 мб. Показано, что корреляционная функция влажности лучше всего описывается выражением вида $\rho_0 \exp(-bs)$, где s-расстояние между точками. Константы ρ_0 и b зависят от сезона уровня и исходного поля, (используемого при анализе). Характерное значение для b^{-1} равно 400 км. Оценка ρ_0, значения которого лежат в пределах между 0,5 и 0,9 дает представление о величине ошибки наблюдения. Полученные результаты используются в системе анализа влажности; даны примеры анализа абсолютной влажности.