KONINKLIJK NEDERLANDS METEOROLOGISCH INSTITUUT

WETENSCHAPPELIJK RAPPORT SCIENTIFIC REPORT

W.R. 78 - 17

G.J. Cats

Surface wind analysis over-land in the Netherlands, based on an optimum interpolation method



Publikatienummer: K.N.M.I. W.R. 78-17 (M.O.)

U.D.C.: 551.552 :

551.501.45

CONTENTS

- I. INTRODUCTION
- II. THEORY OF THE OPTIMUM INTERPOLATION METHOD
 - II.1 List of symbols used
 - II.2 The equations of the optimum interpolation method
 - II.3 The situation in the Netherlands
- III. MODELS OF COVARIANCE AND CLIMATOLOGY
 - III.1 Introduction
 - III.2 The correlation model
 - III.3 The variance model
 - III.4 Long-term averages of u and v
 - III.5 Use of the covariance and climatology models
- IV. VERIFICATION
- V. CONCLUSIONS

ACKNOWLEDGMENTS

REFERENCES

ABSTRACT

This paper describes an analysis scheme for over-land surface winds in the Netherlands. The analysis scheme chosen is based on an optimum interpolation method. The covariances and climatology of wind observations needed in this method are modelled, and model parameters are chosen to fit the sub-synoptic scales encountered in the Netherlands. The optimum interpolation method and the covariance and climatology models are verified simultaneously. The method gives a root-mean-square error in wind speed estimates of 1.5 m/s; the error is more or less normally distributed with maximum values of about 7 m/s. These maximum values are often associated with heavy shower systems.

BASED ON AN OPTIMUM INTERPOLATION METHOD

G.J. Cats

I. INTRODUCTION

Surface wind is a meteorological quantity with many interests. One of the main problems in describing surface winds is the variability, both in space and in time. The variability in time, which for a great deal is due to small-scale turbulence, is greatly reduced by averaging wind observations during an interval of 10 minutes, in conformity with WMO-requirements.

In the Netherlands, a program has been carried out to investigate the space variability of the wind. A fairly dense routine observation network is available (see figure 1). The average distance between the stations is about 40 km. Wieringa (1976) developed an objective method to reduce the influence of anemometer sheltering by nearby obstacles, thereby extending the representativity range of the wind observations.

This report describes a subsequent step in this research, namely an objective wind analysis, for which a version of the optimum interpolation method developed by Gandin (1963) was chosen because of its performance, which is optimal in a least-square-error sense. This method has already been applied frequently to height and wind analyses in synoptic-scale models (e.g. Schlatter, 1975), and an application to surface wind analysis has been given by Nordlund (1976), who, however, did not concentrate on the small-distance behaviour.

In the second chapter the theory of the optimum interpolation method is briefly outlined, and a special formulation applicable for the situation in the Netherlands is given. The third chapter describes the models for climatological averages of wind

observations and their covariances. Then follow the verification of the method and the conclusions.

Throughout this report the term "surface wind" denotes the wind at a height of 10 metres, with wind speeds corrected for local sheltering, according to the method developed by Wieringa (1976).

II. THEORY OF THE OPTIMUM INTERPOLATION METHOD

II.1 List of symbols used

- c covariance of observations
- d distance to the coast (i.e. the smallest distance to the coast-line)
- l correlation length
- n number of observation stations
- u wind component in W-E direction
- v wind component in S-N direction
- W weighting factor
- x,y coordinates
- γ correlation of observations
- Γ standard deviation of observations
- δ_{ij} =1 if i=j (Kronecker delta) =0 if i $\neq j$
- Δ deviations
- λ length scale: $\lambda = 20$ km
- σ root-mean-square error
- φ latitude

Indices:

subscripts:

- a in the analysis point
- i,j in the station i,j, i,j=1...n
- r,s denote the subset of values taken by i,j corresponding to stations with missing observations

superscripts:

e estimated value

g value of the guess field

o observed value

t true value

u for observations of u

v for observations of v

< > denotes a long-term time average

II.2 The equations of the optimum interpolation method

An excellent derivation of the analysis equations has been given by Lorenc et al. (1977). Some formulae are restated here.

An analysis procedure is a way to obtain estimated values Φ_a^e for some meteorological quantity Φ in a certain point (analysis point a) from observations of Φ and/or other meteorological quantities at a and/or at other observation points. It is assumed that one has a guess-field for Φ , Φ^g in all points (both in a and in the observation points i). This guess-field may be obtained from climatology, or from model calculations. (If such information is not available, one could start with a guess-field identically equal to zero). In the following an analysis scheme for the u-component of the wind field is described.

One assumes the deviation of the estimated value u_a^e from a guessed value u_a^g to be a linear combination of the observed deviations at the stations i, i=1...n:

$$u_{a}^{e} - u_{a}^{g} = \sum_{i=1}^{n} W_{ia} (u_{i}^{o} - u_{i}^{g}),$$
 (1)

where W are the weights to be determined. Thus:

$$u_{a}^{e} - u_{a}^{t} = u_{a}^{g} - u_{a}^{t} + \sum_{i=1}^{n} W_{ia} (u_{i}^{o} - u_{i}^{g})$$
 (2)

where superscripts t denote true values.

The method is based on a minimalization of the variance of the left hand side of equation (2); this variance reads:

$$(\sigma_a^e)^2 = \langle (u_a^e - u_a^t)^2 \rangle = (\sigma_a^g)^2 + 2 \sum_{i=1}^n w_{ia} \langle (u_a^g - u_a^t)(u_i^o - u_i^g) \rangle +$$

+
$$\sum_{i,j=1}^{n} w_{ia} w_{ja} < (u_{i}^{o} - u_{i}^{g})(u_{j}^{o} - u_{j}^{g}) >$$
 (3)

In the following the guess-field is chosen to be the climatological average (e.g. $u_a^g = \langle u_a^o \rangle$), being the most practicle at the moment.

Now, under the assumption that the observational error at the analysis point $u_a^0 - u_a^t$ is not correlated with $(u_i^0 - u_i^g)$, the observed deviation from the guess-field at station i, equation (3) can be transformed into

$$(\sigma_a^e)^2 = (\sigma_a^g)^2 + \sum_{i,j=1}^n w_{ia}w_{ja}c_{ij} - 2\sum_{i=1}^n w_{ia}c_{ia}$$
 (4)

where c_{ij} is the covariance of the observations at the stations i and j, and c_{ia} the covariance of the observations at the station i and the analysis point:

$$c_{ij} = \langle (u_i^o - \langle u_i^o \rangle)(u_j^o - \langle u_j^o \rangle) \rangle$$
 and $c_{ia} = \langle (u_i^o - \langle u_i^o \rangle)(u_a^o - \langle u_a^o \rangle) \rangle$.

Minimizing the right hand side of equation (4) with respect to W we find the system of n linear equations in W:

$$c_{ia} = \sum_{j=1}^{n} W_{ja}c_{ij} \qquad (i=1...n)$$
(5)

which can be solved for W_{ia} , (i=1...n), if the c_{ia} are known.

In general, at the analysis point no observations will be available. Therefore the c_{ia} have to be modelled from the measured c_{ij} . This modelling is described in the third chapter.

If the analysis point coincides with a measuring point, say j_a , then still the c_{ia} do not coincide exactly with the c_{ij_a} , due to measuring errors. For, the c_{ia} in principle could be obtained by performing observations at the point a; another anemometer at the same place j_a would not yield exactly the same wind speed and direction. Otherwise stated: the c_{ia} do not include variances due to measuring errors, whereas the c_{ij} do.

If, however, it is assumed that the measuring errors are zero, then $c_{ia} = c_{ij}$ for all i, when the analysis point a coincides with the point j_a . Equation (5) then has the trivial solution:

$$W_{j_a} = \delta_{j,j_a}$$
 ($\delta = Kronecker-delta$),

and equation (1) becomes
$$u_a^e = u_{j_a}^o$$
.

Therefore, the analysis scheme exactly reproduces the observed values in the observation points, apart from the observational error.

The analysis described above applies of course equally well to the S-N component of the wind, v. It is even possible to use the observations of v to analyze u, and vice versa. E.g., equation (2) will then read:

$$u_{a}^{e} - u_{a}^{t} = u_{a}^{g} - u_{a}^{t} + \sum_{i=1}^{n} (W_{ia}(u_{i}^{o} - u_{i}^{g}) + W_{ia}^{i}(v_{i}^{o} - v_{i}^{g}))$$
 (2')

The subsequent equations are all changed in an obvious way.

II.3 The situation in the Netherlands

In the Netherlands a fairly dense observational network exists, and on the whole very few observations are missing or unreliable. To obtain the weights W_{ia} equation (5) has to be solved, which requires inverting the n x n matrix c_{ij}. This matrix depends only on long-term averages of products of observations at the stations, and not on the analysis point or on the actual observations at

the stations. It therefore saves computation time, if the matrix c can be inverted once and for all. This indeed can be done, if provisions are made for the rare cases of missing observations.

Missing observations can be incorporated into the formulae 1 to 4, if it is imposed that they are given zero weight. We use the Lagrange undetermined multipliers technique. Thus, instead of the left hand side of equation (4),

 $(\sigma_a^e)^2 - 2\sum_{r}^m \lambda_{r} W_{ra}$ has to be minimized. In here, the summation, denoted by r, is over stations with the missing observations; m is the number of missing observations; λ_{r} is a Lagrange multiplier. This leads to:

$$\sum_{j=1}^{n} W_{ja} c_{ij} - c_{ia} - \sum_{r} \lambda_{r} \delta_{ri} = 0$$
 (6)

which indeed can be solved by inverting the matrix c_{ij} . The Lagrange multipliers λ_r follow from $w_{ra}=0$, thus

$$0 = \sum_{j=1}^{n} ((c^{-1})_{rj} [c_{ja} + \sum_{s} \lambda_{s} \delta_{js}]) =$$

$$= \sum_{j=1}^{n} (c^{-1})_{rj} c_{ja} + \sum_{s} (c^{-1})_{rs} \lambda_{s}$$
(7)

The index s also runs through the missing stations indices. Note that the first term on the right hand side of equation (7) would equal the weight W_{ia} , if no constraints were imposed (see equation (5)).

Equation (7) can be solved by inverting the matrix $(c^{-1})_{rs}$, which is a m x m submatrix of c^{-1} . The inversion of this matrix is much cheaper than the inversion of c_{ij} , when m \ll n. Therefore the procedure described in this section is attractive only in cases where many stations are present which never change their observational methods or position significantly and seldom fail to make observations.

III. MODELS OF COVARIANCE AND CLIMATOLOGY

III.1 Introduction

The covariance $c_{ij} = \langle (u_i^o - \langle u_i^o \rangle)(u_j^o - \langle u_j^o \rangle) \rangle$ is usually written as

$$c_{ij} = r_i r_j \gamma_{ij} \tag{8}$$

where Γ_i is the standard deviation of the observations at station i, and γ_{ij} is the correlation.

The Netherlands are situated on the coast. Covariances c_{ij} are expected to depend on the distance between the stations i and j and on the distances of the stations i and j to the coast. In fact, it appeared possible to incorporate the influence of the mutual distance of the stations into the correlation γ_{ij} , and the influence of the distance to the coast into the variances Γ_i^2 . Therefore the models describe the variances Γ_i^2 and the correlations γ_{ij} separately.

Climatological averages are used as the guess-field in the present version of the optimum interpolation technique. Therefore some attention is paid to these in the fourth section of this chapter.

III.2 The correlation model

In figure 2 the correlations between 1200 GMT observations $(\gamma_{ij} = <(u_i^0 - < u_i^0 >)(u_j^0 - < u_j^0 >) > /\sqrt{<(u_i^0 - < u_i^0 >)^2 > <(u_j^0 - < u_j^0 >)^2 >)})$

of the u-components at 19 stations (see figure 1) are plotted as a function of the distance r_{ij} between the stations. The number of observations was 1389. In order to detect a dependence of γ_{ij} on the distance to the coast, the stations have been divided into three classes, according to their distance to the coast. From figure 2 it is inferred that the influence of this distance is not great. Further, the correlations between the v-components did not differ much from the u-component correlations.

Two equations were tried to fit the observations of γ_{ij} :

a)
$$\Upsilon_{ij} = \Upsilon_{o} \exp(-r_{ij}/1)$$
b)
$$\Upsilon_{ij} = \Upsilon'_{o} \exp(-(r_{ij}/1')^{2})$$
(9)

Both equations were fitted by a linear regression between $\ln \gamma_{ij}$ and r resp. r^2 . Both regressions explained 80 % of the variance of $\ln \gamma_{ij}$. The regression constants are:

a)
$$\gamma_0 = 0.955$$
 and $l = 1150 \text{ km}$
b) $\gamma_0' = 0.911$ and $l' = 520 \text{ km}$ (10)

In synoptic-scale problems an equation of the form (9b) is used, because γ_{ij} is derived from heights correlations by means of a geostrophic relation. Therefore γ_{ij} is differentiable in the origin. "Correlation distances" 1, then, appear to be of the order of 1280 km (Lorenc et al., 1977).

A correlation distance 1, which is of the same order of magnitude, is found here for the equation (9a). Therefore both equations have a feature in common with the corresponding large-scale equation. The first equation was preferred, because the analysis-results turn out to be slightly better (see chapter IV).

As surface wind is highly correlated with synoptic-scale weather systems, it is not surprising to find a large correlation distance. Yet, this does not imply that subsynoptic-scale phenomena are smoothed out by the analysis method: if they are significantly present in the observations, they will be detected, conformable to the statements following equation (5).

The correlation γ_{ij} is not equal to unity for zero distances due to measuring errors. From the fact that γ_0 =0.955 it follows that the root-mean-square measuring error amounts to 20 % of the standard deviation of the observations from the long-term average Γ_i .

In this research the correlations between u-components as well as between v-components were modelled as $\gamma_0 \exp(-r_{ij}/1)$, with l=1150 km and γ_0 =0.955. Thus the correlations are isotropic.

The correlation between u_i and v_j was observed to be small, therefore it was assumed to be zero. With this assumption, the w_{ia} in equation (2') are also zero.

III.3 The variance model

The variance of the wind observations at a station is among other things determined by the position of the station (e.g. Oort and Rasmusson, 1971, give a dependence on latitude). Usually winds are stronger near the coast, and consequently the variance of uand v-components will be higher in coastal regions than inland. It is felt that the influence of the coast disappears almost entirely at distances to the coast greater than about 50 km. For this reason we chose the influence of the distance to the coast to be modelled as $tgh(d/\lambda)$, where $tgh(x) = \frac{exp(x) - exp(-x)}{exp(x) + exp(-x)}$, d is the distance to the nearest coastline, and $\lambda = 20$ km.

The variance of the u-component was linearly fitted to the latitude and $tgh(d/\lambda)$. The best fit was:

$$(\Gamma_{i}^{u})^{2} = 0.62 \, y_{i}/\lambda - 7.1 \, tgh(d_{i}/\lambda) + 24.7$$
 (11)

(explained variance 64 %). $(\Gamma_i^u)^2$ is measured in $(m/s)^2$. The coördinate system (x,y) is chosen such that the origin is at $51^{\circ}58^{\circ}N$, $4^{\circ}56^{\circ}E$, in the centre of the Netherlands. The x-axis points east and the y-axis north.

The variance of the v-component $\mathbf{r_i^{v^2}}$ was approximated by 0.86 $\mathbf{r_i^{u^2}}$.

According to the preceding section the observational error amounts to 0.2 Γ_{i}^{u} . From equation (11) it follows that Γ_{i}^{u} and Γ_{i}^{v} are of the order of 5 m/s; thus the observations have a standard error of about 1 m/s in each component. In table 1 Γ_{i}^{u} , Γ_{i}^{v} and their latitudinal derivatives are compared with synoptic-scale observations.

Table 1 Order-of-magnitude comparison of variances and long-term averages from this research and from the tables of Oort and Rasmusson (1971).

Quantity	Order of magnitude			
	Oort and Rasmusson (52°N)	This research		
$\mathbf{r_i^u}^2 (\mathbf{m/s})^2$	33	24		
$\mathbf{r_i^{v^2}} (\mathbf{m/s})^2$	32	20		
$\frac{1}{r_i^{u^2}} \frac{\partial r_i^{u^2}}{\partial \varphi} (rad^{-1})$	5	8 .		
$\frac{1}{r_i^{v^2}} \frac{\partial r^{v^2}}{\partial \varphi} (rad^{-1})$	3	7		
<u<sub>i> (m/s) < v_i> (m/s)</u<sub>	1.5	1.5		
$\langle v_i \rangle$ (m/s)	0.4	0.7		
$\frac{1}{\langle u_i \rangle} \frac{\partial \langle u_i \rangle}{\partial \varphi} (rad^{-1})$	-4	-4		
$\frac{1}{\langle v_i \rangle} \frac{\partial \langle v_i \rangle}{\partial \phi} $ (rad ⁻¹)	-6	-1 8		

III.4 Long-term averages of u and v

In this section the models for the climatological values for u and v are given.

It appeared that $< u_i>$ and $< v_i>$ in the Netherlands average 1.49 m/s and 0.65 m/s respectively. The 19 individual station values of $< u_i>$ and $< v_i>$ had standard deviations of 0.20 m/s. Instead of using constant values for $< u_i>$ and $< v_i>$, one could choose to use a somewhat more detailed model:

$$\langle u_i \rangle$$
 = 0.03 x_i/λ - 0.02 y_i/λ - 0.49 $tgh(d_i/\lambda)$ + 1.79 (explained variance 45 %)

$$\langle v_i \rangle$$
 = 0.07 x_i/λ - 0.04 y_i/λ - 0.17 $tgh(d_i/\lambda)$ + 0.75 (explained variance 61 %)

 $(u_i,v_i \text{ in m/s})$, which fitted the observed $< u_i>$ and $< v_i>$ slightly better. Therefore this model was used in the present analysis.

III.5 Use of the covariance and climatology models

The models that have been given for c_{ij} and $< v_i^o>$ and $< v_i^o>$ have to be applied in the optimum interpolation method only, in order to obtain an estimate for these quantities in the analysis point. However, usually the models are also used for the description of the c_{ij} and $< v_i^o>$ and $< v_i^o>$ in the observation points themselves; it is hoped that this procedure eliminates errors in the c_{ij} that arise from local influences, from finiteness of data series, etc. It has the further advantage that newly established stations can immediately be incorporated in the analysis scheme. In the present case, the same procedure has been followed.

In order to apply the analysis scheme outside the Netherlands, or improve the analysis in border areas, it is necessary to use observations at sea and in neighbouring countries. Therefore the correlation, variance and climatology models have to be extended to regions outside the Netherlands.

IV. VERIFICATION

The analysis system as described in Chapter II, using the models of Chapter III, has been verified in the following way:

For each day of the four years 1971-1974, at 1200 GMT and at 0000 GMT, 19 analyses were performed, in which each of the 19 stations in figure 1 was treated in turn as the analysis point. The analyses were based on the observations at the other 18 stations. The basic analysis equation (1) has been applied for both u- and v-components. The weights w have been determined from equation (6), with the covariances c_{ij} and c_{ij} given by equations (8), (9a), (10a) and (11). The Lagrange multiplier λ_r followed from equation (7). In order to assess the sensitivity for the particular correlation model used, also equations (9b) and (10b) were applied to 0000 GMT. From the estimated u- and v-components wind speed and direction estimates were obtained in the analysis point. These were compared with the observations in the analysis point. This procedure was repeated with a selection of 12 stations out of these 19 (with analyses based on the observations at 11 stations). Estimation errors generally turned out to be normally distributed, with standard deviations as summarized in table 2.

The values in table 2 have to be compared with the observation error of 1 m/s in each wind component which corresponds to a standard observation error of about 1 m/s in the wind speed, and of about 15° in the wind direction. Maximum errors in wind speed estimates can be as high as 7 m/s. A short investigation showed that these high values are associated with local heavy (thunder)storms.

The analysis scheme is not sensitive to the correlation equation used. The equations (9a) and (10a) appear to be only slightly better.

The analysis scheme took 0.05 seconds of computation time for each analysis point, on the average 0.2 seconds for each inversion of the matrix $(c^{-1})_{ij}$ and 1 second for the inversion of the 19 x 19 matrix c_{ij} on a Burroughs B6700 computer.

Table 2 Results of the verification of the analysis system.

		19 stations			12 stations	
		1200 GMT	0000	GMT	1200 GMT	0000 GMT
			equations (7a),(8a)	equations (7b),(8b)		
Δ ff	(m/s)	1.4	1.4	1.5	1.5	1.4
∆ dd	(°)	19	19	21	24	21
$\frac{\overline{\Delta_{ \vec{u} }}}{ \vec{u} }$	(m/s)	1.8	1.6	1.7	2.0	1.7
maxff	(m/s)	6.8	6.0	6.5	6.9	6.9

Notations:

$\overline{\Delta_{ff}}$	root-mean-square error in wind speed estimates.
Δ _{dd} -	root-mean-square error in wind direction estimates.
$\Delta \overrightarrow{u} $	average speed of the vector difference between measured and estimated values.
max _{ff}	maximum error in wind speed estimates.

The figures in the table are averages over 19 resp. 12 tests,

in which one station was used as an analysis point; only the observations at the other stations were used to form the analysis.

V. CONCLUSIONS

From the results in Chapter IV it follows that a modified optimum interpolation method, developed by Gandin (1963), constitutes an analysis scheme which gives good results on the average, especially for wind speed. Occasionally, however, the deviation between measured and estimated wind speeds can be as high as 7 m/s due to heavy rainshower conditions. Surface wind direction can be estimated only within a standard error of 20°.

The covariances needed in the optimum interpolation method are given by

with Γ_i the square root of the variance at station i, depending on latitude and distance to the coast of station i, and γ_{ij} the correlation between the observations at station i and station j, depending on the distance between these stations.

We chose to describe $\gamma_{i,j}$ as

$$\gamma_{ij} = \gamma_0 \exp(-r/1),$$

thus isotropic, and the same for u- and v-components of the wind. By linearly fitting $\ln \gamma_{ij}$ to r it is found that $\gamma_{o} = 0.955$ and l = 1150 km. This correlation distance l has the same order of magnitude as the distances in large-scale models.

From an extrapolation of the correlations to zero distances between stations it was found that the observational error in surface wind measurements has a root-mean-square value of 1 m/s in each component after application of the shelter-correction for wind speeds. (Wieringa, 1976).

The verification calculations with 12 stations do not show much difference with the calculations with 19 stations. It is concluded that the number of stations can be reduced without losing much information on the average; this is connected with the fact that the average distance between the stations (some 40 km) is much smaller than the correlation distance 1 (1200 km).

Although the correlation distance is determined by synopticscale phenomena, the analysis scheme detects subsynoptic-scale phenomena if they are significantly present. Reduction of the number of stations leads to a less significant determination of such phenomena, and may therefore influence the analysis occasionally.

ACKNOWLEDGMENTS

I would like to thank Mr. J. van Maanen for many helpful and stimulating discussions, and Mr. S. Kruizinga for some interesting remarks.

REFERENCES

- Gandin, L.S.: 1963, Objective Analysis of meteorological fields.

 Translated from Russian by the Israel programme for scientific translations, 1965.
- Lorenc, A., I. Rutherford and G. Larsen: 1977, The ECMWF Analysis and Data Assimilation Scheme: Analysis of mass and wind fields. European Centre for Medium Range Weather Forecasts Technical Report No. 6.
- Nordlund, G.G.: 1976, On the Use of Objective Wind Analysis for

 Mesoscale Dispersion Calculations. Proceedings

 of the Third Symposium on Atmospheric Turbulence,

 Diffusion and Air Quality, October 1976, Raleigh

 (N.C.), American Meteorological Society.
- Oort, A.H., and E.M. Rasmusson: 1971, Atmospheric Calculation Statistics. NOAA Professional Paper No. 5.
- Schlatter, T.W.: 1975, Some experiments with a multivariate Statistical Objective Analysis Scheme. Monthly Weather Review 103, 246-256.
- Wieringa, J.: 1976, An Objective Exposure Correction Method for Average Wind Speed Measured at a Sheltered Location. Quart. J. Roy. Meteor. Soc. 102, 241-253.

FIGURE CAPTIONS

- Figure 1 The location of the 19 wind stations. The division into three classes according to the distance to the coast is indicated by the heavy dashed lines. The 12 selected stations (see Chapter IV) are denoted by crosses.
- Figure 2 Correlation coefficients of shelter-corrected 10-metre u-component observations as a function of the mutual distance of the stations, for the combinations of station/coast-distance classes (see figure 1). Full lines denote linear best fits, for the left graph to the correlation coefficients of stations from the class combinations I-I, II-II and III-III, and for the right graph for the other combinations. The broken line is the simultaneous best fit to both graphs.



