

## **The Initial Field Constructions for Numerical Weather Prediction in the Low Latitudinal Regions by a New Mixed Type Autocorrelation Function**

KUANG-HORNG WANG<sup>1</sup>, SHING-CHUNG ONN<sup>1</sup>, CHUNG-YI TSENG<sup>2</sup> and TAI-HWA HOR<sup>1</sup>

(Manuscript received 27 July 1995, in final form 21 December 1995)

### **ABSTRACT**

A univariate statistical scheme for the analysis of geopotential height fields has been developed based upon optimum interpolation. The data archived from a set of subjective analyses were used to compute the required correlation functions. The computed height autocorrelation functions were fitted by an appropriate function which would definitely reflect the characteristics of the statistical structure in the low latitudinal regions. Rather than representing the model by the negative squared exponential form proposed by Gandin (1963), a representation consisting of a polynomial exponential function was used to fully trace the finer structure of the statistical quantities. The analyzed geopotential height fields were obtained through the computation of the geopotential height autocorrelation model. As a measurement of accuracy, this scheme has been tested at target stations. The root mean squares between the analysed data and those obtained from radiosonde data have been carried out to verify its performance as well.

(Key words: Numerical weather prediction, Correlation function, Objective analysis, Optimum interpolation)

### **1. INTRODUCTION**

This paper describes the tests of a univariate statistical objective scheme intended for use with numerical weather prediction models in the low latitudinal regions. In the past few decades, many scientists have shown considerable interest in the research of the initial field constructions for meteorological variables. The initial stage of optimum interpolation was first carried out using univariate variables, such as the geopotential height, the wind component, the temperature and the dew point temperature to compute their statistical quantities. Later, Gandin (1963) proposed a multivariate statistical analysis scheme for meteorological

---

<sup>1</sup> Department of Applied Physics, Chung Cheng Institute of Technology, Ta-Hsi, Tao-Yuan, Taiwan, R.O.C.

<sup>2</sup> Institute of Physics, Academia Sinica, Nankang, Taipei, Taiwan, R.O.C.

applications. Difficulties he met were due to the lack of information on the cross-correlation elements that had not been determined at that time. Meanwhile, however, Buell (1958, 1960, 1972a), Gandin (1964) Petersen and Middleton (1964) and Petersen (1973c) found out that correlations between the height and wind components are related to geostrophic approximations at mid- and high latitudes. Since then, Kluge (1970), Rutherford (1973) and Schlatter (1975, 1976) have used the multivariate objective analysis scheme in weather prediction models. Additionally and almost at the same time, Eddy (1973) and Petersen (1973c) proposed a multivariate statistical scheme such that, in the vertical, the final analyzed geopotential height fields are subject to thermal wind relations.

In this paper, the univariate optimum interpolation in the low latitudinal regions and its application in initial field constructions are the primary concern. More specifically, this paper attempts to construct a 2D autocorrelation model which is suitable for the analysis of geopotential height fields in the present research area. By using the manual subjective analysis data sets which are 500 hPa, 1200 GMT during the winter months (December through February) of 1977 to 1983, it is hoped that the scheme will demonstrate itself as being capable of practical implementation and of use in operational initial height field constructions for numerical weather prediction.

The remainder of this paper is organized as follows: Section 2 of this paper presents the definitions of various correlation functions, the construction of the h-h autocorrelation matrix as well as the modeling of the autocorrelation functions in terms of distance. Section 3 lays out the mathematical treatment of these algorithms in optimum interpolation. Section 4 demonstrates the ability of the analysis scheme to fit the observation data, the response of the scheme to various inputs and modifications and characteristics of the analysis scheme are also mentioned. Finally, in Section 5, the findings are summarized.

## 2. ANALYSES

The procedure followed here is specified to the problem at hand, namely, the statistical structure constructions of the geopotential height fields and its application in the initial height field analysis for numerical weather prediction models. Specifically, the interest in this paper is in the optimum interpolation of scattered station data interpolate to regular grids simultaneously. As these data are univariate in themselves, they may include a spatial and temporal set of meteorological variables such as the geopotential height, the scalar wind component, the temperature and the dew point temperature as well.

### 2.1 Correlation Function

In the method of optimum interpolation, the statistical structure of the meteorological fields is the main role in the analysis process. Autocorrelation and cross-correlation functions are usually selected to represent statistical quantities. The important properties are briefly mentioned below:  $f(r)$  can be taken as the meteorological variable; it is defined as the sum of the mean  $\bar{f}(r)$  and the deviation  $f'(r)$ , where  $r$  is the geographic point with respect to a reference point. The cross-covariance function  $C_{fg}$  for variables  $f$  and  $g$  is given by:

$$C_{fg}(r_1, r_2) = \overline{f'(r_1)g'(r_2)}, \quad (1)$$

where  $r_1$  and  $r_2$  are two different geographic points. Furthermore, the autocovariance function  $C_{ff}(r_1, r_2)$  of the univariate  $f$  can also be defined from Equation (1).

The mean square of the anomaly of the quantity  $f$  on the point  $r_1$  is written as:

$$C_{ff}(r_1, r_1) = \overline{f'^2}(r_1) = D_f(r_1) \quad (2)$$

The square root of  $D_f$  is sometimes called the variance of  $f$  at that point.

The cross-correlation function of variables  $f$  and  $g$  on two distinct points  $r_1$  and  $r_2$  is defined as:

$$\rho_{fg}(r_1, r_2) = C_{fg}(r_1, r_2) / \sqrt{D_f(r_1)D_g(r_2)} \quad (3).$$

The autocorrelation function of the same variable is defined when  $f = g$ , which is in the form of:

$$\rho_{ff}(r_1, r_2) = C_{ff}(r_1, r_2) / \sqrt{(C_{ff}(r_1, r_1)C_{ff}(r_2, r_2))} \quad (4).$$

When the covariance or correlation function depends not upon the coordinate of the points but rather upon the distance alone, it can be said that the function satisfies the homogeneous and isotropic conditions.

In order to check to what extent the homogeneity and isotropy hypothesis is to be satisfied, it is convenient to proceed as follows: fixing one of the points, the correlation functions for all possible positions of the second point are calculated. If the isolines of such autocorrelation functions are circles, it means that the hypothesis of homogeneity and isotropy are met.

As Gandin suggested in his paper, the station distribution can be selected arbitrarily. In order to find the patterns of correlation functions (among  $h$ ,  $u$  and  $v$  variables) between Taipei and all the other 77 stations, some of the stations must be put on the right hand side of Taiwan Island which is bordered by the vast Pacific Ocean. The data on these prespecified points were archived from the historical Japanese weather map and have been analysed by many experienced meteorologists there. During the readout of these prespecified station data, the task was mainly done by careful subjective analyses. In Figure 1, the stations used in these experiments are numbered in order of increasing distance from Taipei, with the city of Taipei numbered as 58 and lying almost in the center of the station network. In this experiment, the correlations were calculated from a set of 108 cases on each point. Each observations were separated by a five-day interval, thereby making this kind of pick up independent to ensure isolation among cases.

The correlation patterns resulting from Equations (3) and (4) for the computation of the autocorrelations and cross-correlations among the three variables  $h$ ,  $u$  and  $v$  are shown in Figure 2 (refer to Onn, Wang, Tseng 1993). It is obvious that the  $\langle(h-h)\rangle$  correlation satisfies the conditions of homogeneity and isotropy, while  $\langle(h-u)\rangle$ ,  $\langle(h-v)\rangle$ ,  $\langle(u-h)\rangle$ ,  $\langle(v-h)\rangle$ ,  $\langle(u-v)\rangle$ ,  $\langle(v-u)\rangle$  are neither isotropic nor homogeneous.

For the large scale meteorological phenomena, the  $\langle(h-h)\rangle$  autocorrelation is indeed true in satisfying the homogeneous and isotropic conditions for the order of some thousands of kilometers on some isobaric levels. Studies investigated by others have shown that for the geopotential height, temperature, humidity, etc., the assumption is fulfilled with sufficient accuracy up to a distance of several thousands of kilometers. In addition, this assumption substantially simplifies the problems at hand here. The autocorrelation functions of the wind

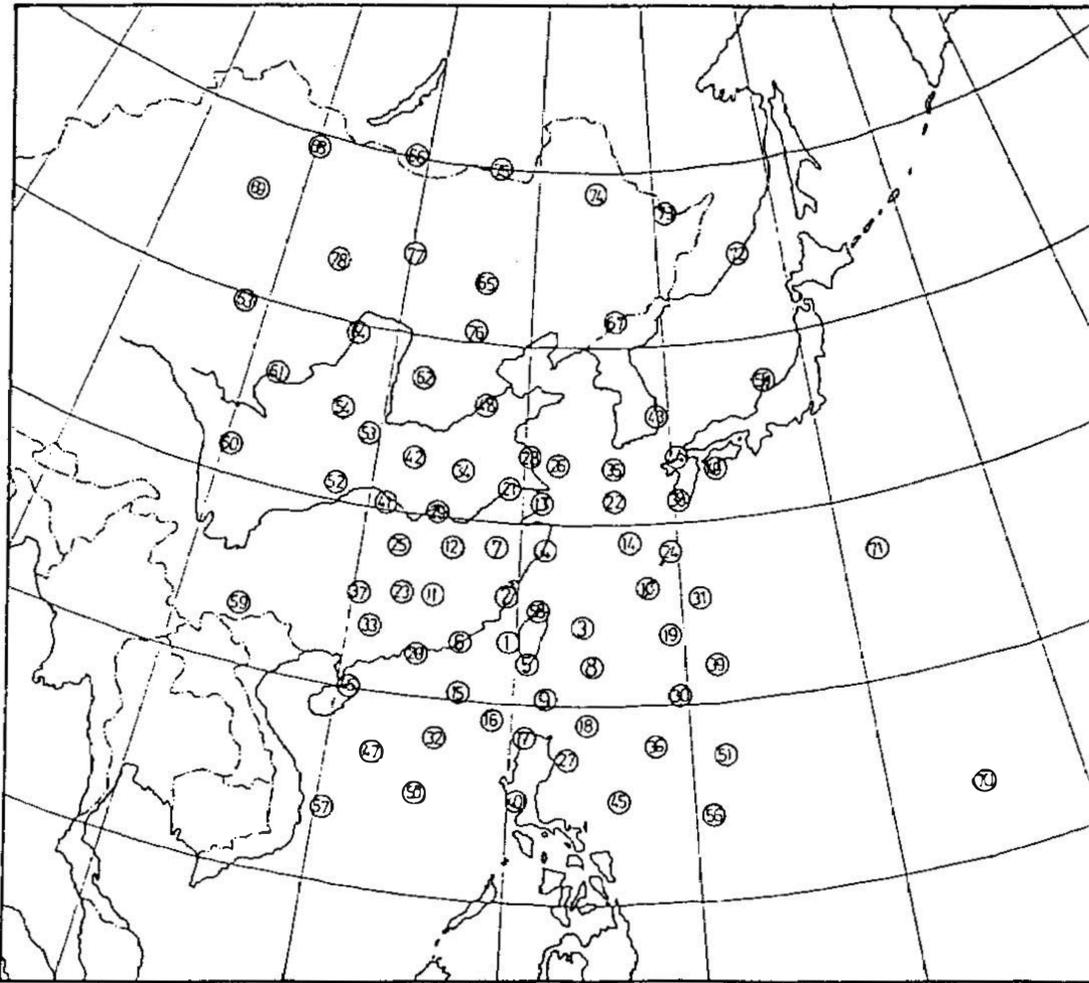


Fig. 1. Network of stations used in the analysis. The stations are numbered in order of increasing distance from Taipei.

components  $\langle(u-u)\rangle$  and  $\langle(v-v)\rangle$  are a little bit deformed from true circle, and the patterns with the major axis were shown as being oriented toward the zonal direction in the case of the  $u$  component but toward the meridional direction for the  $v$  component. Such differences do not strictly satisfied the conditions of homogeneity and isotropy.

The test of the analysis scheme was based upon the autocorrelation model computed from the geopotential height fields in the geographic domain. In this section, the autocorrelation computation of this univariate variable, the modeling of these correlations and its applications in constructing initial geopotential height fields are discussed.

Autocorrelations of the geopotential height field were computed for  $\langle(h_i-h_{fi})(h_j-h_{fj})\rangle$  where  $i$  and  $j$  range over 78 stations in the geographic domain.

The bracket symbol in  $\langle(h'_i-h'_j)\rangle$  denotes the statistical averaging over many cases. The definition gives:

$$\langle(h_i - h_{fi})(h_j - h_{fj})\rangle = \frac{(\sum_{t=1}^{NT} (z_{it} - \bar{z}_i) * (z_{jt} - \bar{z}_j))}{\left(\sqrt{\sum_{t=1}^{NT} (z_{it} - \bar{z}_i)^2} * \sqrt{\sum_{t=1}^{NT} (z_{jt} - \bar{z}_j)^2}\right)}.$$

The ensemble mean is given by:

$$\bar{z}_i = \sum_{t=1}^{NT} z_{it} / NT.$$

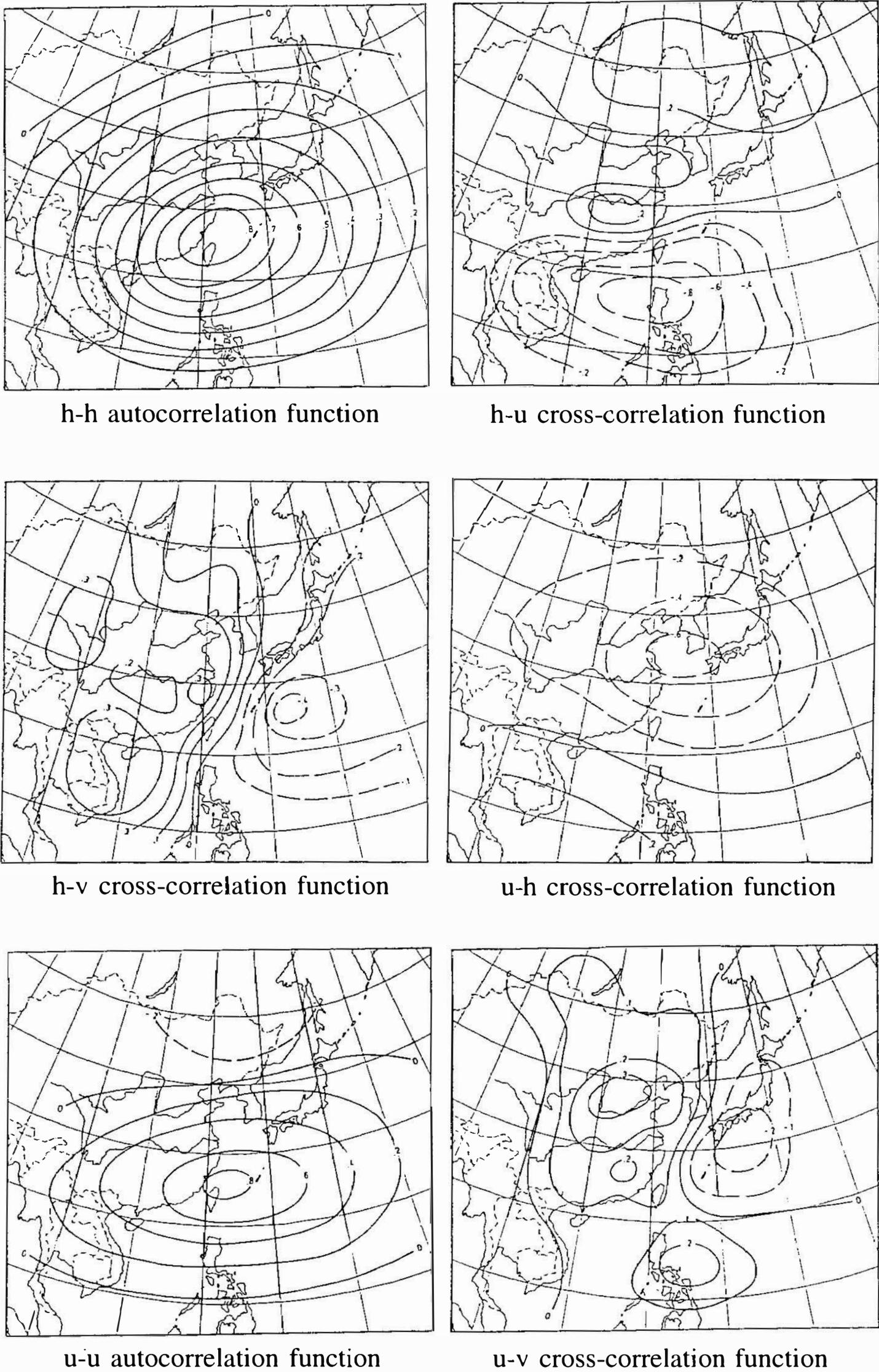


Fig. 2. Correlations among the variables h, u and v. The master station is Taipei.

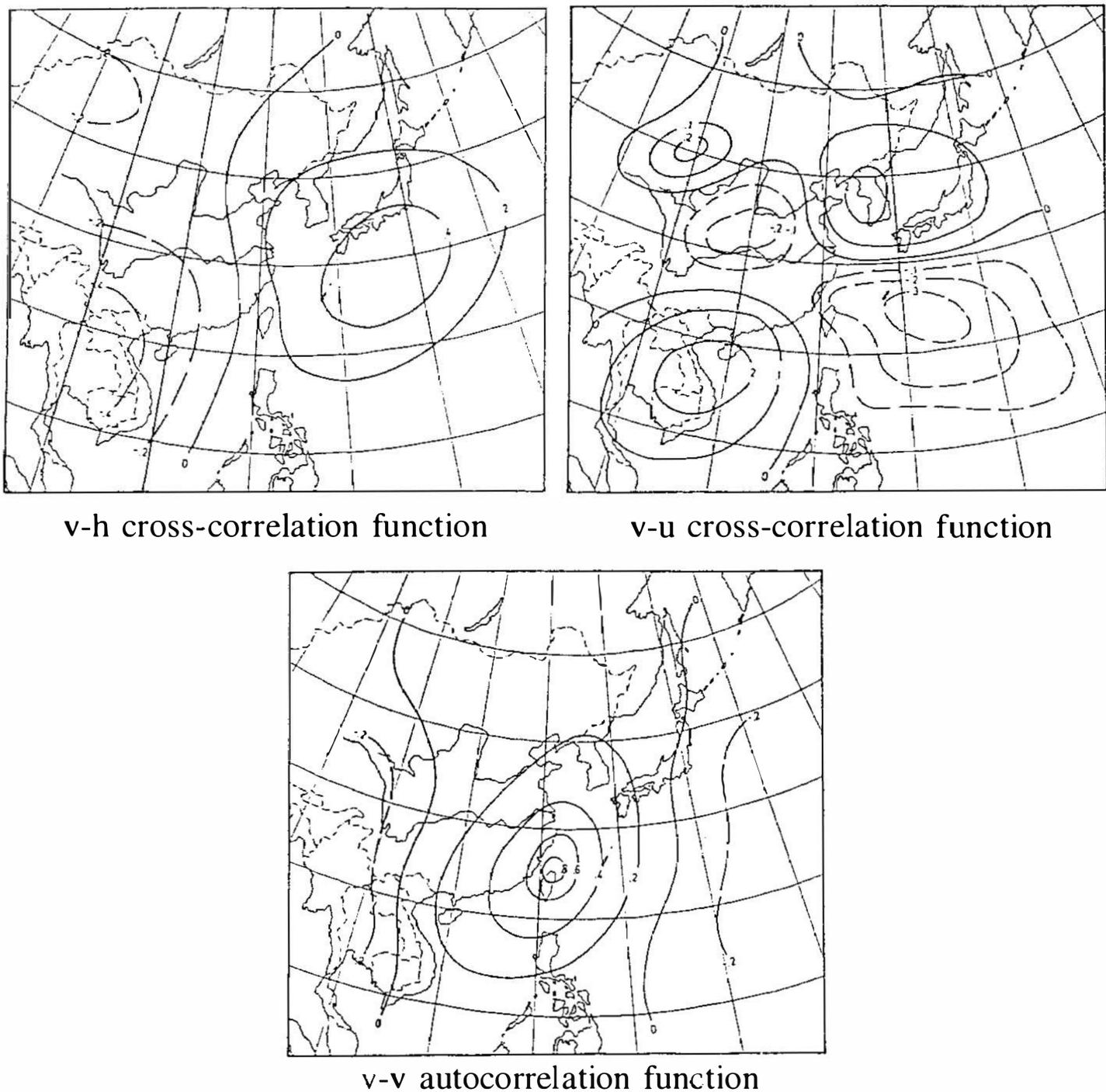


Fig. 2. (Continued)

In this way, the ensemble mean of the long term data (108 cases involved) in the present study were selected as the initial guess  $h_{fi}$ . The autocorrelation functions for every possible pair points are plotted as a function of separation distance. Because this is a symmetric matrix, 78 stations or  $(78 \times 77) / 2 + 78$  possible station pairs were considered. The resultant 3081 autocorrelations for the station pairs are presented in Figure 3, (refer to Onn, Wang and Tseng 1993). The figure shows that the qualitative agreement is the same as others obtained from the mid- and high latitude regions of the earth; however the lines of equal correlation are slightly different. The autocorrelation functions of the geopotential heights generally decrease more slowly with an increasing separation distances in the farther distance than in the near distance.

It is important to consider these autocorrelations as a two-point function. The full correlation arrays were replaced by arrays of correlations averaged over intervals in the separation distance before being fit to the autocorrelation model. In this way, the computer time for the fitting process can be reduced to an acceptable level, while much of the heterogeneity within these data can be eliminated. In calculating the distance averaged data for these autocorrelation functions, it is reasonable to use the classification of the results based on the

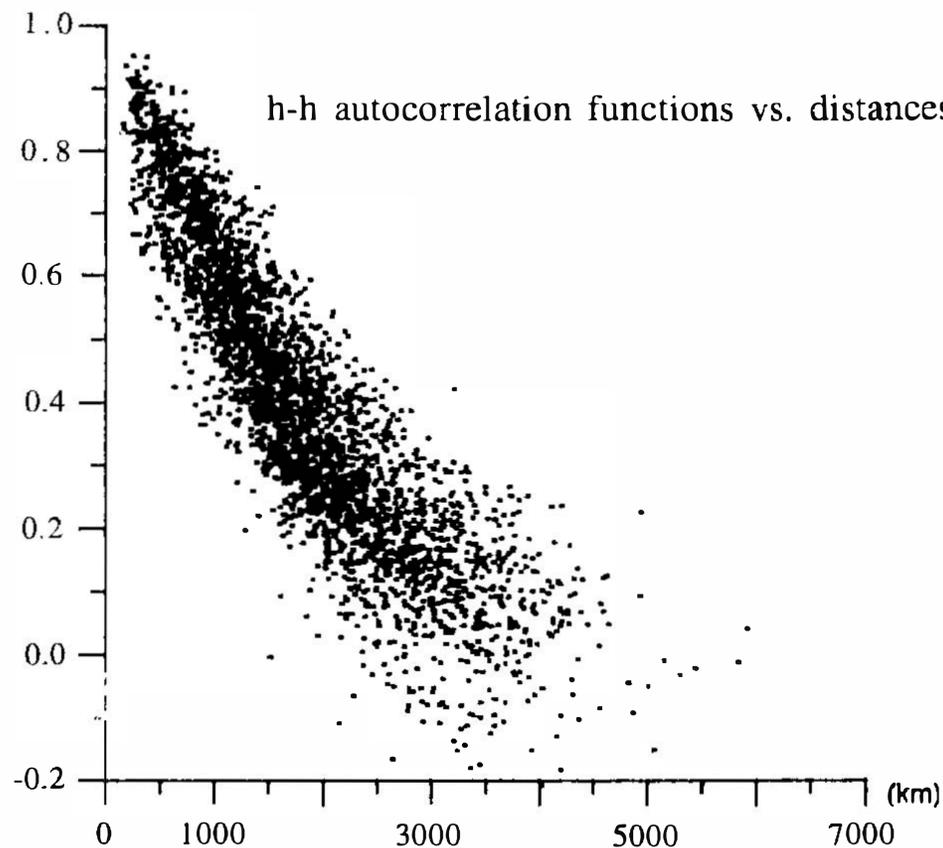


Fig. 3. Autocorrelations for height fields as a function of separation distance (Km) among the station pairs for a total of 78 stations. (The climatological mean is chosen as the first guess.)

gradation of the distance. Usually the gradation of about 200-250 km was taken to be the unit interval. Points corresponding to negative correlations are not appropriate for the curve fitting of the exponential function, and therefore were not considered in the least-squared curve fitting process.

In the past few decades, the Gaussian function  $\rho(r) = a \exp(-br^2)$  has been adopted to fit the height-height autocorrelation data in the least squared sense. Improving the specifications of the autocorrelation model has been approached both observationally and theoretically, and extensive literature now exists on the subject. A number of functional forms have also been proposed for the continuous two-dimensional homogeneous isotropic correlation function.

## 2.2 The Least-squared Method for Construction of the Height Autocorrelation Model

In this paper, the new model that is proposed to analyse the distance averaged autocorrelation data is a function combining the polynomial and exponential form.

$$\rho(r) = A(r)e^{B(r)} = \left( \sum_{i=0}^{ma} a_i r^i \right) \exp \left( \sum_{i=0}^{mb} b_i r^i \right). \quad (5)$$

where  $ma$  and  $mb$  are two arbitrary positive integers. The resulting sum of the squared errors between the correlation model and the correlation data is given by:

$$\sum_{j=1}^N \varepsilon_j^2 = \sum_{j=1}^N \ln \rho_j - \left[ \left( \sum_{i=0}^{mb} b_i d_j^i \right) + \left( \sum_{i=0}^{ma} a_i d_j^i \right) \right]^2, \quad (6)$$

where  $(d_j, \rho_j)$  represent the characteristic distance and the associated distance averaged autocorrelation data respectively. By differentiating Equation (6) with respect to  $a_k$  and  $b_k$ ,

$$F_k = \partial \left( \sum_{j=1}^N \varepsilon_j^2 \right) / \partial a_k = \sum_{j=1}^N \left( \frac{d_j^k}{\sum_{i=0}^{ma} a_i d_j^i} \varepsilon_j \right), \quad (7)$$

and

$$G_k = \partial \left( \sum_{j=1}^N \varepsilon_j^2 \right) / \partial b_k = \sum_{j=1}^N d_j \varepsilon_j. \quad (8)$$

From Equations (7) and (8), the  $(ma + mb + 2)$  nonlinear algebraic equations can be solved simultaneously using the Newton-Raphson method by choosing the arbitrary initial value for  $a_k$  and  $b_k$ . The final solution is considered to be consistent when both of the following conditions are met:

$$\text{Max} \left[ a_k^{\text{iter}+1} - a_k^{\text{iter}} \right] < 10^{-5} \quad \text{where } k = 0, \text{ ma}$$

$$\text{Max} \left[ b_k^{\text{iter}+1} - b_k^{\text{iter}} \right] < 10^{-5} \quad \text{where } k = 0, \text{ mb.}$$

The convergence processes of the maximum errors for  $a_k$  and  $b_k$  are indicated in Figure 4. The curve fitting of the distance averaged data for the autocorrelation model is shown in Figure 5. The respective coefficients for  $a_k$  and  $b_k$  of order 5 are also shown in Table 1.

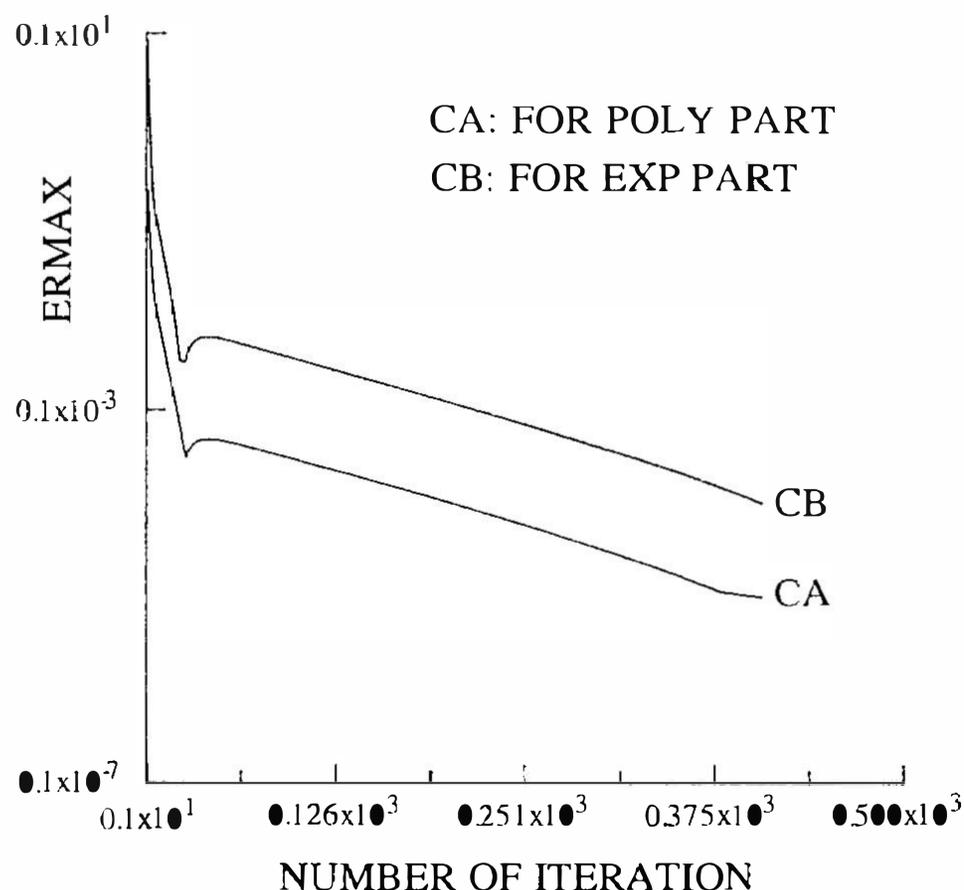


Fig. 4. The convergence process of maximum error for coefficients, where  $k=1,5$ .

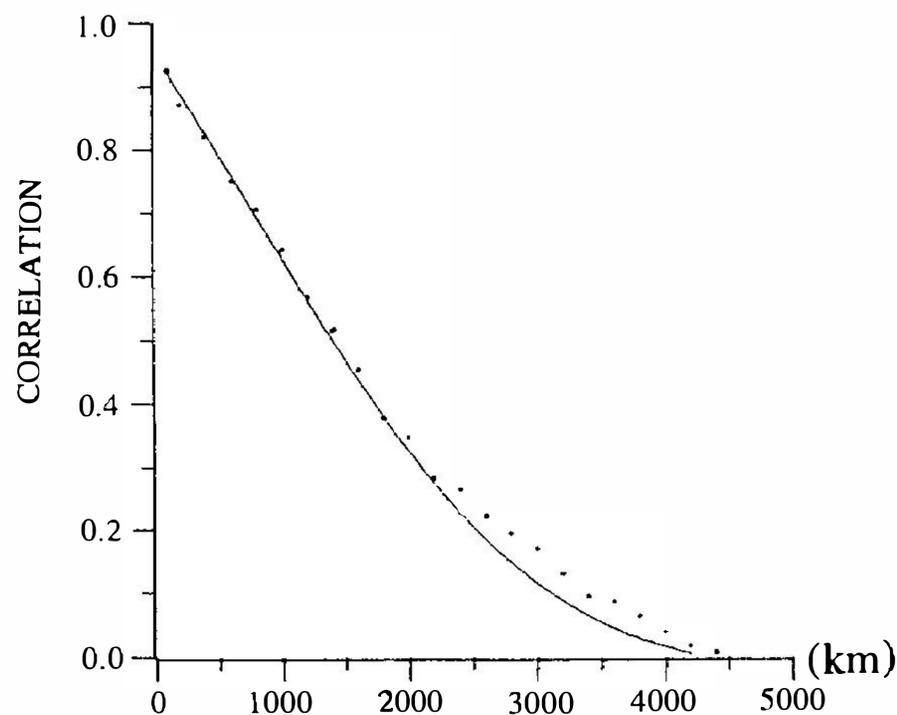


Fig. 5. The mixed-typed autocorrelation model and distance-averaged autocorrelation data.

Table 1. Coefficients for  $m_a = 5$  and  $m_b = 5$ .

$a_0$	0.99998E-01	$b_0$	0.22635E+01
$a_1$	0.23054E-02	$b_1$	-0.35200E+00
$a_2$	-0.33989E-03	$b_2$	-0.68385E-01
$a_3$	-0.11815E-02	$b_3$	-0.71690E-02
$a_4$	-0.10476E-04	$b_4$	0.12955E-02
$a_5$	0.46725E-04	$b_5$	0.91694E-03

### 3. OPTIMUM INTERPOLATION

The analyzed geopotential height value  $\hat{z}_0$  at the specific grid  $o$  is represented in the form of a linear combination of the values  $z_i$  on the neighboring observational points. Such a linear combination may be expressed as:

$$\hat{z}_0 = \sum_{i=1}^N \alpha_i z_i, \quad (9)$$

where  $\alpha_i$  represents the weighting factor at the observation point  $i$  and  $N$  is the number of stations used for interpolation. The weights  $\alpha_i$  were determined from the required condition such that the above equation yields the best results when averaging is done over a large number of cases. In other words, the mean square of errors,

$$E = \overline{(\hat{z}_0 - z_0)^2}, \quad (10)$$

represent a minimum, where  $z_0$  is the true value at grid  $o$ , and the bar denotes the statistical averaging. Substituting Equation (9) into Equation (10) yields:

$$E = \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j \overline{z_i z_j} - 2 \sum_{i=1}^N \alpha_i \overline{z_i z_0} + \overline{z_0^2}. \quad (11)$$

The condition for the minimum of  $E$  with respect to all weight  $\alpha_k$ , such that

$$\frac{\partial E}{\partial \alpha_k} = 0 \quad (k = 1, 2, \dots, N), \quad (12)$$

becomes the form:

$$\sum_{j=1}^N \overline{z_i z_j} \alpha_j = \overline{z_i z_0} \quad (i = 1, 2, \dots, N). \quad (13)$$

It can be shown that the conditions of Equation (13) are both necessary and sufficient for the minimum of  $E$ . This yields a system of simultaneous  $N$  linear algebraic equations to solve the weights of the associated stations.

In order to determine the weights from the system, the mean product of  $\overline{z_i z_j}$  and  $\overline{z_i z_0}$  are required. However, since  $z_0$  is never known in the physical domain, it is better to use the expression below.

Let

$$\begin{aligned} z_i &= \overline{z_i} + z'_i \\ z_0 &= \overline{z_0} + z'_0 \end{aligned} \quad (14)$$

where the bar and the prime denote the statistical ensemble mean and the deviation from its mean respectively. Substituting these into Equation (13), it becomes

$$\sum_{j=1}^N \overline{z'_i z'_j} \alpha_j = \overline{z'_i z'_0} \quad (i = 1, 2, \dots, N). \quad (15)$$

Unlike Equation (13), in Equation (15) these formulas have the mean products of the values not of the elements themselves, but of the deviations from the ensemble mean. It should be noticed that when (15) is substituted into (11) and let the result divided by the variance  $C_{00}$ , the relative interpolation error can also be given by

$$E' = 1 - \sum_{m=1}^N \rho_{m0} \alpha_m.$$

This quantity can be denoted as a measurement of accuracy in the analysis scheme.

Let  $\mathcal{R}_0$ ,  $\mathcal{R}_i$  be the variance of the geopotential height fields on the grid point  $o$  and the observation point  $i$  respectively:

$$\mathcal{R}_0 = \overline{(z'_0)^2} \quad ; \quad \mathcal{R}_i = \overline{(z'_i)^2} \quad (16)$$

The autocovariance function can be defined as:

$$\varphi_{ij} = \overline{z'_i z'_j}. \quad (17)$$

Similarly, the autocorrelation function can be expressed in terms of the autocovariance and variance, which is in the form of:

$$\rho_{ij} = \frac{\varphi_{ij}}{\sqrt{\mathcal{R}_i \mathcal{R}_j}}. \quad (18)$$

Equation (15) now becomes

$$\sum_{j=1}^N \rho_{ij} \alpha_j = \rho_{io}, \quad (i = 1, 2, \dots, N). \quad (19)$$

where  $i$  and  $j$  are the observing points, and  $o$  is the target grid point. Once the members of the observing points are known, then the coordinates of the grid point and the surrounding observing points are used to compute the relative distance among these points. The optimum weight  $\alpha_i$  associated with each of the observation points  $i$  are then computed from Equation (19).

The deviation of the analyzed geopotential height can thus be obtained from:

$$\hat{z}'_0 = \sum_{k=1}^N \alpha_k z'_k. \quad (20)$$

The analyzed  $\hat{z}_0$  is the sum of the deviation and its respective climatological mean  $\bar{\hat{z}}_0$ , thus:

$$\hat{z}_0 = \bar{\hat{z}}_0 + \hat{z}'_0. \quad (21)$$

At this point, the analyzed geopotential height on the two dimensional grids can be obtained. In practice, for the univariate OI scheme, there are two ways to construct the autocorrelation function in Equation (19). First, as for the explicit form, the autocorrelation in (19) is constructed by using the distance-dependent autocorrelation model. The optimum interpolation coefficients in (19) can be obtained directly from the computation of the relative distances among the stations and the grids. Second, regarding the implicit form, the autocorrelation function in (19) is constructed by using the observational data on the stations and the analysed data on the target grids; numerical weather prediction model output data are also valid accordingly. In fact, the explicit autocorrelation model (distance dependent autocorrelation model) is not a necessary requirement for the construction of the initial field analysis. The optimum interpolation coefficients associated with the specific grid can be obtained directly from the nearby real observational data to compute the correlation functions among the pairs of grids and stations. However, in this way (i.e., the implicit form), the memory requirements would be large if the global grid data are computed. By using the distance dependent autocorrelation model, there are no such problems.

The optimum interpolations can be easily done by calculating the relative distances among the grid and the stations. The optimum coefficients associated with the nearby stations can thus be solved from the simultaneous equations.

#### 4. RESULTS

Excellent evaluation for the test analysis scheme was based upon the root mean squares between the observational data and the analysed data from the scheme previously discussed.

The univariate statistical analysis scheme with autocorrelation calculated in section 2 was applied at 500 hPa to compute the root mean square differences between observations and analysed values for the 108 cases involved.

Both the explicit and implicit forms are used under different conditions. In solving Equation (19), the observational data are used to compute the left hand side of the element  $\rho_{ij}$  while the proposed autocorrelation model is used to compute the right hand side of this element  $\rho_{i0}$ . The analysed target grids include 25 stations, ranging from the observation index points 1 to 24, and the Taipei station was indexed as 58, the climatological mean was chosen as the first guess for the computation on the left hand side. The root mean square differences between the analyzed and observed values were computed for the selected 108 days involved.

Table 2 shows how the results depend upon the number of reports used to estimate the univariate geopotential height at these target stations.

In Table 3, the relative interpolation error  $E'$  was computed using the observational data for the analyses of the previous 25 target stations. Since the quantity denotes the measurement of accuracy, the observational data instead of the explicit autocorrelation model would be

Table 2. Root mean square differences of the geopotential height fields between the analysed and observed values(m).

Number of Stations Used for Interpolation		Geopotential Height Fields																								
		First Guess												Climatology												
		Root-Mean-Squares																								
		58	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1		21	29	22	21	28	20	18	19	20	17	22	15	20	30	25	16	15	16	19	20	30	27	27	16	26
2		20	20	20	17	21	17	16	14	16	16	18	15	13	25	23	12	13	15	16	20	28	26	24	15	20
3		17	20	16	12	20	16	15	14	14	13	16	14	12	24	23	12	12	13	15	16	28	23	23	14	20
5		16	20	16	11	17	14	14	13	13	13	12	14	10	20	22	11	11	13	15	14	28	21	20	14	16
7		16	19	14	10	16	14	14	13	13	13	12	13	10	19	21	11	10	13	14	14	28	20	20	14	16
8		16	19	14	10	16	13	13	13	13	12	12	13	10	19	21	11	10	13	14	14	28	20	18	14	16
10		16	19	14	10	16	13	13	13	12	12	11	13	10	18	21	11	10	12	14	14	27	18	18	14	16
15		16	19	13	10	15	13	13	12	11	11	11	12	10	17	20	10	9	12	12	14	26	17	17	13	15
17		15	19	13	10	15	12	13	12	11	11	11	12	10	17	20	10	9	12	12	13	26	17	17	13	15
20		15	18	13	10	15	12	12	12	11	11	11	12	10	17	20	9	9	12	11	13	26	17	17	13	15

Table 3. The relative interpolation error of the geopotential height fields computed based upon the nearby stations used for interpolation.

Geopotential Height Fields

Number of Stations Used for Interpolation  $E' = 1 - \sum_{k=1}^{N_0} \rho_{k0} \alpha_k$  First Guess Climatology

	58	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	0.25	0.45	0.25	0.31	0.28	0.30	0.23	0.14	0.31	0.29	0.21	0.12	0.17	0.27	0.19	0.16	0.29	0.31	0.39	0.24	0.50	0.24	0.17	0.12	0.24
2	0.24	0.22	0.20	0.18	0.16	0.23	0.18	0.08	0.20	0.23	0.13	0.12	0.07	0.19	0.16	0.14	0.18	0.31	0.28	0.22	0.45	0.21	0.14	0.11	0.18
3	0.18	0.22	0.13	0.10	0.14	0.20	0.15	0.07	0.16	0.15	0.10	0.11	0.06	0.18	0.15	0.13	0.16	0.23	0.25	0.15	0.44	0.16	0.12	0.10	0.12
5	0.17	0.21	0.12	0.08	0.10	0.16	0.13	0.07	0.14	0.15	0.06	0.10	0.05	0.12	0.14	0.11	0.12	0.22	0.23	0.11	0.44	0.14	0.09	0.09	0.12
7	0.17	0.21	0.10	0.07	0.10	0.15	0.13	0.06	0.12	0.15	0.06	0.09	0.04	0.11	0.14	0.11	0.12	0.22	0.21	0.11	0.44	0.13	0.09	0.09	0.12
8	0.17	0.21	0.10	0.07	0.09	0.14	0.12	0.06	0.12	0.15	0.06	0.09	0.04	0.11	0.14	0.11	0.12	0.22	0.21	0.11	0.43	0.13	0.08	0.09	0.12
10	0.17	0.20	0.10	0.07	0.09	0.13	0.12	0.06	0.11	0.14	0.05	0.09	0.04	0.10	0.13	0.10	0.12	0.20	0.20	0.11	0.40	0.11	0.08	0.09	0.11
15	0.16	0.20	0.09	0.07	0.09	0.12	0.11	0.06	0.10	0.11	0.05	0.08	0.04	0.09	0.12	0.09	0.09	0.19	0.16	0.11	0.38	0.10	0.07	0.08	0.11
17	0.15	0.20	0.09	0.06	0.09	0.12	0.11	0.06	0.10	0.11	0.05	0.08	0.04	0.09	0.12	0.09	0.09	0.18	0.15	0.11	0.38	0.10	0.07	0.08	0.09
20	0.15	0.18	0.09	0.06	0.08	0.12	0.11	0.06	0.10	0.11	0.05	0.08	0.04	0.09	0.11	0.08	0.09	0.18	0.13	0.10	0.37	0.09	0.07	0.08	0.09

more appropriate for the computation of  $\rho_{i0}$  in the relative interpolation error computation. It can be easily seen that the root mean squares and the relative interpolation errors definitely decrease as more stations are used in the analysis scheme. Gandin (1963) assumed that the root mean squares would decrease indefinitely as more stations are used in the interpolation process, but in fact all measurements are subject to error. Instrumental errors especially are always inherent in instrument or manual operations. Additional reports in the estimation improve the accuracy of the estimate but only up to a certain limit. Normally, for a report providing high-quality data, most of the information pertinent to the analysis can be obtained from the nearest 6-8 stations. The addition of more stations does not significantly improve the estimate. The previous results can also be shown in Tables 4 and 5 which are the analysis of the relative interpolation error computation for the u and v components of the wind fields.

Table 4. The relative interpolation error of the u-component wind fields computed based upon the nearby stations used for interpolation.

U Component Wind Fields

Number of Stations Used for Interpolation  $E' = 1 - \sum_{k=1}^{N_0} \rho_{k0} \alpha_k$  First Guess Climatology

	58	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	0.63	0.82	0.63	0.64	0.63	0.85	0.64	0.64	0.66	0.60	0.42	0.57	0.40	0.47	0.72	0.66	0.74	0.66	0.48	0.49	0.49	0.74	0.42	0.47	
2	0.58	0.61	0.55	0.57	0.32	0.76	0.42	0.29	0.52	0.63	0.54	0.42	0.49	0.34	0.42	0.58	0.61	0.63	0.60	0.42	0.28	0.33	0.59	0.33	0.42
3	0.55	0.56	0.45	0.56	0.30	0.74	0.41	0.29	0.46	0.59	0.42	0.34	0.33	0.34	0.38	0.50	0.49	0.51	0.55	0.40	0.27	0.29	0.50	0.29	0.40
5	0.49	0.53	0.42	0.52	0.28	0.73	0.34	0.27	0.41	0.49	0.35	0.33	0.29	0.31	0.37	0.46	0.42	0.50	0.54	0.36	0.27	0.27	0.48	0.27	0.34
7	0.49	0.51	0.40	0.51	0.28	0.67	0.33	0.26	0.38	0.49	0.33	0.31	0.29	0.29	0.36	0.44	0.41	0.48	0.53	0.34	0.25	0.26	0.47	0.27	0.34
8	0.49	0.49	0.39	0.46	0.28	0.58	0.33	0.26	0.37	0.49	0.32	0.31	0.29	0.28	0.36	0.43	0.41	0.48	0.53	0.33	0.24	0.26	0.42	0.27	0.34
10	0.48	0.46	0.35	0.44	0.27	0.56	0.33	0.26	0.37	0.46	0.31	0.31	0.28	0.28	0.35	0.40	0.39	0.48	0.43	0.32	0.24	0.24	0.41	0.24	0.33
15	0.44	0.45	0.34	0.41	0.24	0.55	0.32	0.25	0.28	0.31	0.30	0.29	0.28	0.27	0.32	0.39	0.33	0.46	0.42	0.31	0.24	0.23	0.38	0.24	0.33
17	0.44	0.45	0.34	0.41	0.23	0.55	0.31	0.24	0.28	0.30	0.28	0.29	0.27	0.27	0.32	0.39	0.31	0.43	0.41	0.31	0.23	0.22	0.36	0.23	0.32
20	0.43	0.43	0.33	0.40	0.23	0.54	0.30	0.23	0.26	0.30	0.26	0.28	0.27	0.26	0.29	0.39	0.29	0.43	0.39	0.31	0.22	0.21	0.36	0.22	0.31

Table 5. The relative interpolation error of the v-component wind fields computed based upon the nearby stations used for interpolation.

V Component Wind Fields

Number of Stations Used for Interpolation

$$E' = 1 - \frac{N_p}{\sum_{i=1}^{N_p} \rho_{i0} \alpha_i}$$

First Guess Climatology

	58	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	0.45	0.66	0.45	0.59	0.62	0.56	0.51	0.36	0.59	0.72	0.80	0.17	0.64	0.45	0.18	0.27	0.72	0.46	0.67	0.43	0.51	0.78	0.59	0.17	0.18
2	0.35	0.43	0.44	0.37	0.20	0.47	0.34	0.18	0.50	0.51	0.78	0.16	0.34	0.44	0.18	0.26	0.42	0.44	0.61	0.42	0.40	0.41	0.47	0.17	0.18
3	0.28	0.39	0.36	0.26	0.19	0.45	0.34	0.17	0.30	0.25	0.51	0.14	0.17	0.43	0.17	0.26	0.40	0.29	0.61	0.40	0.38	0.36	0.33	0.14	0.18
5	0.28	0.39	0.32	0.20	0.16	0.44	0.28	0.17	0.29	0.24	0.27	0.13	0.17	0.36	0.17	0.25	0.37	0.29	0.37	0.34	0.37	0.31	0.32	0.14	0.17
7	0.28	0.38	0.24	0.20	0.15	0.44	0.26	0.17	0.28	0.23	0.27	0.12	0.16	0.26	0.16	0.25	0.37	0.22	0.37	0.34	0.32	0.30	0.32	0.13	0.17
8	0.28	0.38	0.24	0.20	0.15	0.28	0.26	0.16	0.28	0.23	0.26	0.12	0.16	0.26	0.16	0.25	0.36	0.21	0.36	0.34	0.32	0.29	0.25	0.13	0.17
10	0.27	0.37	0.22	0.19	0.15	0.27	0.26	0.16	0.28	0.22	0.25	0.12	0.16	0.25	0.16	0.24	0.35	0.21	0.34	0.33	0.32	0.25	0.25	0.13	0.16
15	0.25	0.35	0.19	0.18	0.14	0.25	0.23	0.16	0.25	0.20	0.24	0.12	0.15	0.21	0.16	0.23	0.33	0.20	0.32	0.31	0.28	0.24	0.24	0.13	0.16
17	0.24	0.35	0.19	0.18	0.14	0.24	0.23	0.15	0.25	0.20	0.24	0.12	0.14	0.21	0.15	0.22	0.32	0.19	0.31	0.29	0.28	0.23	0.23	0.13	0.16
20	0.23	0.34	0.18	0.17	0.14	0.24	0.22	0.14	0.22	0.20	0.22	0.11	0.11	0.21	0.15	0.20	0.31	0.18	0.30	0.28	0.25	0.23	0.23	0.13	0.16

The 500 hPa geopotential heights at Taipei were analysed based upon a single observation of height. The experiments were conducted repeatedly, and each time the observation station report was chosen to be farther from Taipei. From Figure 6, it is seen that as the distance of the station is farther from Taipei, the root mean square is increasingly higher than the previously near station. This implies that the more distant observations are only slightly correlated with the grid points and have little influence on the estimates.

It is important to make a comparison between the classical Gandin form and the new proposed mixed-type form. The distance averaged autocorrelation functions of the 500 hPa geopotential height fields are fitted to the negative exponential form as proposed by Gandin

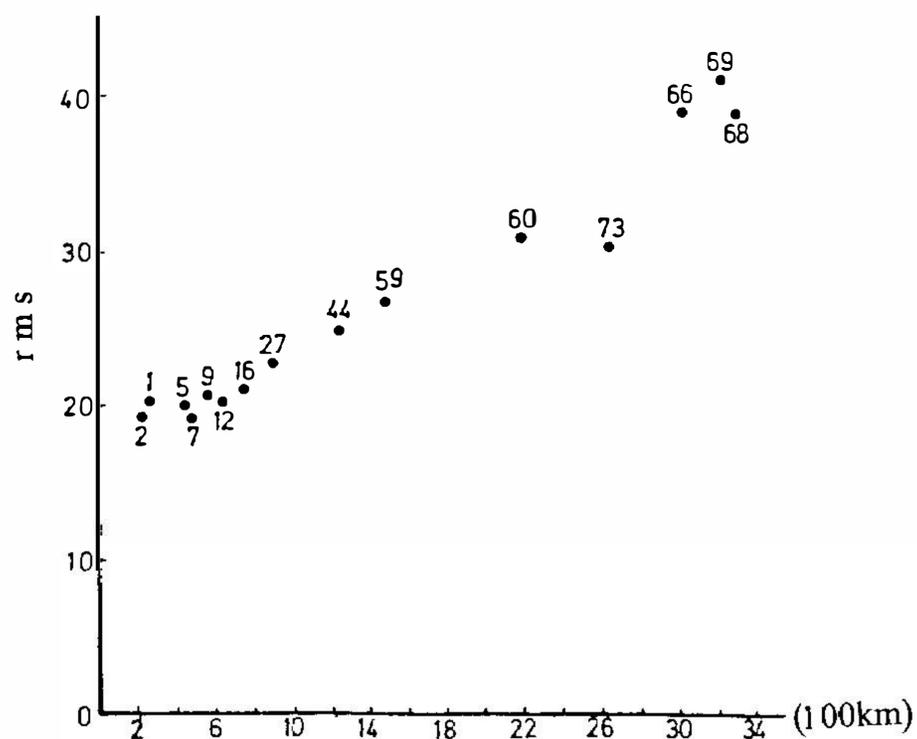


Fig. 6. Root-mean-square differences between the analyzed and observed values of height (m) at Taipei for the 108 cases involved. The experiment was conducted repeatedly. Each time the observation point was chosen to be farther from Taipei.

(1963). Figure 7 shows the results with  $a = 0.87654$  and  $b = 0.31543$ ; the optimum constants are determined by the same least square method as that of the mixed-typed autocorrelation models. Accordingly; it is obvious that due to the fixed form of the negative exponential autocorrelation model, the curve fitting is not as good as the new proposed mixed-type autocorrelation model.

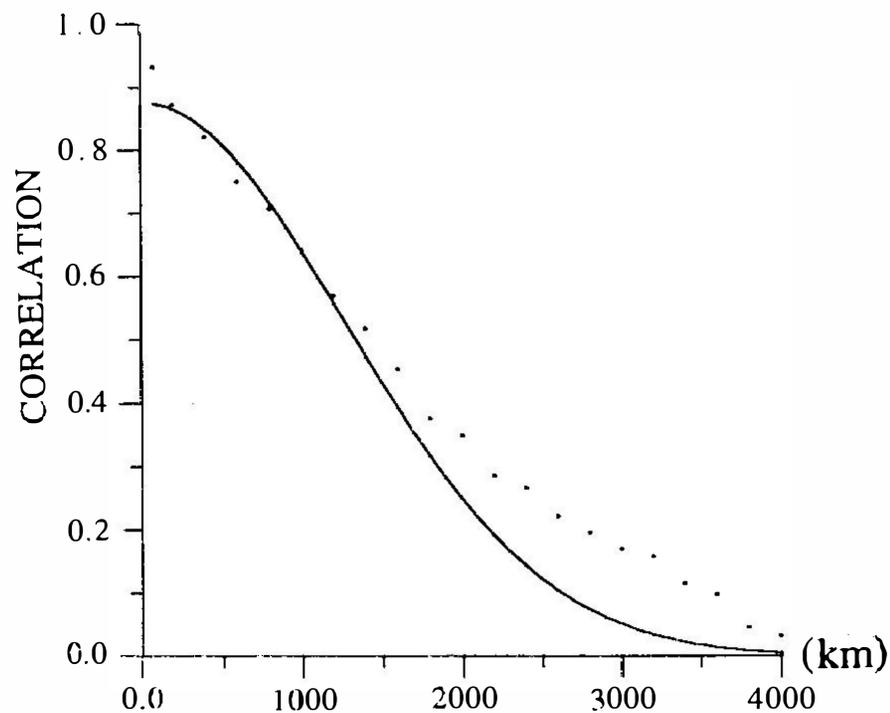


Fig. 7. The negative exponential autocorrelation model and distance-averaged autocorrelation data.

In order to determine the superior capability of the two forms in analyzing the geopotential height fields, the two autocorrelation models were also utilized to compute the root mean squares between the analysed and the observational values. In Figures 8 and 9, the eight stations are used for interpolation. The root mean squares for the  $i$ -th station were computed using the formula:

$$\text{Sig}(i) = \sqrt{\sum_{j=1}^{108} (z_{ij} - \hat{z}_{ij})^2 / 108}.$$

Figure 8 is for the classical Gandin form, while Figure 9 is for the proposed mixed-typed form. Comparing the two figures, it is evident that the amplitudes of Figure 9 are smaller than those of Figure 8 in estimating the root mean squares of the total 78 stations. This implies that the proposed mixed-typed autocorrelation model has a better capability in analyzing the geopotential height fields.

## 5. CONCLUSIONS

The statistical structure of the three meteorological variables ( $h$ ,  $u$  and  $v$ ) have been derived by computing the autocorrelation and cross-correlation functions, and the geopotential height autocorrelation pairs among the 78 stations have also been computed. These pairs were grouped in 200-kilometer distance intervals from the range of zero to six thousand kilometers. The resulting autocorrelation functions were then plotted against station separation; the new

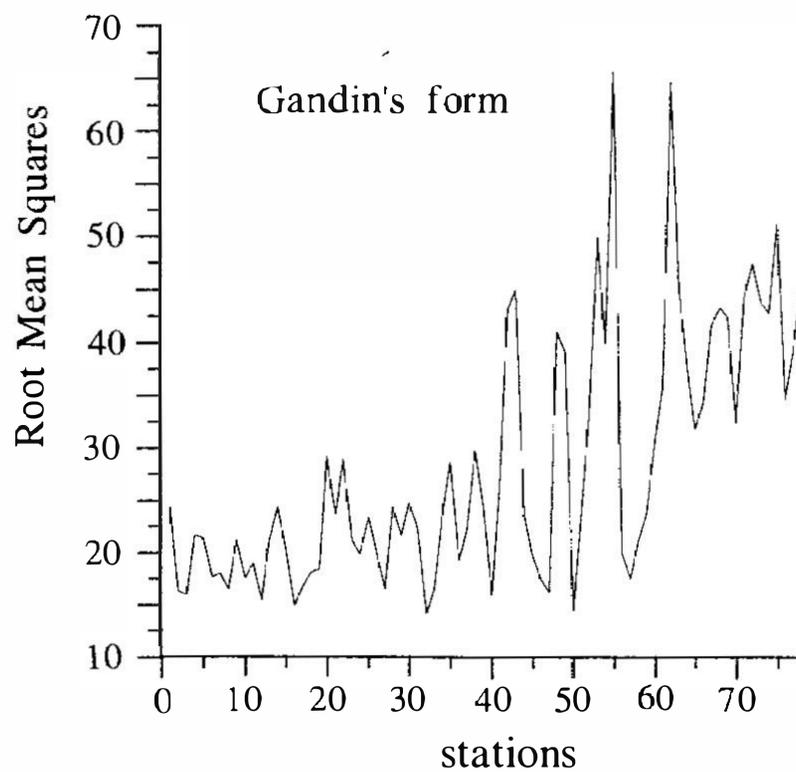


Fig. 8. The root-mean-square differences between the analyzed values and observed values for the 78 stations. Each station include 108 cases. The nearest 8 stations were used for interpolation (Gandin form was used in this study)

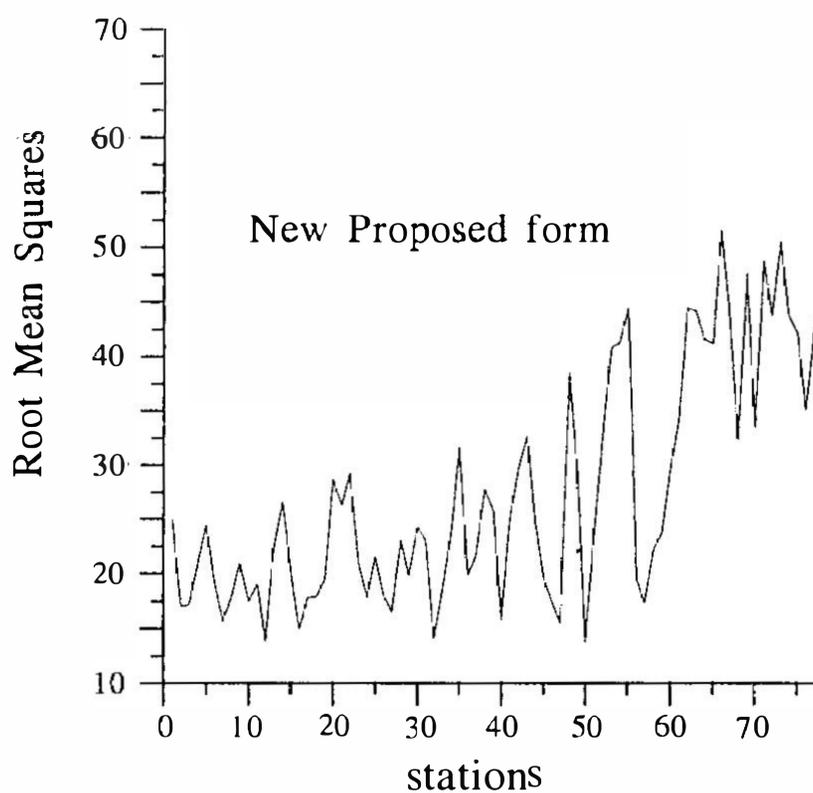


Fig. 9. The root-mean-square differences between the analyzed values and observed values for the 78 stations. Each station include 108 cases. The nearest 8 stations were used for interpolation new (The proposed form was used in this study)

autocorrelation model was fitted to the scatter diagram in the least squared sense. The following conclusions can be made:

- (1) The scatter pattern shows that the clustering does normally distribute along the exponential function patterns with the magnitude of the correlation coefficients decreasing slowly with distance.
- (2) Zero correlation coefficients occur at separations of some three thousand and eight hundred kilometers. The distance averaged data are shown to spread along the exponential

functions, a finding can be determined by the new autocorrelation model in the least square sense which is a function of the relative distance only.

- (3) The statistical correlation functions of the geopotential heights and wind components derived from the low- and mid latitudinal regions have been found to be extremely useful in the initial field constructions for numerical weather prediction. The original data base are from historical observational data. Thus, it is likely that the autocorrelation model for the geopotential height fields would definitely reflect the statistical properties in this research area. By adding the order of the polynomial terms, the autocorrelation model proposed in this paper proves to be more powerful in tracing the finer structure of the autocorrelation functions, especially in the low latitudinal region where the large area is surrounded by the sea. The accuracy of the autocorrelation model is important for the full range of separation distances between the target points and the interpolation stations.
- (4) Because of the similar patterns in autocorrelation data distributions for the mid- and low latitudinal regions, it can be anticipated that the algorithms for the autocorrelation model are also valid for different latitudes of the earth as well.

**Acknowledgments** The first author would like to thank Dr. W-S. Kau, Dr. LinHo, Dr. H-C. Kuo, Dr. C-S. Lee, Dr. J.-D. Jou, Dr. H.-H. Hou, Department of Atmospheric Science, National Taiwan University and Dr. S-C. Lin, Dr. C-S. Chen, Dr. G-R. Liu, Institute of Atmospheric Physics, National Central University, Dr. S.-B. Yu, Institute of Earth Sciences, Academia Sinica, R.O.C. and Dr. M-D. Chen, Research and Development Center, Central Weather Bureau, R.O.C., The comments from two anonymous reviewers are greatly appreciated.

## REFERENCES

- Achtemeier, G. L., 1989: Modification of a successive correction objective analysis for Improved Derivative Calculations. *Mon. Wea. Rev.*, **117**, 79-86.
- Baker, W. E., 1983: Objective analysis and assimilation of observation data from FGGE. *Mon. Wea. Rev.*, **112**, 1544-1561.
- Balgovirid, R., A. Dalcher, M. Ghil, and E. Kalnay, 1983: A stochastic-dynamic model for the spatial structure of forecast error statistics. *Mon. Wea. Rev.*, **111**, 701-722.
- Belousov, S. L., L. S. Gandin, and S. A. Mashkovich, 1968: Computer Processing of Meteorological Data. Israel Program for Statistical Translation, 210pp.
- Bengtsson, L., and N. Gustavsson, 1972: Assimilation of non-synoptic observations. *Tellus*, **24**, 383-399.
- Bengtsson, L., 1975: 4-Dimensional assimilation of meteorological observations. GARP Publ. Ser., No. 15, 76pp. [Available from WMO, Case Postale No.5, CH 1211, Geneva 20, Switzerland.]
- Bengtsson, L., 1976: Initial Data and Some Practical Aspects of Weather Forecasting and Forecasts, Models, Problems and Systems, Vol. 1, National Center for Atmospheric Research, Boulder, Colorado.

- Bergman, K. H., 1979: Multivariate analysis of temperatures and wind fields using optimum interpolation. *Mon. Wea. Rev.*, **107**, 1423-1444.
- Bergthorsson, P., and B. Doos, 1955: Numerical weather map analysis, *Tellus*, **7**, 329-340.
- Bouttier, F., 1994: A dynamical estimation of forecast error covariances in an assimilation. *Mon. Wea. Rev.*, **122**, 2376-2390.
- Buell, C. E., 1958: The correlation between wind and height on the isobaric surface. *J. Meteor.*, **15**, 502-512.
- Eddy, A., 1967: The statistical objective analysis of scalar data fields. *J. Appl. Meteor.*, **6**, 597-609.
- Eddy, A., 1973: The objective analysis of atmospheric structure. *J. Meteor. Soc. Japan*, **51**, 450-457.
- Eliassen, A., 1954: Provisional report on the calculation of spatial covariance and autocorrelation of the pressure field. Videnskaps-Akademiets Institutt for Vaerog Klimaforskning, Report No.5, 11pp.
- Gandin, L. S., 1963: Objective Analysis of meteorological fields *Gidrometeorologicheskoe Iza-tel'stvo*, Leningrad. Translated from Russian, Israel Program for Scientific Translations, Jerusalem, 1965, 242pp.
- Gandin, L. S., 1964: On the optimum interpolation of vector fields. *Trudy Glavnoi Ge-ofizicheskoi observatorii (GGO)*, Imeni A.I. Voeikova, Leningrad, No. 165, 47-49.
- Gilchrist, B., and Cressman, G. 1954: An experiment in objective analysis. *Tellus*, **6**, 309-318.
- Haltiner, G. J., and Williams, R. T., 1980: *Numerical Prediction and Dynamic Meteorology*, John-Wiley, New York, 477pp.
- Hagenson, P. L., 1983: Review and evaluation of methods for the objective analysis of meteorological variables. *Papers Meteor Res.*, **5**, 113-133.
- Hildebrand, F. B., 1965: *Methods of Applied Mathematics*. Prentice-Hall, Englewood, N.J., 362pp.
- Inman, R. L., 1970: *Papers on Operational Objective Analysis Scheme at NSSFC*. NOAA Tech. Memo. ERL NSSL-51, 91pp.
- Lorenc, A., 1979: *Meteorological Data Analysis Lecture Note, No. 3*. European Centre for Medium Range Weather Forecasting, Reading, England, 68pp.
- Lonnberg, P., and Hollingsworth, A. 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part II: The covariance of height and wind errors. *Tellus*, **38A**, 137-161.
- Lorenc, A., 1979: *Meteorological Data Analysis Lecture Note, No. 3*. European Centre for Medium Range Weather Forecasting, Reading, England, 68pp.
- Lorenc, A., 1981: A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.*, **109**, 701-721.
- Lorenc, A., 1988a: A practical approximation to optimal four-dimensional data assimilation. *Mon. Wea. Rev.*, **116**, 730-745.

- Lorenc, A., 1988b: Optimal nonlinear objective analysis. *Q. J. Roy. Meteor. Soc.*, **114**, 205-240.
- Mitchell, H. L., 1990: Comment on "Monitoring of observation and analysis quality by a data assimilation system". *Mon. Wea. Rev.*, **118**, 1927-1928.
- Nuss, W. A., and Titley, D. W. 1993: Use of multiquadric interpolation for meteorological objective analysis. *Mon. Wea. Rev.*, **122**, 1611-1631.
- Ooyama K. V., 1987: Scale-controlled objective analysis. *Mon. Wea. Rev.*, **115**, 2479-2506.
- Perrie. W., and B. Toulany, 1989: Correlations of sea level pressure fields for objective analysis. *Mon. Wea. Rev.*, **117**, 1965-1974.
- Ramamurthy. M. K., and F. H. Carr, 1987: Four-dimensional data assimilation in the monsoon region. Part I: Experiments with wind data. *Mon. Wea. Rev.*, **115**, 1678-1706.
- Onn, S. C., K. H. Wang, G. L. Chang, and C. Y. Tseng, 1992: Numerical simulation of the correlations with the general polynomial and exponential forms for the climatic statistical structure. Proc. 40th Summer Computer Simulation Conference, July 27-30, Reno, Nevada, U.S.A., 199-206.
- Onn, S. C., K. H. Wang, and C. Y. Tseng, 1993: A computational method for constructing the missing data of some observational stations using the climatic statistical structure. Proc. 73rd Annual Meeting of the American Meteorological Society, the 8th Conference on Applied Climatology, 116-121.
- Onn, S. C., K. H. Wang, and C. Y. Tseng, 1994: Numerical simulation for the climatic statistical structure with the new correlation functions of the extended polynomial form and the extended exponential form, COSPAR, *Adv. Space Res.*, **14**, 77-81.
- Schlatter, T. W., 1975: Some experiments with a multivariate statistical objective analysis scheme. *Mon. Wea. Rev.*, **103**, 246-257.
- Twomey, S., 1970: Information content and indirect sensing measurements. *J. Atmos. Sci.*, **27**, 515-518.
- Uden, P., 1989: Tropical data assimilation and analysis of divergence. *Mon. Wea. Rev.*, 2495-2517.
- Weber, R. O., and P. Talkner, 1993: Some remarks on spatial correlation function models. *Mon. Wea. Rev.*, **121**, 2611-2617.