On the Correlation Fields and the Predicting Possibilities by Statistical Methods at 500 mb Level

by

M. Magata, T. Ozawa, E. Suzuki, T. Fujita and K. Tomatsu

Meteorological Research Institute, Tokyo

(Received January 28, 1958)

Abstract

Cross correlation coefficients between the 500 mb contour height at about $50^{\circ}\text{N } 140^{\circ}\text{E}$ and those at the twenty five points in its vicinity, were calculated from the data of the Januaries 1948—1952 and the features of the correlation fields were investigated. The contour heights at the several points having the relatively stable correlation were selected as predictors and the 24-hour predictions of the 500 mb contour height at about $50^{\circ}\text{N } 140^{\circ}\text{E}$ were performed by the method of linear regression. When time lag τ is relatively large, the correlation coefficients at above twenty five points become small and unstable, and so cross correlation coefficients of lag 1—10 days between the 500 mb contour height at about $50^{\circ}\text{N } 140^{\circ}\text{E}$ and those at many points over the northern hemisphere, were calculated and the behaviours of the distributions of correlation coefficients were discussed. At the same time, the theoretical correlation field was obtained from the linearized vorticity equation and compared with the actual correlation fields, and the features of the correlation field were discussed by the method of spectrum analysis. Some considerations on the nonlinear system were also made.

1. Introduction

Statistical weather predictions of the contour height have been studied by many authors such as WADSWORTH and BRYAN (1948), WADSWORTH (1951) WHITE and PALSON (1955), MALONE (1955), LORENZ (1956), and WHITE and collaborators (1957). MALONE (1955) represented the contour height pattern by the orthogonal polynomials and predicted the pattern by the theory of linear regression on the coefficients of polynomials. LORENZ (1956) proposed an efficient prediction method using the empirical orthogonal functions. WHITE and PALSON (1955) used the three constant pressure levels 1000 mb, 500 mb and 200 mb and selected as predictors the contour heights at fifteen irregularly spaced points on each of the three pressure levels and performed the prediction of the contour heights at above forty-five points. In this paper the cross correlation coefficients between the 500 mb contour height at about $50^{\circ}\text{N } 140^{\circ}\text{E}$ and those at many points over the northern hemisphere were calculated and the behaviours of the distributions of the correlation coefficients and the pre-

dicting possibilities of the 500 mb contour height at 50°N 140°E , were investigated.

At first the correlation coefficients of lag 1–5 days between the 500 mb contour height at about 50°N 140°E and those at twenty five points in its vicinity, were computed from the data of the Januaries 1949–1952, and the same correlations were also computed from each of the Januaries 1948–1952. The features of correlation fields were investigated for every year and it was found that in order to obtain good accuracy of prediction, it is necessary to take the difference between the correlation fields of every year into account. And it is desired to classify all the correlation fields of every year according to their features, but it is difficult to find an objective rule of classification. In this paper two kinds of prediction formulas were formed: the one from the data of the Januaries 1949–1952 and the other from each January of several years. It has not yet been decided which formula is better.

In order to find an area in which there is a good correlation for the 500 mb height at 50°N 140°E when the time lag is relatively large, the correlation fields of lag 1–10 days were calculated over the northern hemisphere. And it was found that the area of good correlation moves westwards, as time lag becomes larger. On the other hand, it is desired that a theoretical correlation field may be calculated from the equations of motion and compared with the correlation fields obtained from the observed data. But it is difficult to deal with the nonlinear equations of motion. Therefore a theoretical correlation field was calculated from the linearized vorticity equation and it was found to be a good approximation of the real. The relations between the correlation field of the linear system and that of the real, were considered by the method of spectrum analysis. Lastly, it seems to be convenient to investigate the physical and statistical features of the departures of the contour height from the linear system.

2. Characteristics of the correlation field in the neighbourhood of Japan

Fig. 1. Location of the grid points used for the derivation of correlation field. A pair of the numbers in the left and upper edge specifies a grid point.

The locations of the grid points used for computation of the correlation coefficients, the intervals of which are 1200 km, are shown in Fig. 1. Hereafter, a position of grid point will be indicated by the numbers on the upper left edge of Fig. 1. The correlation fields of lag 1–5 days between 500 mb height at the point (5,5) near 50°N , 140°E and those at the surrounding 25 points are shown in Fig. 2. The case of January 1948 is shown in Fig. 2(a), Jan. 1949 in Fig. 2(b), Jan. 1950 in Fig. 2(c), Jan. 1951 in Fig. 2(d), Jan. 1952 in Fig. 2(e) and the case for the Januaries 1949–1952 is in Fig. 2(f).

A correlation coefficient was calculated by the following formula and then, the isolines of the correlation coefficients were drawn,

Fig. 2. Correlation field of the lag 0-5 days between the 500 mb height at the grid point (5,5) marked by black spot and those at the surrounding 25 points. The case of January 1948 is indicated in (a), Jan. 1949 in (b), Jan. 1950 in (c), Jan. 1951 in (d), Jan. 1952 in (e) and the four Januaries 1949-1952 in (f) respectively.

$$r_{ij}(\tau) = \frac{\sum_{t=1}^{N-\tau} (Z_{55}(t+\tau) - \bar{Z}_{55}) (Z_{ij}(t) - \bar{Z}_{ij})}{\sqrt{\sum_{t=1}^{N-\tau} (Z_{55}(t+\tau) - \bar{Z}_{55})^2 \cdot \sum_{t=1}^{N-\tau} (Z_{ij}(t) - \bar{Z}_{ij})^2}} \quad (i, j=1, 3, 5, 7, 9) \\ (\tau=0, 1, \dots, 5),$$

where $Z_{55}(t)$ is the height of 500 mb surface in the grid point (5,5) at time t , $Z_{ij}(t)$ are those in the surrounding 25 points, N is the number of data, τ is time lag, i, j , are the numbers of grid point and bar means arithmetic average for each datum. It is found from Fig. 2 (f) that the simultaneous correlation field for the four Januaries 1949—1952 has a nearly concentric circle-shaped pattern about the point (5,5), and according as the time lag τ increases, the absolute values of correlations become smaller and a center of maximum correlation slowly moves westward. However, if we examine the correlation field for each year in detail, a considerable difference would be found. For example, the characteristics mentioned above are very clear in the case of Jan. 1950, but in the other cases, especially for the lag 3, 4, and 5 days, there are considerable irregularities and it is difficult to find any regularity.

Though there is a common character to all the years in the correlation fields of the lag 0—2 days, the bias of each year from the field for the four months of 1949—1952, seems to be considerable. When a prediction formula has been derived by the multiple regression method based on the correlation field for the four months of the Januaries 1949—1952, the larger the foregoing bias in the year to be predicted is, the larger the error of prediction would become.

Judging from the correlation fields in Fig. 2, it may be expected that 24 hr prediction will be successful to some extent.

3. On the case of 24-hr prediction

In order to predict the height at the grid point (5,5) by means of the multiple regression method based on the foregoing correlation field, simultaneous equations with 25 unknowns should be solved. This solving procedure requires a great deal of labour, and hence the 500 mb heights at the grid points, to which the predictand is highly correlated in each year, were exclusively selected as predictors. Consequently, the 500 mb heights Z_{33} , Z_{35} , Z_{53} , Z_{55} , Z_{73} and Z_{75} in the grid points (3, 3), (3, 5), (5, 3), (5, 5), (7, 3) and (7, 5) were respectively adopted as predictors, and the 500 mb height at the point (5, 5) was selected as a predictand. That is to say, the multiple regression equation was defined in the following form:

$$(1) \quad Z_{55}(t+1) = A_{00} + A_{33}Z_{33}(t) + A_{35}Z_{35}(t) + A_{53}Z_{53}(t) + A_{55}Z_{55}(t) + A_{73}Z_{73}(t) + A_{75}Z_{75}(t).$$

The regression coefficients A_{ij} in this equation were determined by the least square method. These coefficients were derived from two kinds of data, namely the daily heights of the four months of the Januaries 1949—1952 and of January of each year.

i) The prediction based on the data of 1949—1952

The general situation of the correlation field in this case is shown in Fig. 2(f). The simple and multiple correlation coefficients between the predictand and the predictors which have been selected in the equation (1) are as follows:

Predictand	Predictor	Simple corr. coeff.	Multiple corr. coeff.
$Z_{55}(t+1)$	$Z_{33}(t)$	0.11	0.81
	$Z_{35}(t)$	0.32	
	$Z_{53}(t)$	0.61	
	$Z_{55}(t)$	0.71	
	$Z_{73}(t)$	0.46	
	$Z_{75}(t)$	0.45	

The regression coefficients in the equation (1) were determined by the simple corr. coeff. mentioned above and the 24-hr prediction of the 500 mb height at the point of (5, 5) was executed for thirty independent data, which were composed of the daily height of January 1948. The result of prediction was verified by the corr. coeff. between the predicted values and the observed. In our case, the corr. coeff. between the observed and the predicted heights was 0.57 and the corr. coeff. between the observed and the predicted height changes was 0.55. Though these results were not so good, the prediction for the dependent data, namely, the daily heights of January 1952 of course showed a good result, that is, the corr. coeff. between the observed and the predicted heights was 0.78 and the corr. coeff. between the observed and the predicted height changes was also the same. These results tell us that the number of data has been so insufficient that our data have to be enlarged. However, one of the methods of improving our prediction formula is to take into consideration the proper correlation field for each year. In the following, the regression coefficients in the equation (1) were derived from the data of January in several years and we predicted the 500 mb heights of February in those years as the independent data.

ii) The prediction based on the data of January 1948

The general situation of the correlation field in this case was shown in Fig. 2(a). The simple and multiple correlation coefficients between the predictand and the predictors are as follows:

Predictand	Predictor	Simple corr. coeff.	Multiple corr. coeff.
$Z_{55}(t+1)$	$Z_{33}(t)$	-0.07	0.71
	$Z_{35}(t)$	0.31	
	$Z_{53}(t)$	0.47	
	$Z_{55}(t)$	0.51	
	$Z_{73}(t)$	0.13	
	$Z_{75}(t)$	0.14	

Thus, using these simple corr. coeff., the regression coefficients in the equation (1) were determined and the independent data of February 1948 were predicted. The corr. coeff. between the predicted daily height changes and the observed has come up to 0.77.

To verify whether such a method gives good result in prediction also for other years, we shall next try the same kind of prediction from the data of January 1950.

iii) The prediction based on the data of January 1950

The situation of the correlation field in this case is shown in Fig. 2(c). The simple and multiple correlation coefficients between the predictand and the predictors are as follows:

Predictand	Predictor	Simple corr. coeff.	Multiple corr. coeff.
$Z_{55}(t+1)$	$Z_{33}(t)$	0.33	0.87
	$Z_{35}(t)$	0.57	
	$Z_{53}(t)$	0.75	
	$Z_{55}(t)$	0.71	
	$Z_{73}(t)$	0.55	
	$Z_{75}(t)$	0.57	

By using these values, the regression coefficients in the formula (1) were determined, and 24-hr prediction was performed for the data of February 1950. As a result, the corr. coeff. between the observed and the predicted height changes is 0.52 and this value is not so good as that obtained for 1948.

It may be considered that the prediction formula based on the data of the January of any one year may be applied to the January of any other year. Thus, the prediction formulas were obtained from the data of January 1948 and the prediction for January 1949 was performed by use of it. The corr. coeff. between observed and predicted height changes was 0.57. Furthermore, using the prediction formula of January 1950, the prediction for January 1952 was performed, and the corr. coeff. between the observed and the predicted height changes was 0.74. From the results obtained above, it may be expected that the good prediction result is obtained by means of the objective classification of all the correlation fields.

Next, in the prediction formula (1) based on the data of January 1949–1952, the correlation coefficient between $Z_{55}(t+1)$ and $Z_{33}(t)$ was small, and so we used the height $Z_{11}(t-3)$ at the grid point (1,1) in stead of the height $Z_{33}(t)$, and furthermore added the $Z_{39}(t-2)$ to the predictors. As shown in Fig. 2, the corr. coeff. between $Z_{55}(t+1)$ and $Z_{11}(t-3)$ is higher and more stable than at $Z_{33}(t)$, and thus the following prediction formula can be produced:

$$(2) \quad Z_{55}(t+1) = B_{00} + B_{11}Z_{11}(t-3) + B_{35}Z_{35}(t) + B_{53}Z_{53}(t) + B_{55}Z_{55}(t) \\ + B_{73}Z_{73}(t) + B_{75}Z_{75}(t) + B_{39}Z_{39}(t-2).$$

In order to determine the coeff. B_{ij} of formula (2), the data of January 1949–1952 were used. In this case, the corr. coeff. between the predictand and the predictors are as follows:

Predictand	Predictor	Simple corr. coeff.	Multiple corr. coeff.
$Z_{55}(t+1)$	$Z_{11}(t-3)$	-0.48	0.82
	$Z_{35}(t)$	0.32	
	$Z_{53}(t)$	0.61	
	$Z_{55}(t)$	0.71	
	$Z_{73}(t)$	0.46	
	$Z_{75}(t)$	0.45	
	$Z_{39}(t-2)$	-0.21	

Using the values obtained above, the coeff. B_{ij} were determined, and the 24-hr prediction was performed for January 1948. This result was verified by means of the corr. coeff. between the observed and the predicted values. The coeff. for the

Fig. 3. 72-hr prediction for the period of February 1–10 based on the data of Jan. 1949–1952. Full lines and dashed lines show the observed and predicted values respectively. And (a) is the prediction of the height and (b) the prediction of the height change.

observed and the predicted heights was 0.62, and for the height changes was 0.48. Comparing the above result with the result obtained from the prediction formula (1), though the corr. coeff. of verification for the heights obtained by formula (2) is better than that of formula (1), the corr. coeff. for the height changes conversely is worse. Using the same prediction formula as mentioned above, the predictions for the period of February 1–10, 1949–1952 were performed and the corr. coeff. between the predicted and the observed heights was 0.75, and that for the height changes was 0.42.

Now, with respect to the method of the verification of the predicting result, it is a problem whether the corr. coeff. for the heights is better or for the height changes. The correlation for the height changes seems to be better in the case of 24-hr prediction. Using the correlation field of lag 3 days, the 72-hr prediction was performed for the period of February 1–10, 1949–1952 by means of the same method. The corr. coeff. for the heights was 0.73 and the corr. coeff. for the height changes was 0.65. This predicting result seems to be fairly good. The predicted and the observed values are shown in Fig. 3. As may be understood from Fig. 3, the value 0.73 of the corr. coeff. for the heights depend upon the trend caused by the difference of means by years and it appears that the predicting results for the height is not good. The correlation for the height changes is also relatively good, but as may be seen from Fig. 3, in order to verify the predicted results, it seems to be necessary to consider the variance of the height change and the corr. coeff. simultaneously.

4. Features of correlation field on the Northern Hemisphere

As may be seen from Fig. 2, when time lag τ becomes larger than 3 days, no areas of good correlation emerge in this region. And so the correlation fields will be spread over the Northern Hemisphere. Fig. 4 shows the distributions of the correlation coefficients of lag 1–10 days between the 500 mb height at (5, 5) and those at each point along the 50°N. In Fig. 4, (a) is calculated from the data of January 1951, (b) January 1950 and (c) January 1949. As may be seen from this figure, an area which shows a good correlation, moves westwards, as lag τ becomes larger. But this moving velocity and the areas of good correlation are different every year. And it cannot be expected that there is an area of good correlation when the correlations were calculated from the data of the Januaries of many years in the case of a relatively large τ . Therefore the behaviours of the correlation fields must be investigated for every year. Thus the correlation coefficients of lag 1–10 days between the 500 mb height at (5, 5) and that at each point over the Northern Hemisphere, were calculated from the data of January 1951. Fig. 5 shows the distributions of correlation coefficients calculated. The most outstanding feature is that an area in the vicinity of (5, 5), which shows a good correlation when $\tau=0$, gradually moves westwards and in the case of $\tau=10$ day, an area of good correlation emerges in the vicinity of 60°E. It may be owing to the lack of network stations on the China continent that in the case of lag $\tau=3, 4, 5$, and 6 days, this area of positive correlation is not clear. Next in the case of $\tau=0$, there is an area of negative correlation in the vicinity of the 100°E, and in the case of $\tau=5$ this area emerges at about 60°E and in the case of $\tau=10$ day it reaches to the 35°E. As may be seen from the above results, the main areas of good correlation moves

January 1951 (a)

Long. Lat.	0	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180
0	08	-00	15	50	68	60	33	-02	25	54	62	53	-24	27	97	72	37	-08	-32
1	-13	-19	34	64	71	41	02	-30	-40	-60	-64	-34	09	57	63	44	14	-14	-32
2	-34	-26	41	67	52	11	-28	-38	-43	-54	-41	01	23	39	22	27	06	-24	-39
3	-55	-28	47	55	25	-14	-41	-47	-50	-47	-13	23	20	05	03	12	-12	-35	-37
4	-58	-13	52	37	07	-33	-59	-63	-48	-27	03	23	15	-12	-19	-12	-31	-38	-28
5	-42	01	43	18	-14	-45	-66	-64	-28	06	26	25	13	-13	-37	-28	-43	-36	-22
6	-35	-03	13	-08	-39	-58	-57	-24	14	30	26	13	-02	-36	-49	-27	-29	-26	-16
7	-21	-05	02	-31	-58	-67	-29	26	42	33	20	-02	-16	-46	-41	-21	-14	-18	-08
8	01	15	-11	-48	-56	-45	14	56	44	26	-03	-31	-41	-49	-13	-04	-07	-04	05
9	22	31	-14	-48	-47	-02	48	56	22	10	-12	-40	-47	-29	-00	-10	-04	00	00
10	59	57	-12	-36	-16	39	61	43	03	-08	-21	-31	-28	-10	10	-01	12	-13	-20

Fig. 4. Cross correlations of lag 1—10 days between the 500 mb heights at about 50°N 140°E and at each point along 50°N. (a) was calculated for January 1951, (b) January 1950, (c) January 1949 and (d) obtained from the linear theory. Above tables show the values of the correlation coefficients ($\times 10^{-2}$).

Fig. 5. Correlation fields over the Northern Hemisphere. That is, the distributions of cross correlations of lag τ of 1—10 days between the 500 mb height at the point shown by a black spot and the heights at many points over the hemisphere. Full lines show the areas of positive correlation and dashed lines the areas of negative correlation.

westwards at the rate of about $8^\circ/\text{day}$. The area of positive correlation at about 40°E in the case of $\tau=0$, slowly moves westwards and on the west half area of the Northern Hemisphere there is no clear feature. And it may be seen that the influence area to the 500 mb height pattern of the vicinity of Japan for ten days seems to be the east half area of the Northern Hemisphere.

It is difficult to obtain an area in which there is a stable correlation through many years when time lag is relatively large, and so extended prediction can not be performed by the same method as the 24hr prediction mentioned in Section 3.

5. Statistical properties of dynamically linear model

It is very interesting to explain the various situations of the correlation fields mentioned already by the dynamical or thermodynamical method, but we cannot induce the general explicit form of correlation function from the nonlinear equations of motion, hence we shall induce the correlation function from the linearized vorticity equation as follows:

The linearized form of vorticity equation is (assuming the motion to be uniform)

$$(3) \quad \frac{\partial}{\partial t} \left(\frac{\partial^2 Z}{\partial \lambda^2} \right) + A \frac{\partial^3 Z}{\partial \lambda^3} + 2(A + \omega) \cos^2 \varphi \frac{\partial Z}{\partial \lambda} = 0,$$

where

φ = latitude

λ = longitude

Z = isobaric height

ω = angular velocity of earth rotation

A = angular velocity of general current

and the solution of equation (3) is generally written as follows:

$$(4) \quad Z = \sum_{n=0}^{\infty} \{A_n \cos n(\lambda - C_n t) + B_n \sin n(\lambda - C_n t)\},$$

where,

A_n, B_n = Amplitudes of cosine and sine wave of wave number n respectively when $t=0$

$$(5) \quad C_n = A - \frac{2(A + \omega) \cos^2 \varphi}{n^2}.$$

Equation (4) can also be rewritten as follows:

$$(6) \quad Z = \sum_{n=0}^{\infty} \{a_n(t) \cos n\lambda + b_n(t) \sin n\lambda\}$$

where

$$(7) \quad \begin{aligned} a_0(t) &= A_0, & b_0(t) &= 0, \\ a_n(t) &= A_n \cos nC_n t - B_n \sin nC_n t \quad (n=1, 2, \dots), \\ b_n(t) &= A_n \sin nC_n t + B_n \cos nC_n t \quad (n=1, 2, \dots). \end{aligned}$$

Generally, mean, variance and correlation coefficient of $f(t)$ and $g(t)$ in the interval $[-\infty, 0]$, are defined by the next formulas respectively.

$$\begin{aligned}
 (8) \quad \bar{f} &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T}^0 f(t) dt, \\
 \sigma^2(f) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T}^0 \{f(t) - \bar{f}\}^2 dt, \\
 \rho(f, g) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T}^0 \{f(t) - \bar{f}\} \{g(t) - \bar{g}\} dt / \sigma(f) \cdot \sigma(g).
 \end{aligned}$$

First, such statistical quantities for the amplitude in (7) become as follows:

$$(9) \quad \bar{a}_n = \bar{b}_n = 0 \quad (n=1, 2, \dots)$$

$$(10) \quad \sigma^2(a_n) = \sigma^2(b_n) = \frac{A_n^2 + B_n^2}{2} \quad (n=1, 2, \dots)$$

$$\begin{aligned}
 (11) \quad \rho(a_m(t-\tau), a_n(t)) &= \rho(b_m(t-\tau), b_n(t)) = \rho(a_m(t-\tau), b_n(t)) = \rho(b_m(t-\tau), a_n(t)) = 0 \quad (m \neq n) \\
 \rho(a_m(t-\tau), a_m(t)) &= \rho(b_m(t-\tau), b_m(t)) = \cos mC_m\tau \quad (m=1, 2, \dots) \\
 \rho(a_m(t-\tau), b_m(t)) &= \rho(b_m(t-\tau), a_m(t)) = -\sin mC_m\tau \quad (m=1, 2, \dots)
 \end{aligned}$$

Secondly, similar relations for the contour height Z are

$$(12) \quad \bar{Z} = A_0$$

$$(13) \quad \sigma^2(Z) = \frac{1}{2} \sum_{n=1}^{\infty} (A_n^2 + B_n^2),$$

$$(14) \quad \rho(Z(t, \lambda_1), Z(t-\tau, \lambda_2)) = \frac{\sum_{n=1}^{\infty} (A_n^2 + B_n^2) \cos n(\xi + C_n\tau)}{\sum_{n=1}^{\infty} (A_n^2 + B_n^2)},$$

where

λ_1, λ_2 = two arbitrary longitudes

$$\xi = \lambda_2 - \lambda_1,$$

$$A_n^2 + B_n^2 = a_n^2 + b_n^2. \quad (n=1, 2, \dots)$$

From equation (14), $\rho(Z(t, \lambda_1), Z(t-\tau, \lambda_2))$ may be regarded as a function of distance $\xi = \lambda_2 - \lambda_1$ and lag τ , and we shall denote it by $\rho(\xi, \tau)$ hereafter.

Moreover,

$$(15) \quad F(n) = (a_n^2 + b_n^2) / \sum_{n=1}^{\infty} (a_n^2 + b_n^2), \quad (n=1, 2, \dots)$$

is the spectral function for wave number n , and is connected to $\rho(\xi, \tau)$ by the next FOURIER transformation.

$$\begin{aligned}
 (16) \quad F(n) \cos nC_n\tau &= \frac{1}{\pi} \int_0^{2\pi} \rho(\xi, \tau) \cos n\xi d\xi, \quad (n=1, 2, \dots) \\
 F(n) \sin nC_n\tau &= \frac{1}{\pi} \int_0^{2\pi} \rho(\xi, \tau) \sin n\xi d\xi. \quad (n=1, 2, \dots)
 \end{aligned}$$

At last, the correlation coefficients between Z and a_n, b_n become

$$(17) \quad \rho(Z(t, \lambda), a_n(t-\tau)) = \frac{\sqrt{A_n^2 + B_n^2} \cos n(\lambda - C_n\tau)}{\sqrt{\sum_{m=1}^{\infty} (A_m^2 + B_m^2)}},$$

$$\rho(Z(t, \lambda), b_n(t-\tau)) = \frac{\sqrt{A_n^2 + B_n^2} \sin n(\lambda - C_n \tau)}{\sqrt{\sum_{m=1}^{\infty} (A_m^2 + B_m^2)}}, \quad (n=1, 2, \dots)$$

From the results mentioned above, the height Z can be expressed in terms of the above a_n, b_n for $t-\tau$ by the linear regression method as follows:

$$Z(t, \lambda) = A_0 + \sum_{n=1}^m \{a_n(t-\tau) \cos n(\lambda - C_n \tau) + b_n(t-\tau) \sin n(\lambda - C_n \tau)\}.$$

This relation is reduced to (6) when m tends to infinity. In the next section, we shall investigate the possibility that the observed correlation field may be explained by the correlation field calculated from the dynamically linear model.

6. Analysis of the correlation fields at 500 mb surface

First, we compare the equation (14) with the actual cross correlations calculated from the 500 mb height along the 50°N in the Northern Hemisphere.

KUBOTA and IIDA gained the result shown in Table 1, for the mean value of amplitude and the mean phase velocity of wave number n , by Fourier Analysis of the height anomaly of 500 mb surface along 50°N in the Northern Hemisphere from January 1 to February 5, 1949.

Table 1. The mean amplitudes and the mean phase velocities of the travelling waves obtained from the data of 500 mb level of January 1—February 5, 1949.

	$\sqrt{A_n^2 + B_n^2}$	C_n
$n=1$	210 ft	-22 degree/day
2	250	-13
3	250	0
4	240	4
5	140	7
6	170	9
7	110	8

Therefore, we calculated the correlation coefficients of (14) by substituting the results in this table and using

$$\begin{aligned} \lambda_1 &= 140^\circ\text{E}, \\ \lambda_2 &= 0^\circ, 10^\circ, \dots, 360^\circ, \\ \tau &= 0, 2, 4, 6, 8, 10. \end{aligned}$$

Fig. 4(d) shows this correlation field. The main feature of correlation field shown in Fig. 4. (c) can be approximately explained by Fig. 4 (d). As may be seen from (14), the waves in the correlation fields move to the east when the waves in the 500 mb height along 50°N move to the west. The eastward movement of the wave pattern in the correlation field can be clearly seen in each field of Fig. 4, and we shall try to discuss in detail the correlation field for 1951.

We can get the next equation by putting $\tau=0$ in (14)

$$\begin{aligned} (18) \quad \rho(\xi, 0) &= \rho(Z(t, \lambda_1), Z(t, \lambda_2)) \\ &= \sum_{n=1}^{\infty} (A_n^2 + B_n^2) \cos n\xi / \sum_{n=1}^{\infty} (A_n^2 + B_n^2) = \sum_{n=1}^{\infty} F(n) \cos n\xi. \end{aligned}$$

From this equation, we can induce spectral function $F(n)$ by cosine transformation of $\rho(\xi, 0)$, but of course, such relation will not be exactly satisfied for the actual correlation function and spectral function.

Fig. 6. Spectral distribution for each wave number n along 50°N in January 1951.

In order to examine the possibility of the above relations being satisfied, we calculated spectral function $F(n)$ by cosine transformation of the actual correlation field for January 1951, and showed these results in Fig. 6.

And, $\rho(\xi, 0)$ were calculated from (18) by using the above-obtained $F(n)$ ($n=1, 2, \dots, 7$). These $\rho(\xi, 0)$ are drawn with the actual correlation field in Fig. 7. The correlation coefficient between two curves in Fig. 7 is 0.80

and shows a relatively good fitting. But the correlation field calculated from (18) is theoretically symmetric with regard to $\xi=0$ and the actual curve is not.

Fig. 7. Cross correlations $\rho(\xi, 0)$ along 50°N in January 1951. Full line shows the calculated cross correlations $\rho(\xi, 0)$ gained by composing waves from wave number 1 to 7. Dotted line shows the actual cross correlations $\rho(\xi, 0)$.

Next, when $\tau=0$, the phase velocities C_n ($n=1, 2, \dots$) disappear from the formulas (16) and we cannot discuss them. Using the formulas (16) in the case when time lag $\tau=1, 5$ and 10, we calculated the amplitudes and the phase velocities C_n for each wave number $n=1, 2, 3, \dots, 7$ from the correlation fields for January 1951. Fig. 8 (a) and (b) show these amplitudes and phase velocities respectively. When the wave number n was fixed, the same amplitude and phase velocity should be gained for any value of time lag τ as the linearity mentioned already is satisfied for actual data. But as may be seen from Fig. 8, they are different according to the various values of τ , and this result seems to be caused by the assumption of linearity.

Using the amplitudes and the phase velocities obtained above when $\tau=1$, and 5 days, we calculated the correlation functions of (14) and compared with the actual correlation field. Fig. 9 (a) and (b) show these two kinds of correlation fields when $\tau=1$ and 5 days.

Fig. 8. Amplitudes $\sqrt{A_n^2 + B_n^2}$ and phase velocities C_n for each wave number n ($n=1, 2, \dots, 7$) in January 1951.

Fig. 9. Calculated cross correlations from the amplitudes and phase velocities of wave number $n=1 \sim 7$ (Full lines) and actual cross correlations (Dotted lines) for time lag $\tau=1$ and 5 days.

Lastly, we refer to the variations of the correlation coefficients with the time lags.

Correlograms of the 500 mb height at 50°N 140°E for the January of 1949, 1950 and 1951 when $\tau=1, 2, \dots, 10$ days, are shown in Fig. 10. An the mean curve of the correlograms for the above three years,

$$(19) \quad \rho_\tau = e^{-0.74\tau} \cos \frac{\pi}{6} \tau,$$

was fitted by the method of least square.

Correlogram calculated from (14) using the mean amplitudes and phase velocities obtained from the data of January 1949, is also drawn in Fig. 10. This curve is a little different from (19). Moreover, cross correlation coefficients between the 500 mb

Fig. 10. Correlograms of height Z at 50°N 140°E in 1949, 1950 and 1951. Full line shows the estimated correlogram from the mean amplitudes and phase velocities in January 1949. Dashed line shows that of $\rho_{\tau} = e^{-0.74\tau} \cdot \cos \frac{\pi}{6} \tau$.

height at 50°N 140°E and that at each point along 50°N when $\tau = 0, 1, 2, \dots, 10$ are shown in Fig. 11. We can also see that both ridge and trough of the correlation pattern move gradually to the east.

From the above analysis, the limit that the features of correlation fields may be explained by the dynamically linear theory, was ascertained. Since it is difficult to obtain the correlation function from the nonlinear equation, some consideration on the relationship between the correlation field and the nonlinear model will be given in the following section.

7. Relationship between the correlation field at 500 mb level and the nonlinear model

The outstanding feature of the linear pattern to be given by the equation (4) is that both amplitudes and phase velocities are independent of time. Making use of the coefficients obtained by harmonic analysis of daily 500 mb height of 70°E to 180°E along 50°N , for the data of January 1951, the variation with time of the mean values of the height in the interval and the amplitudes of wave numbers 1–3, are shown in Fig. 12. The waves of wave numbers 1, 2 and 3 have the wave lengths of

Fig. 11. Cross correlation patterns between the height at 50°N , 140°E and that of each point along 50°N .

Dashed lines show the eastward movements of the trough and ridge in such patterns.

120°, 60° and 40° by longitude respectively. Then monthly means and standard deviations are shown in Table 2.

It is recognized in this table that the monthly mean and variation of the amplitude

of wave number 1 are far larger than the remaining wave numbers. The situation of the 500 mb height level of 70°E to 180°E along 50°N is almost interpreted by the behaviour of the wave of wave number 1, and for instance there is a remarkable relationship between the variation of the 500 mb contour height at 50°N, 140°E and the variation of the amplitude of wave number 1, and the correlation coefficient between them is 0.75. This corresponds with the existence of a stationary cyclonic pattern at 500 mb height level in the neighbourhood of Japan in the winter season.

As may be seen from Fig. 12, the outstanding feature in variations of the amplitudes seems to be the variation of wave number 1. Then the distributions of the mean phase velocities of wave number 1, 2 are shown in Table 3.

Table 2. The monthly mean amplitudes and the standard deviations for the wave by harmonic analysis of the 500 mb height from 70°E to 180°E along 50°N of January 1951.

Wave number	1	2	3	4	5
Monthly mean (ft)	550	188	131	88	70
Standard deviation (ft)	252	93	64	52	47

Fig. 12. The variation with time of the mean value of the height and the amplitudes of wave number, 1, 2 and 3, calculated by harmonic analysis of the 500 mb height from 70°E to 180°E along 50°N of January, 1951.

Table 3. The distribution of the daily mean phase velocities of the wave number 1 and 2.

W. N.	class(deg/day)	<-3	-3~-2	-2~-1	-1~0	0~1	1~2	2~3	3<	Mean(deg./day)
1	freq.	2	1	1	2	16	3	2	2	1 ₀
	class(deg/day)	-15~-10 -10~-5 -5~0 0~5 5~10 10<								
2	freq.	1	0	3	10	7	4			8°

Fig. 13. Cross correlation for the field of the residual height. The solid line shows the real cross correlation for the above field, $\rho\{Z'(t, 70^\circ\text{E}), Z'(t+\tau, \lambda_j)\}$, for time lag $\tau=0, 1, \dots, 5$, $\lambda_j=80^\circ\text{E}, \dots, 180^\circ\text{E}$. The thin line shows the cross correlation obtained by the linear theory.

On a glance at this table, it is found that wave number 1 is almost stationary and monthly mean velocity of wave number 2 is about 8 deg./day eastwards. As mentioned above, the variation of the amplitude of wave number 1 with time is much larger than the remaining waves, and the behaviour of the wave of wave number 1 can not be considered to be linear and the non-linear effect of the wave of wave number 1 for the height is larger than the remaining waves. Therefore the new height distribution in the interval of 70°E to 180°E along 50°N was made by subtracting the wave of wave number 1 from the observed 500 mb height. The new height distribution is called the field of the residual height. The field of the residual height Z' is represented as follows:

$$Z'(t, \lambda) = Z(t, \lambda) - a_1(t) \cos \lambda - b_1(t) \sin \lambda$$

where $Z(t, \lambda)$ shows 500 mb height and $a_1(t) \cos \lambda + b_1(t) \sin \lambda$ is the wave of wave number 1.

The cross correlation coefficients $\rho(Z'(t, 70^\circ\text{E}), Z'(t+\tau, \lambda_j))$ ($\lambda_j=80^\circ\text{E}, 90^\circ\text{E}, \dots, 180^\circ\text{E}$) are shown by the solid line in Fig. 13. The variation of the wave of wave number 2, the phase velocity of which has about 8 deg./day eastwards, dominates in about 50% of the variation for the field of the residual height. And so applying the linear model mentioned in the previous section for this field, the correlation function was calculated from the following equation:

$$\rho\{Z'(t, 70^\circ\text{E}), Z'(t+\tau, \lambda_j)\} = \frac{A_2^2 + B_2^2}{\sum_{n=2}^6 A_n^2 + B_n^2} \cos 2(\xi + C_n\tau),$$

where $C_n: 8 \text{ deg./day}$, $A_n^2 + B_n^2$: monthly mean square value of amplitude.

The result obtained is shown by the thin line in Fig. 13. Comparing with the real correlation field for the field of the residual height and the calculated correlation field, it was found that there is a good correspondence for the time lag $\tau=2$, 3 and 4 days especially. In consequence, it may be considered that the linear theory is pertinent to the field of the residual height,

As mentioned above, the nonlinear features of the height pattern may be understood as the time changes of the amplitude and the phase velocity. We shall, in future, investigate into their relations to the synoptic patterns and the physical quantities.

Acknowledgements—The writers wish to express their hearty thanks to Dr. H. ARAKAWA for his encouragement. They also wish to express their hearty thanks to the staff members of the Forecast Research Laboratory, Meteorological Research Institute in Tokyo for their kind help, and to Mr. N. HONDA and Miss. T. INOUE for their earnest assistance in numerical computation and figure drawing.

References

- KUBOTA, S. and M. IIDA, 1954: Statistical characteristics of the atmospheric disturbances. *Pap. Met. Geophys.*, 5, 22-34.
- LORENZ, E.N., 1956: Empirical orthogonal functions and statistical weather prediction. Scientific Report No. 1, statistical forecasting project. Contract No. AF 19 (604) 1566.
- MALONE, T.F., 1955: Application of statistical methods in weather prediction. *Proc. Nat. Acad. Sci.*, 41, 806-815.
- MALONE, T.F., 1956: Application of synoptic climatology to weather prediction. Chapter 28. PETTERSEN, S: *Weather analysis and Forecasting*. 2nd Ed. Vol. II. New York, Mc Graw-Hill Book Co., 238-255.
- WADSWORTH, G.P., and J.G. BRYAN, 1948: Short range and extended forecasting by statistical methods. *Air. Wea. Serv. Tech. Rep. No.* 105-37, 202.
- WADSWORTH, G.P., 1951: Application of statistical methods to weather forecasting. *Compendium of Meteorology*, Boston, Amer. Meteor. Soc., 849-855.
- WHITE, R.M., and W.C. PALSON, 1955: On the forecasting possibilities of empirical influence functions. *J. Meteor.*, 12, 478-485.
- WHITE, R.M., R.C. DERBY, D.S. COOLEY, and F.A. SEAVER, 1957: Hemispherical prediction by statistical techniques, *J. Meteor.*, 14, 448-457.

500 mb 面の相関場と統計的予報法について

曲田光夫, 小沢 正, 鈴木栄一, 藤田敏夫, 戸松喜一

日本附近の 500 mb 等圧面高度分布についての統計的予報の研究の第一段階として, その可能性について検討した。1 日乃至 10 日の時間のずれを考慮した種々の相互相関係数によつて得られた相関場の特性およびその相関場に対して力学的に線型な系からもたらされる統計的性質との関係等について調べ, 又 1 日, 3 日予報について簡単な統計的予報法を試みた。

まず 1948 年から 1952 年までの 5 カ年の 1 月の資料を用い, 東経 140° , 北緯 50° の地点の 500 mb 等圧面高度とその周囲 1200 km 間隔の 25 格子点 (範囲: 東経 70° — 180° , 北緯 30° — 65°) の 500 mb 等圧面高度との同時相関及び 1 日～5 日ずれの相互相関係数を計算した。これらは各年ごと及び 4 年間通した相関係数の 2 種類について求めた。得られた相関場の特性を吟味し, さらに統計的予報法の簡単な試みとして, 東経 140° , 北緯 50° の地点の 500 mb 等圧面高度に対して各年共通して比較的高く安定した相関係数を示す数地点の 500 mb 等圧面高度をその予報要因として採用し, 線型回帰法により 1 日, 3 日予報を実施し, 独立な資料について予報の検証をも行つた。予報精度をさらに向上させるには各年の相関場の特性を把握し, それらを予報法のなかに導入する必要性が明かにされた。つぎに上述の相関場の範囲を拡張し, 1951 年 1 月について東経 140° 北緯 50° の地点の 500 mb 高度と同緯度圏上の経度 10° 間隔の 36 地点の各 500 mb 高度との時間差 1 日～10 日ずれの相互相関係数が計算され, それらの相関場は特徴的な分布および移動を示している。同時にそれらの相関場に対し, 力学的に線型な系から導かれた理論的相関場との対応が検討され, 力学的に線型のモデルが平均的にかなりよい近似になつていることが示された。又スペクトル分解により相関場における波の特性の検討がなされた。なお力学的に非線型な系についての若干の考察をおこなつた。