

T H E U N I V E R S I T Y O F M I C H I G A N

COLLEGE OF ENGINEERING

Department of Meteorology and Oceanography

Technical Report

DEPICTING STOCHASTIC DYNAMIC FORECASTS

Edward S. Epstein

Rex J. Fleming\*

ORA Project 03743

supported by:

NATIONAL SCIENCE FOUNDATION

GRANT NO. GA-19248

WASHINGTON, D.C.

administered through:

OFFICE OF RESEARCH ADMINISTRATION

ANN ARBOR

November 1970

---

\*Captain, United States Air Force

## DEPICTING STOCHASTIC DYNAMIC FORECASTS

by E.S. Epstein & R.J. Fleming

### ERRATA

Page 4, line 3 from bottom: Read " $\psi$ " for "x"

Page 17: last line on page should conclude:  
"...the small ellipse labeled " $\Psi|T$ ".

Page 21: The following sentences should immediately follow the sentence ending on line 10.

"Any position for the wave is credible, and an amplitude very near zero is a distinct possibility. Nevertheless, even here the temperature and stream-function waves are so closely correlated that one can be sure that the waves, if they have any sizable amplitude at all, will be very nearly in phase and will still give rise to negligible heat flux. In that sense, at least, they are not entirely unpredictable."

Engn  
UMR  
1370

This report was prepared for presentation at the International Conference on Meteorology, Tel Aviv, November 30 - December 4, 1970, sponsored by the American Meteorological Society and the Israel Meteorological Society.

In a recent study Epstein (1969) proposed a method of prediction that recognizes the inevitability of uncertainty in initial conditions, and produces forecasts that contain explicit statements of uncertainty in the predicted state. The amount of information contained in a complete statement of the uncertainty of a complicated field is extremely large. Although the numerical representation of variances and covariances by a square symmetrical matrix is straightforward, the information it contains is not readily comprehended. The graphical representation of this information is far more complex but can also be much more meaningful. It is the purpose of this paper to present a number of forms of graphical results of stochastic predictions, both to demonstrate how information on uncertainty can be represented and also to illustrate the value of specific information of uncertainty.

For our illustrative purposes we will consider stochastic predictions of a model in which the mean flow in an infinite channel is represented by a stream function and the temperature field by departures from a mean value. These two fields are generally represented graphically as in Figure 1. Boundary conditions are that the flow is periodic in the x-direction and there is no flow across the walls at the northern and southern boundaries. These are evident in the figure.

The particular fields shown here are represented exactly by 28 parameters, 14 for the stream function and 14 for the temperatures. (There are two modes in the y-direction and wave-numbers 0, 1, 2, and 3 in the x-direction. Two parameters are required to represent each wave/mode combination, except for wave 0 which requires but one term.) Indeed these fields are a particular

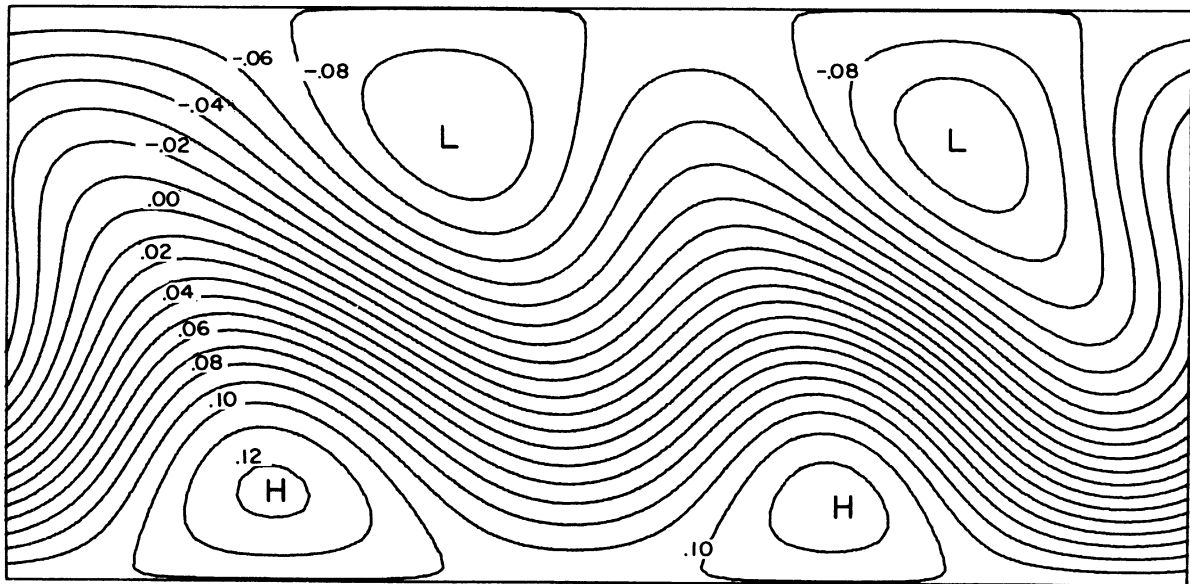
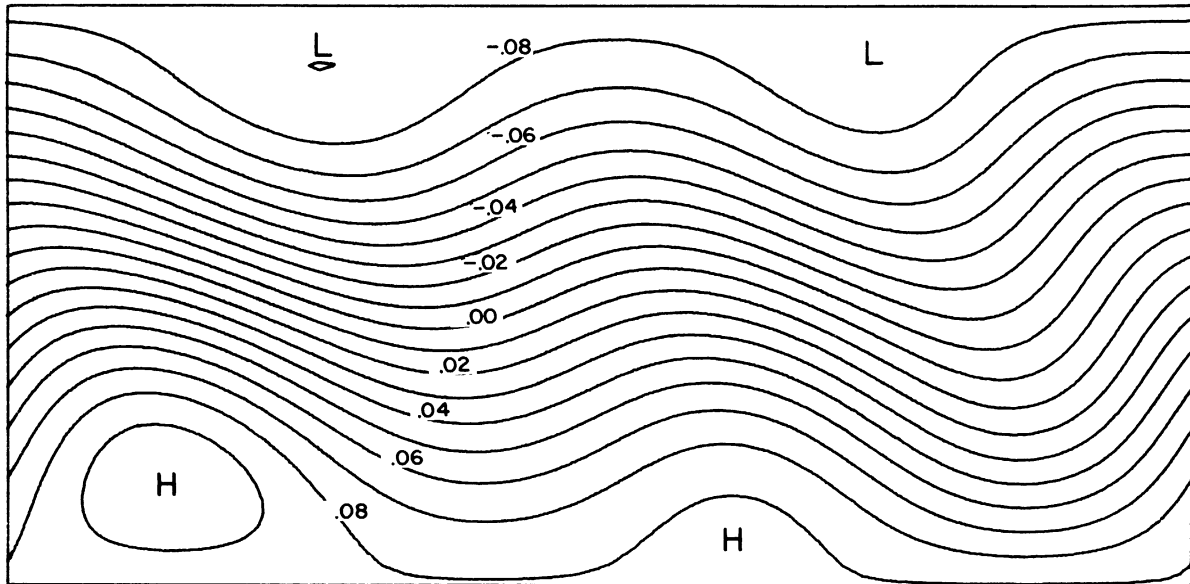


Figure 1. Stream function for the mean flow (lower chart) and for the shear flow (upper chart). These two fields define the state of the two-level model. The stream lines for the shear flow are equivalent to mean isotherms. Units are nondimensional.

realization of a two-level quasi-geostrophic model which we have been using to study various aspects of stochastic dynamic prediction and stochastic analysis. The model includes crude forms of diabatic and frictional effects.

In a deterministic analysis or forecast, the state of the model atmosphere would be represented by a vector of 28 terms. In stochastic procedures one must deal not only with the 28 expected values—or means—of the parameters, but also with their variances and covariances—a total of 434 terms in all. It requires 28 terms to produce the maps in Figure 1. There is no single graphical representation (in two or three dimensions) that could illustrate all the information of the 434 terms. There are several representations, however, that would be very useful and meaningful to the meteorologist. The illustrations that follow will refer specifically to a stochastic dynamic prediction that is to be verified against the map in Figure 1. We will be illustrating the kinds of statements about the atmosphere that stochastic predictions allow. In this particular case the prediction is based on simulated observations, containing random errors, made 24 hours earlier at an array of 30 stations that was also chosen at random. The standard errors of the simulated observations were .003 in the units in which the stream function and temperature are given in Figure 1. If we relate the error in the "observation" of the stream function to an error of 12m in the measurement of the height of a constant pressure surface, then the total range of "height" on these maps would be about 800m. This implies further that the "error" in the temperature observations is about 1°K and the range of temperatures on Figure 1 is about 50°K. In all figures we use the original nondimensional units, but as very crude rules of thumb, one might multiply the values of the stream functions by 4000 to get height

differences in meters, and multiply the nondimensional temperatures by 300 to get temperature differences in degree Kelvin. The total dimensions of the region shown in the figures is correspondingly approximately 6300 x 12600 km.

The predicted expected values of the stream function and temperature field are shown in Figure 2. These are represented algebraically by

$$E[\psi(x,y)] = \sum_{i=1}^{14} B_i F_i(x,y)$$

$$E[T(x,y)] = \sum_{i=1}^{14} B_{i+14} F_i(x,y)$$

where the  $B_i$  are the (predicted) expected values of the parameters with the convention that the first 14 terms represent the stream field and the second set of 14 terms refer to the temperatures. The  $F_i$  are a set of functions that satisfy our boundary conditions and are orthogonal over the region.

Since stochastic results include the variances and covariances of the  $B_i$ , it is relatively easy to calculate the variances (or standard deviations) of linear combinations of the  $B_i$ . Thus

$$\text{var}[\psi(x,y)] = \sum_{i=1}^{14} \sum_{j=1}^{14} \text{cov}(B_i, B_j) F_i(x,y) F_j(x,y)$$

and similarly for the variance of the temperature field. Isopleths of the standard deviations of  $\psi$  and  $T$  are shown in Figure 3.

Note that the maximum standard error of estimate is about equal to the interval between contours on the mean charts and is several times the standard

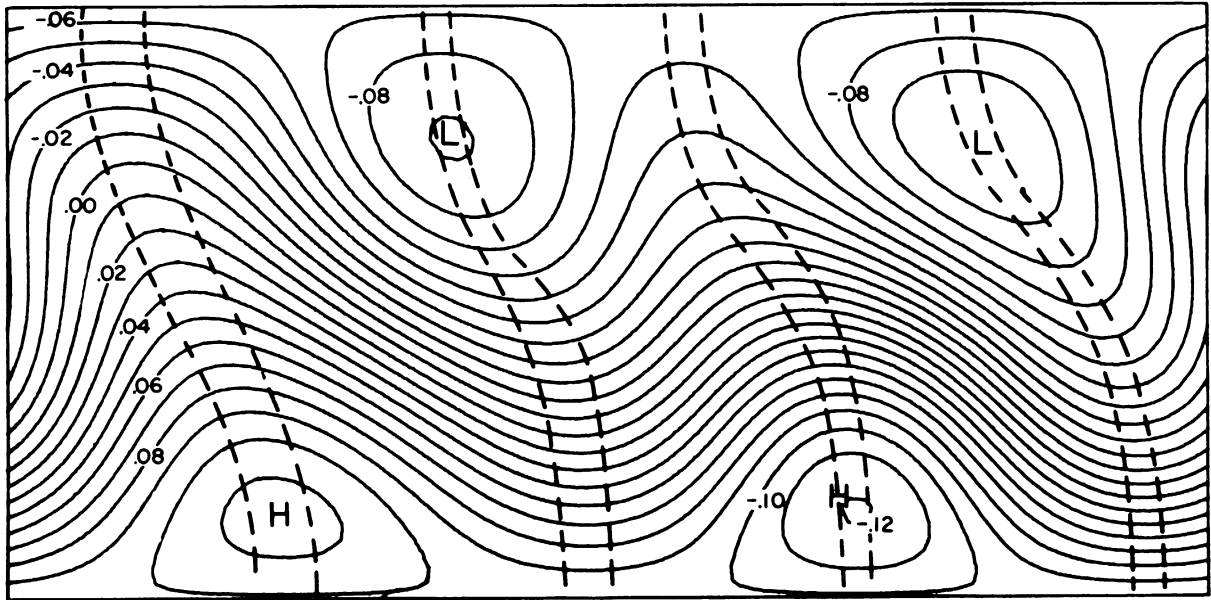
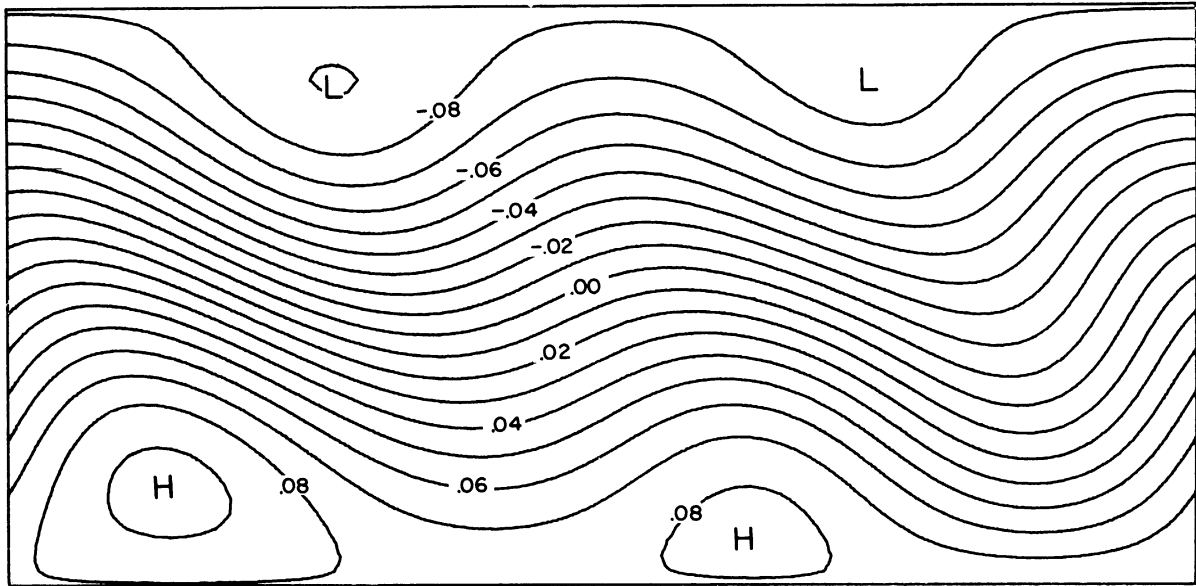


Figure 2. Stochastic dynamic prediction of the mean stream function (lower chart) and temperature (upper chart) fields. The dashed lines represent uncertainty in the positions of the ridges and troughs and are explained later in the text.



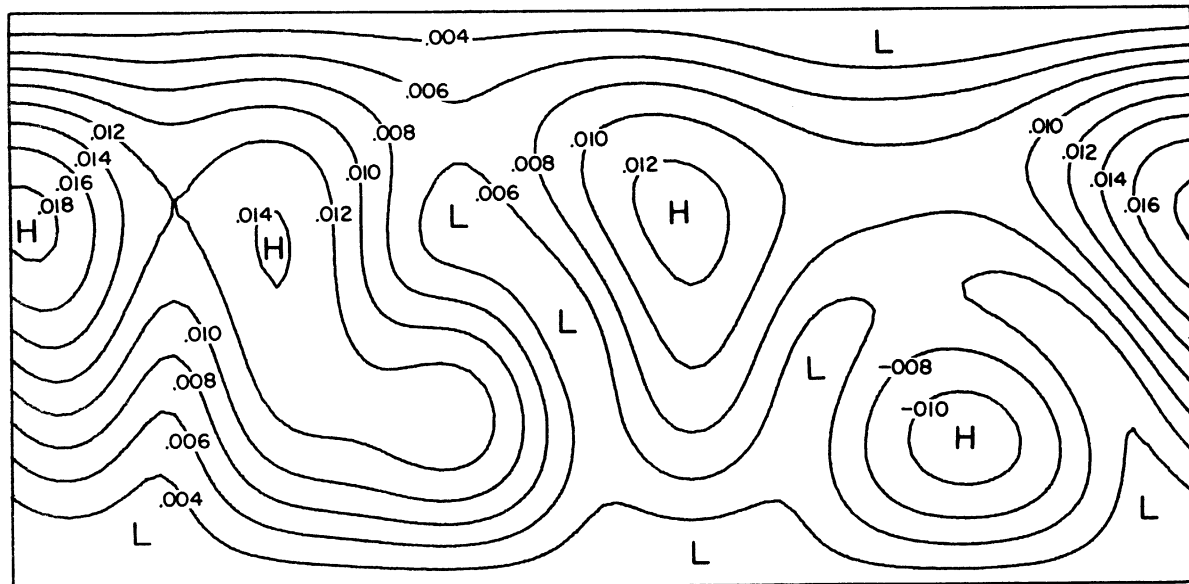
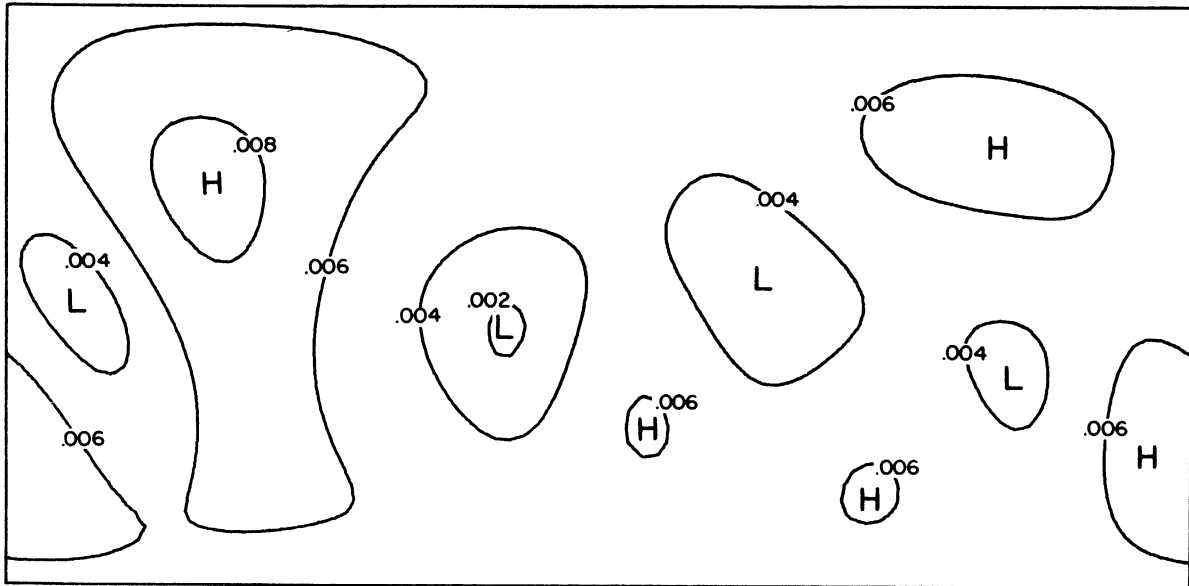


Figure 3. Fields of standard deviation of the forecast temperatures (upper chart) and stream function (lower chart).

deviation of a single observation. One can expect that errors will usually be less than one standard deviation, but occasionally errors as large as two or three standard deviations may occur. To some degree the maximum uncertainty seems to be where the gradients are largest, but this is not the entire picture at all. The origin of the uncertainty lies in paucity of observations and the uncertainty of the measurements. The patterns of uncertainty reflect to a large degree the locations of the observations. The amount of uncertainty, however, has grown considerably since the observations were made. Initially the standard deviation was less than .003 almost everywhere, and necessarily at all observation points, but the standard deviation of the prediction is almost everywhere greater than .003.

Figure 4 gives the actual errors of the expected values, the differences between the maps of Figures 1 and 2. These patterns depend explicitly on values actually observed, so that the patterns do not particularly resemble those in Figure 3. Still the magnitude of the differences are within the bounds expected on the basis of the calculated standard deviations. Note that the patterns of error in the two fields are similar, reflecting the strong correlation between the mean flow and shear flow that one would expect from the physical model. This is observed in spite of the fact that the errors in the initial temperature and stream function observations were independent.

It can also be added that the errors in Figure 4 are generally smaller than they would have been if we had made a deterministic rather than a stochastic forecast based on the actual observations. Stochastic prediction tends to minimize the root mean square error.

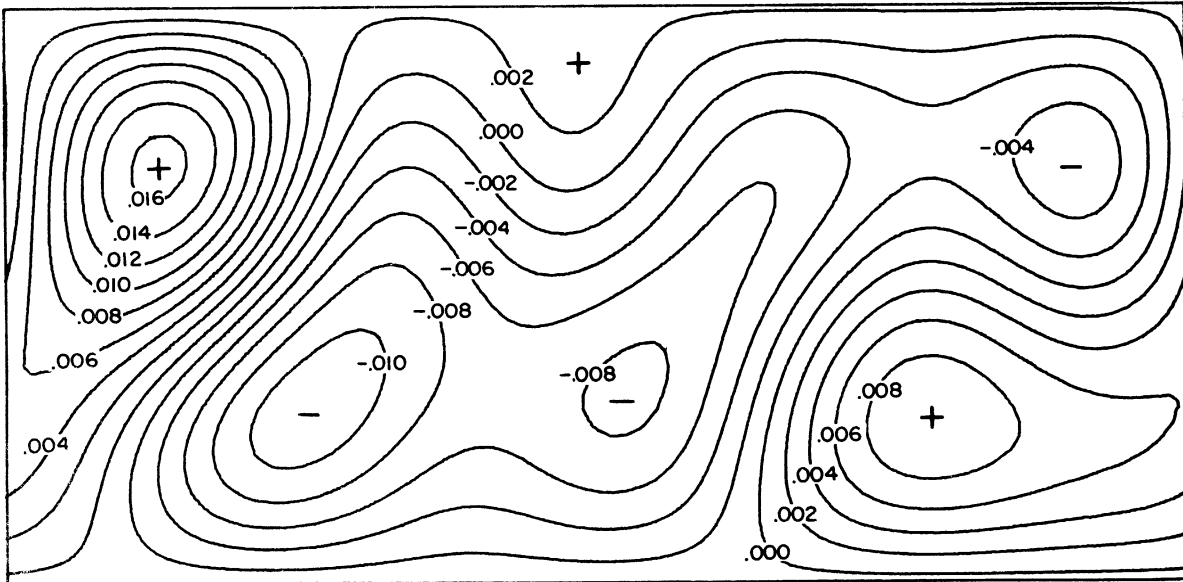
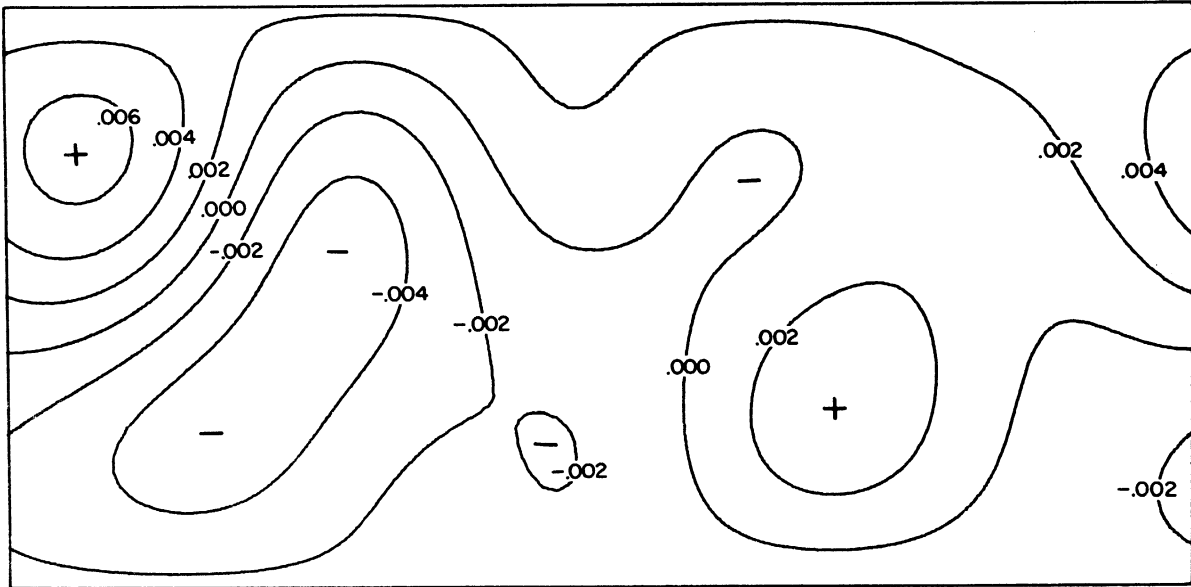


Figure 4. Departures of the forecast expected values of the temperatures (upper chart) and stream function (lower chart) from their true values. These charts are differences between those of Figures 1 and 2.

Not only are there strong correlations between the mean and shear flows, but within each field there will be strong correlations from place-to-place of the stream function or of the temperature. Figure 5 shows this. It is the auto-correlation of the  $\psi$ -field with the value of  $\psi$  at an arbitrarily chosen point. The large values of the correlation coefficient that occur at considerable distance from the chosen point are worth noting. They imply that an observation of  $\psi$  at one point will give considerable information about the field elsewhere, not so much because it is a directly relevant measure, but because it would say which, of many possible states of the model, are the most reasonable.

The wind field is one that is frequently of interest to the meteorologist. Especially he tends to be concerned with the meridional flow. It is simple to derive the v-component of the wind from the stream function as

$$E[v(x,y)] = E\left(\frac{\partial\psi}{\partial x}\right) = \sum_{i=1}^{14} B_i \frac{\partial F_i}{\partial x}$$

$$\text{and } \text{var}(v) = \sum_{i=1}^{14} \sum_{j=1}^{14} \text{cov}(B_i B_j) \frac{\partial F_i}{\partial x} \frac{\partial F_j}{\partial x}$$

These fields, again as predicted, are given in Figure 6. A consistent scaling of the nondimensional units would imply that .01 in these units corresponds to a wind in the vicinity of  $2 \text{ m sec}^{-1}$ . Note that the greatest uncertainty (a standard deviation of about .07) does not correspond to a region of maximum meridional wind, but occurs in the vicinity of a ridge where the v-component of the wind is small. This implies that placement of that ridge, and the positions of the other ridges and troughs, are uncertain—but to a measurable extent. The dashed lines in Figure 2 are lines along which  $E(v)/[\text{var}(v)]^{1/2} = \pm 1$ .

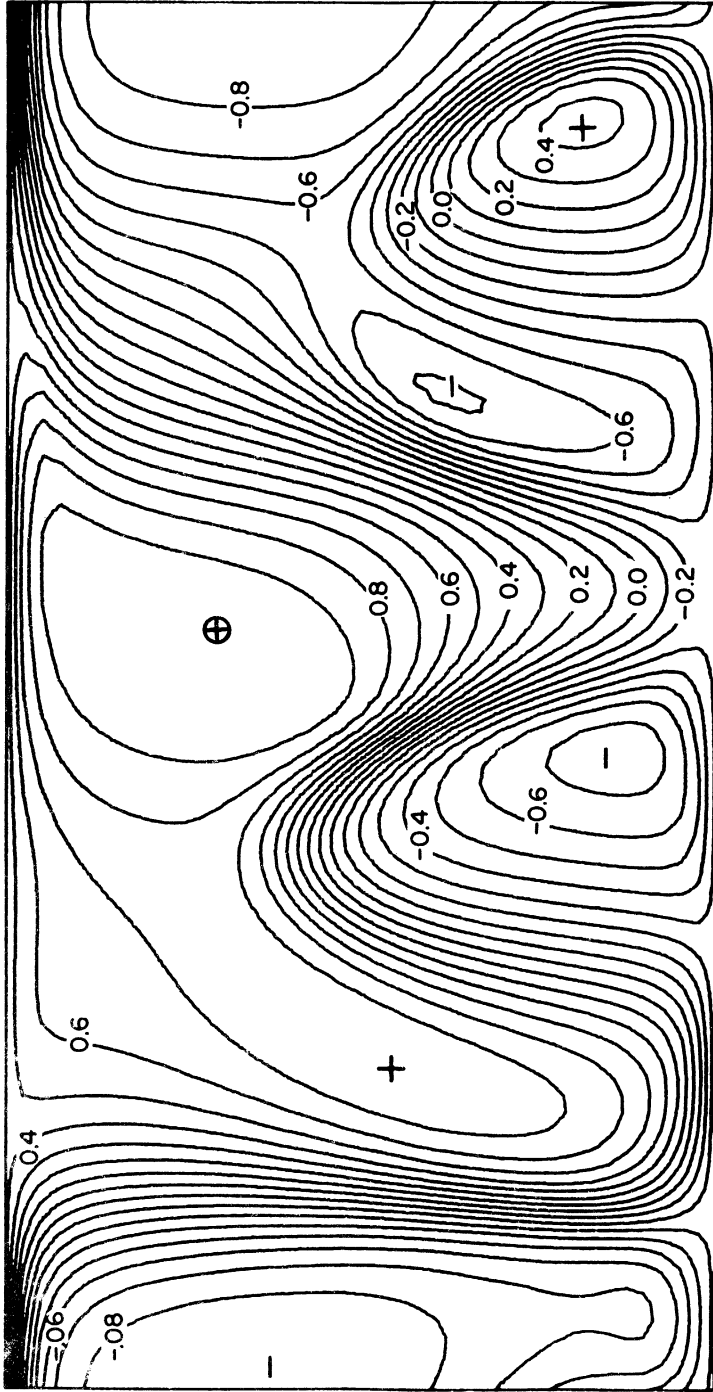


Figure 5. Autocorrelation coefficient of the forecast stream function and its value at the indicated (arbitrarily chosen) point.

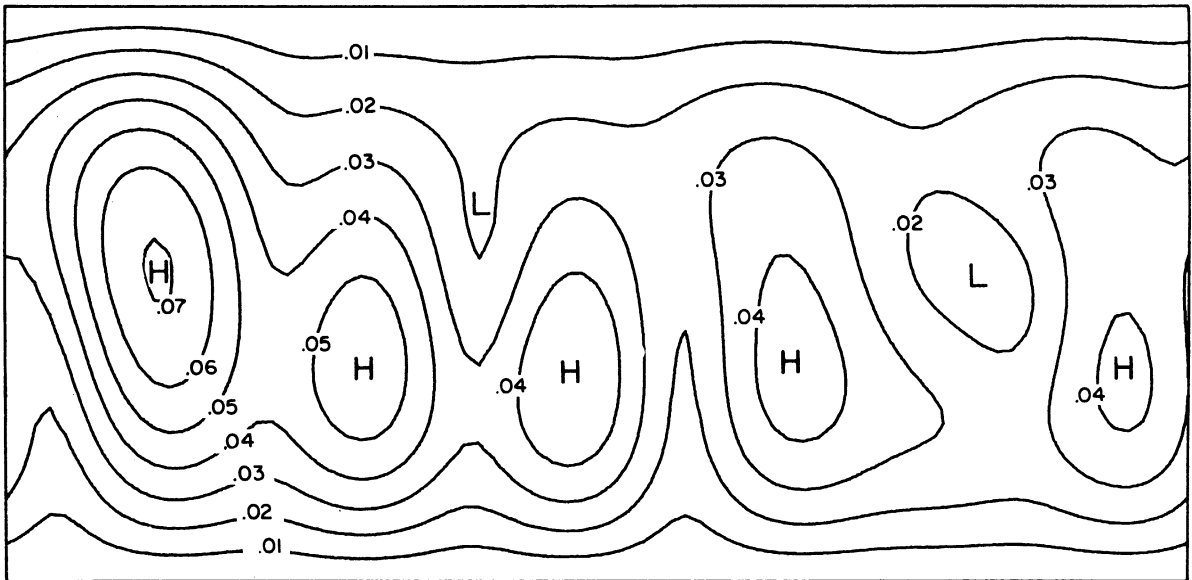
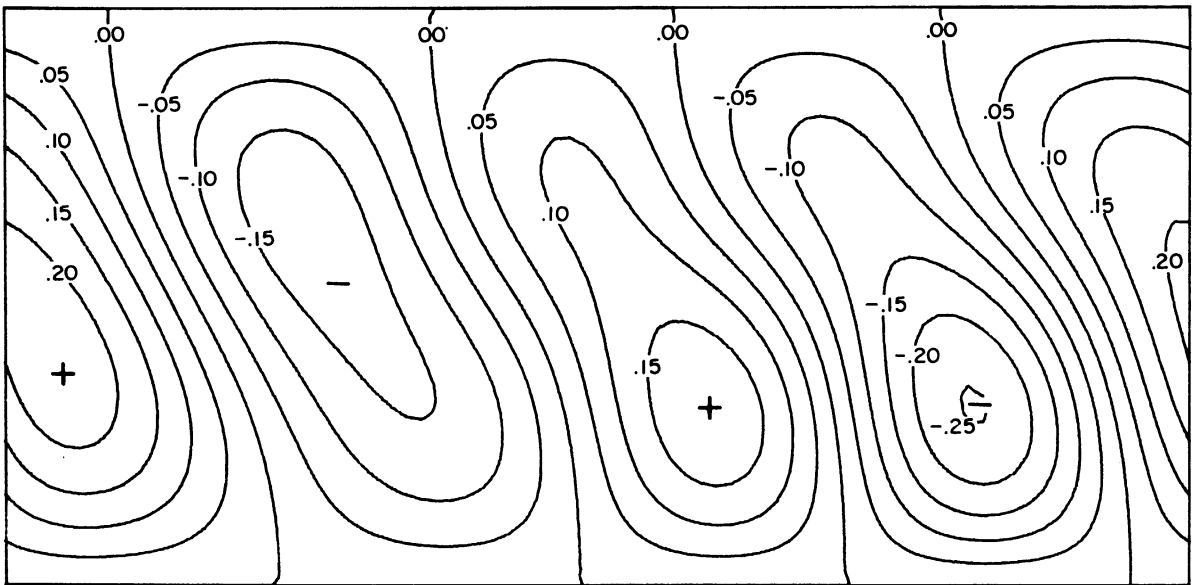


Figure 6. The forecast expected meridional wind speed (upper chart) and its standard deviation (lower chart).

It is more likely than not that the ridge and trough lines will lie between the pairs of dashed lines.

Winds are more readily represented by vectors than by their scalar components. Figure 7 is a view of the forecast wind field. A wind vector of length equal to the distance between grid points corresponds to a wind speed of about  $75 \text{ m sec}^{-1}$ . The ellipses centered on the endpoints of the vectors toward which the winds are blowing require some explanation.

Our stochastic procedures provide forecasts of first and second moments, but not of the complete joint probability distribution. It is known, for example, that this joint probability distribution will not in general be multivariate normal. Nevertheless, for purposes of illustration, it is useful to represent the joint distribution as though it were multivariate normal with the given first and second moments. It is a property of multivariate normal distributions that the marginal distributions of linear combinations of the parameters are also distributed multivariate normal. In particular, since  $u$  and  $v$  are both such linear combinations, it is possible to treat the joint distribution of the wind components at each point as though they were bivariate normal. This allows us to construct "credible" ellipses, such that there is a specific joint probability of the true values of the wind components lying within the ellipses. In Figure 7 each ellipse includes a 50% credible region. The probability that the end-point of each vector lies within its ellipse is 50%. Note that if we knew that the wind at one particular grid point lay outside its ellipse, then we would have to revise our statements about the wind at all other grid points, since they are all, in general, correlated to one another. Figure 7 takes into consideration the correlation of the  $u$  and  $v$  wind components

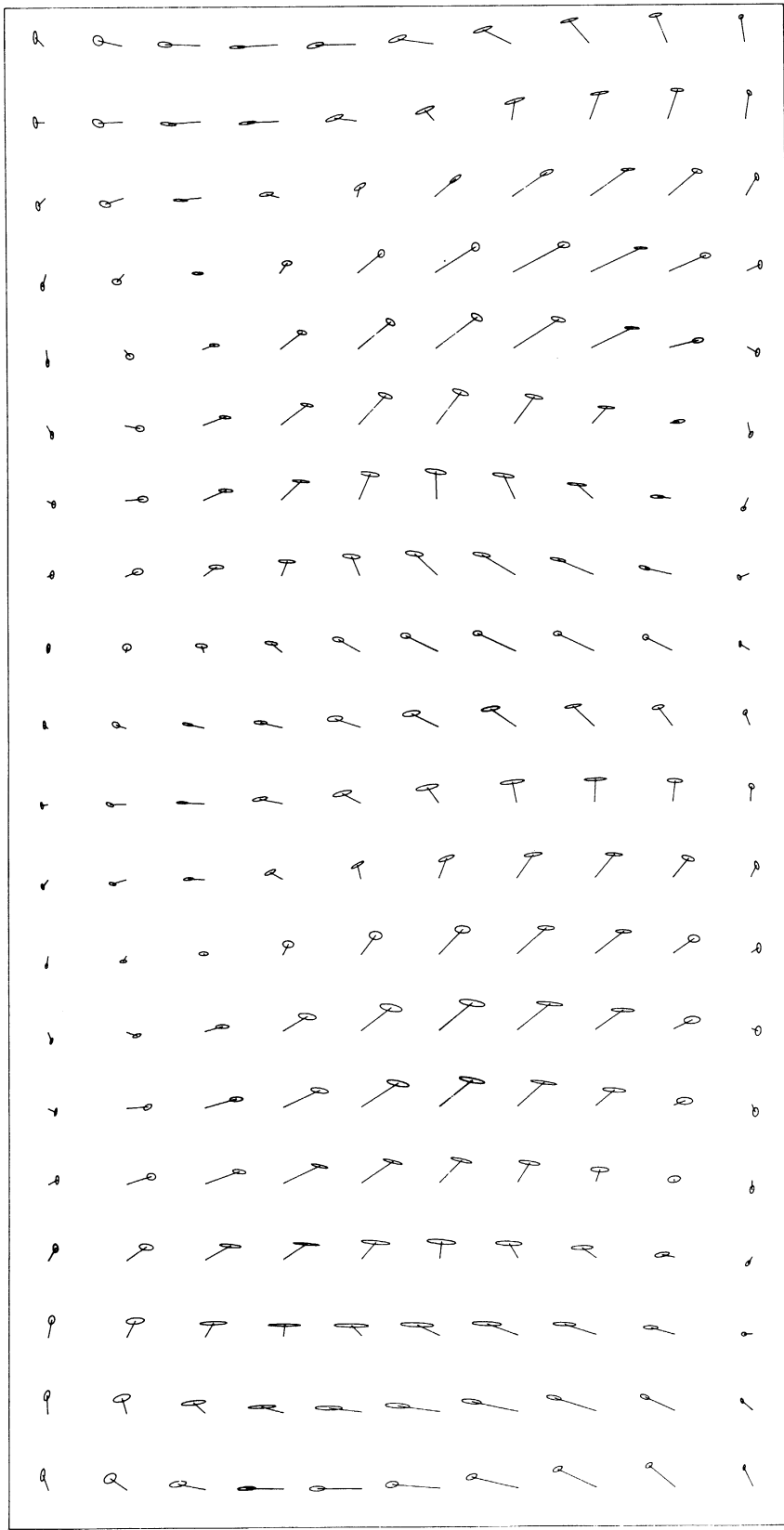


Figure 7. The forecast wind field, indicating both the expected winds and their uncertainty. There is a 50% probability that the true wind vectors terminate within the ellipses drawn around the endpoint of the expected wind vectors.

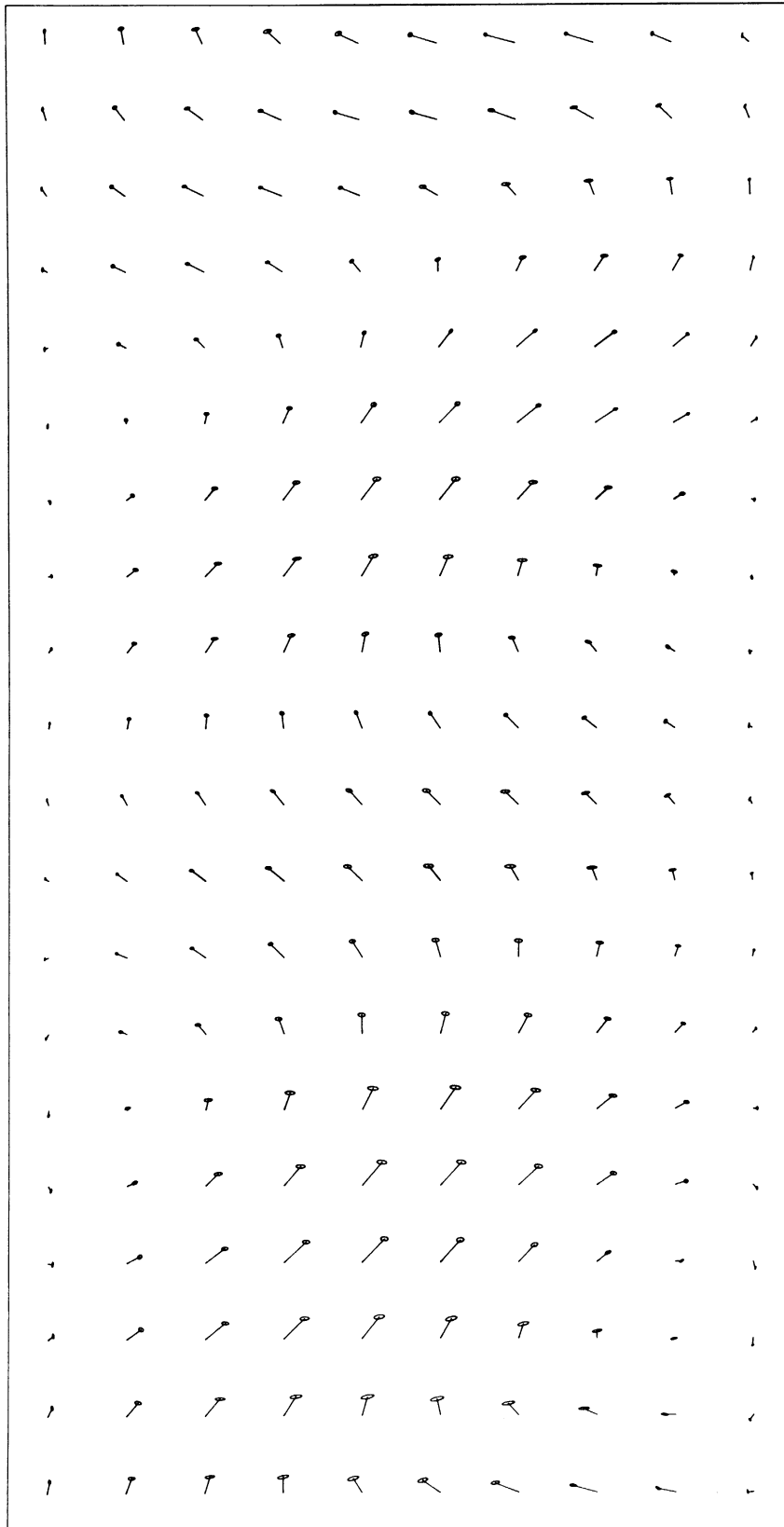


at each point, but not the interrelations among the winds at different points.

The general structure of the wind field is immediately apparent in Figure 7. It is also apparent that in some places the wind is much better known than elsewhere. For example there are "southwest" winds in the "south-central" section of our space domain and some "northwest" winds further to the "east" that show very little uncertainty. On the other hand the winds in the vicinity of the strong ridge near the "western" boundary show a great deal of uncertainty in their meridional components, although not in their zonal components. This is reflected in the uncertainty of the position of the ridge (cf. Figure 2a). One notices that in general we seem to have more confidence in our forecasts of the zonal component of the wind than the meridional component.

Figure 8 is similar to Figure 7, except that the vectors are thermal winds, derived from the temperature field. We had already seen, from Figure 3, that the uncertainty of the temperatures was less than that of the streamfunction, and this is substantiated by our finding smaller ellipses of uncertainty in Figure 8 than in Figure 7. But note that while the charts of Figure 3 deal with the uncertainty of  $\psi$  and  $T$ , the ellipses in Figures 7 and 8 refer to the uncertainty in the gradients of these fields.

The use of credible ellipses also has application to the representation of the forecast in phase space, which is the multidimensional space in which the dependent variables of the model are represented along orthogonal axes. Of course it is generally feasible to draw pictures in only two dimensions at a time. For example, in Figure 9, the axes are the parameters  $B_1$  and  $B_{15}$ , which represent, in effect, the mean zonal wind and temperature gradients. The shape,



X →

← Y

Figure 8. The forecast thermal wind field, represented as in Figure 7.

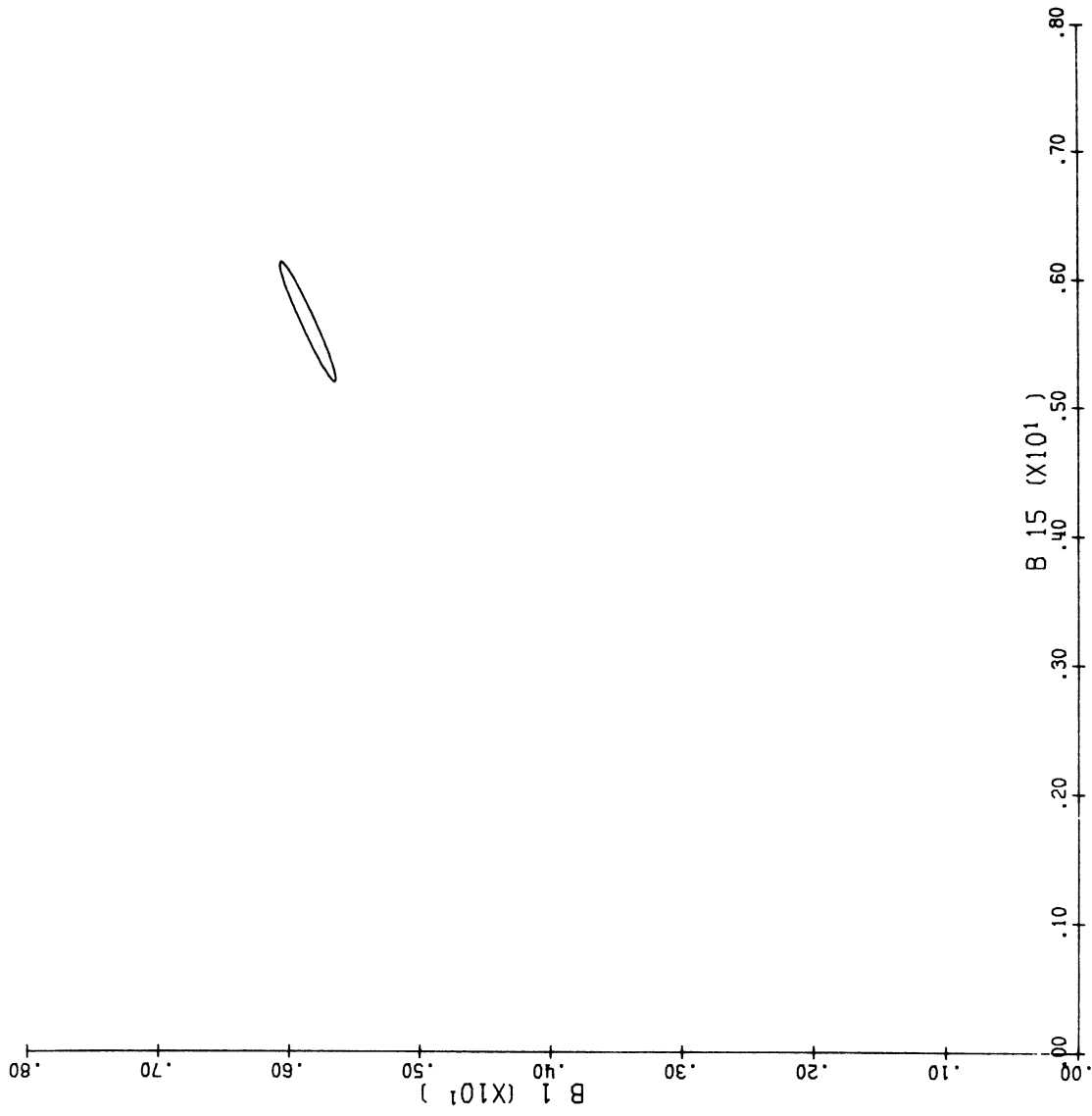


Figure 9. The 50% credible ellipse for the joint distribution of the amplitudes of the mean zonal wind ( $B_1$ ) and temperature gradient ( $B_{15}$ ).

size and orientation of the 50% ellipse tell us not only the limits within which we expect these quantities to lie, but also the extent to which a larger temperature gradient will be accompanied by a stronger zonal wind. If, for example, we observe tomorrow a stronger than expected zonal wind, then we can infer (quantitatively) a stronger than forecast mean zonal temperature gradient, and vice versa.

We can carry this analysis one step further. The curve labeled " $\psi$ " in Figure 10 is the 50% ellipse for the parameter pair which are the coefficients of the sine and cosine terms of the stream function for the longest wave in the model. A vector from the origin to the center of that ellipse represents the phase and amplitude of the "expected" wave. The curve labeled "T" is the 50% ellipse for the corresponding temperature wave.

These two ellipses both represent marginal distributions of the particular pairs of parameters. It is also possible to illustrate various conditional distributions. Just as the sine and cosine components of the wave are correlated, as indicated by the fact that the axes of the ellipse are skewed with regard to the coordinate axes, so also is there information about the temperature wave contained in knowledge of the stream function wave, and vice versa. For example, if the temperature wave were given by the point indicated by a mark on the "T" ellipse, then the 50% ellipse indicating the conditional distribution of the corresponding stream function is the small ellipse labeled

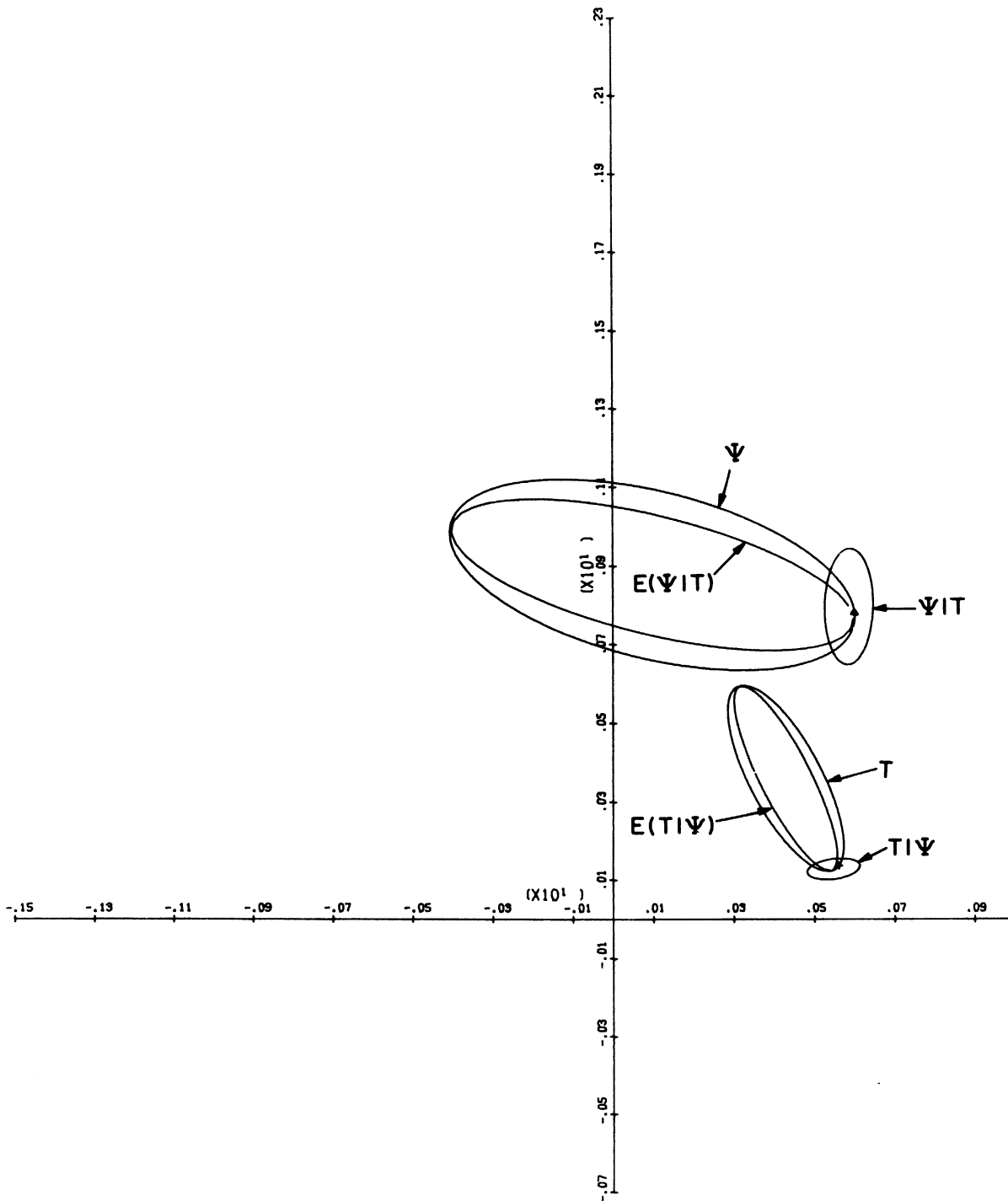


Figure 10. Ellipses representing the marginal and conditional joint distributions of the forecast amplitudes of the stream function and temperature waves of mode 1, wave no. 1. See text for explanation of symbols.

The use of the multivariate normal distribution as a model for representing our knowledge about these parameters also implies homoscedasticity. That is the variances (and covariances) of the conditional distributions are not dependent on the values of the "given" parameters. As the assumed value of "T" changes, the center of the " $\Psi|T$ " ellipse will change, but its size, shape, and orientation will not. Indeed if the indicated point on the "T"-ellipse were to move around that curve, the center of the " $\Psi|T$ " curve would trace out the ellipse labeled " $E(\Psi|T)$ ". The curves labeled " $T|\Psi$ " and " $E(T|\Psi)$ " have entirely symmetric interpretations.

Notice in Figure 10 that the ellipse " $T|\Psi$ " is considerably smaller than the "T" ellipse. This implies a large degree of correlation between the  $\Psi$ - and T-waves. The larger the ellipse " $T|\Psi$ " becomes, the smaller the ellipse " $E(T|\Psi)$ " would become, until, in the limit, " $E(T|\Psi)$ " would shrink to a point in the center of the "T"-ellipse and the " $T|\Psi$ "- and "T"-ellipses would coincide. At the opposite extreme, when knowledge of  $\psi$  completely specifies T, " $T|\Psi$ " shrinks to a point on the "T"-ellipse, and "T" and " $E(T|\Psi)$ " coincide.

In algebraic terms the ellipses are readily written in terms of their mean value vectors

$$\bar{\Psi} = \begin{pmatrix} B_i \\ B_j \end{pmatrix} \text{ and } \bar{T} = \begin{pmatrix} B_{i+1} \\ B_{j+1} \end{pmatrix}$$

where i and j represent an appropriate pairing of indices, and variance-covariance matrices:

$$\Sigma_{\Psi} = \begin{pmatrix} \text{var} (B_i) & \text{cov} (B_i, B_j) \\ \text{cov} (B_i, B_j) & \text{var} (B_j) \end{pmatrix}$$

$$\Sigma_{T\Psi} = \begin{pmatrix} \text{cov}(B_i, B_{i+1k}) & \text{cov}(B_i, B_{j+1k}) \\ \text{cov}(B_j, B_{i+1k}) & \text{cov}(B_j, B_{j+1k}) \end{pmatrix}$$

and

$$\Sigma_T = \begin{pmatrix} \text{var}(B_{i+1k}) & \text{cov}(B_{i+1k}, B_{j+1k}) \\ \text{cov}(B_{i+1k}, B_{j+1k}) & \text{var}(B_{j+1k}) \end{pmatrix}$$

Then the "T" ellipse is given by

$$(T - \bar{T})^* \Sigma_T^{-1} (T - \bar{T}) = C,$$

the "E(T| $\psi$ )" ellipse is

$$(T - \bar{T})^* \Sigma_{T\Psi}^{-1} \Sigma_{\Psi}^{-1} \Sigma_{\Psi T}^{-1} (T - \bar{T}) = C$$

and the "T| $\Psi$ " ellipse is

$$(T - T_*)^* (\Sigma_T - \Sigma_{T\Psi}^* \Sigma_{\Psi}^{-1} \Sigma_{\Psi T})^{-1} (T - T_*) = C.$$

The constant, C, is  $-2 \ln(1-p)$ , where p is the probability that the appropriate vector will lie within the ellipse.

$$T_* = \bar{T} + \Sigma_{\Psi}^{-1} \Sigma_{\Psi T} (\Psi_0 - \bar{\Psi})$$

is the expected value of T, given  $\Psi = \Psi_0$ .

Returning to Figure 10, we see that the conditional distributions provide considerable information not contained in the marginal ellipses. For example, it is apparent that the  $\Psi$ -wave is leading (in the meteorological sense) the T-wave by about  $40^\circ$ , and very likely by at least  $30^\circ$ , but no more than  $50^\circ$ . There is an uncertainty of the position of either wave by as much as  $60^\circ$ , but

relatively little uncertainty in their relative positions. This is of course directly related to the heat flux being accomplished by the wave.

As we examine the phase diagrams for some of the other waves, shown in Figures 11-15, we see that the phase relations are generally known much better than the positions of the waves. The most substantial eddy flux of heat is brought about by wave 2, (Figure 11) mode 1, and is also in the expected "north-west" sense.

In the case of wave 3, mode 1 (Figure 12), the uncertainty is so large compared to the expected amplitude of the wave that it would be reasonable to think of this wave as "unpredictable."

In making these predictions we have used a stochastic model that assumed that the basic zonal heating, which is the ultimate source of the energy in the model, is not known precisely. Its standard derivation has been taken as 10% of its expected value. We have also made a stochastic forecast using the same simulated observations, but assuming no uncertainty in the zonal heating. We find that the major benefit of knowing this forcing term exactly is that the mean zonal wind is more predictable, and knowledge of the phase difference between the wave 2, mode 1  $\psi$ - and T-waves, where the greatest heat flux is being accomplished, is particularly improved. Compare Figures 16 and 17 with Figures 9 and 11. It is an interesting sidelight concerning the model, and the stochastic method, that even though the rate of generation of available potential energy is what is particularly uncertain in the former case, it is largely in the uncertainty of the kinetic energy that this uncertainty appears in the forecast. The model not only converts zonal available potential energy to



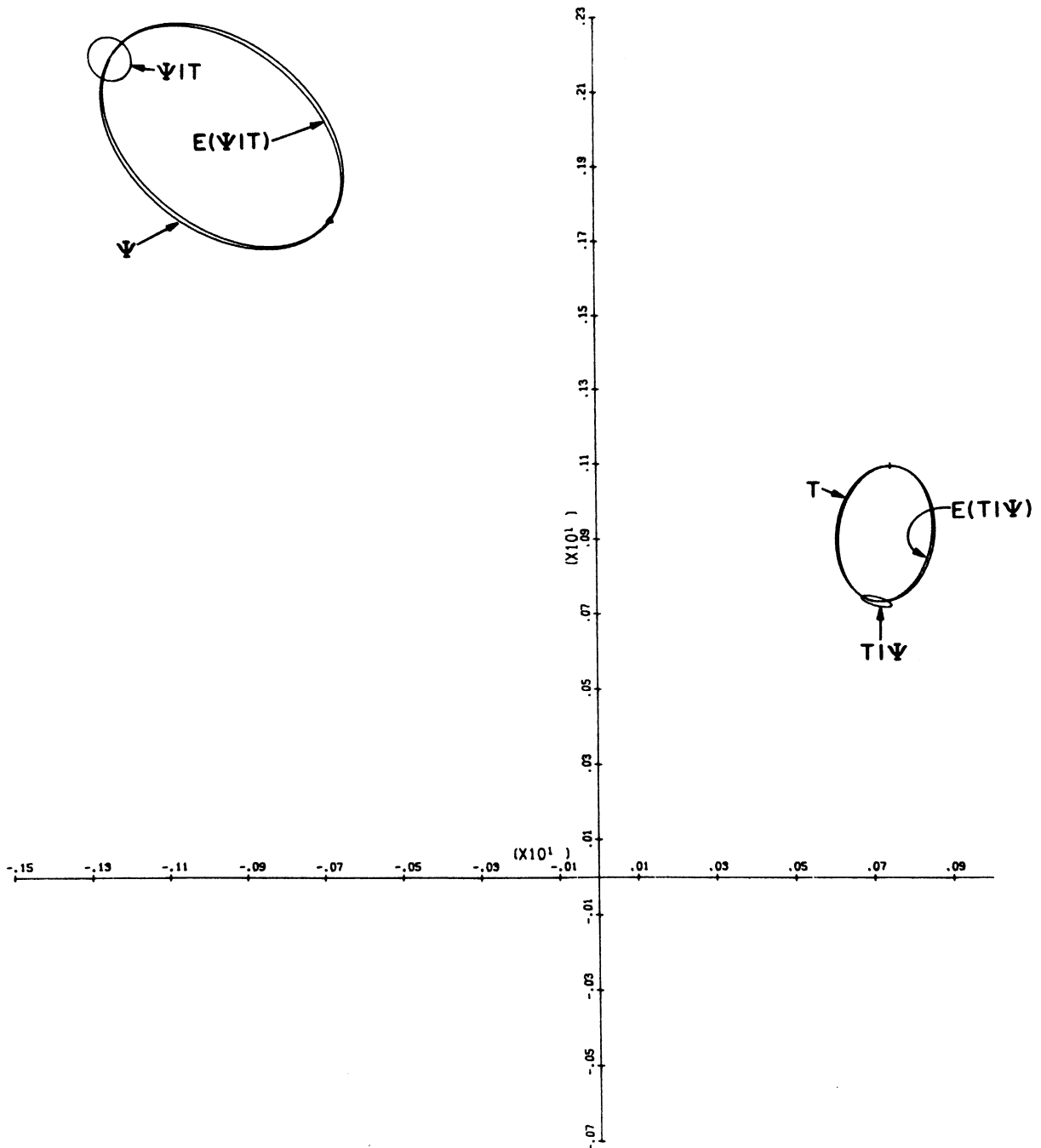


Figure 11. Phase space representation of the forecast of wave no. 2, mode 1.

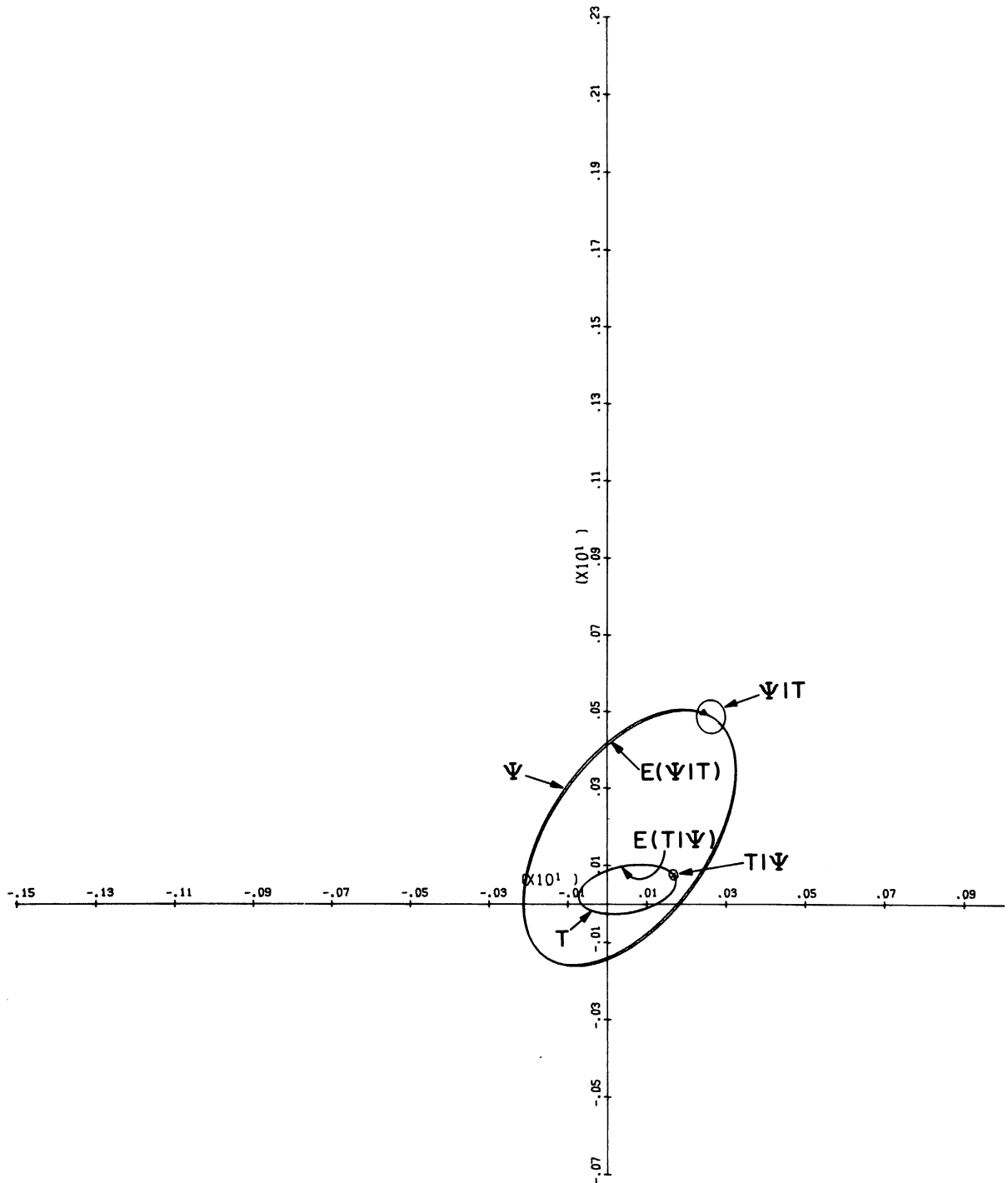


Figure 12. Phase space representation of the forecast of wave no. 3, mode 1.

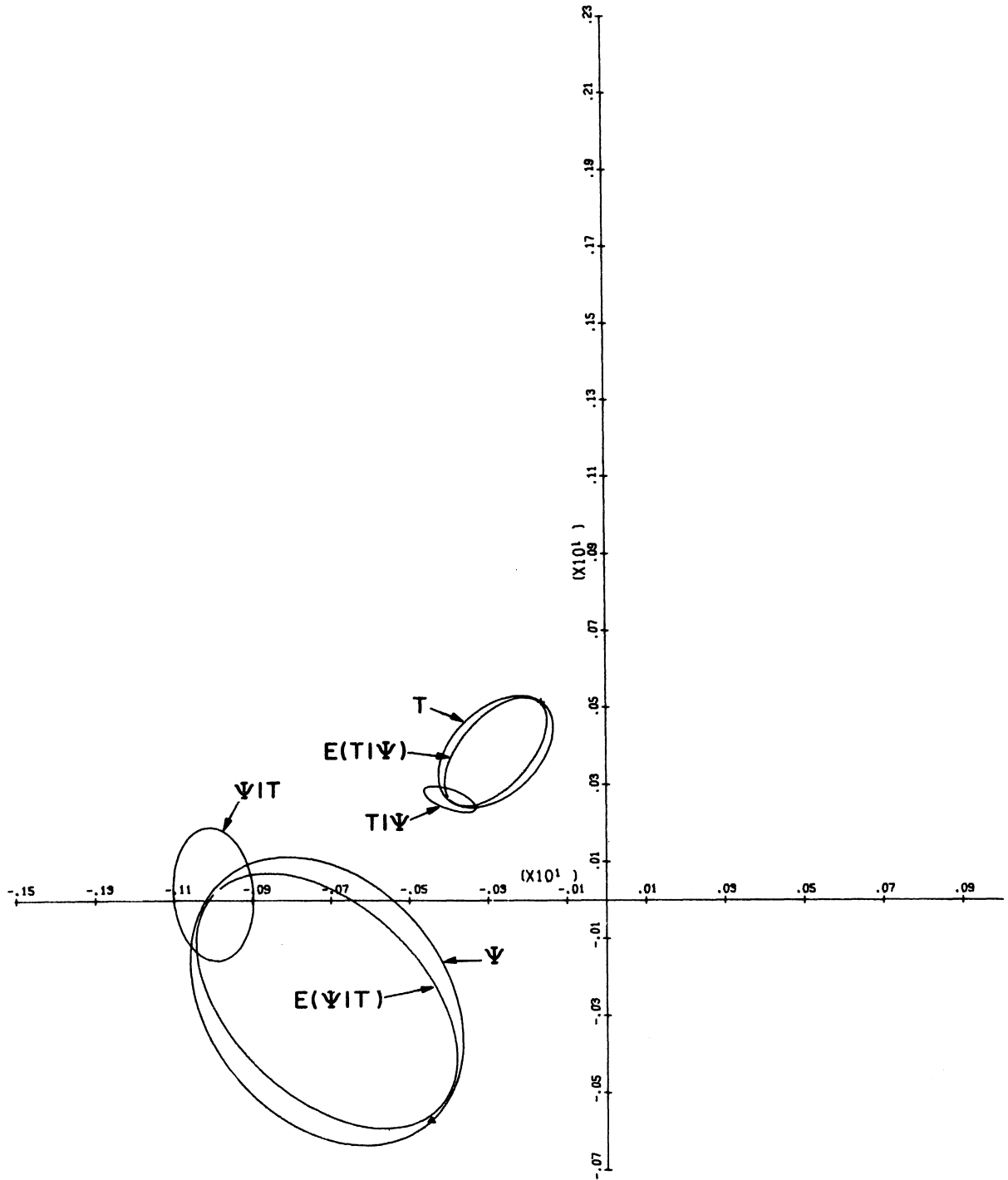


Figure 13. Phase space representation of the forecast of wave no. 1, mode 2.

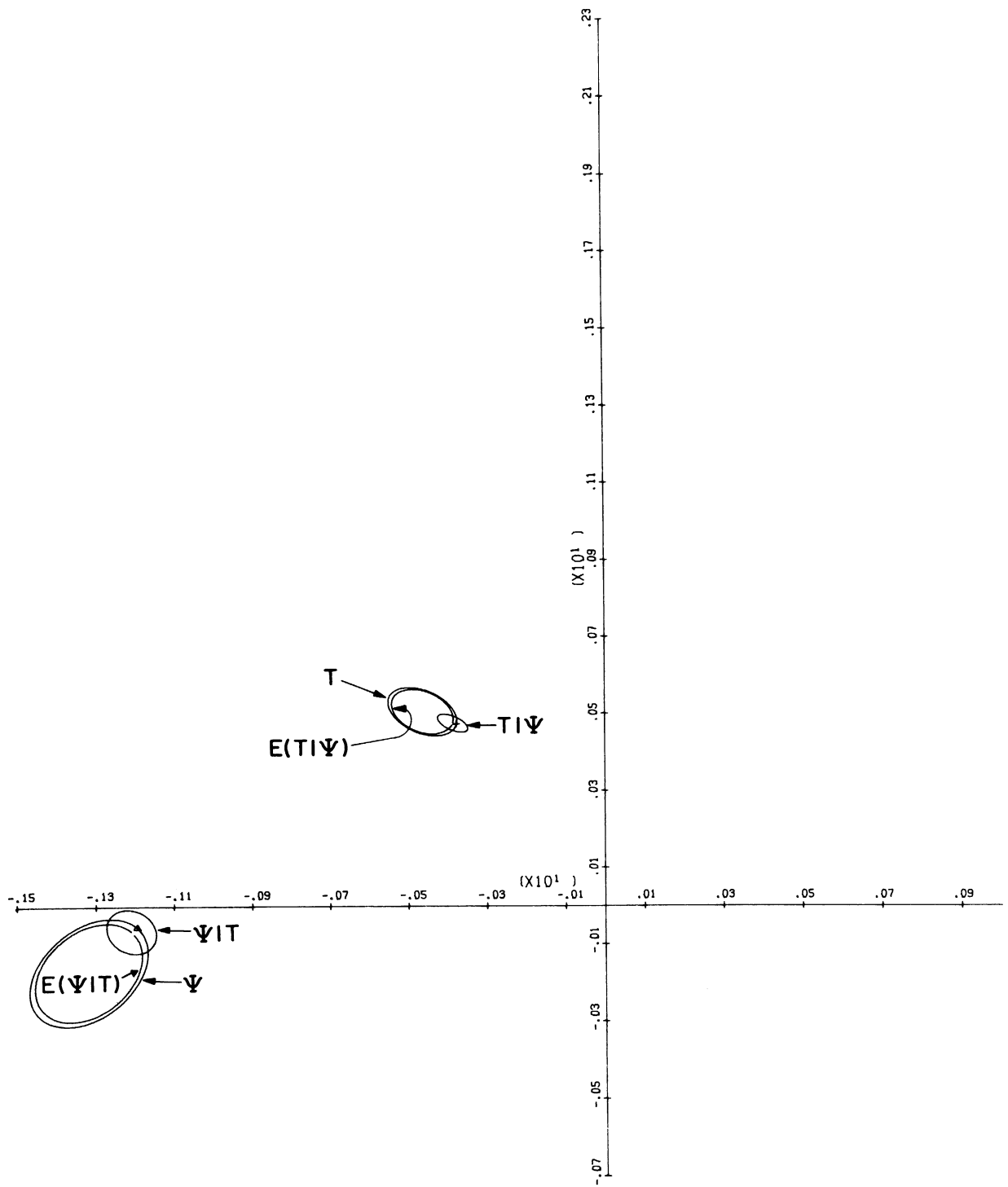


Figure 14. Phase space representation of the forecast of wave no. 2, mode 2.

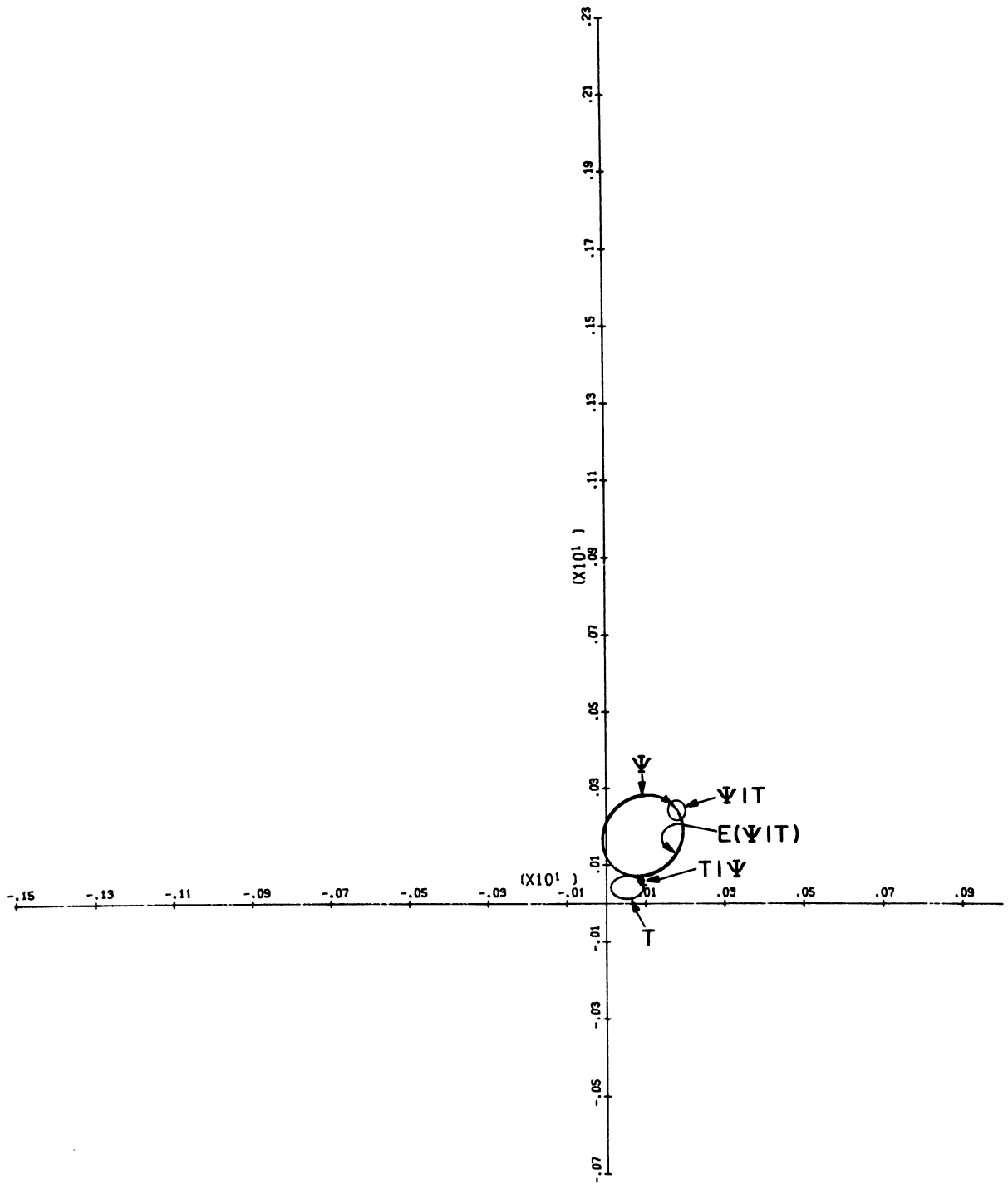


Figure 15. Phase space representation of the forecast of wave no. 3, mode 2.

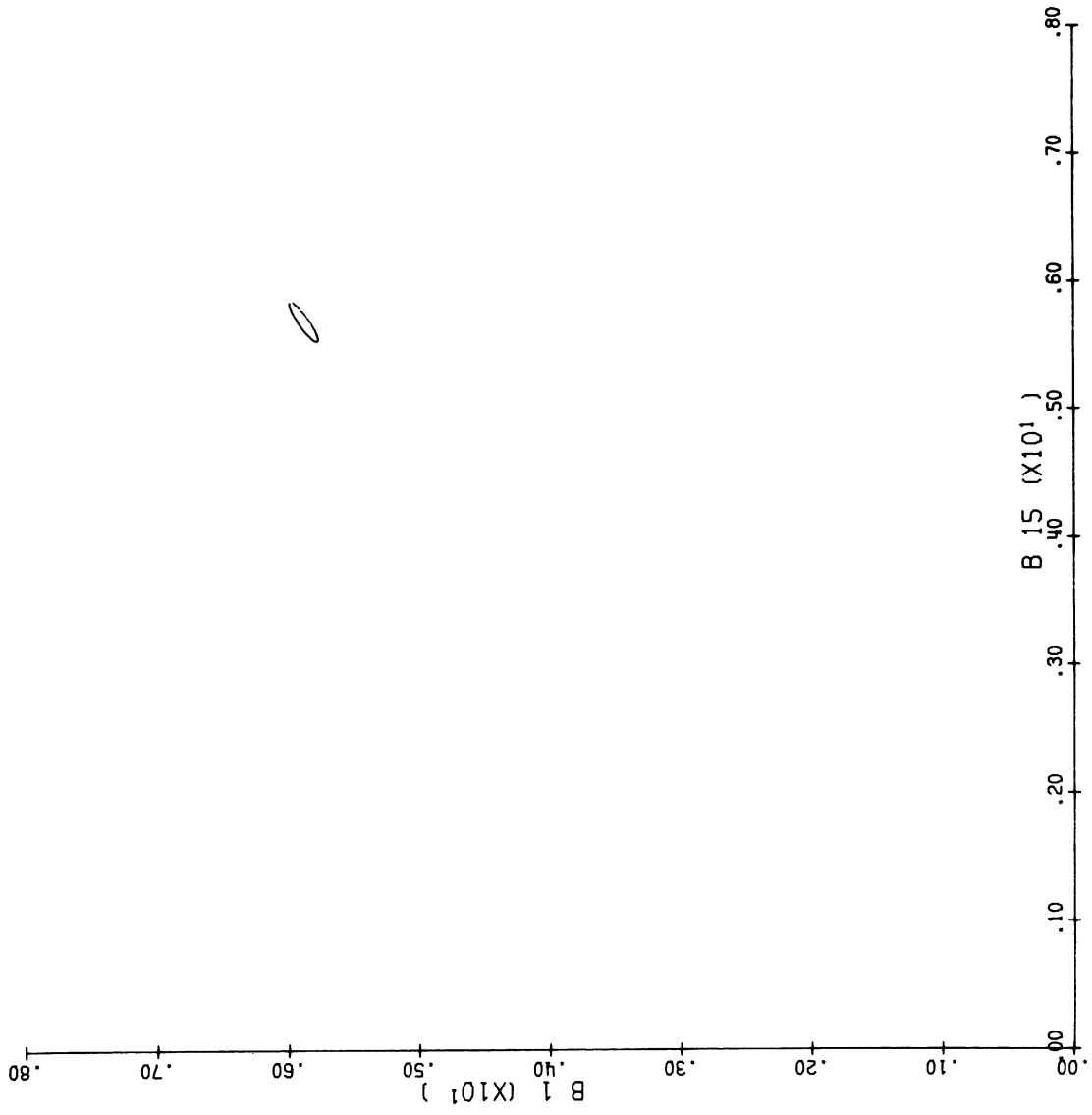


Figure 16. Representation of the joint forecast of the mean zonal wind and temperature gradient in the case of known heating. Compare with Figure 9.

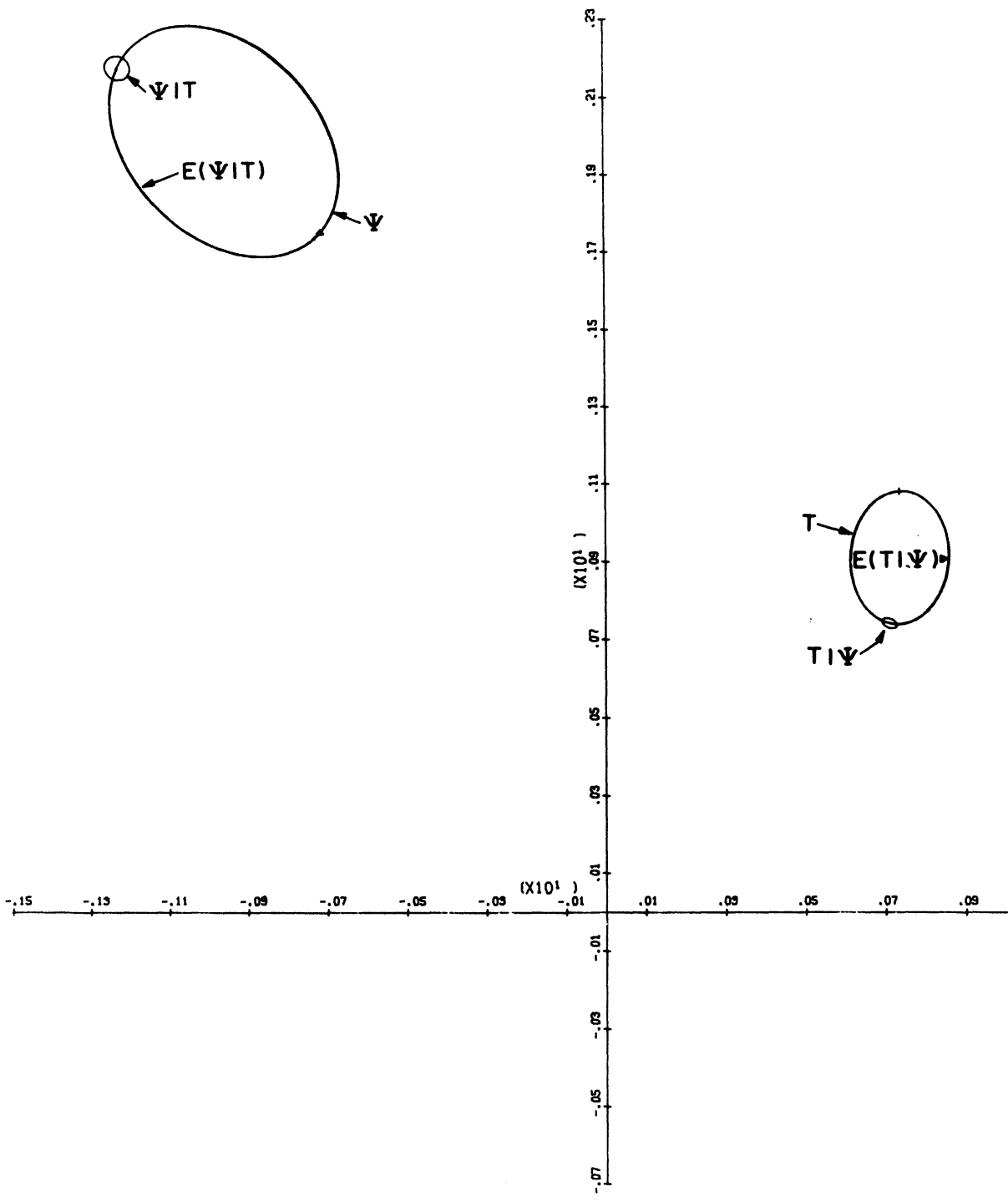


Figure 17. Forecast of wave no. 2, mode 1, in the case of known heating. Compare with Figure 11.

eddy APE, and on to eddy kinetic energy, but it also fully accommodates conversions of "uncertain" components of these categories of energy (Fleming, 1970).

There is almost no limit to the variety and complexity of the information that stochastic predictions can make available to the meteorologist. We hope that this introduction to some of the forms that the presentation of this information can take will serve to demonstrate the value of stochastic information. There is a great deal of significant information that can be effectively communicated and comprehended. We hope to whet the appetite of meteorologists for producing forecasts that contain such information. We must warn though, that operational stochastic prediction is a very formidable task that will require greatly improved computing methods and capabilities. Still we think of it as the way of the future.



## REFERENCES

- Epstein, E.S., 1969: Stochastic dynamic prediction. Tellus, 21, 739-759.
- Fleming, R.J., 1970: Concepts and Implications of Stochastic Dynamic Prediction.  
NCAR Cooperative Thesis No. 22, The University of Michigan and Laboratory  
of Atmospheric Science, NCAR, 171 p.



THE UNIVERSITY OF MICHIGAN

DATE DUE

9-13 2:41