Forecasting Research

Technical Note No. 32

STATISTICAL ASSESSMENT OF THE MONTHLY MEAN ERROR IN THE 3-DAY FORECAST OF THE GLOBAL 500mb HEIGHT FIELD

by
R.A. Bromley and R. Downton

October 1990

Meteorological Office (Met O 11)
London Road, Bracknell, Berkshire RG12 2SZ, England

MET O 11 TECHNICAL NOTE NO 32

Statistical Assessment of the Monthly Mean Error in the 3-day Forecast of the Global 500mb Height Field

by

R.A.Bromley and R.Downton

Meteorological Office, UK.

METEOROLOGICAL OFFICE.
Met.O.11 Technical Note (New Series) No.32

Statistical assessment of the monthly mean error in the 3-day forecast of the global 500mb height field.

Bracknell::::1990:Pp.i+19:30cm:

06450591

551.509.313
551.509.5
551.547.5
519.23

Met O 11,
Meteorological Office,
London Road,
Bracknell, Berkshire,
England, RG12 2SZ.

October 1990

Statistical Assessment of the Monthly Mean Error

in the 3-day Forecast of the Global 500mb Height Field

by R.A.Bromley & R.Downton

1. INTRODUCTION

A standard diagnostic for the performance of a numerical weather prediction (nwp) system is the error of a predicted field, that is the departure of the forecast value from the value that is deduced to have actually occurred. Statistics offers the possibility that if a population of errors is under consideration, the departure of its mean from zero may be calibrated in terms of the variance of the population as a whole. "Student's t", or a variant of it, is the usual form of expression of this calibration: it has been used in recent work by Epstein (1988) where the results are used in a (suitably cautious) assessment of errors in the forecast model of the National Meteorological Centre, Washington.

The need for caution arises because meteorological fields and their errors possess correlations in time and space which may undermine this statistical approach. Preisendorfer and Barnett (1983) have discussed this problem further and have proposed a parametric method for the statistical assessment of sparse populations such as occur in the study of general circulation models. In the assessment of operational forecasting models, there may be sufficient members of a population of errors for an attempt to be made to estimate and allow for the existing correlations. Nevertheless, caution must still be exercised when interpreting the results: even if a stable parameter can be computed, it is unlikely that its distribution will comply with one of the standard statistical distributions (Preisendorfer and Barnett, 1983).

The approach used in the current work follows earlier studies where autoregressive time-series were applied to general circulation models (Katz, 1982; Wilson and Mitchell, 1987) but it has important differences because of the interdependence of the forecasts and the analyses in nwp. The time-series method has been used to identify the time-independent part of the model error at each point on a global grid. A type of t-statistic, here denoted by τ, has been evaluated for the monthly population of errors at each point. The distribution of τ over the globe is clearly still subject to spatial correlations, so that the population of τ must be assessed with care.

After reviewing the monthly mean errors for 1988 in Section 2, the theoretical basis of the investigation is set out in Section 3, including an examination of time-series of different orders for the month of February 1988. A first-order time-series is adequate for the representation of the errors of the 72-hour forecast of the 500mb height field. Global populations of τ for the time-independent part of the error are presented in Section 4: a similar exercise is conducted for some other forecast fields, from lower levels in the model atmosphere, in Section 5. The results are summarised and discussed further in Section 6.

2. MONTHLY MEAN FORECAST ERRORS DURING 1988

The heart of this study is a statistical assessment of the differences between the routine operational forecast of the 500mb height and the verifying analysis of the same field throughout the year 1988. These differences are frequently regarded as the error of the forecast, as will be done here also. (However, it is advisable to remember that the analysis is not the truth, but only an acceptable approximation to it: hence to regard the difference as the 'error' of the model forecast is also only an approximation). The global three-day forecast, produced by the coarse-mesh model of the UK Meteorological Office (Bell and Dickinson, 1987), has been considered. By this stage of the forecast, the pattern of the model errors is well-developed, and they are at least as important as any errors due to the initial analysis (Arpe et al, 1985).

The difference field has been calculated at 12Z for every day of 1988 that was present in the archive of operational forecasts and analyses. Usually only one or two days per calendar month were missing: the actual number of days used in each month is shown in Table 1. The mean monthly difference was then computed as the time-mean of the differences at each point of the model grid. This has been done for each month of 1988.

Results for January, April, July and October of 1988 are shown in Figures 1a, b, c and d respectively. In general the errors in the northern and southern hemispheres at the time of their own winters are larger than those that are found in summer. In the northern hemisphere, the largest monthly mean error exceeds 6 decametres in all months from January to April and from November to December. In the southern hemisphere, the errors are generally larger than in the northern and exceed 6 decametres somewhere in the hemisphere in every month. However, the largest errors in the southern hemisphere are found in the months away from the solstices: in April and May in the southern autumn, and in August to November of the southern spring.

In the UKMO model, the geographical distribution of the mean errors and its variation over the year follows a well-established pattern (White and Adams, 1988), similar to that found in other nwp centres (e.g. Arpe and Simmons, 1988; Bettge, 1983; White, 1988). The most obvious feature is that a majority part of the whole area of the globe is covered by an area of negative error, due to a slight cooling which is already known to be present in the UK forecast model. It is also obvious that the largest errors occur at latitudes polewards of 30° in association with the depression belt in each hemisphere. Topographic effects are thought to be responsible for the elongation of the error structure along the axes of the Rockies and the Andes. The local maximum in the region of the Tibetan plateau is also due to topography, but, since this land mass is above the chosen pressure level (500mb), this feature of the error field is rather artificial.

The standard deviation shows the variability of the errors from day to day taken over each month as a whole. As with the mean error, the largest values are found in association with the depression belt polewards of 30°. In the northern hemisphere the standard deviation usually has its largest values over the North Atlantic or the North Pacific, although large values are also found over Northern Asia and Hudson Bay in some months. The distribution of this field within the southern hemisphere depression belt

appears to be less organized so that local maxima occur in a fairly random fashion. In the results for January, April, July and October (Figures 2a, b, c and d) the areas with standard deviation greater than 6 decametres have been shaded. This shows that, as for the mean error, the standard deviation of the error is generally larger in the southern hemisphere than in the northern, and in each hemisphere it tends to be larger in winter than in summer. Unlike the mean error however, the period with the largest standard deviations (more than 12 decametres) in the southern hemisphere is from May through August to November, coinciding with the southern winter. The northern hemisphere values are also at their largest during winter.

The distributions of the mean errors and of the standard deviation of the errors are further illustrated by the graphs in Figure 3. These show the mean values of these two statistics around each latitude circle of the model, plotted as a function of latitude. The graphs that emerge are consistent with the full geographical distribution of the two statistics in that they show that the extreme values occur in the mid-latitude regions of each hemisphere. However, it is also apparent from the two graphs that the latitudinal dependence of the standard deviation is much stronger than that of the mean so that the ratio of the mean to the standard deviation is largest in the tropics. This suggests that the errors in the tropics may differ from zero by an amount which is relatively greater than for the errors in mid-latitudes. This possibility will be investigated further in the following sections.

3. STATISTICAL TREATMENT OF MODEL ERRORS

3.1. Mean and Variance of the Forecast Errors.

The time series of model errors at a specific location, e_i, is made up of the differences $f_i - a_i$ between the forecast f_i and the verifying analysis for time i. Thus, by definition,

$$e_i = f_i - a_i \qquad (1)$$

The time-mean of the error at this location, $\mu(e)$, is evaluated over the period from i=1 to N, where N is the number of days in the month. By a familiar statistical identity,

$$\mu(e) = \bar{f} - \bar{a} \qquad (2)$$

where $\bar{f}$ and $\bar{a}$ are the mean forecast and the mean analysis respectively over the same period.

Thoughout this study it has been assumed that at a given location the time-mean error $\mu(e)$ is generally non-zero. It is desired to know whether the departure of $\mu(e)$ from zero is excessive, (and therefore, perhaps, susceptible to correction) or whether it is no more than might occur by chance. This may be done using a version of the t-statistic to express the mean error as a ratio of its variance. To do this, however, the variance of the error must also be evaluated. Although the variance is the familiar sum of squares of the departures of each e_i from the mean μ, it is made more complex by the fact that the time-series of either the errors, e_i, or the individual fields, f_i and a_i, possess correlations which must be identified and removed before the independent part of the variance can be computed. The correlation of the fields is contained in the covariance which appears in the standard equation for the variance of the error, $\sigma^2(e)$, in terms of the variances of its constituent parts:-

$$\sigma^2(e) = \sigma^2(f) - 2\,\mathrm{cov}(f,a) + \sigma^2(a) \qquad (3)$$

In climatological work it is acceptable to assume that the forecasts and the verifying analyses are independent so that their covariance $\mathrm{cov}(f,a)$ may be taken to be zero. It is then possible to evaluate $\sigma^2(e)$ in terms of $\sigma^2(f)$ and $\sigma^2(a)$, as was done by Katz (1982). Harr et al (1983) have proceeded from this same assumption to develop a version of the t-statistic which they applied to operational analyses and 72-hour forecasts. However, since there is some skill in the forecast at 72 hours, it seems reasonable to assume that there is some correlation between the two fields, and that therefore the covariance is not zero. Consequently $\sigma^2(e)$ must be evaluated directly from the time-series e_i itself.

3.2. First-Order Autoregressive Time-Series

To evaluate the independent part of the variance from the time-series now requires an approach that will make allowance for the

dependence of each individual member of the series on its predecessors. This is true not only for the series e_i but also for f_i and a_i. However, e_i is expected to be more nearly random than f_i or a_i and so the numerical processing is less complicated: whereas Wilson and Mitchell (1987) found they had to assume at least a second-order autoregressive time-series for the fields f and a, it may be possible to assume a first-order time-series, corresponding to a red-noise spectrum, for the errors. The analysis of the first-order series has been set out by Box and Jenkins (1976, Chapter 3.2) whose approach is repeated here.

It is assumed that the error satisfies the first-order autoregressive process described by

$$e_i = \varphi_{1,1} e_{i-1} + r_i \qquad (4)$$

In this notation, the coefficient $\varphi_{p,j}$ is a constant factor multiplying e_{i-j} in an autoregressive time-series of order p, and r_i is a purely random contribution to the error at time i. In general the coefficients $\varphi_{p,j}$ may be related to the time-lagged autocorrelation coefficients ρ_k through the Yule-Walker relations for j and k ranging from 1 to p. However, in this first-order process, the Yule-Walker equations are much simplified: in particular the 1-day time-lagged autocorrelation is given by the trivial equality

$$\rho_1 = \varphi_{1,1} \qquad (5)$$

It follows that the coefficient $\varphi_{1,1}$ in equation 4 may be determined by computing the 1-day autocorrelation coefficient of the time-series e_i. Once this autocorrelation is known, it may be used to estimate $\mu(r)$, the mean of r_i over the period, by using equation (5) in the time-mean of equation (4) to obtain

$$\mu_1(r) = (1-\rho_1)\, \mu(e) \qquad (6).$$

The subscript on $\mu(r)$ is the order of the series used for the estimate. Similarly, the variance of the series may now be expressed as

$$\sigma^2(e) = \sigma_1^2(r)/(1 - \varphi_{1,1}^2) \qquad (7)$$

and this too may be re-worked to obtain an estimate for the independent part of the variance contributed by the random element r_i:-

$$\sigma_1^2(r) = (1 - \rho_1^2)\, \sigma^2(e) \qquad (8)$$

where once again the subscript on $\sigma(r)$ is the order of the series used for the estimate. Corresponding results will be found in several text-books on statistics: a simple meteorological version was given by Brooks and Carruthers (1953). The variance of r is equivalent to the 'innovation variance' described by Katz (1983).

It is now possible to construct a version of the t-statistic for r_i, the set of independent, random contributions to each error e_i with $\mu(r)$ as the mean value of the set of random contributions and $\sigma^2(r)$ as the independent part of the variance. Since this statistic may contain

contributions which prevent it from being distributed like "Student's t", it will be denoted by the Greek letter τ. The number of degrees of freedom is taken to be the number of members of the set, n, less p for the constraint imposed by the assumption of the pth-order autoregression series, less a further 1 for the constraint imposed by sampling. Hence

$$\tau = \mu_p(r)(n-p-1)^{1/2} / \sigma_p(r) \qquad (9)$$

3.3. Second-Order Autoregressive Time-Series

Before proceeding to use an estimate of τ based on a first-order time-series, it is desirable to investigate whether a higher-order series is more suitable. Using the same notation, a second-order autoregressive series would have the form

$$e_i = \varphi_{2,2} e_{i-2} + \varphi_{2,1} e_{i-1} + r_i \qquad (10).$$

Still following Box and Jenkins (1976), the coefficients $\varphi_{2,1}$ and $\varphi_{2,2}$ may be determined from the Yule-Walker relations for the second-order series, giving

$$\varphi_{2,1} = \frac{\rho_1 (1 - \rho_2)}{1 - \rho_1^2} \qquad (11)$$

$$\varphi_{2,2} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \qquad (12).$$

It will be noted that if $\rho_2=\rho_1^2$ then $\varphi_{2,2}$ is zero and the series reverts to the first-order form with $\varphi_{2,1}=\varphi_{1,1}=\rho_1$. However, it is unlikely that this condition will hold good for the whole of the global distribution of forecast errors in the 500mb height field. The second-order series yields new estimates for the mean and variance of the random part of the error:-

$$\mu_2(r) = (1-\varphi_{2,1}-\varphi_{2,2})\ \mu(e) \qquad (13)$$

$$\sigma_2^2(r) = (1 - \rho_1\varphi_{2,1} - \rho_2\varphi_{2,2})\ \sigma^2(e) \qquad (14)$$

New values of τ appropriate to the second-order time-series may be constructed using these estimates in equation (9) with p set to 2.

3.4. Investigation of February 1988

The choice of p, the order of the autoregressive time-series, was investigated by considering the results for a single month which was assumed to be typical for the year as a whole. February was chosen for the study because it was the first month with a complete record of forecasts in the archive (January had one day missing).

The j-days time-lagged autocovariance of the error of the T+72 forecast of the height field at 500mb was estimated by summing the products $(e_i-\mu)(e_{i-j}-\mu)$ at each grid-point of the field for all 29 days. The summation started on the first day of the month so there was an overlap with the errors for the last j days of the previous month. The time-lagged autocorrelation ρ_j was obtained by dividing the

corresponding autocorrelation by the variance of e_i over the month. This is a general procedure that was eventually used for all months, but in this investigation it was applied to the data for February 1988 to determine both the first- and second-order time-lagged autocorrelation coefficients. The global distributions of the results are shown in Figures 4a and b. It can be seen that in general the condition $\rho_2=\rho_1^2$ does not hold and the investigation must proceed to examine the τ-statistics obtained from the two series. However, it may be noted in passing that there are persistently high values of the autocorrelation (ρ_1 and ρ_2 are both greater than 0.5) in the vicinity of the sub-tropical anticyclones in both hemispheres and also over the Canadian prairies, the Western Sahara and to the south of New Zealand. In contrast, there are also negative values of ρ_1 in meteorologically active areas such as in the baroclinic regions over the Gulf Stream near to Newfoundland and around the Antarctic landmass, and in the convective areas near the Equator over the Indian Ocean and the Western Pacific.

The τ-statistics have been calculated in accordance with equation (9), using means and variances of r_i determined for both the first- and second-order time-series that yielded the autocorrelation coefficients in Figures 4a and b. The first-order results for the spatial distribution of τ are shown in Figure 5a: the second-order results are shown in Figure 5b. For the purposes of the present stage of the study, it is only necessary to note that the geographical distribution of τ in these two charts is virtually the same. Of course, equation (9) ensures that the pattern of positive and negative areas of τ is the same as for the mean error, μ, since all other factors in the formula for τ are positive. However, the correspondence between the two charts is very strong even for the most positive and most negative values of τ. Close examination will reveal that there are some small areas of difference, but these only indicate a random set of both positive and negative changes: there is no systematic difference between the two charts. Hence there is no indication of a meaningful difference between the geographical distributions of τ computed from the first- and second-order time-series. The geographical distribution of τ will be considered further in Section 4.

The results for the two fields of τ may also be compared with the field of τ obtained by using the original fields of model errors e_i without the adjustments imposed by the time-series. To do this the population distributions of τ have been displayed in Figures 6a and 6b for the first- and second-order time-series and in Figure 6c for the unadjusted values of τ. Once again the similarity between the results from the two time-series is very strong. There are some minor differences, especially close to τ =0, where the second-order field appears to have slightly fewer values than does the first-order field, but these are not likely to be important. However, on comparing the unadjusted population (Figure 6c) with the fields from the time-series (Figures 6a and 6b), some more serious differences are seen: there are fewer values of τ near to zero in the unadjusted field and it has more values of τ that are more negative than -3: there are also more values of τ that are more positive than +4, but this difference is not so marked as the others.

It may be concluded from the population distributions and from the geographical distributions of the calculated fields of τ that the use of the time-series has made an improvement to the distributions, producing a more realistic value of τ, because it makes allowance for the temporal correlation of the errors. However, since the use of a time-series of second-order (instead of first-order) appears to have made no great difference, the time-series in the rest of this study have been kept to first-order.

4. ASSESSMENT OF THE COMPUTED POPULATION OF τ

4.1 Population distribution of computed values of τ

The conventional approach to a test of statistical significance for "Student's t" would be to consider only one value of t and then to test whether it might reasonably be expected to have occurred as a random event (Weatherburn, 1962).In the current investigation, where a complete population of values of τ is being considered, the equivalent of this conventional approach would not be appropriate. This is because the distribution to be expected for τ is not known and because the occurrence of some extreme values in the whole population should not be unexpected. For example, if it was known what value of τ would be exceeded with a probability of 1%, then out of the 23232 points that make up the model grid, 232 points would be expected to exceed the known value. It follows that an assessment of the computed values of τ cannot be made by considering the values at individual points in isolation. However, some useful conclusions may be drawn by considering the shape and location of the population distribution as a whole.

The population distribution for the computed values of τ in May 1988 is shown in Figure 7. Two features of this distribution are immediately apparent. First, the maximum is displaced from the origin towards negative values of τ : this is consistent with the well-known cold bias of the 500mb forecasts (E.g. White, 1988; White and Adams, 1988). Secondly, the distribution is lop-sided, even after allowing for the bias of the population as a whole, with many more extreme values on the negative side than on the positive. The numbers of occurrences of a value of τ either greater than +5 or less than -5 have been computed for each month of 1988 and they are set out in Table 1. The results confirm that the wings of the τ-distribution in most months are very similar to those for May, with more than 2000 points from the model grid having τ less than -5. Even after allowing for the bias of the whole distribution, by shifting the τ-axis by between 1.0 and 1.5 units to the negative side, the lop-sided character remains: it could even be measured as the ratio of negative to positive values deviating from the peak by more than a chosen amount, say 3 units. This ratio would favour the negative lobe in all instances. Regardless of the true distribution of τ, this lop-sided character should not occur: it is the clearest indicator of an unexpected deviation of the errors from zero that is available in this study.

4.2. Geographical distribution of t

Besides the global distributions of τ for February 1988, already shown in Figures 5a and b, similar charts of τ have been computed for all months investigated. The results for May 1988 (Figure 8) may also be considered. As already noted (Section 3.4), equation (9) ensures that the pattern of positive and negative values of τ in these figures is the same as for the mean error, μ. The most interesting parts of the figures are those areas with exceptionally large positive or negative values of τ since it is here that the magnitudes of the mean error may be considered to be unexpectedly large and probably much greater than might occur by chance. These areas are located over the Pacific (positive values up to 10) and over the Amazon basin and the Indian Ocean (negative values up to -25).

The chart for May illustrates that the pattern seen in February is persistent: in fact it occurs in a similar form for every month of 1988. In each month of the twelve the most negative values of τ occur over the Amazon basin and Indonesia. The area of values more negative than -5 usually extends out from Indonesia over the Indian Ocean with occasional incursions over East Africa, and out from the Amazon basin over the tropical Atlantic. The area does not reach out very far over the Pacific in any month. (On the contrary, the mean errors over this ocean are generally positive: but the magnitude of the deviation from zero remains unexpectedly large). The general behaviour over 1988 shows that the pattern in the tropics in February is rather weak, although the reason for this has not been discovered. By contrast, May has one of the strongest patterns of the year. The persistence of this pattern over so many months of 1988 suggests the presence of a persistent, non-negligible, source of error operating in the tropical regions of the numerical model.

However, it is evident that if the distribution of τ were indeed random, the occurrence of high values of τ should be found at occasional model points all over the globe. In fact the model is on a latitude-longitude grid and so a random distribution might be expected to have more points with high values of τ near the poles rather than near the equator. The need for appropriate weights to be used is one reason why the τ distributions shown in Figures 6 and 7 are not like the distribution of "Student's t" for the same number of degrees of freedom.

The spatial coherence of the global pattern (Figure 8) is an impressive sign that the feature is not random, but it is also an impediment to further investigation. The τ notation has been adopted because it was likely that correlations between different locations would affect the distribution of this statistic, further distorting it from the distribution expected for "Student's t". A full investigation of the distribution would require the determination of the spatial correlations, something that will not be attempted here.

4.3 Other populations of t

The differences between the first- and second-order calculations of τ have already been reviewed in Section 3.4. However, for completeness, the numbers in the wings of the τ-distribution have also been computed for the uncorrected parameter and for the second-order calculation for February 1988. The results, shown in Table 2, confirm that the use of the time-series has reduced the number of values of τ in the wings of the distribution and that there is little improvement on going from first- to second-order.

Table 2 also shows results for two further calculations. The first is for the month of December 1987. Although the results suggest that this month was more biassed than December 1988, the values for December 1987 are comparable with those for most of the months from 1988. Hence it appears that December 1988 showed a distinct improvement on the other months of the year, and on December of the previous year. Since the latest data assimilation scheme, the analysis correction scheme (Lorenc et al, 1989), was introduced on the last day of November 1988, it is quite likely that this was responsible for the comparative improvement in the results for December.

Finally, Table 2 contains one more set of results for February 1988: these are still computed from errors of the T+72 forecast of the 500mb height, but for the forecast valid at 00Z instead of 12Z. Although there is little to choose between the tabulated results for 12Z and 00Z, the chart of the global distribution of τ obtained from the first-order calculation at 00Z (Figure 9) appears to have more extreme values of τ, both positive and negative, than does the corresponding distribution for 12Z (Figure 5a). This may indicate that a weak diurnal variation in the modelling is causing the error.

5. ASSESSMENT OF τ FOR FORECAST FIELDS AT LOWER LEVELS

The appearance of such a pronounced field of τ in the error of the forecast at 500mb raises the question of what meteorological processes might have been instrumental in causing it. Since the geopotential height in pressure coordinates is calculated as an integral from mean sea-level,

$$z = \frac{R}{g} \int_{500}^{pmsl} T \, d(\ln(p)) \qquad (17),$$

it is possible that the fields of temperature and/or pressure that are needed to evaluate this function may themselves be subject to relatively large errors. Since these fields are all scalars, they may be assessed by means of the same procedure that has been used for the height at 500mb. The assessment has been carried out for three fields:- pressure at mean sea-level (pmsl), and temperature at 850mb and 700mb. The use of pmsl is satisfactory for the present purpose but a more complete analysis would allow for the presence of topography. A first-order time-series for the errors of the forecast was constructed for each of these fields during May 1988 and the field of τ was then computed for the time-independent part of the error in each series, following the method described in Section 2.

The results of these calculations are shown as population distributions in Figures 10a, b and c. The number of points with $\tau>5$ and $\tau<-5$ in each population are shown in Table 3. It can be seen that the two temperature populations are biassed negative much more strongly than is pmsl: this is consistent with the excessive cooling that is known to be present in the forecast model. However, in the distribution for pmsl the negative wing is much more extensive: since the peak of this distribution is quite close to $\tau=0$, the occurrence of such negative values for pmsl is relatively more important than they would be in the distributions for the temperature fields.

The geographical distributions of τ are shown in Figures 11a, b and c.. On comparing these three figures, it is apparent that the distribution for pmsl is quite different from those for temperature. The most negative values of τ for the pmsl errors occur in the same areas as were found for the errors in the 500mb height whereas the most negative values for the temperature errors are found mostly over the Pacific ocean and with a displacement of a few degrees of latitude away from the equator. Some of these temperature errors may be associated with the sub-tropical jets. Although there appears to be excessive cooling over North America at both 850 and 700mb, this effect is artificial, at least in part, as the field in most of this area lies below topography. It is also apparent that in those areas of South America and the Indian Ocean where the pmsl errors are so strongly negative, the temperature errors tend to be positive. (Of course this relation of the errors in the vertical would also have been seen in the fields of the mean errors). Overall these patterns suggest that the error at 500mb is closely related to the error in the pmsl and that where these errors occur, the heating in the atmosphere is acting to reduce them.

The location and shape of the most negative parts of the error pattern suggest a number of processes that may be responsible for them. They occur in regions generally associated with convection and they may also be thought to have the pattern of a planetary wave number two, perhaps arising from a

tidal effect. It may be speculated that the error in the forecast of surface pressure is due to an error in the model's treatment of divergence. This is reinforced by the observation that the high values of τ occur in regions with large values for the vertical velocity (G.H.White, private communication). However, although divergence is a scalar, the model does not use it directly but works in terms of the wind, a vector, instead. This model variable is not suited to the assessment methods that have been exploited in this work. Further investigation of the causes of the forecast errors must therefore be left to other methods: these may involve the use of more appropriate statistical techniques, or else a direct examination of the fields in individual model runs.

6. DISCUSSION AND SUMMARY

This study has developed and interpreted an objective statistical approach to the assessment of the global population of errors of the 72-hour forecast of the 500mb height. The 500mb height is a slightly unusual field to use for investigations which include the tropics: however, it is also a scalar quantity which is therefore suitable for applying the method that has been developed, and it is so well-known that it offers a suitable basis for judging the usefulness of this method. In the event, it has been possible to relate the results for the 500mb field to results for other (scalar) fields.

It has been shown that although the mean error of the forecast shows a distinct dependence on latitude, the variance of the error has an even stronger one. After allowing for the temporal correlation of the model errors, τ, a version of the t-statistic, has been used to draw attention to some features of the error field in the tropics, an area which may have been thought to have small and unimportant forecast errors. The results suggest that the most serious errors in the 500mb height forecast occur over South America and the Indian Ocean. Investigation of other fields shows that the geographical variation is similar for the pmsl errors but that the most serious temperature errors are over the Pacific Ocean. This discrimination between the errors of the different model fields is an indication of the usefulness of the method.

Although the tropics are an area with persistent model fields, it does not follow that the model *errors* should show the same persistence. It is unlikely that a first-order auto-regressive time-series would be sufficient to account for the persistence of model fields themselves. But it has been shown that not only does a first-order series give an adequate representation of the model errors, but also that such small changes as do occur on going to a second-order series are not confined to the tropics. So long as this remains the case, then it is not easy to disregard the implications of the τ-statistics that have been presented for the tropics.

On the other hand, although the spatial coherence of the distribution of τ prevents the attribution of statistical significance to the results, it certainly suggests that some specific mechanism is at work to produce this pattern. Although it is unlikely to be associated with a particular planetary wave, results from a spectral model would be needed to confirm this. Some speculations have been offered about a possible physical mechanism but further investigations are required to substantiate them.

The statistical procedures could have introduced some artificial element into the results: for example, the application of the assessment to 30-day means may have prevented the baroclinic disturbances of mid-latitudes, which have a typical time-scale of only about 5 days, from showing up strongly. However, the choice of the 30-day period is valid because it is the one used in operational nwp centres for intercomparisons of their results. Other periods, for example 90 days (as suggested by G.White, private communication), may also be appropriate, but it may be doubted whether the results would be much affected by an increase in the period in view of the persistence of the pattern in Figures 5a and 7 from one month to another.

The τ-statistic may be simplified by dispensing with the extra factors used in its computation and considering the simpler statistic μ/σ. For many purposes this would be sufficient. However, the full τ-statistic permits some allowance to be made for the number of degrees of freedom in the population for which μ and σ are calculated: this is helpful when it is required to compare results from two populations with different numbers of degrees of freedom.

In summary, it may be stated that the ratio of the mean error to the variance of the error has been shown to be both a useable and a useful statistic so long as due allowance is made for temporal correlations. Other versions of this ratio, computed over other lengths of time may also be useful. As used here, the ratio has helped to demonstrate that the errors of the 500mb height forecast in the tropics are relatively the largest found over the whole globe, and it has been possible to show similar features in related fields.

It remains to question whether the statistical interpretation can be associated with a corresponding meteorological interpretation. The response must admit that since the strict statistical approach yields results from the data - the model fields - without any a priori meteorological assumptions, any association with meteorological effects could be fortuitous. The meteorological view of the errors will be informed by an awareness of the mechanisms that are at work in the global weather systems. (Also, since the errors of a forecast are being discussed, the meteorological view will be weighted by the usefulness of this forecast). However, so long as a global view is retained, the meteorology of the general circulation is such that activity in the tropics is the means by which incoming energy is re-distributed over the whole globe. It follows that it is likely that a model must treat the tropics correctly if it is also to be able to treat the rest of the globe correctly. In this sense it would appear that the meteorological view coincides with the statistical view that has emerged from this work.

ACKNOWLEDGEMENTS

The benefit of several useful conversations with R.S.Bell, A.C.Lorenc, G.Ross, G.H.White and C.A.Wilson during the course of this work is gratefully acknowledged.

REFERENCES

K.Arpe and A.Simmons 'Systematic Errors of the ECMWF Model and Their Dependence on the Numerical Formulation and Representation of Orographic Effects'. CAS/JSC Working Group on Numerical Experimentation Report No 12: Workshop on Systematic Errors in Models of the Atmosphere, Toronto, Ontario pp3-9 (1988).

K.Arpe, A.Hollingsworth, M.S.Tracton, A.C.Lorenc, S.Uppala and P.Kallberg, 'The Response of Numerical Weather Prediction Systems to FGGE Level IIb Data. Part II: Forecast Verifications and Implications for Predictability'. Quarterly Journal of the Royal Meteorological Society, Vol 111, pp67-101 (1985).

R.S.Bell and A.Dickinson, 'The Meteorological Office Operational Numerical Weather Prediction System'. Meteorological Office Scientific Paper No 41 (1987).

T.W.Bettge, 'A Systematic Error Comparison between the ECMWF and NMC Prediction Models'. Monthly Weather Review, Vol 111, pp2385-2389 (1983).

G.E.Box and G.W.Jenkins, 'Time Series Analysis: Forecasting and Control', 575pp. Pub Holden-Day, San Francisco (1976).

C.E.P.Brooks and N.Carruthers, 'Handbook of Statistical Methods in Meteorology', 415pp. Pub Her Majesty's Stationery Office, London (1953).

E.S.Epstein, 'How Systematic are Systematic Errors?'. American Meteorological Society, Eighth Conference on Numerical Weather Prediction, Baltimore, Maryland: Preprint Volume, pp460-465, (1988).

P.A.Harr, T.L.Tsui and L.R.Brody, 'Identification of Systematic Errors in a Numerical Weather Forecast'. Monthly Weather Review, Vol 111, pp1219-1227 (1983).

R.W.Katz 'Statistical Evaluation of Climate Experiments with General Circulation Models: a Parametric Time Series Modelling Approach'. Journal of Atmospheric Sciences, Vol 39, pp1446-1455 (1982).

R.W.Katz 'Procedures for Determining the Statistical Significance of Changes in Variability Simulated by an Atmospheric General Circulation Model'. Oregon State University, Climate Research Institute, Report No 48 (1983).

A.C.Lorenc, R.S.Bell and B.Macpherson, 'The New Meteorological Office Data Assimilation Scheme'. Met O 11 Technical Note No 27 (1989).

R.W.Preisendorfer and T.P.Barnett, 'Numerical Model - Reality Intercomparison Tests using Small-Sample Statistics'. Journal of the Atmospheric Sciences, Vol 40 pp1884-1896 (1983).

C.E.Weatherburn, 'A First Course in Mathematical Statistics', 277pp. Pub Cambridge University Press (1962).

C.A.Wilson and J.F.B.Mitchell, 'Simulated Climate and CO_2-Induced Climate Change over Western Europe'. Climatic Change, Vol 10, pp11-42 (1987).

G.H.White, 'Systematic Performance of the NMC Medium-Range Forecasts 1985-88'. American Meteorological Society, Eighth Conference on Numerical Weather Prediction, Baltimore, Maryland: Preprint Volume, pp466-471, (1988).

P.W.White and W.L.Adams, 'The United Kingdom Meteorological Office Forecasting Model: Seasonal and Interannual Variability of Systematic Errors and Dependence on Model Formulation'. CAS/JSC Working Group on Numerical Experimentation Report No 12: Workshop on Systematic Errors in Models of the Atmosphere, Toronto, Ontario pp10-26 (1988).

FIGURE LEGENDS

Figure 1. Monthly mean error of the 72-hour forecast of the 500mb height at 12Z over the whole globe during 1988. a) January; b) April; c) July; d) October. Negative areas shaded.

Figure 2. Monthly standard deviation of the 72-hour forecast of the 500mb height at 12Z over the whole globe during 1988. a) January; b) April; c) July; d) October. Areas with standard deviation more than 6 decametres shaded.

Figure 3. Latitude means of the monthly mean and the standard deviation of the error of the forecast of the 500mb height at 12Z during February 1988, shown as functions of latitude.

Figure 4. Global distribution of the time-lagged autocorrelation coefficients for the error of the 72-hour forecast of the 500mb height at 12Z during February 1988. a) First-order; b) Second-order. Negative areas shaded.

Figure 5. Global distribution of the τ-statistic for the time-independent part of the error of the 72-hour forecast of the 500mb height at 12Z during February 1988. a) First-order time-series; b) Second-order time-series.

Figure 6. Population distribution of the τ-statistic for the error of the 72-hour forecast of the 500mb height at 12Z during February 1988. a) Time-independent part of the first-order time-series; b) Time-independent part of the second-order time-series; c) Unadjusted time-series.

Figure 7. Population distribution of the τ-statistic for the time-independent part of the first-order time-series for the error of the 72-hour forecast of the 500mb height at 12Z during May 1988.

Figure 8. Global distribution of the τ-statistic for the time-independent part of the first-order time-series for the error of the 72-hour forecast of the 500mb height at 12Z during May 1988. Negative areas shaded, areas with $\tau<-5$ shown darker.

Figure 9. Global distribution of the τ-statistic for the independent part of the first-order time-series for the error of the 72-hour forecast of the 500mb height at 00Z during February 1988. Negative areas shaded, areas with $\tau<-5$ shown darker.

Figure 10. Population distribution of the τ-statistics for the independent part of the first-order time-series for the error of the 72-hour forecasts at 12Z during May 1988. a) Pmsl; b) Temperature at 850mb; c) Temperature at 700mb.

Figure 11. Global distribution of the τ-statistic for the time-independent part of the first-order time-series for the error of the 72-hour forecast at 12Z during May 1988. a) Pmsl; b) Temperature at 850mb; c) Temperature at 700mb. Negative areas shaded, areas with $\tau<-5$ shown darker.

TABLES

Month	Jan	Feb	Mar	Apr	May	Jun	Jly	Aug	Sep	Oct	Nov	Dec
N	30	29	28	30	31	29	31	29	30	31	26	31
ν	28	27	26	28	29	27	29	27	28	29	24	29
$\tau<-5$	2317	1671	2177	3258	2571	2703	2742	2762	2596	2671	2923	1786
$\tau>+5$	178	342	104	77	16	88	160	135	113	38	6	0

Table 1: Monthly data for τ at 12Z, 1988

Month	Feb 1988 uncorrected	Feb 1988 2nd-order	Dec 1987 1st-order	Feb 1988 1st-order
Time	12Z	12Z	12Z	00Z
N	29	29	30	29
ν	28	26	28	27
$\tau<-5$	2802	1760	2537	2190
$\tau>+5$	685	381	30	328

Table 2: Variations of monthly data for τ

Field Level	Pressure Mean Sea-Level	Temperature 850mb	Temperature 700mb
$\tau<-5$	1447	1864	2479
$\tau>+5$	869	332	165

Table 3: Monthly data for τ, various fields, 12Z May 1988

MEAN VALUE 3DY F/C ERRORS H500. 12Z. JAN88
CONTOUR INTERVAL=2.0 DAM
VALID AT 12Z ON 1/1/1988
LEVEL: 500 MB EXPERIMENT NO.: 1 T+72

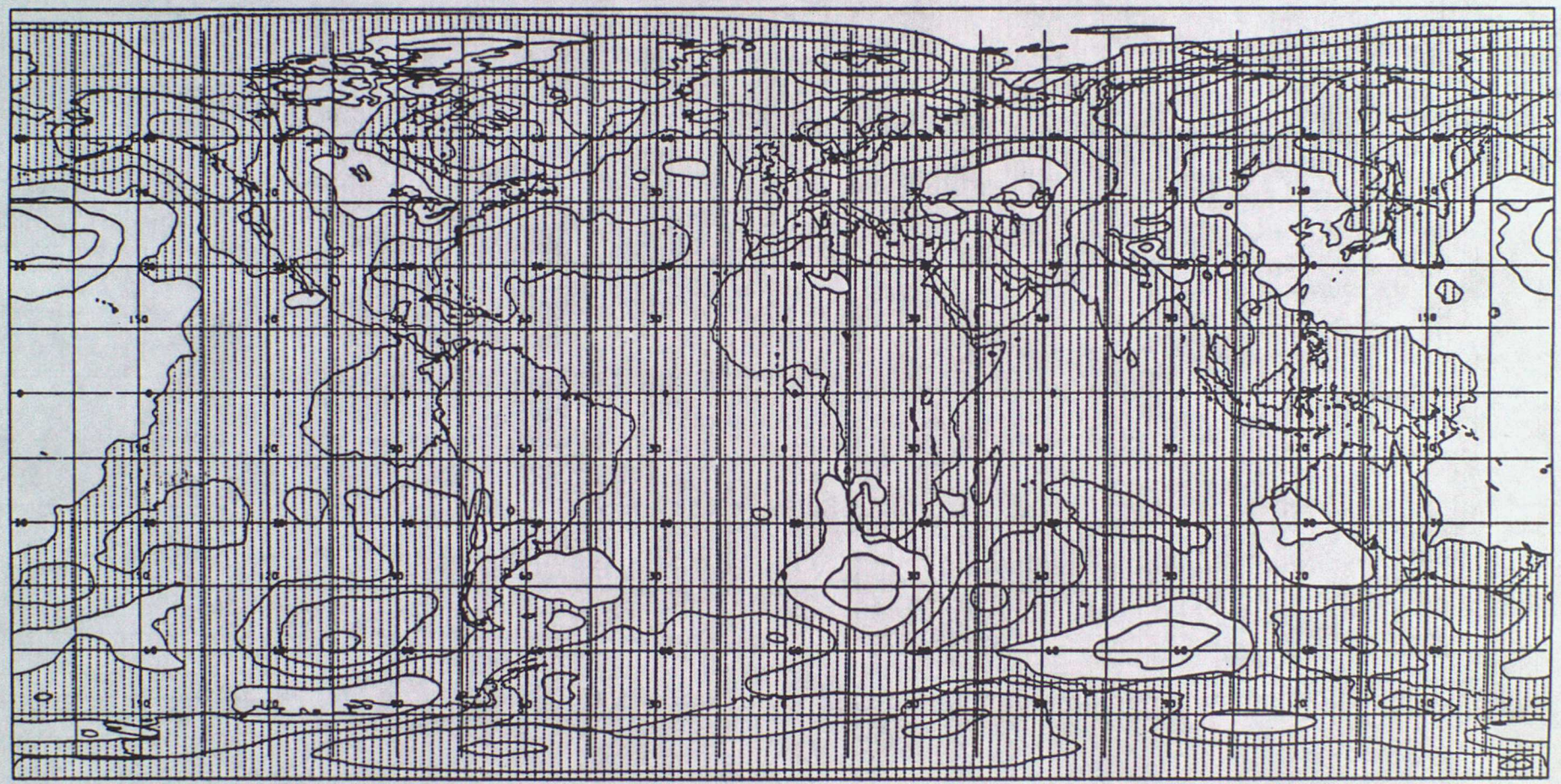


Figure 1a.

MEAN VALUE 3DY F/C ERRORS H500. 12Z. APR88
CONTOUR INTERVAL=2.0 DAM
VALID AT 12Z ON 1/4/1988
LEVEL: 500 MB EXPERIMENT NO.: 1 T+72

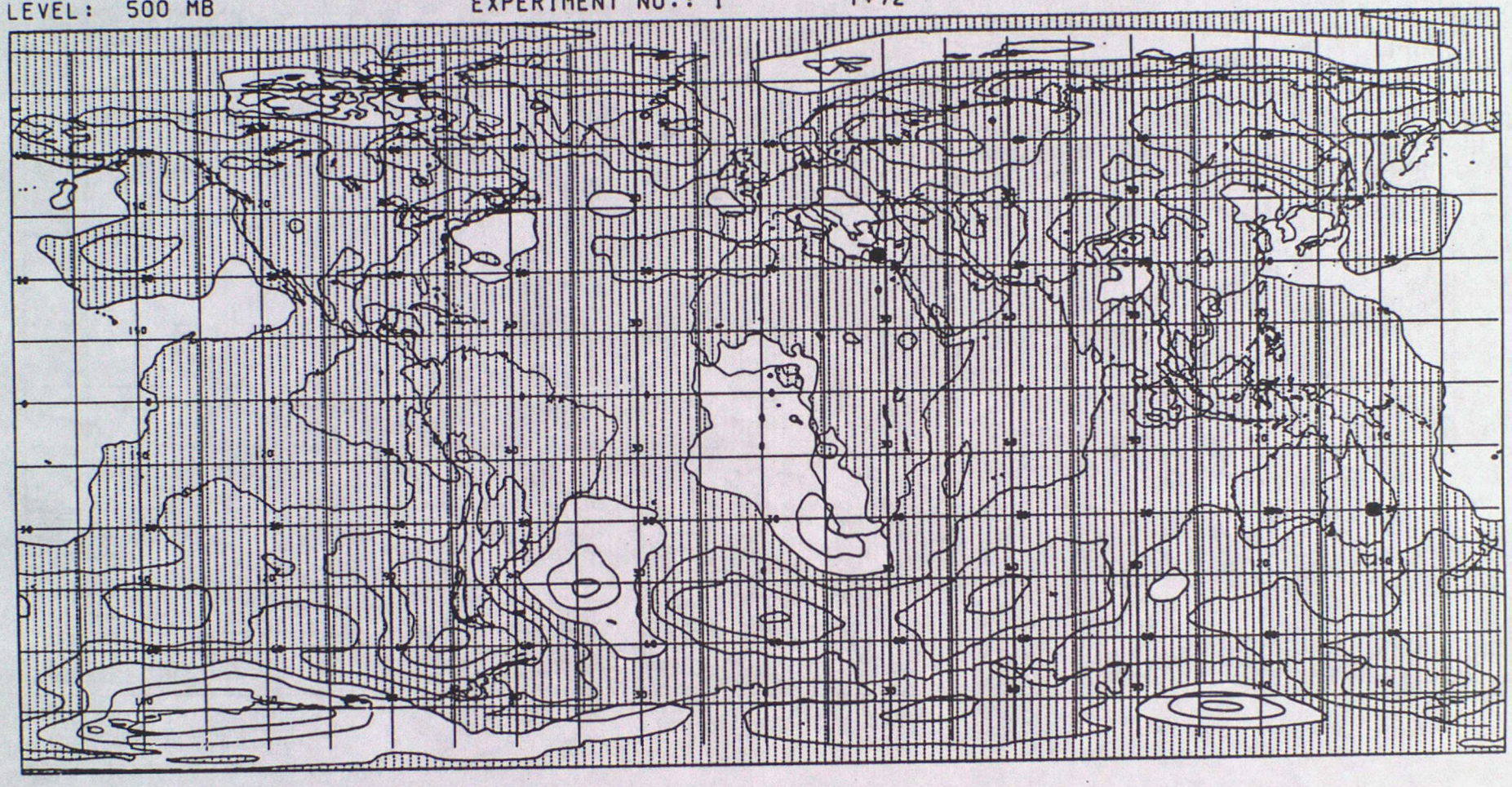


Figure 1b

MEAN VALUE 3DY F/C ERRORS H500. 12Z. JLY88
CONTOUR INTERVAL=2.0 DAM
VALID AT 12Z ON 1/7/1988
LEVEL: 500 MB EXPERIMENT NO.: 1 T+72

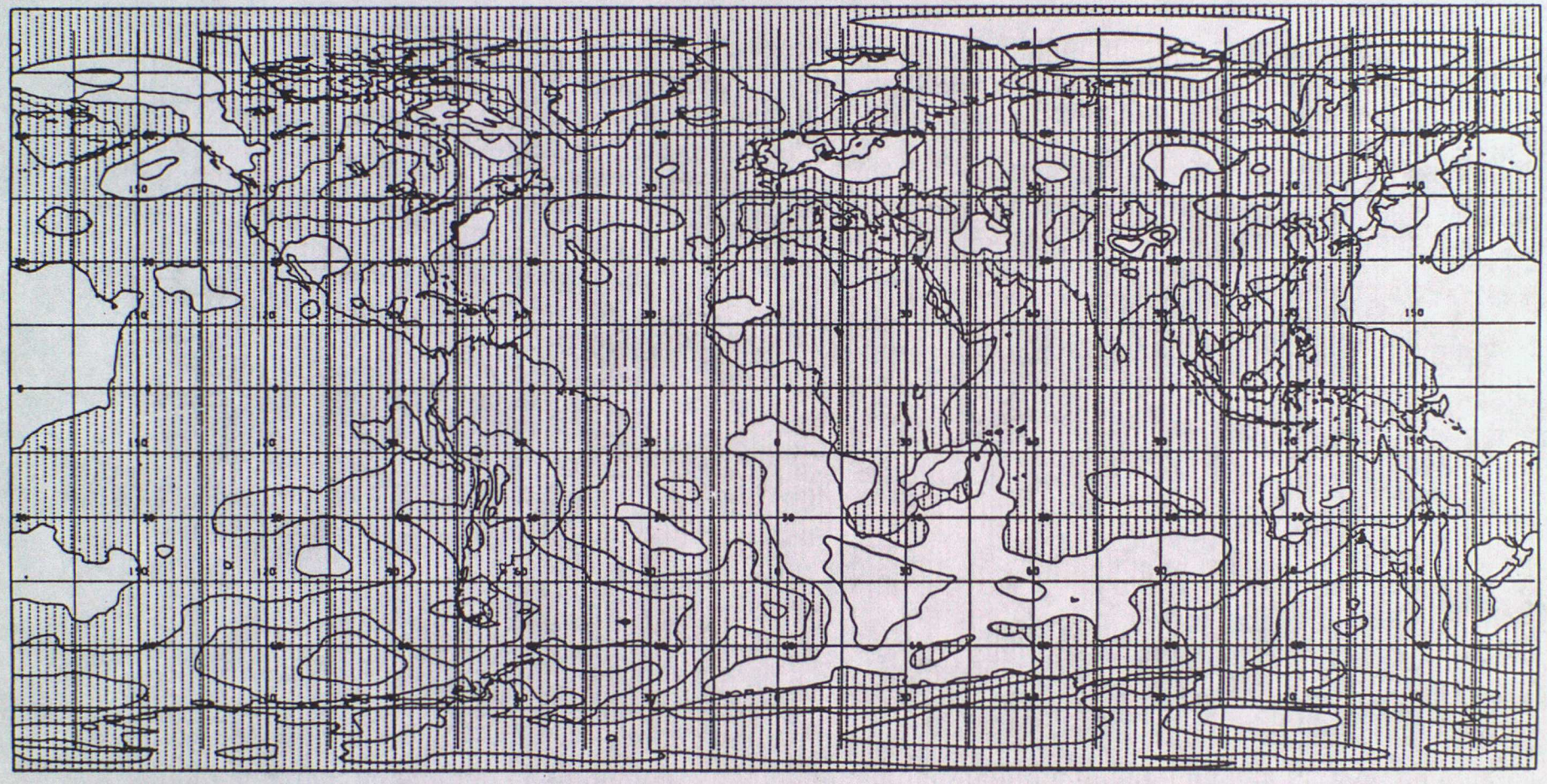


Figure 1c

MEAN VALUE 3DY F/C ERRORS H500. 12Z. OCT88
CONTOUR INTERVAL=2.0 DAM
VALID AT 12Z ON 1/10/1988
LEVEL: 500 MB EXPERIMENT NO.: 1 T+72

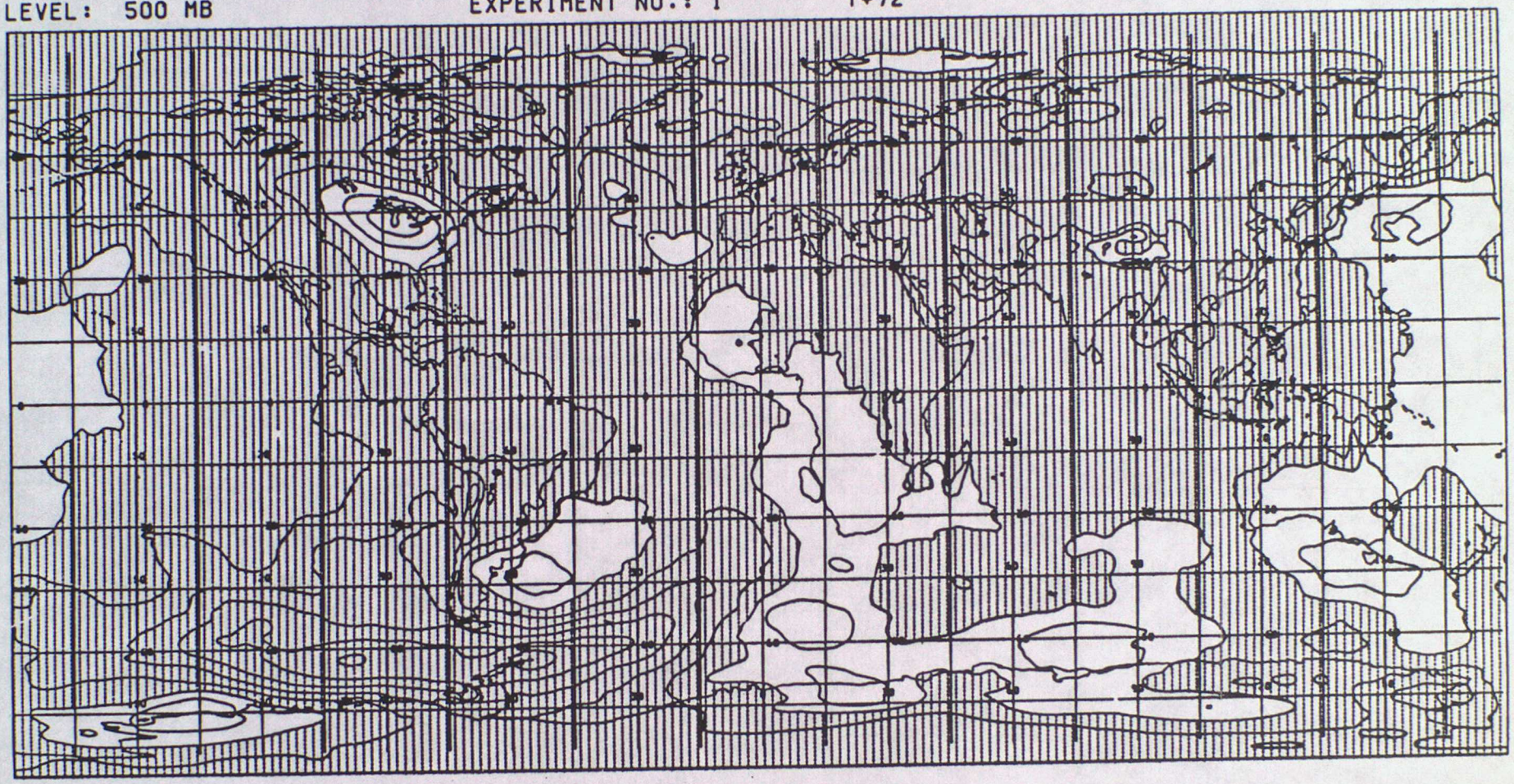


Figure 1d

STANDARD DEVIATION OF 3DY F/C ERRORS H500. 12Z. JAN88
CONTOUR INTERVAL=2.0 DAM
AVERAGE FROM 12Z ON 1/1/1988 TO 12Z ON 31/1/1988
LEVEL: 500 MB EXPERIMENT NO.: 1 T+72

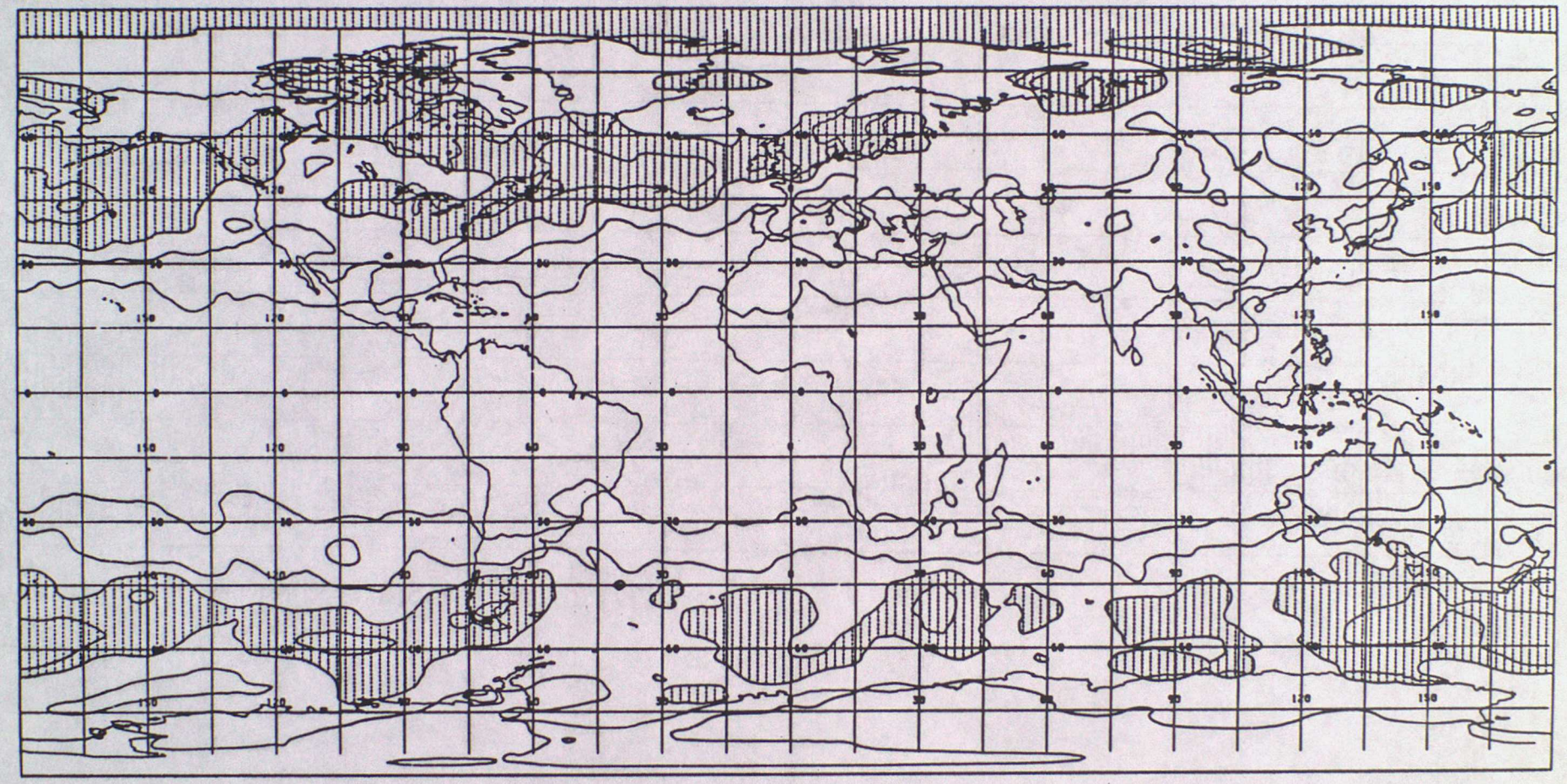


Figure 2a.

STANDARD DEVIATION OF 3DY F/C ERRORS H500. 12Z. APR88
CONTOUR INTERVAL=2.0 DAM
AVERAGE FROM 12Z ON 1/4/1988 TO 12Z ON 30/4/1988
LEVEL: 500 MB EXPERIMENT NO.: 1 T+72

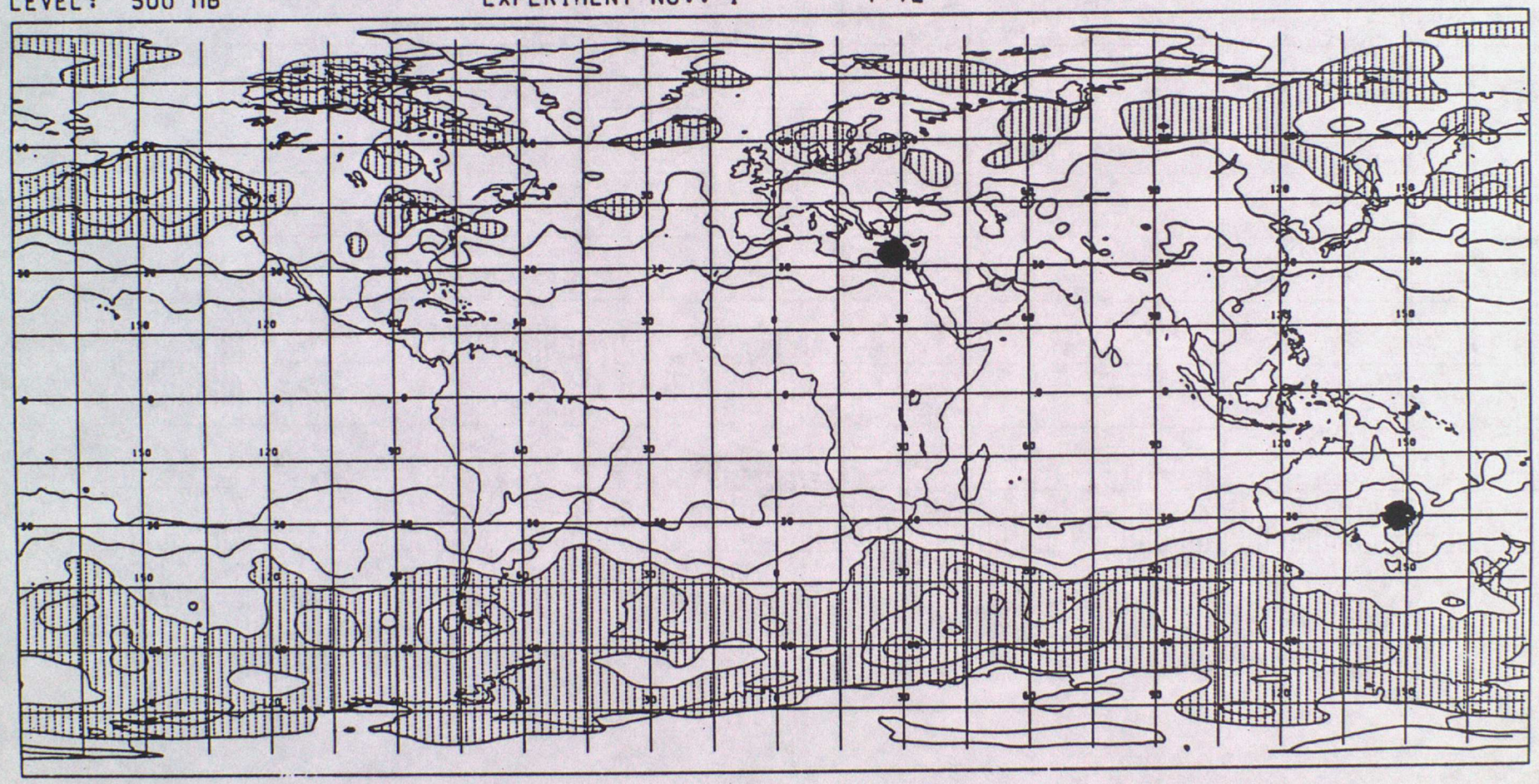


Figure 2b

STANDARD DEVIATION OF 3DY F/C ERRORS H500. 12Z. JLY88
CONTOUR INTERVAL=2.0 DAM
AVERAGE FROM 12Z ON 1/7/1988 TO 12Z ON 31/7/1988
LEVEL: 500 MB EXPERIMENT NO.: 1 T+72

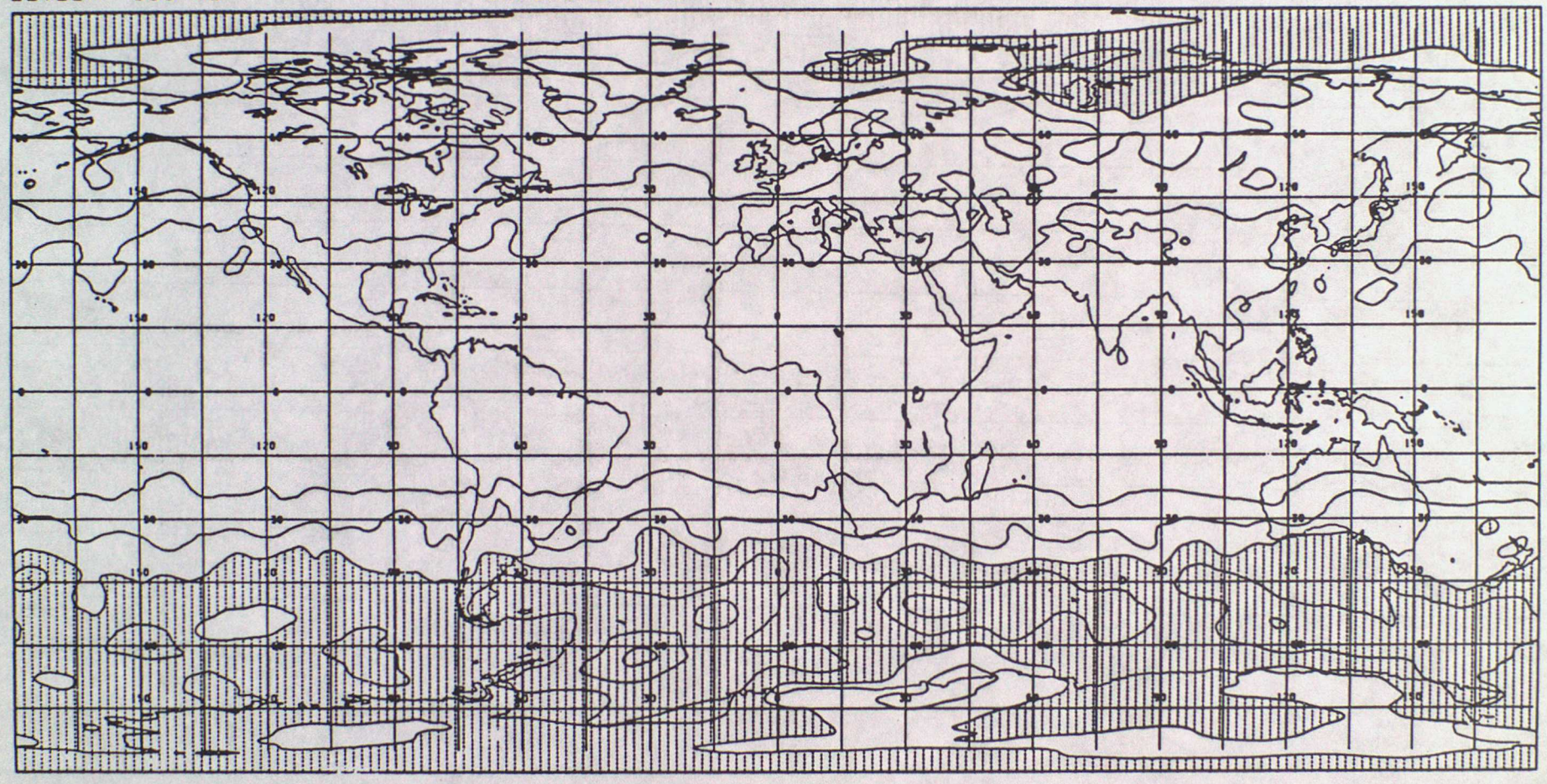


Figure 2c

STANDARD DEVIATION OF 3DY F/C ERRORS H500. 12Z. OCT88
CONTOUR INTERVAL=2.0 DAM
AVERAGE FROM 12Z ON 1/10/1988 TO 12Z ON 31/10/1988
LEVEL: 500 MB EXPERIMENT NO.: 1 T+72

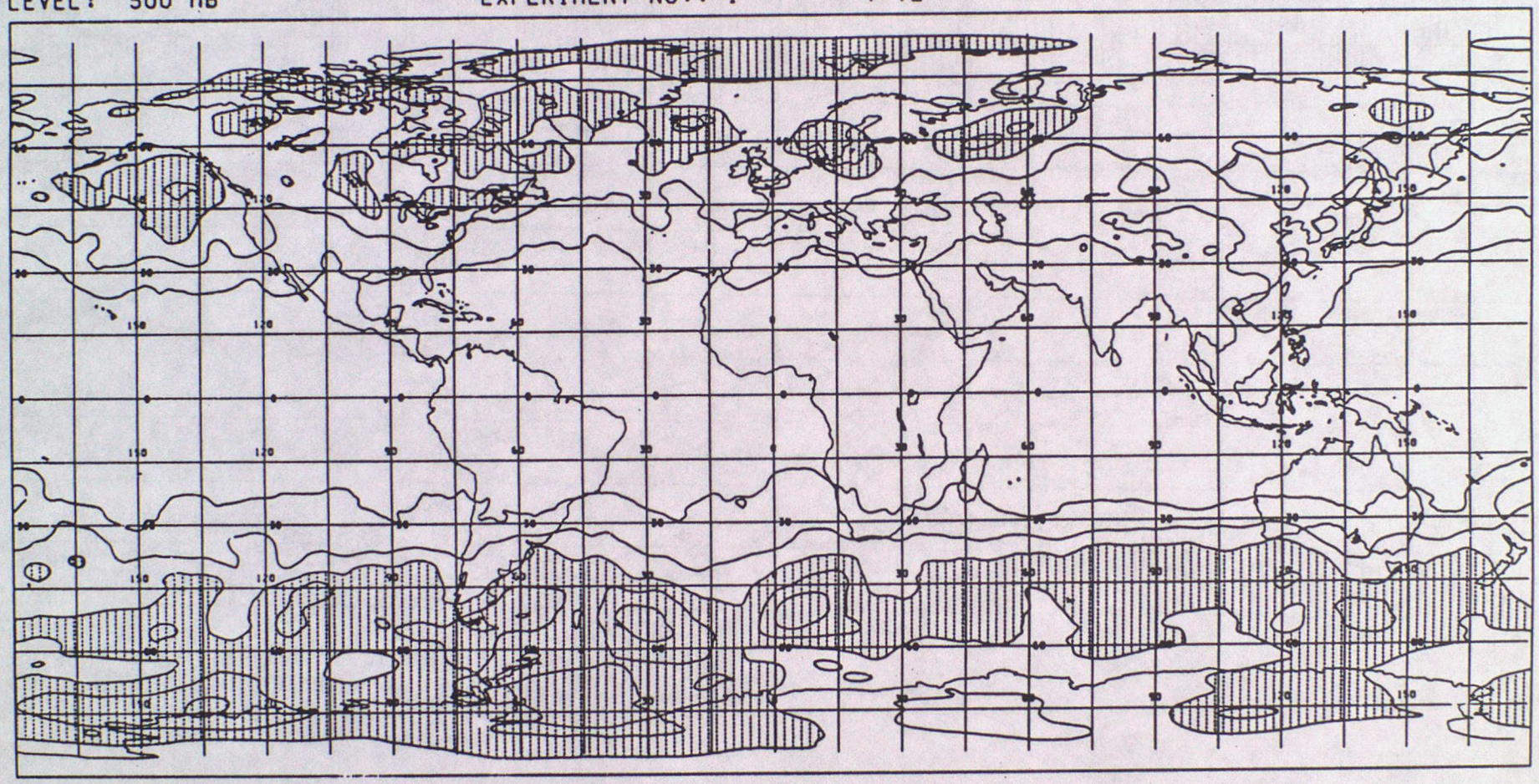


Figure 2d

MEAN & STANDARD DEVIATION OF THE ERROR OF THE H500 T+72 FORECAST, 12Z, FEB 88

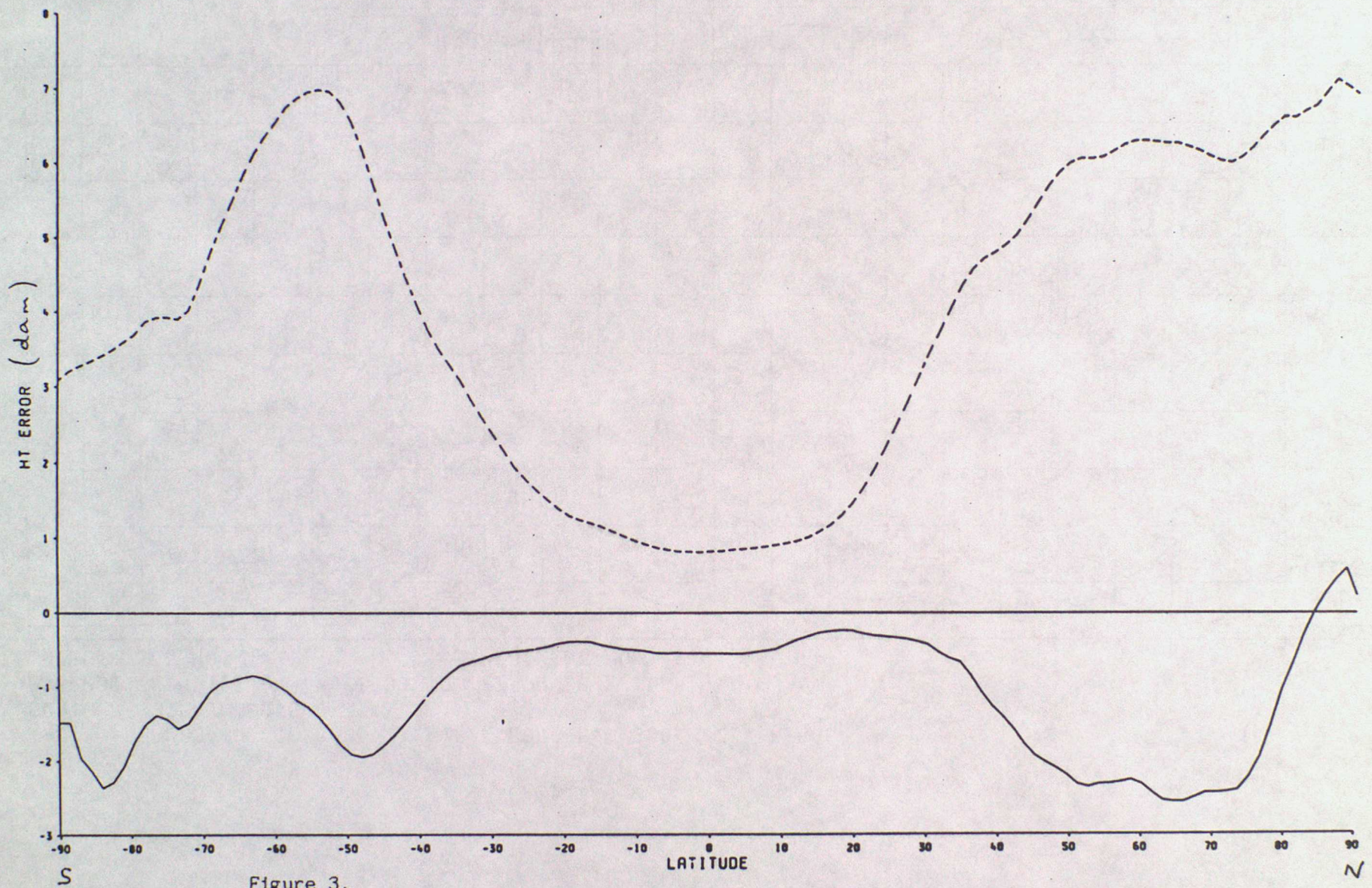


Figure 3.

1-DY TIME-LAG AUTOCORR OF T+72 F/C ERRORS. H500. 12Z. FEB88
CONTOUR INTERVAL=0.5
AVERAGE FROM 12Z ON 1/2/1988 TO 12Z ON 28/2/1988
LEVEL: 500 MB EXPERIMENT NO.: 1 T+72

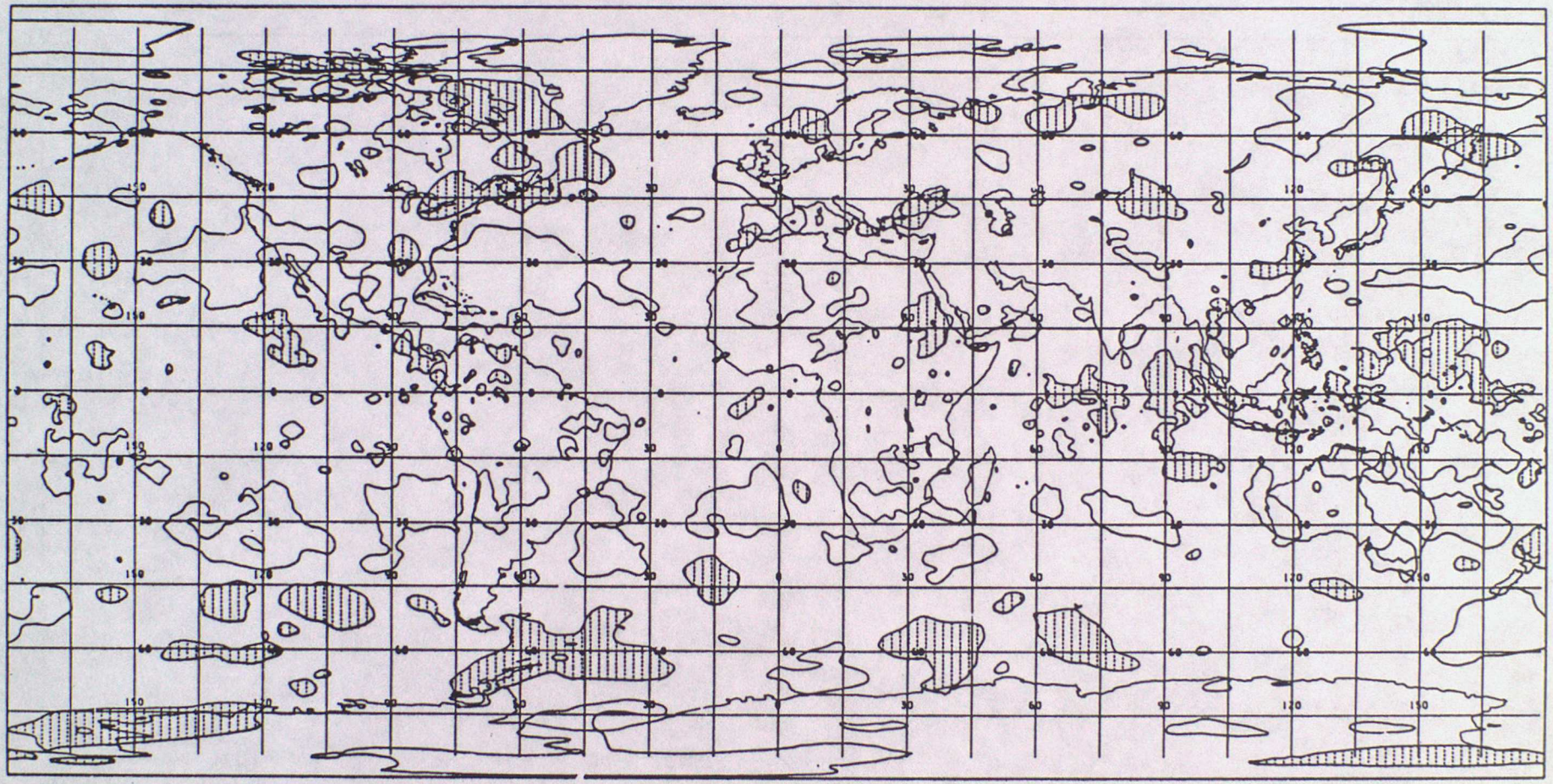


Figure 4a

2-DY TIME-LAG AUTOCORR OF T+72 F/C ERRORS. H500. 12Z. FEB88
CONTOUR INTERVAL=0.5
AVERAGE FROM 12Z ON 1/2/1988 TO 12Z ON 27/2/1988
LEVEL: 500 MB EXPERIMENT NO.: 1 T+72

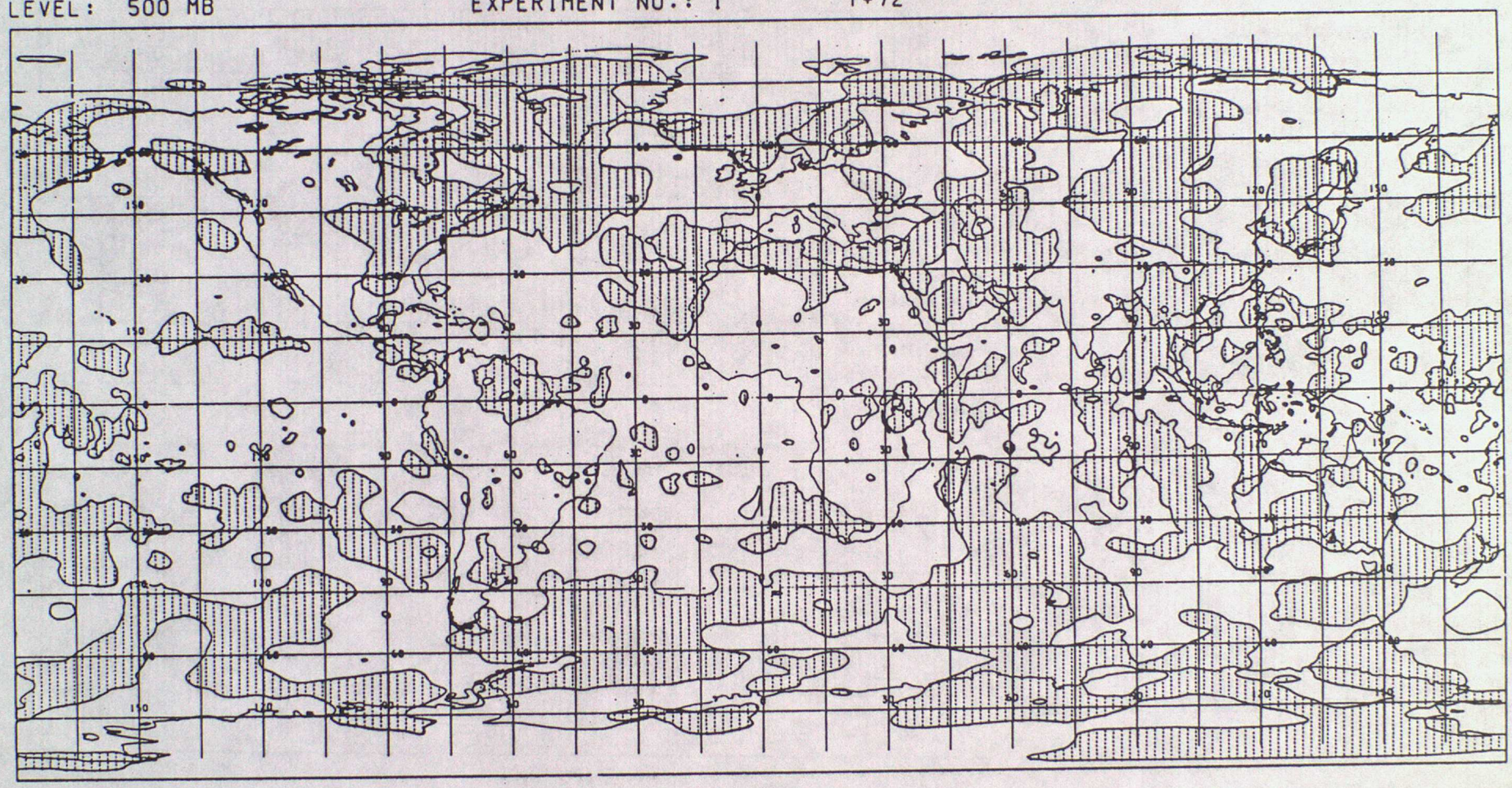


Figure 4b

T-STATISTIC. RANDOM PART OF 3DY F/C ERRORS H500. 12Z. FEB88
CONTOUR INTERVAL=5.0
VALID AT 12Z ON 1/2/1988
LEVEL: 500 MB EXPERIMENT NO.: 1 T+72

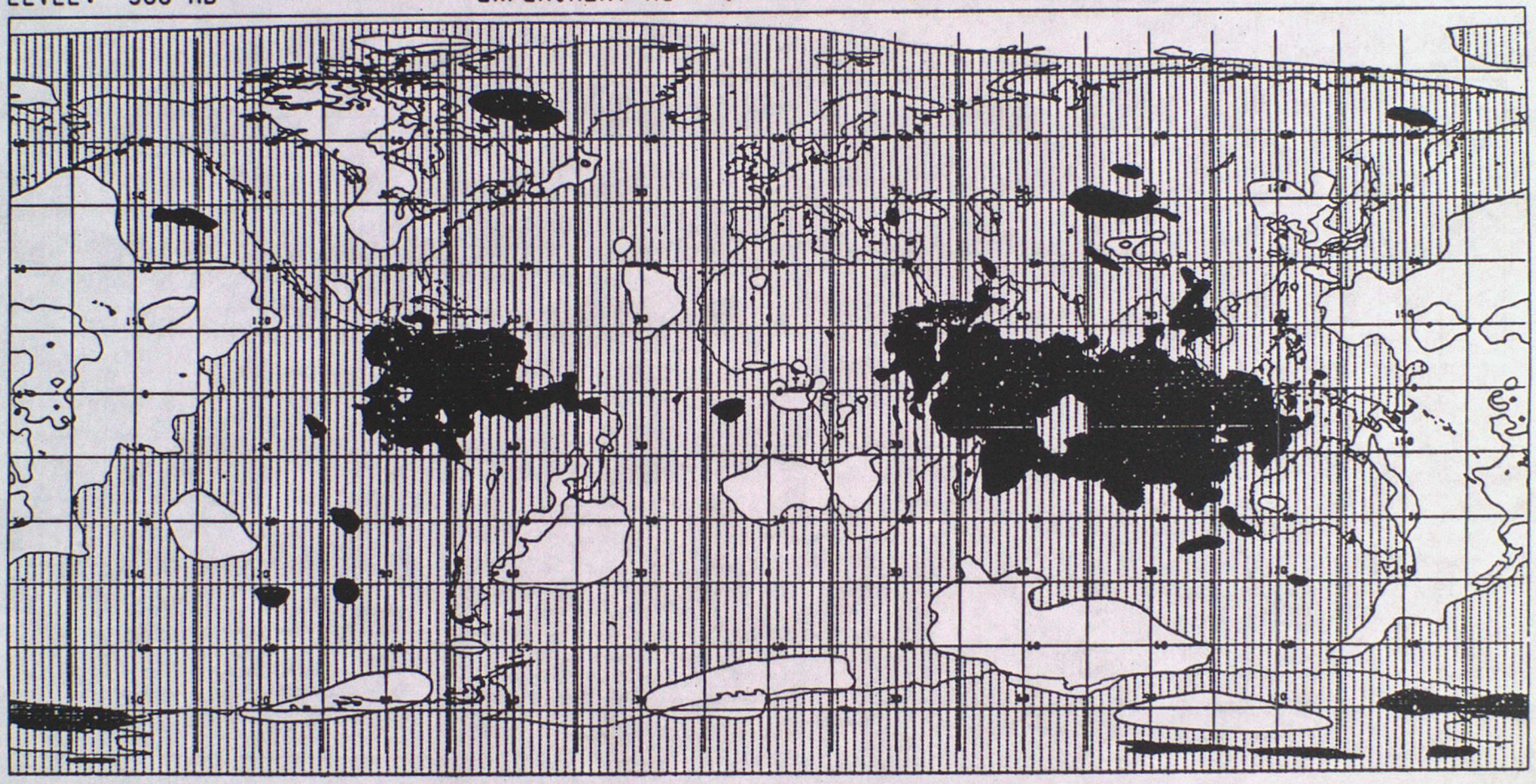


Figure 5a

2ND-ORDER T. RANDOM PART OF 3DY F/C ERRORS H500. 12Z. FEB88
CONTOUR INTERVAL = 5.0
VALID AT 12Z ON 1/2/1988
LEVEL: 500 MB EXPERIMENT NO.: 1 T+72

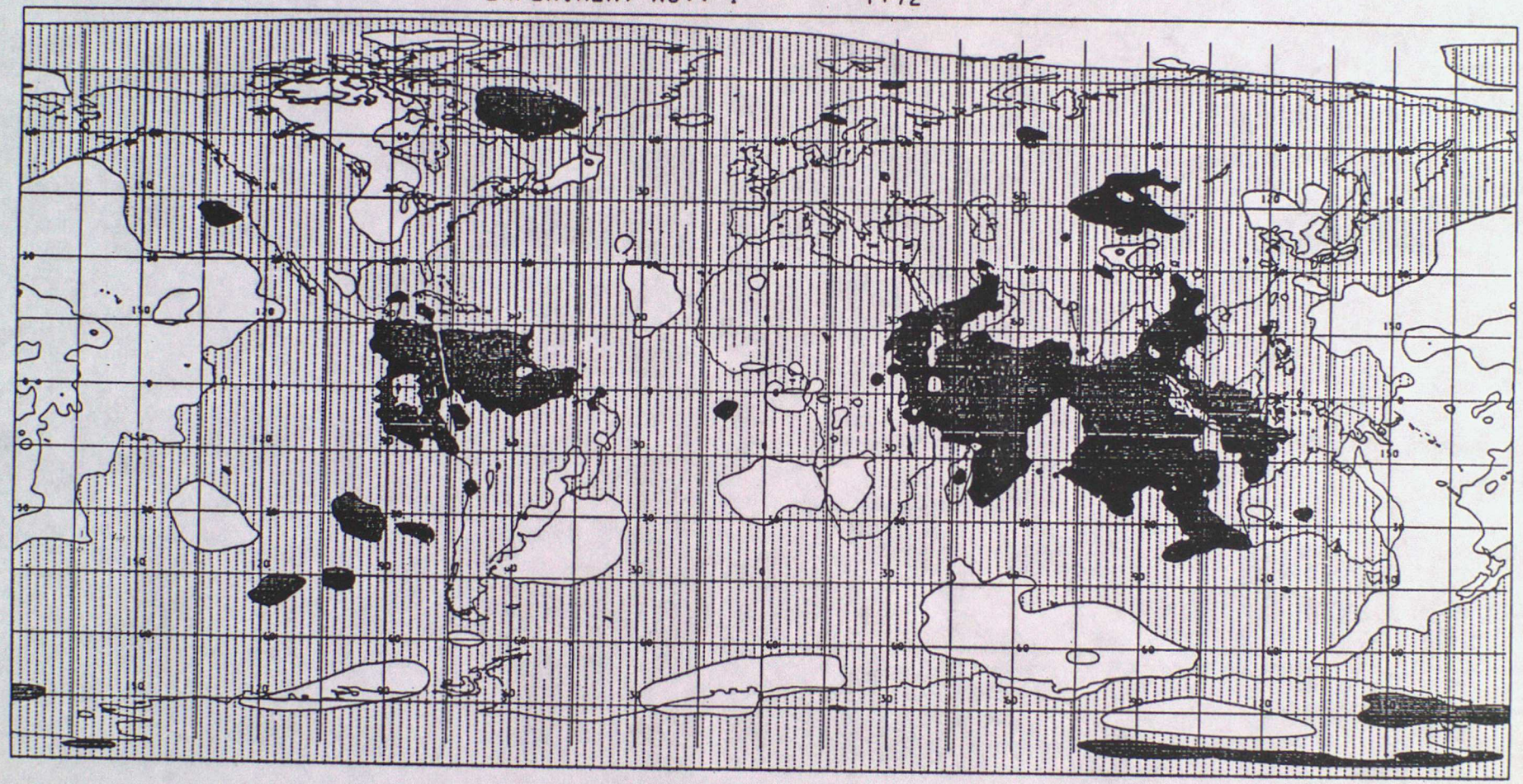


Figure 5b

Figure 6a.

DISTRIBUTION OF T-STATISTIC. 500MB F/C ERROR. 12Z FEB88

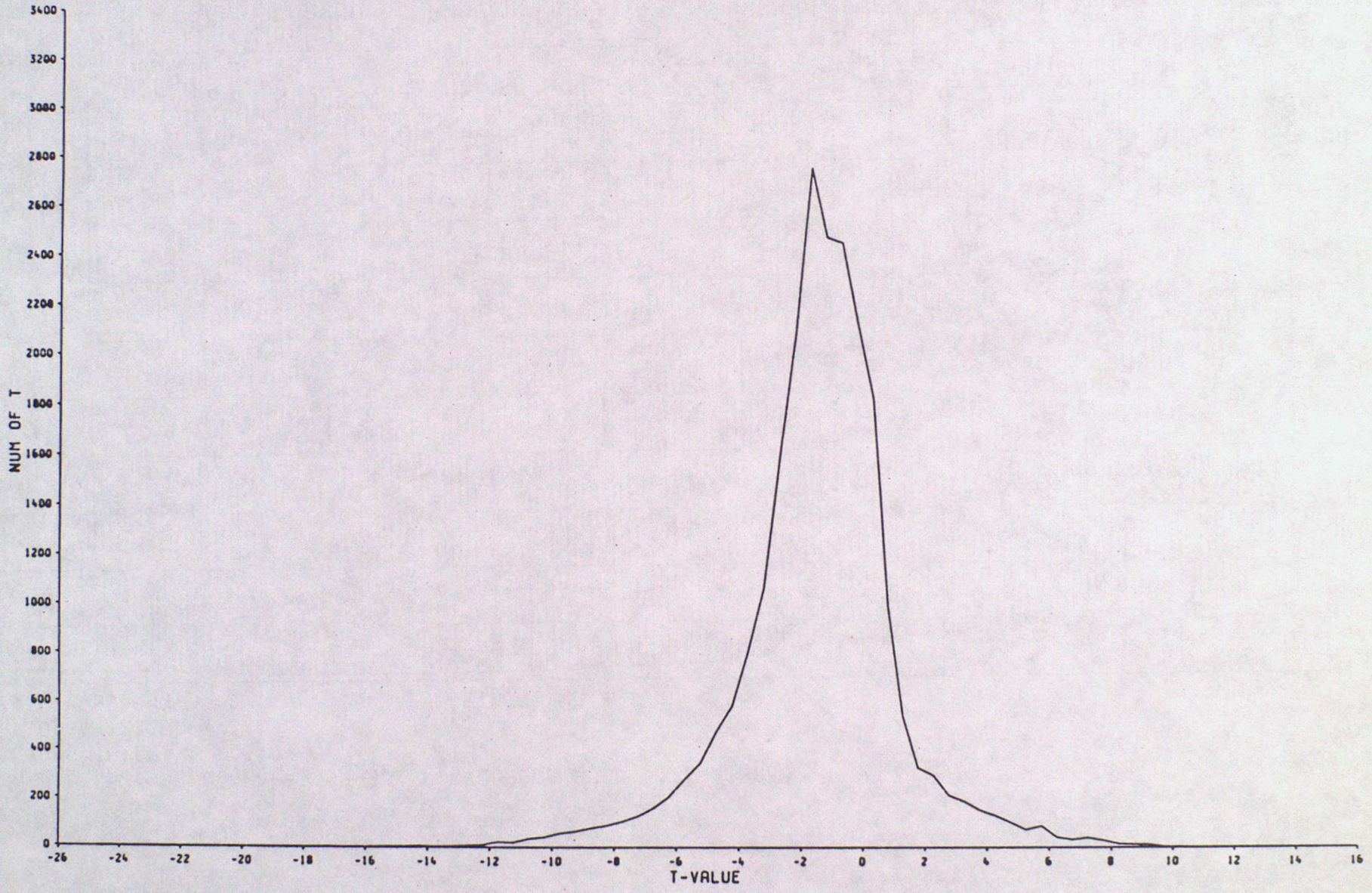


Figure 6b

DISTRIBUTION OF 2-DY T-STAT. 500MB F/C ERROR. 12Z FEB88

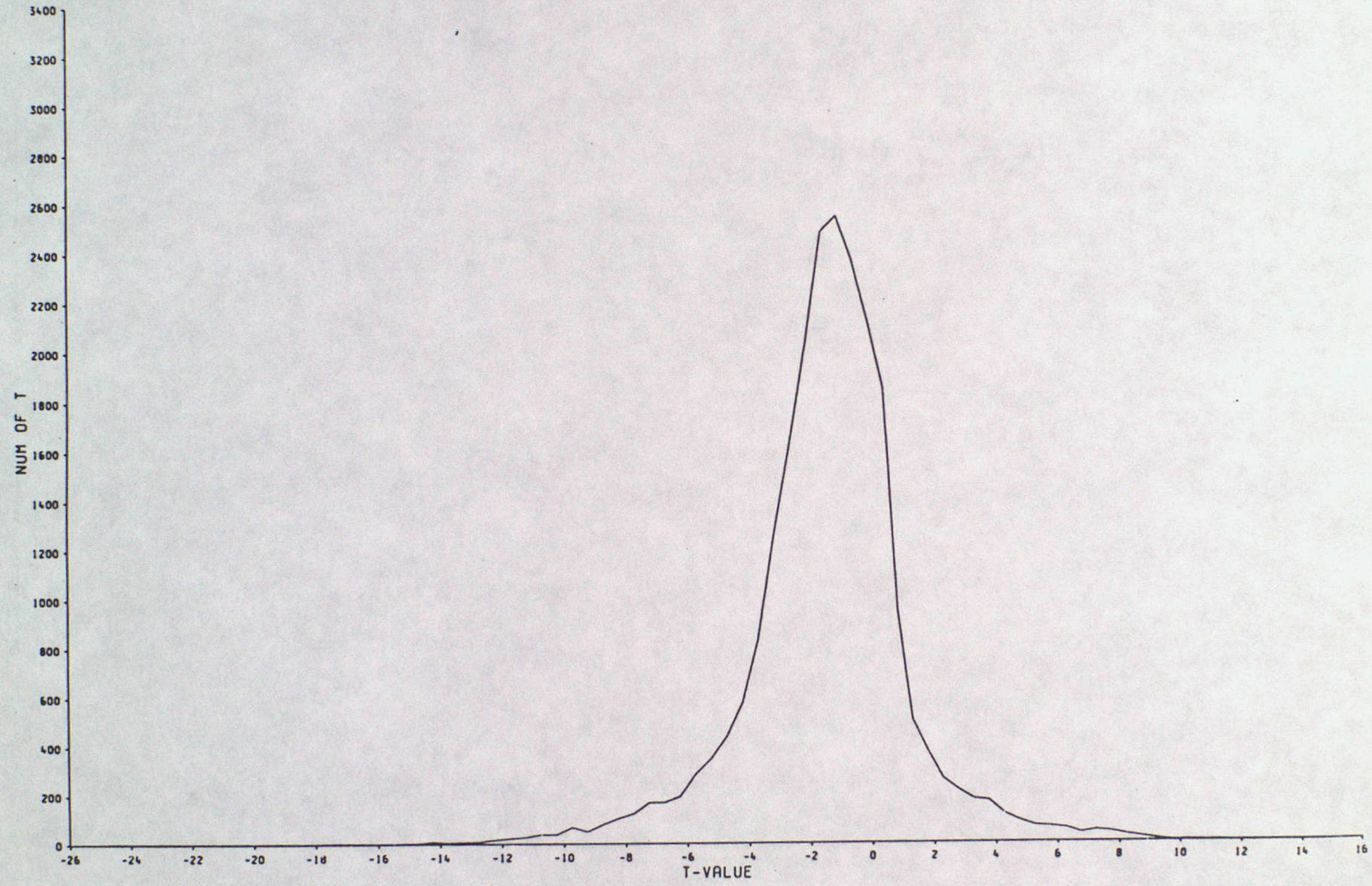


Figure 6c

DISTRIBUTION OF UNCORRECTED τ-STATISTIC, H500, T+72 FORECAST ERROR, 12Z FEB 88

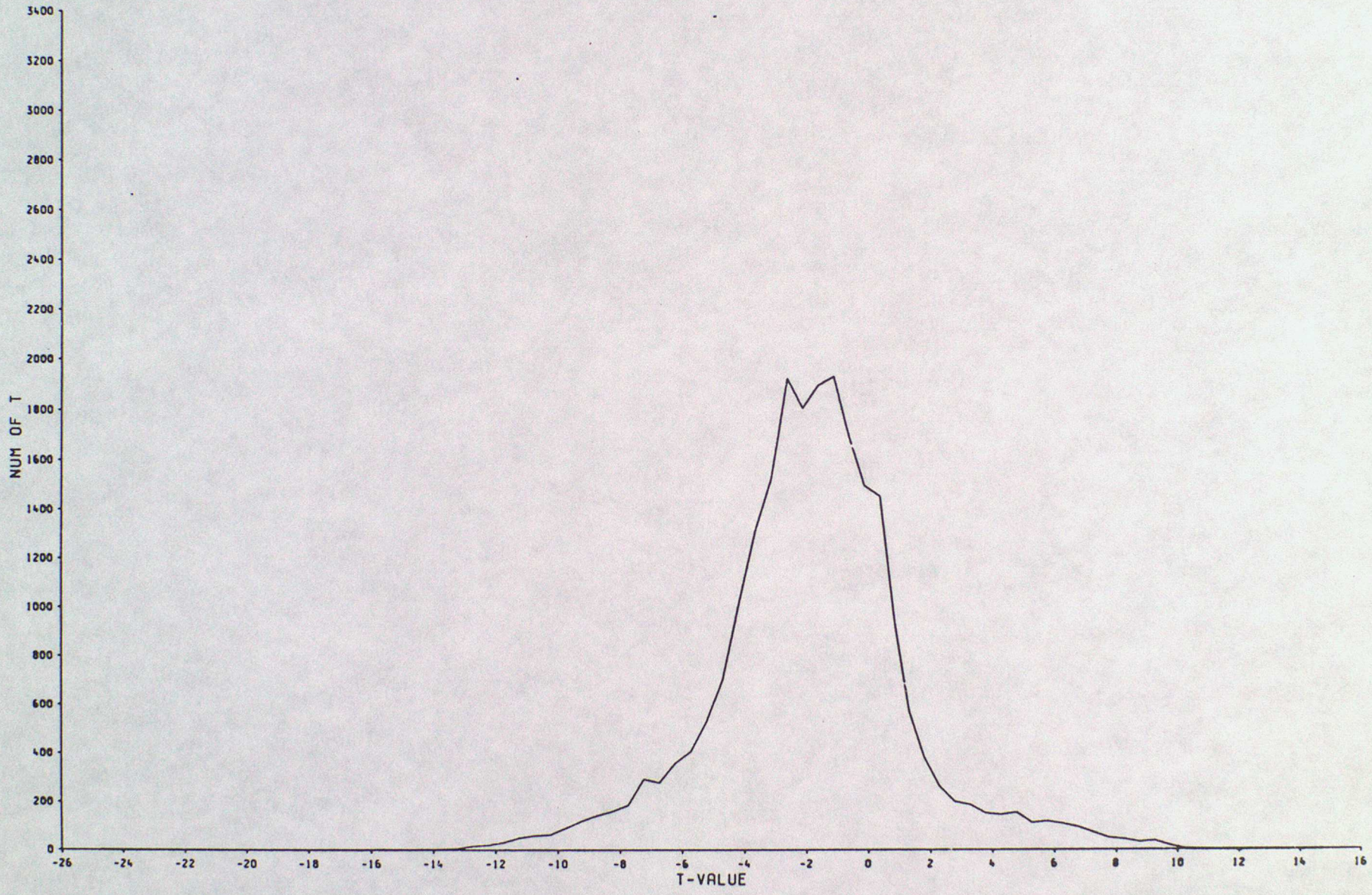
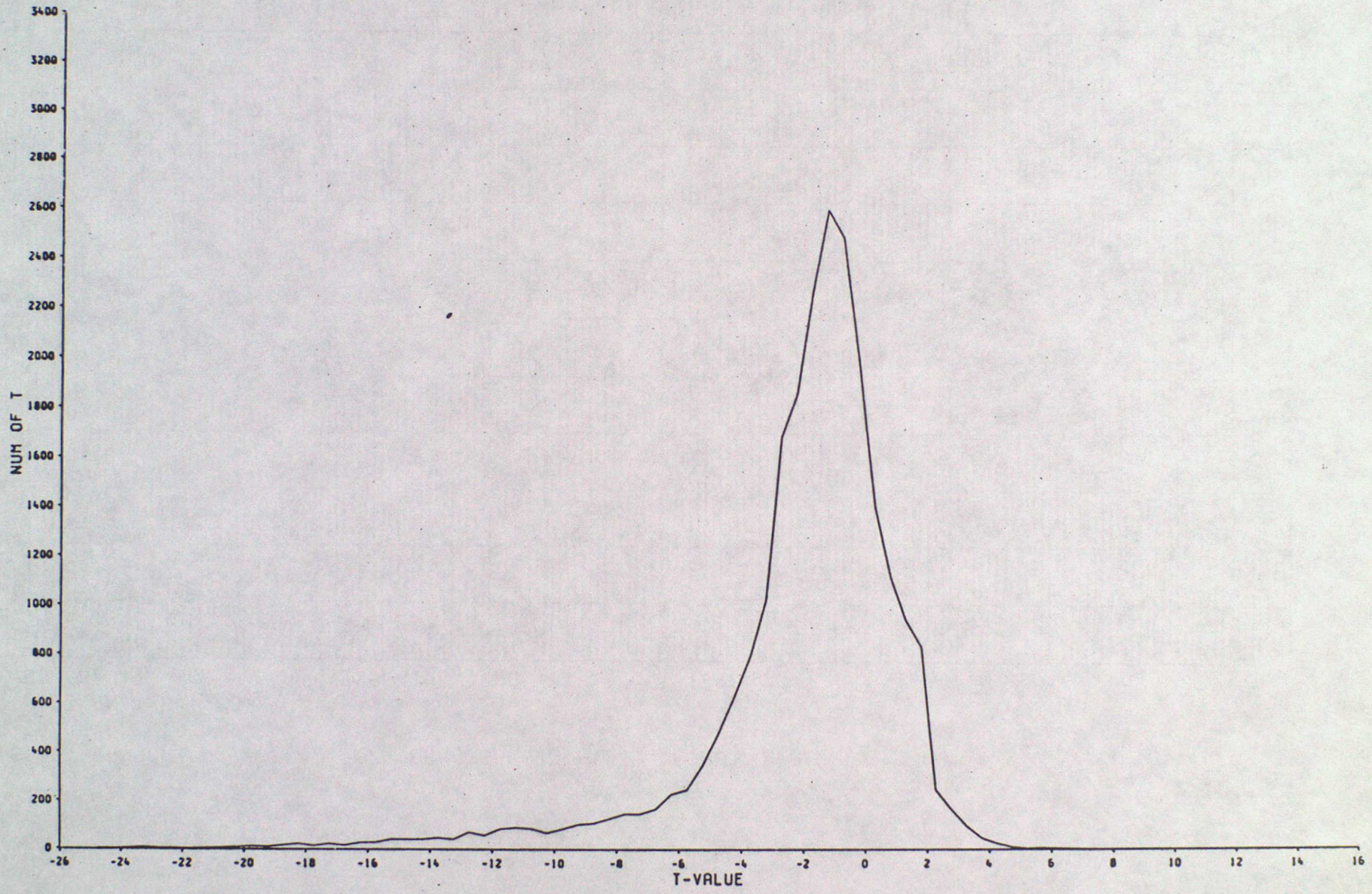


Figure 7.

DISTRIBUTION OF T-STATISTIC. 500MB F/C ERROR. 12Z MAY88



T-STATISTIC. RANDOM PART OF 3DY F/C ERRORS H500. 12Z. MAY88
CONTOUR INTERVAL=5.0
VALID AT 12Z ON 1/5/1988
LEVEL: 500 MB EXPERIMENT NO.: 1 T+72

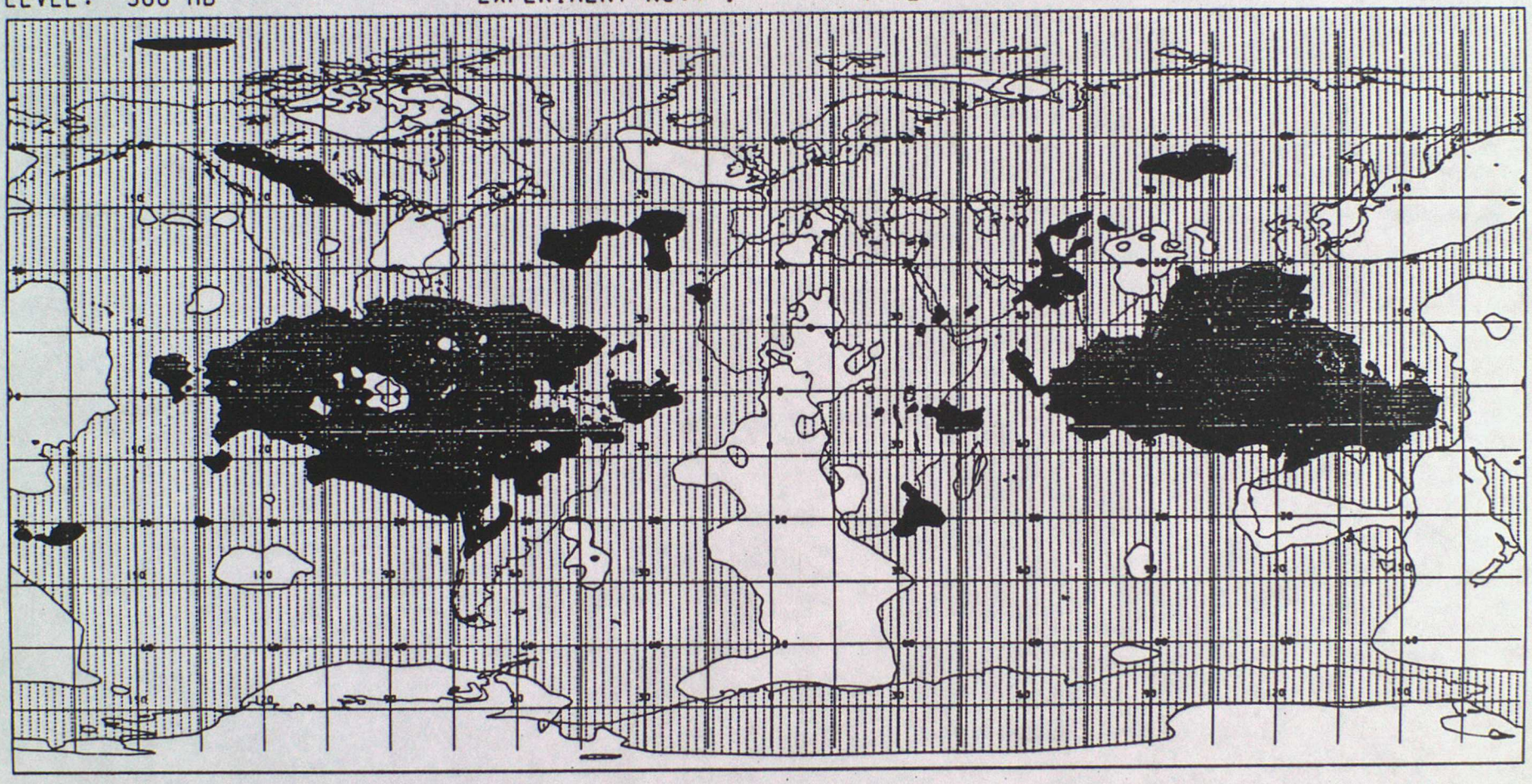


Figure 8

T-STATISTIC RANDOM PART OF 3DY F/C ERRORS H500. 00Z. FEB88
CONTOUR INTERVAL = 5.0
VALID AT 0Z ON 1/2/1988
LEVEL: 500 MB EXPERIMENT NO.: 1 T+72

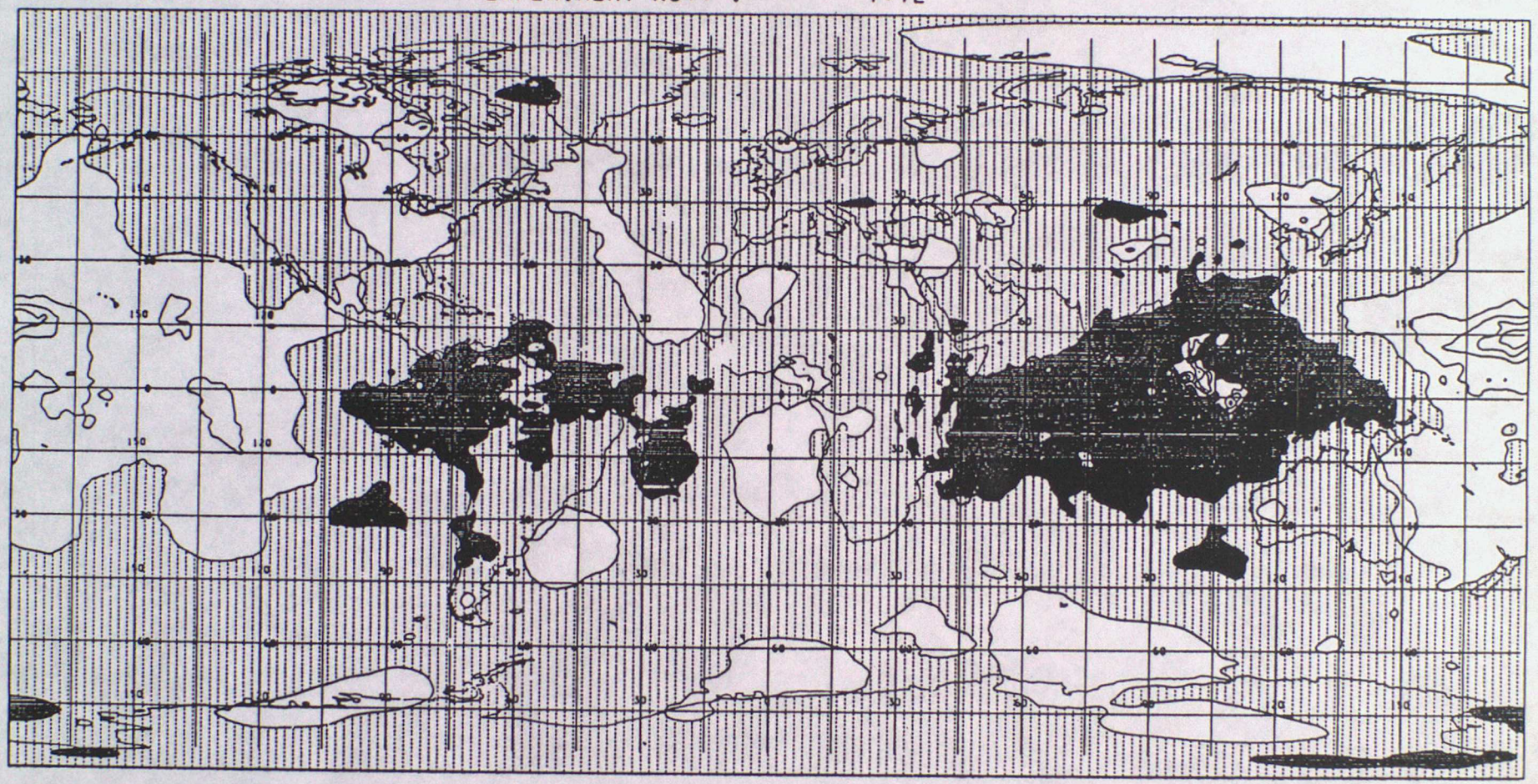


Figure 9

Figure 10a.

DISTRIBUTION OF T-STATISTIC. PMSL F/C ERROR. 12Z MAY88

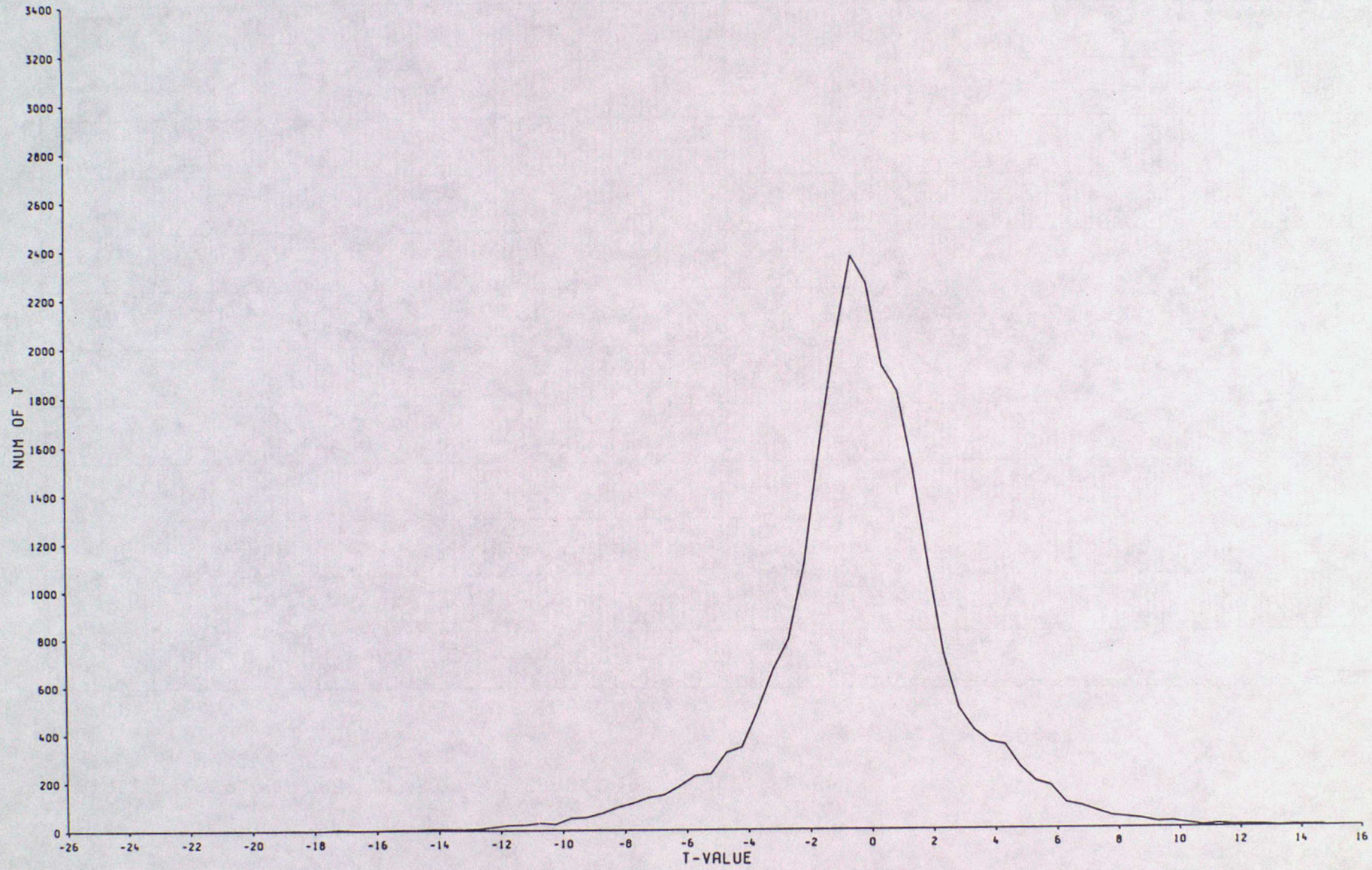


Figure 10b

DISTRIBUTION OF T-STATISTIC, T850 F/C ERROR, 12Z MAY88

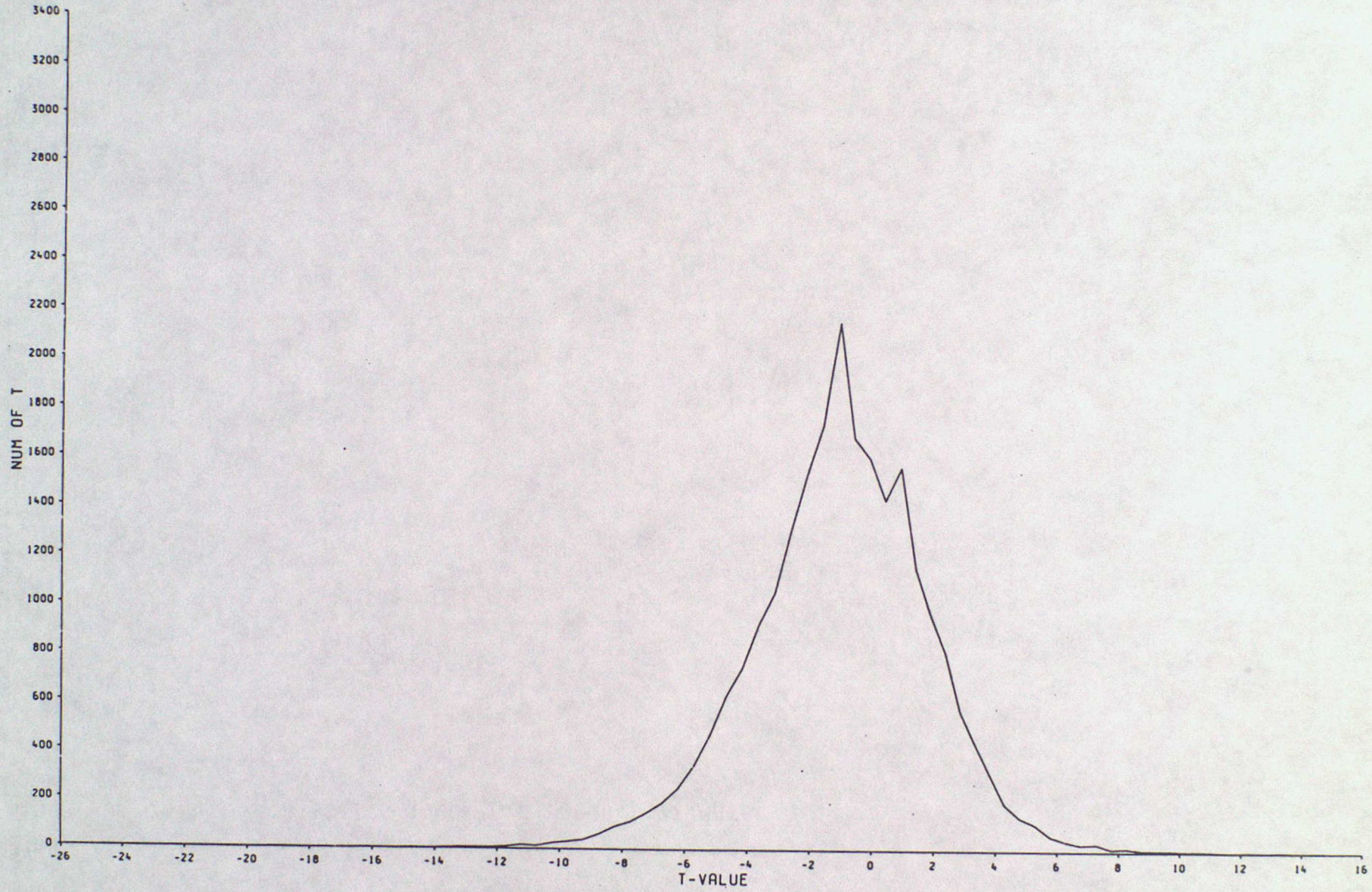
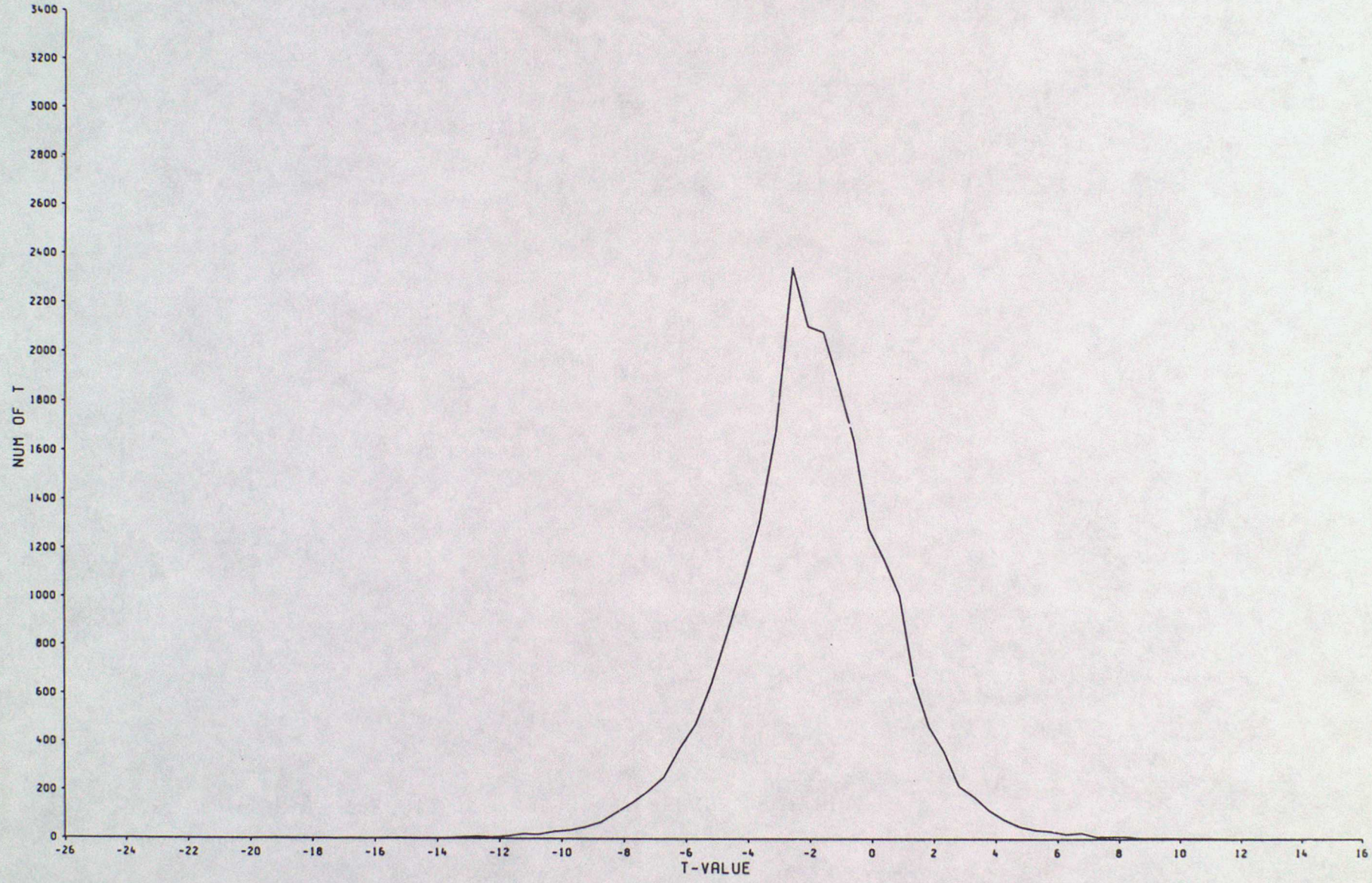


Figure 10c

DISTRIBUTION OF T-STATISTIC, T700 F/C ERROR, 12Z MAY88



T-STATISTIC RANDOM PART OF 3DY F/C ERRORS PMSL. 12Z. MAY88
CONTOUR INTERVAL = 5.0
VALID AT 12Z ON 1/5/1988
SEA LEVEL

EXPERIMENT NO.: 1 T+72

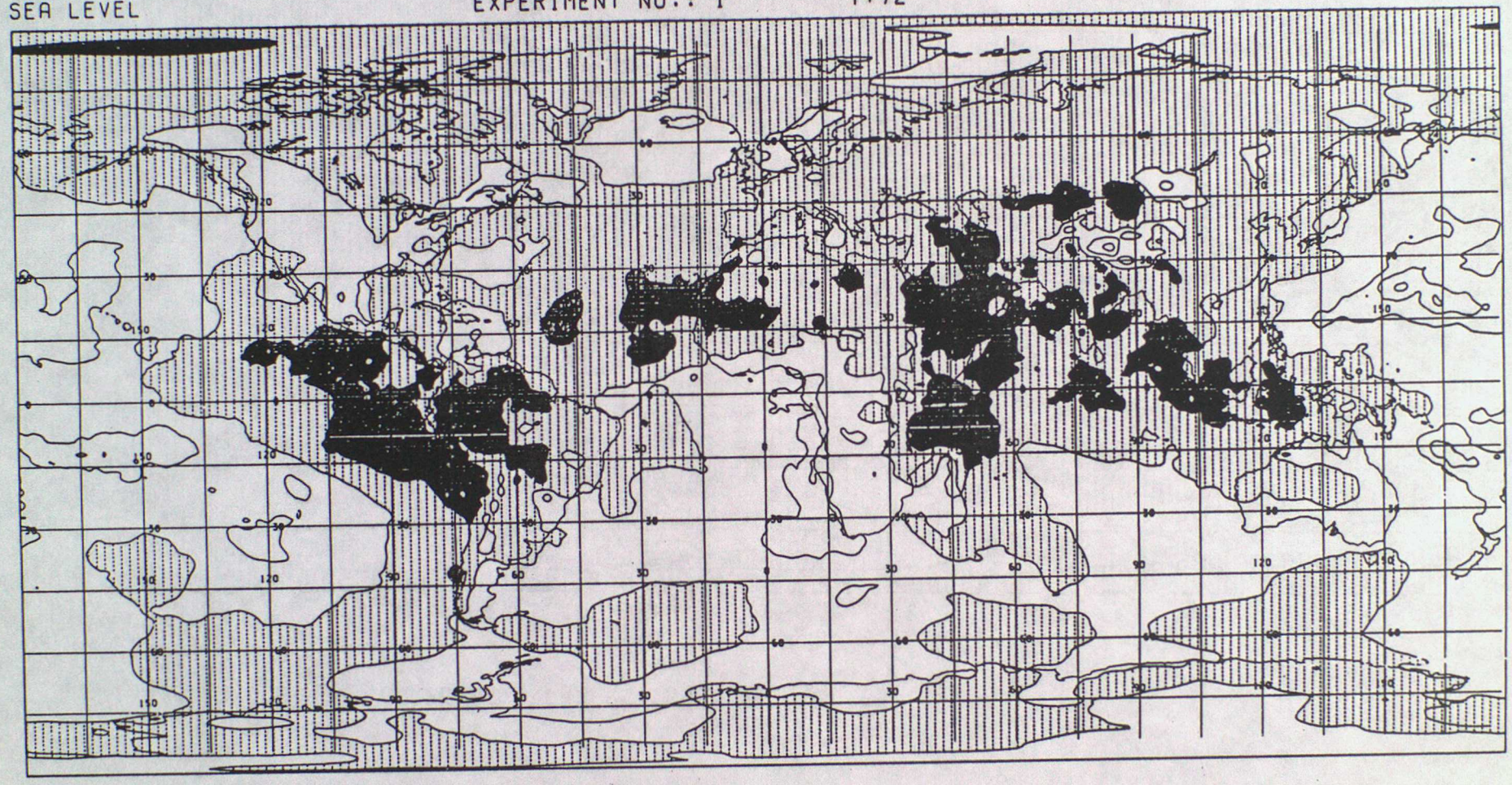


Figure 11a.

T-STATISTIC RANDOM PART OF 3DY F/C ERRORS T850. 12Z. MAY88
CONTOUR INTERVAL = 5.0
VALID AT 12Z ON 1/5/1988
LEVEL: 850 MB EXPERIMENT NO.: 1 T+72

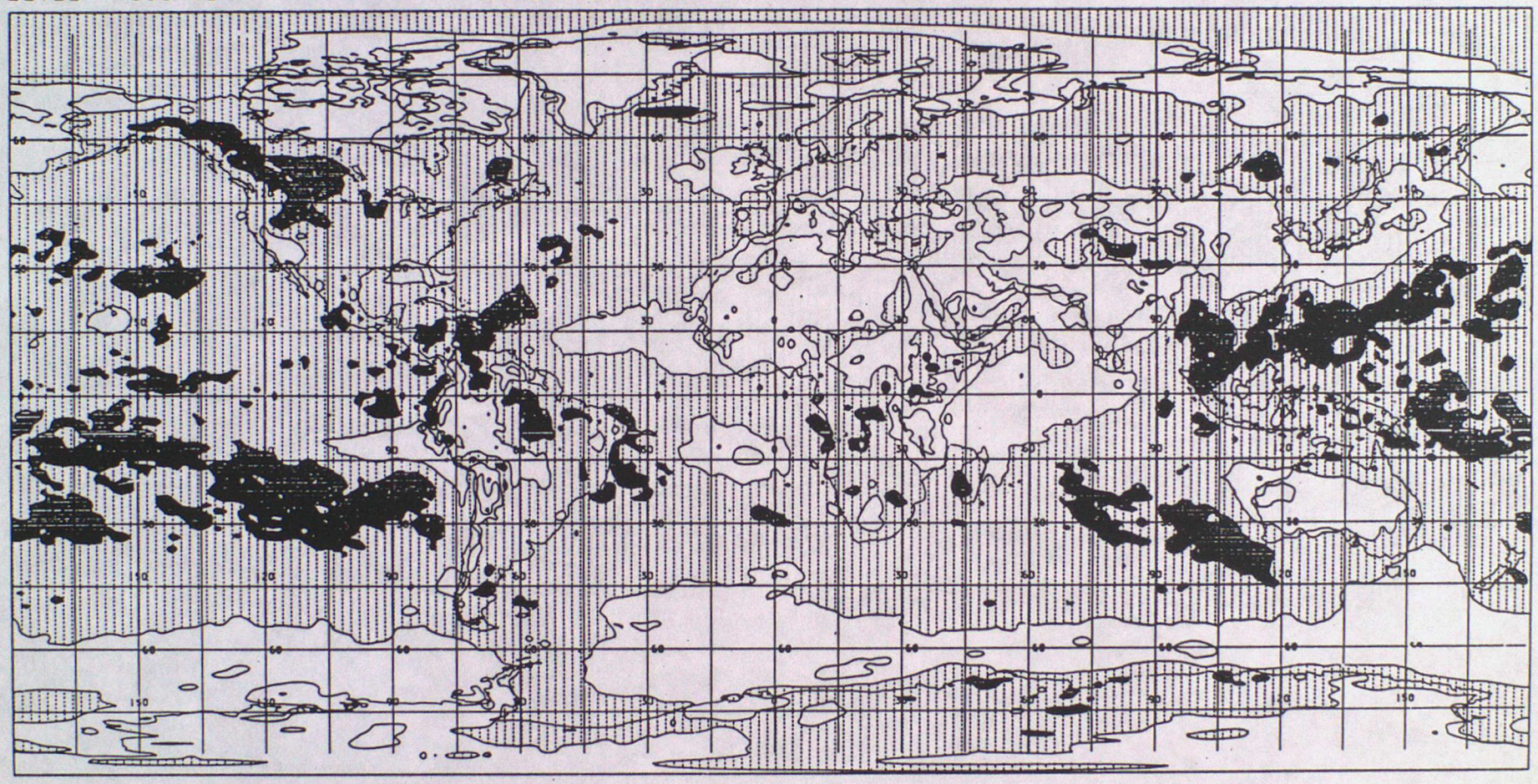


Figure 11b

T-STATISTIC RANDOM PART OF 3DY F/C ERRORS T700. 12Z. MAY88
CONTOUR INTERVAL = 5.0
VALID AT 12Z ON 1/5/1988
LEVEL: 700 MB EXPERIMENT NO.: 1 T+72

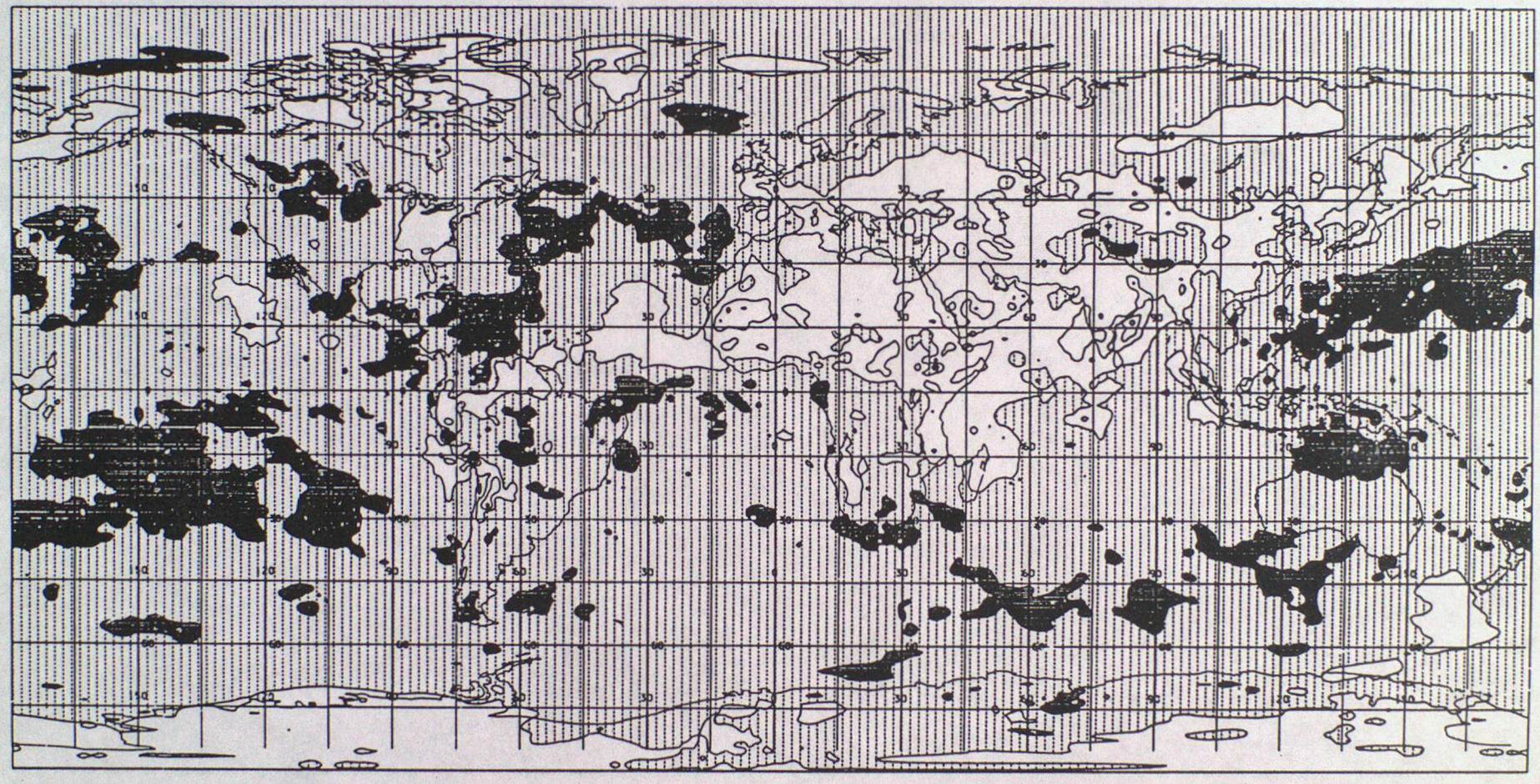


Figure 11c

Recent Met O 11 Technical Notes (New Series)

27.	The New Meteorological Office Data Assimilation Scheme	A.C. Lorenc R.S. Bell B. Macpherson April 1989
28.	Model error structure and estimated analysis accuracy with a network of wind profilers	N.B. Ingleby R.A. Bromley April 1989
29.	Examples of hybrid vertical co-ordinate systems for the unified forecast/climate model.	R. Swinbank July 1989
30.	The Meteorological Office Experimental Mesoscale Numerical Weather Prediction System:July 1989	B.W. Golding July 1989
31.	Tests of the Heun Advection Scheme.	B.L. Marshall July 1989
32.	Statistical Assessment of the Monthly Mean Error in the 3-day Forecast of the Global 500mb Height Field	R.A. Bromley R.A. Downton October 1990
33.	Spatial diagnostics of operational assimiliations using the observation processing database.	P. Jemmer August 1989
34.	Enhancements to the Mesoscale Model and their impact on forecasts.	S.P. Ballard September 1989
35.	Preliminary stratospheric analysis experiments with the Analysis Correction scheme.	R. Swinbank October 1989
36.	A diagnostic study of the impact of Seasat scatterometer winds on numerical weather prediction.	N.B. Ingleby R.A. Bromley October 1989
37.	Extension of the Bayesian Ship Quality Control Scheme to all Surface Data, and a trial of the Quality Control of Land Synops.	C.A. Parrett O.M. Hammon November 1989
38.	The impact of satellite sounding data in the fine-mesh model.	R.S. Bell O. M. Hammon December 1989
39.	NWP systems of the future at the UK Meteorological Office, and the likely impact of windprofiler observations.	A.C. Lorenc November 1988
40.	The Impact of the Interactive Mesoscale Initialisation	B.J.Wright B.W.Golding November 1989
41.	A Trial of the Bayesian Upper Air Quality Control - New Year 1990.	N.B. Ingleby O.M. Hammon November 1990

42.	Dirac - Bergmann constraints and the semigeostrophic equations.	I.Roulstone March 1990
43.	Stratospheric Analysis Experiments with a 20-level sigma - coordinate model.	R. Swinbank May 1990
44.	Dynamic Initialisation by Repeated Insertion of Data.	B. Macpherson September 1990
45.	The Quasi-Equilibrium model version 1.	M.H. Mawson August 1990
46.	The impact of the North Atlantic Tempship observations on global model analyses and forecasts during a case of cyclogenesis.	R.J.Graham July 1990
47.	A Flexible and Efficient Method for Transforming Grid-point Data from One Grid to Another.	B.R.Barwell August 1990
48.	Vertical Interpolation of temperature observations and model data.	R Swinbank and C.A.Wilson September 1990
49.	Interpreting results from the Global and Regional NWP versions of the Unified Model.	C.A. Wilson
50.	A comparison of helicopter reports of icing conditions with mesoscale model forecasts for winter 1989/90.	R.W. Lunnon August 1990.
51.	An assessment of the surface fluxes from the UKMO NWP models. Part I MOMENTUM.	J.O.S. Alves September 1990
52.	An assessment of the surface fluxes from the UKMO NWP models. Part II HEAT AND FRESH WATER.	J.O.S. Alves September 1990
53.	The distribution of Bathy and Tesac reports in the Atlantic during 1988.	E.J.Cane M.J.Bell September 1990
54.	The comparison of winds at two sites at Gatwick airport during early 1990.	A.D. Marklow R.W. Lunnon November 1990
55.	A positive definite advection scheme suitable for the Unified Model.	S.A. Leathwood M.J.P. Cullen November 1990
56.	A comparison of a Hemsby radiosonde ascent with a dedicated flight of the RAE Bedford BAC-111 on 31 st. October 1990.	R.W. Lunnon D.A. Lowe A.D. Marklow January 1991
57.	Transformation of the mesoscale model equations to a generalised vertical coordinate system.	J.A.James January 1991
58.	Trials of increased resolution in space and direction for the wave model.	M.W. Holt March 1991