

European Centre
for Medium Range
Weather Forecasts

Analysis Error Calculations for the FGGE

Internal Report 11
Research Dept.

July 77

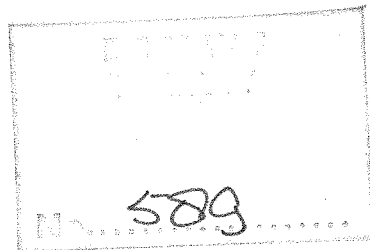
Centre Européen pour les Prévisions Météorologiques
à Moyen Terme

Europäisches Zentrum Für Mittelfristige Wettervorhersagen

Analysis Error Calculations
for the FGGE

By

G. Larsen, C. Little, A. Lorenc, I. Rutherford
EUROPEAN CENTRE FOR MEDIUM RANGE WEATHER FORECASTS, BRACKNELL



Internal Report 11
Research Department

June 1977

NOTE: This paper has not been published
and should be regarded as an
Internal Report from ECMWF Research
Department.
Permission to quote from it should
be obtained from the Deputy Director,
Head of Research at ECMWF.



A B S T R A C T

Global calculations have been made of the statistical error of analyses, by the method of optimum interpolation, for the basic FGGE observing system and for two versions of the augmented system for the special observing periods. The contributions of the two space-based systems, namely temperature soundings from radiance measurements and wind estimates from cloud motions, were also determined. Other calculations were made to determine the sensitivity of these results to the assumed prediction error of the data assimilation model, to the error characteristics of the satellite temperature soundings, and to the data selection algorithm in the analysis program. The results are presented in the form of zonal averages.

1. Introduction

Experiments have been made to determine the theoretical statistical error of analysis (interpolation) for various combinations of observing systems such as those proposed for the First GARP Global Experiment (FGGE). These experiments have simulated the operation of a global data assimilation scheme in which a global prediction model is updated by means of multi-variate optimum interpolation in three dimensions. The analysis error estimates are a natural by-product of the statistical calculations of the analysis scheme. They are sensitive to various statistical input parameters (covariance models) as well as to the data distribution and hence some experiments were also carried out to test this former sensitivity.

The analysis experiments were a simulation in the sense that no data assimilation cycle was actually run nor were any analyses actually made, but rather only the error distributions were calculated for a single "typical" update cycle using just the positions and error characteristics of the data.

The analysis program used was a preliminary version of the one being developed for eventual use at ECMWF for both operations and for FGGE. Since it will be more thoroughly described in a forthcoming Technical Report, only a brief description will be given in Section 2. Section 3 will include a description of the prediction error covariance model. Section 4 will describe the derivation of the data distributions and list the error characteristics assumed for the observations. Section 5 will describe the experiments and discuss the results.

2. The Analysis Program

The analysis method is a fairly straightforward extension to three dimensions of multi-variate optimum interpolation of observed minus forecast differences. Using A to represent any scalar variable and superscripts i , p and o to denote interpolated, predicted and observed values, the basic analysis equation can be written as

$$A_k^i = A_k^p + \sum_{i=1}^N C_{ki} (A_i^o - A_i^p) \quad (1)$$

where subscript k denotes the point of analysis and subscripts $i = 1, N$ denote points of observation. A can be any of the variables Z (geopotential height), ΔZ (geopotential thickness between two specified isobaric levels) or u or v (eastward and northward components of the wind). The points of observation can be anywhere in three dimensional space with respect to the point of analysis which also has arbitrary three-dimensional coordinates. Defining the "error" of A as $a = A - A^t$, where t denotes the "true" value, (1) becomes

$$a_k^i = a_k^p + \sum_{i=1}^N C_{ki} (a_i^o - a_i^p) \quad (2)$$

The mean square error of the interpolated value is then

$$\begin{aligned} \langle a_k^i{}^2 \rangle = & \langle a_k^p{}^2 \rangle + 2 \sum_{i=1}^N C_{ki} \{ \langle a_i^o a_k^p \rangle - \langle a_k^p a_k^p \rangle \} \\ & + \sum_{i=1}^N \sum_{j=1}^N C_{ki} C_{kj} \{ \langle a_i^p a_j^p \rangle - \langle a_i^o a_j^p \rangle - \langle a_i^p a_j^o \rangle \\ & + \langle a_i^o a_j^o \rangle \} \end{aligned} \quad (3)$$

Minimising this with respect to variations of C_{ki} leads to the following set of linear equations for the C_{ki} :

$$\begin{aligned} \sum_{i=1}^N \{ \langle a_i^p a_j^p \rangle - \langle a_i^p a_j^o \rangle - \langle a_i^o a_j^p \rangle + \langle a_i^o a_j^o \rangle \} C_{ki} = & \langle a_j^p a_k^p \rangle \\ - \langle a_j^o a_k^p \rangle ; & k = 1, N. \end{aligned} \quad (4)$$

The minimised mean square error is then

$$\langle a_k^i{}^2 \rangle_{\min} = \langle a_k^p{}^2 \rangle - \sum_{i=1}^N C_{ki} \{ \langle a_i^p a_k^p \rangle - \langle a_i^o a_k^p \rangle \}. \quad (5)$$

The interpolation errors or "analysis" errors discussed in this paper were calculated from expression (5) using the interpolation weights C_{ki} obtained by solving (4). In actual fact, correlation forms of (4) and (5) were used but this has no bearing on the results.

The organisation of the data search and the solution of (4) will be discussed next. The modelling of some of the various covariance terms appearing in (4) and (5) will be discussed in Sections 3 and 4. In the present version the covariances between prediction and observation errors, $\langle a_i^o a_j^p \rangle$ etc., were assumed to be zero, although they could be included if known.

2.1 Organisation of the calculations

Clearly, equation (4) cannot be solved for all the observations available globally, nor is this necessary as long as the covariances appearing in it are small for widely separated locations. Since we will use modelled covariances with this property (an approximation) we can restrict the solution to a region with a radius of at most a few thousand kilometres and often less, depending upon the data density. The limiting factor is the space available to store the left-hand side matrix in (4).

We adopt the principle that data be selected up to that limit, covering a roughly cylindrical influence volume with a radius as large as possible. We then invert the left-hand side matrix, which is independent of the analysis points k , and use it to calculate the weights and errors for as many analysis points as possible near the centre of the influence volume, i.e. for some "analysis volume". The number and placement of the grid-points within the analysis volume is quite arbitrary and independent of the analysis volume definition. In other words, the data-collection and weight-calculation algorithms are quite independent of the analysis grid.

A second principle adopted is that as much as possible of the information common to each analysis volume should be saved in passing to the next volume. In particular, the data and the correlations are carried over.

2.2 Collection of observations

In order to facilitate the collection from mass storage of data for the influence volumes, we first sort the observations into boxes of roughly equal volume and bounded by lines of latitude and longitude. Currently the boxes are 6° of latitude in the north-south direction and approximately the same east-west. These dimensions are easily varied. The sorting proceeds by data type in some order of priority, radiosondes coming first and satellite data last. Boxes are filled up to some limit (currently 10) of observations (an "observation" denoting all the data with the same horizontal coordinates). Surplus observations are placed in an overflow box (or boxes) whose index is then added to a list of neighbouring boxes carried with each box. Each box, which in addition to data and a neighbour list contains the parameters of the covariance model for that region, constitutes one record of a random access input file.

An influence volume consists of an appropriate collection of boxes, at least one box plus all its neighbours but more if necessary. The analysis volume is then the box at the centre. Provision exists for multi-box analysis volumes or for subdivision of a box into several analysis volumes, either horizontally or vertically.

A list of the boxes covering the region of analysis (which may be any subset of the box system) is pre-specified. Each box is then treated in turn, its data plus that of its neighbours and their neighbours, etc. being read in until a given volume of data is reached. The decision whether to form a multi-box analysis volume or to subdivide the central box into several analysis volumes depends upon the density and type of data collected.

The main benefit of three-dimensional as opposed to two-dimensional analysis lies in the treatment of single-level data, soundings with gaps and soundings which require "reference" information such as temperature soundings from satellites. On the other hand, for radio-sonde soundings, which provide complete mass and wind information at each level, there is no advantage in a three-dimensional treatment. In fact it can be shown that, with the

dimensionally separated covariance model used here, off-level data contribute almost no extra information. Moreover, each radiosonde sounding raises the order of the matrix in (4) by $3 \times$ (number of levels). With an upper limit for N of ~ 150 on the present computer, very few such soundings can be treated. Therefore, the present version of the program subdivides the central box into separate levels whenever any radiosonde soundings are present in the data collection. The data selection algorithm, to be described next, ensures that non-radiosonde data receive a three-dimensional treatment.

2.3 Selection of data

Data are selected from the available observations as collected above. In general, a compromise between several requirements has to be reached. There must be a sufficiently overlapping spatial coverage to ensure continuity between adjacent analysis volumes. There should be a reasonable balance between mass and wind data. Very dense data of a single type or for a single level, say, should not be allowed to exclude sparse data of other types. The experiments described in this report were run with a very crude preliminary version consisting of the following steps:

1. Scan the boxes in the order : Central box, neighbouring boxes, overflow central box, neighbours of neighbours.
2. Scan the observations in each box in the order: radiosondes, dropsondes, land surface, ship surface, aircraft, satellite temperature soundings, satellite cloud-track winds.
3. Scan all the levels within the analysis volume for each observation. If no data is found then scan up and down outside these limits until data is found.
4. Accept all data of each type ($V, Z, \Delta Z$) up to a limit of $N/3$ for each type separately or N for the total. In all the work reported here $N = 105$.

Clearly, when the number of available data is large the spatial distribution of selected data will be adversely affected by these arbitrary selection rules, particularly the arbitrary cut-offs mentioned in 4. Some experiments were performed to test the seriousness of this and they will be discussed in Section 5.5.

Future versions will incorporate a preliminary treatment of the data in order to:

1. Identify and separate-out data which need to be considered only locally (i.e. at points within the box).
2. Combine into "super-observations" very dense observation clusters of similar type.

Both of these steps will have the effect of reducing the number of data for a given spatial coverage and hence allow a reasonable selection and coverage to be obtained within the limits imposed. Furthermore a larger limit will be used for critical cases.

3. Prediction Error Covariance Model

Prediction error covariances were modelled as products of a horizontal and a vertical correlation function times a product of standard errors.

i.e.

$$\langle a_i, b_j \rangle = \langle a_i^2 \rangle^{\frac{1}{2}} \mu_h(a_i, b_j) \mu_v(a_i, b_j) \langle b_j^2 \rangle^{\frac{1}{2}}$$

where $a_i, b_j = u, v, Z, \psi$ or ΔZ .

3.1 Horizontal Correlations

It was assumed that :

$$\mu_h(Z_i, Z_j) = \mu_h(\Delta Z_i, \Delta Z_j) = \mu_h(\psi_i, \psi_j) = \exp(-br^2)$$

where r is the chord distance between location i and j , and b is an arbitrary constant (in these experiments $2 \times 10^{-12} \text{ m}^{-2}$).

$$\text{Also, } \mu_h(Z_i, \psi_j) = \mu_h(Z_i, Z_j) \mu_h(Z_j, \psi_j)$$

where $\mu_h(Z_j, \psi_j)$ is an arbitrarily chosen function of latitude, taken here to be

$$\mu_h(Z_j, \psi_j) = \begin{cases} -1 & \phi_j \leq -\phi_c \\ \sin \left(\frac{\pi \phi_j}{2\phi_c} \right) & -\phi_c < \phi_j < \phi_c \\ 1 & \phi_c \leq \phi_j \end{cases}$$

This implies that corrections to wind components and geopotential are geostrophically related poleward of ϕ_c , which was taken to be 20° in these experiments.

Then

$$\mu_h(\psi_i, u_j^{\text{transverse}}) = -\mu_h(u_i^{\text{transverse}}, \psi_j) = \sqrt{2b} \exp(-br^2)$$

$$\mu_h(\psi_i, u_j^{\text{longitudinal}}) = \mu_h(u_i^{\text{longitudinal}}, \psi_j) = 0$$

$$\mu_h(u_i^{\text{transverse}}, u_j^{\text{transverse}}) = (1 - 2br^2) \exp(-br^2)$$

$$\mu_h(u_i^{\text{longitudinal}}, u_j^{\text{longitudinal}}) = \exp(-br^2)$$

$$\mu_h(u_i^{\text{transverse}}, u_j^{\text{longitudinal}}) = (u_i^{\text{longitudinal}}, u_j^{\text{transverse}}) = 0$$

The implication of the above is that all horizontal correlations involving Z , ψ , $u^{\text{transverse}}$ and $u^{\text{longitudinal}}$ are homogeneous and isotropic, i.e. independent of horizontal location or of direction.

Correlations involving eastward and northward wind components (u_i, v_i) and (u_j, v_j) were then calculated from the above by a rotation of coordinates at i and/or j .

3.2 Vertical correlations

It was assumed that all vertical correlations involving ψ , u, v , or Z between two levels i and j are the same for those levels and that they are independent of horizontal position. They can then be tabulated as an $NL \times NL$ symmetric matrix where NL is the number of levels. They were derived from the Z covariance matrix presented by Rutherford (1976) and are given in Table 1.

Vertical covariances involving the thickness ΔZ , (defined as the thickness between level l and level $l - 1$, the level immediately below) were obtained by simply differencing the appropriate rows of the Z covariance matrix. These were then converted to correlation form.

3.3 Variances

Prediction error variances for each variable were modelled as functions of latitude and level only. Since the ratio of winter and summer values of the variance for a given level is about two or three (Rutherford, 1972), the northern hemisphere late-summer values of Hollett (1975) were multiplied by 3 in order to model winter values.

These values, different for each level, were then multiplied by a latitudinal profile function, the same for each level. This function was loosely based on the known latitudinal variation of the variances of the fields themselves (from Oort and Rasmusson, 1971). The profile was adjusted so that the resulting prediction errors were consistent with what is known for mid-latitudes but approached the variance of the fields themselves at the equator, the assumption being that the prediction model will likely be little better than climatology in the tropics.

The relations actually used were the following :

$$\begin{aligned} \langle Z^2 \rangle^{\frac{1}{2}} &= \sigma_Z Z(\phi), & \langle \Delta Z^2 \rangle^{\frac{1}{2}} &= \sigma_{\Delta Z} Z(\phi), \\ \langle \psi^2 \rangle^{\frac{1}{2}} &= 2\Omega g^{-1} \sigma_Z \Psi(\phi), & \langle u^2 \rangle^{\frac{1}{2}} &= \langle v^2 \rangle^{\frac{1}{2}} = (2b)^{\frac{1}{2}} \langle \psi^2 \rangle^{\frac{1}{2}} \end{aligned}$$

where Ω = earth's angular rotation rate,

$$\Psi(\psi) = (18 - 5 \cos(4\phi) + 5|\sin\phi|)/18,$$

$$Z(\psi) = \sin\phi \mu_h(Z, \psi)^{-1} \Psi(\phi).$$

Note that for

$$|\phi| \geq \phi_c, \langle Z^2 \rangle^{\frac{1}{2}} = fg^{-1} \langle \psi^2 \rangle^{\frac{1}{2}} \text{ where } f = 2 \Omega \sin \phi$$

while

$$\lim_{\phi \rightarrow 0} \frac{Z(\phi)}{\Psi(\phi)} = \frac{2\phi_c}{\pi}$$

Graphs of Z, Ψ and $\mu_h(Z\psi)$ are given in Fig. 1 while Fig. 2 shows the resulting profile for the 500 mb vector wind prediction error and also the corresponding profile for deviations from climatology (from Oort and Rasmusson, 1971).

4. Observations

In these experiments three different data distributions were used. The first of these, the Basic Observing System, contains all the components of the normal meteorological observing system as it will exist during FGGE, including temperature soundings from four polar orbiting satellites and cloud-track winds from five geostationary satellites.

The second includes everything in the Basic Observing System plus certain special observing systems which it was originally hoped would be operating during the two Special Observing Periods (SOP's) of the FGGE, as an ideal extension of the basic system. The third is similar to the above except that the Special Observing Systems are reduced to the level that it is expected will actually be available if current commitments are fulfilled.

A description of the various components and their approximate distributions and resolutions are given in Table 2.

Maps of the horizontal positions of the observations for each component are given in Figs. 3 to 15. These maps show only those observations which were actually chosen by the analysis program, not the full set actually generated. Because of the order in which data were chosen and the upper limit on the number that could be used, some of the distributions show gaps in areas where data of a higher priority are very dense. For example, over Europe, where the radiosonde density is very high, only a few surface observations were chosen, even though the number available was actually very large. Similarly over the North Atlantic and to a lesser extent over the Pacific the density of ships taking surface observations and of aircraft reports was so great that very few satellite soundings or satellite winds were actually chosen. Clearly this will have an effect on the calculated impact of satellite data, as we shall see later. On the other hand, in the tropics and in the southern hemisphere the distributions of the

last chosen types (satellite soundings and winds) were essentially as generated.

4.1 Generation of distributions

4.1.1 Sondes and land surface stations

Positions were assigned as listed in WMO Volume A. Data, either Z or V or both were assumed to be available as given in Volume A. In these experiments only those stations reporting at OOGMT were used.

4.1.2 Ships :

Positions were generated at random within 5° squares according to reporting-frequency tables constructed for the distribution on 12 June 1964 (Hanzawa and Tournier, 1968). These tables give the number of reports available in 24 h whereas our experiments require the data available in a typical 6-hour period. Nevertheless they were used without correction. Since most ships report only during daylight hours (i.e. for 2 of the 4 daily synoptics) we have probably over-estimated the number of ships by a factor of two. On the other hand, during FGGE the number of ship reports may be considerably larger than on 12 June 1964.

4.1.3 Aircraft :

Positions were generated at random within 5° squares according to frequency tables constructed from frequencies for 15 days in March 1975 (McDonell, 1976) and 10 days in June 1975 (WMO Meeting on Automated Aircraft Reporting Systems, Geneva, 1975).

4.1.4 Satellite Temperature Soundings:

Four polar orbits of the NIMBUS-6 type were assumed. Each satellite produced 6 soundings perpendicular to its track every 400 km along the track. No positions over land were generated and gaps were left wherever the humidity was high over a minimum thickness of model atmosphere representing thick cloud. (For this purpose the humidity distribution on a particular day in a run of the U.K. Met. Office GCM was used).

4.1.5 Cloud-track winds :

Frequency tables for 5° squares were constructed assuming 5 geostationary satellites of the SMS - GOES type, as specified in WMO publication No. 411. Random positions within these squares were generated but only if model humidity was high, but not

saturated, representing non-overcast model cloud. The distribution of positions was very similar to that obtained in 10 months of actual operation by NESS (C.A. Spohn, personal communication).

4.1.6 S.O.P. Additions

Positions for the various components were generated at random within the geographical limits and at the densities given in Table 2. In the case of aircraft dropsondes, positions were generated along the tracks indicated in Figure 13 (GARP Special Report No. 22).

4.2 Observation Errors

In most cases random errors with the magnitudes given in Table 2 were assumed. For radiosondes, errors were assumed to be vertically correlated but not horizontally. Standard errors as a function of level and correlations between levels are given in Table 3. These are the values calculated by Hollett (1975). Satellite thicknesses were assumed to be correlated both horizontally ($\mu_h(Z, Z) = \exp(-16 \times 10^{-12} r^2)$) after Bergman and Bonner (1976) and vertically. Standard errors and vertical correlations are given in Table 4. The former were calculated hydrostatically from the layer mean temperature errors, also given in the table (extracted from the report of the 13th session of the WGNE, section 3, attachment 5). The correlations are simply crude guesses based on the shapes of the radiance weighting functions.

5. Experiments

The analysis error experiments which were made were of two different types; firstly, experiments to determine the effectiveness of various data distributions and secondly experiments to determine the sensitivity of the results to various assumptions about the errors of prediction and observation.

In all experiments analysis errors were calculated for four standard pressure levels 1000, 850, 500 and 300 mb. Several of the runs were global but many were restricted to the region south of 20° N (The data distributions are identical north of 20°N) and others were even further restricted to south of 20°S (when it was known that there would be little difference in the tropics and northern hemisphere). All analyses were made on a 5° latitude by 5° longitude grid (origin at 2.5° lat., 0° long). For all experiments full two-dimensional maps of the analysis error fields are available. However these are far too detailed and only a selection will be presented here. For most purposes the zonal averages of these are more meaningful and the results will be presented mainly in this form.

In all of the figures the curves labelled "No Observations" are in fact the assumed prediction error. They give the error that would exist if there were no data available for one update cycle (assuming that the prediction error curve is valid). The same prediction error profile has been used for both hemispheres, ignoring the dependence of prediction error on analysis error. The values used are thought to be appropriate for the northern hemisphere winter with good data coverage. Sensitivity of the results to the prediction error profile will be discussed in Section 5.2.

5.1 Analysis Errors for the three FGGE network configurations

Global analysis error fields for the three FGGE configurations of Section 4 were calculated and the zonal averages are displayed in Figures 16 to 23.

At 1000 mb (Fig. 16 and 20) the errors of both height and wind in the two hemispheres are similar equatorward of about 40° . Poleward of 40° the southern hemisphere errors are much larger than in the northern hemisphere, even with the SOP additions which are principally the drifting buoys. The analysis error for the basic system (without buoys) lies about half-way between the SOP curves and the curve for no observations. Southward of about 60°S there are no buoys and the three error curves converge by about 70°S . In the tropics all three configurations give similar height errors, presumably because the prediction error is as small as the errors of the observations. For winds, however, the prediction errors are substantial compared with observation errors and hence the analysis error curves show more separation. This effect increases with altitude.

At 850 mb (Figs. 17 and 21) the picture is broadly similar but the value of the SOP systems in the equatorial zone is more apparent. It appears that the ideal tropical wind observing system provides about double the error reduction given by the expected system.

At 500 mb (Figs. 18 and 22) the curves for the expected and ideal SOP systems are well separated between 70°S and 30°S . The ideal system has the lowest errors and this must be due to the constant level balloons (at 300 mb) in the latter.

At 300 mb (Figs. 19 and 23) the expected SOP system is only slightly better than the Basic system between 70°S and 30°S , again because of the lack of constant level balloons. This suggests that even though there are drifting buoys at the surface providing reference level information for the satellite soundings of temperature (thickness) the errors of the latter are so large that this reference information is ineffective, except in the lower troposphere. This is not to say that the satellite temperature soundings provide no information. The errors with the soundings are still considerably lower than without them (see Section 5.3). However, this error reduction is due simply to the correlations between thickness and geopotential corrections, which provide information on the geopotential even in the absence

of a reference geopotential.

It is interesting that the combination of satellite temperature soundings and winds from constant level balloons is just as effective for geopotential determination as it is for wind.

The value of the balloons is particularly apparent over the Antarctic region south of 60°S at 300 mb and also at 500 mb, whereas at lower levels all the error curves coincide. Again, this suggests that the reference information degrades rapidly away from the reference level which in this case is at 300 mb.

It will be noted that all of the analysis error curves show a peak at about 40°N and at both poles. The former is due to the lack of radiosonde data in the Pacific and Atlantic oceans. These gaps are only poorly filled by satellite soundings, which in our distributions are very few in number (partly because of cloudiness in those areas on the day for which our distributions were modelled, but mostly because of the data selection problems discussed in Sections 2.3 and 5.5). Also, our distributions of cloud-track winds cut off rather sharply at about 40°N and S.

The errors at the poles are also due to data gaps; however, their importance is exaggerated in these plots of the zonal average against a linear latitude scale. They could be filled by small amounts of additional data. In particular, our modelled distribution of satellite temperature soundings cuts off over the Antarctic continent (and all other land areas). It can be expected that soundings of the TIROS-N type will be available for Antarctica (and other land areas) during FGGE, so our results are presumably too pessimistic for the polar regions.

5.2 Sensitivity to prediction error

The experiments discussed above were repeated with a different assumed form for the latitudinal profile of the prediction error. Since the effect was the same for all data distributions and levels only the 500 mb curves for the basic observing system are given, in Figs. 24 and 25. The two curves labelled "no observations" are the prediction errors whereas the curves labelled "basic" and "smaller prediction errors" are the corresponding analysis errors. It can be seen that the effect of changing the prediction error is considerable, especially in the southern hemisphere. However, it is quite small at northern latitudes (north of about 50°) where the errors of the available observations (mostly radiosondes) are much smaller than either of the assumed prediction errors and their density is such that they receive much more weight than the prediction and hence determine the analysis by themselves.

In an actual analysis/prediction cycle the prediction errors must depend on the analysis errors which in turn will depend on the prediction errors. This is an obvious feedback effect which we have neglected. Our results show that the magnitude of the effect of prediction error on analysis error depends upon the data density and the relative size of the observation errors.

Globally averaged, a 29 % change in prediction error led to a 25 % change in analysis error. On the other hand, the effect of analysis errors on prediction errors depends upon many factors. With most models, the error with respect to the real atmosphere grows during a prediction at a rate much larger than the rate of growth of initial errors in a perfect model. This presumably means that the largest part of prediction error growth is due to model-related factors. Hence a doubling of analysis error does not necessarily lead to a doubling of subsequent prediction error. In an actual cycle some sort of equilibrium will be reached, which will depend mainly upon the density and reliability of the observations themselves. Because of the neglected feedback effect this dependence on the data is underestimated in our results quoted in Section 5.1.

5.3 The impact of satellite data

Graphs showing the effectiveness of the space-based observing systems: satellite temperature soundings (denoted SIRS) and cloud-track winds (SWND), both separately and together are given in Figs. 32 to 39. All of these graphs indicate that the impact of satellite data in the northern hemisphere is very small, whereas in the southern hemisphere it is substantial. In fact, at about 70°N our calculations indicate that removal of satellite soundings actually lowers the errors. This is a false result caused by the crudeness of the data selection technique and the limitation on matrix size which meant that with satellite soundings present some of the more significant radiosonde data in neighbouring boxes were left out. This problem is particularly serious when the grid points are near box boundaries, as they are at that latitude.

The failure to show any impact of satellite data in the northern hemisphere is mainly due to the fact that our program simply did not select them north of about 20°N, due to the presence of large numbers of other types of data that were selected first, as discussed in Section 4. Future versions will avoid this problem in the manner already discussed.

At 1000 mb the space-based systems make little contribution to the reduction of analysis error anywhere, although equatorward of about 40° the cloud-track winds (presumably those at 850 mb) do have some effect, particularly on the wind error.

At 850 mb the effect of cloud-track winds on the wind error is very significant while the temperature soundings provide very little error reduction. In the case of height error, neither satellite system is very effective but winds are more effective at low latitudes and temperature soundings at high latitudes.

At 500 mb and 300 mb cloud-track winds provide lower wind errors than temperature soundings, although between 30°S and 60°S the temperature soundings give considerable information. For height errors the temperature soundings are more significant than cloud-track winds, except equatorward of about 30° where

the reverse is true.

In general, the effect of cloud-track winds on height errors is greater than the effect of temperature soundings on wind errors, in spite of the fact that these winds are available (in this data set) at only two levels (850 mb and 300 mb). At low latitudes the temperature soundings provide almost no error reduction at all, simply because their errors are larger than the prediction error and indeed larger than the error of climatology.

As an example of the horizontal distribution of analysis error (normalized by prediction error) with and without satellite data (of both types) we present Figs. 34 and 35, which show these fields for the 500 mb height. The impact in the ocean areas of the southern hemisphere can clearly be seen, as well as the lack of any effect in the northern hemisphere except in a small part of the Pacific, viz. $40^{\circ}\text{N } 160^{\circ}\text{W}$ where a few soundings were in fact selected (see Fig.7).

5.4 Sensitivity to different assumptions about the -----errors of satellite temperature soundings-----

Several sensitivity experiments were made in order to test the dependence of the above results regarding the impact of satellite temperature soundings on the form of the observation errors. These runs were limited to the region south of about 20°S where that impact was largest. The cases treated were :

1. SMALLER SIRS ERR : The standard error was halved but the correlations were left unchanged.
2. BROADER SIRS COR : The horizontal correlation function was broadened by using $\exp(-2 \times 10^{-12} r^2)$ (instead of $\exp(-16 \times 10^{-12} r^2)$).
3. FLAT SIRS COR : The horizontal correlation was given the form $0.5 + 0.5 \exp(-16 \times 10^{-12} r^2)$.
4. NO SIRS VER COR : The vertical correlations were all set to zero.

The 500 mb curves for halved standard error are shown in Figs. 36 and 37. By and large they are about as far below the basic system error curves as the latter are below the curves for no satellite temperature sounding data at all. The effect of lower observation errors increases with height (not shown) just like the impact of the data themselves. The southern hemisphere errors are comparable to the northern hemisphere values equatorward of about 40° , with the halved SIRS error; whereas with the original values they were considerably larger poleward of 20° . However, poleward of 40° the southern hemisphere errors are still considerably larger than in the north.

Although they are not shown here, these calculations were done for the B and C distributions as well. The analysis errors for these were lowered by about the same amount as with the A (basic) distribution. However, at 500 mb and 300 mb the errors for

distribution C (expected) are a little closer to those for distribution B (ideal) than they were in the original calculations. This indicates that with smaller SIRS errors the reference level information at the surface is effective to greater heights and hence the impact of a second reference level is somewhat smaller. However the conclusion that poleward of 40° errors in the southern hemisphere are considerably larger than in the north still stands, even with the ideal system.

The analysis error curves for various forms of SIRS correlation are given in Figs. 38 and 39. In general, these show that the sensitivity to changes in the spatial correlation is quite small. A broader Gaussian form or the introduction of a flat component causes the height analysis error to increase slightly and the wind error to decrease slightly, as one would expect, but the changes computed are really quite insignificant. However, the assumption of zero vertical correlation has a larger effect, lowering both height and wind errors by nearly as much as when the standard error was halved. Clearly, it will be important to determine the correct form for the vertical correlation of SIRS errors. One might expect that these could even be negative between some layers, in which case the effective thickness error for the sum of several layers could be even less than for uncorrelated errors.

5.5 Sensitivity to data selection algorithm

The selection technique employed here only guaranteed that the first 10 observations in a data box were used for the analysis at grid points within that box. The use of data from neighbouring boxes was haphazard, and in some circumstances important nearby data were neglected. The standard global run with the "basic" network was repeated with a slightly revised selection algorithm. In this version, the choice of data was made in the following sequence :

1. All data in the central box (including the overflow box).
2. All radiosonde data in neighbouring boxes .
3. All remaining data in neighbouring boxes.

Otherwise, everything was as previously described. The results for this run, and also the original run, are presented in Figs. 40 and 41.

As far as geopotential height is concerned the two selection algorithms give the same result in the southern hemisphere and in the tropics. The revised algorithm yields lower analysis errors in the northern hemisphere. The effect is particularly noticeable at the higher levels and is negligible at 1000 mb. For the vector wind errors the picture is similar but the southern hemisphere and tropical differences are larger than is the case for height errors.

Clearly, data selection is an important factor where data is dense, and it will be necessary in future versions to refine the technique or to use larger data matrices. Since the sensitivity to selection seems to be smaller in the tropics and southern hemisphere the comparisons between networks in Sections 5.1 and 5.3 are probably still valid, at least as far as this factor is concerned.

6. Conclusions

Our results indicate that only north of about 45°N are there sufficient observations to determine the zonal mean analysis error independently of the assumed prediction error in an analysis / forecast cycle. This means that our values of the absolute error must be treated with caution. However, it also implies that the use of past information, in the form of a prediction from previous cycles, is very important for the construction of accurate analyses. Moreover, it suggests that the prediction model used should be as accurate and as physically realistic as possible.

In view of the many assumptions and simplifications which have to be made in a study of this kind it seems safest to draw only qualitative conclusions from the results. The following statements concerning the relative performance of different observing systems may be supported :

1. During FGGE, analysis errors in the southern hemisphere will be considerably larger, latitude for latitude, than in the northern hemisphere, except near the surface equatorward of 40°S .
2. In the tropics, the expected special observing period configuration, will provide only about half of the error reduction which would be provided by an ideal system of soundings at 500 km average spacing.
3. A system of constant-level balloons at 300 mb poleward of 30°S would provide a significant lowering of upper tropospheric analysis error.
4. The impact of satellite data in the southern hemisphere is substantial (technical difficulties prevent any meaningful impact estimate for the northern hemisphere). Two levels of cloud-track winds provide about as much information as temperature soundings with errors of 2 - 3°C . Halving this error doubles the value of the soundings. Changing the assumed vertical correlation has more effect than changing the horizontal correlation.

References:

- Bergman, K.H. and W.D. Bonner (1976) Analysis error as a function of observation density for satellite temperature soundings with spatially correlated errors. Mon.Wea.Rev., 104, 1308-1316.
- Hanzawa, M. and T.H. Tournier (1968) System for the collection of ships' weather reports. WWW planning report No.25, WMO, Geneva.
- Hollett, S.R. (1975) Three-dimensional spatial correlations of PE forecast errors. Unpublished M.Sc.Thesis, Dept. of Meteor. McGill University, Montreal.
- McDonnell, J.E. (1976) Preparation of the data base for use in an operational environment. Proceedings of the JOC study group conference on four-dimensional data assimilation, Paris, 17-21 November 1975. GARP Report No. 11, 26-36.
- Oort, A.H. and E.M. Rasmusson (1971) Atmospheric circulation statistics. NOAA professional paper 5, U.S. Dept. of Commerce.
- Rutherford, I.D. (1972) Data assimilation by statistical interpolation of forecast error fields. J.Atmos. Sci., 29, 809-815.
- Rutherford, I.D. (1976) An operational three-dimensional multivariate statistical objective analysis scheme. Proceedings of the JOC study group conference on four-dimensional data assimilation, Paris, 17-21 November 1975. GARP Report No. 11, 98-121.

References:

- WMO (1975a) Volume A Observing Stations, record layout explanatory notes and code tables. Publication No. 9 TP.4, WMO, Geneva.
- WMO (1975b) Information on meteorological satellite programs operated by members and organisations. WMO No. 411, WMO, Geneva.
- WMO (1975c) Report of the meeting on automated aircraft reporting systems, Geneva.
- WMO (1976) Report of the third session of the WMO executive committee, inter-governmental panel on the first GARP global experiment. GARP Special Report No. 22, WMO, Geneva.



TABLE 1
PREDICTION ERROR CORRELATIONS

mb	1000	850	700	500	400	300	250	200	150	100
1000	1.00	.75	.56	.40	.30	.18	.11	.04	.03	.07
850	.75	1.00	.92	.66	.51	.35	.25	.14	.08	.14
700	.56	.92	1.00	.82	.66	.50	.39	.24	.18	.23
500	.40	.66	.82	1.00	.93	.77	.62	.42	.32	.36
400	.30	.51	.66	.93	1.00	.90	.71	.45	.35	.44
300	.18	.35	.50	.77	.90	1.00	.89	.63	.51	.54
250	.11	.25	.39	.62	.71	.89	1.00	.88	.74	.56
200	.04	.14	.24	.42	.45	.63	.88	1.00	.94	.58
150	.03	.08	.18	.32	.35	.51	.74	.94	1.00	.72
100	.07	.14	.23	.36	.44	.54	.56	.58	.72	1.00

COMPONENT	COVERAGE	HORIZONTAL RESOLUTION	VERTICAL RESOLUTION	VARIABLES	ERROR CHARACTERISTICS
Sondes	WMO Vol.A.		10 levels 1000-100 mb	V Z	random vert.corrn. (Table 3)
Surface	land:WMO Vol.A. sea: WWW, No. 25		1000 mb	V	random 6ms ⁻¹
Aircraft	(frequency tables:see text)		3 levels 300, 250, 200 mb	Z V	random 12 m random 4ms ⁻¹
Satellite Temperature Soundings	4 polar orbiters	400 km (sea only)	9 layers 1000-100 mb	Z	horiz. corrn.exp(-16x10 ⁻¹² r ²) vert.corrn. (Table 4)
Satellite Cloud- Track Winds	40N-40S 60N-60S		850 mb 300 mb	V	random 3ms ⁻¹ random 6ms

TABLE 2a - Basic Observing System -

TABLE 2b - Ideal SOP Additions -

COMPONENT	COVERAGE	HORIZONTAL RESOLUTION	VERTICAL RESOLUTION	VARIABLES	ERROR CHARACTERISTICS
Extra Tropical Sondes	10N-10S	500 km	10 levels 1000-10 mb	V Z	random 3 ms ⁻¹ vert.corrn.(Table 3)
Constant Level Balloons	0-90S	700 km	300 mb	V	random 1.5ms ⁻¹
Drifting ¹ Buoys	20S-60S sea only	600 km	1000 mb	Z	random 12 m

TABLE 2c - Expected SOP Additions -

Tropical sondes from Ships	10N-10S sea only	700 km	10 levels 1000-100 mb	V Z	random 3ms ⁻¹ vert.corrn.(Table 3)
Aircraft Dropsondes	special tracks		5 levels 1000-400 mb	V Z	as above as above
Constant ¹ Level Balloons	30N-30S		150 mb	V	random 1.5ms ⁻¹
Drifting ¹ Buoys	20S-60S sea only	600 km	1000 mb	Z	random 12 m

¹ Strictly speaking, these systems will not be confined to the SOP's but the deployment plan is to maximize coverage for these periods.

TABLE 3

- Radiosonde height errors and vertical correlations -

	1000	850	700	500	400	300	250	200	150	100
σ_Z (m)	6.0	7.8	9.3	13.1	16.1	21.7	27.3	33.9	35.6	32.2
1000	1.00									
850	.70	1.00								
700	.50	.72	1.00							
500	.39	.56	.74	1.00						
400	.34	.48	.69	.91	1.00					
300	.32	.45	.63	.84	.92	1.00				
250	.29	.42	.60	.80	.85	.94	1.00			
200	.27	.38	.56	.76	.79	.88	.96	1.00		
150	.27	.38	.56	.76	.80	.81	.92	.96	1.00	
100	.27	.41	.56	.71	.76	.79	.75	.74	.83	1.00

TABLE 4

- Satellite thickness errors and vertical correlations -

	1000	850	700	500	400	300	250	200	150
$\sigma_{\Delta Z}$ (m)	13.3	14.2	21.2	13.4	18.1	12.2	15.9	21.9	33.2
σ_T ($^{\circ}$ C)	2.8	2.5	2.15	2.05	2.15	2.28	2.43	2.60	2.80
1000 - 850	1.00								
850 - 500	.50	1.00							
700 - 500	.35	.55	1.00						
500 - 400	.20	.35	.60	1.00					
400 - 300	.10	.20	.35	.60	1.00				
300 - 250	0	.10	.20	.30	.55	1.00			
250 - 200	0	0	.05	.05	.15	.52	1.00		
200 - 150	0	0	0	0	0	.05	.50	1.00	
150 - 100	0	0	0	0	0	0	.0	.50	1.00

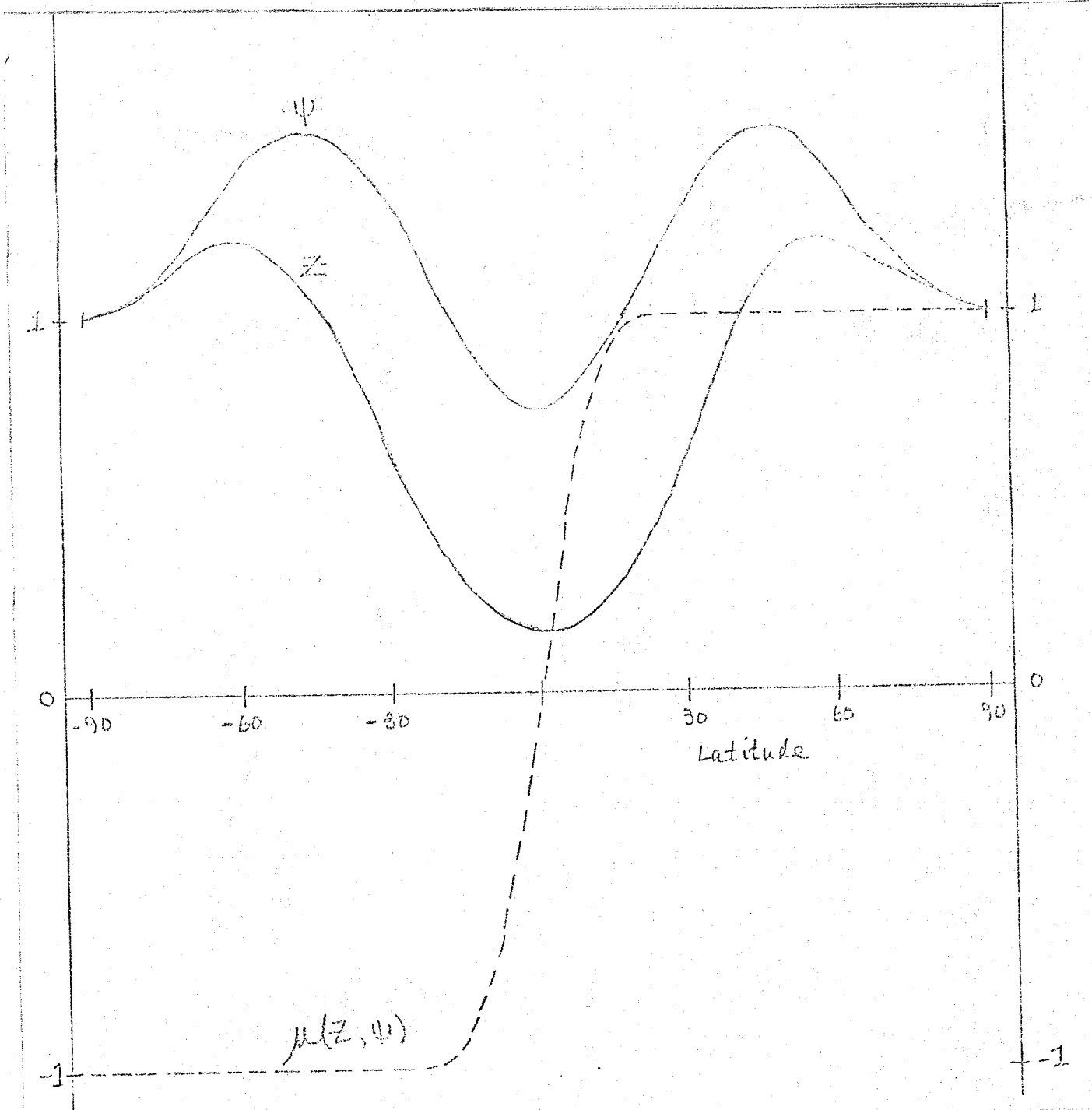


Figure 1 : Latitudinal variation of the prediction error for streamfunction (Ψ) and geopotential (Z) and their correlation (μ).

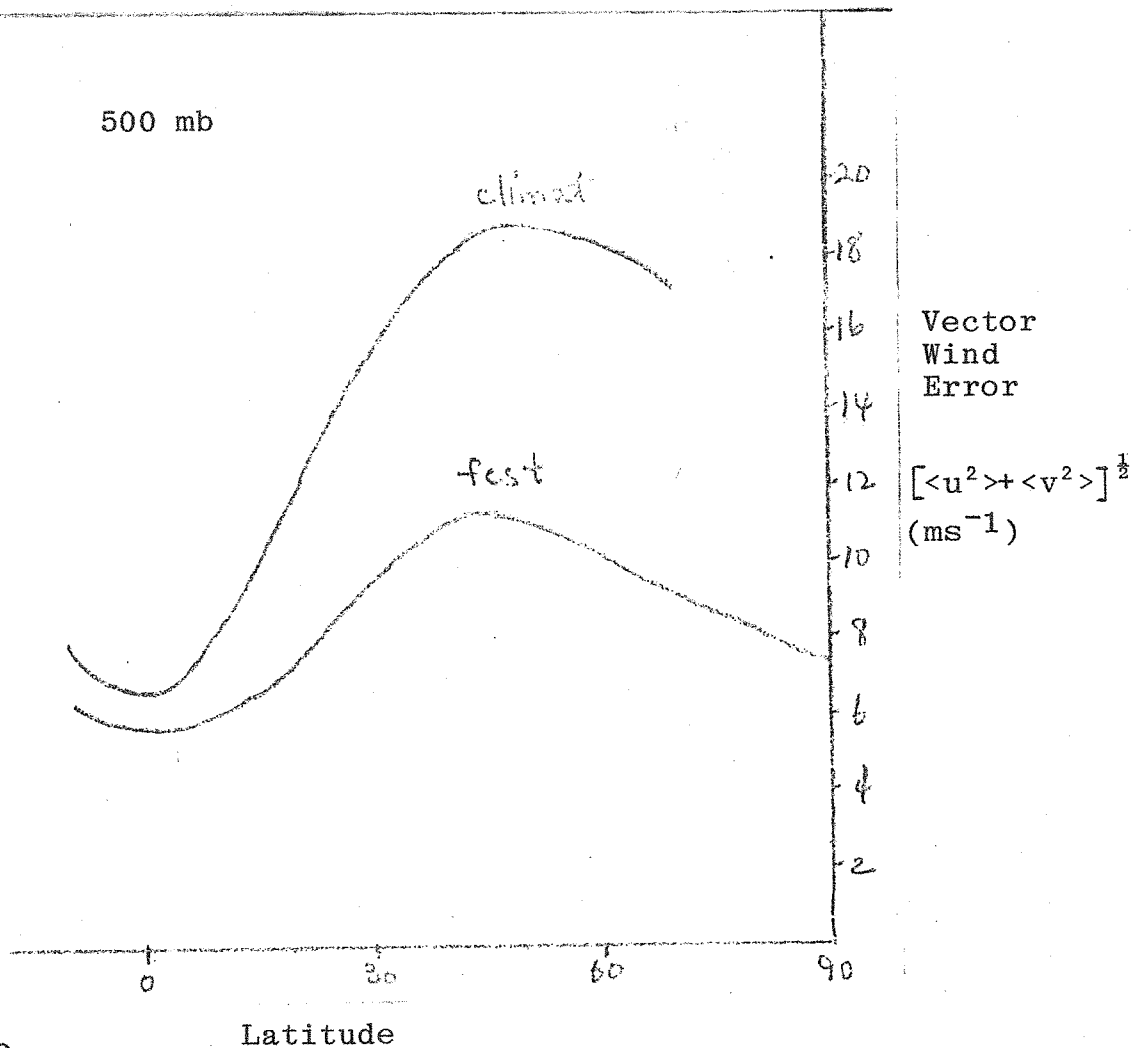
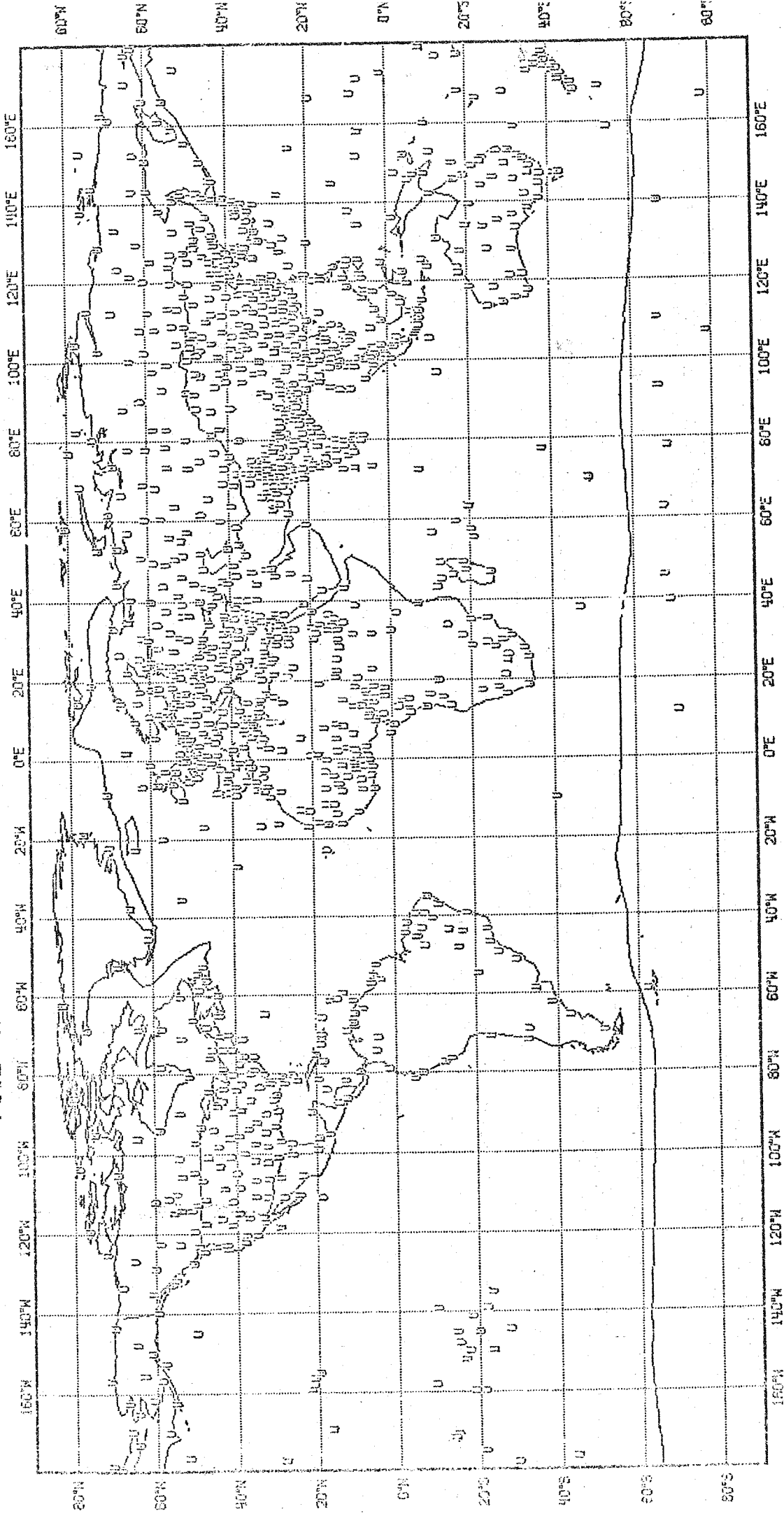


Figure 2 :

Latitudinal variation of the vector wind error at 500 mb.

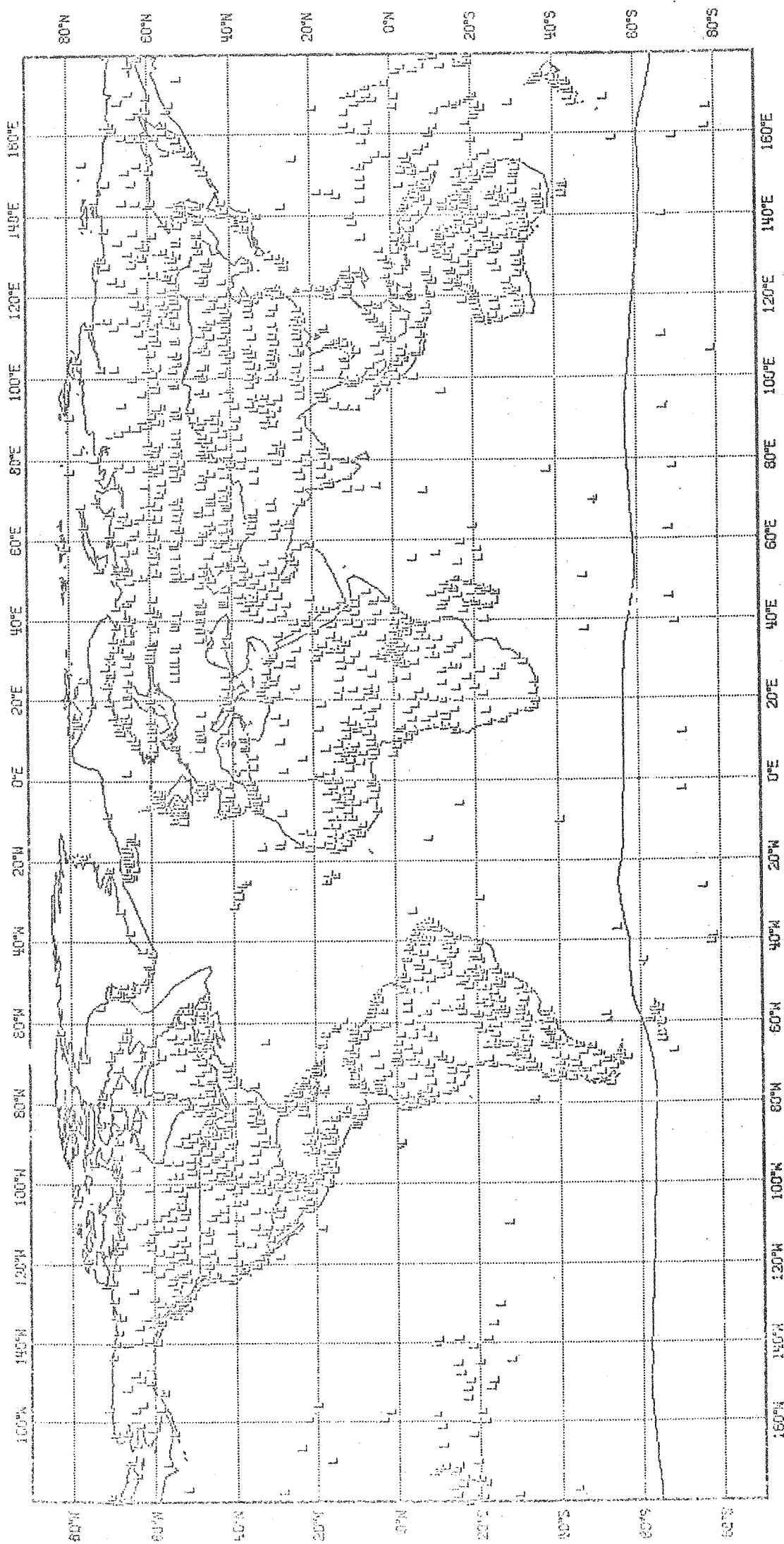
FIGURE 3: DISTRIBUTION OF CHOSEN OBSERVATIONS.



UPSONDES (U)

Figure 3

FGCE #Aa: DISTRIBUTION OF CHOSEN OBSERVATIONS.



LAND SURFACE (L)

Figure 4

FCSE #A# DISTRIBUTION OF CHOSEN OBSERVATIONS.

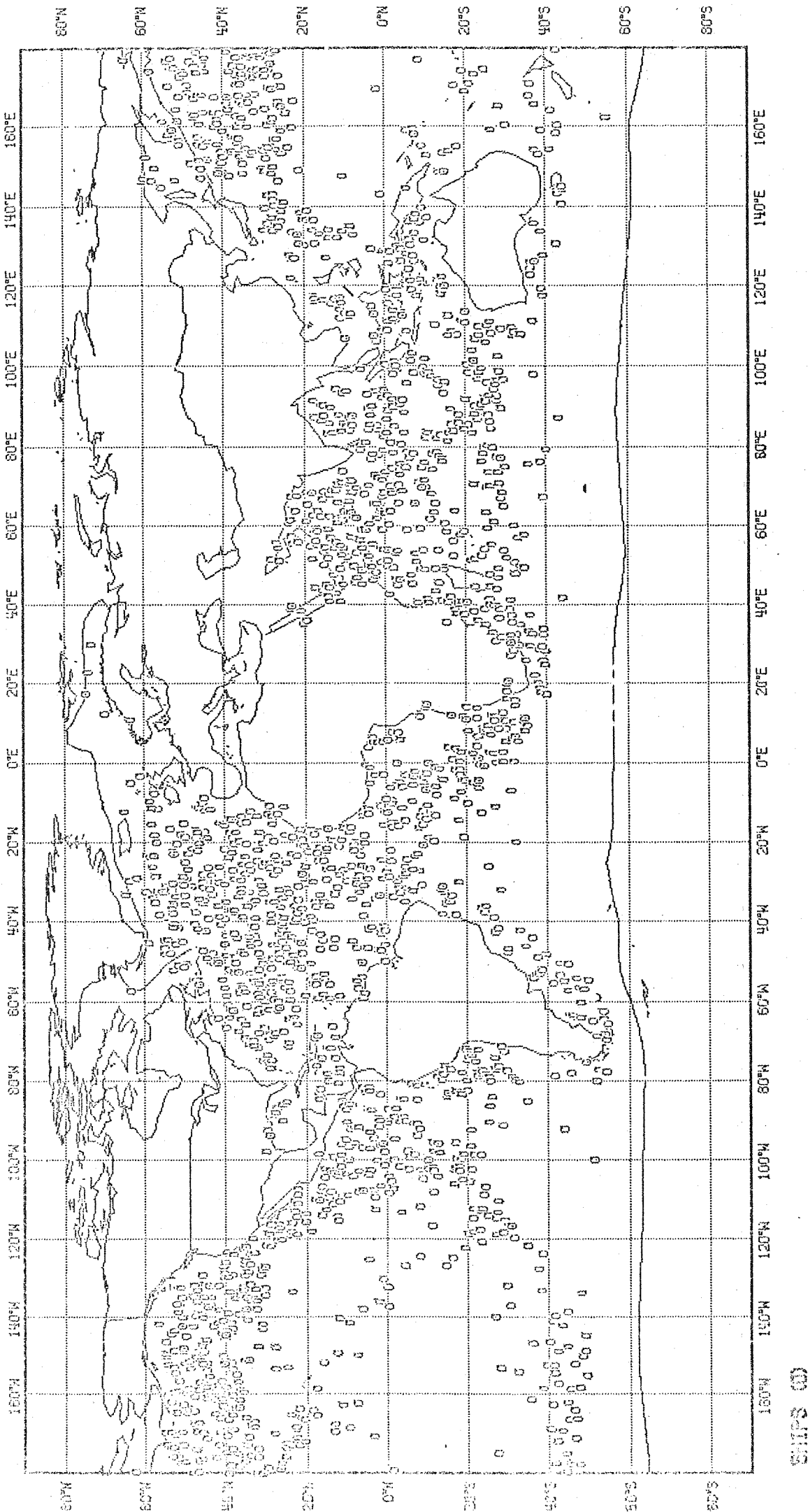
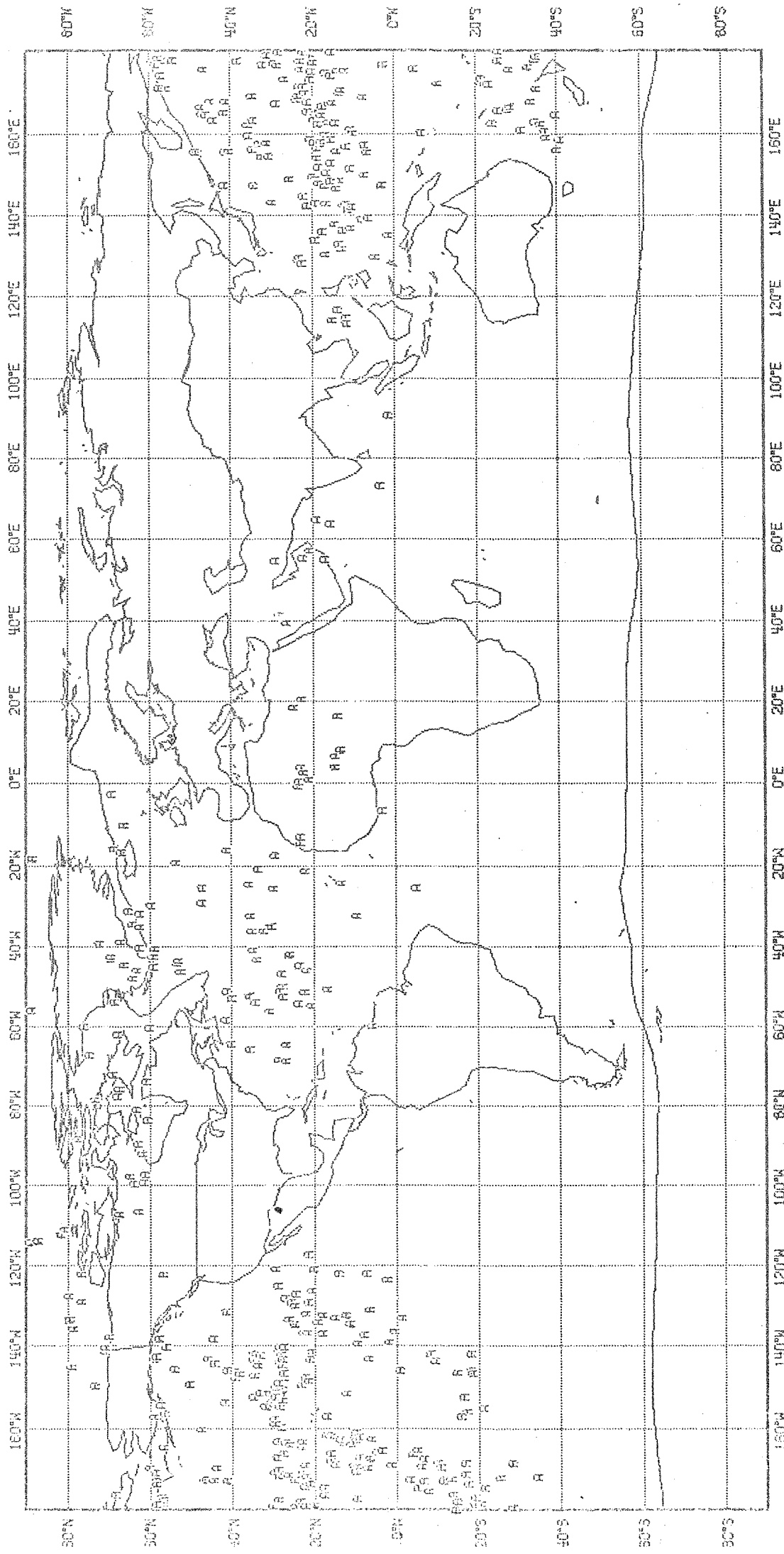


Figure 5

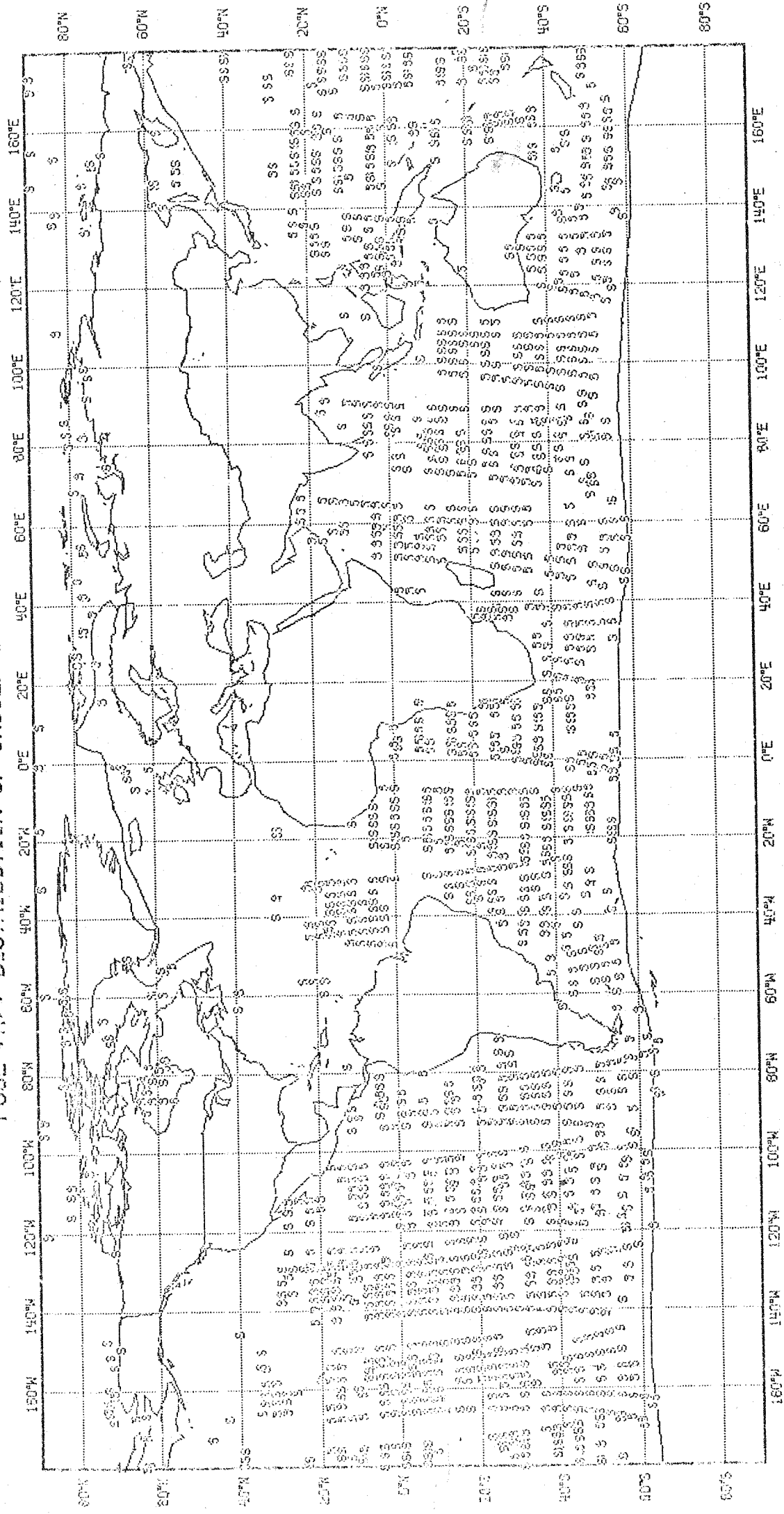
FOCE #R#s DISTRIBUTION OF CHOSEN OBSERVATIONS



ALA REPORTS (A)

Figure 6

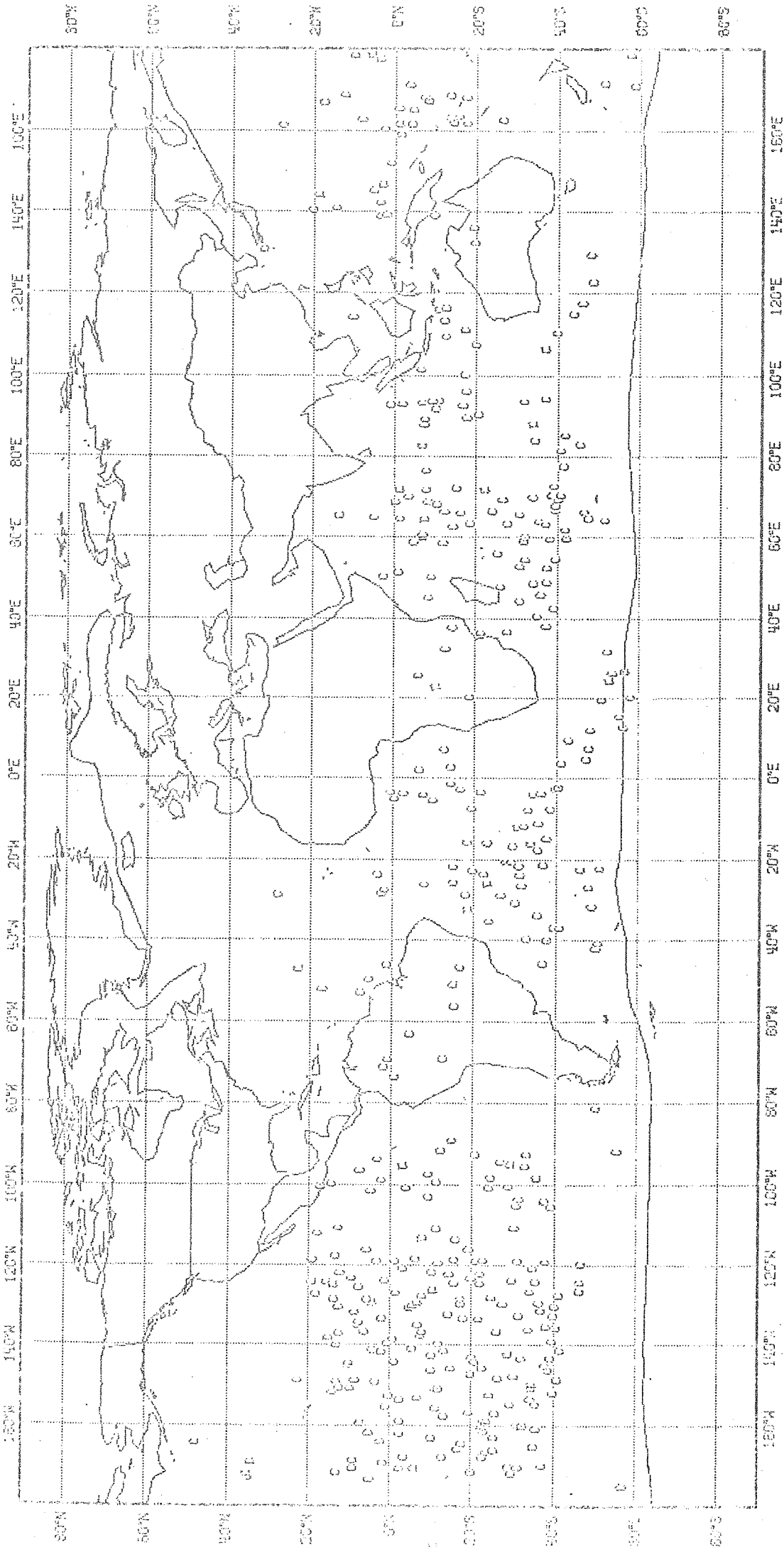
FIGURE 8: DISTRIBUTION OF CHOSEN OBSERVATIONS



SATELLITE SOUNDINGS (S)

Figure 7

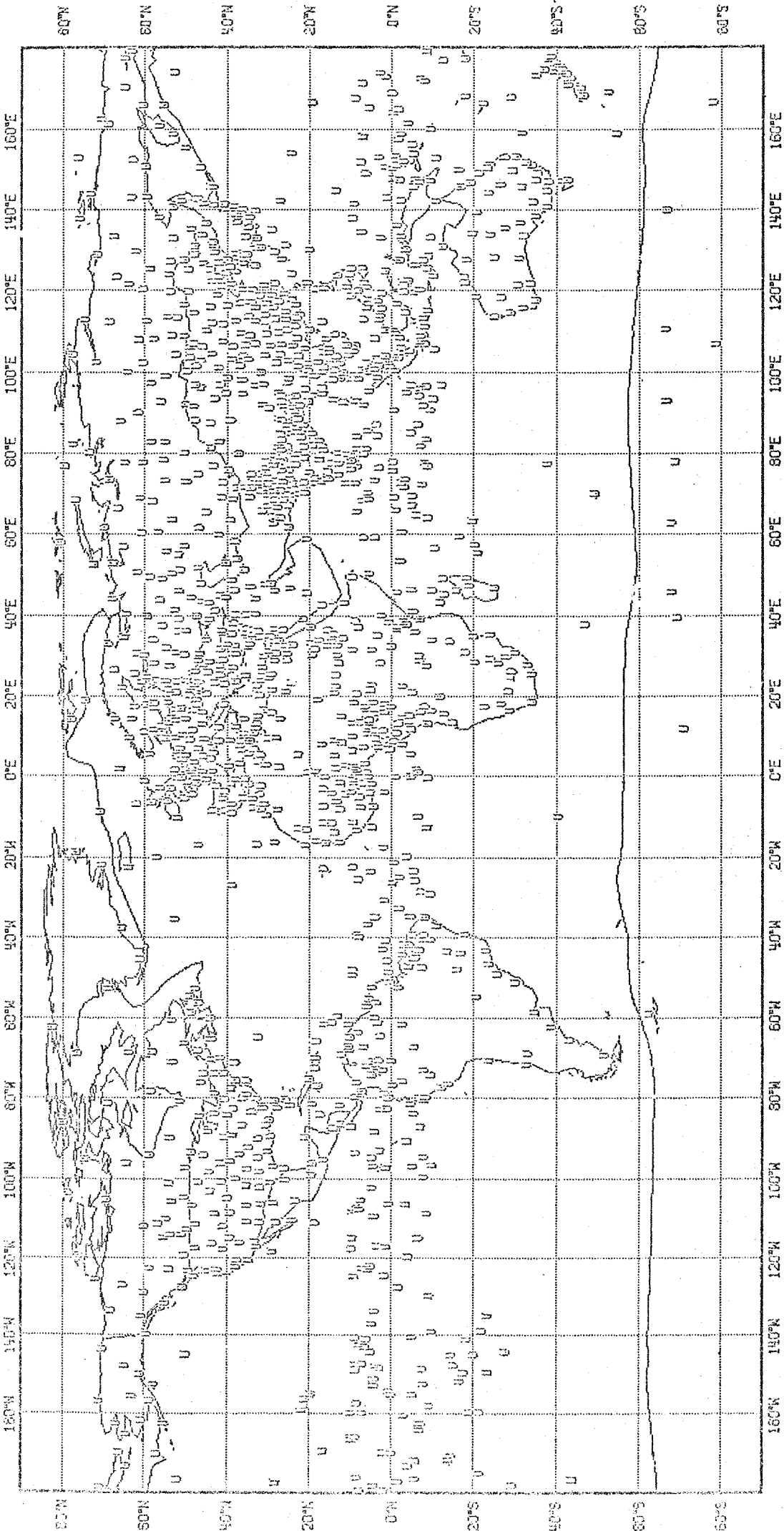
FOUR AREA: DISTRIBUTION OF CHOSEN OBSERVATIONS



CLOUD WINDS (C)

Figure 8

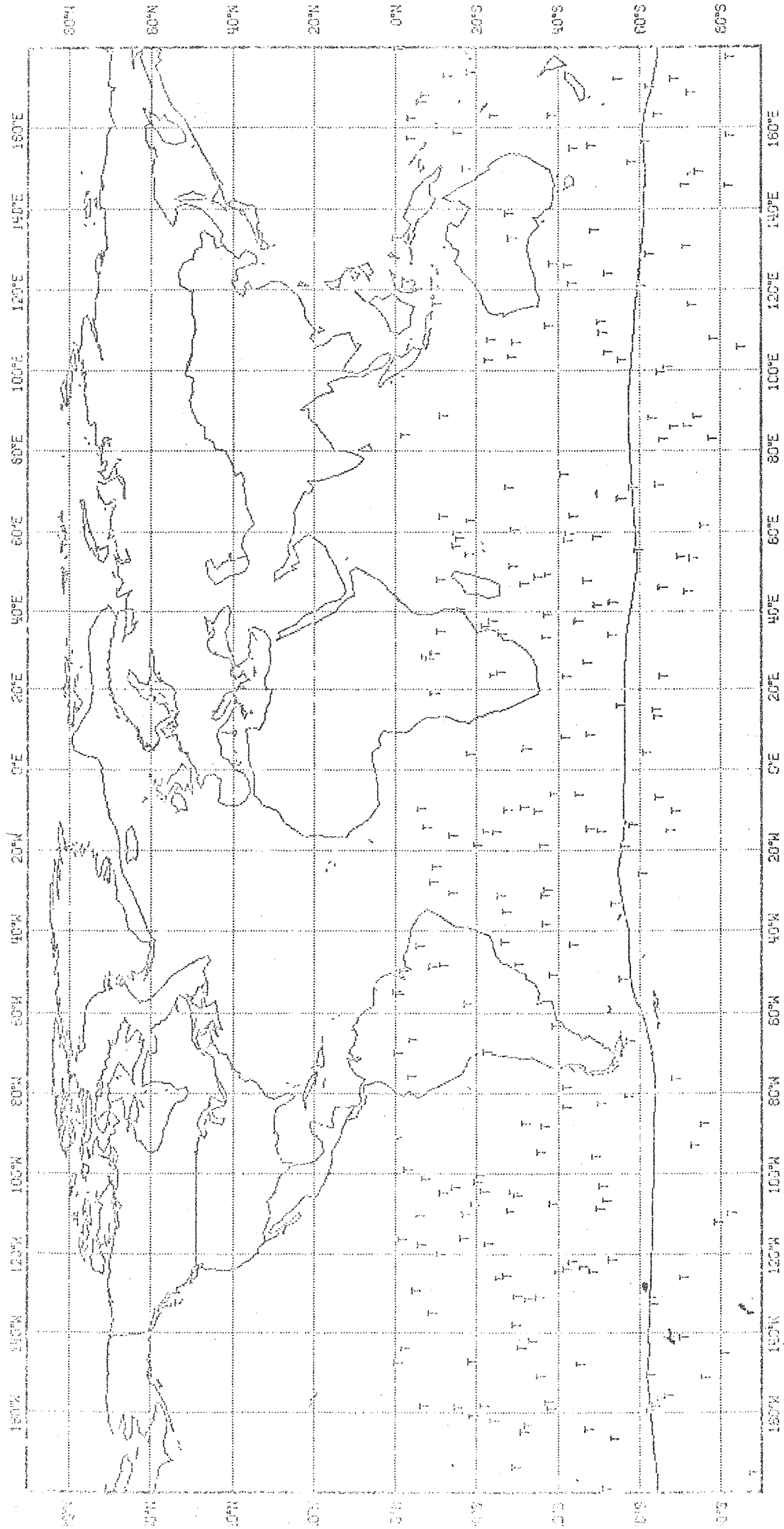
FOGE #B## DISTRIBUTION OF CHOSEN OBSERVATIONS.



UPSONDES (U)

Figure 9

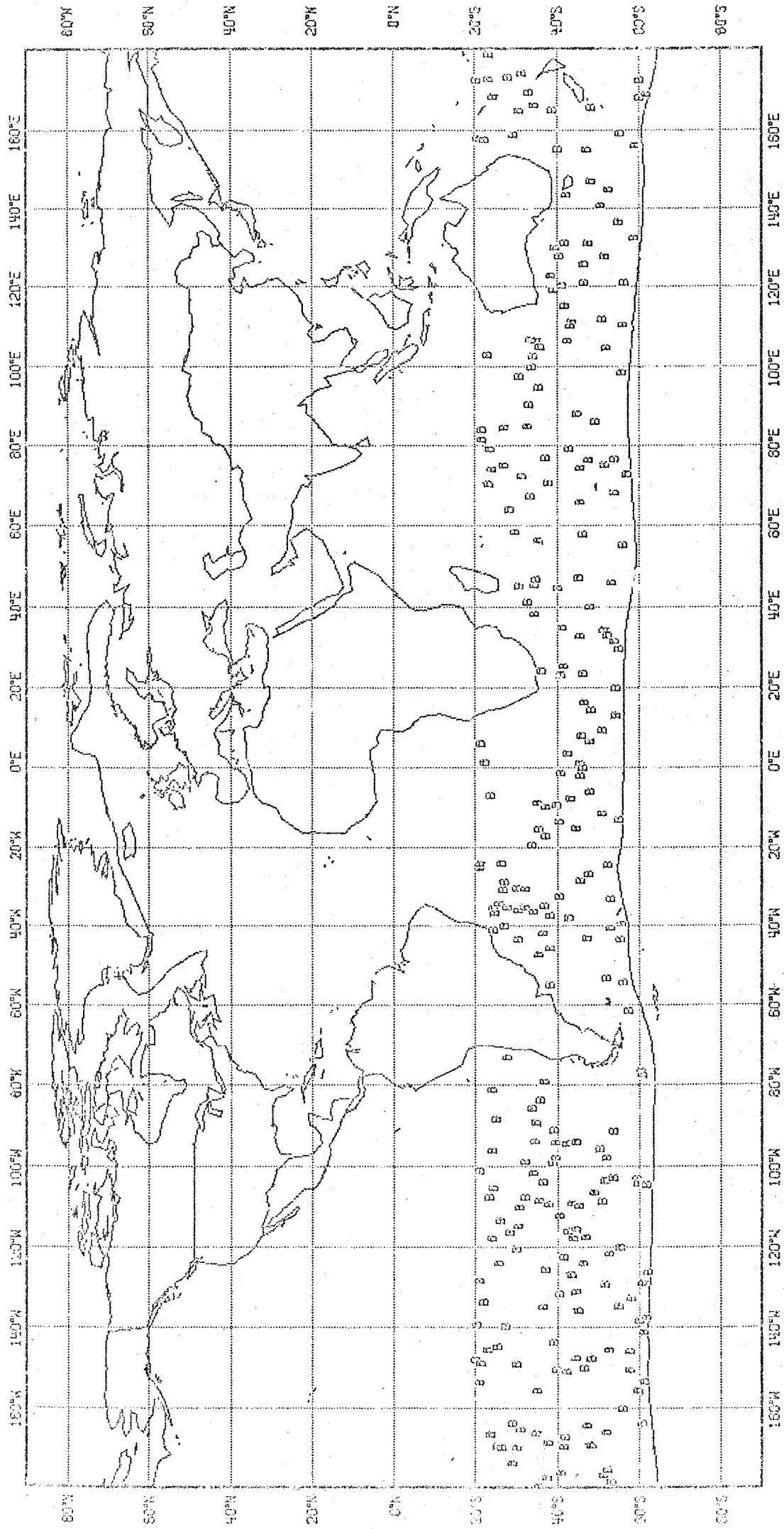
FOSE 40% DISTRIBUTION OF CHOSEN OBSERVATIONS



CONSTANT LEVEL BALLEONS (T)

Figure 10

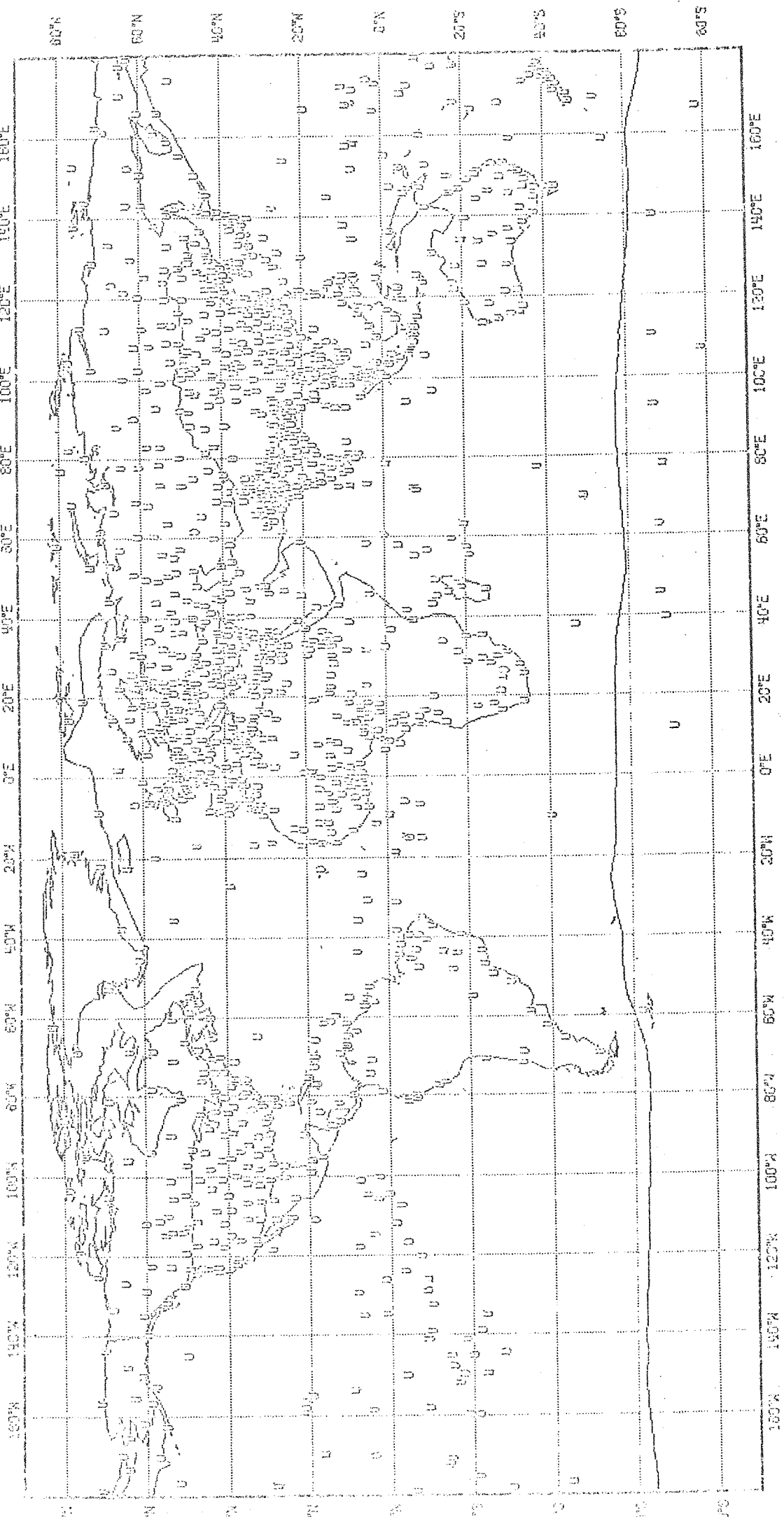
FGOE #B# DISTRIBUTION OF CHOSEN OBSERVATIONS



8005 (B)

Figure 11

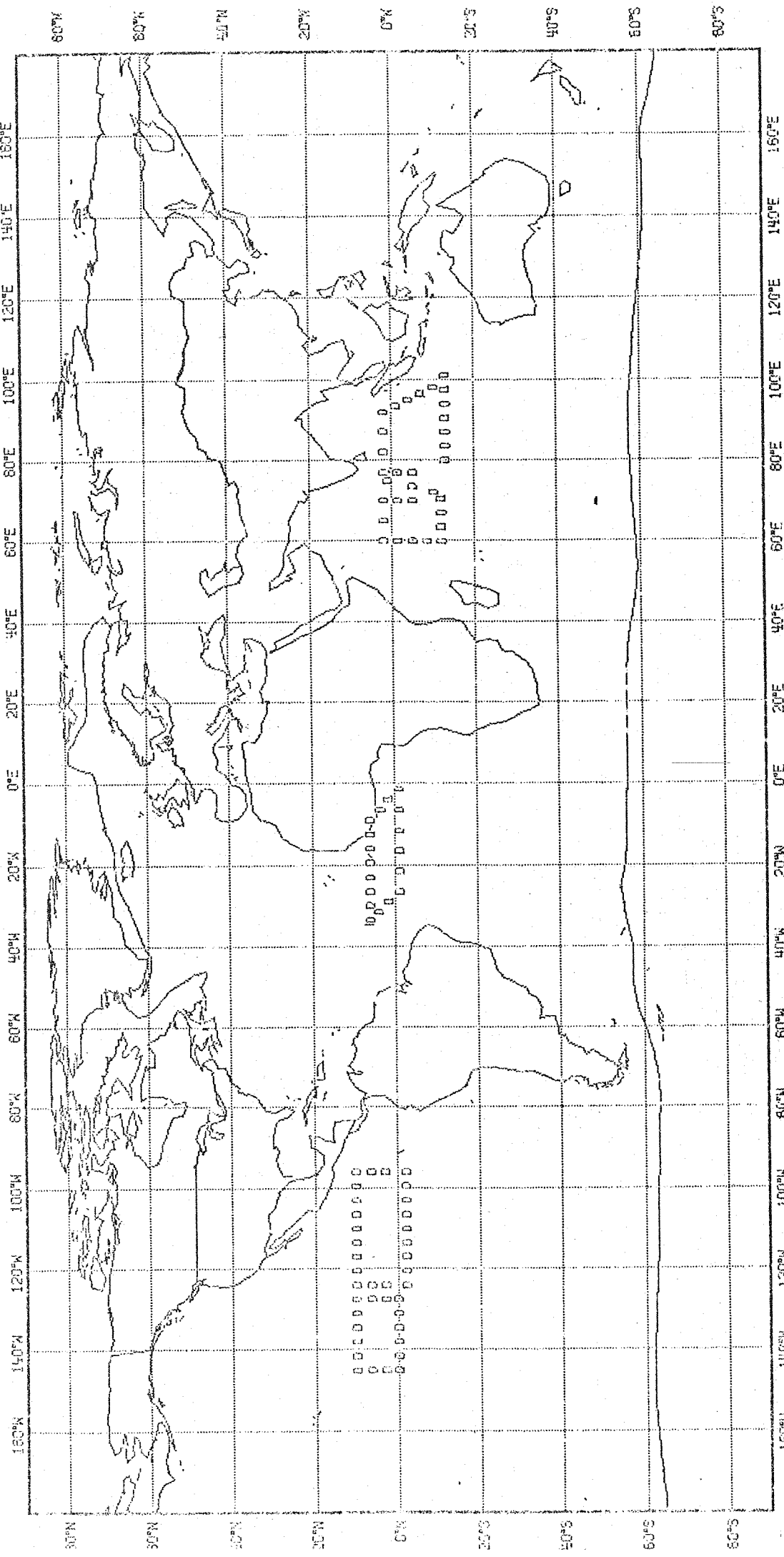
FIGURE 12. DISTRIBUTION OF CHISEN OBSERVATIONS.



UNITED STATES (U)

Figure 12

FGCE #02: DISTRIBUTION OF CHOSEN OBSERVATIONS

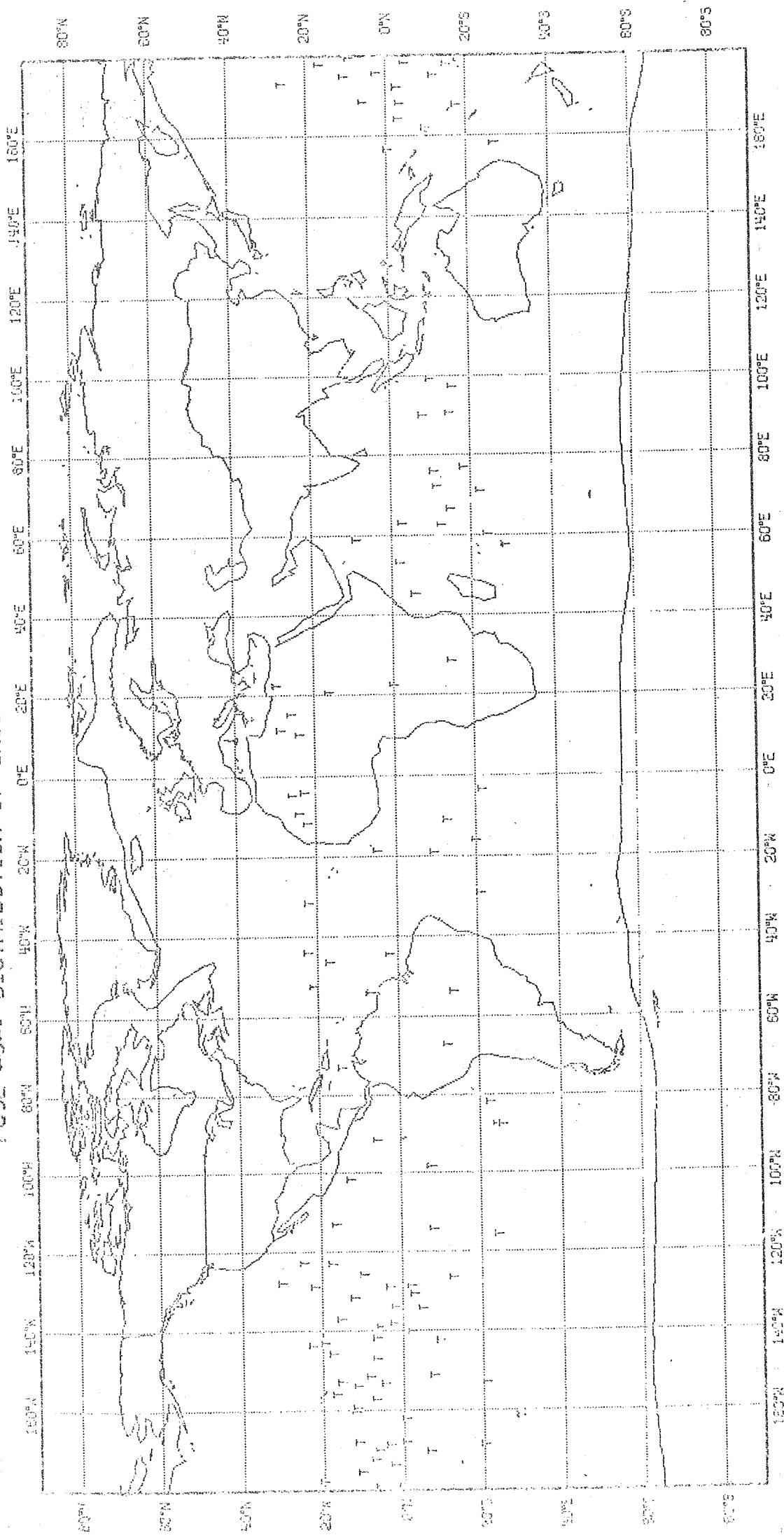


ORIGINATOR'S ID

Figure 13



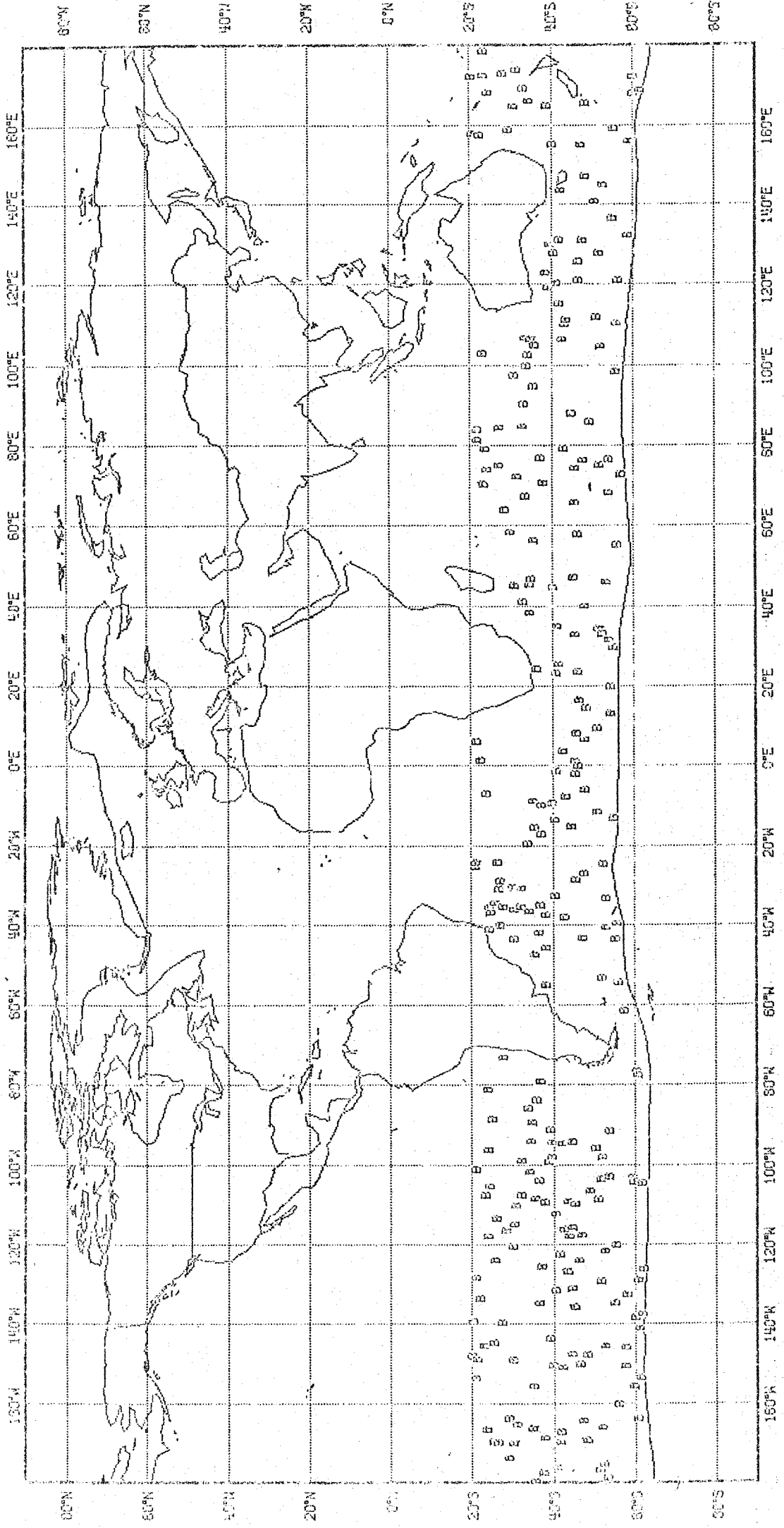
FIGURE 14: DISTRIBUTION OF CHOSEN OBSERVATIONS



CONSTANT LEVEL BALLOONS (T)

Figure 14

FGGE WCHS DISTRIBUTION OF CHOSEN OBSERVATIONS



BUDYS (S)

Figure 15

ABSOLUTE ANALYSIS ERROR OF HEIGHT, 1000MB.

□	BASIC	◇	NO OBSERVATIONS
△	BASIC+IDEAL SOP		
*	BASIC+XPCTD SOP		

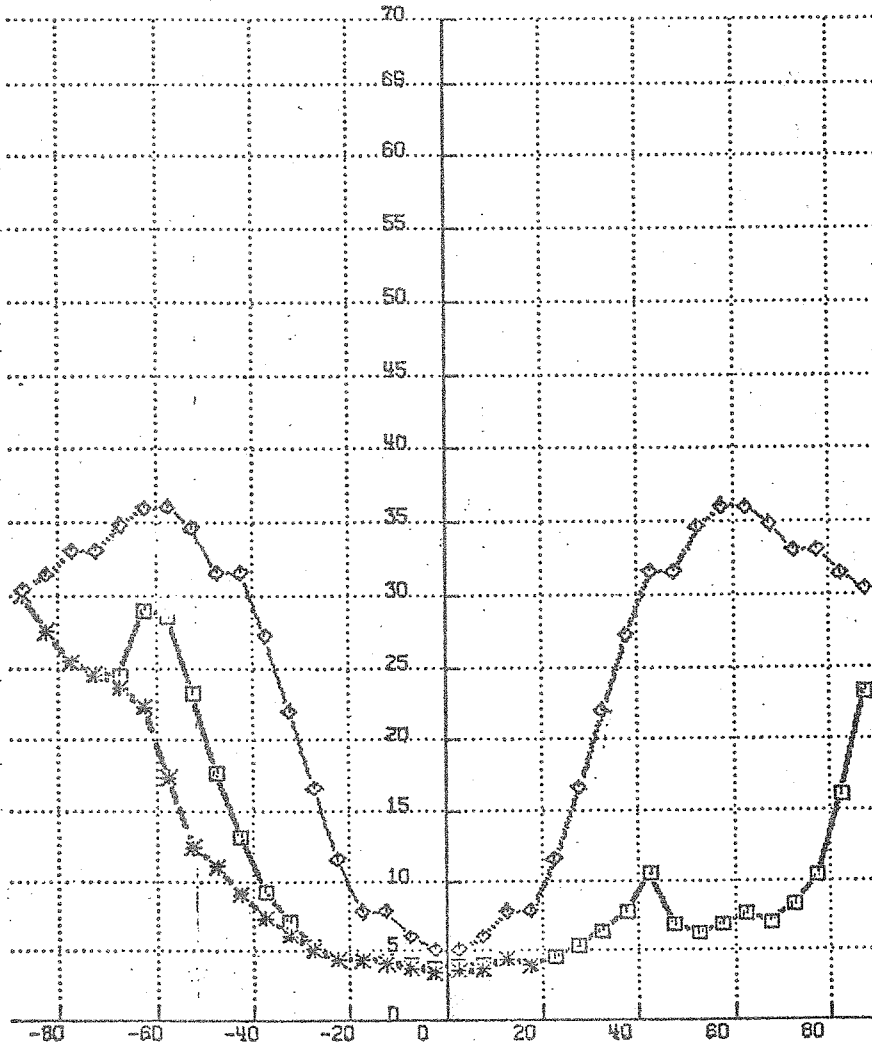


Figure 16

ABSOLUTE ANALYSIS ERROR OF HEIGHT, 850MB.

□—— BASIC
△—— BASIC+IDEAL SOP
*—— BASIC+XPCTD SOP
◇—— NO OBSERVATIONS.

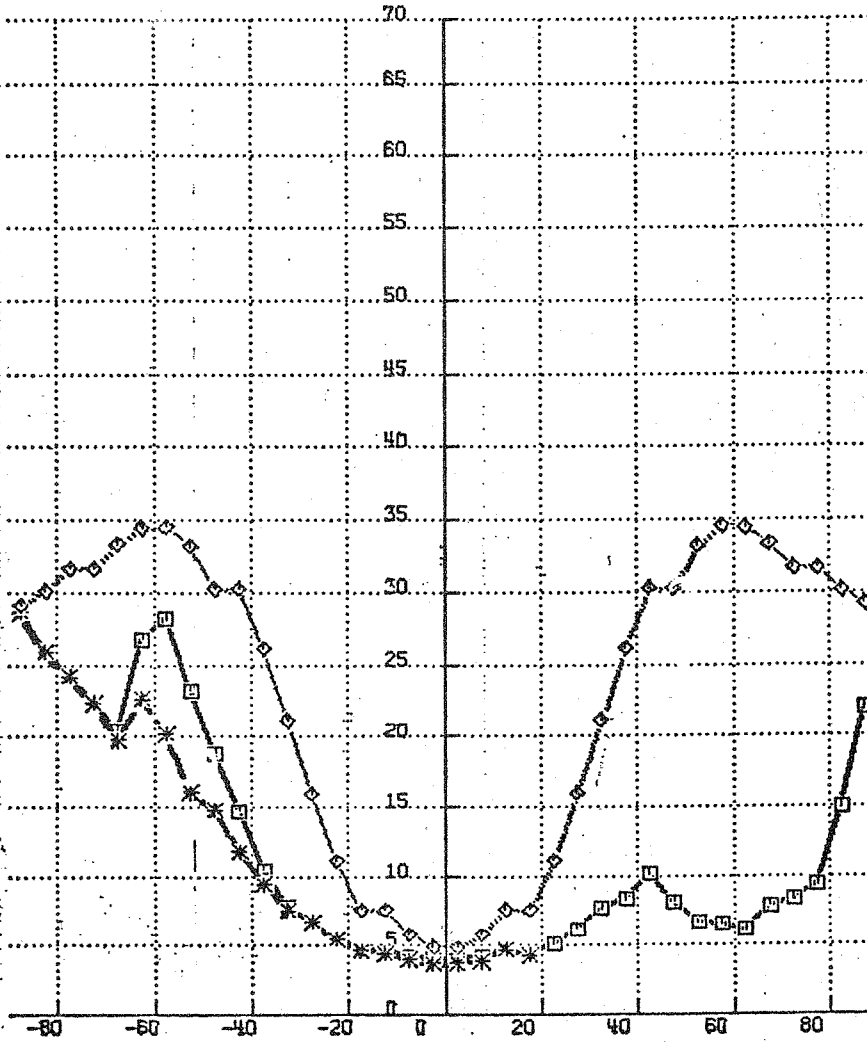


Figure 17

ABSOLUTE ANALYSIS ERROR OF HEIGHT, 500MB.

□—— BASIC
△—— BASIC+IDEAL SQP
*—— BASIC+XPCTD SQP
◇—— NO OBSERVATIONS.

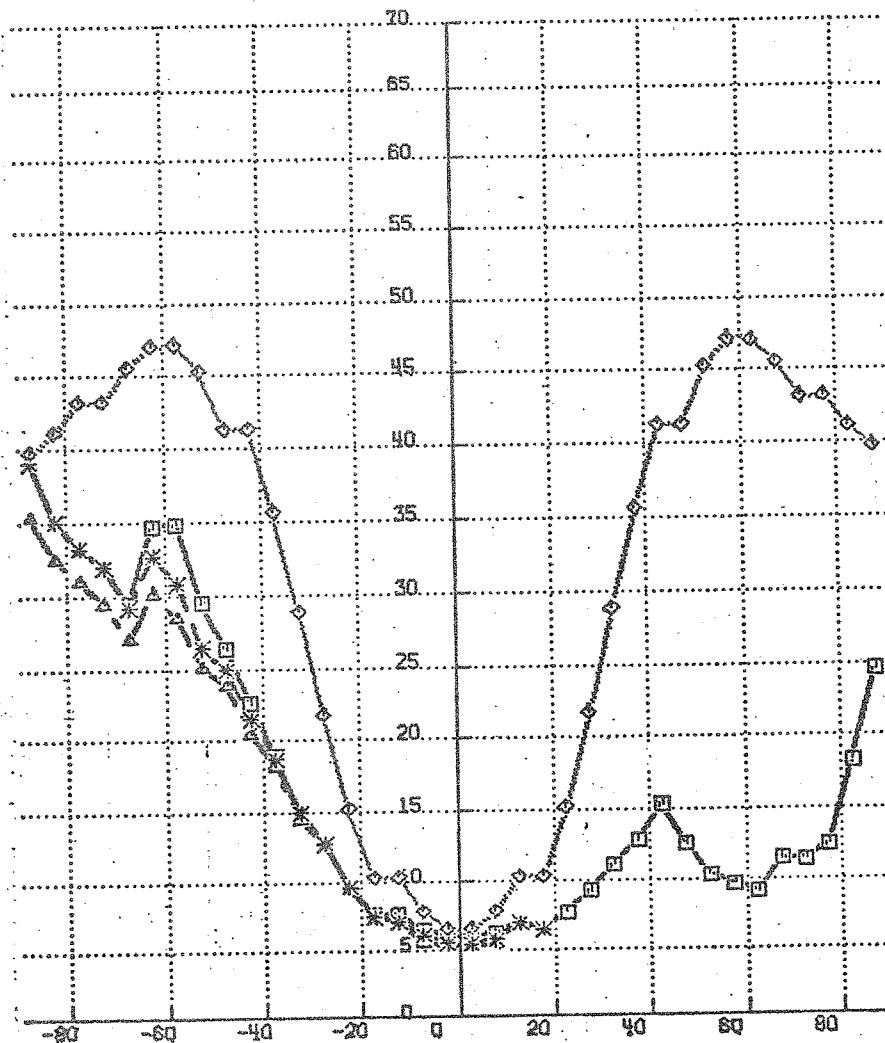


Figure 18

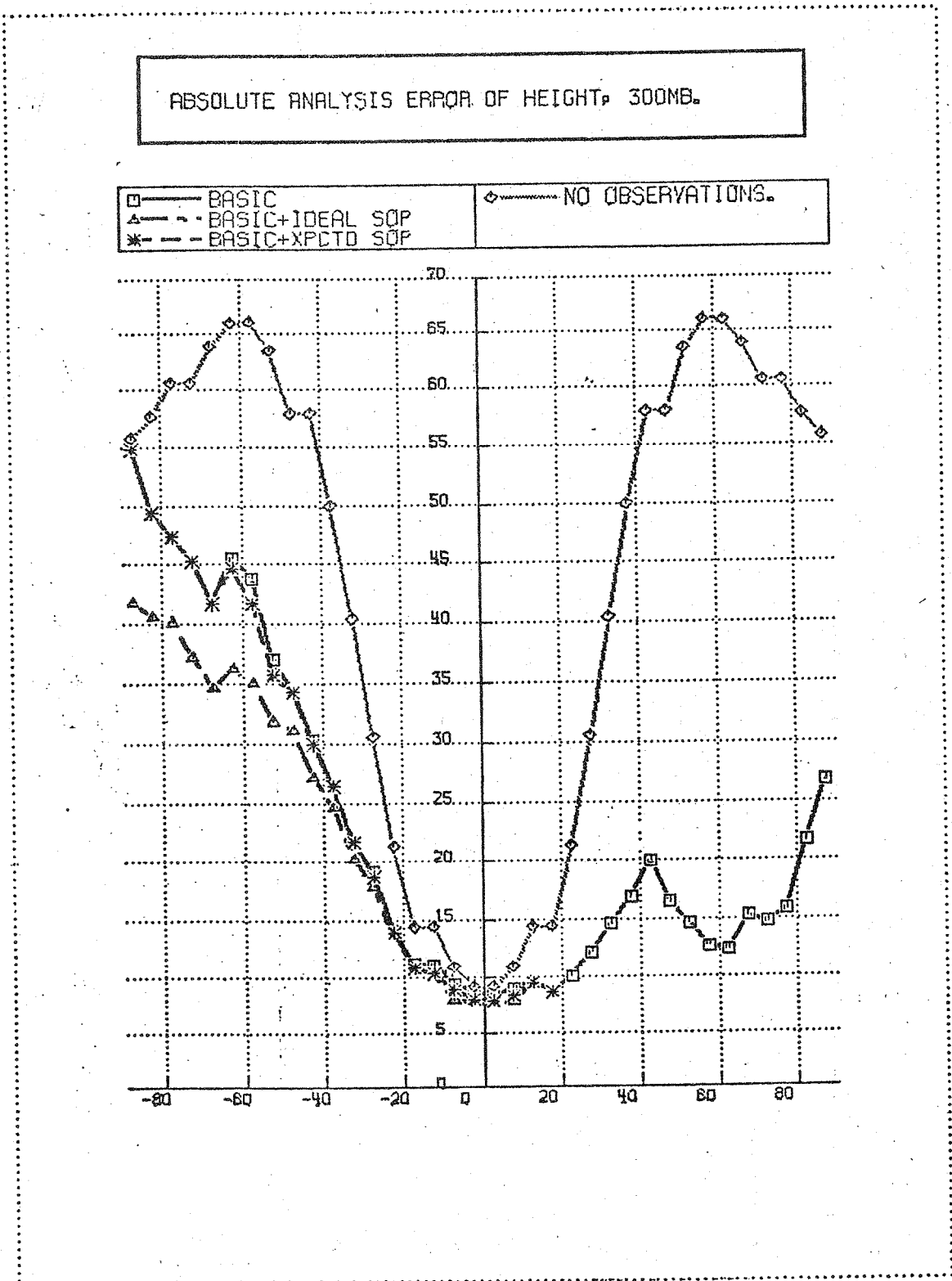


Figure 19

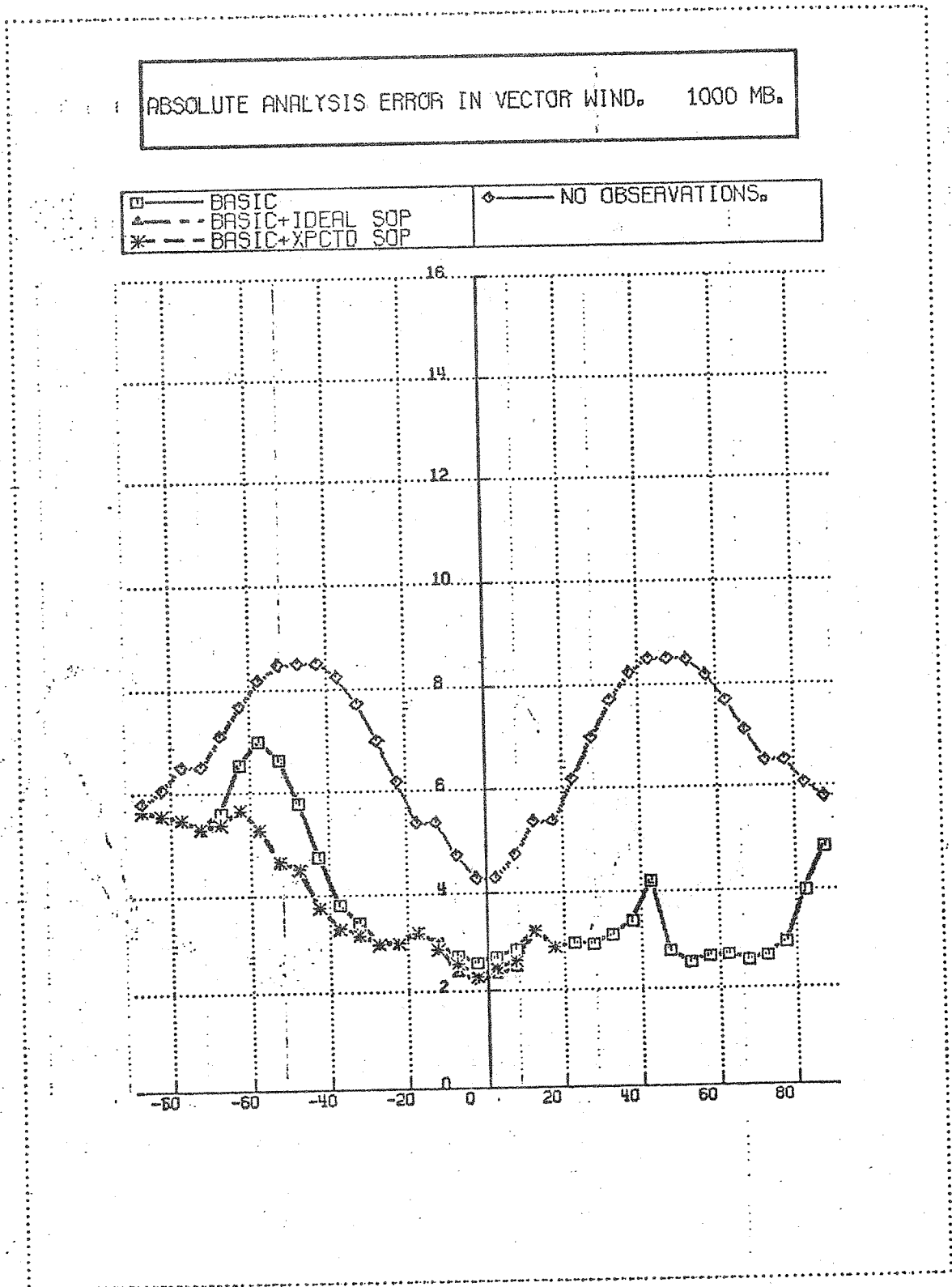


Figure 20

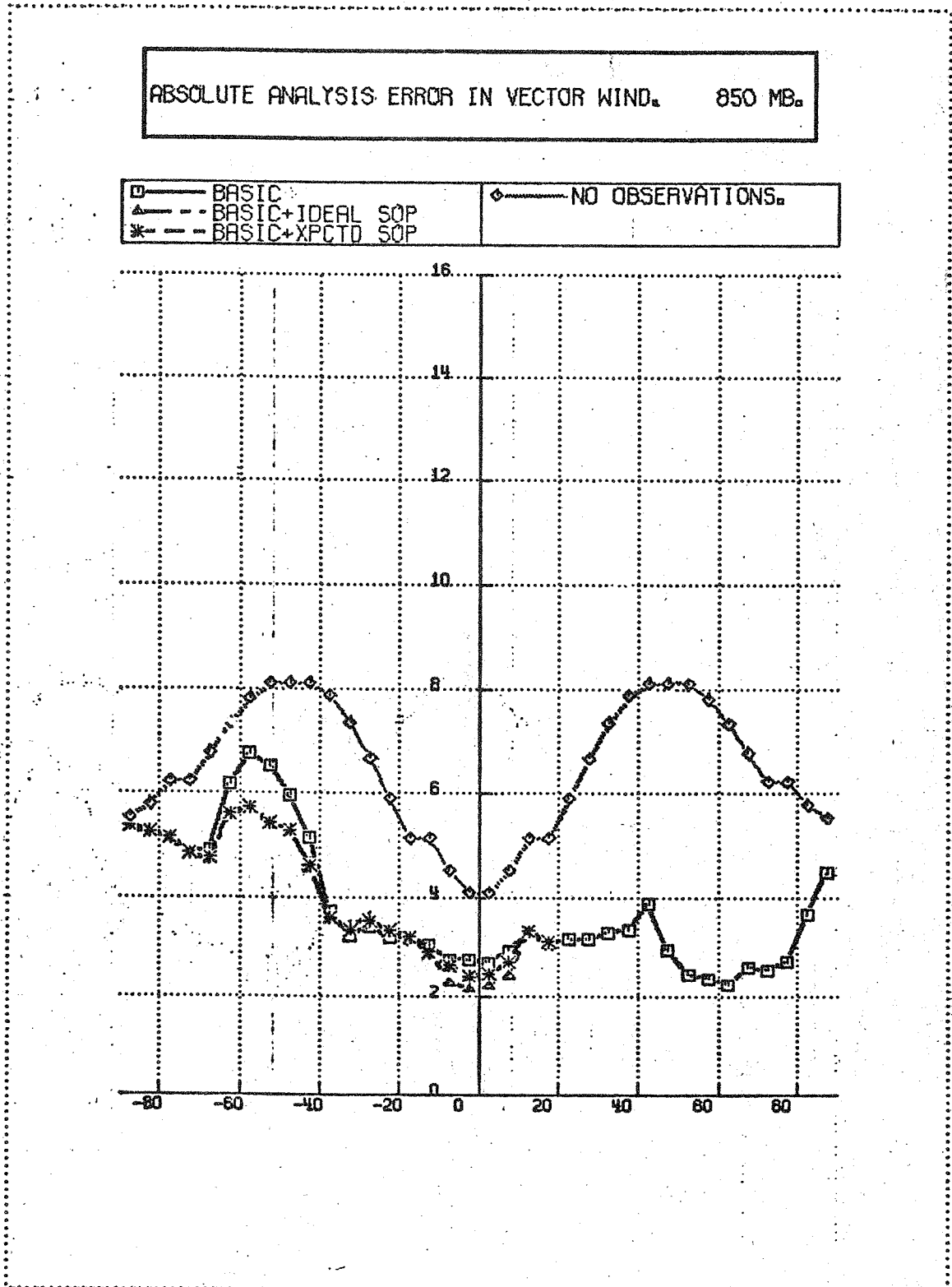


Figure 21

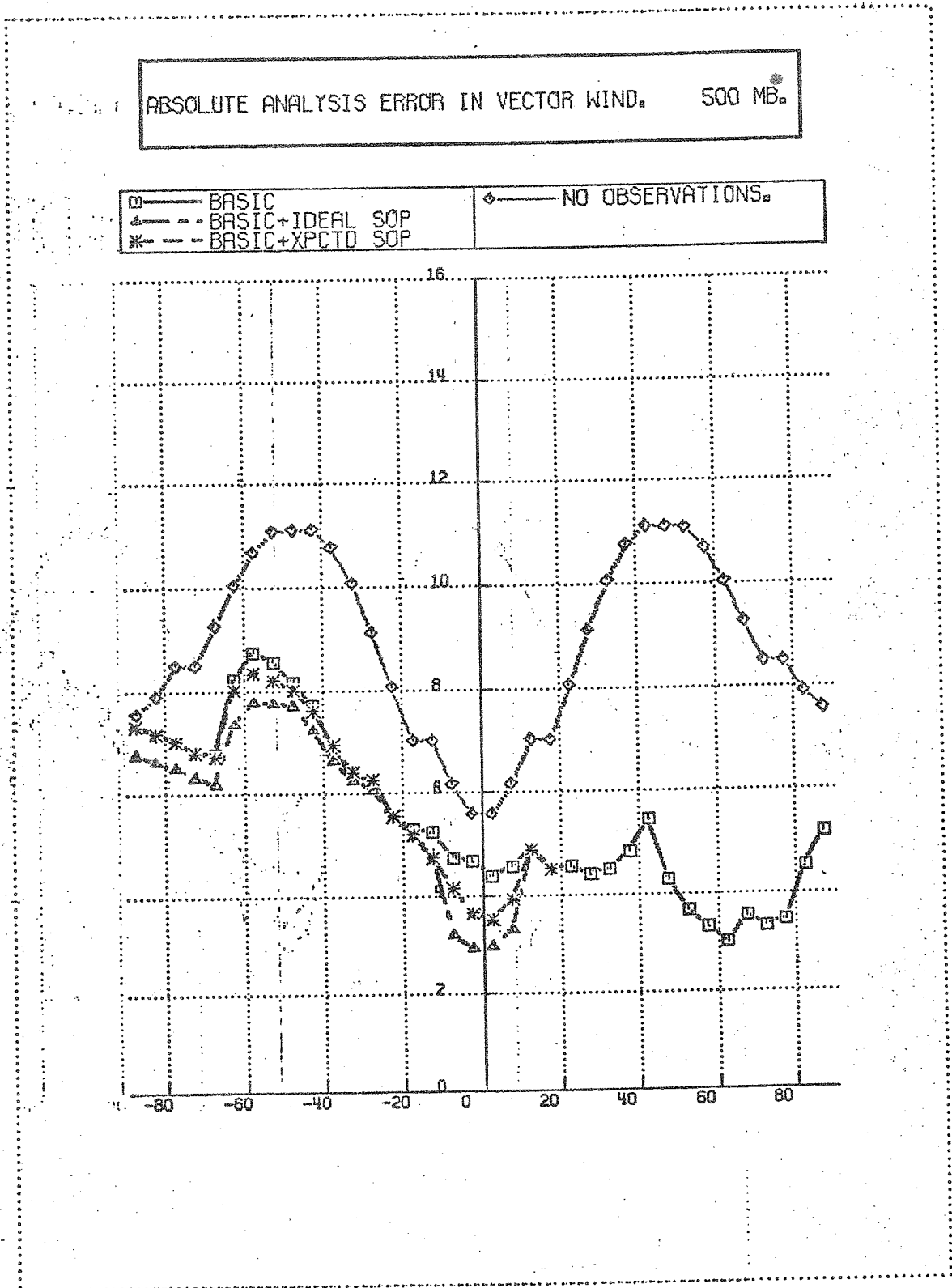


Figure 22

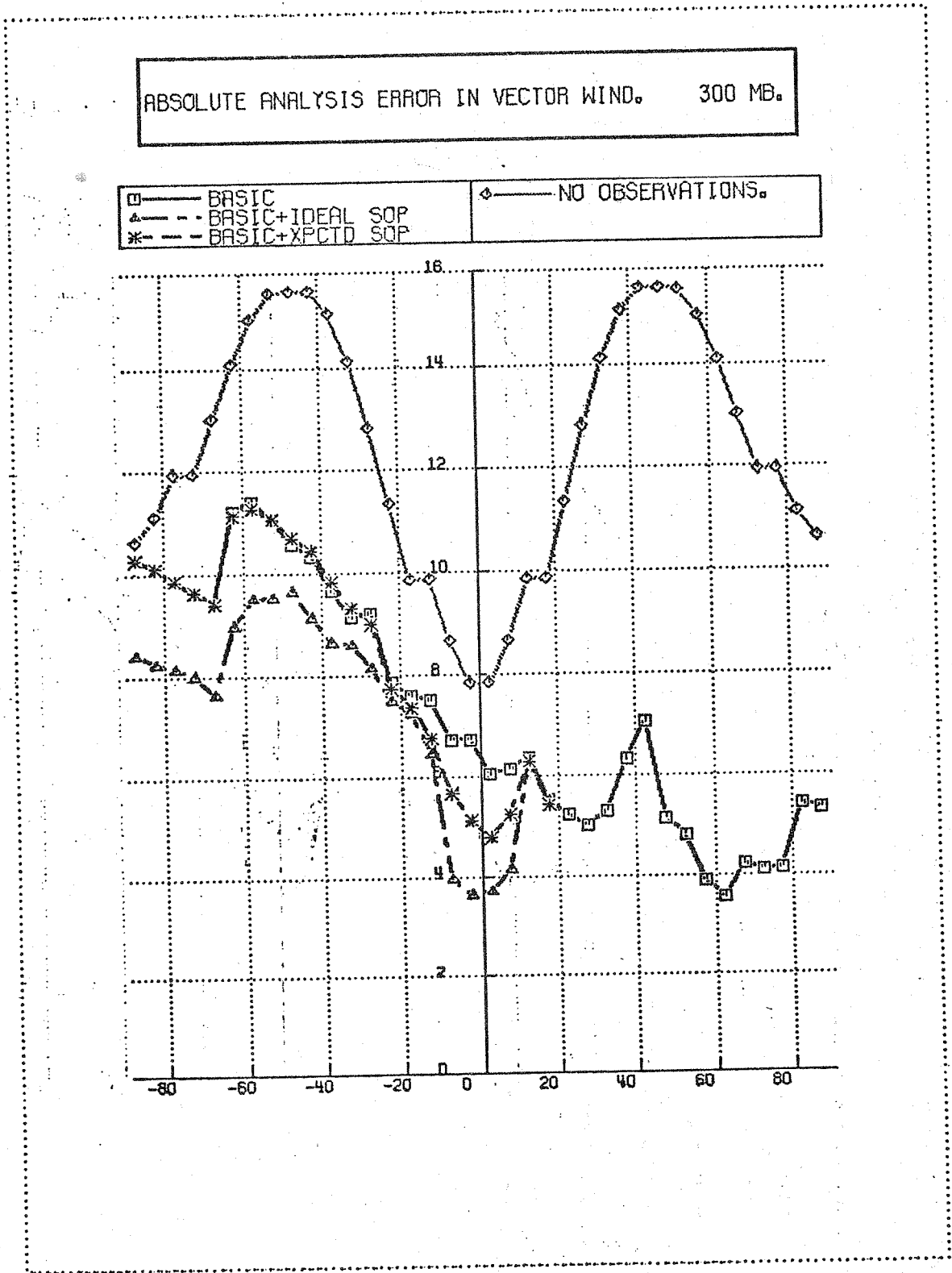


Figure 23

ABSOLUTE ANALYSIS ERROR OF HEIGHT, 500MB.

□ ——— SMALLER PRED ERR	◇ ——— NO OBSERVATIONS.
△ — — — NO OBSERVATIONS.	
* — — — BASIC	

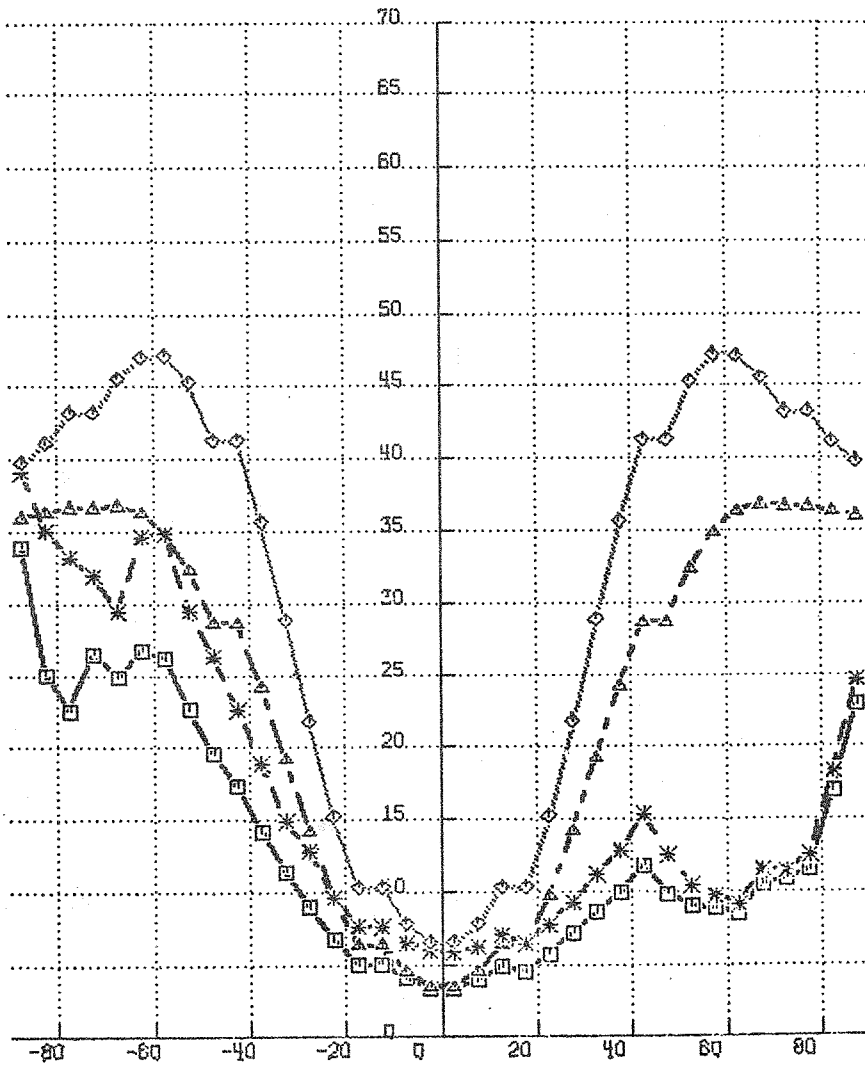


Figure 24

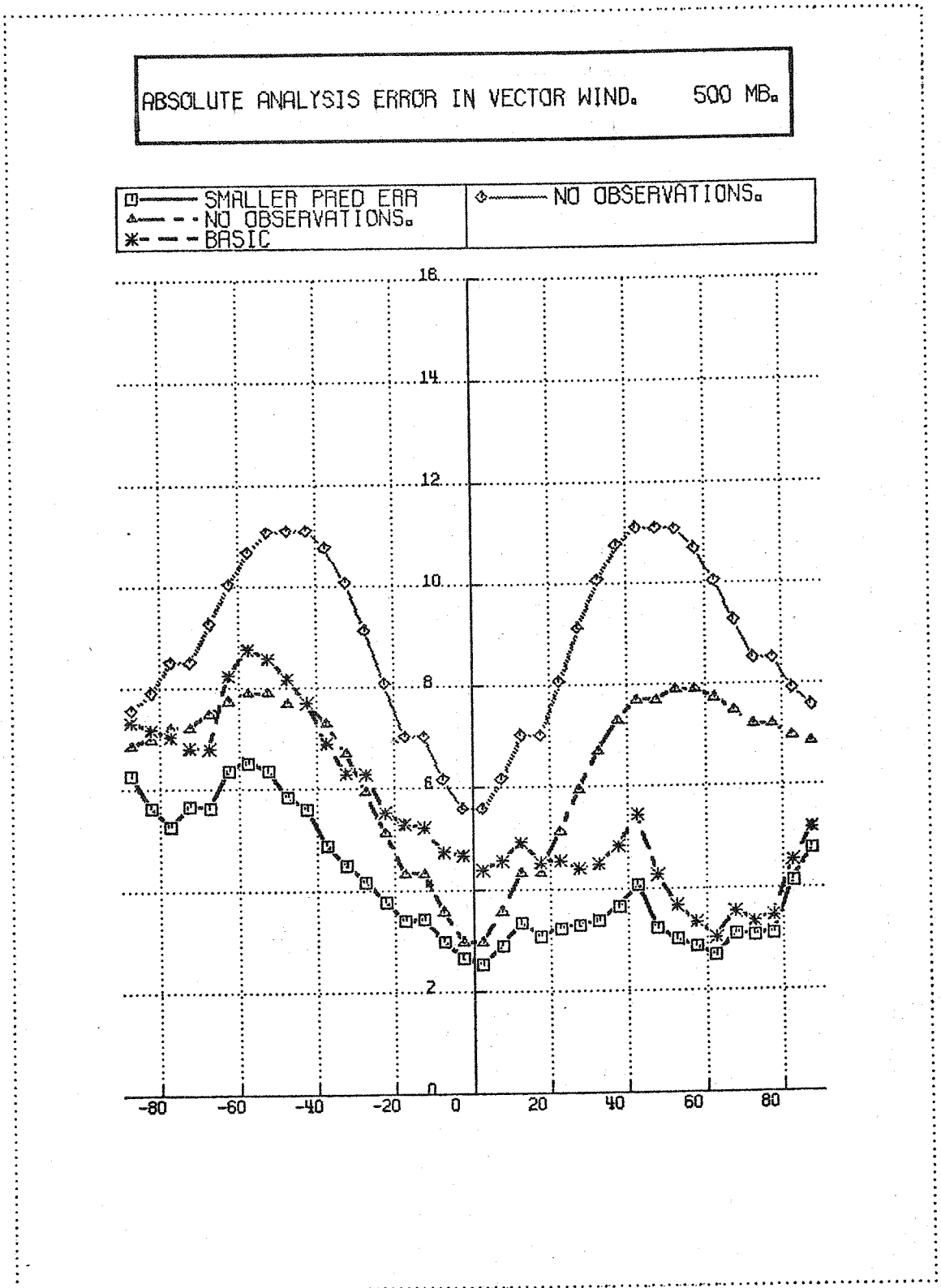


Figure 25

ABSOLUTE ANALYSIS ERROR OF HEIGHT, 1000MB.

□	—	BASIC
△	- -	BASIC-SIRS
*	- -	BASIC-SWND
◇	- - - -	BASIC-SIRS-SWND
x	- - - -	NO OBSERVATIONS.

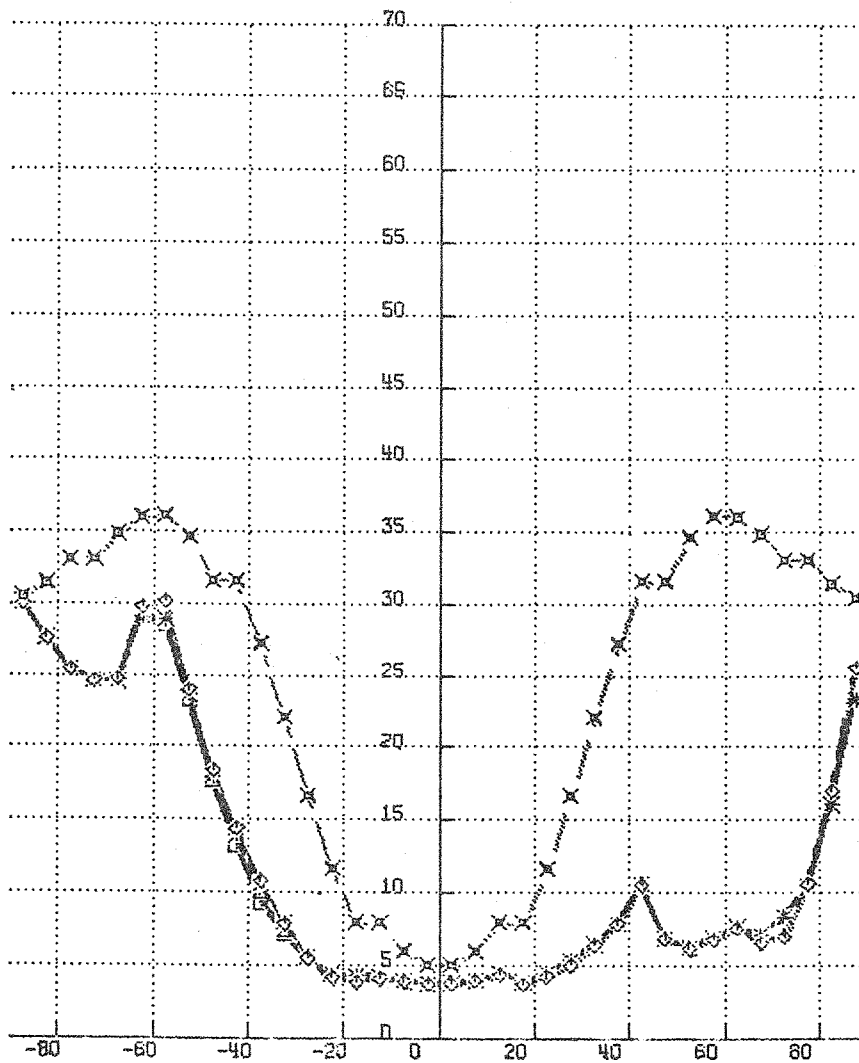


Figure 26

ABSOLUTE ANALYSIS ERROR OF HEIGHT, 850MB.

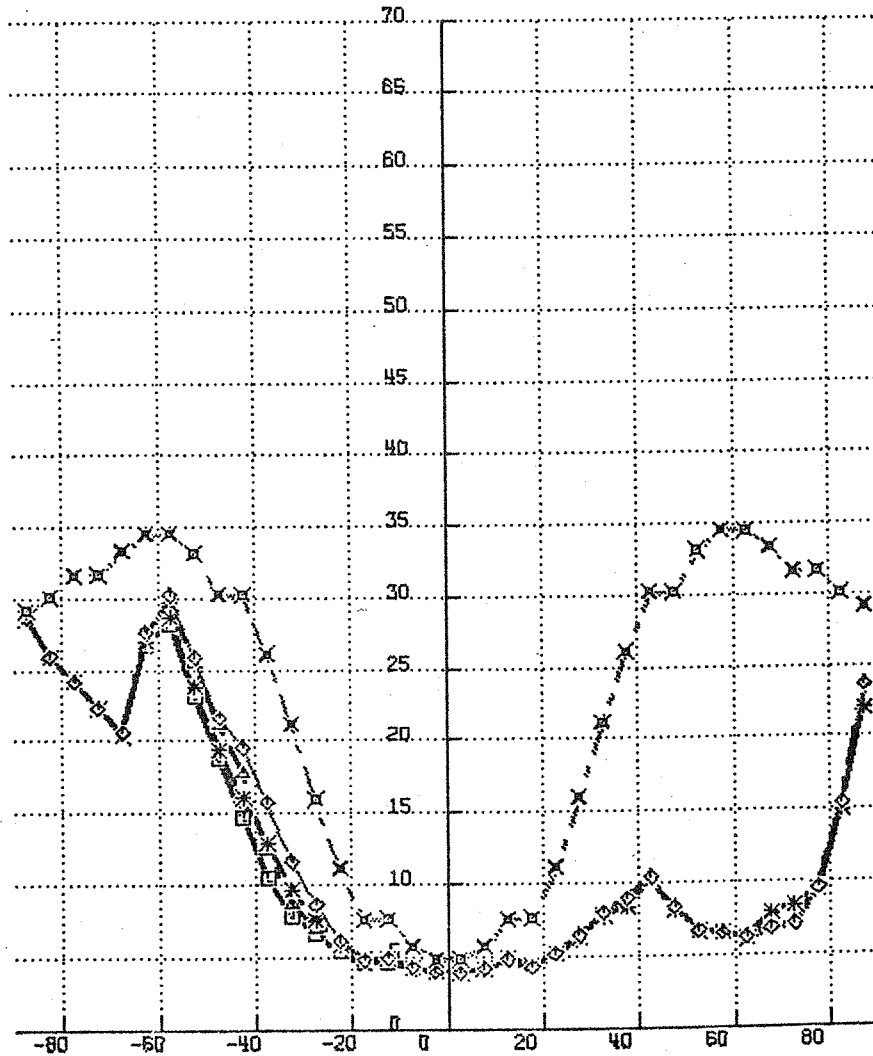
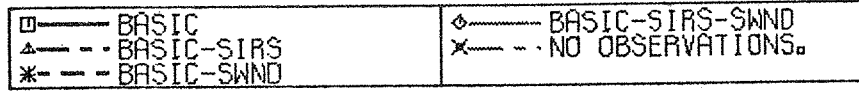


Figure 27

ABSOLUTE ANALYSIS ERROR OF HEIGHT, 500MB.

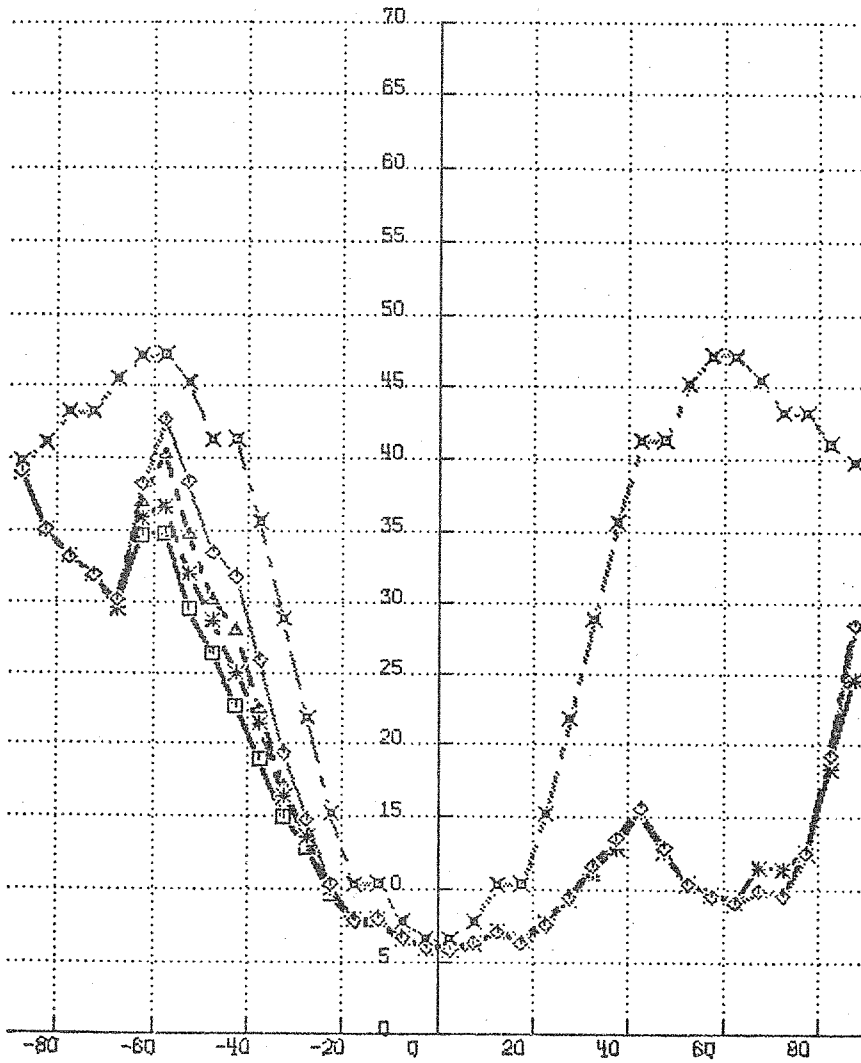
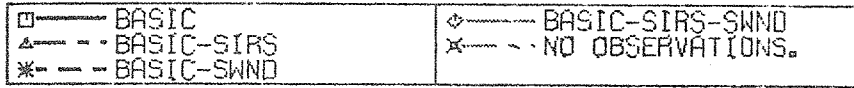


Figure 28

ABSOLUTE ANALYSIS ERROR OF HEIGHT, 300MB.

□ — BASIC	◇ — BASIC-SIRS-SWND
△ — BASIC-SIRS	× — NO OBSERVATIONS.
* — BASIC-SWND	

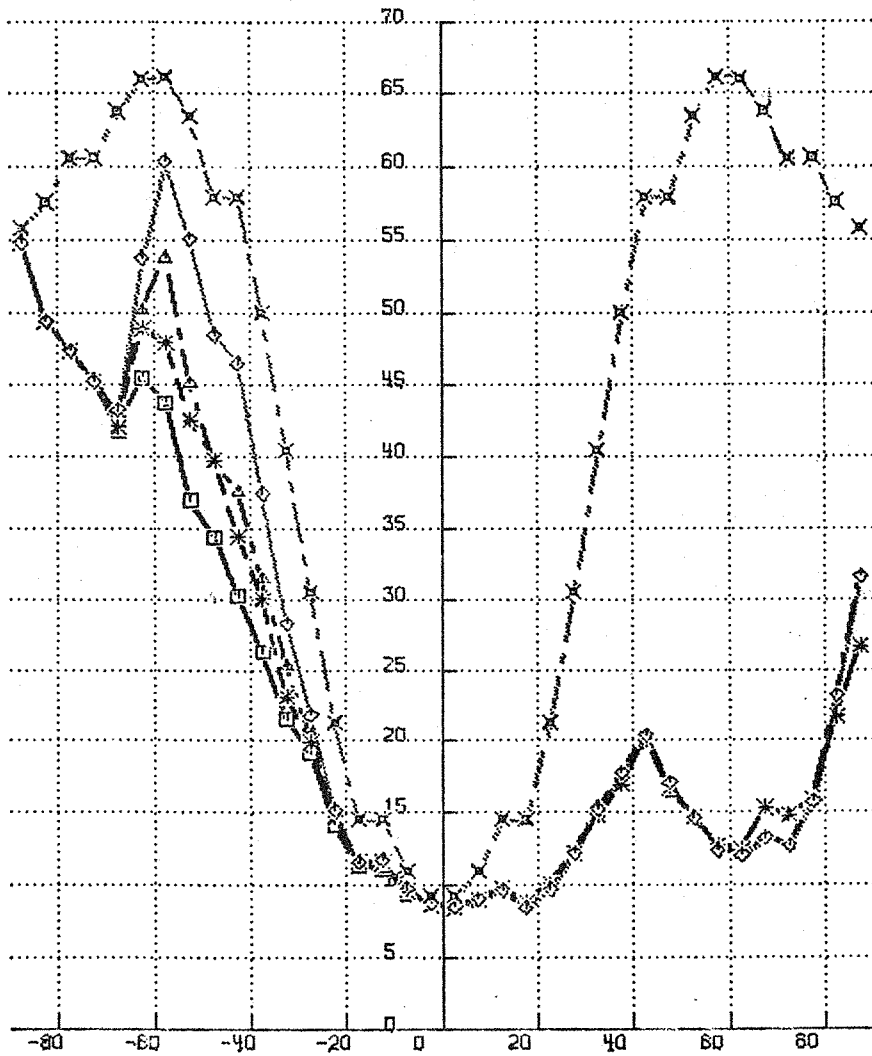


Figure 29

ABSOLUTE ANALYSIS ERROR IN VECTOR WIND. 1000 MB.

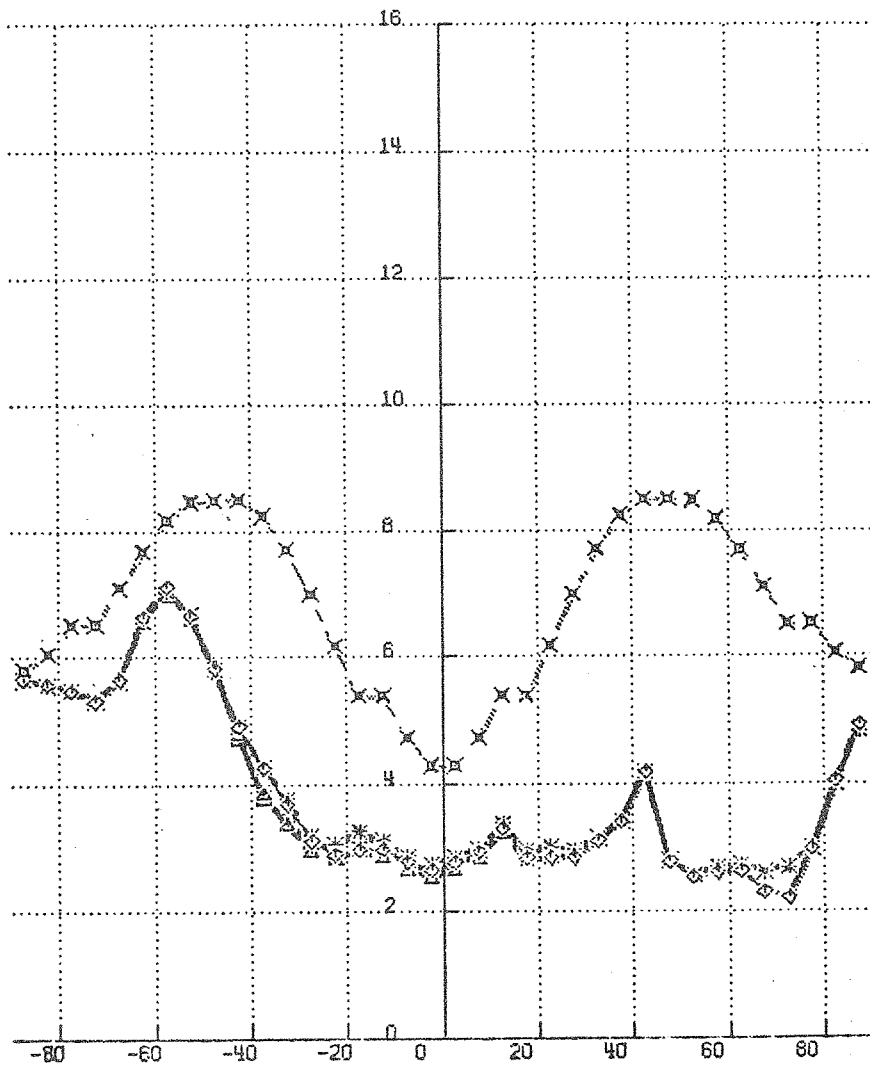
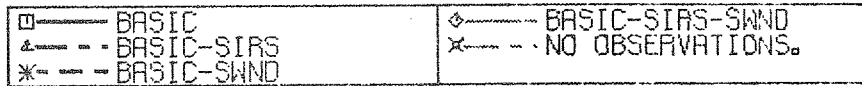


Figure 30



ABSOLUTE ANALYSIS ERROR IN VECTOR WIND. 850 MB.

□	—	BASIC
△	- -	BASIC-SIRS
*	- -	BASIC-SWNO
◇	- - - -	BASIC-SIRS-SWNO
x	- - - -	NO OBSERVATIONS.

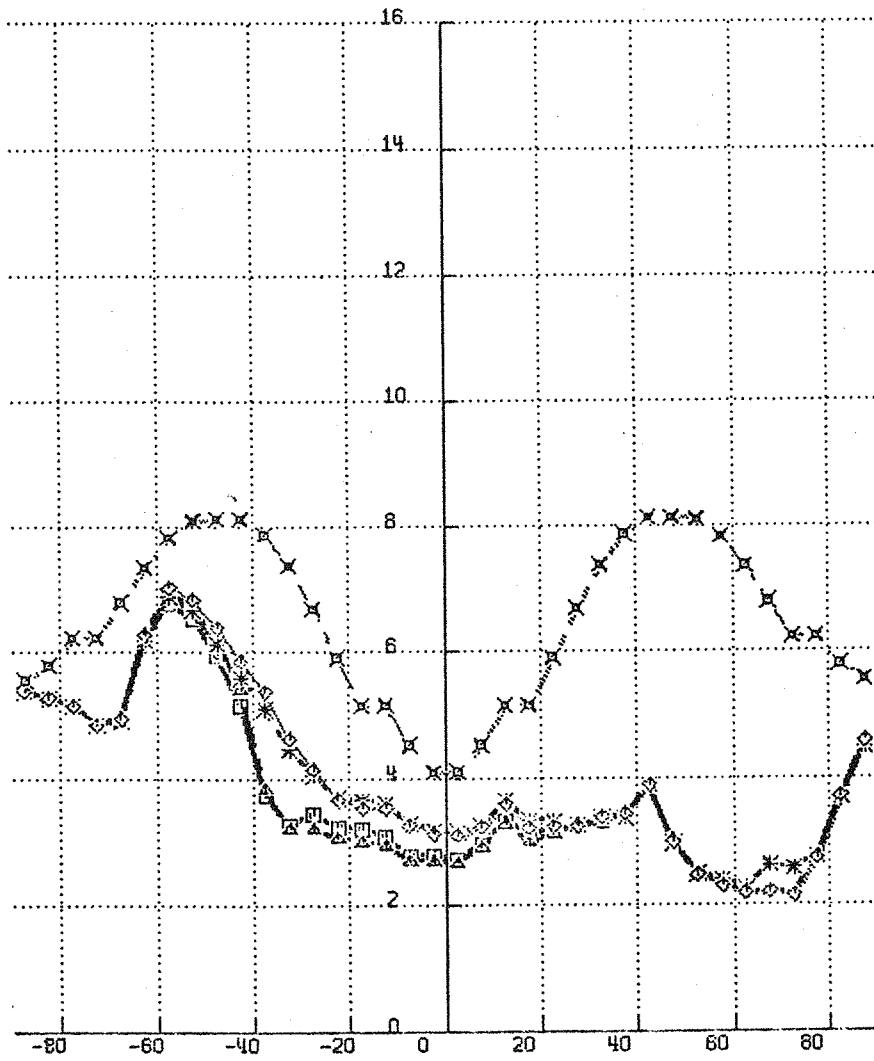


Figure 31

ABSOLUTE ANALYSIS ERROR IN VECTOR WIND. 500 MB.

□	—	BASIC	◇	—	BASIC-SIAS-SWNO
△	- -	BASIC-SIAS	×	- -	NO OBSERVATIONS.
*	- -	BASIC-SWNO			

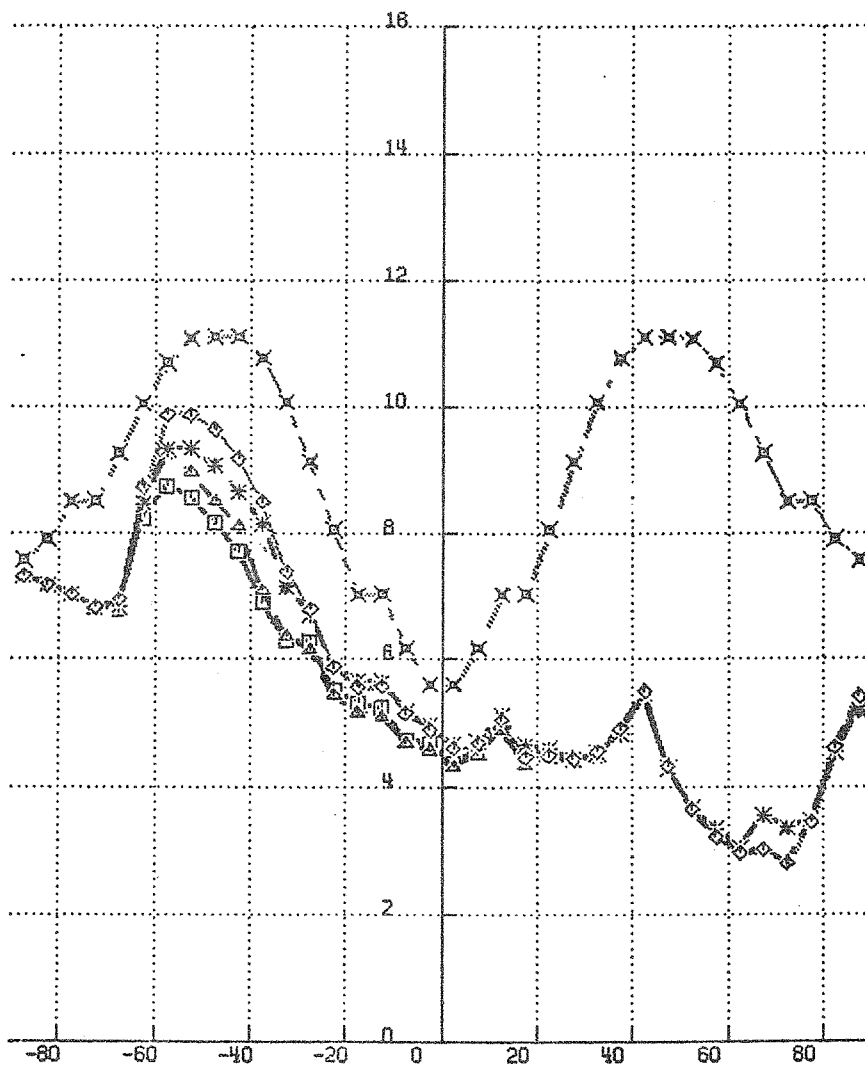


Figure 32

ABSOLUTE ANALYSIS ERROR IN VECTOR WIND. 300 MB.

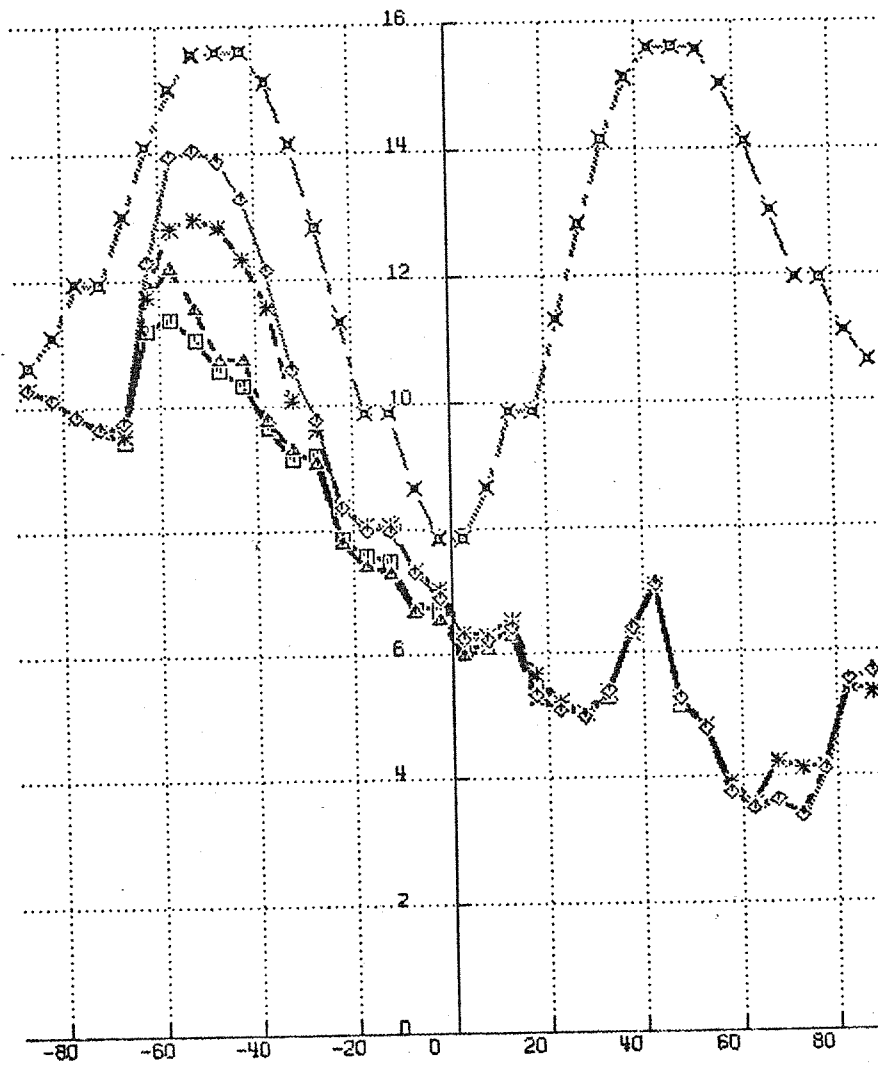
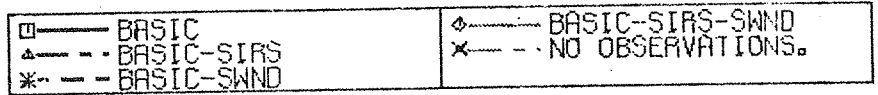


Figure 33

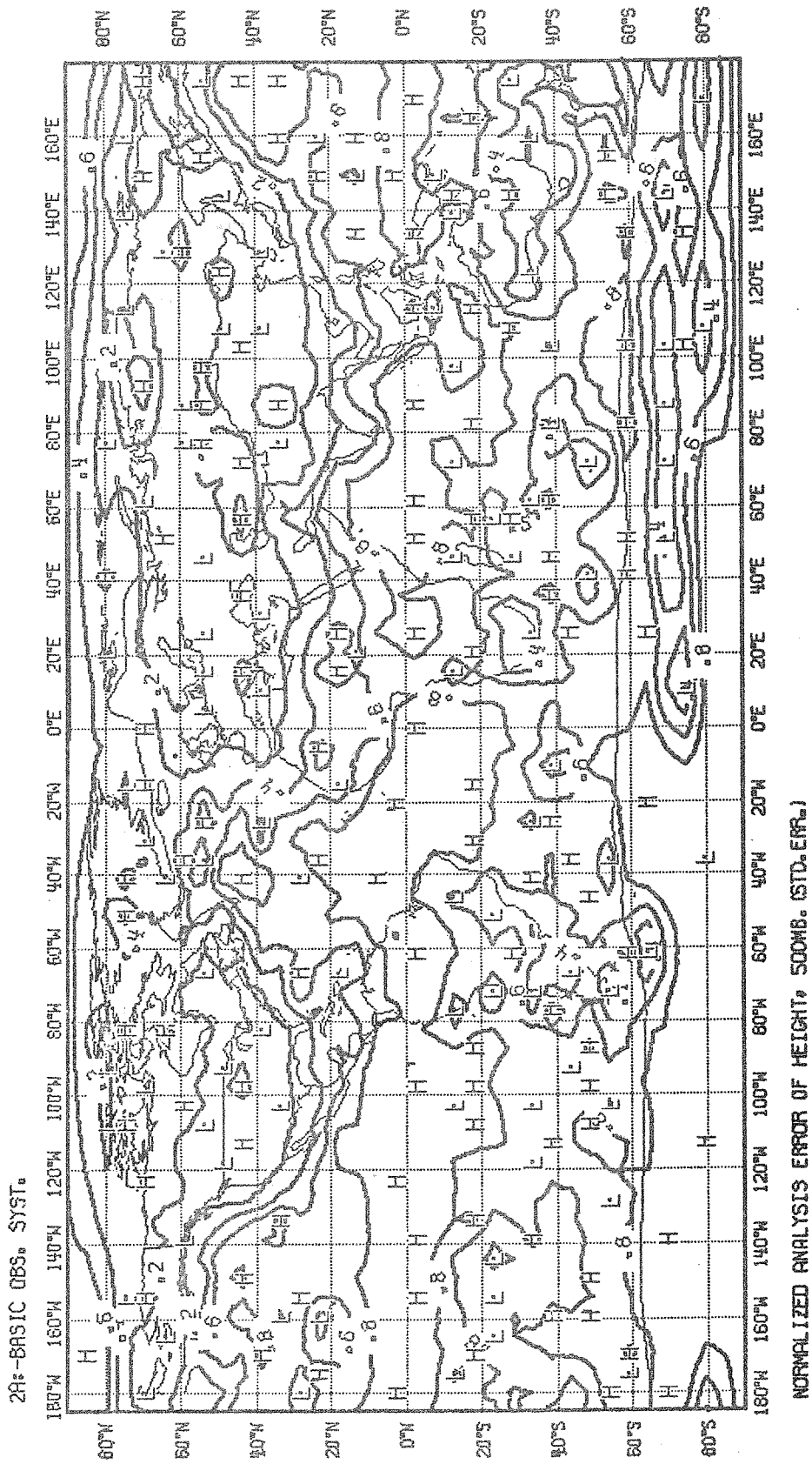


Figure 34

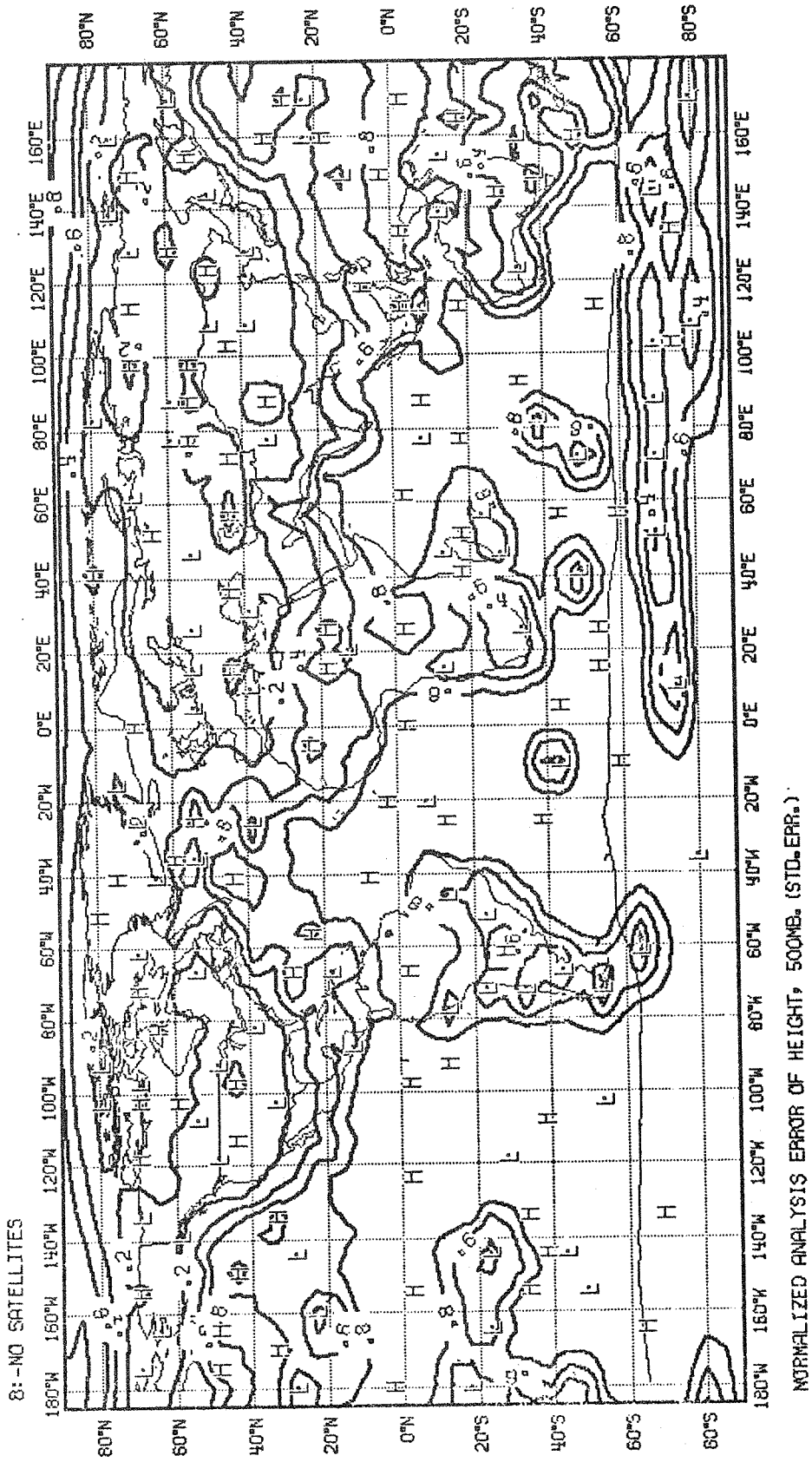


Figure 35

ABSOLUTE ANALYSIS ERROR OF HEIGHT, 500MB.

□——	SMALLER SIRS ERR	◇——	NO OBSERVATIONS.
△——	BASIC		
*——	BASIC-SIRS		

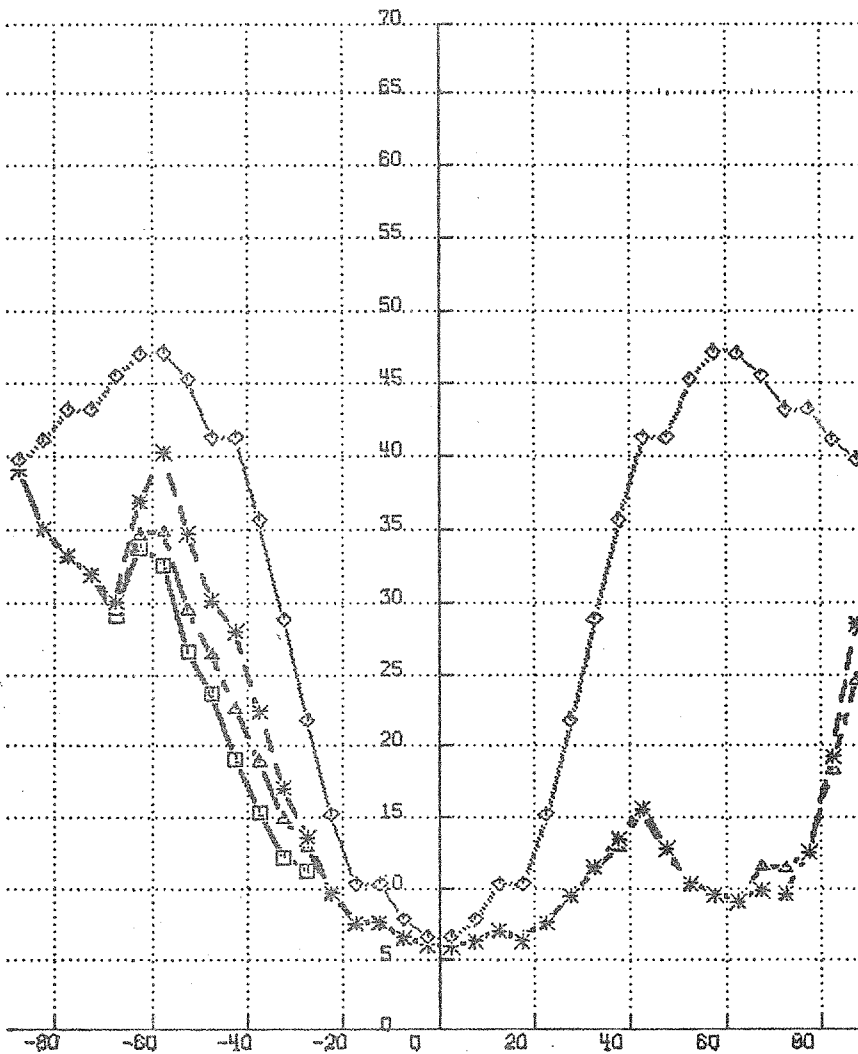


Figure 36

ABSOLUTE ANALYSIS ERROR IN VECTOR WIND. 500 MB.

□ ——— SMALLER SIRS ERR	◇ ——— NO OBSERVATIONS.
△ ——— BASIC	
* ——— BASIC-SIRS	

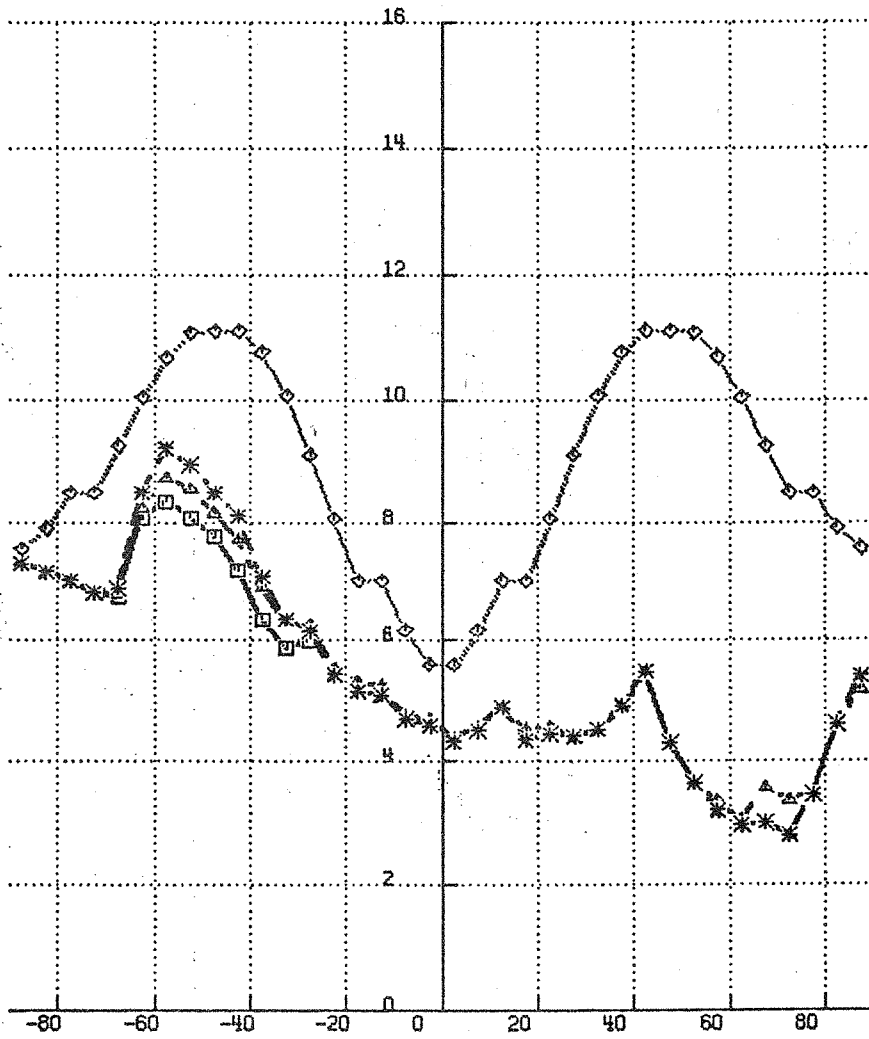


Figure 37

ABSOLUTE ANALYSIS ERROR OF HEIGHT, 500MB.

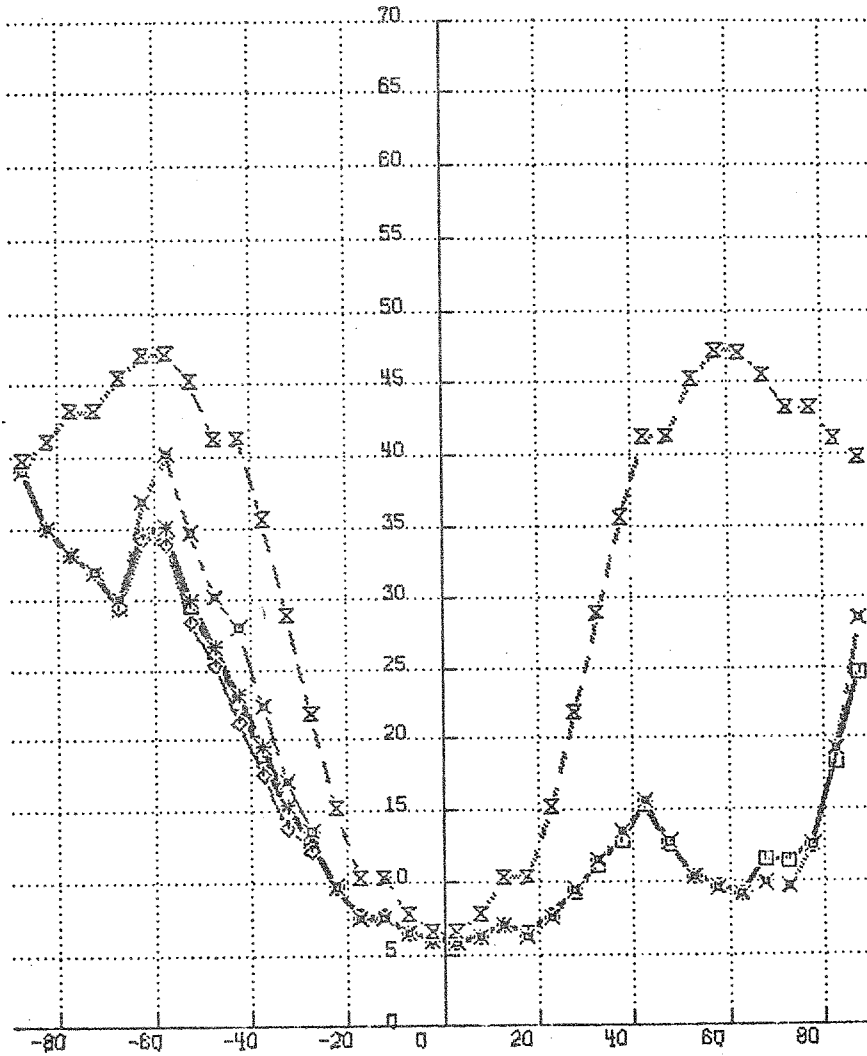
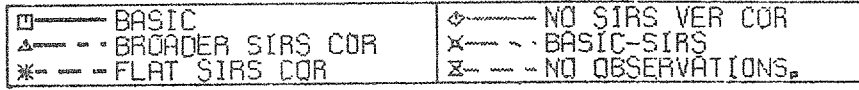


Figure 38

ABSOLUTE ANALYSIS ERROR IN VECTOR WIND. 500 MB.

□	—	BASIC	◇	—	NO SIRS VEA COR
△	- - -	BROADER SIRS COR	x	- - -	BASIC-SIRS
*	- - -	FLAT SIRS COR	⊗	- - -	NO OBSERVATIONS.

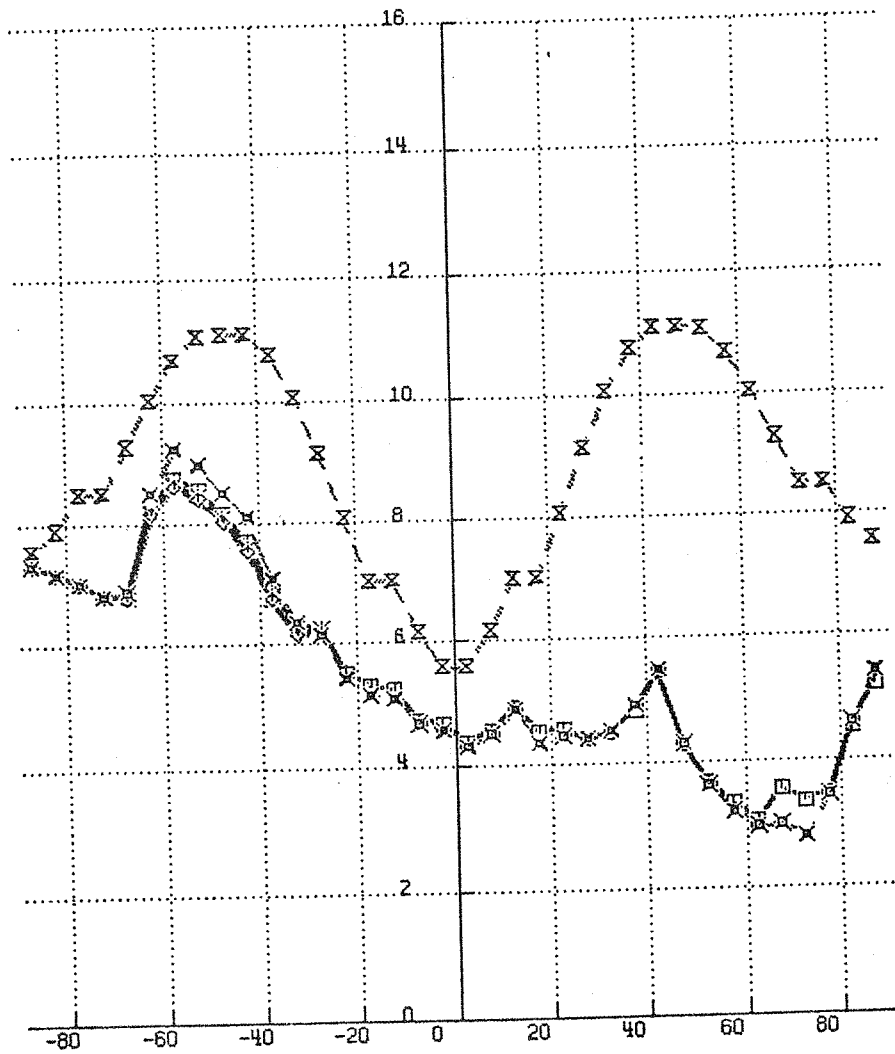


Figure 39

ABSOLUTE ANALYSIS ERROR OF HEIGHT, 500MB.

□ — FGGE A PRIORITY TO R
△ — BASIC
* — NQ OBSERVATIONS.

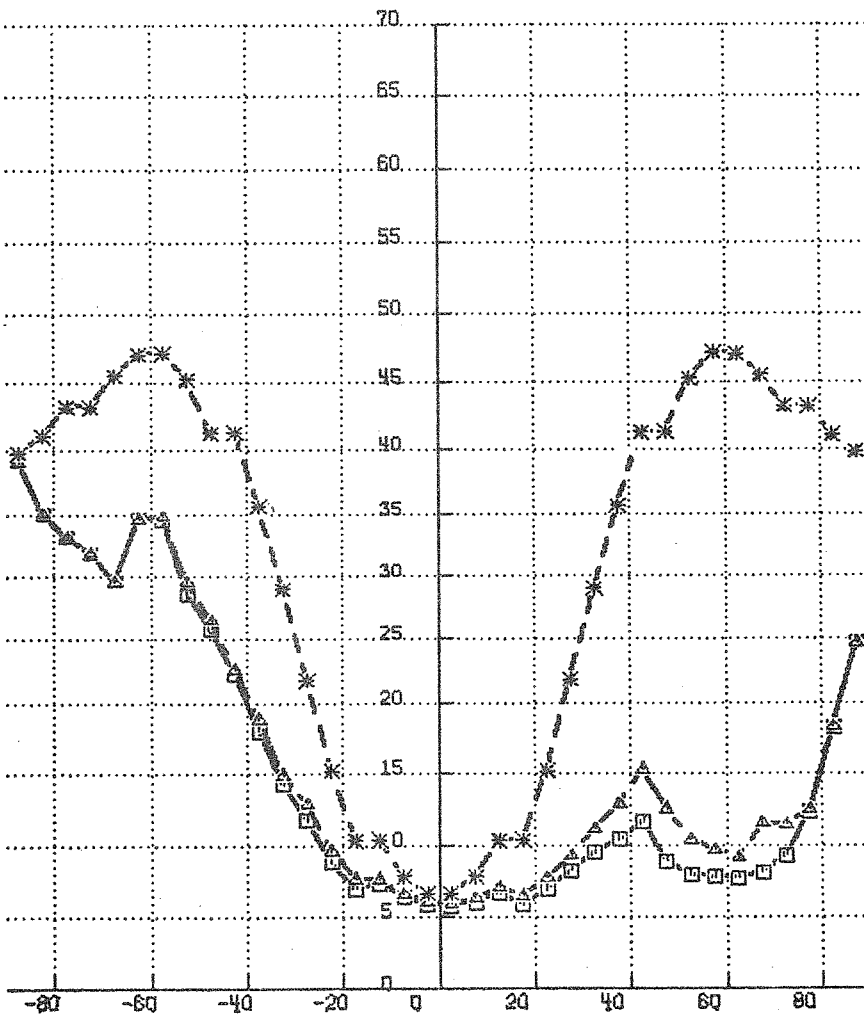


Figure 40

ABSOLUTE ANALYSIS ERROR IN VECTOR WIND. 500 MB.

□ — FGGE A PRIORITY TO A
△ — BASIC
* — NO OBSERVATIONS.

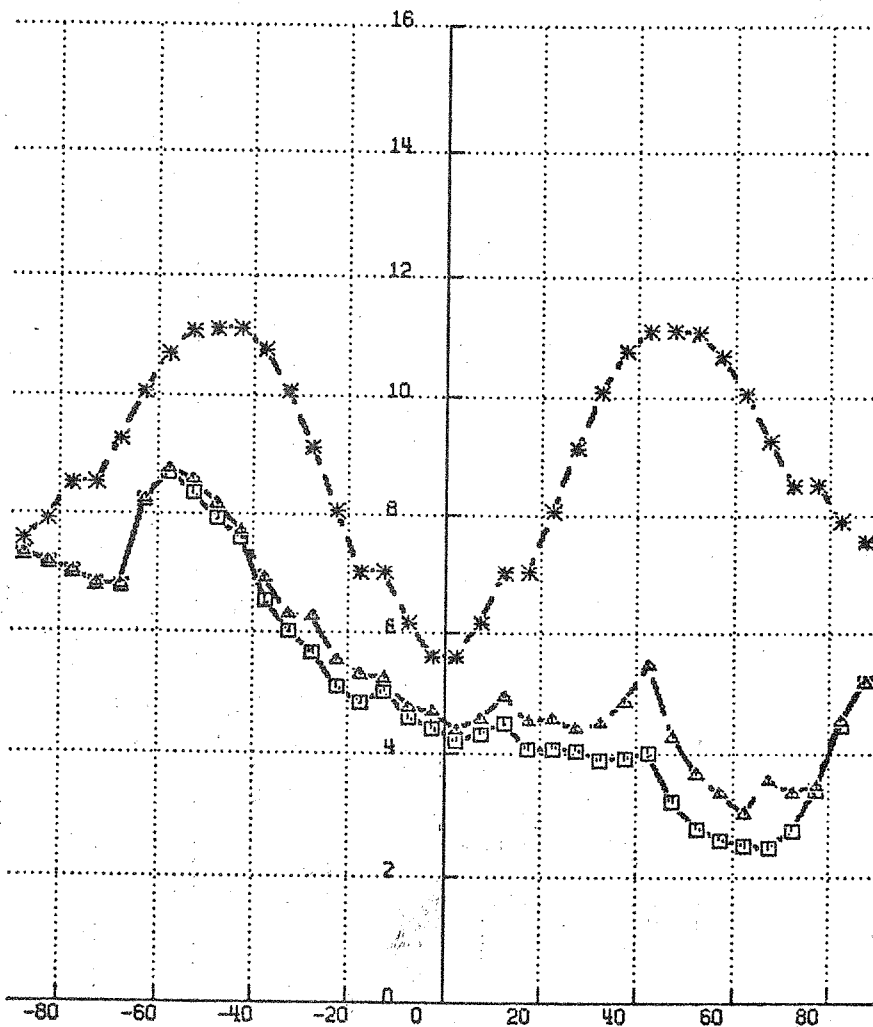


Figure 41

EUROPEAN CENTRE FOR
MEDIUM RANGE WEATHER
FORECASTS

Research Department (RD)
Internal Report No. 11

- No. 1 Users Guide for the G.F.D.L. Model
(November 1976)
- No. 2 The Effect of Replacing Southern Hemispheric Analyses
by Climatology on Medium Range Weather Forecasts
(January 1977)
- No. 3 Test of a Lateral Boundary Relaxation Scheme
in a Barotropic Model
(February 1977)
- No. 4 Parameterisation of the surface fluxes
(February 1977)
- No. 5 An Improved Algorithm for the Direct Solution of
Poisson's Equation over Irregular Regions
(February 1977)
- No. 6 Comparative Extended Range Numerical Integrations
with the E.C.M.W.F. Global Forecasting Model
1: The N24, Non-Adiabatic Experiment
(March 1977)
- No. 7 The E.C.M.W.F. Limited Area Model
(March 1977)
- No. 8 A Comprehensive Radiation Scheme designed for
Fast Computation
(May 1977)
- No. 9 Documentation for the E.C.M.W.F. Grid Point Model
(May 1977)
- No.10 Numerical Tests of Parameterisation Schemes at an
Actual Case of Transformation of Arctic Air
(June 1977)
- No.11 Analysis Error Calculations for the FGGE
(June 1977)

