U. S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
NATIONAL METEOROLOGICAL CENTER

OFFICE NOTE 280

On the Precision of Calculating Atmospheric Energy Parameters from NMC Optimum Interpolation Analyses

Alvin J. Miller
Climate Analysis Center

and

Lauren L. Morone
Development Division

September 1983

Abstract

Utilizing error estimates of wind and temperature that are available from the National Meteorological Center optimum interpolation analysis technique, a Monte Carlo analysis is performed on the data of December 20, 1981 to provide error estimates of various energy budget calculations based on variance and co-variance elements. It is determined that the precision of estimates of integrated hemispheric energy parameters are on the order of 1%. For specific parameters, however, such as momentum transport, temperature variance etc., the calculated precision estimates can be quite variable in pressure and latitude. In regions where the parameter estimates are large and well defined, the standard deviations are better than 10%. Outside of these regions, however, where the values are small, the precision estimates become considerably worse and can be as large as 50-100%.

1. Introduction

A common question concerning atmospheric synoptic analyses asks the degree of accuracy and precision of the various quantities. While it is difficult enough to answer this question in terms of the analyzed parameters such as wind, temperature etc., it is not at all clear how to answer it for cross-product terms such as eddy momentum or heat flux. Recently, this question has attained added significance in that these analyses are the source of many studies on climate change and the possibility of anthropogenic influence on the atmosphere. As such, it is important that any changes and impacts on the analyses by either a change in data base and/or analysis technique be documented and numerically assessed.

For an analysis technique as complicated as the NMC optimum interpolation analysis (McPherson et al, 1979) one approach might be to apply a Monte-Carlo technique to the data. That is, using the estimated errors of the observations, allow the data to vary in a random Gaussian sense, apply the analysis system and examine the different resultant analyses. Unfortunately, such a procedure is very costly in terms of computer time. Therefore, we have adopted a variation on this method.

One of the strong points of the optimum interpolation analysis system is its ability to utilize error estimates of the various data sources to produce error estimates (in a root mean square sense) at each grid-point of the analysis. The details of this system will be discussed below. Simply stated, the estimated error variance assigned to an analyzed grid point value of wind, temperature or height depends on the type and quality of the data used in the analysis of that grid-point. For example, radiosonde winds have smaller assigned errors than aircraft winds and cloud-tracked winds. From this gridded error array, then, it is a relatively simple matter to estimate the precision of the individual analysis.

Our approach is to assume that an individual gridded analysis for a particular day and time is only one of many possible representations of the actual atmosphere at that day and time. We also assume that all the possible representations at a grid point are distributed normally, with mean equal to the analysis value and standard deviation equal to the assigned estimated analysis error. Utilizing a Gaussian random error generator with the above properties (with certain modifications described below to allow for synoptic space scale correlation), alternate gridded fields can be constructed within the error bounds of the intitial observations. From each field, the zonal average and cross-product terms can be calculated (e.g. Miller and Hayden, 1978; Hauser and Miller, 1978) and the differences among the results compared.

In practice we selected December 20, 1981, 00 GMT as representative of both the general circulation and the data coverage during the Northern Hemisphere winter. Using this as our base field, the above approach was done five separate times resulting in five independent "possible representations" of the original map.

One difference from this approach compared to the original "ideal" approach described earlier is that we would expect our results to have greater average variance. That is, we have added variance to the calculations without allowing the analysis system the opportunity to smooth it. The degree to which the analyses are smoothed is, however, of interest by itself and a discussion of this aspect is incorporated within the text.

2. Methodology

a) Analysis Errors

One of the by-products of an optimum interpolation (OI) analysis is an estimate of the error associated with the analysis. It is available as a result of the minimization process of OI (Bergman, 1979). Estimated analysis

errors computed by OI depend on the quality, quantity and distribution of the data. The use, at NMC in the OI analysis procedure of an analytic function to approximate forecast error covariances instead of actual forecast error covariances means that the estimated analysis errors depend also on the covariance model assumed. See Bergman (1979) for more details on forecast error modelling at NMC.

At each grid-point, the analysis is provided with knowledge of the number of observations available for the analysis at that grid-point, the location of those observations in relation to the grid-point and the error of each observation. The analysis procedure uses this information to compute the relative influence or weight that each observation will have in determining the analyzed value of that grid-point. These weights are then used to compute the estimated analysis errors. Thus, the smallest values of analysis errors will be located in regions where a dense network of high quality radiosonde data is available such as North America and Europe. Maximum values of estimated analysis errors would be expected where the data are sparse such as the interior regions of Africa and South America. Areas in which satellite observations predominate such as the oceans, will have analysis error values in between, since satellites provide good data coverage but the observations are considered to be of lower quality than radiosondes.

The amount and distribution of data around a particular grid-point is well known and can be described to an analysis procedure exactly. The relative quality of observations is also well known; for example, that radiosonde temperatures are more accurate than satellite temperatures and radiosonde wind measurements are more accurate than cloud tracked winds. Therefore, the relative values of the estimated analysis errors and the locations of the maximum and minimum values can be considered to be a good representation of the average performance of the analysis.

In order to see the danger associated with attaching special significance to the absolute values of any particular analyses's errors, we must look at the method by which the analysis is made aware of the quality of each observation. The quality of each type of observation is represented by a root-mean-square observational error. Although these values are based on the best measurements available, they still may have uncertainties associated with them. At NMC they are set constant in time and horizontal space and allowed to vary only in the vertical. Thus, limited knowledge about observational errors does not allow them to portray to the analysis, changes in the accuracy of an instrument's measurement due to age, different manufacturers of the instrument, the synoptic regime in which the measurement is taken (except in the case of satellite temperature measurements where distinction is made between clear, partly cloudy and cloudy retrievals) or different procedures used to collect and compute the measured values.

It is evident from this discussion that while the relative values of the estimated analysis error from an OI analysis give a good general idea of the performance of the analysis, the absolute values assigned by the analysis are only as good as user-supplied estimates of observational error. However, since observational errors are generally originally computed as root-mean-square error values from many cases they may be said to represent the average error of the particular instrument and thus, it is reasonable to view the absolute values of the analysis error as representing an average value of the error of the analysis. We are using estimated analysis errors from one day to determine error estimates of energy calculations. The preceeding details of how these estimates are derived implies that the estimated errors from this one case are representative of an average case as well.

For the day examined within this study, December 20, 1981, the estimated analysis errors of temperature and wind speed are depicted in Figure 1 for the

500 mb level, both Northern and Southern Hemispheres. As described above, the values are very much influenced by the data availability. In particular, gaps in coverage exist in low latitudes in the Eastern Atlantic and Northern Africa as well as over the Northeast Pacific Ocean. Such coverage is very typical of this period. We note that satellite data are not utilized over land areas at all, even if there is little or no radiosonde coverage, resulting in relatively large errors over North Africa, South America and Antarctica. At higher altitudes some of the gaps are filled in with cloud-tracked winds or aircraft reports.

At 500 mb, the temperature analysis error estimates are generally about 1-2°C increasing to about 4 degrees in the gap areas and about 5-8°C over the Antarctic Continent. We should note that the estimated analysis error is not permitted to exceed a set climatologically-derived value.

For the wind speed, the estimated analysis errors are on the order of 3 ms^{-1} increasing to about 10 ms^{-1} in the data-sparse regions and about 13 ms^{-1} over Antarctica.

Within Table 1 we present the above described estimated error variances by wave number group as a percentage of the variance within the actual analysis. For planetary to mid-scale waves in the Northern Hemisphere, the error variance is less than 3%. A much more substantial percentage error is present at the smaller wavelengths. In the case of the Southern Hemisphere, which is in the summer season, we see that the percentage error variances are generally much greater. The maximum value of about 28% appears in the smaller wavelengths for temperature.

b) Monte Carlo Procedure

As is evident from Figure 1, the estimated analysis errors have a large scale pattern associated with the availability of data and the type of data. Hence,

within a Monte Carlo type of simulation this must be considered since it means that each 2.5^{o} x 2.5^{o} latitude/longitude grid point is not statistically independent of its neighbor. If we consider that the above is true in both horizontal dimensions as well as in the vertical, our modifications to the Monte Carlo technique will reflect a compromise on achieving an element of reality versus limiting the computer time. This modification to the procedure outlined in the introduction involves analyzing the error fields in longitudinal wave number space up to wave number 24 and, retaining the phase, using the harmonic amplitudes as the value equal to one standard deviation in a Gaussian random error generator. This is done at each wave number up to 24 and at each pressure level independently. Then the error field is transformed from wave number space into a 2.5^{o} gridded field to be added to the original analysis. In this way, an alternate representation of the analysis is created which attempts to account for the fact that the analysis errors are spatially correlated with one another. Using the OI analysis of December 20, 1981 as the base field, 5 independent alternate representations of that analysis were created. Presented below are the results for these 5 runs as they compare against the original analysis as well as against each other.

We recognize at the outset, however, that utilizing this wave-number technique is not without difficulties in the sense that for any particular run the sum of error variances over the wave numbers need not equal the original variance. However, in our view, a major element of this study is the evaluation of the cross-product terms which requires a reasonable consideration of the longitudinal correlation in the error structure.

Before we examine the results, it is necessary to examine the statistical impact of the Monte Carlo procedure on our interpretation. The true value of

an atmospheric variable such as temperature (T^t) may be represented at a point point as:

$$T^t = T \pm \mathcal{E}$$

where T is the analysis value and

$\mathcal{E}$ is the error in the analysis value

Within the optimum interpolation analysis method $\mathcal{E}$ is estimated from the type and amount of available data. When this error is added to the analysis (via the Monte Carlo Technique), the mean and variance around a latitude circle become:

$$[T^t] = [T] + [\mathcal{E}]$$

$$\text{Var}\ (T^t) = \text{Var}\ (T) + \text{Var}\ (\mathcal{E}) + 2\ \text{Cov}\ (T, \mathcal{E})$$

We see, then, that the effect of the zonal average analysis error $[\mathcal{E}]$ on $[T^t]$ can be plus or minus, independent of the T values. For the variances, however, the effect is to add noise to the system, the actual amount dependent on the correlation of the error estimates and the analyzed field. In this study, within each run when the wave number pattern is changed, both the variance and covariance terms are altered as well. When the 5 run average is examined, this average variance will, in general, be greater than the variance of the original field, (i.e. the covariance term will not cancel the Var ($\mathcal{E}$) term) thus providing information on the smoothing included within the analysis system. By examining the variations among the 5 runs, we gain insight into the "noise" level of the variance and covariance calculations.

One element that must be considered within this simulation, however, is the inherent correlation of the observation errors versus the random variation assigned. We have seen, for example, from Figure 1 that the error patterns contain latitudinal as well as longitudinal correlation and we would expect (Bergman, 1979) that a vertical correlation of errors would exist as well. As

described above, we have included only the longitudinal correlation, mainly because it was not clear how to include the other elements in a realistic manner. On one hand, the result is that by allowing the Monte Carlo variations to be independent of latitude and pressure we increase the randomness of the calculations such that when the hemispheric integral values are calculated, the tendency, in general, is to average out the variations. For the zonal available potential energy the above latitudinal variation would increase the value while it would be smoothed by the vertical integration.

On the other hand, for the case of the variations among the 5 runs, the lack of latitudinal and pressure correlation, in effect, provides an increase in the relative number of independent data points. Examination of the results revealed great consistency in latitude and pressure which added confidence to the results. This was the factor that lead to our decision to restrict our study to 5 runs.

3. Results

In Figure 2, we show latitude-pressure cross sections of the analyzed zonal averages of the parameters U, V and T along with the zonal average errors for the same components. We see that the average errors tend to have their largest values in the tropical regions and over Antarctica, again where the data are sparsest. For the zonal wind fields, the errors tend to be greatest where the analyzed values are lowest, thus leading to substantial uncertainty in the poleward gradients. The large errors in the temperature field over Antarctica, about $7^{o}C$, reflect the lack of data in this region.

Within Figure 3, we present information on the variance components. Figure 3a is composed of cross sections of the eddy kinetic energy ($[U^{*2} + V^{*2}]$) and the temperature variance. Figure 3b contains cross-sections of the U, V and T error variances. As might be anticipated from above, the error variances

are greater in the tropics and Antarctica where the analyzed variances tend to be small, thus leading to larger percentage errors. On the other hand, in mid-latitudes of both hemispheres the error variances are considerably smaller than analysis values giving substantial confidence in these areas.

For the overall discussion of results, we will proceed from the general to the specific. In Table 2 we present the hemispheric average values, equator-pole and 1000-100 mb, for the zonal and eddy available potential energy (Lorenz, 1967; Miller et al, 1975) AZ, AE; zonal and eddy kinetic energy, KZ, KE; and the energy transfer terms AZ to AE and KE to KZ; CA and CZ, respectively.

Looking first at the results for the Northern Hemisphere, we see that, on average, the AZ and KZ values agree with the original analysis to better than 1% and even the variation within runs satisfies this standard. As we will see below, this is a result of the gross averaging over the entire hemispheric domain and can be quite different over limited areas. The terms involving direct variance, AE and KE, show an average 2-4% increase over the analysis with, again, a relatively small variation among the runs. Finally, the terms involving cross-covariance elements, CA and CK, are +0.5% and -1.99% respectively, but with somewhat more variation among the runs.

For the Southern Hemisphere, a similar overall pattern of results emerges, but with more variance indicated due to the large error values over Antarctica. We see, then, that the integration over such large domains tends to average out the error variations with the result that the overall errors, both on average and between runs, are within about 5%. The one exception is the Southern Hemisphere average AE at 9.8%.

In Figures 4-7 we present pressure latitude cross sections of the eddy kinetic energy, KE; temperature variance $[T^{*2}]$; eddy momentum transport, $[U^*V^*]$; and eddy heat transport, $[V^*T^*]$, respectively. For each parameter the

original analysis values are on top, the average of the 5 runs in the middle and the standard deviation among runs on the bottom. We note that these parameters were selected for presentation as each involves some element of the longitudinal variance which we have seen from Table 2 is particularly sensitive to the error components.

If we examine these four diagrams collectively for the analysis versus average run value, it is very striking that the overall patterns are very well maintained in both hemispheres. The increase in variance of the average run value is distributed over the entire domain with little overemphasis in any one region save for the expanded temperature variance between 20°S and 30°S.

Looking at the standard deviation among runs, however, we see some very interesting patterns for each parameter. For the eddy kinetic energy, Figure 4, the maximum standard deviation occurs at low latitudes, extending up to mid-latitudes above about 300 mb with another maximum over Antarctica. As a percentage of the actual analyzed values, then, the standard deviation ranges from about 1 to 2% in mid-latitudes to about 10% in the tropics and polar areas. On average, the standard deviation is about 6%. The standard deviation of the temperature variance (Figure 5), on the other hand, shows maxima between 20-30° north and south as well as the anticipated one over Antarctica. As a percentage of the analyzed values, the standard deviation for this parameter ranges from about 5% in mid-to-high latitudes of the Northern Hemisphere to about 50% in the low-latitude regions. On average the standard deviation is about 17%.

For the momentum and heat transport terms, Figures 6 and 7, the situation is confused by the fact that these terms change sign, which leads to very large percentage standard deviations in the areas of near zero transports. Bearing this in mind, we see in Figure 6 that the standard deviations of the momentum transports are about 5-10% in mid-latitudes where the transports are greatest.

Outside of this region the estimate can increase to over 100%. In the case of the sensible heat transport, the standard deviations are about 4% in mid-latitudes of the Northern Hemisphere. Because the analysis values outside this region tend to be small, the standard deviation, as a percentage of the analyzed values, are highly variable, but can be of the order 50-100%.

The point to be gained from the above is that in regions where the parameter estimates are large and well defined, the standard deviation seems to be less than 10%. Outside of these regions, however, where the values are small, the precision estimates become considerably worse and can be as large as 50-100%.

4. Final Remarks

Insofar as the error estimates provided by the optimum interpolation analysis method are representative of the actual errors, we have shown that the estimates of integrated hemispheric energy parameters are, generally, on the order of 1 percent with the analysis "smoothing" the "noise" by about 4% as calculated from the AE and KE terms. For specific parameters, however, the calculated precision estimates can be both pressure and latitude dependent ranging from a few percent to over 100% depending on the value of the parameter itself.

From a practical point of view, this provides a basis for comparing analyses or, as argued in the introduction, the impact of a change in analysis procedure. For two parallel analyses on the same data base, the 95% confidence limit on the difference between parameters can be derived given the formula:

$X = 2\sigma/\sqrt{n}$ where σ is the paired noise estimate and n the number of paired analyses. In practice, if two analyses include the same data, the resultant errors will be correlated so that we do not have truly independent comparisons. For an initial estimate, however, it seems reasonable to ascribe the same error estimate to each of the paired analyses, in that it will be a conservative esti-

mator, and ask how many parallel runs it would require to obtain a prescribed level of precision, X.

For example, if we wish to compare temperature variance in mid-to-high latitudes, a reasonable estimate of the paired noise, from above, would be $[(5\%)^2 + (5\%)^2]^{1/2} = 7.07\%$. and to obtain a 5% precision on the difference would require $((2)(7.07)/5)^2$ independent matchups. Thus, for this particular parameter, 8 parallel runs would have to be compared. One must be careful, however, in that the result depends on the σ value, which we have seen is dependent on both pressure and latitude.

Acknowledgements

This work was supported, in part, by the National Aeronautics and Space Administration. We are also grateful to Dr. James Hoke for his constructive advice on the harmonic analysis approach.

References

Bergman, K., 1979: Multivariate analysis of temperatures and winds using optimum interpolation. Mon. Wea. Rev., 107, 1432-1444.

Hauser, R. K. and A. J. Miller, 1978: A study of the atmospheric energetics of a six-layer operational forecast model, Mon. Wea. Rev., 106, 607-613.

Lorenz, E. N., 1967: The nature and theory of the general circulation of the atmosphere. WMO, No. 218.TP.115, 161pp.

McPherson, R. D., K. H. Bergman, R. E. Kistler, G. E. Rasch, and D. S. Gordon, 1979: The NMC operational global data assimilation system. Mon. Wea. Rev., 107, 1445-1461.

Miller, A. J., W. Collins and D. Dubofsky, 1975: The NMC operational global energy program. Office Note 109, National Meteorological Center, Washington, D.C., 13 pp.

Miller, A. J. and C. M. Hayden, 1978: The impact of satellite derived temperature profiles on the energetics of NMC analysis and forecasts during the August 1975 Data Systems Test. Mon. Wea. Rev., 106, 607-613.

Table 1. Error Variance by Wave Number Group at 500 mb as Percentage of Analysis Variance - Dec. 20, 1981

N.H.

Wave No.	Var T	Var U	Var V
1-4	2-8%	2.1%	2.8%
5-10	2.4%	1.4%	0.7%
11-15	12.6%	8.0%	3.8%

S.H.

Wave No.	Var T	Var U	Var V
1-4	8.0%	3.7%	15.1%
5-10	11.24%	2.4%	1.0%
11-15	27.8%	12.6%	5.3%

TABLE 2 Hemispheric integral values of zonal available potential energy (A-Z), eddy available potential energy (AE), zonal kinetic energy (KZ), eddy kinetic energy (KA), energy transfer AZ to AE (CA) and energy transfer AE to KZ (CK) for Dec. 20, 1981 analysis and 5 Monte Carlo runs.

December 20, 1981 N.H.

	AZ		AE		KZ		KE		CA		CK	
	$\times 10^4$		$\times 10^3$		$\times 10^3$		$\times 10^3$					
RUN 1	487.53		811.72		936.84		991.71		3.93		.673	
RUN 2	487.49		827.98		937.57		989.85		3.95		.652	
RUN 3	487.46		814.63		936.34		986.76		3.93		.649	
RUN 4	487.76		810.55		937.43		986.10		4.02		.659	
RUN 5	487.55		810.20		938.05		987.39		3.89		.670	
AV	487.56	+.002%	815.02	+3.86%	937.25	-.03%	988.36	+2.29%	3.94	+.5%	.6606	-1.99%
SD	0.118	.02%	7.45	0.9%	0.665	.07%	2.35	.24%	0.0477	1.2%	.0106	1.6%
Analysis	487.55		784.73		937.54		964.24		3.92		.674	

S.H.

	AZ		AE		KZ		KE		CA		CK	
RUN 1	331.79		365.68		561.59		624.93		.630		.918	
RUN 2	330.83		357.23		560.64		628.28		.632		.942	
RUN 3	330.62		367.89		560.26		623.74		.663		.907	
RUN 4	331.16		372.85		561.08		626.19		.588		.920	
RUN 5	331.23		367.66		560.83		627.41		.665		.933	
AV	331.13	+.01%	366.26	+9.78%	560.88	-.007%	626.11	+4.21%	.636	+2.75%	.924	-.32%
SD	0.446	.13%	5.70	1.6%	0.497	.09%	1.83	.29%	0.031	4.9%	0.0137	1.5%
Analysis	331.09		333.63		560.92		600.84		.619		.927	

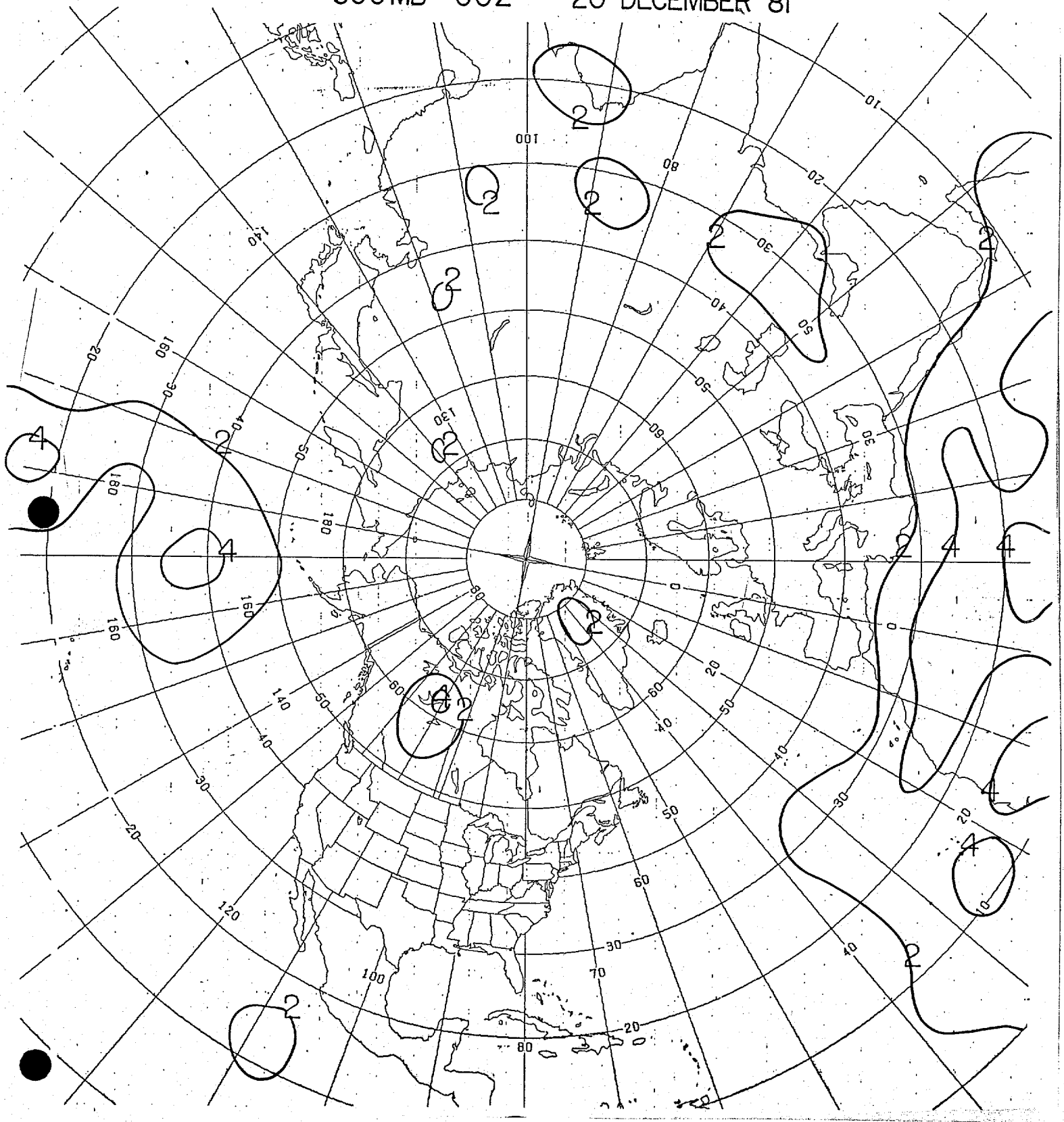

Figure 1a) Estimated Northern Hemisphere temperature analysis errors (°c) at 500 mb for December 20, 1981.

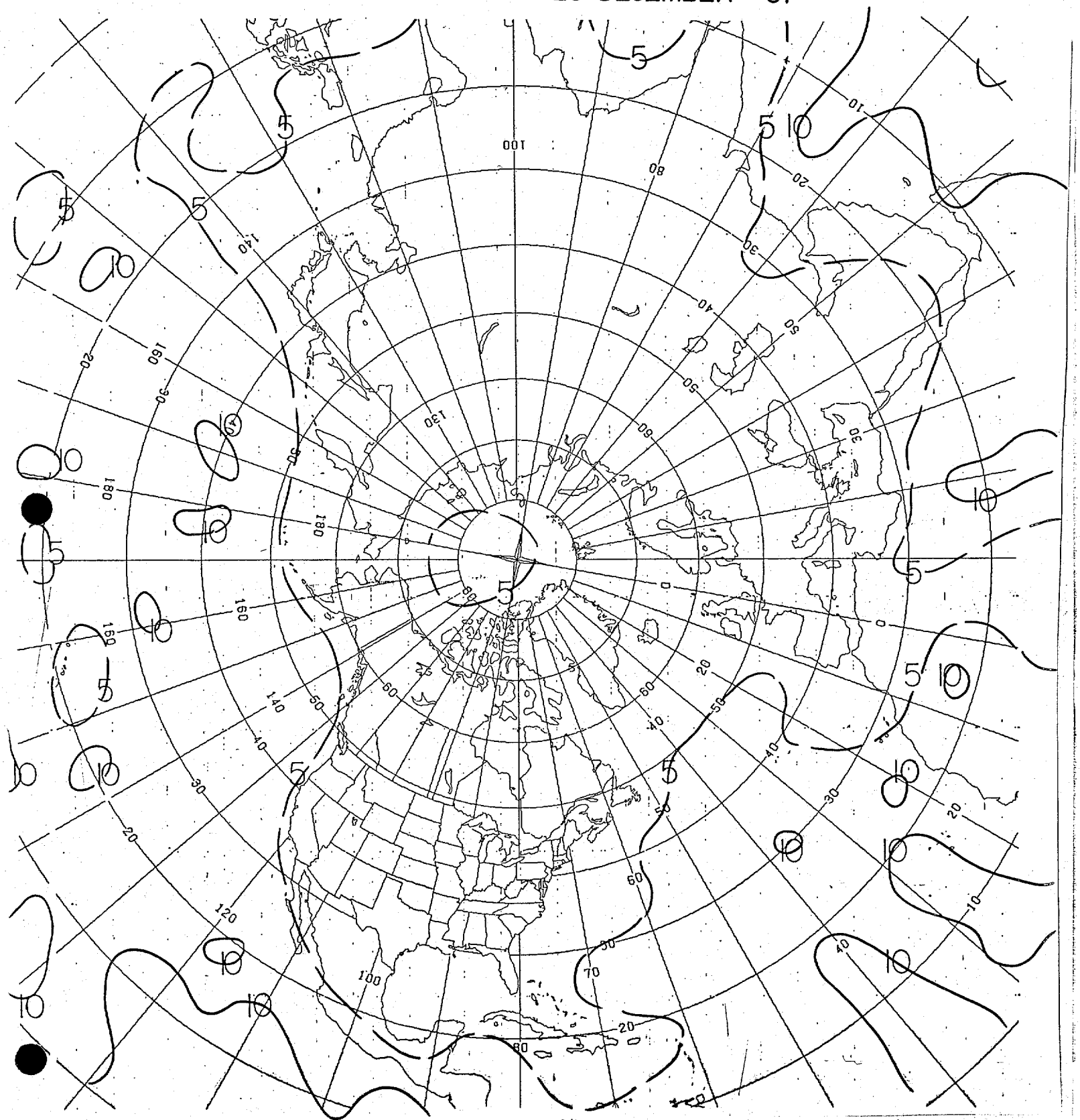

Figure 1b) Estimated Northern Hemisphere wind speed analysis errors (ms^{-1}) at 500 mb for December 20, 1981.

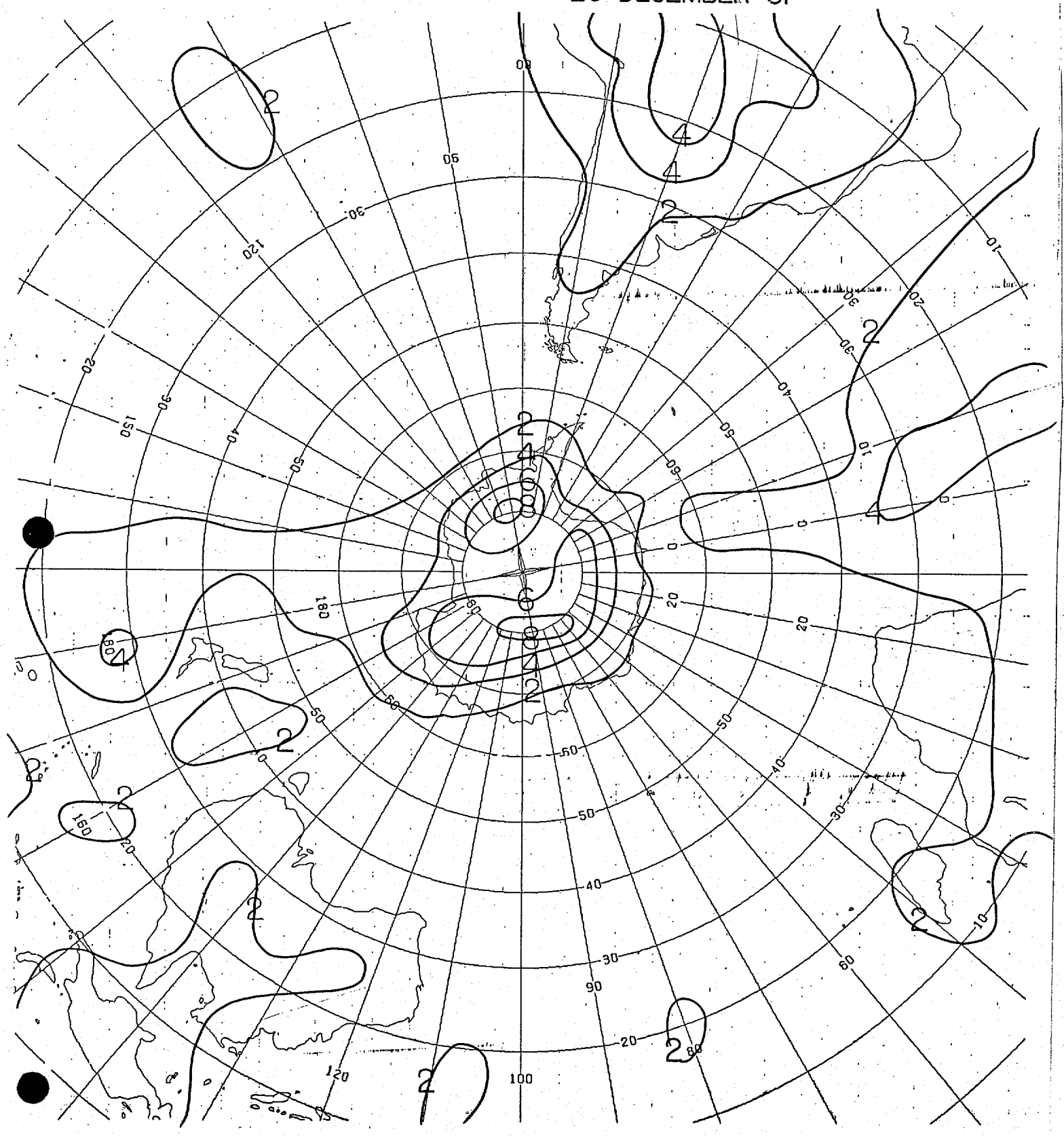

Figure 1c) Same as Figure 1a) for Southern Hemisphere.

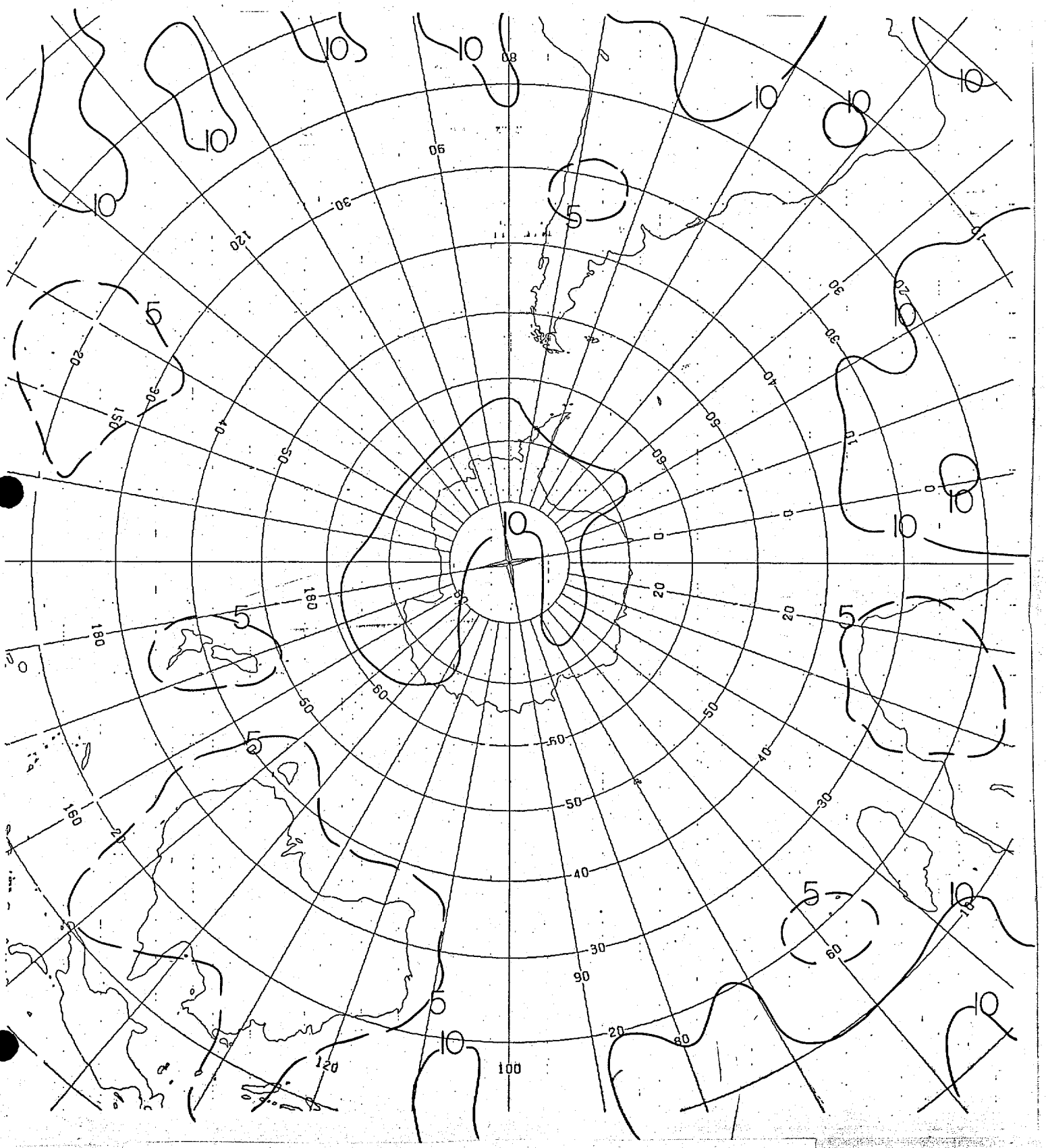

Figure 1d) Same as Figure 1b) for Southern Hemisphere.

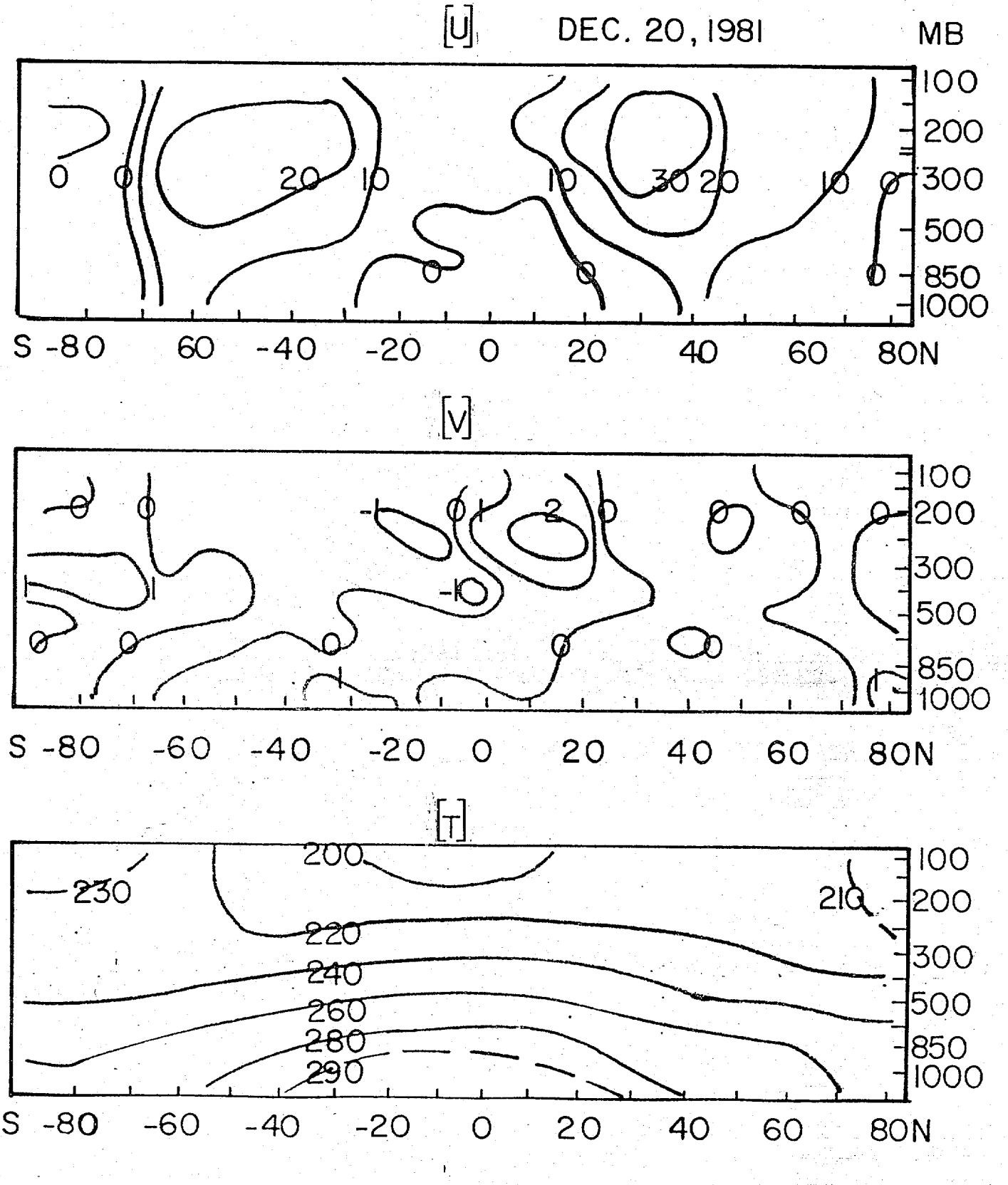

Figure 2a) Latitude-pressure cross sections of analysis of zonal averages of zonal wind, (ms^{-1}) meridional wind (ms^{-1}) and temperature (^{o}K).

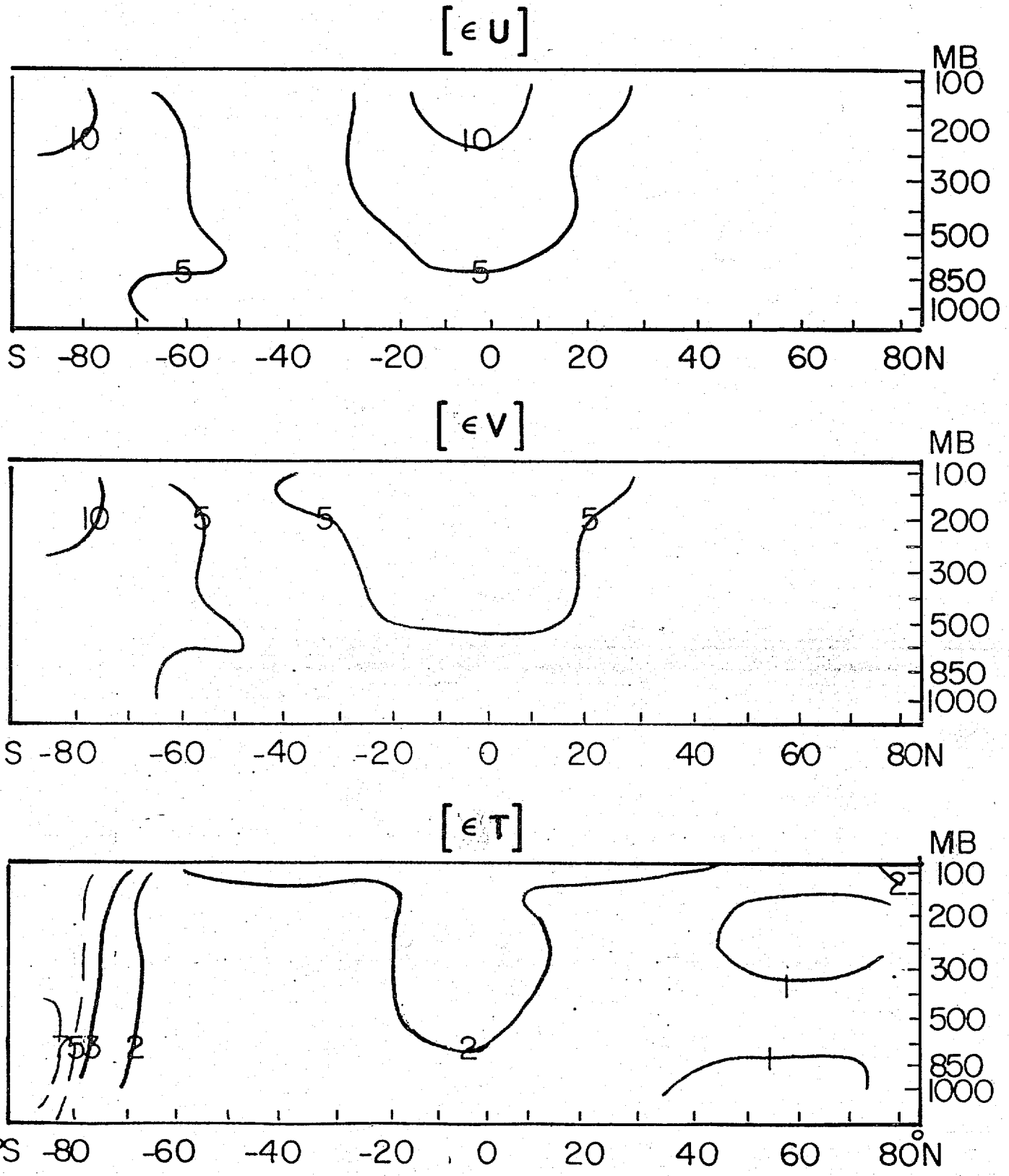

Figure 2b) Same as Figure 2a) for zonal average estimated errors.

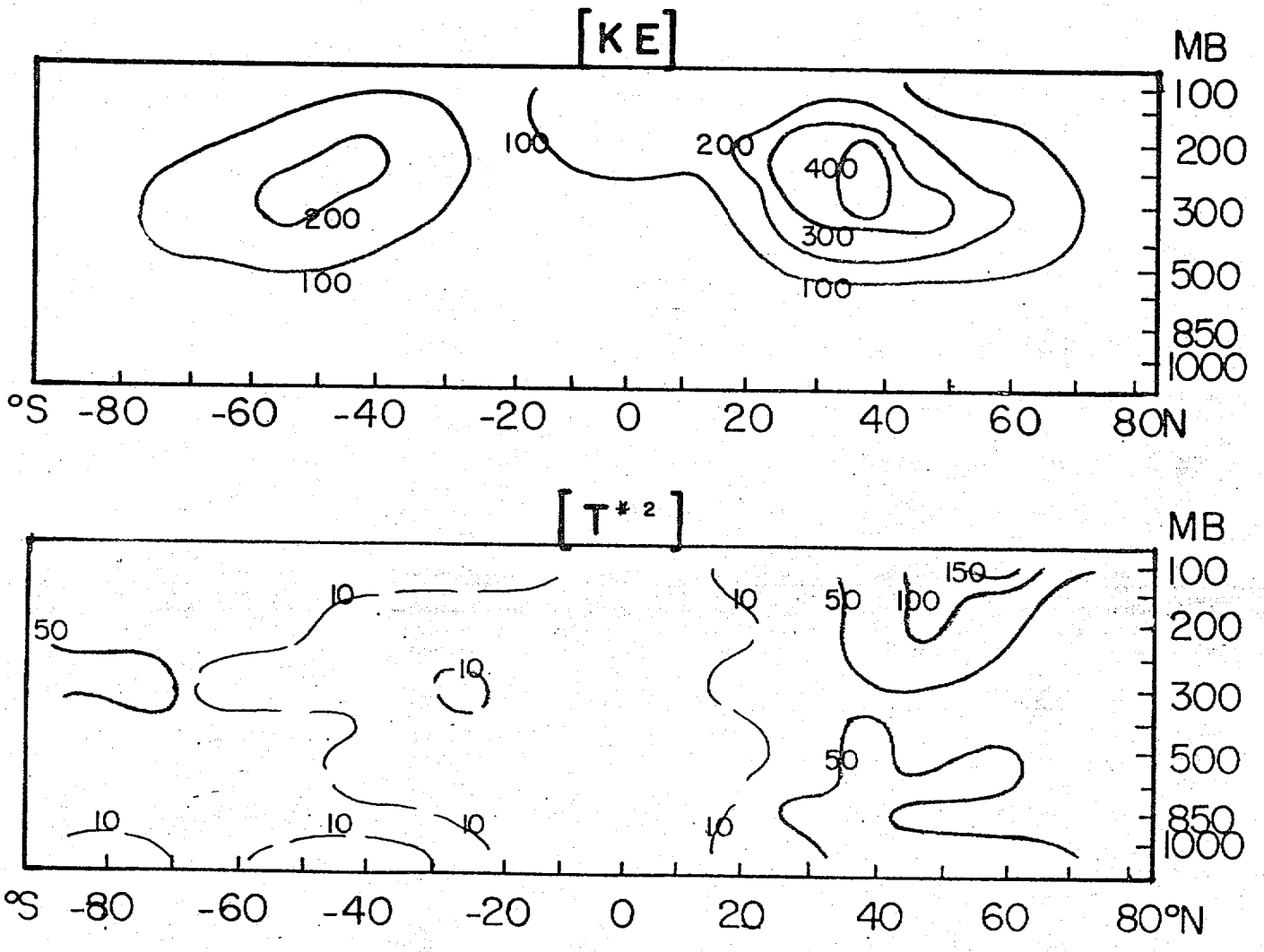

Figure 3a) Latitude-pressure cross sections of analysis eddy kinetic energy (J kg^{-1}) and temperature variance ($^{o}K^{2}$).

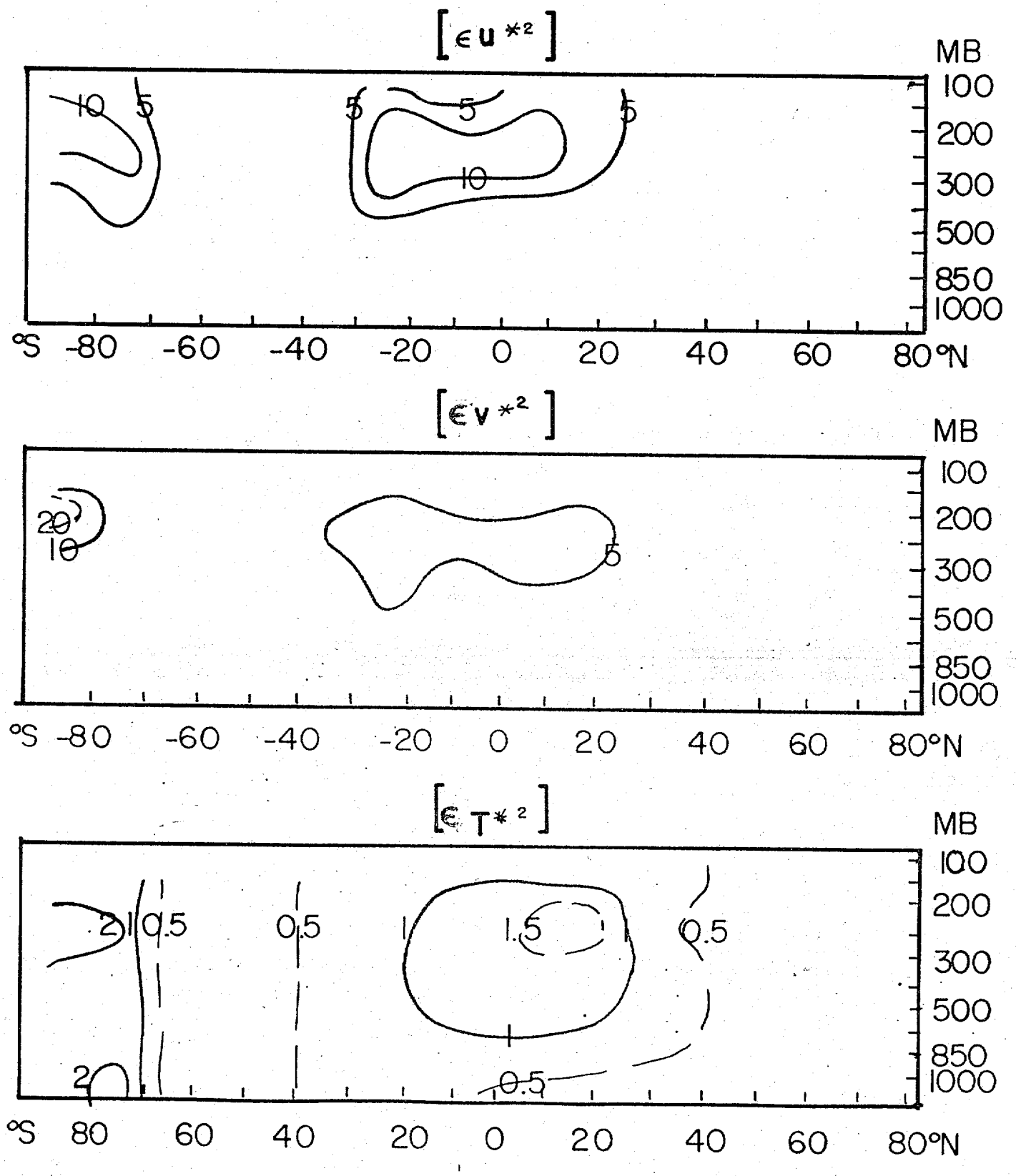

Figure 3b) Same as Figure 3a for error variances of zonal wind (m^2s^2), meridional wind (m^2s^2) and temperature ($^oK^2$).

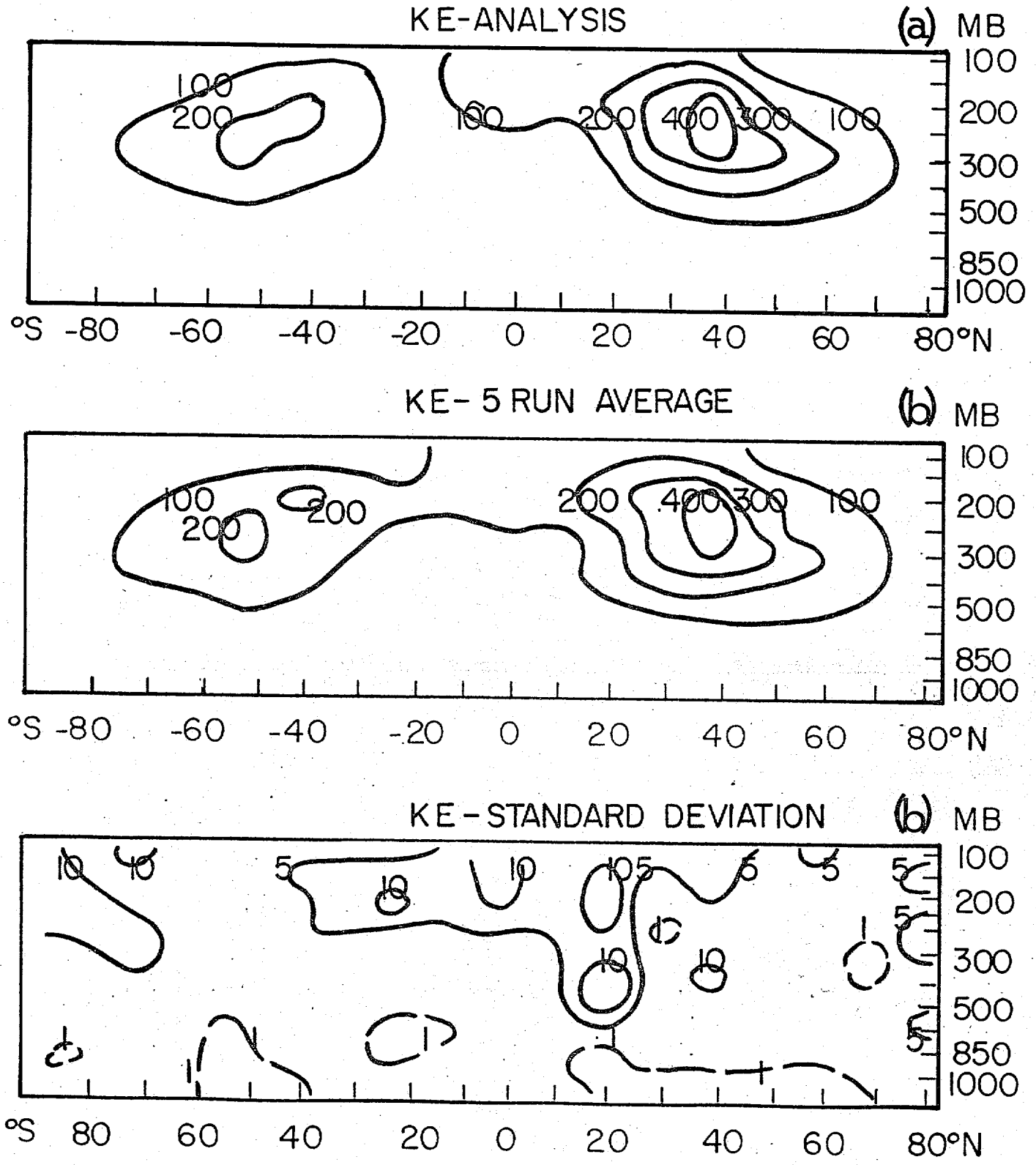

Figure 4) Latitude-presure cross sections of eddy kinetic energy (KE). analyzed value is presented on top, the average of the 5 Monte-Carlo runs in the middle and the standard deviation among runs on the bottom.

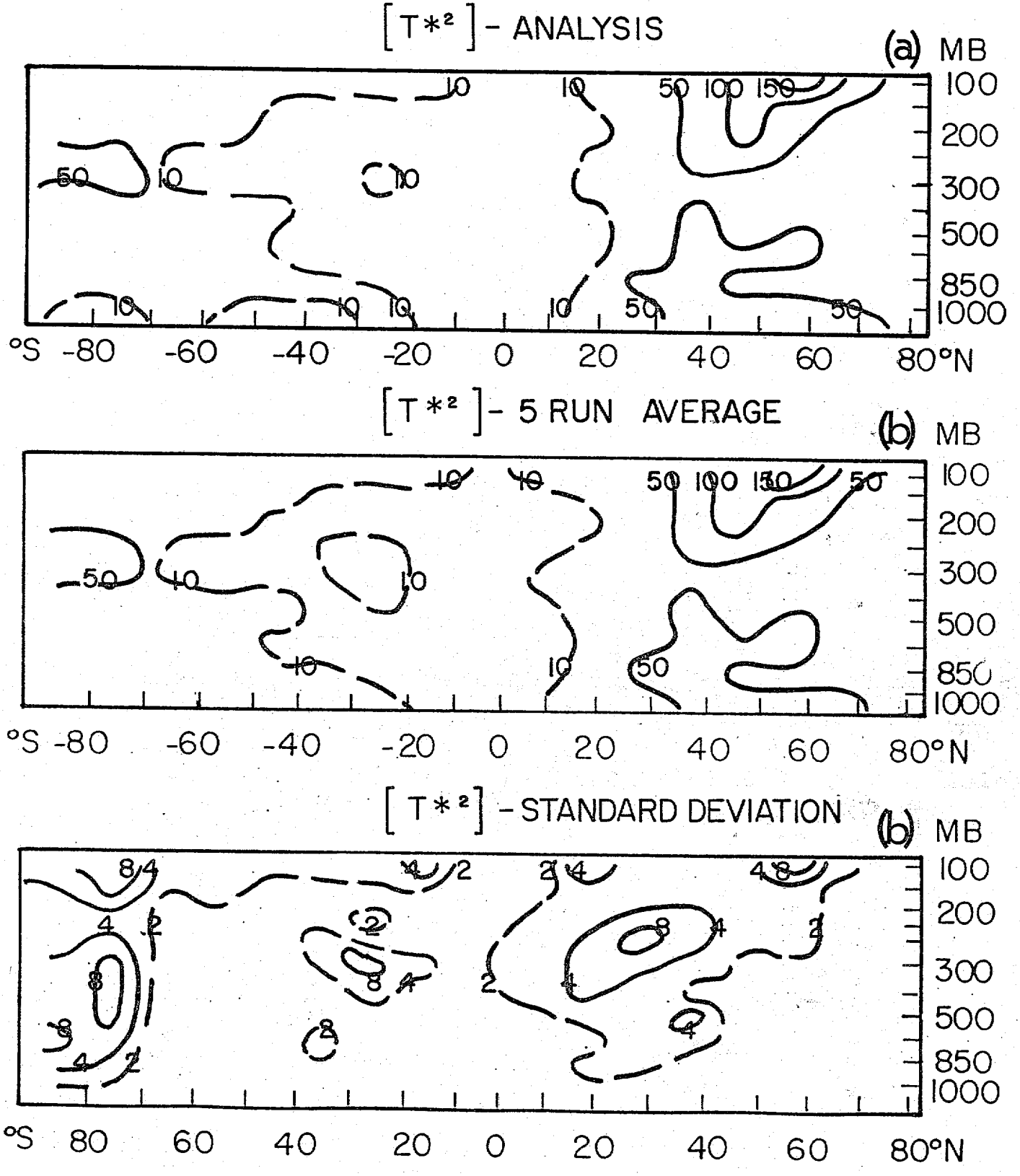

Figure 5) Same as Figure 4 for temperature variance.

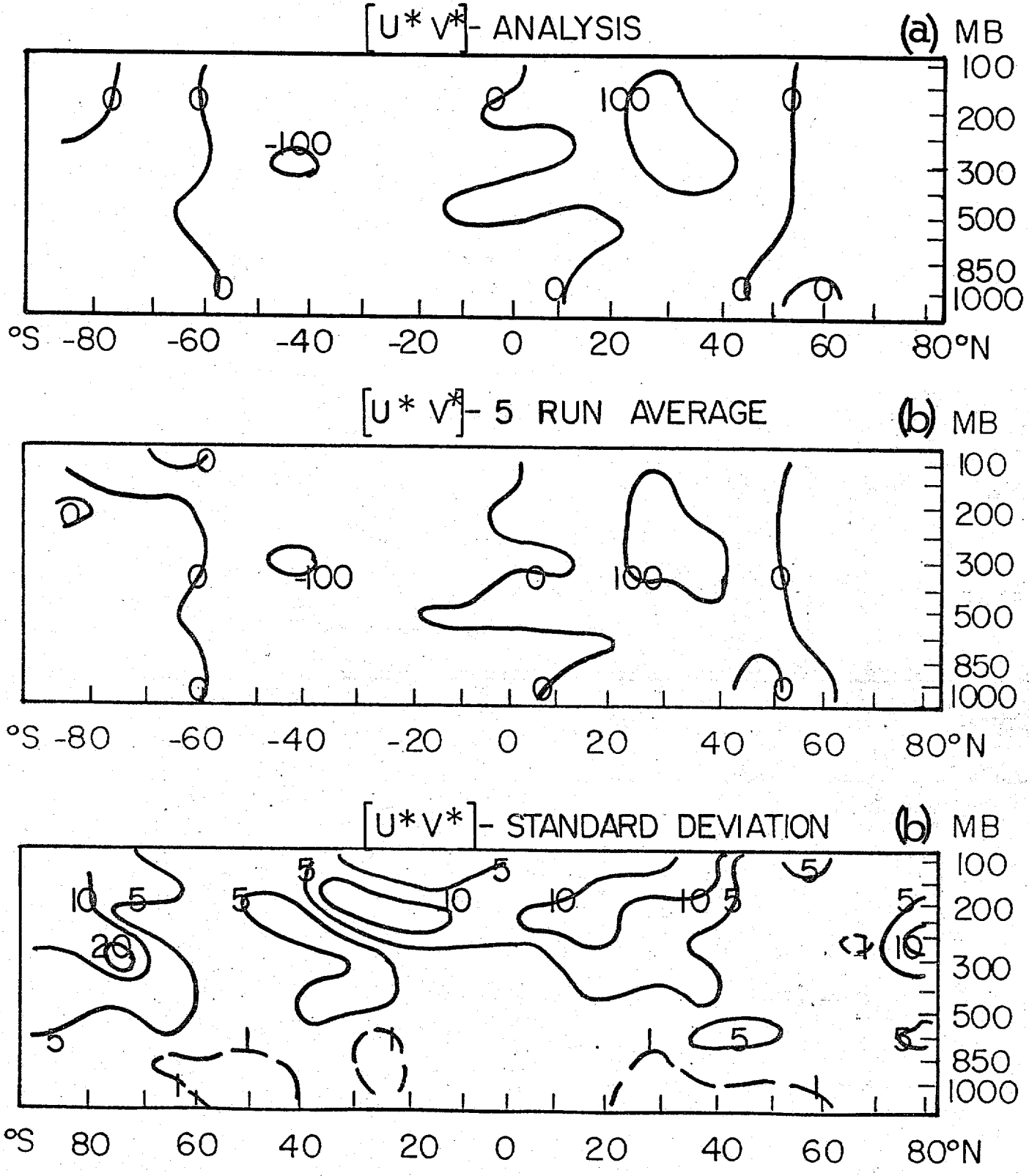

Figure 6) Same as Figure 4 for eddy momentum transport.

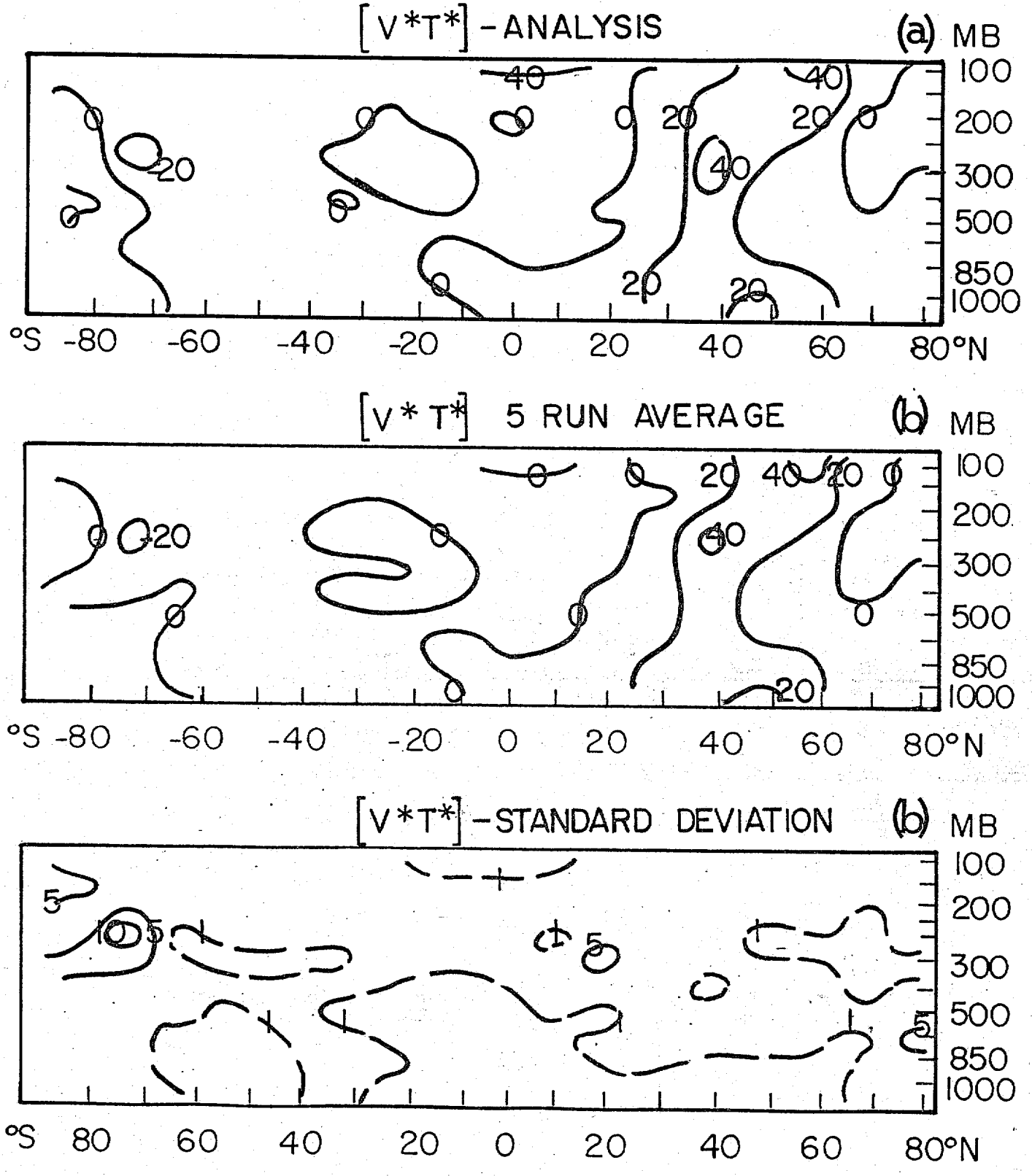

Figure 7) SAme as Figure 4 for eddy heat tranport.