The Near-Real-Time, Global, Four-Dimensional Analysis Experiment During the GATE Period, Part I

K. MIYAKODA, L. UMSCHEID, D. H. LEE, J. SIRUTIS, R. LUSEN AND F. PRATTE

Geophysical Fluid Dynamics Laboratory/NOAA, Princeton University, Princeton, N. J. 08540

(Manuscript received 7 August 1975, in revised form 12 December 1975)

ABSTRACT

Global upper air and surface data for the entire GATE period from 15 June to 24 September 1974, were collected by the Data Assimilation Branch of NMC and mailed to GFDL. After processing these data, a four-dimensional analysis technique was applied for the entire GATE period, using a global numerical model. For a selected period, several different versions of the data processing scheme were tested. The resulting analyses were compared with each other and with the objective analysis of NMC in Washington D. C., and ANMRC in Melbourne. Overall, the analyses for the extratropics were satisfactory for the Northern Hemisphere, and to a lesser extent, for the Southern Hemisphere, though flow patterns are somewhat excessively smoothed. The analyses for the tropics were not of the same quality as those for the extratropics, and yet they were much improved compared with those of several years ago. A noteworthy point is that tropical cyclones were successfully represented in several cases.

1. Introduction

As one of the GATE research activities, it was planned to conduct a near-real-time sequential dynamical analysis of meteorological fields on a global extent as well as for the limited GATE area [see the report of ISMG-International Scientific Management Group of GATE (1972) and Houghton and Parker (1974)]. The application of the so-called four-dimensional analysis to the GATE data set is not only pertinent but also very significant in view of the unique opportunity to utilize enhanced tropical observations. During Phases 1, 2 and 3 of the GATE period (each of 21-day duration), more than 30 ships were mobilized from various nations and deployed over the tropical Atlantic Ocean. In addition, the SMS-A satellite was launched to support the GATE field missions, and provided cloud photographs and, thereby, cloud-tracked winds. Such a dense observing network was unprecedented in history, and should not be expected to occur again in the next ten years. Although these intense observations were confined to a small portion of the Atlantic Ocean, this paper concerns the global analysis. It is our intention to proceed to the limited-domain tropical analysis in the near future; this will, however, bear a close connection with the present study.

For the analysis of meteorological fields, the tropics present unique and formidable problems. It would not be cynical to say that we have not yet reached a stage where we are able to make satisfactory tropical weather maps that are reasonably accurate on the synoptic scale. Other than the paucity of observations, there are several factors which make it difficult to do so. For one thing, the tropical flow is relatively independent of the pressure field due to the weakness of the Coriolis force; in addition, complex physical processes, such as the liberation of condensation heat, have an appreciable influence on the flow field. Probably due to the combination of these effects, the relatively small-scale disturbances are more important constituents in the tropical flow patterns than in those of mid and high latitudes, and the disturbances seem to be less strongly coupled in the vertical as well as in the horizontal. It has long been a question, therefore, as to what degree of spatial resolution is needed for an adequate observation network in the tropics, and what accuracies for observations of wind and temperature are required to produce a usable initial condition for tropical forecasts (see Report of the Fifth Session of JOC of GARP, February, 1971). Answers to these questions are fundamental to further exploration of the behavior of the tropical atmosphere and, therefore, to unraveling the intricate mechanisms of tropical disturbances.

As a technique to make use of all the available data scattered unevenly not only in three-dimensional space but also in time, a four-dimensional assimilation concept and scheme were proposed [GARP Study Conference at Stockholm, 1967; Charney *et al.*, 1969; see also the reviews of Kasahara (1972) and Bengtsson (1975)]. The principle is to use the numerical atmospheric model for the data analysis and to obtain a data set internally consistent among various parameters in the dynamical context. So far the method has been mostly applied to hypothetical data generated by the same model[1] ("identical twin transplant"), and these tests demonstrated that the approach is quite promising.

[1] Except, for example, Halem *et al.* (1974, personal communication).

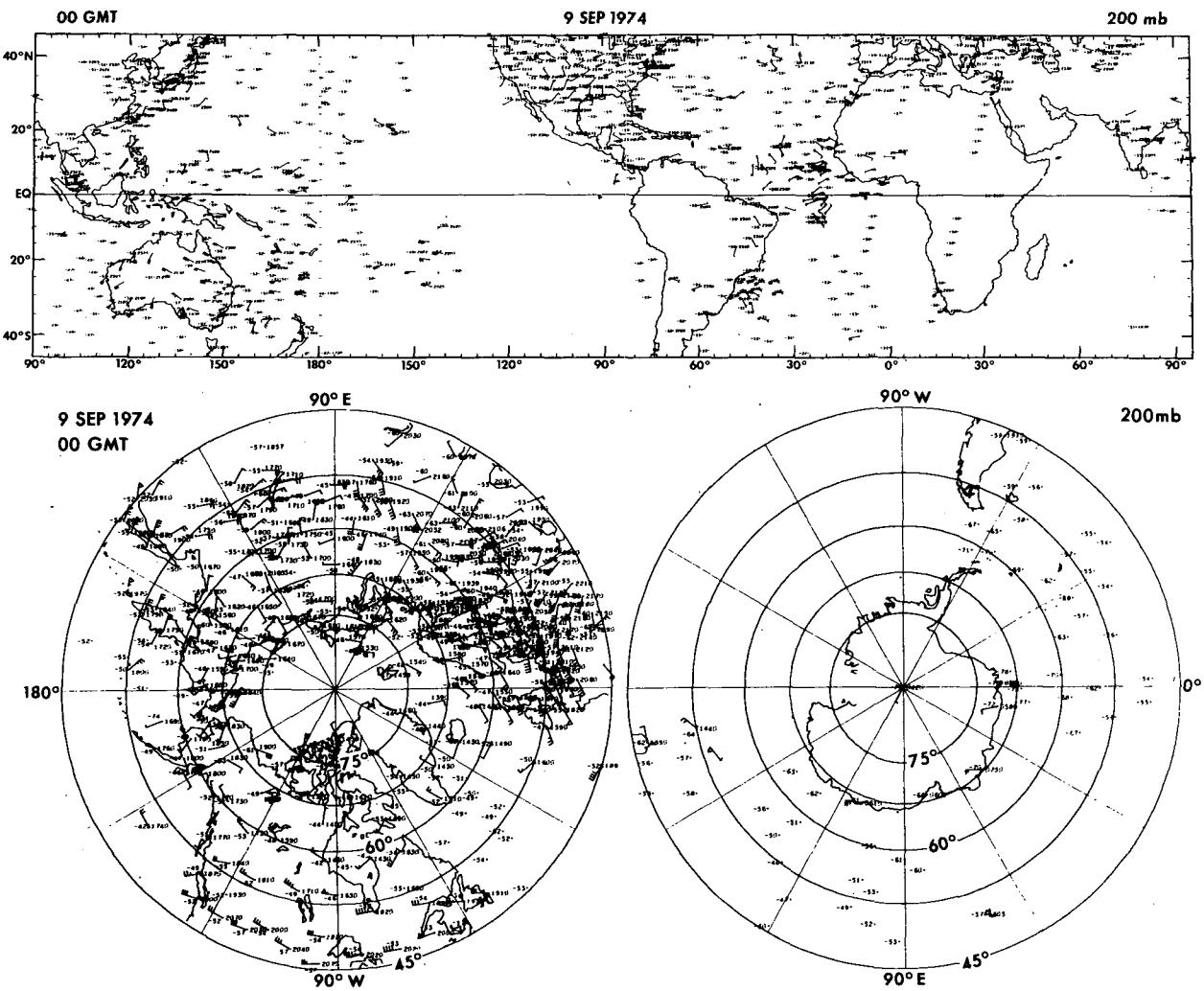

FIG. 1a (see legend p. 564).

It was anticipated, however, that the application of the method to real data might pose entirely different problems (Morel *et al.*, 1971). In fact, prior to GATE, we expended a great deal of effort to apply the four-dimensional analysis to the BDS [Basic Data Set of 1969, proposed by JOC of GARP; see Thompson (1972)], and realized that the treatment of real data indeed requires substantial alterations to the conventional method. A simple injection of raw data resulted in its outright rejection by the model, leaving violent "shocks" in its solution. This undesirable effect occurs, first of all, because the observed data are not dynamically compatible with the particular numerical model employed, and second, because the real data often contain errors of various sources. It is not a trivial matter to handle such raw data and assimilate them smoothly in the analysis.

Based on the experience gained so far, we will present here practicable and workable methods of four-dimensional analysis. We will then discuss the analysis resulting from the implementation of the methods to the GATE data set, and investigate the feasibility of the approach not only for the tropics but also for the extratropics.

2. Data

Global meteorological data for the GATE period of 101 days from 15 June to 24 September 1974, were collected by the Data Assimilation Branch of NMC, Maryland, with a cutoff time of 12 h after the observation time instead of the operational cutoff time of 10 h. Magnetic tapes containing all data were mailed from NMC to GFDL continuously during the GATE period [and to the Center for Experimental Design and Data Analysis (CEDDA) for archiving at the National Climate Center].

The data consist of the conventional upper air and surface data including VTPR clear air column vertical soundings of the polar oribiting satellite NOAA-2,

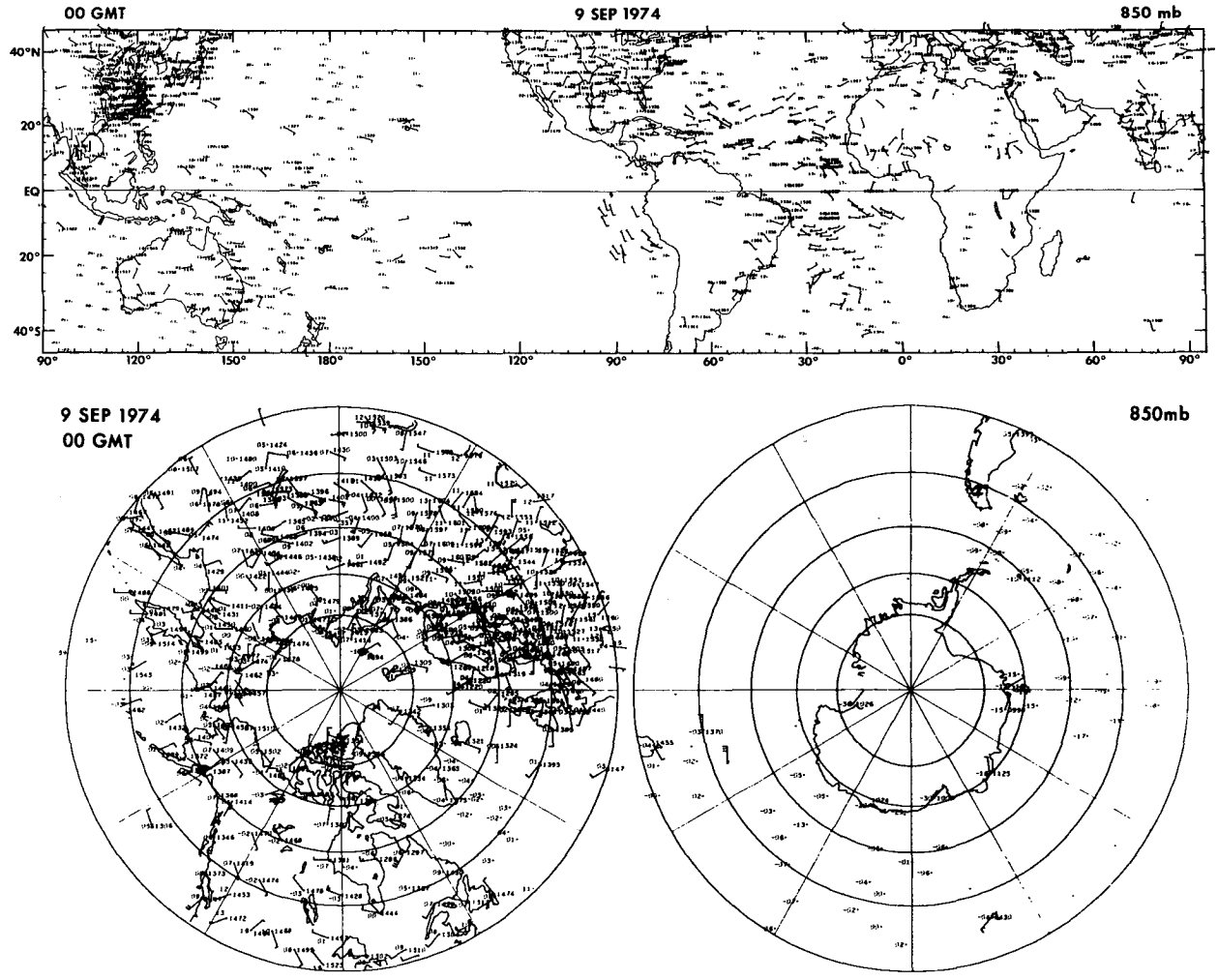

FIG. 1b (see legend p. 564).

the cloud-tracked wind data from the geosynchronous satellites ATS-3 and SMS-A, as well as the GATE B-scale and A-scale ship reports. The statistics on the data acquisition in this experiment for the upper air and the surface data are shown in "data inventories," Tables A1 and A2 of Appendix A.

Examples of data received at a particular time of day are shown on a Mercator projection map for the zonal belt between 45°N and 45°S, and on stereographic projection maps for the regions poleward from 45° latitude for the Northern and Southern Hemispheres (Figs. 1a–c). On the upper air data maps (Figs. 1a and 1b), winds, temperature, geopotential heights, and dew-point temperatures are plotted. Wind vectors are indicated by arrows with wind barbs; one full barb corresponds to 10 m s^{-1}. Letters at the origin of the arrows indicate the type of wind observations; rawinsonde or pibal data simply have a dot at the point; C is a cloud-tracked wind; A is a commercial or reconnaissance aircraft report; and S is a ship report (mostly GATE dedicated ships). Compared with the BDS in 1969, the area distribution of data in the Southern Hemisphere has not been appreciably improved (see Thompson, 1972). On the surface maps, small circles or dots mark the observation stations or platforms. It can be seen that there are still large data-void areas, e.g., the east tropical Pacific, the South Indian Ocean, and a large portion of the Southern Hemisphere.

Enlarged data maps for the GATE area also are shown later (Fig. 21). In Fig. 1, the GATE B-scale ships can be seen clustered west of the bulge of West Africa. All these data are the so-called "level II data" in GARP terminology. In addition to these data, NMC supplied us, on our request, the NMC "level III data," which are the global analyses of temperature and wind and other parameters. These NMC level III data were in general available twice a day at 0000 and 1200 GMT, and were used for comparison together with the level III data (map) of ANMRC (Australian Numerical Meteorology Research Centre) in Melbourne.

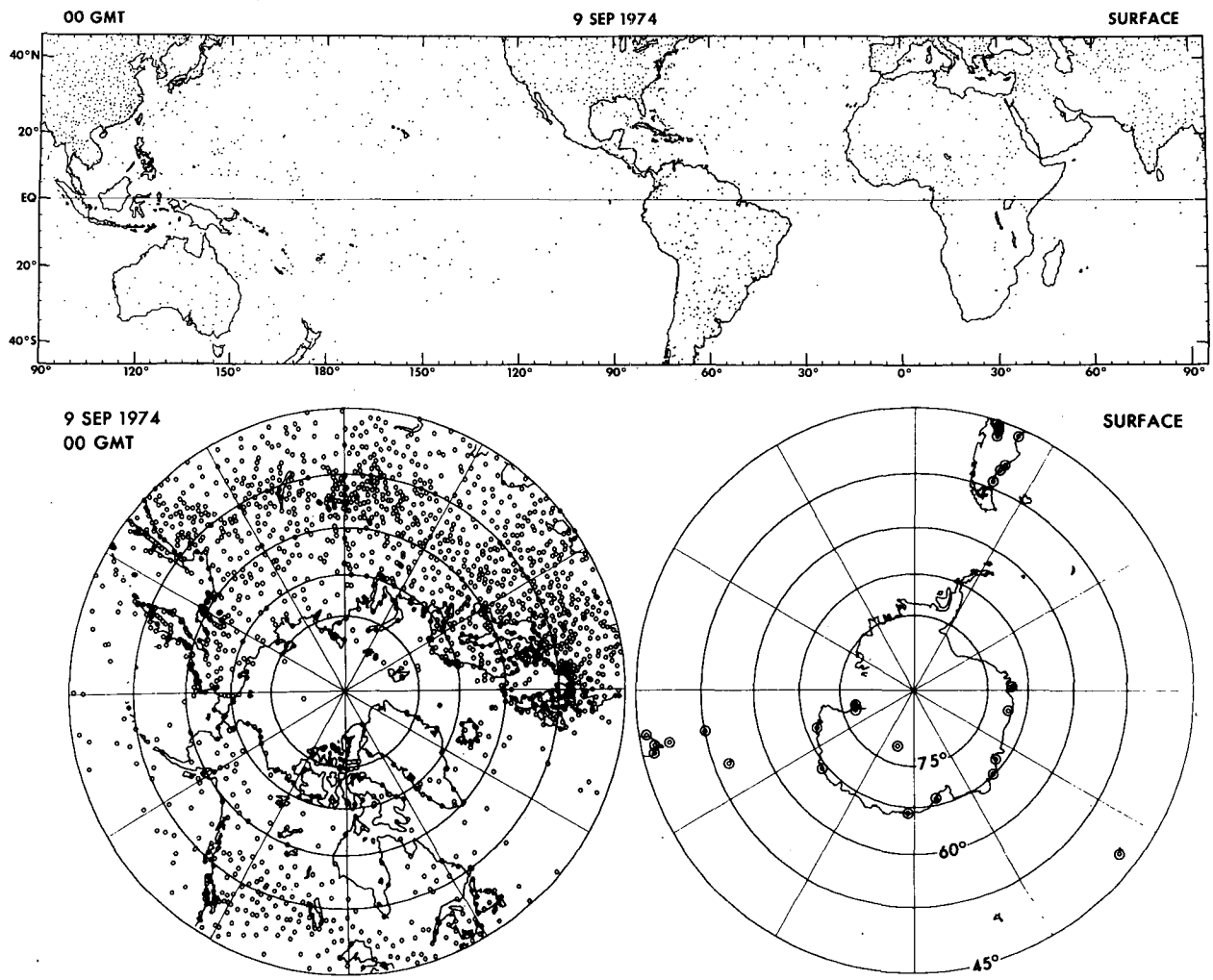

FIG. 1c.

FIG. 1. Examples of data maps for 00 GMT±3 h, 9 September 1974. (a) 200 mb map includes the data in the layer between 150 and 275 mb; (b) 850 mb map includes those between 775 and 900 mb; and (c) the surface map shows the data distribution with small circles or dots.

3. Four-dimensional analysis

The observed data were sequentially processed and prepared for insertion into a global numerical model. The model was run for 101 model days starting from 15 June 1974. While the marching calculation was proceeding, the observed meteorological data were injected into the model, and the information was assimilated into the model's solution. A schematic diagram of these processes will be shown later.

a. Global numerical model

The global model used for the data assimilation was the GFDL global, finite difference, general circulation model with 9 vertical levels and approximately 220 km horizontal resolution on a modified Kurihara grid.

The space derivatives are estimated by the finite difference operators of Kurihara and Holloway (1967). The modified Kurihara grid, in which the grid points have been increased considerably at polar latitudes in the zonal direction compared with the original Kurihara grid, has been used extensively at GFDL since 1971; it is illustrated in Fig. 2. This modification was introduced for the purpose of improving numerical accuracy in the polar regions since serious truncation errors were experienced in the original Kurihara grid. The horizontal resolution, as measured by the number of grid points between pole and equator, is 48 in the present case. For this resolution the total number of grid points on a whole global plane is 10 440 in the modified Kurihara grid, compared to 9218 in the original Kurihara grid. In order to use a large time interval in the numerical integration ($\Delta t = 200$ s), "spectral filtering" is applied near the poles (Umscheid and Sankar-Rao, 1971).

The individual physical processes common in general circulation models are all included except cloud prediction and its feedback effect to the radiation calculation.

The sun's zenith angle varies from June through September, but the diurnal variation is not incorporated for reasons of economy. The sea surface temperature is specified as a function of space and time, which was interpolated in time from the four seasonal normals [the same data as used by Manabe *et al.* (1974)]. Sea ice limit lines are also specified for the four months (see Fig. 3 for an example). Ice temperatures are also given externally. Ensembled cumulus convection is parameterized by a moist convective adjustment, and the planetary boundary layer process is treated by the mixing length method without the effect of non-neutral stratification.

The performance of this model or related models can be found in the literature, e.g., Manabe *et al.* (1970) and Miyakoda *et al.* (1971, 1974), for mid and high latitude prediction and tropical prediction, respectively, with the original Kurihara grid model; in addition, Umscheid and Bannon (prepared for publication) have compared the numerical accuracy and efficiency of the original Kurihara, the modified Kurihara, and the latitude-longitude grid models.

b. Insertion data

Parameters used for updating were winds and temperatures, excluding winds in the planetary boundary layer, and surface pressure. Water vapor data were used only for the calculation of surface pressure from upper air data.

The management of such a large and varied volume of data was not an easy task (though it is done daily in the operational centers). One of the most cumbersome and yet compulsory processes is the quality control of the data. After a preliminary check of the data for format and other obvious errors, a second quality control was performed by a conventional "objective analysis," which is based essentially upon two-dimensional interpolation (see Quality Check II in Fig. 5). Using all available data, an objective analysis was done with Welsh's (GFDL) scheme [a similar program was used by Oort and Rasmusson (1971); for the practical details of the method see this paper]. If the individual raw data deviated considerably from the resulting analysis, the data were rejected. This check, and accordingly, this preliminary analysis were conducted for all parameters at all 11 selected pressure levels at 6 or 12 h intervals. The number of rejected data depends upon the toss-out criterion for the specific parameter at the specific level.

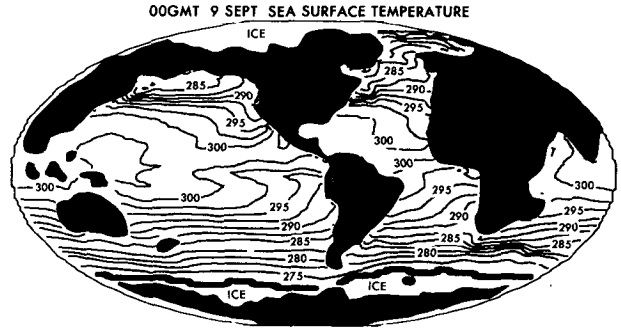

FIG. 3. Example of the specified sea surface temperature (K) for 00 GMT 9 September. Ice limits are indicated.

After these quality checks, however, we noticed that a considerable number of bad data were still not excluded. The third quality check will be discussed later in the subsection on assimilation schemes.

During most of the GATE period, we used one version of insertion data (Version 1 below), but for a selected period (4–17 September), we tested three different versions of insertion data:

Version 1. Upper air data only with direct replacement.
Version 2. Upper air and surface data with direct replacement.
Version 3. Station upper air data with optimum interpolation and the rest of upper air data and surface data with direct replacement.

In the first two versions, the insertion data at a grid point are taken directly (without any modification) from the nearest observation, or average of nearby observations, whenever and wherever they are available. This is called "direct replacement" (see Appendix B). On the other hand, in the last version, the upper air insertion data at a grid point is determined by the "optimum interpolation" technique (Gandin, 1963; Petersen and Middleton, 1963; Eddy, 1967). However, this process was applied only to the radiosonde, rawinsonde and pibal station data for the reason

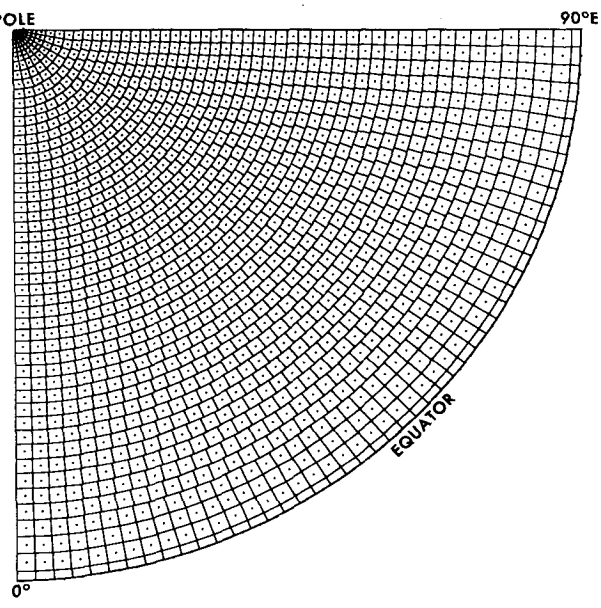

FIG. 2. The modified Kurihara grid of resolution $N=48$.

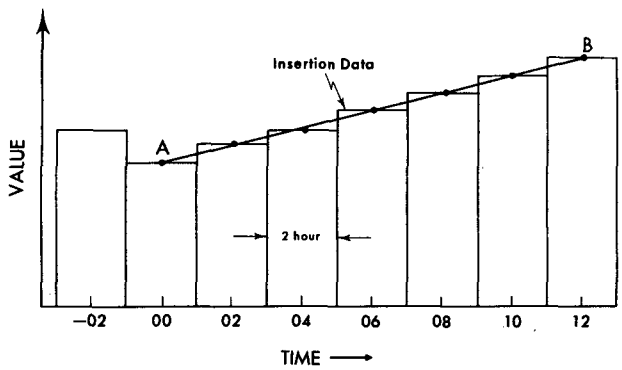

FIG. 4. Schematic illustration of time interpolation of data for insertion. Using the observed data given at 00 and 06 GMT, an interpolation is made in time to 2 h intervals. Within each 2 h interval, the same data are used for the insertion.

described in Appendix B. Among the observed surface parameters, we employed only the pressure value in this experiment. These pressures were reduced to the surface of the model's topography with the aid of the surface temperature after excluding the combined effect of terrain height and diurnal variation.

In addition to the data processing mentioned above, we interpolated the grid-point data in time to 2 h intervals, using the 6 h interval data at both ends or the 12 h interval data, if the former were not available (see Fig. 4). This arrangement of data makes the assimilation smooth and accurate, as will be discussed below.

Note that no process to insure geostrophic balance between the wind and temperature was included, and no "bogus" data were used whatsoever.

c. *Insertion and assimilation scheme*

The scheme of updating is the forward, continuous method with continuous application of Euler-backward time differencing [for terminology, see Bengtsson (1975)]. There are two groups of insertion data described above, in terms of time continuity: data interpolated to 2 h intervals, which are therefore continuous in time, and the rest of the data, which are not interpolated to 2 h intervals and accordingly are less continuous in time In the present four-dimensional analysis technique, the same data at one grid point is inserted repeatedly every time step for a 2 h interval (beginning 1 h before the valid time) so that the inserted data are forcefully assimilated by the model. Otherwise, the speed of assimilation is much slower; this fact was obtained from the rate of error reduction in observation system simulation experiment (Rutherford and Asselin, 1972; Gordon and Lusen, 1976), and also in studies of initialization with the four-dimensional assimilation technique. Every sixth synoptic hour, i.e., at 00, 06, 12 and 18 GMT, a mass of new data is entered, and as a result, the "shocking" is increased appreciably at these times (see Figs. 6 and 7). The reason for using the time interval of 2 h instead of 3 h, for example, is not profound, but rather that 2 h is almost the lowest limit to the time interval required to represent the diurnal variation; however, in this particular experiment we did not include the diurnal effect.

Another aspect which is very important and therefore worth mentioning is the data compatibility check included in the present scheme. This quality control of the insertion data is made by comparing them with the current model solution. If the discrepancy between the two is beyond the tolerance criterion, the inserted data are modified to differ only by the criterion. The philosophy of this arrangement is that the incompatible data are likely to be bad data or that, if they are correct, such data will create a large shocking, and therefore it is better to avoid them. The tolerance values currently employed are: 12 m s^{-1}, 15°C, and 20 mb for each component of the wind speed, for temperature, and for surface pressure, respectively.

Finally, we must emphasize that in order to accelerate the geostrophic adjustment process, Euler-backward time differencing (Matsuno, 1966; Kurihara, 1965) was applied continuously for every time step. [This method was used first for four-dimensional analysis by Charney *et al.* (1969).] The application of this method is expensive in computer time, but presently this process seems necessary.

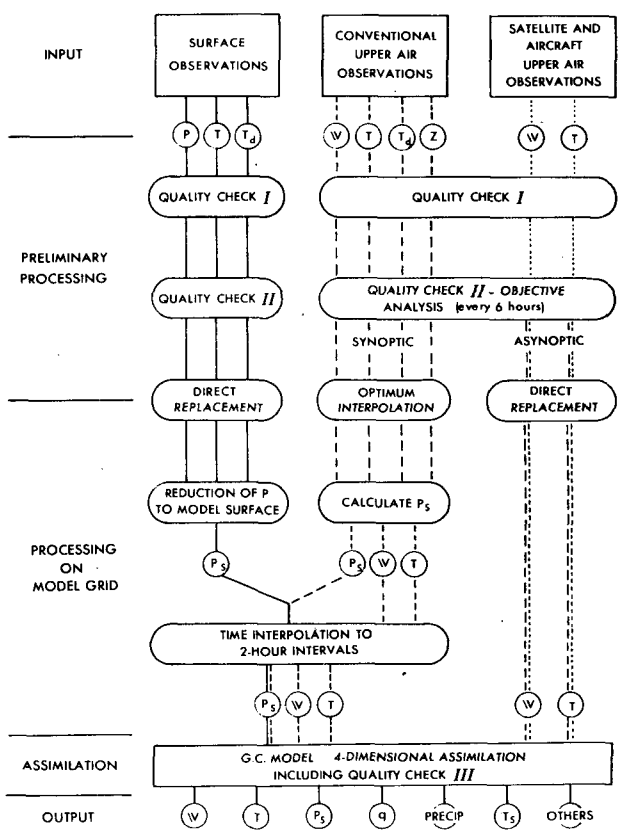

FIG. 5. Flow diagram of data processing and assimilation for Version 3. Flow is from the top down.

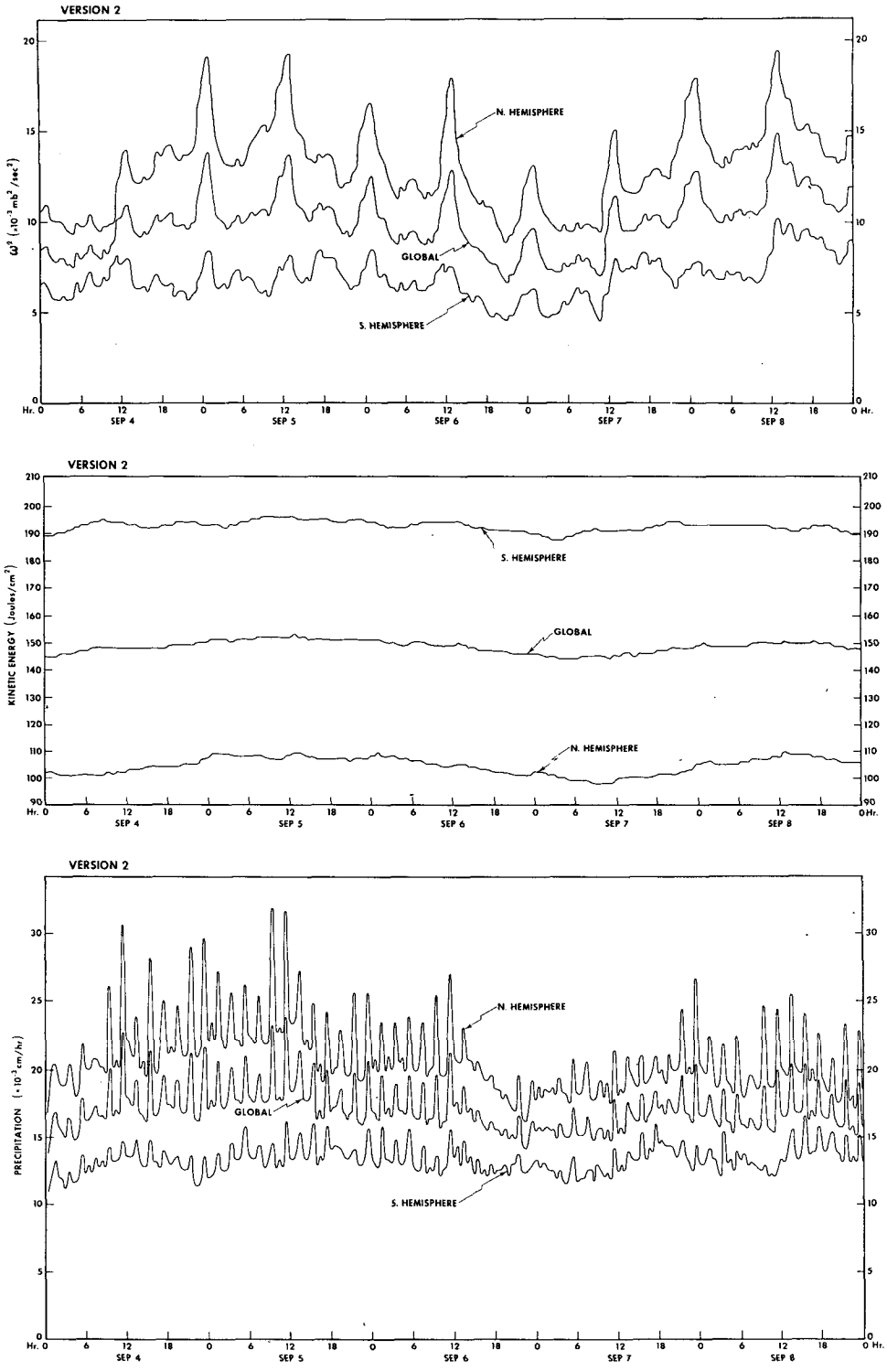

FIG. 6a. Impact of the data insertion on the model demonstrated by time series of $\overline{\omega^2}^v$, kinetic energy and the rate of precipitation in Version 2 from 4–8 September 1974.

d. Diagrams of data processing and assimilation

To summarize the above, we schematically illustrate the data handling for Version 3 in a flow diagram (Fig. 5). The diagram is divided into five stages: the input of data, the preliminary processing, the processing on model grid, the assimilation, and the output of data.

The input data are divided into three general categories: surface observations, conventional upper air observations, and satellite and aircraft observations. The notations are **V** for the wind vector, T for temperature, p for pressure, z for the geopotential, T_d for dew-point temperature, q for the mixing ratio of water vapor, and p_s for the surface pressure at the height Z_* of the model topography. The parameters that were actually used for the four-dimensional analysis were **V**, T and p_s.

The main function of the preliminary processing is the quality control of the input data and classification of different parameters. Throughout the preliminary processing phase the upper air data are carried on 13 pressure levels, i.e., 4, 10, 30, 50, 100, 150, 200, 300, 500, 700, 850, 1000 and 1150 mb. The interpolation to these levels is accomplished just after Quality Check I. After Quality Check II, the input upper air data are sorted into two groups according to synoptic and asynoptic times, where the synoptic times are 00, 06, 12 and 18 GMT. The model's grid is first introduced in the optimum interpolation or direct replacement step. Then the model's topography is introduced for the estimation of p_s. The data are assigned to the model's vertical levels by vertical interpolation just before the insertion into the model.

Finally, the output data for various parameters are obtained at model grid points.

e. Time variation of the assimilated variables

The global three-dimensional average of ω^2, where $\omega = dp/dt$, is normally calculated as a measure of the degree of geostrophic imbalance or "shocking." Fig. 6 shows an example of time series of $\overline{\omega^2}^v$, for the Northern Hemisphere, the Southern Hemisphere and the global average, compared with the number of insertion data in Version 2 for a selected period, where $\overline{(\,)}^v$ means the three-dimensional average.

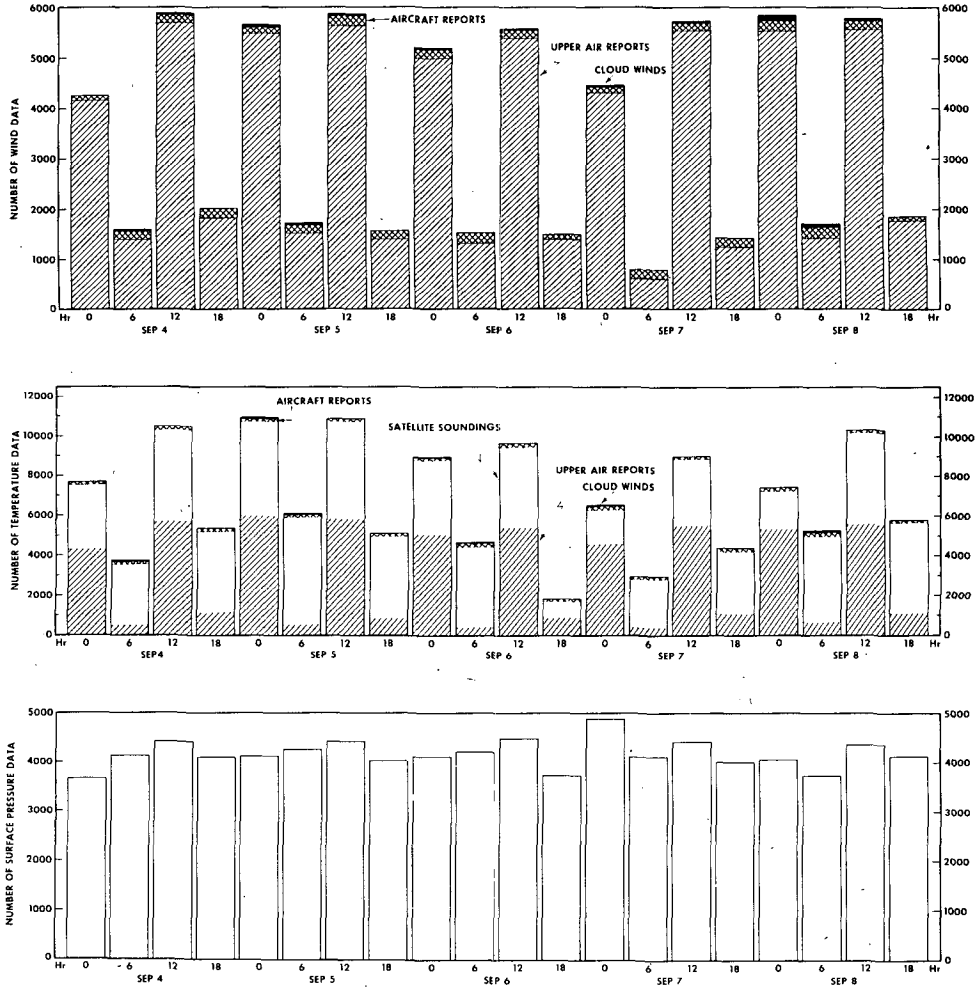

FIG. 6b. Counts of the received data for wind and temperature summed in the vertical and classified by the modes of measurement. Note in panel 2 that cloud wind data are accompanied by temperature data.

Kinetic energy $\int \rho \cdot \frac{1}{2}(u^2+v^2)dV$, and the rate of spatially averaged precipitation are also shown. The data available for insertion are divided into 6 h intervals and classified by their modes of measurement. Wind data are provided from conventional observations, cloud-tracking, and commercial and reconnaissance aircraft reports. Temperature data are from conventional radiosonde, satellite vertical soundings, aircraft reports, and cloud-top measurements from SMS-A. The quantity of data reported in Fig. 6 includes contributions from the 11 selected levels. Conventional observations and the satellite vertical soundings generally yield data at all 11 levels (excluding 4 and 1150 mb), and the cloud-tracked winds, cloud-top temperatures and aircraft reports are often given at only one or two levels. Since the counts shown here were obtained at an early stage of data processing, the data were not yet sorted according to the model's vertical levels. The number of actual insertion data for this period can be found in Fig. 7. From a comparison with Fig. 6, we see that $\overline{\omega^2}^v$ is larger

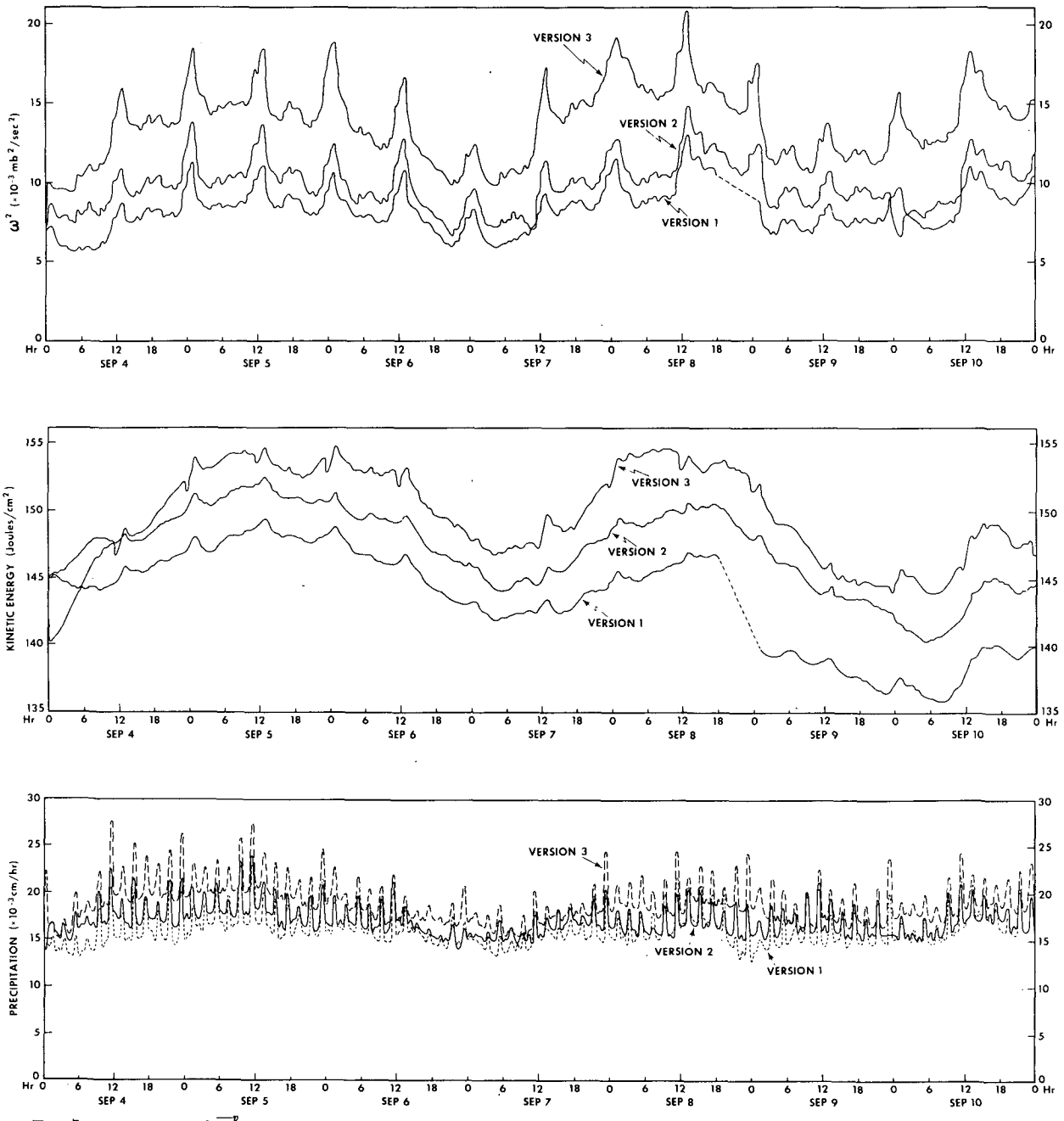

FIG. 7a. Time series of $\overline{\omega^2}^v$, kinetic energy and precipitation for the three versions of insertion data from 4–10 September 1974.

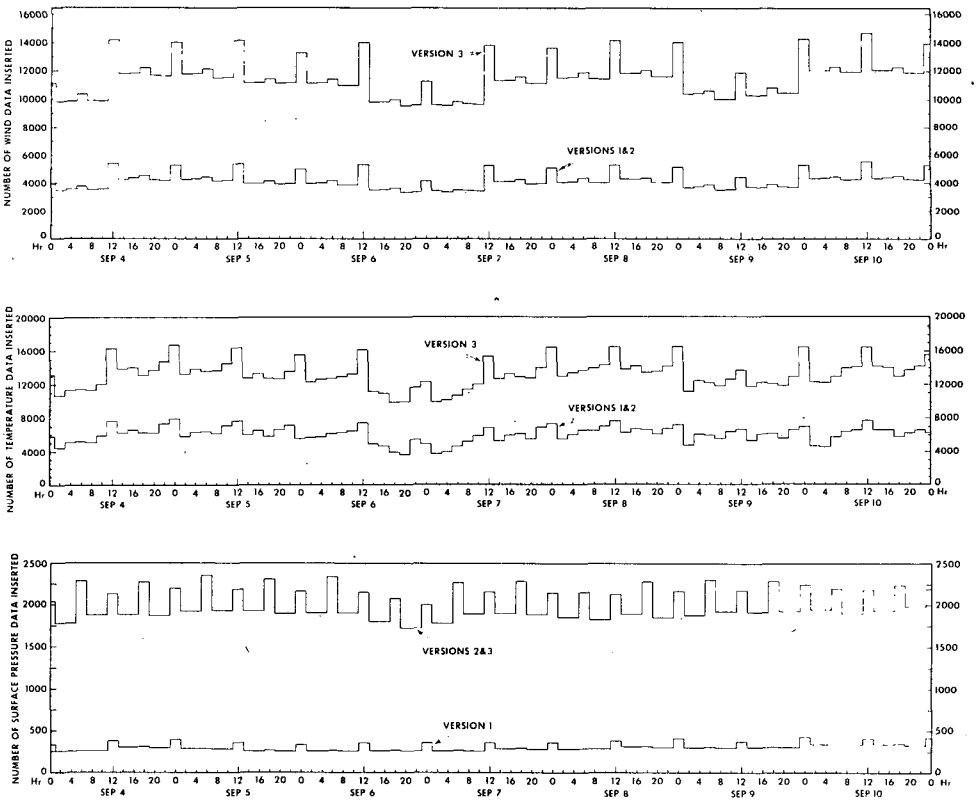

FIG. 7b. Number of insertion data for wind, temperature and surface pressure for the three versions.

in the Northern Hemisphere than in the Southern Hemisphere, reflecting the fact that the insertion data were numerous in the Northern Hemisphere. Yet the kinetic energy is greater in the Southern than in the Northern Hemisphere, since it was winter in the Southern Hemisphere. For a similar reason the rate of precipitation is greater in the Northern Hemisphere. $\overline{\omega^2}^v$ shows major peaks every 12 h and minor peaks every 6 h in between, indicating that at these times, there were simultaneously inserted a large number of data which were not interpolated in time.

Fig. 7 shows the difference in $\overline{\omega^2}^v$, kinetic energy, and the rate of precipitation for the three versions of data insertion. Included in this diagram are the counts of actual insertion data for wind and temperature (summation at 9 vertical levels) and surface pressure. As expected, the yields of insertion data are largest in Version 3 by a factor of about 2.4, due to optimum interpolation. $\overline{\omega^2}^v$ is increased in Version 2 compared with Version 1 due to the intense insertion of surface data. But the increase of $\overline{\omega^2}^v$ is more appreciable in Version 3 due to the large number of upper air data insertions.

The differences in the magnitude of $\overline{\omega^2}^v$ for the different versions are perhaps plausible. But it may be somewhat puzzling that the magnitudes of kinetic energy and of precipitation are different for the three versions. This may imply that these values in the four-dimensional analysis are composed of two components, i.e., the meteorologically significant part and the part produced artificially due to the gravity waves induced by the data-insertion "shocking." However, the percentage of the second component is very small and the perturbations seem to be confined to the small-scale features. Therefore, the analyzed patterns of any variables are filtered by local smoothing and the resultant fields have proven to be very reasonable in comparison with reference fields in the preliminary hypothetical data studies. It is noted that the maps shown here were all filtered by local spatial smoothing.

4. Results of the four-dimensional analyses

For the selected period from 4 through 17 September 1974 the three versions of analyses were compared with each other, and with the analyses of NMC in Washington, D. C., and ANMRC in Melbourne whenever the analyses were available. The NMC maps are the global analyses based on the Hough function method (Flattery, 1970), which is a three-dimensional objective analysis using the 12 h forecast of the NMC 8-level global model (Stackpole *et al.*, 1974) as the initial guess. These analyses were obtained separately from the operational analyses, using the data with the delayed cutoff time, which are supposedly the same data as used in our four-dimensional analyses. The ANMRC

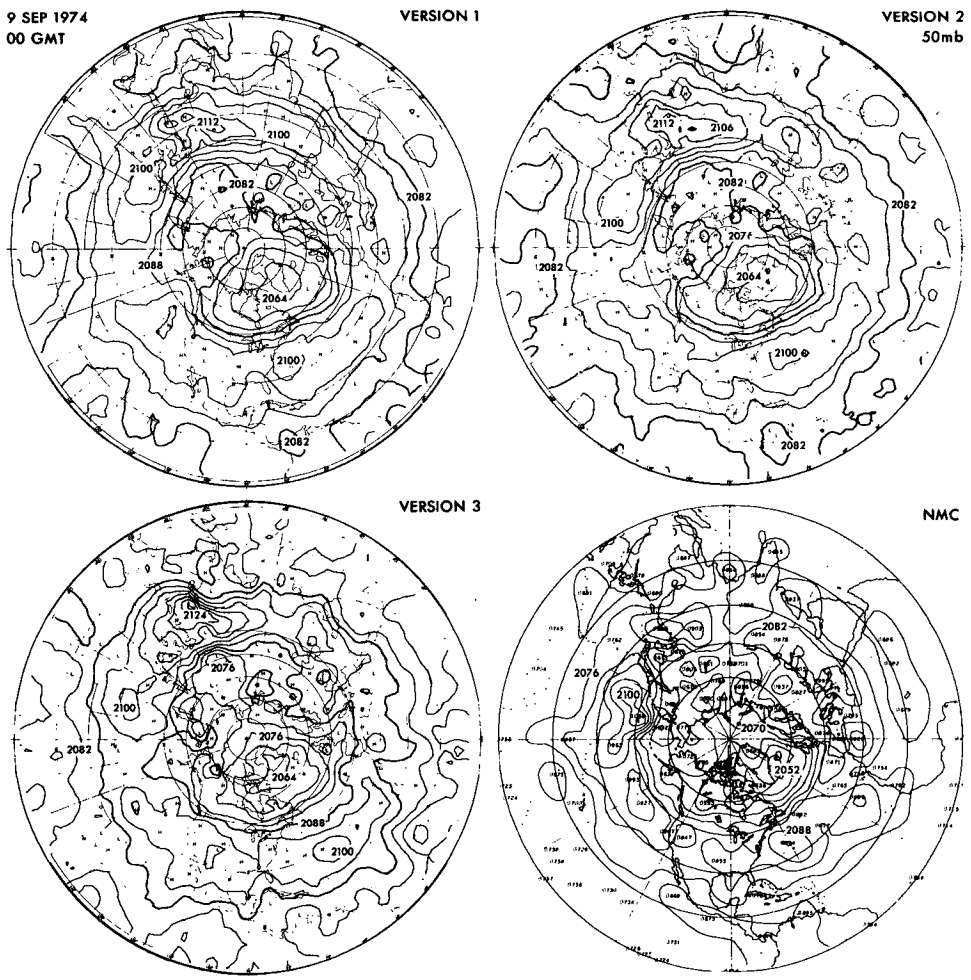

FIG. 8a. Geopotential height contours (dkm) for the Northern Hemisphere at 00 GMT 9 September 1974 for Versions 1, 2, 3 and NMC at 50 mb, contour interval 60 m.

analyses are produced by Kelly (1975); the explanation of his scheme will be given later.

a. The Northern Hemisphere extratropics

Fig. 8 presents the geopotential height contours on stereographic projection maps for 00 GMT, 9 September 1974, at 50, 200, 500 and 1000 mb levels. Maps of the plotted data corresponding to this time are given in Fig. 1. In general it appears that the analyses are improved from Version 1 to Version 2, and little difference is noticeable between Versions 2 and 3.

Let us next compare them with those of NMC. An overall similarity at all levels may be seen except for the stratosphere. One can identify almost all cyclones and anticyclones, or troughs and ridges in middle and high latitudes, and their locations are almost exactly the same for all analyses. Yet it is clearly noticeable that the NMC map at 200 mb has more concentrated contour lines than any of the three versions of the GFDL four-dimensional analysis, where Version 3 is most similar to NMC with respect to the contour interval. Overall, in the four-dimensional analysis, the contour lines are smoother and the cyclones are shallower than those in the NMC map. It is likely that the stronger winds are most realistic; this aspect seems to be particularly important to forecasting. It is not known whether the smoothness is due to the technique of data assimilation or the spatial resolution of the model. As will be discussed later, however, more effective application of the optimum interpolation may intensify the degree of concentration of the contour lines.

It is interesting to note that the 1000 mb maps of Version 1 surprisingly resemble the others in spite of the fact that this version did not use surface data. The similarity is more pronounced in the maps at the higher levels. The differences in the height patterns between different versions can hardly be detected.

On the 50 mb maps, easterlies predominate at this time of year, and the geopotential height fields have low relief. Accordingly, the GFDL analyses do not agree with those of NMC.

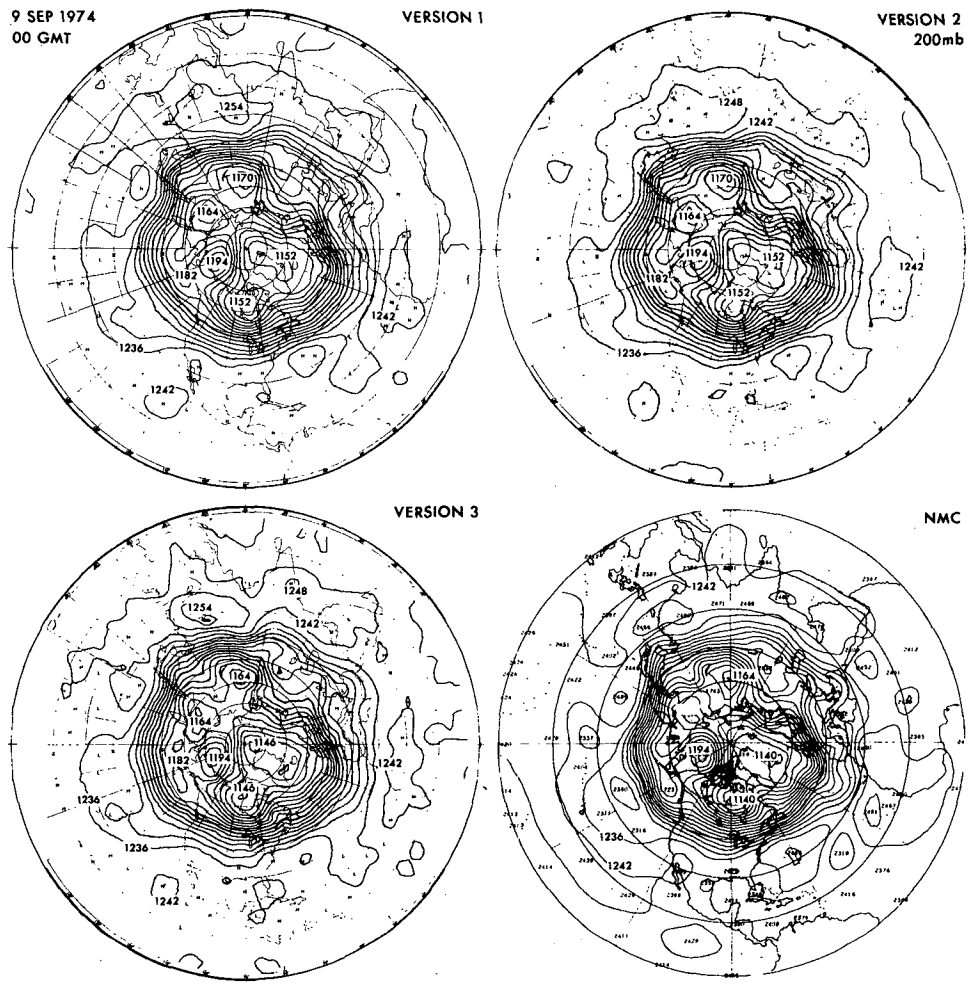

FIG. 8b. As in Fig. 8a except for 200 mb.

b. The Southern Hemisphere extratropics

Fig. 9 is the same as Fig. 8 but for the Southern Hemisphere. The differences between the maps of Version 1 and Versions 2 or 3 are relatively larger in the Southern Hemisphere than in the Northern, particularly for the 1000 mb maps; it appears that the impact of surface data on the analysis is greater in this hemisphere than in the Northern. Likewise, the effect of the optimum inter polationis also more appreciable in these charts than in the Northern Hemisphere. This is probably because the observed data are so scarce in the Southern Hemisphere that any effort to increase data has a significant affect on the analysis. The location of intense wind areas in Versions 1 and 2 are quite different from NMC, but Version 3 does come close to NMC on the 200 and 500 mb maps, for example.

One noteworthy difference, however, between our present analysis and that of NMC is that the Flattery analyses have many vortices at the high latitudes around Antarctica, where data are critically sparse.

The four-dimensional analyses are also compared with Melbourne's analysis in Fig. 10. These maps were produced by using the operational analysis of WMC (World Meteorological Center) Melbourne as the first guess and then modifying the short-wave components on the maps so as to fit the observed cloud patterns (Kelly, 1975). It seems that the NMC and ANMRC analyses are similar in middle latitudes. In particular, the troughs and ridges are more pronounced in the NMC and ANMRC analyses compared with any of the four-dimensional analyses. However, the vortices which are located in circumpolar latitudes in the NMC analysis are missing in the ANMRC analysis as well as in the four-dimensional analysis. It may be interesting to note that the trough at 120°W longitude is represented equally well by the Melbourne and GFDL Version 3 analyses despite the data-sparse region. Intensification of the troughs and ridges in the four-dimensional analyses may perhaps be achieved by more effective use of optimum interpolation.

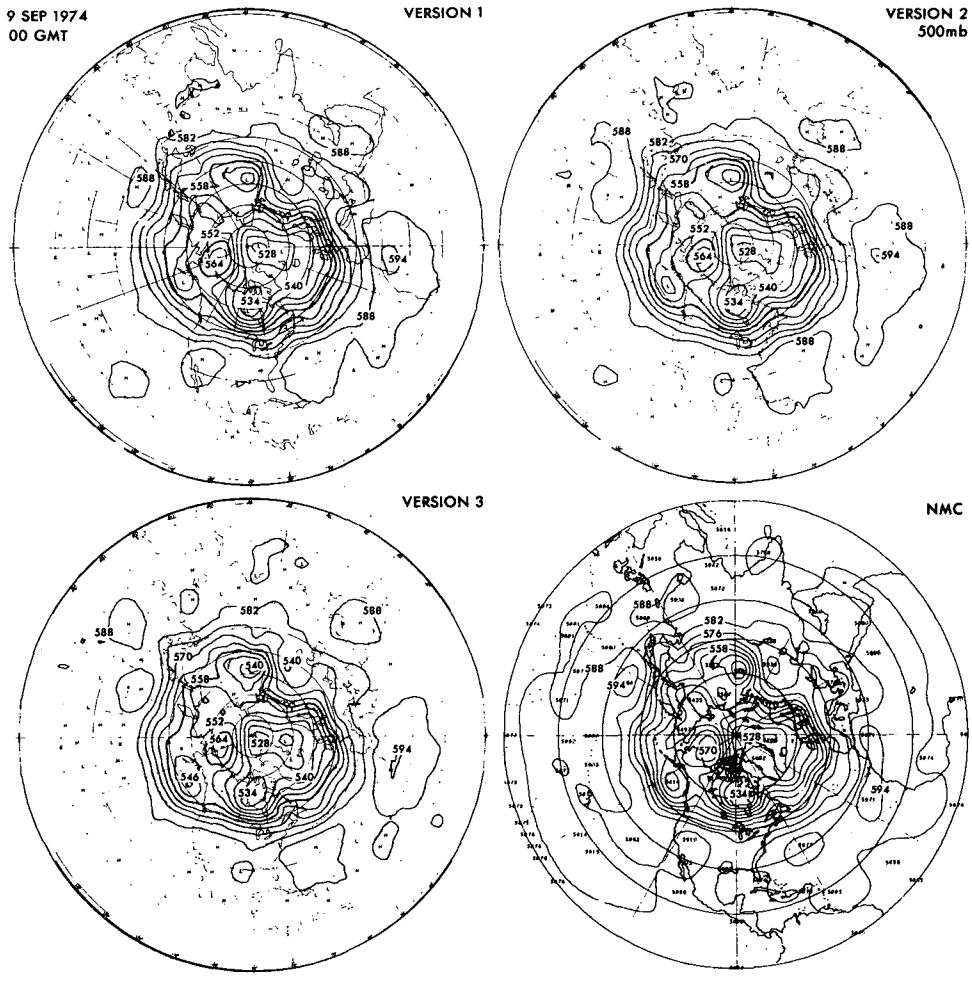

FIG. 8c. As in Fig. 8a except for 500 mb.

c. Tropics

The tropical analyses are illustrated here by streamlines on Mercator projection maps, although this zonal belt region also includes some of the temperate latitudes. The procedure for drawing these streamlines was described by Miyakoda *et al.* (1974). The examples shown in Figs. 11 and 12 are the 200 and 850 mb flow patterns, respectively, at 00 GMT 6 September 1974. In addition to the three versions of four-dimensional analysis, the NMC analysis is also shown.

It may be seen that Versions 1, 2 and 3 are similar to each other, but different from the NMC analysis, although there are common features in the large-scale patterns and in the area of good data coverage. Overall, the resemblance is noticeable in the 200 mb maps, but not as noticeable in the 850 mb maps, so far as the tropics are concerned. Large disagreements occur in the areas where no or few data are available. For example, consider the tropical areas in the 200 mb maps in the equatorial Pacific from the Phoenix Islands to the Galapagos Islands and also over equatorial Africa. A most striking difference is that the 850 mb flow pattern in the four-dimensional analysis has straight flow in the trade-wind region, whereas the NMC analysis includes many vortices; this feature is similar to that at the high latitudes in the Southern Hemisphere. It appears that in the Flattery analysis the rotational component of flow is dominant and the divergent component is missing; this point will be mentioned later.

In the 200 mb streamline patterns of 6 September, the mid-Atlantic and mid-Pacific troughs are pronounced. Beneath the tip of the mid-Atlantic trough, hurricane Carmen is located over the Yucatan Peninsula, the hurricane manifesting itself in the analyses as anticyclonic outflow at the 200 mb level. This hurricane later crossed the Gulf of Mexico and hit land on the coast of Louisiana on 8 September (see the stereographic map on 9 September in Fig. 7). As seen in the 850 mb streamline maps, hurricane Carmen is somewhat obscure in Version 1, whereas it is well

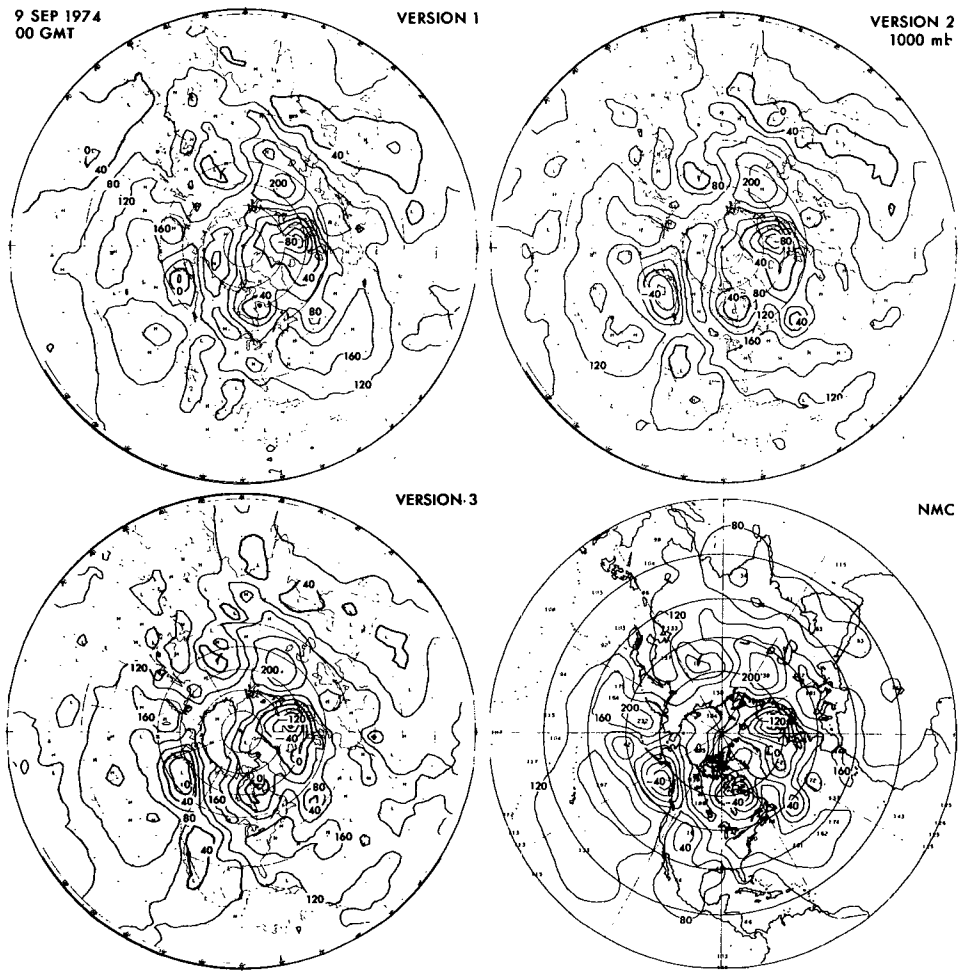

FIG. 8d. As in Fig. 8a except for 1000 mb, contour interval 40 m.

reproduced in Versions 2 and 3. It is noted that the equatorial flow patterns at 200 mb seem to be closely related to the distribution of land and sea in the equatorial belt.

It should be stressed, compared with the data acquisition of ten years ago (Miyakoda *et al.*, 1974), that the situation during the GATE period was remarkably improved. The high density of data in this experiment allows the production of less subjective (or less model-dependent) and, therefore, more reliable analyses.

d. Time means of the analyses

In order to see some statistical characteristics of the analyses, we calculated the time averages of the wind vectors, temperature, moisture, etc., for various analyses. Some of these results are shown below. The averaging period is 13 days from 00 GMT 4 September through 00 GMT 17 September and the treated analyses are taken at 12 h intervals.

The wind vector is expressed by two scalar variables—the streamfunction ψ and the velocity potential χ. The streamfunction and the velocity potential are defined by

$$\left.\begin{aligned} u &= -\frac{\partial \psi}{a\partial\varphi} + \frac{\partial \chi}{a\cos\varphi\partial\lambda} \\ v &= -\frac{\partial \psi}{a\cos\varphi\partial\lambda} + \frac{\partial \chi}{a\partial\varphi} \end{aligned}\right\},$$

where u and v are the wind speed in the zonal and meridional directions, respectively, a is the radius of the earth, and λ and φ are the longitude and latitude, respectively. The boundary conditions are

$$\left.\begin{aligned} \psi_N &= \bar{\psi}_0 + \frac{gZ_N}{2\Omega\sin\varphi} \\ \chi_N &= 0 \end{aligned}\right\} \quad \text{at} \quad \varphi=\varphi_1,$$

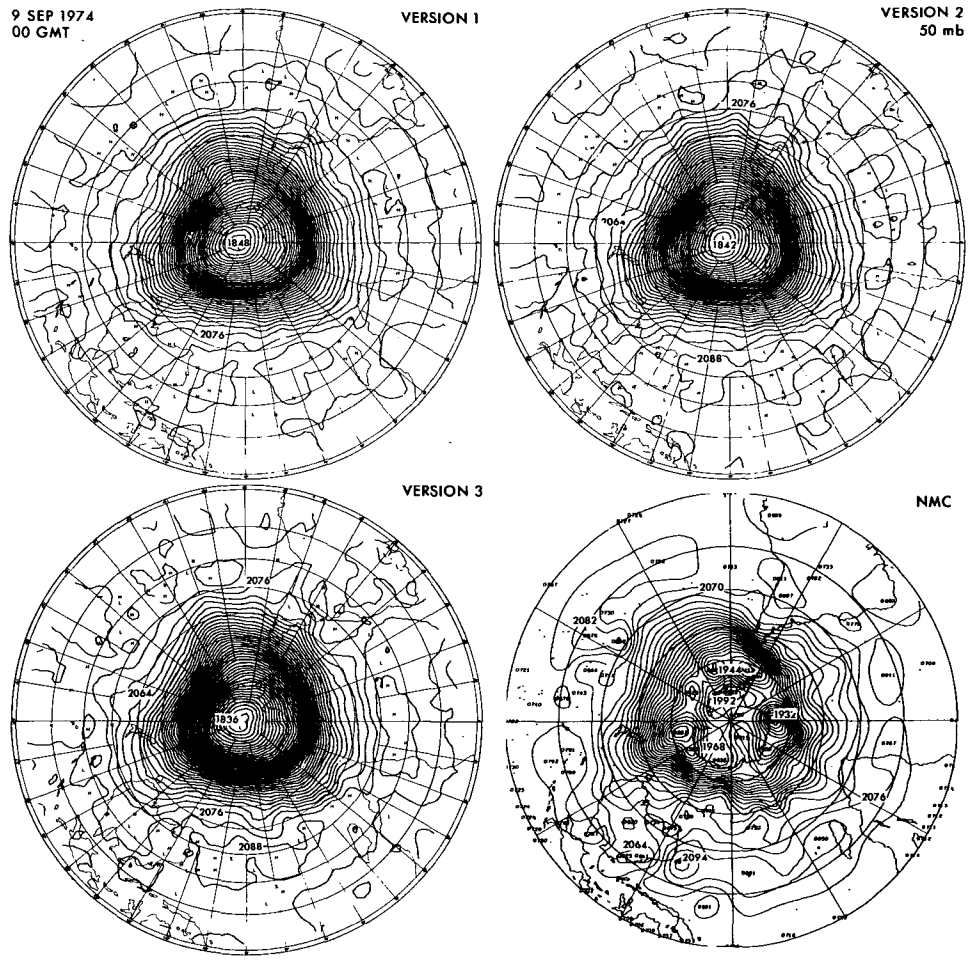

FIG. 9a. As in Fig. 8a except for the Southern Hemisphere.

and

$$\left.\begin{aligned}\psi_s &= -\bar{\psi}_0 - \frac{gZ_S}{2\Omega \sin\varphi}\\ \chi_s &= 0\end{aligned}\right\} \quad \text{at} \quad \varphi = -\varphi_1,$$

where

$$\bar{\psi}_0 = \frac{a}{4\pi}\int_{-\varphi_1}^{\varphi_1}\int_0^{2\pi} u\,d\lambda\,d\varphi,$$

Z is the geopotential height, g the acceleration of gravity, Ω the earth's rotation, φ_1 is taken as 45° latitude, and the subscripts N and S indicate the northern and southern boundaries, respectively. The scheme to solve ψ numerically was described in Miyakoda *et al.* (1974), and χ is likewise obtained in a similar manner.

Figs. 13 and 14 are time-average maps of ψ and χ, respectively, at 50, 200, 500 and 850 mb for Version 3 and NMC. It was noticed that the χ component in Flattery's analysis appeared to be zero or negligibly small, so that only the Version 3 charts of χ are presented here.

The maps of ψ include the typical features of the tropical general circulation, such as the tropical easterly jet, the mid-Atlantic trough and mid-Pacific trough in the upper troposphere, mid-oceanic subtropical anticyclones, African easterly jet, the trade easterly winds, the Somali low-level jet in the lower troposphere, and so forth.

Most outstanding at the 200 mb level in the summer Northern Hemisphere are the tropospheric mid-oceanic troughs tilting from northeast to southwest, and two major anticyclones, i.e., the Tibetan anticyclone and the Mexican high (Krishnamurti, 1971a). In the Southern Hemisphere, where it is winter, strong westerlies prevail, forming five weak troughs.

At 850 mb in the Northern Hemisphere, two major anticyclonic centers are located in subtropic latitudes at this time of the season, alternating with cyclonic regions. It is interesting to note that a cyclonic region in the western hemispheric tropics exists over the

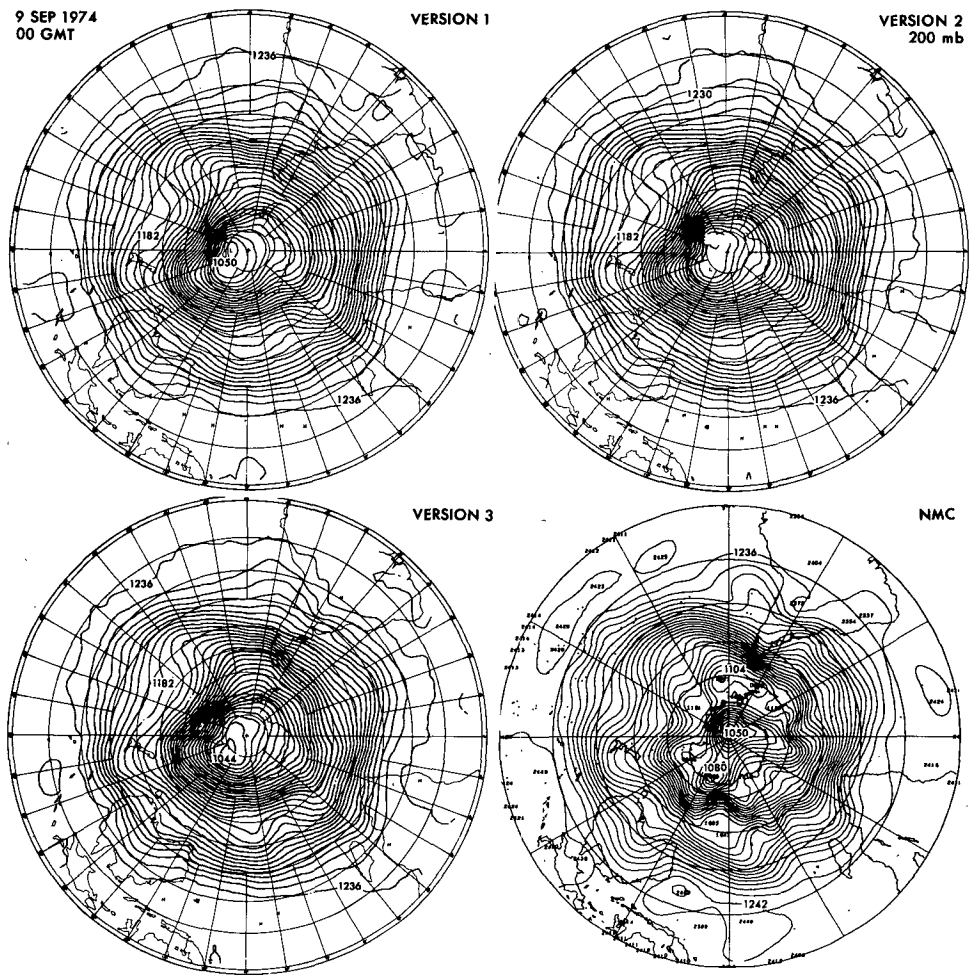

FIG. 9b. As in Fig. 8b except for the Southern Hemisphere.

Caribbean and east Pacific, which corresponds to the hurricane genetic region, and a cyclonic region in the Pacific Ocean extends from Southeast Asia through the Marshall Islands area, corresponding to the typhoon genetic region. As mentioned earlier, these cyclonic regions are beneath the southwestern portion of the mid-oceanic troughs in the upper troposphere, say 200 mb. This point will be discussed further in Part II. In the Southern Hemisphere, five anticyclones can be identified, including three major ones.

The streamfunction patterns at 200 and 500 mb agree well between Version 3 and NMC except along the equator in the eastern Pacific and Indian Oceans. Some discrepancies may be noticed in the 850 mb maps. First, the trades are considerably stronger in Version 3 than in the NMC maps; second, the flow near the equator over the Andes mountains is deflected on the NMC chart, whereas the flow is straight in Version 3 despite the presence of the mountain barrier; third, the African easterly jet is observed in Version 3, whereas it is very weak on the NMC map; and finally, the Somali jet is found in Version 3 but not in the NMC maps (few data were available in this area). In this connection, the streamfunction maps over the GATE area in summer, 1972, obtained by Tripoli and Krishnamurti (1975), showed that flow deflection over the Andes is appreciable, whereas their Somali jet resembles that in Version 3.

The χ maps in Fig. 14 show the flow of the divergent component, where the wind direction associated with χ (perpendicular to the contours) is indicated by arrows. The pronounced aspects are that the divergent component of the wind is appreciably stronger at 850 and 200 mb than at other levels, i.e., 50 and 500 mb, and that the wind intensity of χ component is much smaller than that of the ψ component even in the tropical region; the ratio is about $\frac{1}{2}$ at 850 mb. Looking at Fig. 14, one may see that a large center of outflow at 200 mb is located over the eastern part of China at 30°N, which is therefore somewhat shifted from the average summertime position of the Tibetan anticyclone (see Krishnamurti, 1971a,b; Newell *et al.*, 1972; Sadler

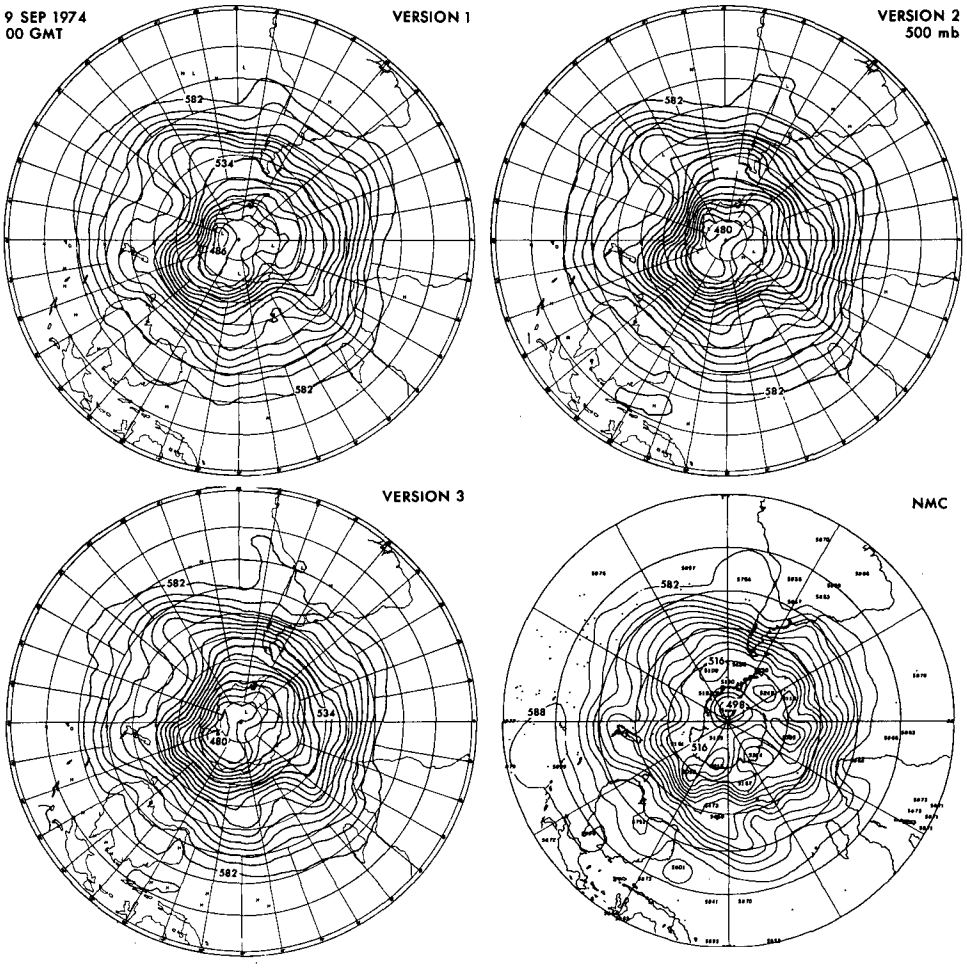

FIG. 9c. As in Fig. 8c except for the Southern Hemisphere.

and Harris, 1970). The reason is not clear; possibly it is peculiar to this year. One of the intriguing facts is that at 200 mb a center of convergence is located over the tropical Atlantic, near the GATE A/B-scale ship array area. The divergence pattern at 200 mb obtained by Krishnamurti (1971b) also shows that the GATE area corresponds to a center of convergence.

At 850 mb, the main convergence regions are associated with the Asian southwest monsoon and the intertropical convergence zones in the northern portion of the equatorial Pacific. This inflow seems to occupy a deep layer through the 500 mb level with outflow at about the 200 mb level and above. On the other hand, the GATE area corresponds to a center of divergence at 850 mb, which is compensated by convergence at 200 mb and even at 50 mb. Another convergence center at 850 mb is found over the eastern portion of China and the east coast of Africa corresponding to the Somali low level jet, though the latter is associated with a shallow circulation as seen in the vertical distribution of divergence.

It is of special interest to compare the χ field obtained by Tripoli and Krishnamurti (1975), although their averaging period was the three months of June, July and August in 1972, and their domain was limited to the GATE area (note that their definition of χ has a sign opposite to that in this paper). Our strong convergence center near the Red Sea and the divergence center near Madagascar, which are apparently associated with the Somali Jet, coincide with those on their map. The divergence center over the GATE A/B scale area and the convergence center over the South American Continent also agree with Tripoli and Krishnamurti. The differences in their map, however, are that a weak convergence center is located along the ITCZ at 10°N over the Atlantic Ocean, and that a large divergence center is found over Brazil.

e. Differences between the analyses

In order to obtain quantitative values of the differences between the analyses, we calculated the root-

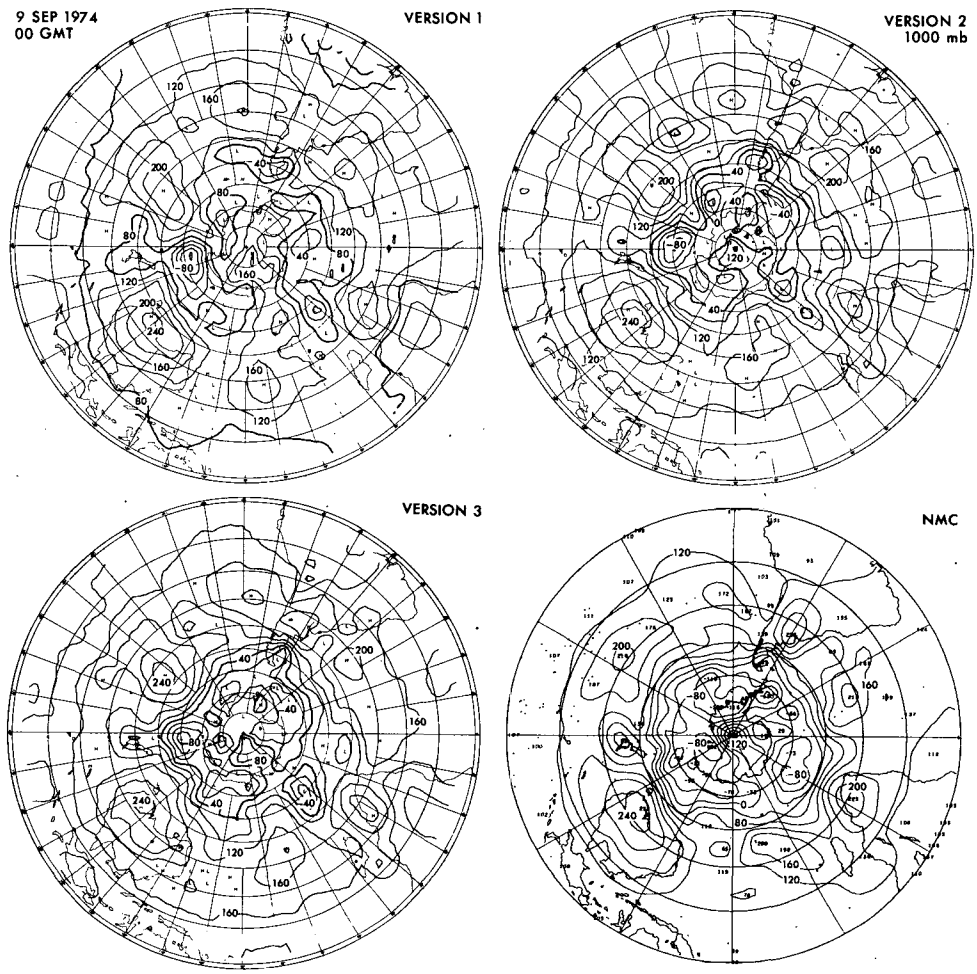

FIG. 9d. As in Fig. 8d except for the Southern Hemisphere.

mean-square (rms) differences of wind vectors and temperatures between two analyses and averaged the differences for the 13-day period.

The rms of temperature difference is calculated based on the formula

$$[\overline{(T_1-T_2)^2}^t]^{\frac{1}{2}}, \tag{4.1}$$

where the subscripts 1 and 2 refer to the two different analyses, and $\overline{(\ )}^t$ indicates the time average for 13 days. The zonal mean of rms of temperature difference is calculated using

$$\left[\overline{\overline{(T_1-T_2)^2}^t}^{\lambda}\right]^{\frac{1}{2}}, \tag{4.2}$$

where $\overline{(\ )}^{\lambda}$ denotes the zonal average over 360° of longitude.

Similarly, the rms of wind vector difference is obtained from

$$[\overline{(U_1-U_2)^2}^t+\overline{(V_1-V_2)^2}^t]^{\frac{1}{2}} \tag{4.3}$$

and

$$\left[\overline{\overline{(U_1-U_2)^2}^t}^{\lambda}+\overline{\overline{(V_1-V_2)^2}^t}^{\lambda}\right]^{\frac{1}{2}}. \tag{4.4}$$

The first comparisons were made between Version 3 and Version 1. Fig. 15 shows the meridional charts for the zonal mean of the rms differences. The overall wind difference is less than 6 m s^{-1} below the tropopause in the Northern Hemisphere; a maximum difference of 8.8 m s^{-1} is found in the stratosphere. Another maximum is observed in the lower part of the troposphere. This is understandable, since Version 1 did not use surface data as did Version 3. In the winter Southern Hemisphere, the difference appears to be considerably larger (the maximum is 10.8 m s^{-1}) lending support to the contention that the impact of surface data as well as the optimum interpolation is much more effective and appreciable for the Southern Hemisphere, particularly in winter. It is worthwhile to point out that the wind differences are maximum at the 200 mb level at the equator, i.e., 7.5 m s^{-1}. This agrees qualitatively with

Gordon *et al.* (1972), who found that the wind error appeared very large at the same location although their experiment was not for a real data analysis but for an observation system simulation study with hypothetical data.

The temperature differences are in general less than 1.5°C for most of the Northern Hemisphere. Larger differences appear near the earth's surface (2°C) and in the area of the Southern Ocean (5°C). Fig. 15 seems to indicate that the analysis difference comes not only from differences in analysis technique and data availability, but also strongly reflects some dynamical characteristics; in other words, the error distributions in general are partly the consequences of dynamical instability and the response of the respective elements.

The second comparison was made between the Version 3 and NMC analyses. Fig. 16 shows the meridional charts of zonal means of the rms differences of wind and temperature. Compared with the previous case, the discrepancies are much larger. In particular, it is enormous in the Southern Hemisphere between 30°S and the pole, the maximum being 55.4 m s^{-1} at 60°S at the 50 mb level. The main reason for this huge difference is obviously due to the lack of data in those regions but very probably also due to the shortcomings of the analysis scheme itself. Another maximum difference, 15.9 m s^{-1}, is observed at the equator, a value which is almost the same as the magnitude of the natural variability (see Gordon *et al.*, 1972). The overall difference in the middle and lower troposphere of the Northern Hemisphere is 4–6 m s^{-1}; the difference in the subtropical jet region is 10 m s^{-1}. The temperature difference in Fig. 16b has extremely large discrepancies in the Southern Hemisphere; one is near the surface of

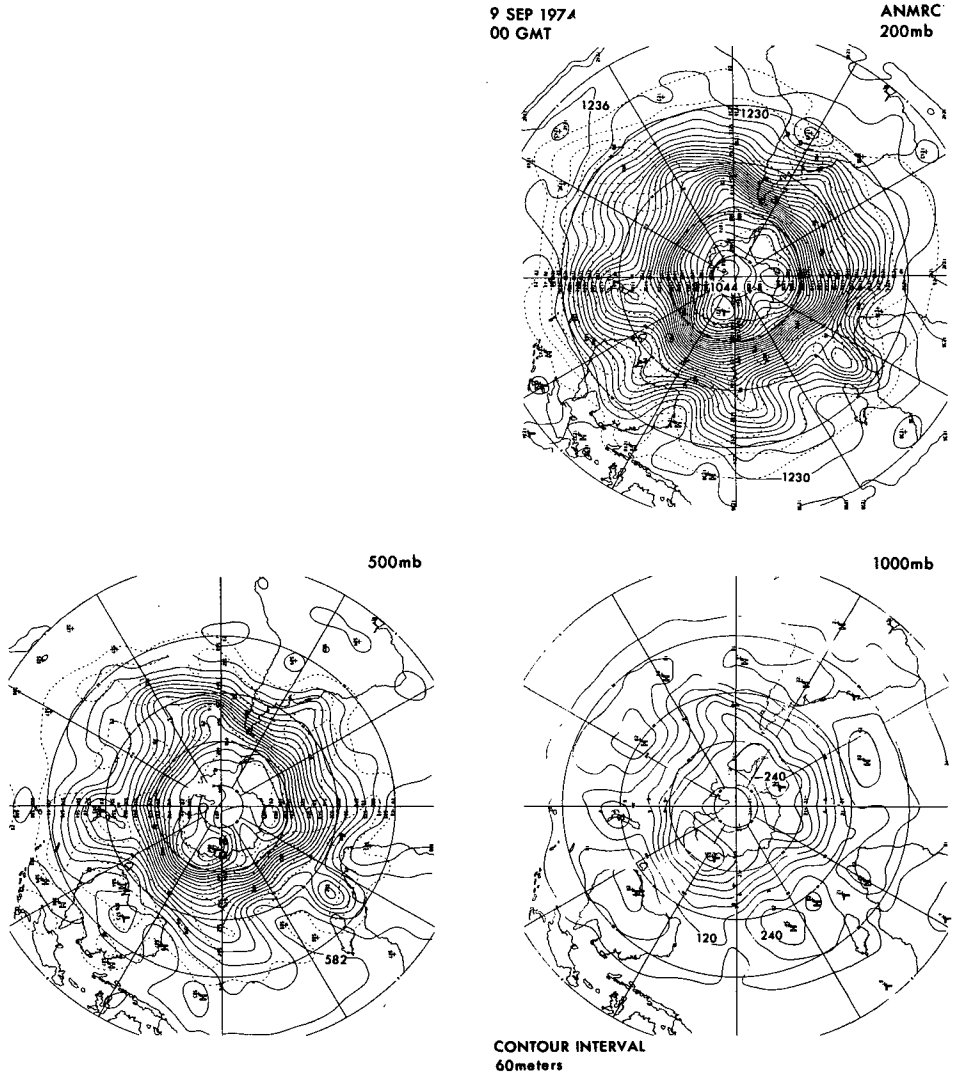

FIG. 10. Geopotential height charts for the Southern Hemisphere at 00 GMT 9 September 1974, produced at Melbourne, Australia. The contour intervals are all 60 m; therefore, the interval is different from the maps of Fig. 8d.

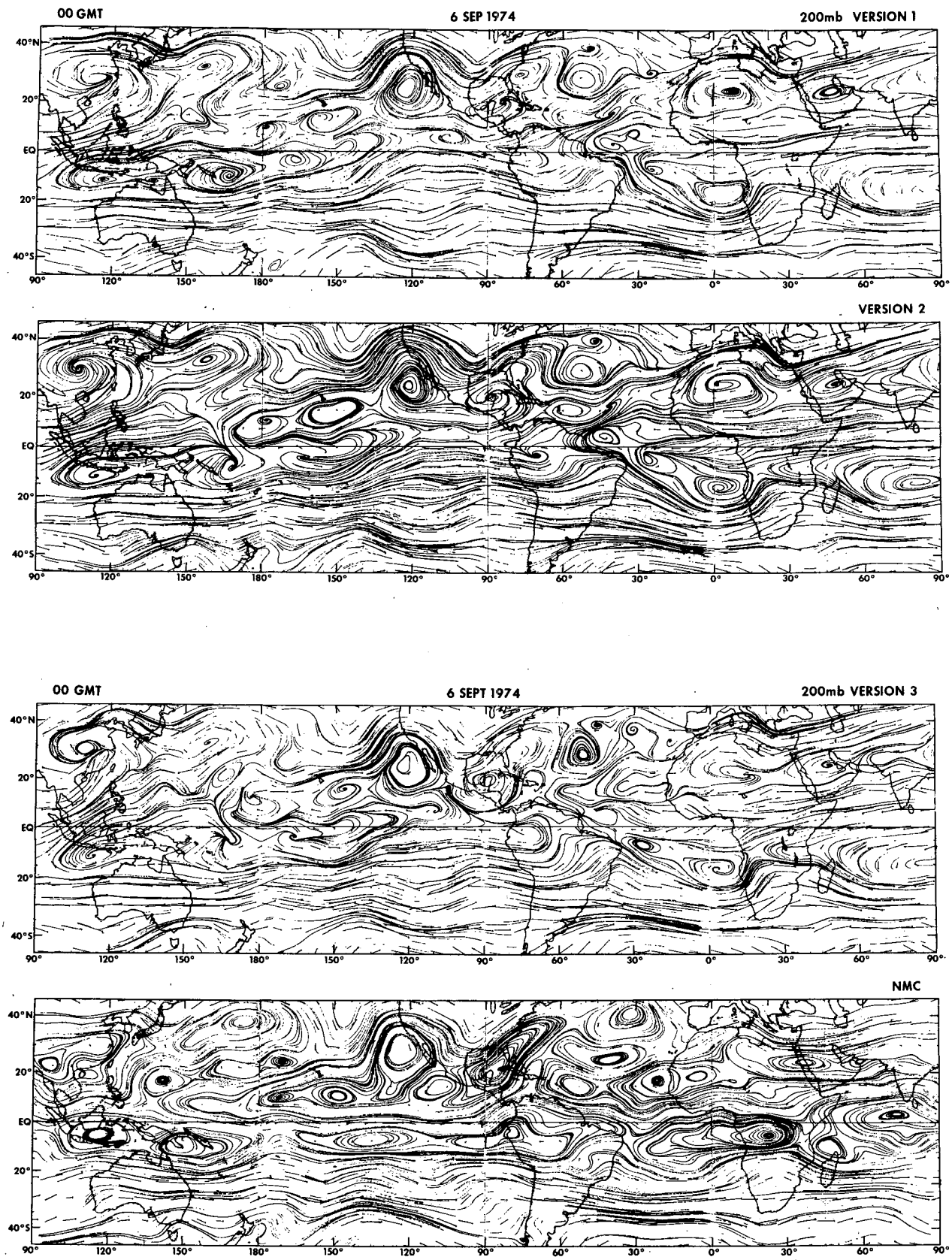

FIG. 11. Streamlines on a Mercator projection at 200 mb level at 00 GMT 6 Septebmer 1974 for Versions 1, 2, 3 and NMC.

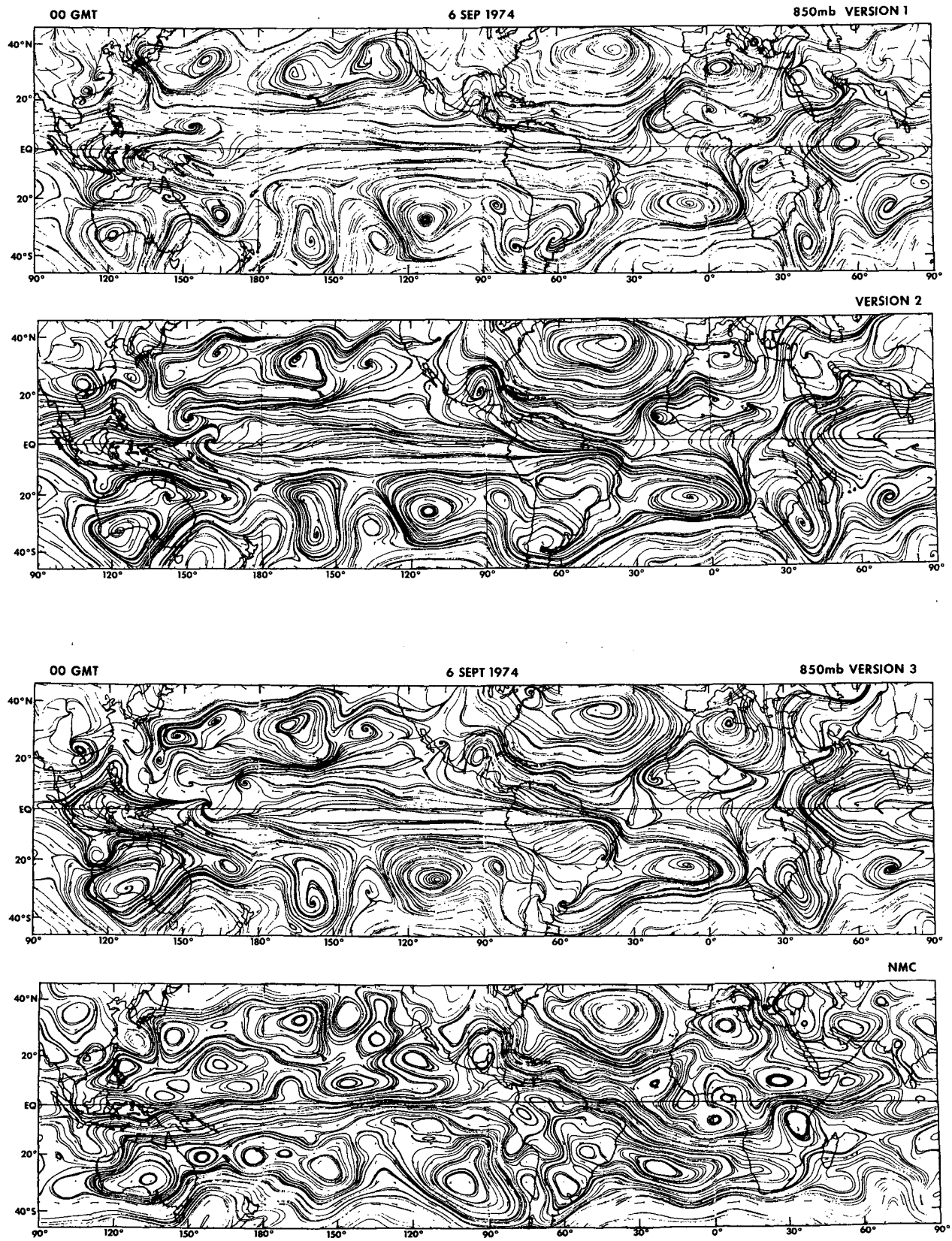

FIG. 12. As in Fig. 11 but for the 850 mb level.

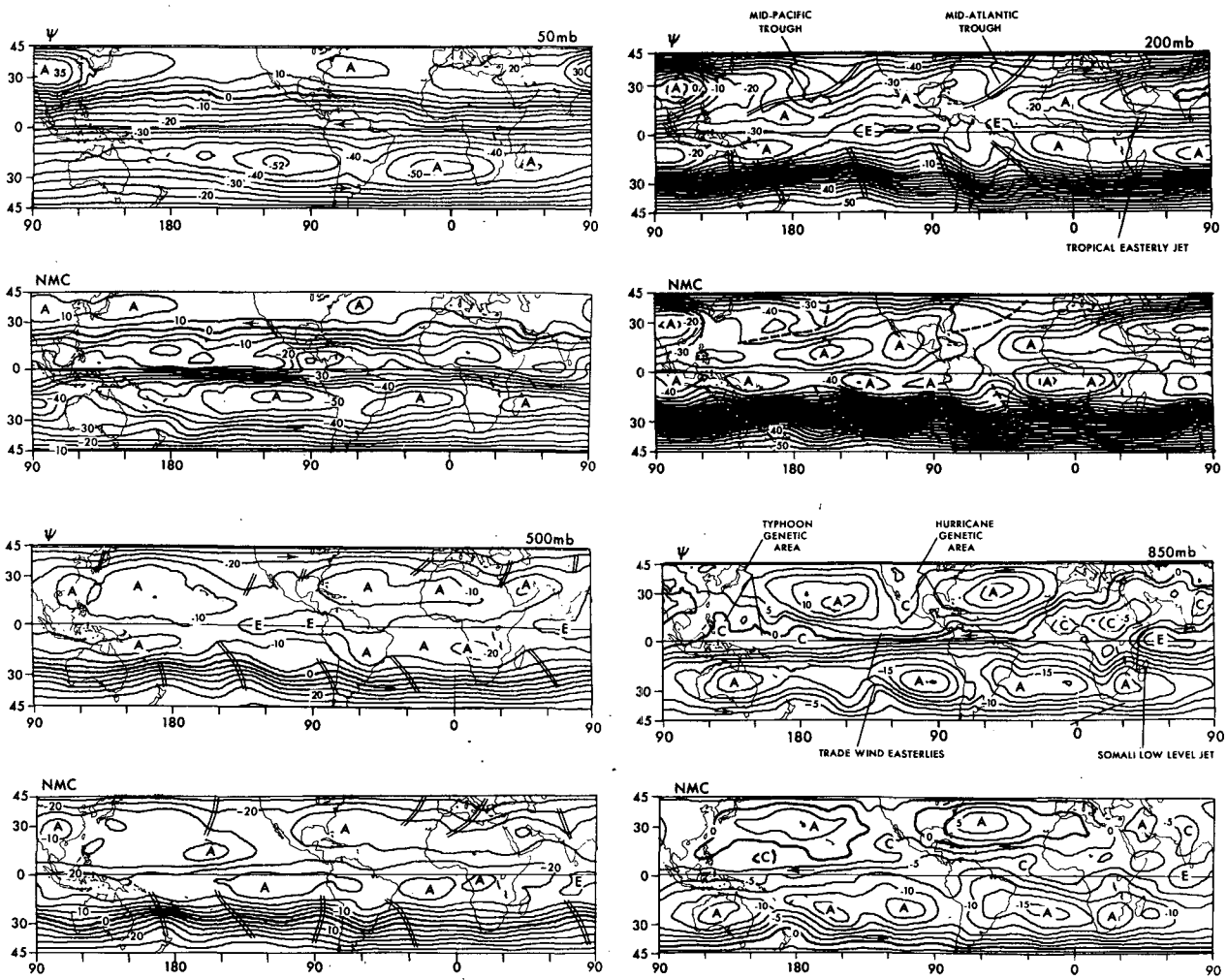

FIG. 13. Streamfunction ψ on a Mercator projection, averaged for 13 days from 4–17 September 1974 for Version 3 and NMC. The contour intervals are 5×10^{10} cm^2 s^{-1} in (a) 50 mb, (b) 200 mb and (c) 500 mb and 2.5×10^{10} cm^2 s^{-1} in (d) 850 mb. A is an anticyclonic center, C a cyclonic center, and E an equatorial circulation center.

the Antarctic and the other is above the 200 mb level. The overall difference in the middle and lower troposphere is 0.6–2°C.

Fig. 17 is the horizontal distribution of the wind difference on Mercator projection maps at the 200 and 850 mb levels. At 200 mb the largest difference may be seen in the following areas: (i) the east equatorial Pacific, (ii) the equatorial Indian Ocean, (iii) the Southern Hemisphere except Australia and New Zealand areas, (iv) China and (v) the North Pacific. These areas are data-sparse regions except China (it was found that a large amount of upper tropospheric reports from this area were lost in the reception process).

On the 850 mb chart, the difference is relatively small, reflecting the small natural variability of wind at this level. Yet some difference is noticeable in areas such as the east equatorial Pacific, equatorial Africa, the equatorial Indian Ocean, and the Phillipine–New Guinea area. It should be stressed that the difference in both levels is small over the Atlantic Ocean due to the special observational efforts.

f. Comparison with the observed wind data

One of the most direct type of verification of an analysis field is that of comparing the analysis with the original observations at stations. To make such a test, the wind vectors observed by radiosonde, rawinsonde and pibal were averaged for a 13-day period from 4–17 September, and were compared with the flow patterns in the four-dimensional analyses averaged over the same period. Fig. 18 is the result of the time-averaged streamlines of Version 3, on which are superposed the observed wind arrows for 200 and 850 mb. As can be seen the wind directions almost exactly agree with the observations.

We show the isotach patterns of Version 3 and the observed wind in Figs. 19 and 20. The comparison of

wind intensity is somewhat difficult, since the observed data are not well distributed spatially. There is overall agreement except that the isotach patterns in the four-dimensional analyses tend to be more smoothly distributed compared with observations, as one would expect.

We now turn our attention to the GATE area and the comparison of the instantaneous wind fields. Fig. 21 is a set of streamline maps of Version 3 over the tropical Atlantic area at 00 GMT 16 September 1974, on which are superposed the plotted data received at NMC in real time. These maps are typical examples of data acquisition, although the cloud-tracked winds, normally given at the 900 mb level, are not included.

One may see from these maps how the original data were used in the four-dimensional analyses, i.e., how consistent or inconsistent the resultant streamlines are to the observed data. At the same time, one may also see how closely or loosely the tropical flow fields are coupled in the vertical. Since this analysis is supposed to be consistent in time, the inertia from flow patterns at earlier times would be appreciable. In other words, information from cloud-tracked winds injected at 18 GMT, for example, is implicitly included in Fig. 21; thus, the analyzed streamlines may or may not be consistent with the observed wind at 00 GMT. In addition, there is another factor which creates deviation of the resultant streamlines from the wind data: the present four-dimensional analyses did not include the effect of diurnal variations.

During the field phase of GATE, synoptic charts were drawn at Dakar to provide synoptic-scale weather

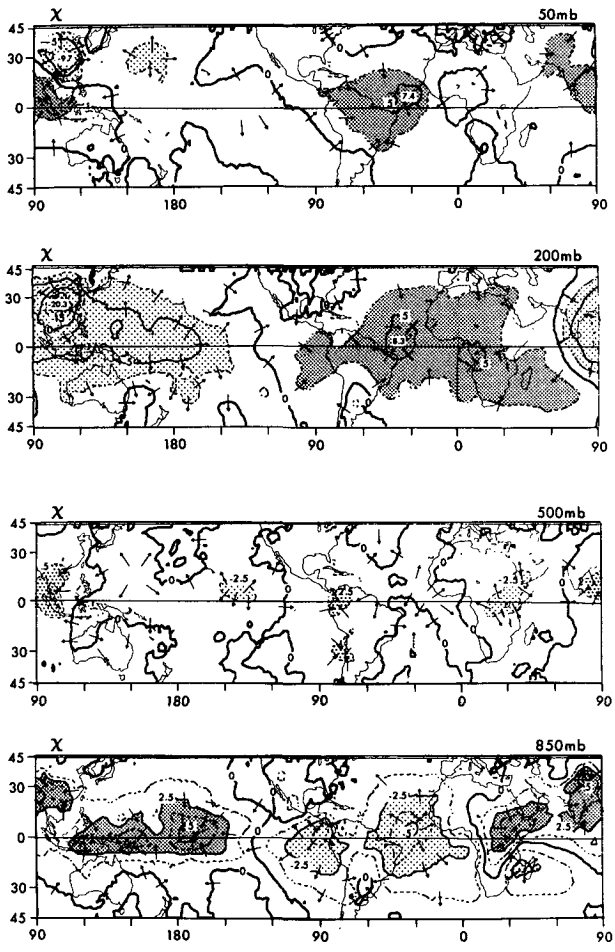

FIG. 14. Velocity potential χ for the same period as in Fig. 12. The contour intervals of the solid and dashed lines are 5×10^{10} cm^2 s^{-1} in (a) 50 mb, (b) 200 mb and (c) 500 mb and 2.5×10^{10} cm^2 s^{-1} in (d) 850 mb. Dashed lines are drawn between the solid lines of 5×10^{10} cm^2 s^{-1} interval. Directions of the divergent component of the wind are indicated. The densely stippled areas show the regions in which χ is larger than 2.5×10^{10} cm^2 s^{-1}, and the sparsely stippled areas are the regions in which χ is less than -2.5×10^{10} cm^2 s^{-1}.

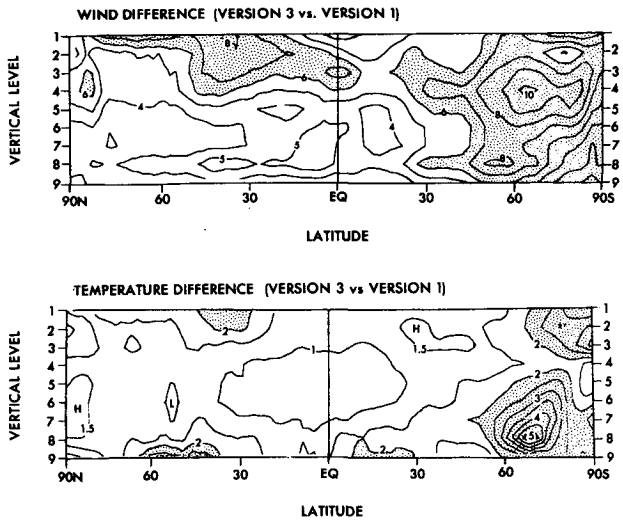

FIG. 15. Latitude-height distribution of the rms difference between Version 3 and Version 1 for (a) wind vector (m s^{-1}) and (b) temperature (°C). Levels 1 through 9 in the ordinate are approximately 9, 74, 188, 336, 500, 664, 811, 926 and 991 mb.

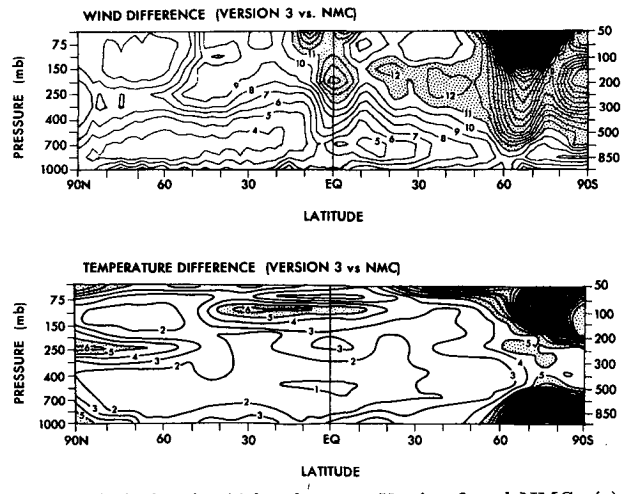

FIG. 16. As in Fig. 14 but between Version 3 and NMC: (a) wind difference; areas in which the difference is greater than 11 m s^{-1} are shaded; (b) temperature difference; areas in which the difference is greater than 5°C are shaded.

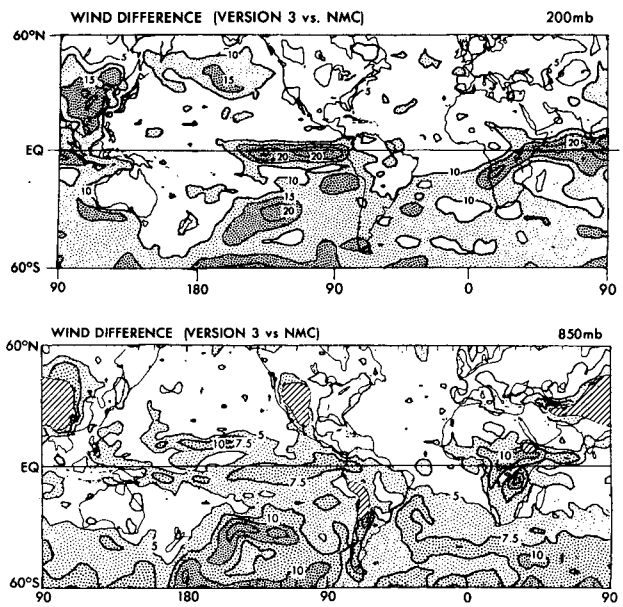

FIG. 17. The rms difference of wind vector (m s^{-1}) between Version 3 and NMC: (a) 200 mb level; areas in which the difference is between 10 and 15 m s^{-1} are sparsely stippled, and areas in which the difference is greater than 15 m s^{-1} are densely stippled; (b) 850 mb level; areas in which the difference is between 5 and 10 m s^{-1} are sparsely stippled, and areas in which the difference is greater than 10 m s^{-1} are densely stippled.

information to the aircraft mission selection; they were subsequently reanalyzed a few days later at Dakar.

The Dakar analyses include pronounced small-scale disturbances, particularly in the 850 and 700 mb charts (the scale is 200–600 km). This type of disturbance is normally called a mesoscale disturbance, and is presumably associated with the activities of cloud clusters. The Dakar analyses tend to be very consistent with the original data, whereas the four-dimensional analyses sometimes ignore some data. This occurs in particular when two pieces of data are close to each other geographically and yet differ in value. This means that the model's resolution of 220 km is too coarse for resolving this scale of disturbance. In addition, so far as this area is concerned, the data acquisition at Dakar was much better than that at NMC. The Dakar maps contain more data not only over the African continent but also over the A/B-scale network. A survey we made for a limited period revealed that only 40–50% of GATE ship data were received at NMC; thus, only that amount of data was used in our analysis.

However, it is also worth mentioning that large-scale features, say greater than 1000 km, are similar in all these analyses. We also compared one of our analyses with an objective analysis produced by Drs. Gilchrist and Graystone's group at Bracknell, England, with

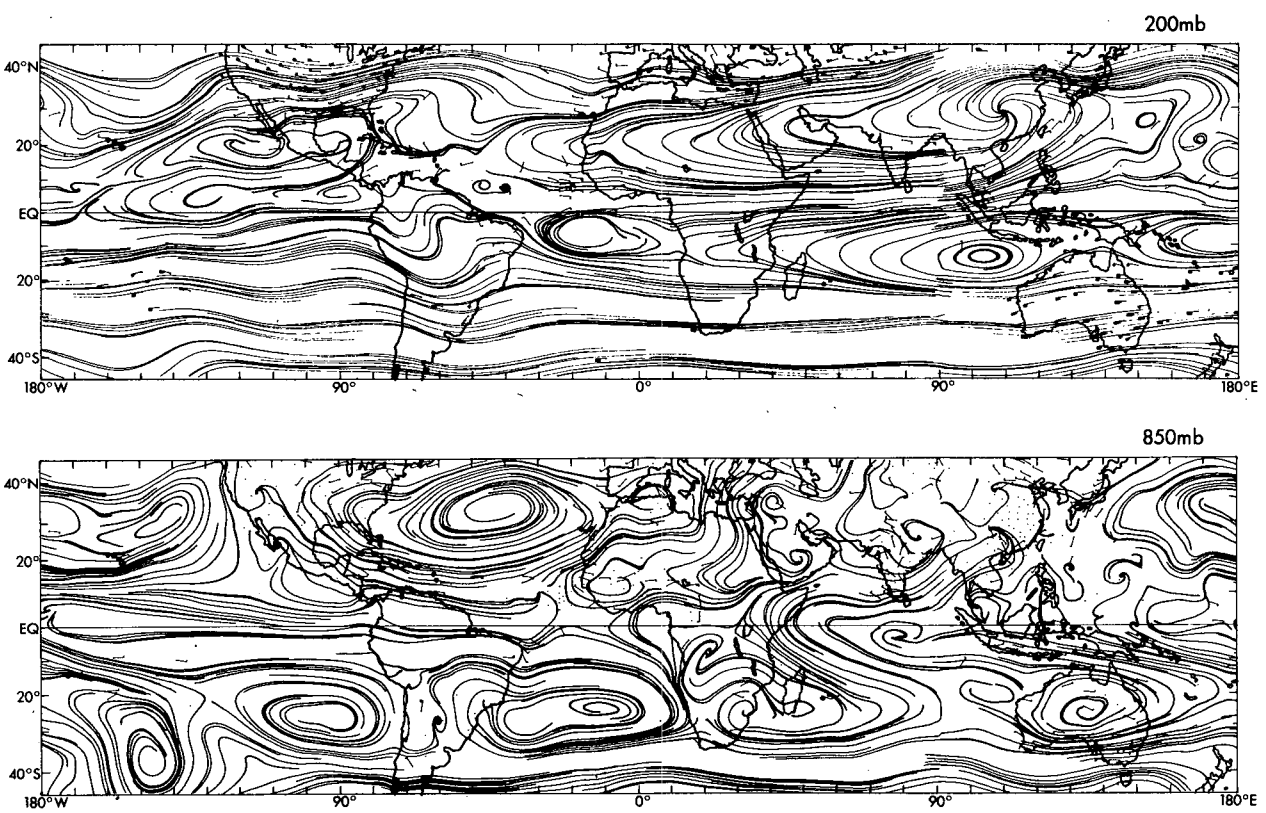

FIG. 18. Averaged wind arrows of radiosonde and rawinsonde observation and the streamlines of Version 3 for the period from 4 to 17 September 1974: (a) 200 mb, (b) 850 mb.

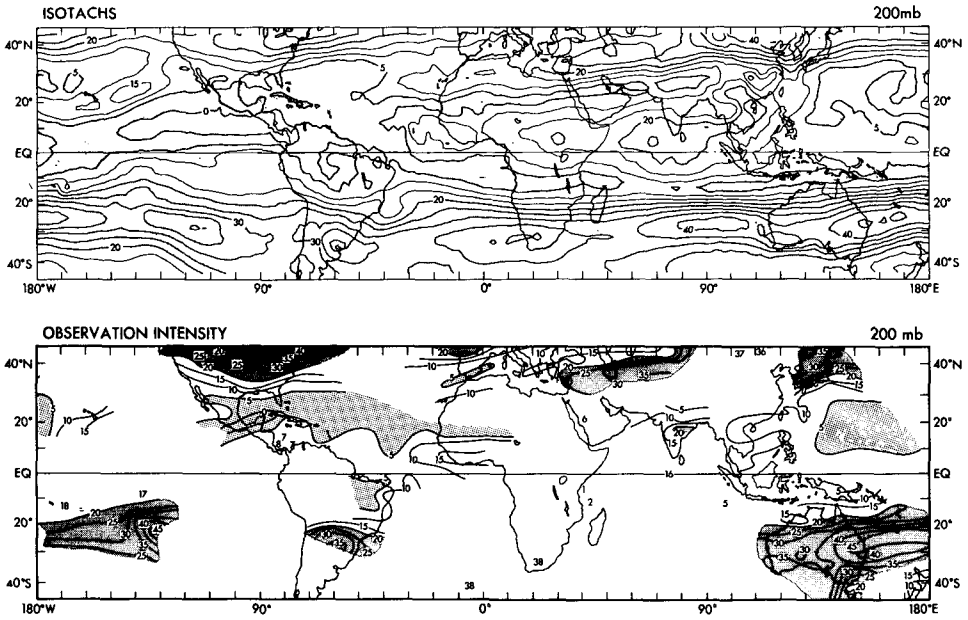

FIG. 19. Isotachs (a) of Version 3 and the averaged wind intensities (b) of radiosonde and rawinsonde observation at 200 mb, for the same period as in Fig. 18.

Sasaki's variational method; there was overall agreement between the two analyses.

This comparison seems to indicate that the tropics apparently contain small-scale disturbances of 200–600 km, but these disturbances can only be resolved by a finer resolution model with a sufficient amount of data. Perhaps the limited-domain model of small grid size, say 40 km, can handle these scales of disturbance with the four-dimensional analysis, though it will be beyond the model's capability to capture the internal structure of the disturbances.

5. Remarks

a. The characteristics of data assimilation

In the past 8 years a number of experiments involving four-dimensional analysis have been performed in many meteorological research centers, and much

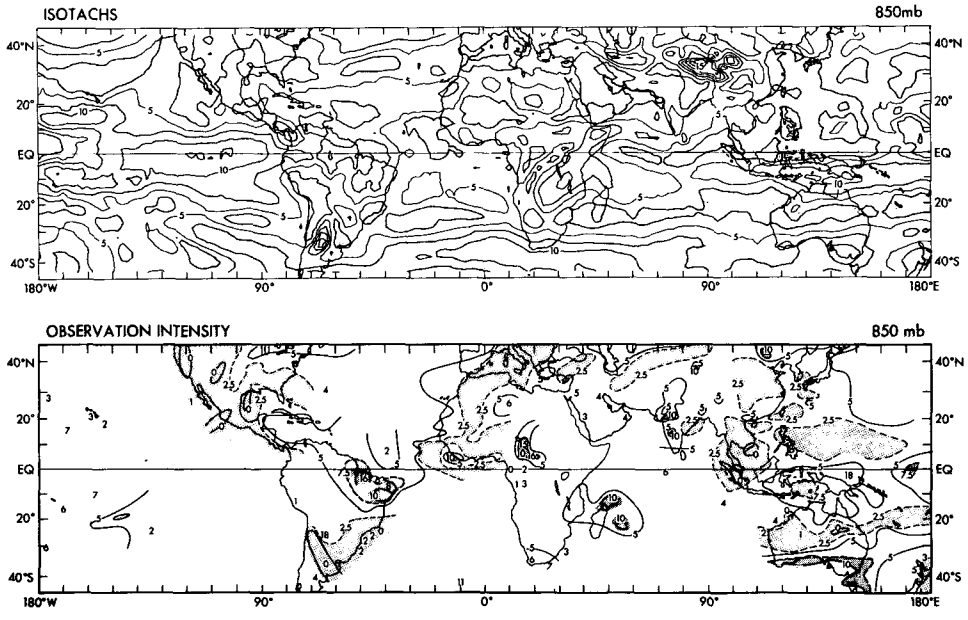

FIG. 20. As in Fig. 19 but for 850 mb.

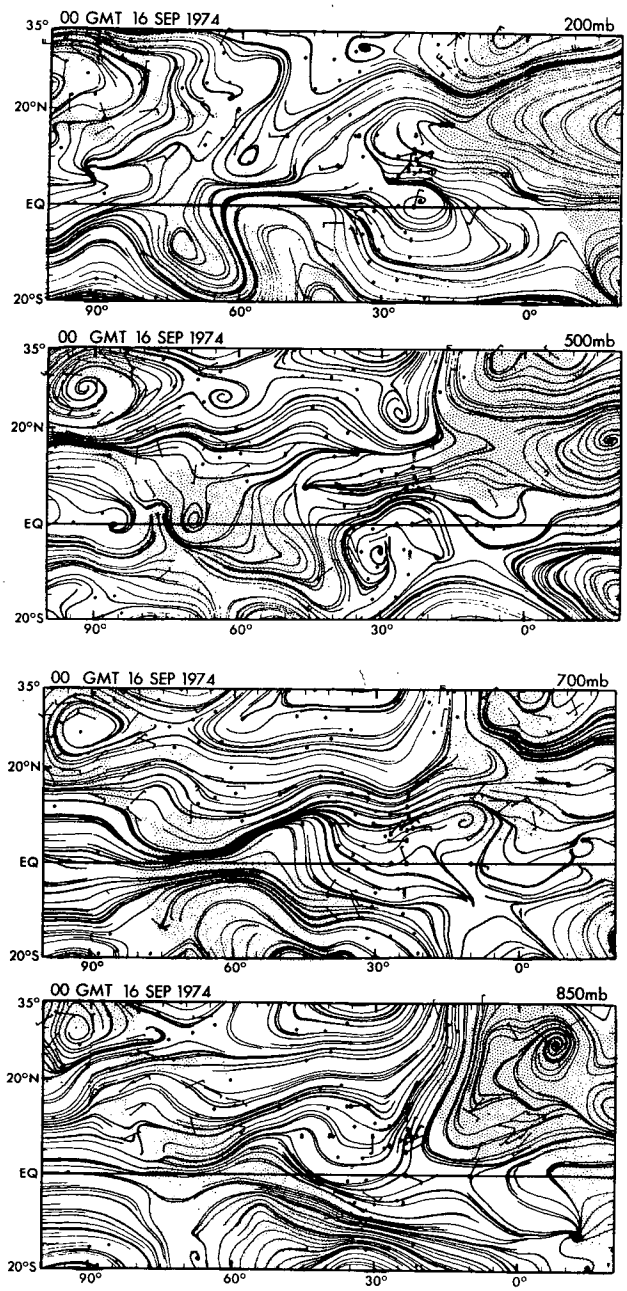

FIG. 21. The received data and the streamlines of Version 3 in the GATE area for 00 GMT, 16 September 1974, at 200, 500, 700 and 850 mb levels. For the wind, one full barb corresponds to 10 m s^{-1}. S at the point of the arrow indicates a ship observation. Temperature observations including the satellite soundings are indicated by small stars.

knowledge has been accumulated (see, for example, Morel and Talagrand, 1974, and the review of Bengtsson, 1975). Reflecting these experiences, let us summarize information obtained in our studies:

1) More data of good quality give better analyses, and the resultant analyses becomes less model-dependent.

2) More data insertion creates more dynamical imbalance (particularly geostrophic) in the model's solutions, and accordingly, causes excitation of gravity waves.

3) Intermittent insertion (forward) gives more shocking, and the analysis turns out to have spuriously smooth patterns for the large-scale flow, though there are many wiggles in the small-scale features.

4) The model can smoothly "phase-in" and "phase-out" a specified value by using continuous insertion in time. For this purpose it is effective to interpolate the observed data to 2 h intervals from 6 or 12 h interval data, and insert the interpolated data into the model continuously.

5) The preliminary application of hydrostatic balance to the insertion data is compulsory.

6) Even without imposing geostrophic balance, wind and temperature data can be entered separately into the model without exciting excessive shocking. [As Bengtsson and Gustavsson (1971) and Hayden (1973) asserted, however, the dynamical adjustment would be faster and the accuracy of the analysis may be increased, if the local geostrophic balance is applied to the insertion data beforehand. But the practical application is difficult. Rutherford (personal communication), Schlatter (1975) and Bergman (personal communication) have proposed the multivariate optimum interpolation scheme, which includes the geostrophic relation to some extent.]

7) Quality control of the insertion is essential. An advantage of a four-dimensional analysis is that a quality check can be done easily and effectively by comparing the insertion data with the model's solution.

It is stressed, however, that the shocking itself should not be a matter of strong concern, unless it is excessive. The wiggles in the flow patterns are not necessarily the manifestation of a poor analysis. What is important is to obtain a reasonable and accurate analysis of the large-scale features. If more data are available, the degree of model-dependence is decreased. Nevertheless, it should be one of the advantages of four-dimensional analysis that it can handle the data gap regions. We feel, moreover, that although the analysis produced by the model in the data-poor regions is not completely accurate, it is dynamically consistent and probably contains more useful information than a sheer guess or the climatological norm.

b. Insertion of surface data

The upper level height patterns in Versions 1 and 2 are almost the same in the mid and high latitudes of the Northern Hemisphere. This point was speculated, to some extent, in the study of redundant variables (Smagorinsky *et al.*, 1970). However, it is also true that for the Southern Hemisphere and particularly the tropics, the resulting analyses in Versions 1, 2 and 3 are

noticeably different. Since these regions have poor data coverage, any data are valuable and the impact on the result is appreciable. This fact was pointed out in numerical experiments of observation system simulation in terms of a "reference level" [NCAR (Kasahara and Williamson, 1972; Baumhefner and Julian, 1972); and GISS (Jastrow, 1972)]. The former reached the conclusion that "the observations of surface pressure over land only are sufficient for updating in the tropics and Northern Hemisphere when compared to observations of surface pressure everywhere." In our case, the situation is somewhat different; even without the surface data over land, the upper level analysis of the Northern Hemisphere is almost the same as NMC's. Concerning the Southern Hemisphere, however, a similar conclusion to that of Kasahara and Williamson (1972) was reached by Gauntlett and Seaman (1974) in their reference-level experiment of the real-data, four-dimensional analysis; if the mean sea level pressure data were not inserted, "the assimilation system failed to maintain the intensity and in certain instances the identity of synoptic systems" in the upper levels.

Jastrow (1972) also mentioned that the surface data for the zonal belt between 50 and 65°S are particularly needed, because the area is mostly covered with clouds which hinders the use of surface temperature determination by the microwave technique.

c. Optimum interpolation

The advantages of this method are (i) to provide a reasonable and therefore accurate interpolation of the original data; (ii) to increase the amount of grid-point insertion data—using optimum interpolation, the total number of insertion data were increased two- or three-fold on the average, so far as the data collection range we used is concerned; (iii) to provide a rational and optimal system to utilize the observed data, incorporating information on the reliability of the data (however, we did not utilize this property in this study); and (iv) despite a large volume of data, to provide relatively less shock-producing data by spreading out in space the observed impulse (Bengtsson and Gustavsson, 1971; Rutherford, 1972).

Our current technique of insertion of the optimally interpolated data is by no means the best method. As is discussed in Appendix B, we encountered a serious problem of error contamination. However, the method has the potential of estimating the error and taking into account the reliability of the data, thereby utilizing the interpolated data with the proper weight according to the error estimate.

For this reason, we presently apply the optimum interpolation only to station data. We plan to modify this arrangement by assigning a weighting factor to each insertion data which reflects its dependability.

6. Conclusions

This experiment revealed that the four-dimensional analysis approach is viable. In particular, Version 3 seemed to give the best results, though the differences among the analyses from the three versions are delicate. For the mid and high latitudes of the Northern Hemisphere, the analysis results appear satisfactory, although the flow patterns are somewhat smoother than the NMC analyses; the final evaluation should await the results of forecasts using the analyses as initial conditions. For the Southern Hemisphere and the tropics, the four-dimensional analysis technique will provide a useful tool for monitoring the evolution of atmospheric disturbances. Tropical cyclones were, to some extent, represented even with a model of about 220 km horizontal resolution, when adequate data were inserted.

Acknowledgments. The authors wish to acknowledge Dr. J. Smagorinsky, who took the leadership of this project and supported the research enthusiastically. The success of an undertaking of this scale required the contributions and cooperation of many people at various locations. They wish to thank Drs. F. Shuman, W. Bonner and A. Desmaris, and his group, who gathered the data at NMC and kindly sent the magnetic tapes to GFDL. They also extend grateful acknowledgment to Drs. D. Sargeant, J. Rasmussen, and Mr. T. Kaneshige of the NOAA GATE Office. They are also indebted to Mrs. Erica Hasslacher, Messrs. J. Welsh, P. Bannon, T. Reek, L. Schroyer, J. Conner, P. Tunison, Dr. G. Cobb, Mrs. B. Williams, Dr. A. H. Oort, Messrs. H. Frazier, F. Uveges, A. Clark, R. Strickler, W. Stern, D. Hembree, I. Shulman, J. Chludzinski, A. Rosati and J. Ryan. The manuscript was checked by Mr. G. Kelly, Drs. T. C. Gordon, I. Simmonds and W. Bonner. Some data computer programs and the level III data for comparison were supplied by the courtesy of Drs. A. Gilchrist, L. Gates, W. Washington, W. Bourke, Mr. R. H. Clarke, Dr. J. Mahlman, Messrs. D. Hahn, G. Kelly and Dr. D. Gauntlett.

APPENDIX A

Data Acquisition

Tables A1 and A2 indicate the statistics for the data received by the Data Assimilation Branch (DAB) of NMC at the 12 h cutoff time through GTS (Global Telecommunication System) during the GATE period. These statistics were calculated at GFDL, using the magnetic tapes mailed from NMC on a near-real-time basis.

The figures in the tables are the mean daily counts of the data obtained for each of the three GATE phases. Figures without parentheses are for the globe and those in parentheses for the GATE area. For example, in Table A1, for radiosonde data at 850 mb, the count

TABLE A1. Average daily counts of global upper air data. Figures in parentheses are for the GATE area.

Data type	Pressure levels (mb) 850	700	500	400	300	250	200	150	100
					Phase 1				
Radiosonde	1414	1378	1343	1334	1263	1210	1182	1141	1103
T	(73)	(73)	(72)	(68)	(67)	(65)	(63)	(63)	(61)
Rawin or pibal	2022	1895	1700	1473	1412	1325	1253	1173	1064
V	(135)	(114)	(80)	(71)	(72)	(66)	(62)	(58)	(55)
Satellite vertical sounding	1686	1686	1686	1686	1686	1686	1686	1686	1686
T	(127)	(127)	(127)	(127)	(127)	(127)	(127)	(127)	(127)
Cloud wind	376	0	0	0	0	0	95	0	0
V	(115)	0	0	0	0	0	(25)	0	0
Aircraft	1	5	14	22	226	647	486	1	0
V, *T*	(0)	(0)	(0)	(1)	(5)	(16)	(4)	(0)	(0)
					Phase 2				
Radiosonde	1486	1438	1394	1381	1303	1241	1212	1169	1135
T	(130)	(132)	(130)	(124)	(121)	(119)	(112)	(110)	(105)
Rawin or pibal	2063	1939	1760	1542	1481	1397	1326	1232	1109
V	(191)	(170)	(133)	(122)	(121)	(113)	(108)	(102)	(95)
Satellite vertical sounding	1638	1638	1638	1638	1638	1638	1638	1638	1638
T	(129)	(129)	(129)	(129)	(129)	(129)	(129)	(129)	(129)
Cloud wind	404	0	0	0	0	128	0	0	0
V	(120)	(0)	(0)	(0)	(0)	(42)	(0)	(0)	(0)
Aircraft	1	4	2	15	221	650	467	0	0
V, *T*	(0)	(0)	(0)	(0)	(4)	(12)	(8)	(0)	(0)
					Phase 3				
Radiosonde	1514	1479	1435	1407	1355	1306	1276	1234	1186
T	(136)	(137)	(135)	(131)	(128)	(126)	(120)	(119)	(111)
Rawin or pibal	2058	1947	1753	1526	1464	1387	1324	1238	1121
V	(182)	(166)	(131)	(120)	(121)	(113)	(109)	(104)	(96)
Satellite vertical sounding	1407	1407	1407	1407	1407	1407	1407	1407	1407
T	(111)	(111)	(111)	(111)	(111)	(111)	(111)	(111)	(111)
Cloud wind	329	1	6	19	72	38	20	1	0
V	(101)	(1)	(2)	(5)	(14)	(17)	(12)	(0)	(0)
Aircraft	1	5	4	29	190	582	411	1	0
V, *T*	(0)	(1)	(0)	(0)	(3)	(14)	(5)	(0)	(0)

Phase 1: 26 June to 16 July.
Phase 2: 28 July to 17 August.
Phase 3: 30 August to 19 September.

is 1414. This is the amount of temperature data per day taken by radiosonde observations, averaged for the 21 days of the GATE Phase 1 in the layer between the surface and 775 mb over the entire globe.

The ATS-3 cloud-tracked wind data were given for most of the GATE period at 200 and 850 mb. However, after 19 August 1974, cloud heights were more precisely specified. Most lower level cloud wind data were then

TABLE A2. Average daily counts of global surface data. Figures in parentheses are for the GATE area.

Phase 1	17 586 (1 542)
Phase 2	17 926 (1 572)
Phase 3	17 591 (1 543)

given at the 900 mb level. Of the total surface data, about 85% are for the Northern Hemisphere. The reception and acquisition of GATE B-scale and A-scale ship data were not satisfactory until the second week of August, and only a few GATE ship data (less than 50%) were available throughout the entire GATE period. [See the data acquisition at Bracknell: Spackman and Farmer (1975).]

It should be noted that similar tables were presented previously for a different period, i.e., March and April, 1973, by Bonner and the Staff of DAB. In that case, the counts of the data were given at 00 and 12 GMT, separately, whereas here the counts are for a one day total. It appears that the radiosonde data reception is considerably greater during the GATE period, compared with the case of 1973.

APPENDIX B

Optimum Interpolation for Obtaining Insertion Data

In order to determine the insertion data at a grid point from the surrounding observed data, two methods were used: *direct replacement* and *optimum interpolation*. The former method is simply to employ the nearest observation [the station A in (a) of Fig. B1] and to use this as the insertion data at grid point 0. The latter method is to use several observation stations, if available, and to interpolate an optimal value at the point 0 from those data. It is the main purpose of this appendix to discuss only the latter. It is recommended that reference be made to Gandin (1963) for the basic theory and further details.

Consider a parameter—temperature T, for example. To determine T at the grid point 0, we define a small circular domain around the point with radius R [see (b) of Fig. B1] and look for observations within the domain. If no observation is available, there is no insertion data. On the other hand, if there are observations, the value at the point 0 is interpolated using the formula

$$T_0'=\sum_{i=1}^{N} p_i T_i', \tag{B1}$$

where

$$T_i'=T_i-\tilde{T}. \tag{B2}$$

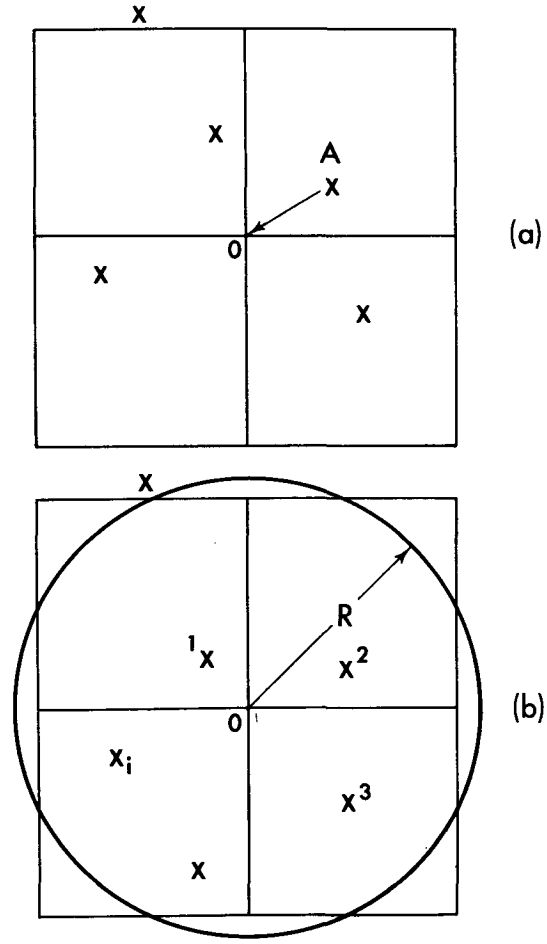

FIG. B1. Direct replacement (a) and optimum interpolation (b).

Here $\tilde{T}$ is the climatological norm, T_0 the value at the point 0, T_i the value at the station i, and the primed variables are the deviations from the climatological norm; p_i is the weighting coefficient for the data at the station i; and N is the total number of stations inside the domain of radius R.

According to theory, p_i is chosen so as to minimize the mean-square error, and in fact, p_i is determined by solving the simultaneous algebraic equations

$$\sum_{i=1}^{N} p_i\mu_{ij}=\mu_{0j}, \quad j=1, \ 2, \ \ldots, N, \tag{B3}$$

where μ_{ij} is the correlation coefficient, such that

$$\mu_{ij}=\frac{\overline{T_i' T_j'}}{[\overline{T_i'^2}\,\overline{T_j'^2}]^{\frac{1}{2}}}, \tag{B4}$$

the bar indicating the climatological mean.

In practice, local isotropy and homogeneity were assumed for the correlation function μ_{ij}; it is dependent on the distance between two stations, i and j, i.e., $\mu_{ij}=\mu(r)$, r being the separation distance. The func-

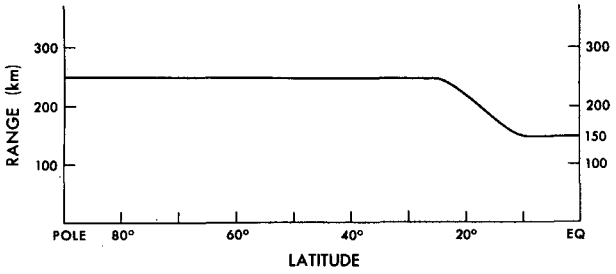

FIG. B2. Radius of the data influence as a function of latitude.

tional forms of $\mu(r)$ are different for temperature, wind and water vapor, but the same for surface pressure and temperature. We used the formulas proposed by Gandin (1963). The climatological norms were obtained from the monthly means of a two-dimensional objective analysis. They are functions of the three-dimensional space coordinates.

In order to solve the Eqs. (B3), a matrix method was used. The data used for this calculation were always limited to no more than 8 observations, i.e., $N \leqslant 8$, as was suggested by Gandin. It is important to filter out the observations which are spatially too close to each other; otherwise, the matrix calculation turns out to be numerically ill-conditioned. Optimum interpolation was carried out for all parameters at all 11 selected pressure levels at 6 h intervals. All observations except the data discarded in the preceding quality control were utilized.

In our preliminary test of the four-dimensional analysis, we encountered an interesting problem related to optimum interpolation, which is worthwhile to report here. If a radius of 500 or 300 km everywhere was used, the four-dimensional analysis (in this test, not only station data but also data from other sources were used) appears unreasonable, especially in the tropics. This problem seems to appear for two different reasons.

In mid and high latitudes, large errors in raw data which were not discarded in the preceding quality control were spread in space by optimum interpolation and caused serious damage to the analysis. On the other hand, in the versions of data processing that used "direct replacement," the trouble did not occur, probably because the erroneous data were sporadic and, therefore, were safely removed in the final quality check of comparing the insertion data with the model's solutions. To avoid this trouble, it seems necessary to adopt a smaller radius R, unless such bad data can be discarded earlier in the processing.

In the tropics, unrealistic spotty patterns in the flow fields were produced, indicating that a representative domain of a single datum taken was too large. In other words, in our test studies, a single datum was excessively stressed by being spread out spatially, even though it was correct. In view of the fact that disturbances in the tropics have smaller space scales, it is speculated that this trouble was caused by employing inadequate correlation functions. (We used the functions for mid latitude everywhere.) This problem had already been raised by Gandin (personal communication; see Kolugina and Kartasheva, 1969), and Alaka and Elvander (1972); the correlation function varies appreciably with latitude. [Gordon and Lusen (1976) also noticed the localized influence of tropical observations in the study of the observation system simulation for the GATE network.] In this particular experiment, instead of using a variable correlation function, we used, as an ad hoc arrangement, a different radius R, depending on latitude, as shown in Fig. B2. R is 250 km from 25° poleward, 150 km in the equatorial region, and smoothly connected in between. This value may seem to be extremely small, since the correlation curve of temperature, for example, becomes zero at 1500 km distance. However, this is our first trial, and as mentioned in the main text, yields of insertion data have been increased by 2.4 times even with this small range of R, causing a large "shocking" to the assimilation.

Finally, it is noted that the idea of using optimum interpolation for the rational design of observing networks was first presented by Gandin (1961). Also, the method has already been used for dynamical analysis in past studies (e.g., Miyakoda and Talagrand, 1971: Talagrand and Miyakoda, 1971; Bengtsson and Gustavsson, 1971; Rutherford 1972), and has more recently been extensively discussed from somewhat different points of view by Schlatter (1975) and Bengtsson (1975). The question remains, however, as to whether optimum interpolation is indeed significantly superior to all other interpolation schemes.

REFERENCES

Alaka, M. A., and R. C. Elvander, 1972: Optimum interpolation from observations of mixed quality. *Mon. Wea. Rev.*, **100**, 612–624.

Baumhefner, D., and P. Julian, 1972: The reference-level problem: Its location and use in numerical weather predictions. *J. Atmos. Sci.*, **29**, 285–299.

——, and N. Gustavsson, 1971: An experiment in the assimilation of data in dynamical analysis. *Tellus*, **23**, 328–336.

Bengtsson, L., 1975: Four-dimensional data assimilation of meteorological observations. WMO/ICSU,GARP Publ. Ser. No. 15, 76 pp. [Secretariat of the World Meteorological Organization, Geneva, Switzerland].

——, and N. Gustavsson, 1971: An experiment in the assimilation of data in dynamical analysis. *Tellus*, **23**, 328–336.

Charney, J., M. Halem and R. Jastrow, 1969: Use of incomplete historical data to infer the present state of the atmosphere. *J. Atmos. Sci.*, **26**, 1160–1163.

Eddy, A., 1967: The statistical objective analysis of scalar data fields. *J. Appl. Meteor.*, **6**, 597–609.

Flattery, Thomas W., 1970: Spectral models for global analysis and forecasting. *Proc. Sixth AWS Technical Exchange Conf.*, U. S. Naval Academy, 21–24 September, Air Weather Service Tech. Report 242, 42–54.

Gandin, L. S., 1961: On the principles of rational distribution of meteorological stations. *Tr. Gl. Geofiz. Observ.*, No. 3.

——, 1963: *Objective Analysis of Meteorological Fields*. Gidrometeor. Izdatel, Leningrad [Israel Program for Scientific Translations, Jerusalem, 1968, 242 pp.].

GARP, 1967: The global atmospheric research programme. Report of the Study Conference, Stockholm, 28 June–11 July, ICSU/IUGG and WMO [Secretariat of the World Meteorological Organization, Geneva, Switzerland].

——, 1972: Experiment design proposal for the GARP Atlantic Tropical Experiment. WMO/ICSU, GATE Report No. 1, 188 pp. [Secretariat of the World Meteorological Organization, Geneva, Switzerland].

Gauntlett, D. J., and R. S. Seaman, 1974: Four-dimensional data assimilation experiments in the Southern Hemisphere. *J. Appl. Meteor.*, **13**, 845–853.

Gordon, C. T., and R. Lusen, 1976: Observing systems simulation experiments relevant to GATE (submitted for publication, *Mon. Wea. Rev.*).

——, L. Umscheid, Jr. and K. Miyakoda, 1972: Simulation experiments for determining wind data requirements in the tropics. *J. Atmos. Sci.*, **29**, 1064–1075.

Hayden, C., 1973: Experiments in the four-dimensional assimilation of Nimbus-4 SIRS data. *J. Appl. Meteor.*, **12**, 425–436.

Houghton, D. D., and D. E. Parker, 1974: The synoptic-scale sub-program for the GARP Atlantic Tropical Experiment. WMO/ICSU, GATE Report No. 6 [Secretariat of the World Meteorological Organization, Geneva, Switzerland].

Jastrow, R., 1972: Observing system simulation studies. Report of the seventh session of the Joint Organizing Committee GARP, WMO/ICSU [Secretariat of the World Meteorological Organization, Geneva, Switzerland].

——, and M. Halem, 1970: Simulation studies related to GARP. *Bull. Amer. Meteor. Soc.*, **51**, 490–513.

Kasahara, A., 1972: Simulation experiments for meteorological observing systems for GARP. *Bull. Amer. Meteor. Soc.*, **53**, 252–264.

—— and D. Williamson, 1972: Evaluation of tropical wind and reference pressure measurements: Numerical experiments for observing systems. *Tellus*, **24**, 100–115.

Kelly, G. A. M., 1975: Interpretation of satellite cloud mosaics for Southern Hemispheric analysis and reference level specification (submitted for publication, *Mon. Wea. Rev.*).

Kolugina, T. P., and M. V. Kartasheva, 1969: On the statistical characteristics of wind field in the equatorial zone and middle latitudes. *Nauch. Issled. Inst. Gidrometeor. Prib.*, No. 39.

Krishnamurti, T. N., 1971a: Observational study of the tropical upper tropospheric motion field during the Northern Hemisphere summer. *J. Appl. Meteor.*, **10**, 1066–1096.

——, 1971b: Tropical east–west circulations during the northern summer. *J. Atmos. Sci.*, **28**, 1342–1347.

Kurihara, Y., 1965: Numerical integration of the primitive equations on a spherical grid. *Mon. Wea. Rev.*, **93**, 399–415.

—— and J. L. Holloway, Jr., 1967: Numerical integration of a nine-level global primitive equations model formulated by the box method. *Mon. Wea. Rev.*, **95**, 509–530.

Manabe, S., J. L. Holloway, Jr. and H. M. Stone, 1970: Tropical circulation in a time-integration of a global model of the atmosphere. *J. Atmos. Sci.*, **27**, 580–613.

——, D. G. Hahn, and J. L. Holloway, Jr., 1974: The seasonal variation of the tropical circulation as simulated by a global model of the atmosphere. *J. Atmos. Sci.*, **31**, 43–83.

Matsuno, T., 1966: Numerical integrations of primitive equations by use of a simulated backward difference method. *J. Meteor. Soc. Japan*, **44**, 76–84.

Miyakoda, K., and O. Talagrand, 1971: The assimilation of past data in dynamical analysis. I. *Tellus*, **23**, 310–317.

——, R. W. Moyer, H. Stambler, R. H. Clarke and R. F. Strickler, 1971: A prediction experiment with a global model of the Kurihara-grid. *J. Meteor. Soc. Japan*, **49**, Special Issue, 521–536.

——, J. C. Sadler and G. D. Hembree, 1974: An experimental prediction of the tropical atmosphere for the case of March 1965. *Mon. Wea. Rev.*, **102**, 571–591.

Morel, P., and O. Talagrand, 1974: Dynamic approach to meteorological data assimilation. *Tellus*, **26**, 334–343.

——, G. Lefevre and G. Rabreau, 1971: On initialization and nonsynoptic data assimilation. *Tellus*, **23**, 197–206.

Newell, R. E., J. W. Kidson, D. G. Vincent and G. J. Boer, 1972: *The General Circulation of the Tropical Atmosphere and Interaction with Extratropical Latitudes*, Vol. 1. The MIT Press, 248 pp.

Oort, A. H., and E. M. Rasmusson, 1971: Atmospheric Circulation Statistics. NOAA Prof. Paper No. 5, U. S. Dept. of Commerce, 323 pp. [NTIS Order No. COM-72-50295].

Petersen, D. P., and D. Middleton, 1963: On representative observations. *Tellus*, **15**, 387–405.

Rutherford, I. D., 1972: Data assimilation by statistical interpolation of forecast error fields. *J. Atmos. Sci.*, **29**, 809–815.

——, and R. Asselin, 1972: Adjustment of the wind field to geopotential data in a primitive equations model. *J. Atmos. Sci.*, **29**, 1059–1063.

Sadler, J. C., and B. E. Harris, 1970: The mean tropospheric circulation and cloudiness over Southeast Asia and neighboring areas. Sci. Rept. No. 1, Hawaii Institute of Geophysics.

Schlatter, T. W., 1975: Some experiments with a multivariate statistical objective analysis scheme. *Mon. Wea. Rev.*, **103**, 246–257.

Smagorinsky, J., K. Miyakoda and R. F. Strickler, 1970: The relative importance of variables in initial conditions for dynamical weather prediction. *Tellus*, **22**, 141–157.

Spackman, E. A., and S. F. G. Farmer, 1975: Synoptic-scale data received at Bracknell during GATE. Meteor. Office 20, Tech. Note No. 36.

Stackpole, J. D., L. W. Vanderman and F. G. Shuman, 1974: The NMC 8-layer global primitive equation model on a latitude-longitude grid. Modelling for the First GARP Global Experiment, WMO/ICSU, GARP Publ. Ser., No. 14, 79–93 [Secretariat of the World Meteorological Organization, Geneva, Switzerland].

Talagrand, O., and K. Miyakoda, 1971: The assimilation of past data in dynamical analysis. II. *Tellus*, **23**, 318–327.

Thompson, T., 1972: The Basic Data Set Project. WMO/ICSU, GARP Publ. Ser., No. 9, 99 pp. [Secretariat of the World Meteorological Organization, Geneva, Switzerland].

Tripoli, G. J., and T. N. Krishnamurti, 1975: Low-level flows over the GATE area during summer 1972. *Mon. Wea. Rev.*, **103**, 197–216.

Umscheid, L., Jr., and M. Sankar-Rao, 1971: Further tests of a grid system for global numerical prediction. *Mon. Wea. Rev.*, **99**, 686–690.

Williamson, D., and A. Kasahara, 1971: Adaptation of meteorological variables forced by updating. *J. Atmos. Sci.*, **28**, 1313–1324.