Monsoon forecasting on parallel computers

U. N. Sinha, V. R. Sarasamma, S. Rajalakshmy, K. R. Subramanian, P. V. R. Bharadwaj, C. S. Chandrashekar, T. N. Venkatesh, R. Sunder, B. K. Basu*, Sulochana Gadgil** and A. Raju**

Flosolver Unit, National Aerospace Laboratories, Bangalore 560 017, India
*National Centre for Medium Range Weather Forecasting, Department of Science and Technology, Technology Bhavan, New Delhi 110 016, India
**Centre for Atmospheric Sciences, Indian Institute of Science, Bangalore 560 012, India

The Global Circulation Model T-80 of the National Centre for Medium Range Weather Forecasting (NCMRWF), DST, which is currently running on their Cray XMP/216, has been parallelized on the Flosolver Mk3. The Flosolver output agrees very well with the Cray output. Post-processing has also been completed and a one-day forecast takes approximately an hour and a half on the 4-processor Flosolver Mk3 whereas the 2-processor Cray machine takes about 15 minutes. By increasing the number of processors, rewriting the code for the i860 processor (the code in its present version is very Cray-specific) and making some minor modification in the CPU card, it shall be possible to significantly improve the time taken for weather forcasting. The present effort shows that it is possible to do research on T-80 or similar models on a Flosolver-type of machine requiring only a modest financial outlay.

WEATHER and climate have fascinated sc . ntists, statesmen and poets alike since time immemorial, and systematic observations of the meteorological elements with modern instruments have been made for over a century. However, major advances in our understanding of the dynamics of the global atmosphere have occurred primarily in the last fifty years. A series of papers in the late forties and early fifties led to the elucidation of the basic physics of the general circulation of the atmosphere. Jule Charney showed that the complex response of the atmosphere to the heating by sun's radiation is a manifestation of an instability of a rotating fluid with latitudinally varying temperature induced by the solar insolation: the so-called baroclinic instability. (It turns out that almost all the important phenomena in the atmosphere – including the cumulus cloud which gives us rain – arise from different kinds of instabilities. This is what makes the problem of prediction of weather and climate intrinsically more difficult than prediction of planetary orbits for example.) With the identification of the important mechanisms for large scale circulation of the global atmosphere it became possible to overcome the problem faced by Richardson[1] in his first numerical integration of the equations governing the atmosphere (due to the unrealistic dominance by inertio-gravity waves) and simplify the highly nonlinear governing equations by filtering out these waves. Around this time, John von Neumann wanted to use the problem of meteorological forecasting by numerical processes as a test of the newly developed electronic computer and formulated the Princeton Meteorology project. At the outset he called a meeting at Princeton in 1946 to gain the support of the meteorological community for the computer that the Princeton computer group could produce[2]. The first atmospheric models which were integrated numerically on the electronic computer for generation of meteorological predictions were developed by the efforts of the pioneers such as Charney, von Neumann and Phillips at Princeton in the early fifties. The development of general circulation models progressed rapidly thereafter. However the early models could generate reasonable predictions only for the mid-latitudinal atmosphere for which the major mechanisms, viz. the baroclinic instability had already been understood. The tropics remained an enigma.

The last three decades have witnessed tremendous advances in our understanding of the atmospheric circulation and convection of the tropical regions in general and the monsoons in particular. The weather-element of greatest interest in the tropics is of course the rainfall. With the advent of satellites in the mid-sixties, it became possible to literally see the genesis and propagation of cloud-bands/rainbelts on a day-to-day basis on the global scale – including over the traditionally data-sparse oceans – over which most of the convective systems are born. Secondly, the nature of the multiple interactions leading to the organization of cumulus clouds (typically a few kilometers in extent) in low-depressions which are several hundred kilometers in extent was elucidated by theoretical studies. The development of general circulation models incorporating these complex interactions progressed rapidly in the last two decades in parallel with the computer revolution.

The progress on all these fronts led to the development of rather realistic models of the global atmosphere by the early eighties. The early models could generate reasonable predictions only over the short range of 24–48 hours and primarily over the mid-latitudes. Now prediction over the medium range of 3–10 days (i.e.

almost up to the Lorenzian limit of predictability, viz. the time-scale over which two solutions with a small difference in initial conditions become as different as two random solutions due to the growth of instabilities) over the entire globe came within the realm of possibility. The success of the European Centre for Medium Range Forecasting set up in the early seventies in the very first decade, triggered the setting up of the National Centre for Medium Range Weather Forecasting (NCMRWF) at Delhi for generating the all-important predictions of the monsoon. This Centre was equipped with the first supercomputer in the country to run atmospheric general circulation models. The general circulation model developed at one of the leading centres in the world, viz. the National Meteorological Centre in US will shortly be made operational in the Centre.

The tasks ahead will be to assess and improve its performance in generating rainfall predictions over the Indian region on the one hand and keep ensuring computer facilities of adequate power for the increased demands of the modified model on the other. The first is being addressed by Indian meteorologists, particularly at the NCMRWF. To get a feel for the computer requirements, it would suffice to say that the sustained speed of 100 MFLOPS will be termed moderate, random access memory (RAM) availability of 100 MBytes will be called manageable and a disk storage of tens of GBytes will be considered marginal for an operational forecast. These requirements could be met either by importing computers of higher power or alternatively by harnessing the indigenous effort in parallel computers. The results of an exercise reported here appear to be sufficiently encouraging to work towards meeting the demands of medium range forecasting by the second mode if once again the progress in computers (this time parallel) and atmospheric models go hand in hand with positive feedbacks to each other. A perspective of this effort is described in the next section.

Initiative of parallel computing for weather calculation in India

Considering that the technology of parallel computing in the last decade has reached a state of maturity and holds the promise of being an alternative to conventional supercomputing, on 23 November 1992 a brain-storming session on 'Future Supercomputing Strategy for Medium Range Weather Forecasting'[3] was convened to explore this alternative and the summary record described the consensus in the following words:

'The consensus emerging from the session was that parallel processing systems have not yet become proven tools for operational medium range weather forecasting within the country or abroad. However R&D needs to be fostered within the country. It would therefore, be desirable to continue to provide support for this activity with which NCMRWF should be actively associated, but without hampering its goal of operational medium range weather forecasting'.

Subsequently, on 14 June 1993, the initial phase of the project 'Implementation of GCM T-80 Model (Global Circulation Model, T-80 version) on Flosolver' was formally approved[4]. The present paper describes the parallelization of the above mentioned code on the 4-processor Flosolver Mk3, one of the series of parallel computers designed and built at National Aerospace Laboratories (NAL).

Brief description of the model

All atmospheric prediction models solve a set of nonlinear partial differential equations to predict the future state of the atmosphere from a given initial state. Since the domain of atmospheric flow is bounded at the bottom by the surface of the earth, exchange of properties takes place at this surface and it is necessary to prescribe appropriate boundary conditions or values for various quantities. The bottom topography plays an important role in controlling the airflow not only close to the ground but also at upper levels through induced vertical motion and momentum transfer by gravity waves. Present day atmospheric models have moisture as one of the variables and take into account diabatic processes like evaporation and condensation. All physical processes involving moisture and others like radiation, turbulence, gravity wave drag, land surface processes, etc. are parametrized in terms of the variables in the model. The governing equations for the present model are discussed below.

The model is based on the conservation laws for mass, momentum, energy and moisture. The momentum equations are replaced by the vorticity and divergence equations so that the spectral technique can be applied in the horizontal direction. The vertical co-ordinate is $\sigma = p/p_*$, where p is the layer pressure and p_* the surface pressure. Finite differences are used to approximate the differential operators in the vertical direction. The governing equations are cast in the form of evolution equations for the model variables temperature (T), surface pressure (p_*), specific humidity (q), divergence (D) and vorticity (η). The model prognostic equations are given below.

The thermodynamic equation is

$$\frac{\mathrm{d}\ln\theta}{\mathrm{d}t} = \frac{H}{C_p T},$$

where H is the heating rate per unit mass and θ is the potential temperature. This is rewritten to give the following equation for temperature

$$\frac{\partial T}{\partial t} = -\vec{V}\cdot\nabla T + \kappa T\left(\frac{\partial}{\partial T} + \vec{V}\cdot\nabla\right)\times$$

$$\ln p_* + \frac{H}{C_p} - \pi\dot{\sigma}\frac{\partial}{\partial\sigma}(T/\pi), \qquad (1)$$

where $T = \pi\theta$, $\pi = p_*^{\kappa}$, $\kappa = R/C_p$ and ∇ is the horizontal gradient in the system.
The continuity equation is

$$\frac{\partial \ln p_*}{\partial t} + \vec{V}\cdot\nabla\ln p_* + \nabla\cdot\vec{V} + \frac{\partial\dot{\sigma}}{\partial\sigma} = 0$$

On integrating this equation with the boundary conditions

$$\dot{\sigma}(0) = \dot{\sigma}(1) = 0,$$

we get the equation for surface pressure,

$$\frac{\partial \ln p_*}{\partial t} = -\int_0^1 (\nabla\cdot\vec{V} + \vec{V}\cdot\nabla\ln p_*)\, d\sigma \qquad (2)$$

The conservation law for moisture is

$$\frac{dq}{dt} = S,$$

where q is the specific humidity and S represents the sources and sinks. This equation is written in the following form.

$$\frac{\partial q}{\partial t} = -\vec{V}\cdot\nabla q - \dot{\sigma}\frac{\partial q}{\partial\sigma} + S. \qquad (3)$$

The equations for divergence and vorticity are obtained from the momentum equation by taking the dot and cross products respectively.
The equation for the divergence D is

$$\frac{\partial D}{\partial t} = \frac{1}{a\cos^2\varphi}\left(\frac{\partial B}{\partial\lambda} - \cos\varphi\frac{\partial A}{\partial\varphi}\right) -$$

$$\nabla^2(E + \Phi + RT_0\ln p_*) \qquad (4)$$

and the equation for the vorticity is

$$\frac{\partial\eta}{\partial t} = \frac{-1}{a\cos^2\varphi}\left(\frac{\partial A}{\partial\lambda} + \cos\varphi\frac{\partial B}{\partial\varphi}\right), \qquad (5)$$

where $D = \nabla\cdot\vec{V}$; $\eta = f + \zeta_k$; $\zeta_k = \vec{k}\cdot\nabla\times\vec{V}$; $A = \eta U + (RT/a)\cos\varphi\,(\partial\ln p_*/\partial\varphi) + \dot{\sigma}(\partial V/\partial\sigma) - \cos\varphi F_\varphi$; $B = \eta V - (RT/a)(\partial\ln p_*/\partial\lambda) - \dot{\sigma}(\partial U/\partial\sigma) + \cos\varphi F_\lambda$; $U = u\cos\varphi$; $V = v\cos\varphi$; $E = \vec{V}\cdot\vec{V}/2$; Φ = geopotential height; f = Coriolis parameter; a = radius of the earth; φ = latitude; λ = longitude; F_φ = zonal component of the dissipative processes; F_λ = meridional component of the dissipative processes.

The classical method of solving nonlinear differential equations is to approximate the derivatives by finite differences and to solve the resultant difference equations by appropriate numerical methods. Such a method has, however, well-known inherent deficiencies. Since errors in finite difference atmospheric models are mostly from the approximation of the spatial derivatives, the spectral technique was formulated in which the horizontal variations of the atmospheric variables are approximated by a truncated series expansion in orthogonal functions. The spatial derivatives can now be evaluated exactly by analytic differentiation but with an accuracy limited by the number of terms in the truncated series expansion (horizontal resolution). The choice of the orthogonal functions has now become standard and almost all centres working on atmospheric modelling use spherical harmonics as the orthogonal basis. These functions are solutions of the horizontal part of the Laplace equation over a sphere.

In the spectral method, the nonlinear terms can be expressed in terms of a truncated series of spherical harmonics. The coefficients of this series, known as 'Interaction Coefficients' are too expensive to be computed online and too many in number to be stored in the memory. Hence, the alternate 'transform method' is used to compute the nonlinear terms in which the field variables are brought from the spectral space to a discrete latitude/longitude grid and the nonlinear products are formed before going back to the spectral space. The Legendre and Fast Fourier Transforms are used to move from one space to another.

The spectral model used in the present study has a horizontal resolution of 80 waves in triangular truncation and 18 atmospheric layers in the vertical. The Fourier space again has 80 waves in the E-W on 128 quasi-equidistant Gaussian latitudes which exclude both the equator and the poles. In the physical grid space the Fourier waves are projected onto 256 equispaced longitude points on each Gaussian latitude. The number of longitude points has to be much larger than the azimuthal wave number in order to resolve the shorter waves produced by nonlinear interaction. The forecast model is fully diabatic and includes parametrization of all known physical processes like radiation, convection, turbulent boundary layer processes, land–air interaction, surface friction, etc. The salient technical features of the model are included in Table 1 and mathematical and numerical details are described in ref. 5.

Parallelizing strategy

The salient features of Flosolver Mk3 on which the code is parallelized are described below.

Table 1. Model description

Dynamics	
Horizontal resolution	Spectral 80 waves in triangular truncation
Vertical resolution	18 layers in sigma coordinates
Gaussian grid	256 equally spaced points on each of the 128 quasi-equally spaced Gaussian latitudes
Time-step	15 minutes
Time integration scheme	Semi-implicit for divergence, logarithm of surface pressure and temperature
	Explicit with semi-implicit correction for vorticity and specific humidity
Diffusion	Wave number dependent Laplacian formulation. No diffusion up to wave number 44
Topography	Mean topography derived from US Navy 10-min resolution field
Gravity wave drag	Surface flux parametrization of Pierrehumbert and vertical deposition of momentum by Lindzen
Physics	
Clouds	Interactive clouds due to Slingo
Radiation	Short wave formulation of Lacis and Hansen, Long wave formulation of Fels and Schwarzkopf. Water vapour, ozone, carbon dioxide and clouds included
Deep convection	Kuo type formulation
Shallow convection	Tiedtke's formulation
Stratified condensation	Adjustment of supersaturation by moist adiabatic process
Vertical turbulent eddy transfer	Stability and height-dependent diffusion coefficient
Surface layer fluxes	Monin–Obukhov type formulation
Soil temperature	3-layer soil model
Soil moisture	Bucket type hydrology
Computation grid	162 equally spaced points on 82 Gaussian latitudes for radiations. Same as dynamics grid for rest of physics
Lower boundary climatological fields	
Sea surface temperature	Monthly mean
Surface albedo	Seasonal mean
Soil moisture	Monthly mean
Sea ice extent	Monthly mean
Snow depth	Monthly mean
Surface roughness	Seasonal mean
Plant resistance	Seasonal mean
Deep subsoil temperature	Annual mean

Flosolver Mk3

Flosolver Mk3 is the latest in the series of Flosolver parallel computers, developed at the National Aerospace Laboratories[6-8]. Flosolver Mk3 is based on i860 RISC processors with an on-board memory of 64MB per processor on MULTIBUS II (Figures 1 and 2). The four-processor Flosolver was used for the present parallelization of the T-80 code. The communication between the processors is handled through the Message Passing Coprocessor (MPC) and high speed Direct Memory Access (DMA) available on each of the processors. One of the processors acts as the host and it runs under standard UNIX. The other processors are termed as the Processing Elements (PEs) and run under Concurrent Executive – the parallel operating system of Flosolver. Details of programming on the Flosolver can be found in refs 9, 10. The sustained performance of the four-processor Flosolver Mk3 on a wide variety of codes is around 25 MFLOPS.

Parallelism in the model

The model computations are done in three distinct spaces namely the two-dimensional global spectral space, the one-dimensional Fourier space (spectral in azimuthal direction) for each latitude and the two-dimensional latitude/longitude grid space. Since the basis functions in the spectral expansion are orthogonal to each other, all linear operations in the spectral space involve only the selected basis function and such computations can be done in parallel. Part of the nonlinear terms are computed in the Fourier space while the rest along with the physical parametrization computations are carried out on the longitude/latitude grid. The natural flow of the model code is to compute the Fourier coefficients for each latitude, complete calculation in Fourier space and then move to the grid space through the discrete Fast Fourier Transform (FFT). After completion of computations in the grid space, mostly related to physical parametrizations which work independently at each grid point along a vertical column, atmospheric variables are again moved to the Fourier space through FFT. Finally, at the end of each time step of integration, full spectral coefficients are again formed by summing up the contributions from all the latitudes. Except for this last operation all computations involve quantities specific to one latitude only. Thus, a natural scheme of parallelization is to distribute the latitudes to different processors. This essentially implies a domain-splitting in the Fourier space only. In the present approach, the latitudes have been divided among the processors.

The inter-processor communication can be minimized if the processors have large enough memories to store the full range of spectral coefficients. In this case the only non-parallelizable part of the code is the summing up of the contributions from each latitude to form the spectral coefficients and their distribution to the processors. Keeping this in view memory requirement is not reduced.

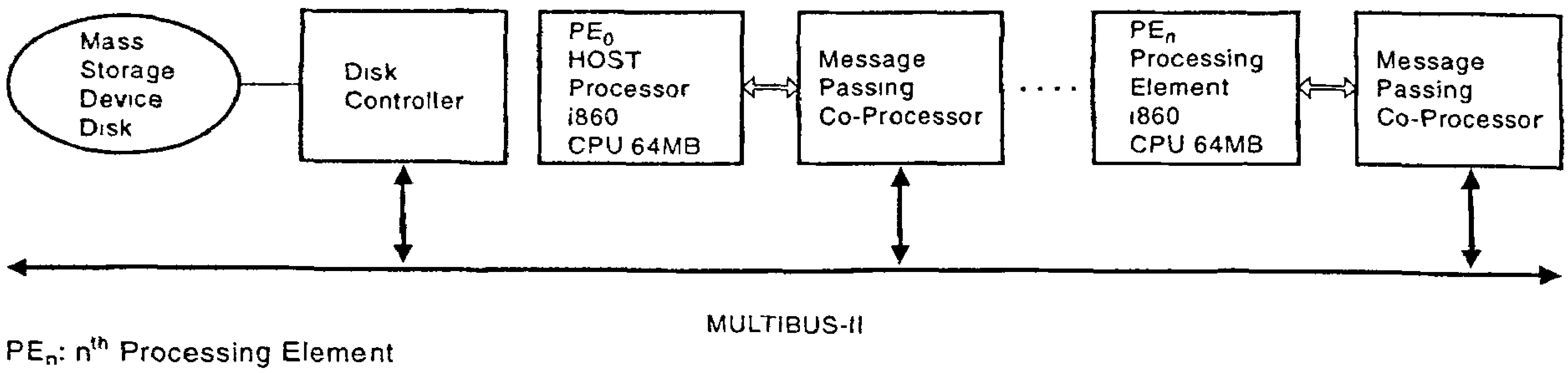

Figure 1. A schematic diagram of the architecture of the Flosolver.

Inter-processor communication

Spectral methods usually entail global communication. This increases the communication overhead significantly. The significant overhead comes from the set up time of message passing mechanisms. To reduce it, the first processor assembles the aggregate data for passing to the next processor. The next processor receives the data and does the summing of the latitude-dependent variables and passes them to the next one. When the last processor completes the assembling, the global data are ready and the forward loop is completed. In the backward loop, these data are sent to all the processors and thus the global communication is completed.

Parallelizing and debugging aspects

When the size of the code is large, portability of even a sequential code assumes frightening dimensions. Every machine has its own specific features like exception handling, I/O calls, precision details, machine specific calls, etc. Parallelizing a large code, therefore, needs all the more extra care. A logical step is to have a sequential calculation made on the same processor used in the parallel computing. The added advantage of such an approach is that it provides necessary intermediate results at every stage, which can be effectively used for debugging of the parallel computations.

Contour of the parallel code

The parallel code for Flosolver consists of two parts: (a) Host part – the part of the parallel code which runs on the Host, (b) PE part – the part of the parallel code which runs on the PEs. A separate directory is dedicated for developing and running Host and PE parts separately. The Host processor has access to I/O devices such as disk, etc., and is used both for doing I/O and calculations whereas the PEs do only computations.

In the present case to gain computational efficiency, the host is freed of computations and it performs only I/O. For ease of debugging and transparency to the users, the sequential code is altered marginally. Details of the changes made for parallelization have been discussed in ref. 11. Essentially latitudes are divided among the processors and interprocessor communication is used to update the global variables.

Reliability of calculations on the Flosolver

When the number of arithmetical operations runs into billions, the precision of calculations and propagation of round off errors determine the reliability of the calculations.

A detailed report of comparison between Cray output and Flosolver output has been made[11] and acceptability of the results computed on the Flosolver has been confirmed.

Timing details

A model for weather forecast cannot be used in an operational mode if calculations for a day cannot be made in a reasonable or acceptable time. The position for 4-processor Flosolver is shown in Table 2 (this is the worst case because code is not optimized to use the salient features of i860).

Post-processing and graphic visualization

The output file is so massive that it is very difficult to get a physical picture from the data obtained from the calculations. Post-processing the file becomes mandatory for such large calculations. Post-processing for five days' integration was done and the results have been presented in pictorial form. New graphic tools have been devised and used on the Flosolver to present the output in colour for quick and easy appreciation by the user. Figure 3 shows the temperature contours as

Table 2. Timing details

No. of days of forecast	CPU time 4-processor Flosolver	I/O time	Wall time	CPU time 2-processor Cray
1	1 h 35 min	6 min	1 h 41 min	14 min 31 s
2	3 h 09 min	12 min	3 h 21 min	
3	4 h 39 min	18 min	4 h 57 min	
5	7 h 45 min	30 min	8 h 15 min	

Figure 2. Front view of Flosolver Mk3.

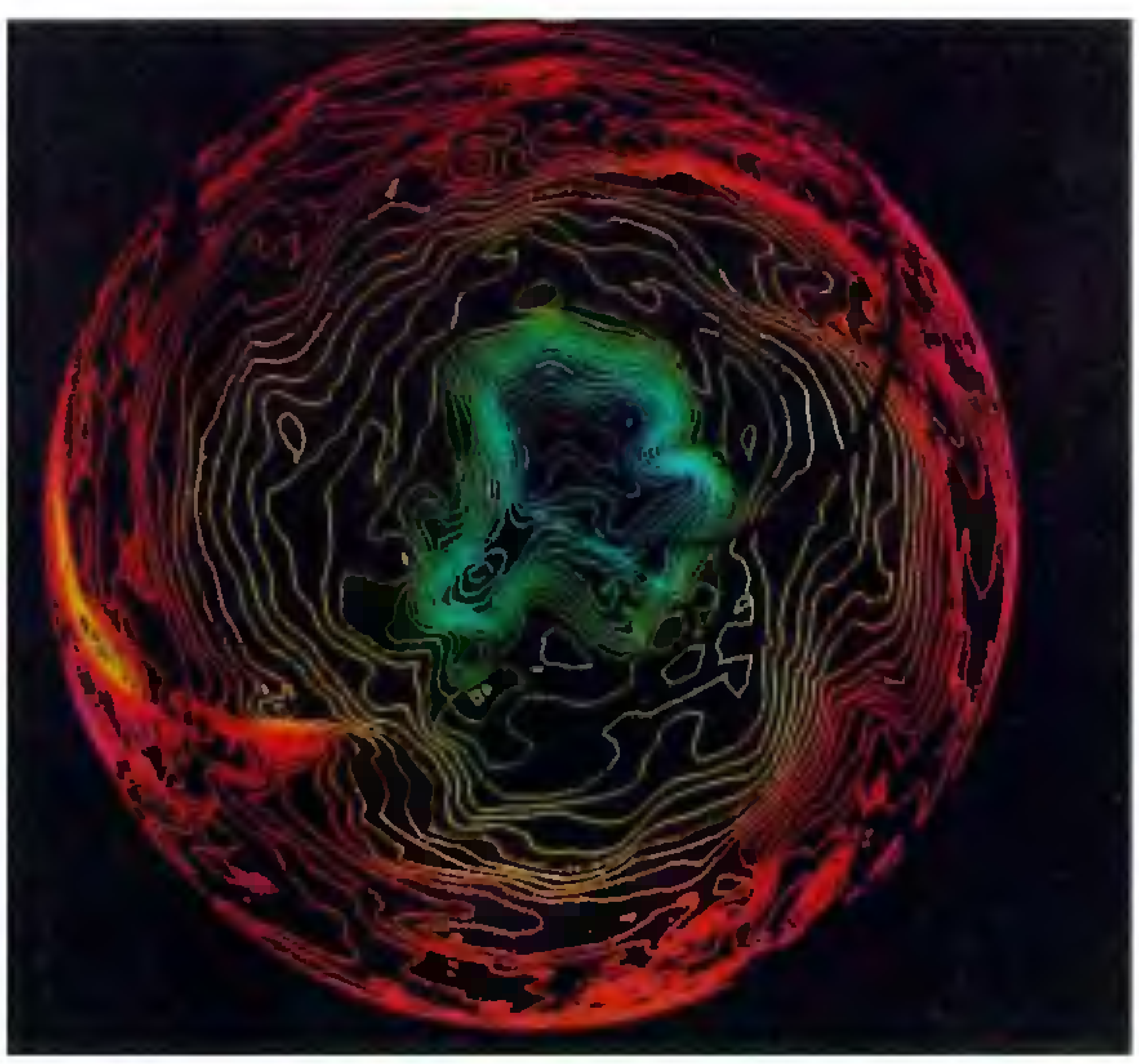

Figure 3. Temperature distribution on 4 February 1993 viewed from the polar axis.

seen from the polar axis. The Rossby waves responsible for most of the weather in the mid-latitudes are clearly seen. Figure 4 shows the temperature distribution for different orientations of the globe. Figure 5 shows the temperature distribution for a typical day at various levels in the standard format. Note the prominent Tibetian plateau and the cold tongue over the northeastern parts

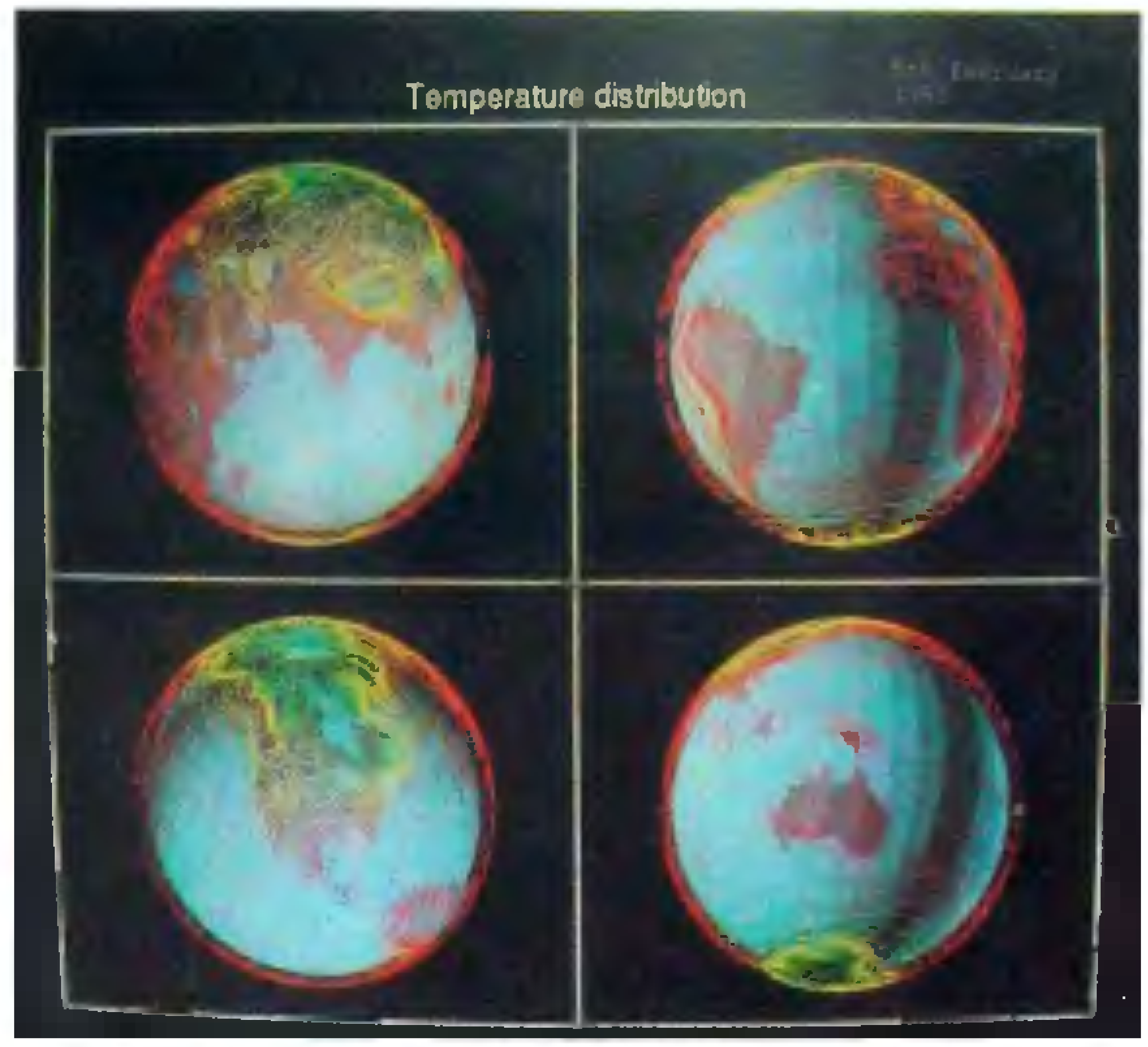

Figure 4. Temperature distribution for different orientations of the globe.

of US. The distortion at the poles cannot be avoided in these plots. Figure 6 shows the surface pressure distribution for the Indian subcontinent. The graphical processing has the feature of extracting the data relevant to a specific region and this can be seen in Figure 6.

Conclusion

Even with the unoptimized code running on the 4-processor Flosolver Mk3 in which the cache memory of the processor remains practically unused, the calculation with GCM T-80 can be made in a reasonable and realistic time frame. The code in the present form is highly Cray specific and this has been retained primarily to keep the familiarity of the user intact. If a Flosolver-specific code were to be written, CPU time of the order of fifty per cent would be easily saved. However, writing of such a code demands a very significant investment of manhours. It would approximately require the concerted effort of four programmers spanning over a period of six to eight months.

The graphical capability developed has greatly facilitated the understanding of the computed results. It would not be an exaggeration to state that colour graphics enables quick appreciation of even subtle variations in the solution.

The timings will be improved by using an increased number of processors. The number of processors can be increased to sixteen without any modification of the existing setup. In addition, there is enough scope to double the performance of the existing CPU card through

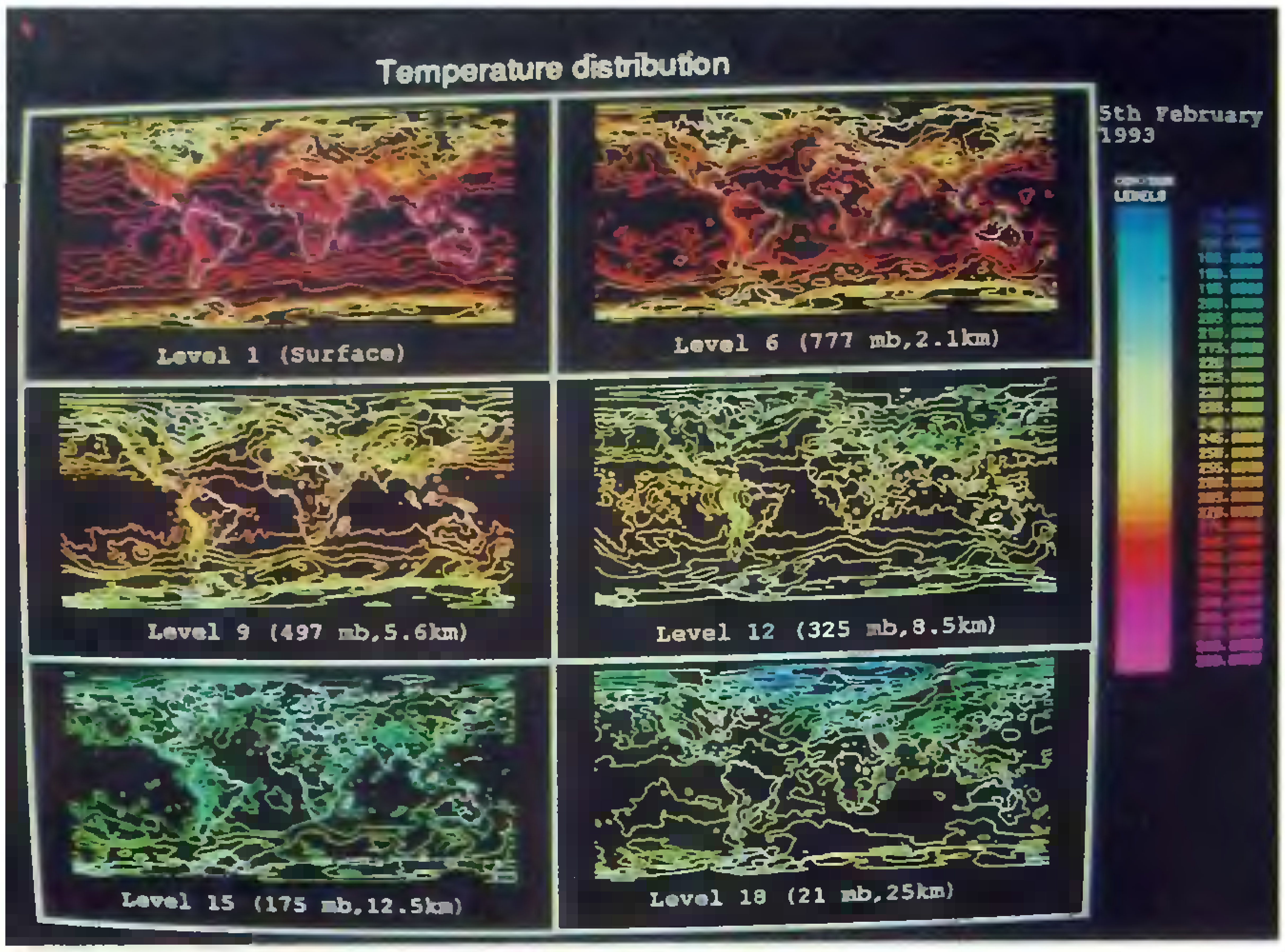

Figure 5. Temperature distribution for a typical day at various levels.

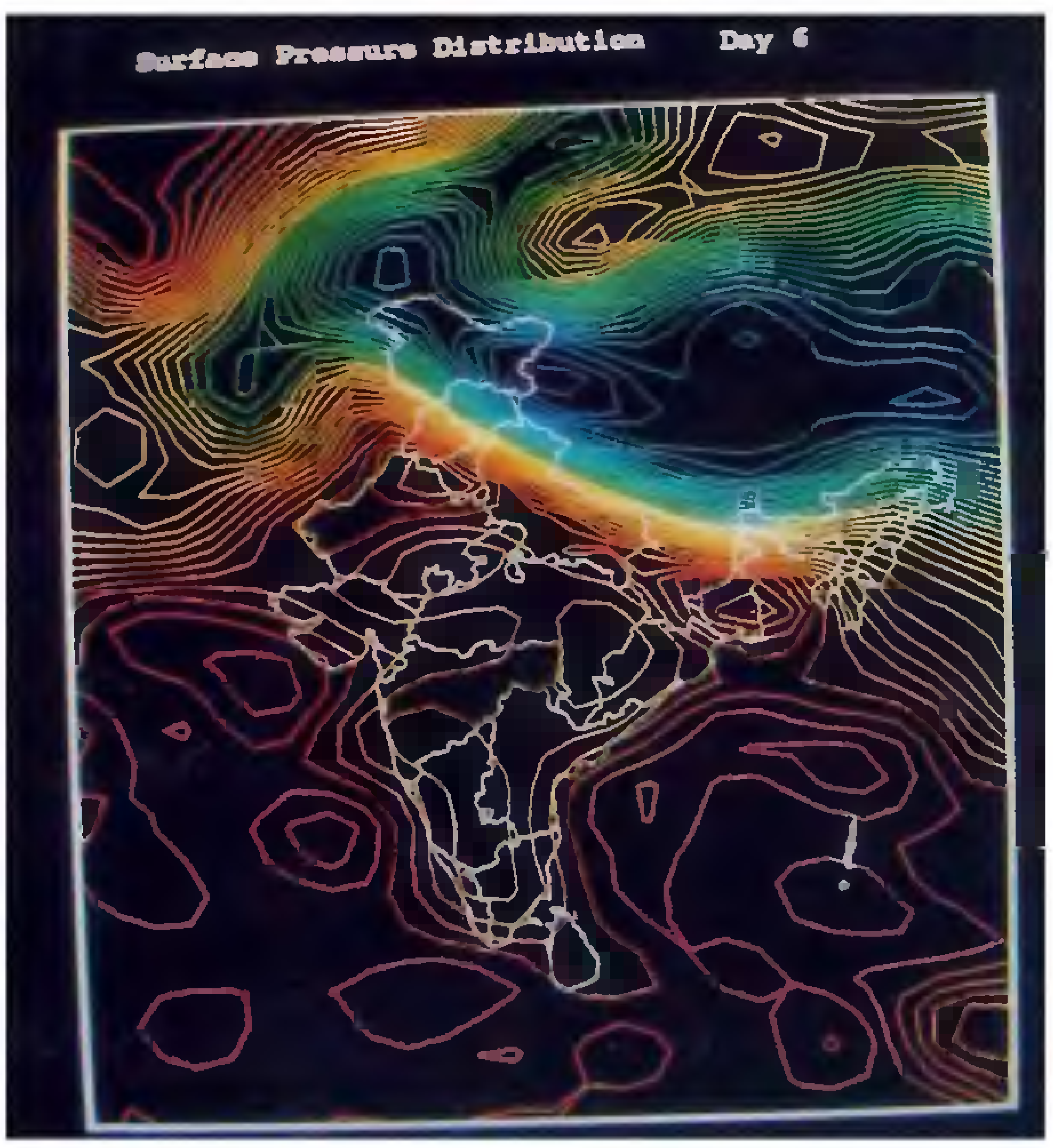

Figure 6. Surface pressure distribution for Indian subcontinent.

minor hardware modifications. But the present effort shows that research using the T-80 or similar models can be carried out with only the modest financial outlay involved in acquiring a Flosolver Mk3.

1. Richardson, L. F., *Weather Prediction by Numerical Process*, Cambridge University Press, 1922, p. 236. (Also, Chapman, S., Dover Publications, 1965.)
2. Wiin-Nielsen, A., *Tellus*, 43 AB, 1991, 4, 37–52.
3. Summary Record of the Brain-storming Session on 'Future Supercomputing Strategy for Medium Range Weather Forecasting' held in the Department of Science and Technology (DST), under the chairmanship of Prof. P. Rama Rao.
4. Official Notification, 1993, NMRF 6/13/93, Govt. of India, Department of Science and Technology for implementation of GCM T-80 code.
5. Kalnay, E., Joseph, G. Sela, Campana, K. A., Basu, B. K., Schwarzkopf, M. D., Long, P. E. Caplan, P. M. and Alpert, J. C., January, 1988, USA Documentation of research version of the NMC Medium Range Forecasting model.
6. Sinha, U. N., Deshpande, M. D. and Sarasamma, V. R., *Curr. Sci.*, 1988, 57, 1277–1285.
7. Sinha, U. N., Deshpande, M. D. and Sarasamma, V. R., *Super Computer*, July 1989, VI(4), 37–42.
8. Sinha, U. N. *et al.*, PD ST 9133, NAL, Bangalore, 1991.
9. Ravishankar Nanjundaiah, Murugan, M. and Sinha, U. N., PD CF 9212, NAL, Bangalore, 1992.
10. Basu, A. J., PD CF 9309, NAL, Bangalore, 1993.
11. Sinha, U. N. *et al.* PD FS 9403, NAL, Bangalore, 1994.

ACKNOWLEDGEMENTS. We, the team members of Flosolver, thank Prof. R. Narasimha for his vision, support and imagination. We also thank Dr K. N. Raju, Director, NAL, Dr S. K. Mishra and the community of meteorologists. Financial assistance under Multi Disciplinary Research in Science and Technology (SERC) from Department of Science and Technology (DST) is gratefully acknowledged. We would also like to thank Dr S. Bhogle for his editorial help.

Received 20 May 1994; accepted 25 May 1994