Operational mesoscale atmospheric dispersion prediction using a parallel computing cluster

C V Srinivas[1,*], R Venkatesan[1], N V Muralidharan[1], Someshwar Das[2], Hari Dass[3] and P Eswara Kumar[3]

[1] *Radiological Safety Division, SE Group, Indira Gandhi Centre for Atomic Research, Kalpakkam 603 102, India.*
[2] *National Centre for Medium Range Weather Forecasting, New Delhi 110 003, India.*
[3] *Institute of Mathematical Sciences, Chennai 600 113, India.*
[*] *e-mail: cvsri@igcar.ernet.in*

An operational atmospheric dispersion prediction system is implemented on a cluster supercomputer for Online Emergency Response at the Kalpakkam nuclear site. This numerical system constitutes a parallel version of a nested grid meso-scale meteorological model MM5 coupled to a random walk particle dispersion model FLEXPART. The system provides 48-hour forecast of the local weather and radioactive plume dispersion due to hypothetical airborne releases in a range of 100 km around the site. The parallel code was implemented on different cluster configurations like distributed and shared memory systems. A 16-node dual Xeon distributed memory gigabit ethernet cluster has been found sufficient for operational applications. The runtime of a triple nested domain MM5 is about 4 h for a 24 h forecast. The system had been operated continuously for a few months and results were ported on the IMSc home page.

Initial and periodic boundary condition data for MM5 are provided by NCMRWF, New Delhi. An alternative source is found to be NCEP, USA. These two sources provide the input data to the operational models at different spatial and temporal resolutions using different assimilation methods. A comparative study on the results of forecast is presented using these two data sources for present operational use. Improvement is noticed in rainfall forecasts that used NCEP data, probably because of its high spatial and temporal resolution.

1. Introduction

An online Radiological Emergency Response System for Kalpakkam nuclear site is under development at IGCAR to serve as a 'live tool' for radiological emergency response due to inadvertent air-borne effluents. The purpose of the system is to assess and predict the local weather condition, the ensuing atmospheric dispersion and the consequent environmental radioactive dose for unit quantity of releases. The plume distribution pattern and dose forecast in terms of direction, distance range and the levels of the plume dose form the basis for decision making in emergency response. The spatial database of the region for the cadastral, transport and other infrastructure information is created using remote sensing techniques and Geographical Information System (GIS). Integration of dispersion prognosis with the region's spatial database would provide objective decision making tools for emergency planning.

Essentially it requires a realistic regional weather model that simulates and forecasts in a reasonable time limit. Networked PC clusters otherwise called 'Beowulf' have been widely used in recent times for high performance scientific computations at an affordable cost. Cluster computing technique along with open source operating systems such as Linux is an appropriate cost effective solution for achieving the above objective.

Keywords. Operational mesoscale dispersion model; cluster computer; MM5; RSL; FLIC.

J. Earth Syst. Sci. **115**, No. 3, June 2006, pp. 315–332
© Printed in India.

This paper presents the results of the implementation of the meso-scale regional weather forecast model MM5 coupled with a dispersion model FLEXPART on various high performance parallel computers. For the model to run in operational mode, uninterrupted availability of data on initial/boundary conditions is essential. These data are available from two sources namely NCEP, USA and NCMRWF, India at different spatio-temporal resolutions. The first section of the paper briefly deals with the models MM5 and FLEXPART; the implementation of the parallel version of MM5 on different parallel computers and the results of runtime are discussed in the next sections. The paper also discuses the differences between the results when data from the above two sources are used in model initialization and integration.

The runtime of the model is reduced to about 4 hours for a 24-hour prediction on a dual Xeon Gigabit Ethernet cluster, which is sufficient for the operational model. It is feasible to construct such a cluster with optimum number of processors from off-the-shelf components. The combined weather and dispersion code was executed continuously for three months in the Scali pilot cluster was IMSc and the system was found to be consistent and suitable for operational use. The operational model results using NCEP and NCMRWF data sets for initialization are evaluated taking a dry and wet weather situation. While no significant differences are found in many of the predicted meteorological parameters, improvements are noticed in the rainfall when NCEP data is used. This may be because of high spatial resolution in the initial and boundary conditions obtained from NCEP data for the model operation.

2. Brief description of the models MM5 and FLEXPART

Realistic simulation of dispersion and environmental radioactive dose estimations in the off-site long-range distances (> 10 km) requires the application of prognostic meso-scale atmospheric models for the prediction of the time and space varying meteorological fields. A community developed Regional Weather Forecast model (MM5) coupled to an advanced Random Walk Particle dispersion model (FLEXPART) is used for the emergency response. The MM5 is a non-hydrostatic meso-scale meteorological model and incorporates detailed physics for various processes, i.e., convection, radiation, micro-physics, atmospheric boundary layer and surface layer energy exchange processes (Grell *et al* 1994). The model can be configured with suitable horizontal and vertical resolutions in a terrain following vertical coordinate system for the simulation of meso-scale atmospheric phenomena such as cyclones, thunder storms, land-sea breezes, etc. As the model predicts the three-dimensional meteorological fields in a high horizontal resolution, it is chosen for forecasting the wind field in a meso-scale range of 100 km around Kalpakkam nuclear site. The model facilitates dynamical and numerical aliasing of the various scales of circulation viz., synoptic scale (> 1000 km), meso-scale (> 10 km) and micro-scale (< 1 km) with nested domain setup so that meso-scale phenomena are simulated realistically. A 3-level interactive nested domain in 3:1 grid size ratio at 18, 6 and 2 km resolutions in the horizontal and 26 levels in the vertical with suitable computational time-steps is selected in the model and prediction is made in all the three spatial scales. Ten of the vertical levels are chosen in the lower atmosphere. The surface boundary conditions refer to the description of land surface where the energy exchange takes place between the atmosphere and the earth. The upper air boundary refers to the relatively stationary state of the atmosphere at the top of the model, i.e., the height of troposphere in this particular case, for the period of mesoscale model integration. The lateral boundary values are kept time dependent for the inner domains and as relaxation for the outer domain. The quasi-stationary terrain data of topography, landuse, soil texture and deep soil temperature for bottom boundary condition are taken from USGS and FAO. A 2nd order turbulence closure scheme (Burk and Thomson 1989) is used for the planetary boundary layer. For the first domain Grell convective parameterization is used. The microphysics for moisture is calculated using a simple ice scheme (Dudhia 1989). The other options used in the model are listed in table 1. The model is initialized with a set of surface and upper air meteorological observations in analysed form, which is discussed subsequently.

An open source Lagrangian Particle dispersion model, FLEXPART, developed at Technical University of Munich, Germany (Stohl 1999) for research and operational use is employed for the simulation of the transport and diffusion of contaminants in the atmosphere and their deposition on the ground. It is numerically more efficient and accurate and incorporates better turbulence description than the Gaussian Plume/Puff models usually employed in regulatory framework. The meteorological fields needed by the dispersion model are predicted using the hydrodynamic model MM5. These are the wind field (u, v, w) and the boundary layer parameters (roughness length, friction velocity, convective scaling velocity, monin obukhov length, etc.). The mesoscale model is validated for realistic simulation of weather in

Table 1. *Description of MM5 configuration at IGCAR for 48-hour forecast.*

Model elements	Components	Specifications
Grid	Horizontal 3 domains (extents, no. grids, resolution)	Staggered Arakawa B-grid Outer Domain-1 (71.16°–86.25° E, 7.18°–21.58° N), 100 × 100, 18 km Middle Domain-2 (72.88°–77.62° E, 10.66° N–15.6° N), 100 × 100, 6 km Inner Domain-3 (74.3° E–74.87° E, 12.12°–13.77° N)100 × 100, 2 km
	Vertical	26 sigma levels (terrain following) (10 levels near surface – 990, 985, 980, 975, 970, 960, 950, 925, 900, 875, 850 hPa)
	Topography	USGS (Interpolated depending on resolution)
	Vegetation	USGS 25 categories
Dynamics		Non-hydrostatic Two-way nesting (interactive nests)
	Prognostic variables	Horizontal (u, v), vertical (w) winds, temperature (t), specific humidity (q), perturbation pressure (p_s), optional: TKE, cloud water, rain, ice, snow, etc.
	Time differencing	Semi-implicit Domain-1 – 54 s, Domain-2 – 18 s and Domain-3 – 6 s
	Horizontal diffusion	Fourth order for inner domain Second order for the coarser domain
Physics	Deep convection	Grell (simplified Arakawa–Schubert's scheme) only for Domain-1
	Grid scale resolved moisture	Dudhia's scheme (simple ice scheme)
	Radiation	Simple cooling depending on temperature
	Planetary boundary layer	Burk-Thomson level 1.5 TKE scheme
	Land surface processes	Force–Restore Ground temperature prediction
Boundary conditions	Lateral boundary	Time-dependent/nest for inner domains Relaxation/inflow-outflow for the mother domain
	Upper boundary	Radiative condition

the larger region, land–sea breeze circulation in the Kalpakkam region and the associated thermal internal boundary layer formation (Srinivas *et al* 2004; Srinivas and Venkatesan 2005).

3. Implementation of parallel MM5 on high performance clusters

The weather model MM5 consists of a finite difference formulation of the time-dependent fundamental equations of the atmospheric flow dynamics plus physics computations for the simulation of clouds, radiation transfer, moist convection, and turbulence exchange in the PBL, etc., in a cubic-three dimensional region representing the atmosphere. As the model is computationally very intensive, minimization of the model runtime is crucial for emergency application. Usually supercomputers like CRAY are used for executing the operational model. In recent times distributed-memory machines are becoming increasingly common. With the advent of network cluster technology and development of open source operating system Linux and other supporting software like message passing interface (MPI), PC clusters have been demonstrated as an efficient and alternative way of achieving high performance scientific computing in an affordable manner (Savarese and Sterling 1999; Baker and Buyya 1999; Wang *et al* 2005). Parallel MM5 model was developed with the same source approach by Argonne National Laboratory (ANL) using a Runtime System Library (RSL) and a Fortran Loop and Index Converter (FLIC) (Michalakes 1997). RSL is a Fortran-callable parallel runtime system library for implementing regular-grid climate models with nesting on distributed memory parallel computers. It performs domain specification, decomposition over processors and remapping, intra- and inter-domain communication, local computation on each processor domain and distributed I/O. The FLIC is a fortran pre-compiler that generates the modified code with loop and index transformations from global to local view of memory. Thus, the parallel code maps the three-dimensional domain on to a two-dimensional array of processors and computes a part of the model on each processor with inter processor communication for data exchange.

The parallel model has been successfully implemented and runtime tested on a single board 8 processor shared memory Xeon server at IGCAR, 9 node dual Xeon Gigabit Ethernet network cluster, 8 node dual Xeon Scali pilot cluster with SCI interconnect and on 16 node dual Xeon with SCI interconnect on the 144 node 'KABRU' at the Institute of Mathematical Sciences (IMSc), Chennai. KABRU is a Tera flop cluster on which the dispersion system is implemented in operational mode. This cluster with a sustained performance of 1002.3 Gflops on 13th October 2004, ranks 257 among the top P500 high speed computing systems in the world (internet source http://www.imsc.ernet.in). The model was compiled using the f90 compilers; and used the open source MPI software from Argonne National Laboratory in all the above-mentioned machines except in the Scali clusters in which the Scali MPI was used.

4. Application of NCEP/NCMRWF data for MM5 initialization

Since MM5 is a limited area weather forecast model, it requires the specification of periodic boundary conditions over the prediction time in addition to initial meteorological conditions. The initial state is specified from a set of observations at the surface and at different heights in the upper atmosphere in the form of analyses provided by a global analyses and forecast system. The boundary conditions are updated periodically using the forecasts of a global weather model. In the operational model these data are to be provided as inputs in real-time. These data are available in open source from National Centre for Medium Range Weather Forecast (NCMRWF) and alternatively from National Centre for Environmental Prediction (NCEP). The global surface and upper air observations are transmitted by global telecommunication system (GTS) to the forecast centers. The global data assimilation is carried out by a spectral statistical interpolation (SSI) scheme. The GDAS at NCEP is updated continuously throughout the day. At NCEP the global forecast is prepared by a global spectral model MRF with triangular truncation at 382 waves and 64 unequally spaced levels (T382), with an equivalent horizontal resolution 0.5×0.5 degree latitude/longitude (Kanamitsu 1989; Kalnay *et al* 1990). It is a primitive equation model and incorporates the physics for the processes of radiation, gravity-wave drag, convective precipitation, shallow-convection, non-convective precipitation, horizontal diffusion, planetary boundary layer and surface energy processes. The current version of MRF incorporates several advanced physics for the above processes. The model is run four times per day (00, 06, 12 and 18 UTC) and the forecast is made at every 3 h intervals up to 7.5 days.

At NCMRWF the global analysis is prepared with observations corresponding to 00Z using SSI scheme. The global forecast model (T-80) (with 80 waves and 18 vertical levels) uses this analysis as initial condition and produces a global forecast up to 7 days at 12 h intervals (figure 1). The T80 is a similar model to MRF and follows the same dynamics. However, it has the physics of the earlier 1988 version MRF model. The inputs for MM5 are created after spectral transform to a Gaussian grid space and then, interpolation from a Gaussian to a regular grid equivalent to 1.5° latitude–longitude (∼150 km) resolution. The T80 model is run once in a day (corresponding to 00 UTC) and the inputs to MM5 are provided in the model 'pregrid' format (figure 1).

5. Results

5.1 *Runtime performance of the model*

Choosing the number of processors in the north–south and east–west directions suitably on each machine the domain decomposition over the processors in the parallel model is incorporated. The processor layout is chosen as 2×4 in PIII Xeon, 2×8 on 8 node Pilot cluster and 4×8 on 16 nodes on Kabru cluster. The model runtime for a day's weather prediction using the model configuration in table 1 is about 8 hours on an 8-processor shared memory parallel PIII Xeon server, 4 hours on 8-node Pilot Cluster and 2.5 hours on 16 nodes of Kabru cluster (table 2). The sequential model runtime for this model configuration is found to be 22 h on single P4 processor IBM PC. A cost effective way of building small but high performance clusters is LAN clusters. The runtime of the parallel MM5 with 2×5 processor layout using 5 nodes on gigabit LAN Xeon cluster at IMSc has been found to be 5.25 h, which is nearer to the runtime on 16-processor Pilot cluster. Since it is feasible to build such gigabit switch cluster indigenously from off-the-shelf components, the scalability of the parallel MM5 code is tested on gigabit switch cluster to arrive at optimum processors for the operational runs of the model. The test runs are made with the parallel MM5 benchmark case as well as the realistic MM5 used at IGCAR. In the benchmark parallel MM5 the domain, grid points, etc., used are simpler than what are used in realistic simulations. This bench mark case uses

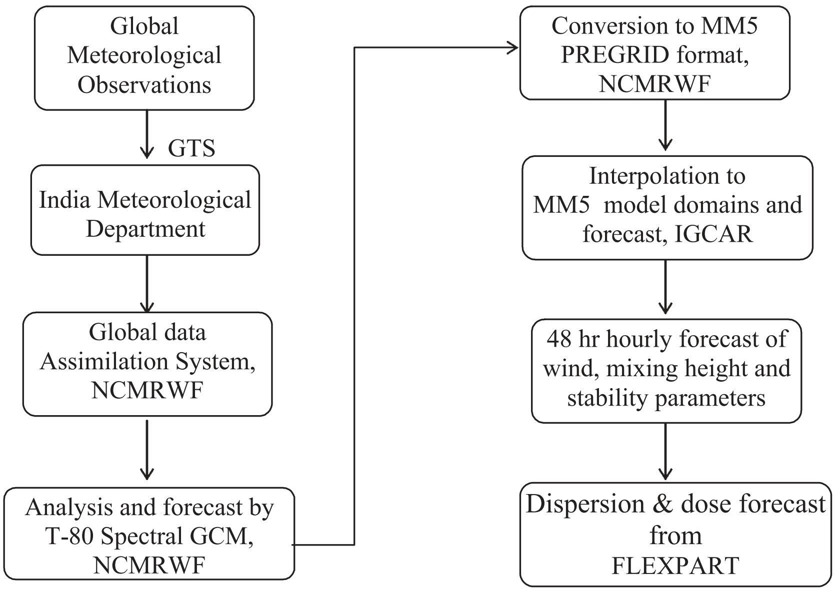

Figure 1. Flowchart of operational real-time atmospheric dispersion system at IGCAR.

Table 2. *Runtime performance on different computing platforms.*

Machines available	Configuration	Processor specifications	OS and other software	MM5 model configuration	Runtime for 24 h forecast
IBM – PC	Desk top PC	Intel P-IV 1.6 GHz 1 GB SDRAM	Linux OS with Portland f90 and gcc compilers	3 level nested domain with minimum details of physics and turbulence	24 h
Xeon Server/CC, IGC	Model: BULL Express 5800 180-Ra7 Single board shared memory (SMS)	8 no. of Intel P-III Xeon 700 MHz SMP, 32 KB L1 Cache and 2 MB L2 ECC Cache per proc., 128 MB L3 cache. Memory: 4 GB (8 × 512)	Linux OS with Portland f90 and gcc compilers MPICH 1.2.6 MPI library	-do-	8 h
Pilot Cluster/IIMSc	Distributed memory (8 nodes/16 CPUs)	Intel Dual Xeon, 2.4 GHz, 2 GB DDR SDRAM per node Dolphin 3D SCI Interconnect	Linux OS with Portland f90 and gcc compilers SCALI MPI library	-do-	4 h
KABRU Cluster/IIMSc	Distributed memory (16 nodes/32 CPUs)	Intel Dual Xeon, 2.4 GHz, 2 GB DDR SDRAM per node Dolphin 3D SCI Interconnect	-do-	-do-	2.5 h
Switch mode cluster	Distributed memory (5 nodes/10 CPUs)	Dual Xeon 2.4 GHz, 1 GB DDRAM, gigabit ethernet switch	Linux OS with Portland f90 and gcc compilers MPICH 1.2.6 MPI library	-do-	5.25 h

112×136 grids, 33 vertical levels, time step of 81 sec, MRF PBL scheme, Grell convection scheme, mixed phase microphysics, multiplayer soil model and cloud interaction radiation scheme. It is seen that the runtime does not reduce linearly with increase in the number of processors (figure 2). It is a well-known fact that the use of multiple processors in parallel speeds up a task by an amount

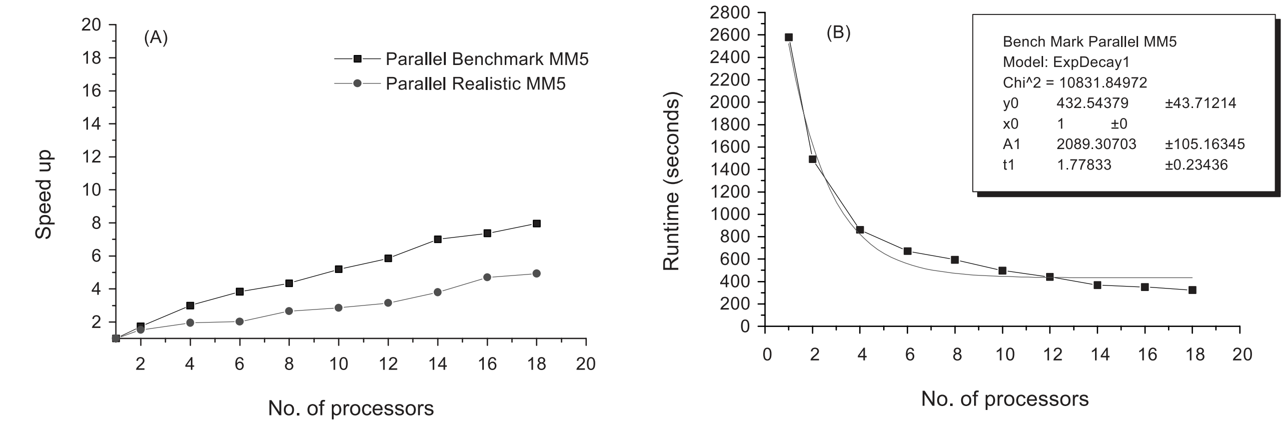

Figure 2. Runtime results of parallel MM5 on Xeon gigabit switch cluster **(A)** speed-up and **(B)** runtime with number of processors used.

depending on the parallel efficiency and is problem-dependent. However, since MM5 is a CFD model known to be scalable on a number of computing architectures the performance parameters are calculated on the present cluster. The speed up (S) is calculated as $S = t/t_p$, where t is the runtime on a single processor (sequential run), and t_p is the runtime on 'p' processors. The efficiency (E) is calculated as $E = S/P$, where 'p' is the total number of processors used in the calculation. Thus in the benchmark case the speed up observed is roughly about 8 and the efficiency achieved is about 44% (figure 2A) which is similar to the results obtained on similar network clusters elsewhere. It is to be noted here that the scalability of any parallel code depends on the network technology used. On a common switched gigabit ethernet we have experimentally found that MM5 scales well up to 18 processors, above which the bandwidth and latency of the devices limit the performance. The reduction in runtime of benchmark MM5 case as a function of number of processors is shown in figure 2(B). The Gaussian fit of the runtime results is given by $T_o = T_{\text{sat}} + T1 \cdot \text{Exp}(-n/\lambda)$ where T_{sat} is 432 s and $\lambda = 1.78$. The plot indicates a saturation in runtime at about 18 processors beyond which no significant improvement is expected in runtime on gigabit ethernet switch cluster. For the realistic MM5 used at IGCAR the speed up and efficiency are noticed to be 4.93 and 29.3% respectively. This is due to the application of multiple nested grids and more advanced physics for boundary layer, etc., in the model in operational mode. Roughly, a reduction of 5 times in runtime is achieved using 9 dual Xeon nodes (18 processors) as against a single node sequential run. Based on the above runtime results a cluster computer facility with 9-node Dual Xeon is commissioned at IGCAR for model operation.

5.2 *Operationalisation of MM5 and FLEXPART*

To operationalise the model, the initial/boundary condition data are to be made available to the model at specified periodic intervals in an uninterrupted way. The daily analysis and 48 h forecast fields from the T80 model based on 00 GMT observations are downloaded from NCMRWF ftp server at 17:30 IST. The MM5 model is initialized with data corresponding to 00 GMT (05:30 IST) and integrated for 48 h, the boundary conditions are updated every 12 h. The model predicts the 48 h regional scale and local range meteorological fields at 1 h interval, which is then used by the dispersion model FLEXPART for radioactive plume forecast and dose estimation. FLEXPART takes the three-dimensional wind, mixing height, stability parameters, rainfall and the surface boundary data of terrain/land use from MM5 for dispersion/deposition calculation. Figure 1 shows the schematic flow chart of the procedure for weather and radioactive dose forecast in the off-site long range. The results of the daily forecast for the weather and dispersion are displayed initially on the IMSc web page (http://www.imsc.res.in/~kabru/mm5.htm). The daily forecasts consist of a set of important parameters like sea level pressure, surface wind, temperature, humidity, rainfall and the plume dispersion forecast in terms of concentration and different forms of radioactive dose (figure 3). The processes of the data downloads, initial conditions preparation, meso-scale model initialization and integration, dispersion calculation using weather forecast and the outputs of predicted results in graphical formats are automated using Unix shell scripting language. The model is run operationally by the event scheduler daemon 'cron' everyday and it is found that the model is very

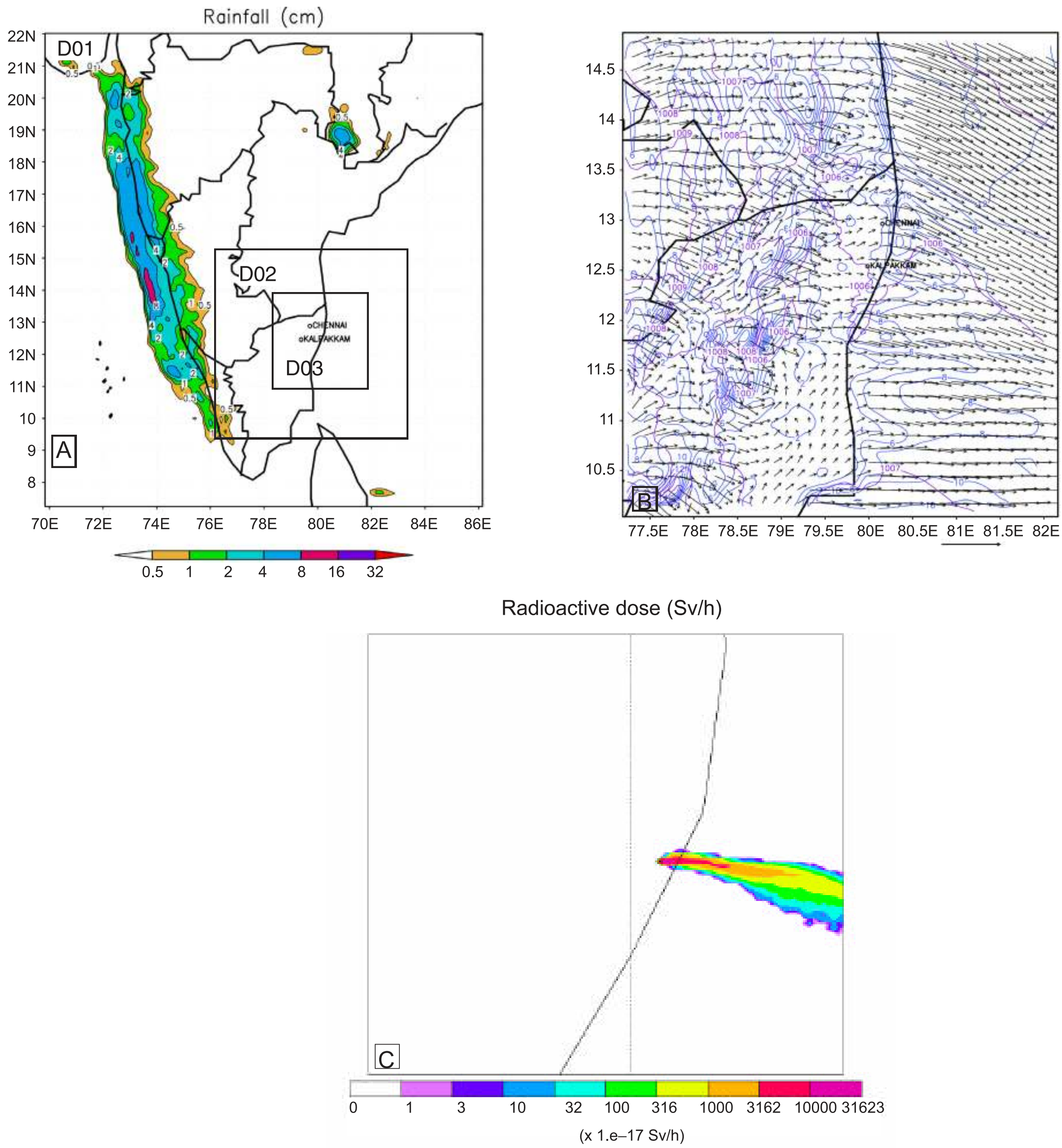

Figure 3. Daily forecasts from the operational model. **(A)** Rainfall in south Indian Peninsula in 1st domain, **(B)** Sea level pressure (in hPa) and surface wind field in the Chennai region in 2nd domain, and **(C)** Simulated radioactive dose (in Sv/h) in the Kalpakkam region, the colour shades are in logarithmic scale.

consistent in its operation by giving unhindered forecasts.

5.3 *Evaluation of meteorological data from NCEP/NCMRWF for operational model*

The data for real-time applications from NCEP, USA is available at a spatial resolution of 1° latitude–longitude ($\sim 100 \times 100$ km) and at a higher time frequency (every 6 h) than the NCMRWF data. Difference also lies in the number of observations incorporated in the analyses procedure. It is known that the quality of the initial and boundary values provided to meso-scale models influence their simulations, which in turn influence the atmospheric dispersion and dose simulations by dispersion models. Initial case studies with MM5 are carried out with NCEP data for simulated flow characteristics and are validated with mini-Sodar and tower based data (Srinivas *et al* 2004, 2005). In this context, a preliminary case study is undertaken to compare and evaluate the NCEP and NCMRWF data sets for their application in operational model. Two simulations with MM5 are made using NCEP and NCMRWF data sets for dry and wet weather situations and using

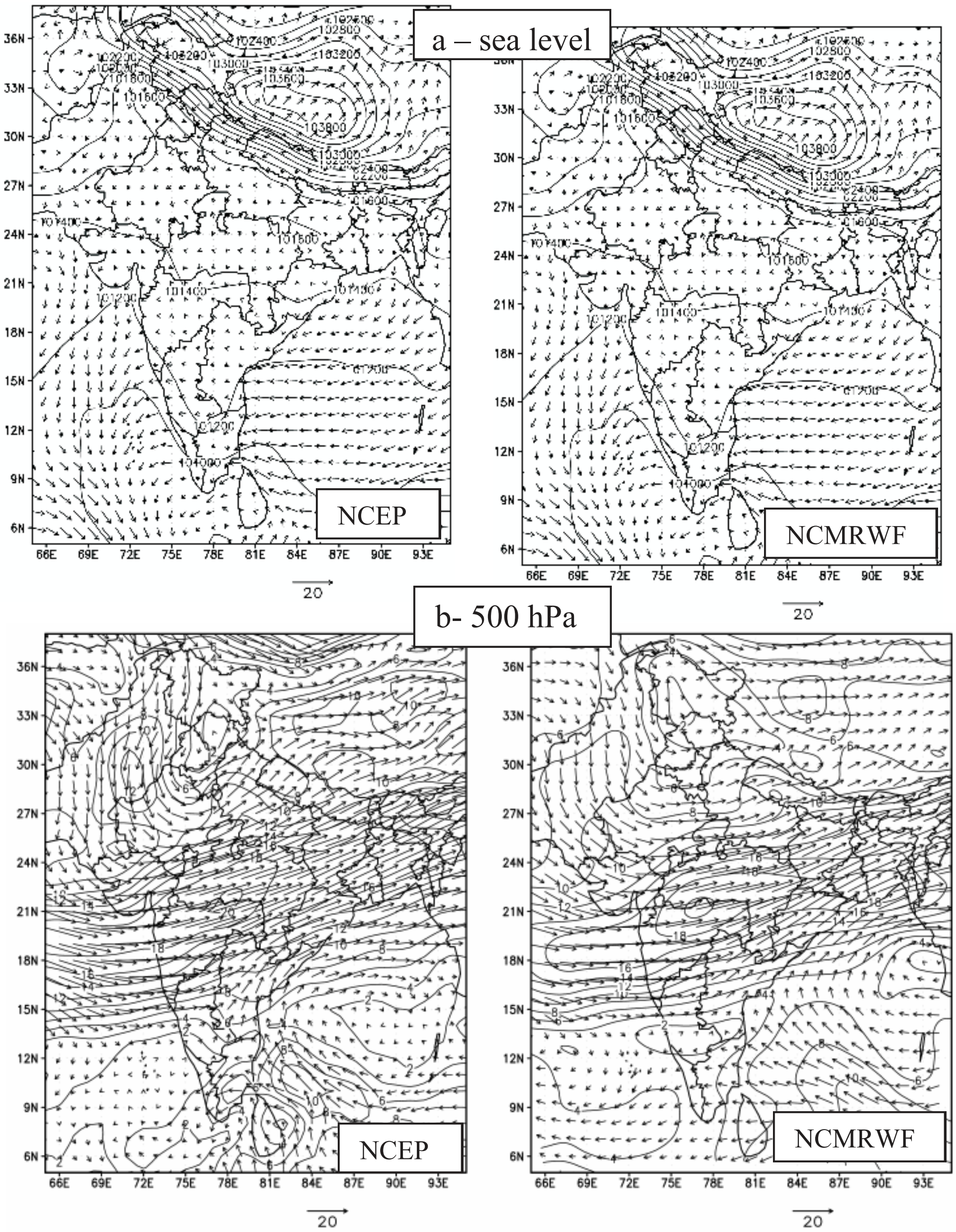

Figure 4. Initial conditions to MM5 from NCEP/NCMRWF sources at 00Z on 29.10.2004. **(a)** Sea level pressure and surface wind. **(b)** Winds at 500 hPa level.

the model configuration used at NCMRWF (internet source: http://www.ncmrwf.gov.in/mm5-new-web-des.htm), the results from the 2nd domain with 30 km resolution over the Indian region are compared. GFS pressure level gridded binary data (pgrbf) products are used for simulation with NCEP data. The wet weather case is studied for 29–31 October 2004 during the active phase of the northeast monsoon in Tamil Nadu and the dry weather case is studied for 7–9 February 2005 in the post monsoon winter period. The boundary conditions are updated uniformly at 12-h interval during model integration while using either data. The initial conditions corresponding to these cases are compared first to find the differences among the above data sets. Model predicted parameters of pressure (p), temperature (t), wind (U), relative humidity (RH), geopotential height (h) at standard pressure levels 925, 850, 500, 200 hPa and precipitation (rf) are compared with radiosonde observations using statistical indices taken from Anthes *et al* (1989). These are the root mean square error (RMSE) between the forecast and the observations; the bias score, which measures the model tendency to systematically overestimate or underestimate a parameter or event, the correlation coefficients between forecast and observations and the threat score for rainfall, which measures

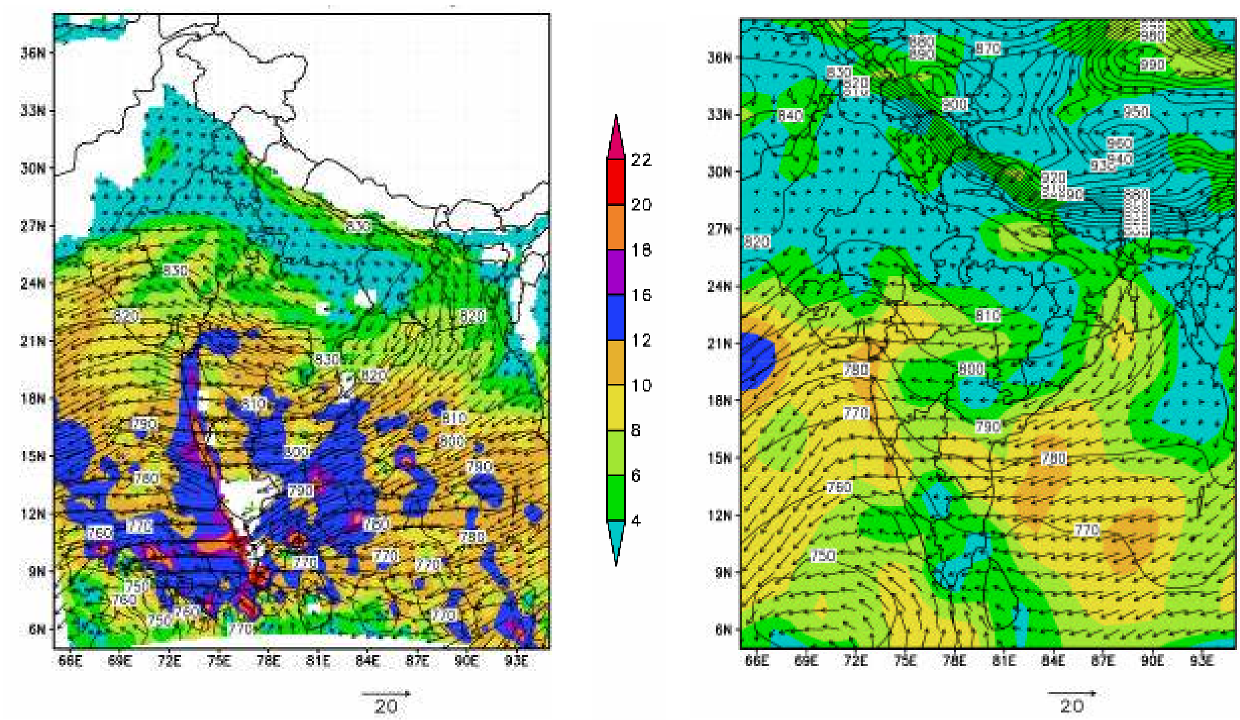

Figure 5. Geopotential height (m) and winds (ms^{-1}) at 925 hPa level at 00Z (06:00 IST) on 31 October 2004. **Left:** Simulation with NCEP data, **right:** NCEP analysis.

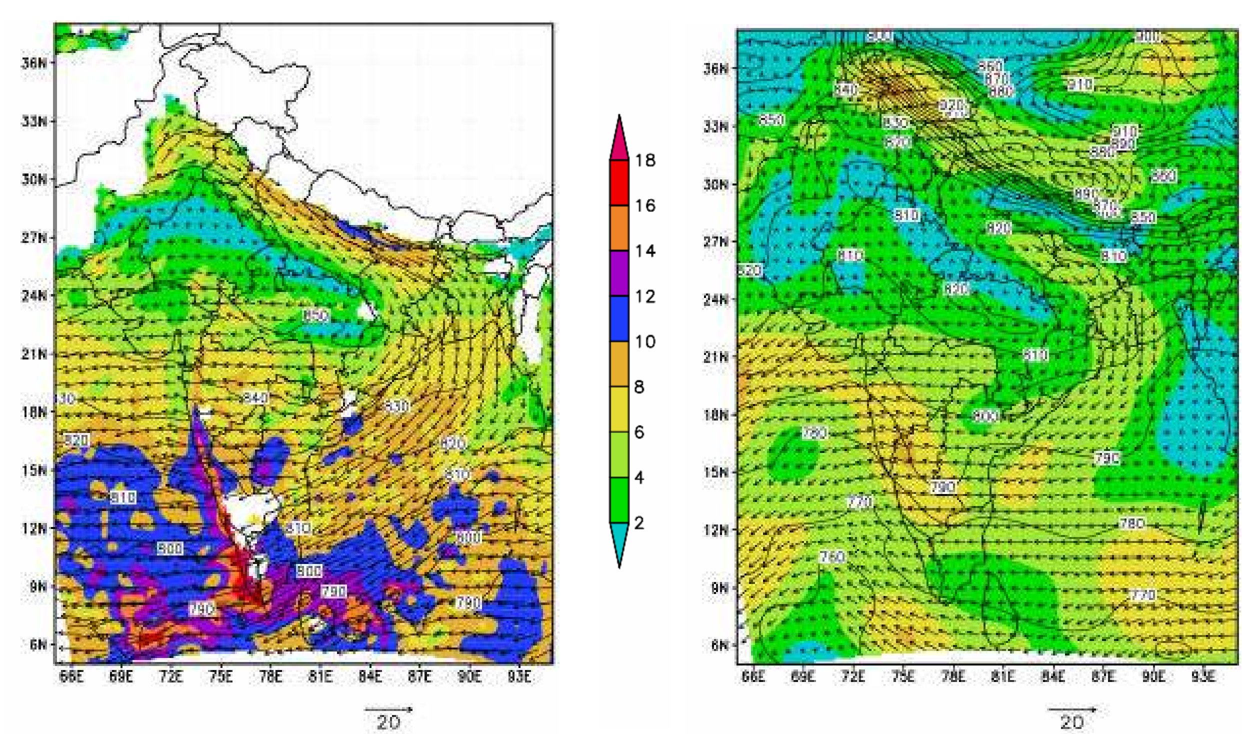

Figure 6. Geopotential height (m) and winds (ms^{-1}) at 925 hPa level at 00Z (06:00 IST) on 31 October 2004. **Left:** Simulation with NCMRWF data, **right:** NCMRWF analysis.

the model ability to forecast rainfall, classified into categories.

Both the NCEP and NCMRWF data sets are found to represent in their analysis the major features of circulation although some differences are seen to exist in their pattern. A well-marked low-pressure trough over southern peninsula, low-level cyclonic circulation west of Sri Lanka (up

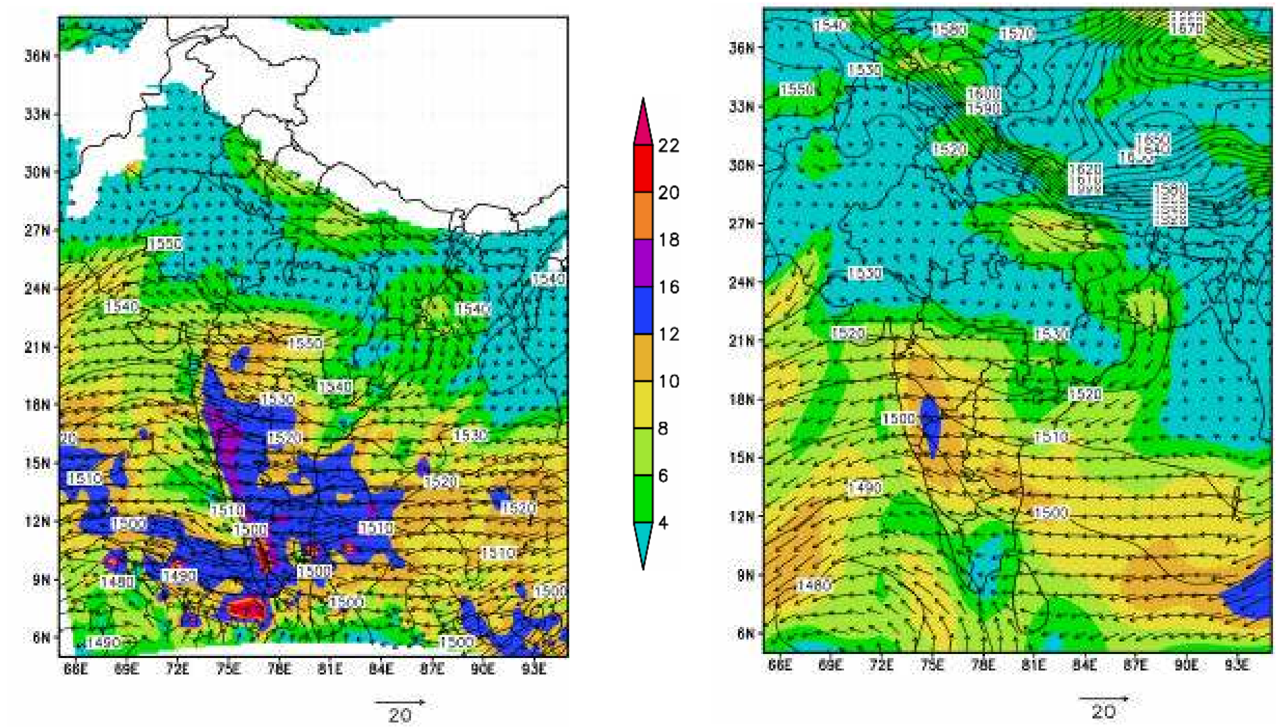

Figure 7. Geopotential height (m) and wind (ms^{-1}) at 850 hPa level at 00Z (06:00 IST) on 31 October 2004. **Left:** Simulation with NCEP data, **right:** NCEP analysis.

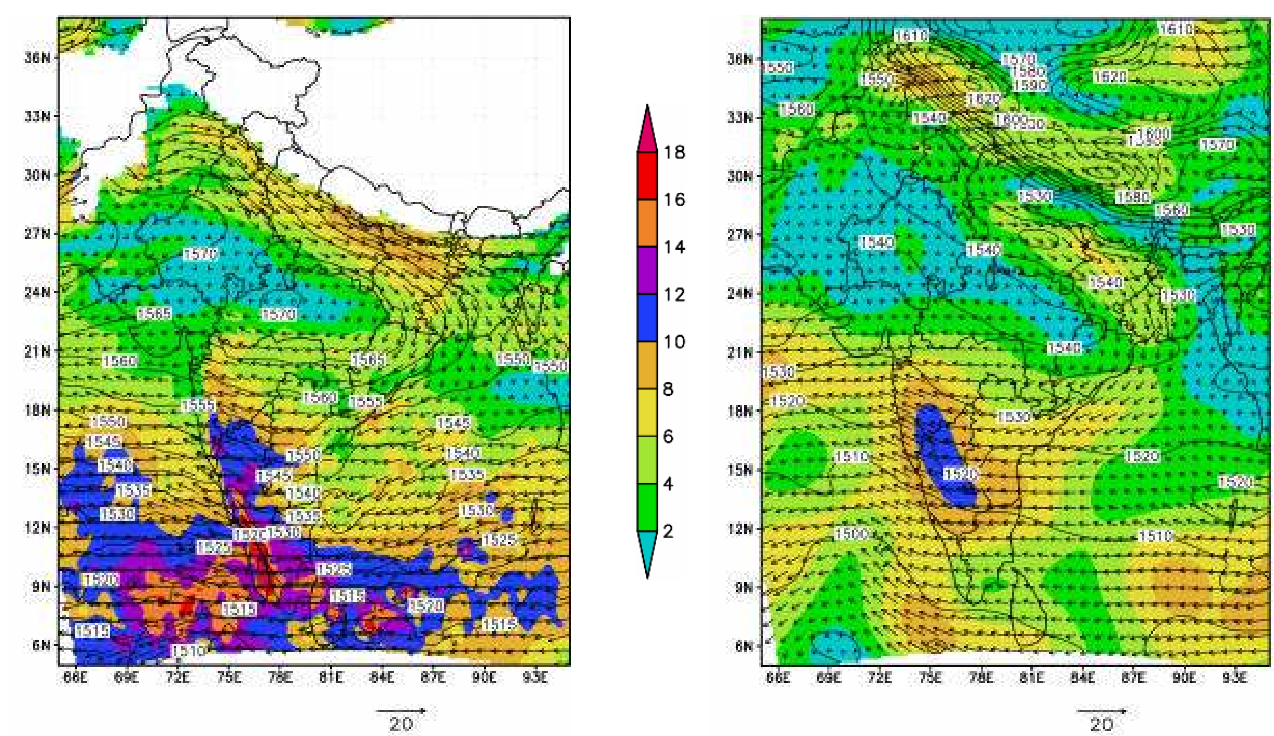

Figure 8. Geopotential height (m) and wind (ms^{-1}) at 850 hPa level at 00Z (06:00 IST) on 31 October 2004. **Left:** Simulation with NCMRWF data, **right:** NCMRWF analysis.

to 850 hPa) over the equatorial Indian Ocean, center of high pressure (1038 hPa) over Tibetan plateau, low-level easterly winds and upper level ($>$500 hPa) westerlies could be noticed in both the initial condition data from the two sources for 29 October, 2004 (figure 4). It is also seen that the

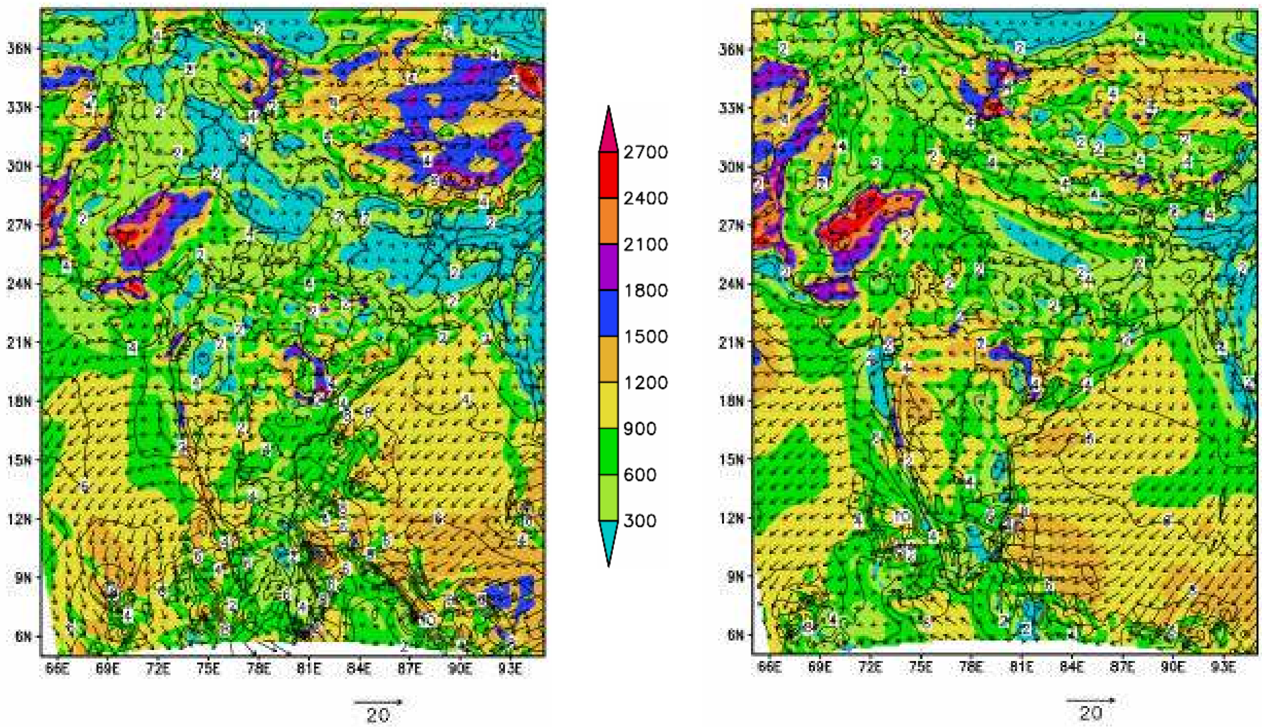

Figure 9. MM5 simulated surface wind (ms^{-1}) and PBL height (m) at 12 Z on 30 October 2004. **Left:** simulation with NCEP data, **right:** simulation with NCMRWF data.

simulated wind, temperature, geopotential height, relative humidity and PBL height have nearly similar patterns/values in the simulations with NCEP and NCMRWF data excepting a few differences (figures 5 to 9). It is seen that while the model geopotential height at 925 hpa and 850 hpa levels agree closely with the corresponding analyses in the case of simulation with NCEP data, it is over-predicted by about 30 m in the case of simulation with NCMRWF data. The simulated wind pattern is easterly/northeasterly in the lower levels (up to 850 hPa) and westerly above 500 hPa and agrees with the analyses. The model winds in both the cases of simulation are generally over-predicted in the southern part of the domain, and the differences are seen considerably more in the simulation with NCMRWF data.

Composite statistical skill scores based on the model results for the dry and wet weather cases are presented in table 3. The magnitudes and trends of correlation, RMSE and BIAS with respect to observations are almost similar in both the runs (table 3). MM5 runs with NCEP data have shown slightly better correlation, relatively smaller RMSE and BIAS than those with the NCMRWF data except in some cases such as temperature and wind forecasts at 36 hours, where the scores are relatively better for the runs with NCMRWF data. This is because of the fact that the NCMRWF analysis is at coarser resolution. Moreover, it may also be kept in mind that the validity of any statistical score is good only when the sample size is greater than or equal to 30 cases. The present scores are based on two cases (a dry and a wet weather condition).

It is known that the quality of the predicted variables p, t, U is superior to precipitation which exhibits higher spatial and temporal variability. A poor correlation and large RMSE error is seen in the predicted 24 h and 48 h accumulated rainfall in both the cases of model simulation probably due to the comparison of model grid (30 km) averaged data with point rainfall observations (tables 4 and 5). It is seen that the simulated rainfall using NCEP data is over-predicted during the first day (29 October) and a positive bias (8.08 cm) is seen in the analysis. It is under-predicted on the 2nd day (30 October) the bias being −11.0 cm. The rainfall is under-predicted on both the 1st and 2nd days in the simulation with NCMRWF data the bias scores being −23.7, −16.2 respectively. The model has considerably under-predicted rainfall with NCMRWF data (figure 10) compared to the simulation with NCEP data. This is seen from very low mean, high negative bias and large RMSE error in 24 h, 48 h accumulated rainfall simulated with NCMRWF data. This may be again due to the differences in the resolutions of the initial and boundary conditions. The threat scores for rainfall are seen to be higher with NCEP data for light and heavy rainfall events whereas they are found better

Table 3. *Composite statistical skill scores of MM5 model runs using NCEP/NCMRWF initial conditions for 00Z, 29 October, 2004 and 00Z, 2 February, 2005.*

		Runs with NCEP data			Runs with NCMRWF data		
Parameter	Level	r	RMSE	BIAS	r	RMSE	BIAS
24 h forecast							
t	925 hPa	0.525	3.067	−0.801	0.594	2.792	−0.960
	850 hPa	0.536	2.805	−1.475	0.405	3.389	−1.915
	500 hPa	0.827	2.495	0.628	0.840	2.473	0.077
	200 hPa	−0.183	8.869	−0.317	0.330	8.510	−1.050
h	925 hPa	0.707	23.836	16.754	0.730	29.285	25.405
	850 hPa	0.506	22.069	13.724	0.330	27.014	20.123
	500 hPa	0.761	46.798	15.345	0.750	48.630	15.294
	200 hPa	0.680	128.290	21.066	0.650	132.800	17.220
RH	925 hPa	0.461	23.089	7.972	0.293	26.329	11.383
	850 hPa	0.489	22.467	5.754	0.486	23.546	7.065
	500 hPa	0.861	17.918	−0.919	0.662	28.737	−10.229
UV	925 hPa	0.167	4.433	0.968	0.333	3.459	0.221
	850 hPa	0.365	3.316	0.398	0.154	3.709	−0.402
	500 hPa	0.819	5.133	−0.737	0.793	5.098	−0.020
	200 hPa	0.575	13.116	3.684	0.880	7.060	2.470
DIR	925 hPa	0.169	93.300	−13.700	0.004	104.575	9.915
	850 hPa	0.701	71.170	14.460	0.769	67.965	27.405
	500 hPa	0.311	69.540	−2.310	0.243	76.725	13.085
	200 hPa	0.942	30.780	−18.380	0.840	42.540	−10.660
36 h forecast							
t	925 hPa	0.495	4.602	−3.359	0.474	4.700	−3.454
	850 hPa	0.461	3.933	−1.633	0.461	3.933	−1.854
	500 hPa	0.327	4.368	1.302	0.336	4.450	0.066
	200 hPa	−0.136	3.740	1.496	−0.033	3.613	0.782
h	925 hPa	0.724	55.193	41.048	0.734	52.827	38.518
	850 hPa	0.548	40.446	31.238	0.446	40.485	27.744
	500 hPa	0.291	69.231	31.841	0.270	67.364	18.344
	200 hPa	0.776	133.359	64.029	0.877	102.702	29.192
RH	925 hPa	0.190	32.827	22.331	−0.070	35.227	22.708
	850 hPa	0.650	20.258	10.324	0.412	28.549	6.645
	500 hPa	0.242	40.555	6.632	0.255	34.284	−1.021
UV	925 hPa	0.505	3.701	1.434	0.459	3.385	0.897
	850 hPa	0.545	3.553	1.254	0.602	3.089	1.136
	500 hPa	0.592	6.267	−1.357	0.443	6.391	−0.021
	200 hPa	0.772	10.180	2.538	0.304	17.160	4.577
DIR	925 hPa	0.508	81.271	−1.161	0.558	82.645	12.540
	850 hPa	0.749	64.619	21.044	0.750	66.328	28.457
	500 hPa	0.413	80.418	−27.818	0.178	104.846	−22.149
	200 hPa	0.681	57.926	7.270	0.520	50.814	4.085

Note: t – temperature, h – geopotential height, RH – relative humidity, UV – wind speed, DIR–wind direction.

with NCMRWF data for no rain and moderate rain events (table 6). A comparison of the vertical structure of temperature, humidity and wind between the model simulation and radiosonde observation for the model grid at Chennai is presented in figure 11. While the temperature structure is almost identical in both the simulations, humidity and wind are seen better simulated with NCEP data, the differences between the model values and observation being less in the simulation with NCEP data. Thus some improvement is expected while using NCEP data in the high-resolution simulations for atmospheric dispersion studies.

From an examination of the initial conditions in the case of 7 February 2005, a low pressure (1008 hPa) is seen over Jammu & Kashmir in the NCEP analysis, which is not present in the NCMRWF data (figure 12). However, the high pressure over the Tibetan region, ridge of high pressure over the peninsular India, easterly surface winds

Table 4. *Observed and predicted rainfall for 29–31 October, 2004.*

Station	Measured Rainfall (in mm) 29–30 Oct'04	Measured Rainfall (in mm) 30–31 Oct'04	Simulated rainfall with NCEP data (in mm) 29–30 Oct'04	Simulated rainfall with NCEP data (in mm) 30–31 Oct'04	Simulated rainfall with NCMRWF data (in mm) 29–30 Oct'04	Simulated rainfall with NCMRWF data (in mm) 30–31 Oct'04
Adiramapattinam	53.0	13.00	76.35	41.06	2.59	2.04
Coimbattore	4.0	2.00	1.57	4.32	3.64	2.66
Coonoor	30.0	19.00	3.99	0.00	0.54	0.00
Cuddalore	40.0	50.00	37.07	13.98	0.61	3.65
Chennai	51.0	46.00	151.48	3.64	68.71	7.59
Kanyakumari	32.0	0.00	12.22	20.24	0.80	54.22
Kodaikanal	6.0	2.00	36.96	14.66	18.71	12.88
Madhurai	9.0	2.50	103.46	23.29	7.46	13.10
Nagapattinam	155.0	312.00	23.67	19.85	8.44	1.53
Palayamkottai	13.00	0.00	39.80	6.46	6.40	31.21
Pamban	2.00	4.00	11.86	15.92	0.02	2.69
Pondicherry	45.00	40.00	86.45	15.12	1.83	3.48
Salem	22.00	9.00	18.81	6.73	0.20	0.11
Tiruchhirapalli	16.00	18.00	33.98	14.85	0.05	12.26
Tirupattur	18.00	4.00	13.85	3.99	15.99	2.97
Tondi	28.00	15.00	152.44	61.67	1.28	4.24
Tuticorin	17.00	0.00	92.61	4.07	2.27	12.23
Ooty	5.00	0.00	7.67	6.27	17.60	1.39
Valaparai	2.00	0.00	14.64	1.65	20.74	2.05
Vedaranyam	167.00	2.00	33.16	28.74	0.27	0.35
Vellore	10.00	13.00	18.37	2.72	3.51	19.83
Trivandrum	24.00	4.00	4.09	12.29	0.09	0.73
Minicoy	24.00	19.00	0.01	1.53	0.00	1.77
Cochin	2.00	8.00	0.17	0.00	0.01	0.12
Alappuzha	0.00	17.00	2.44	0.00	0.03	0.00

Table 5. *Statistical skill scores of the MM5 model runs for rainfall using NCEP/NCMRWF initial conditions for 00Z, 29 October 2004.*

Total rainfall for	Simulation with NCEP data					Simulation with NCMRWF data			
	Mean of observed	Mean	r	RMSE	BIAS	Mean	r	RMSE	BIAS
24 h	31.00	39.08	0.123	57.29	8.08	7.27	0.022	49.384	−23.728
2nd day 24 h	23.98	12.92	0.096	61.70	−11.0	7.72	−0.15	65.442	−16.256
48 h	55.06	52.06	0.115	101.10	−3.05	14.99	−0.056	103.11	−40.065

Note: r – correlation coefficient; RMS – root mean square error; BIAS – bias error.

on east coast, and westerly surface winds over the northwestern parts are seen in both the data sets. The winds are anticyclonic in the lower levels up to 850 hPa and are strong westerly above 500 hPa (not shown). A few differences are noticed in the patterns of the weather in the simulations with NCEP and NCMRWF data sets. A large difference (about 6 ms^{-1}) of surface winds is seen at 1800 IST 8 February 2005 over the southern peninsula and adjoining oceans in the simulation with NCMRWF data (figure 14). The diverging wind pattern over central Arabian Sea could be more clearly seen in simulation with MM5 and corresponding NCEP analyses data than in the simulation with NCMRWF data and its corresponding analyses. Except this, the results for winds, geopotential height, temperature and relative humidity at 925, 850, 500 and 200 hPa levels do not have much difference (figures 13 and 14). It is also seen that the simulated parameters agree with the analyses closely in both the cases of simulation. Hence, for the dry weather case model results are not much different while using either of the data sets in simulation.

Thus, there exists only a little difference in the model simulation when initial data from NCEP and NCMRWF are used. The difference is

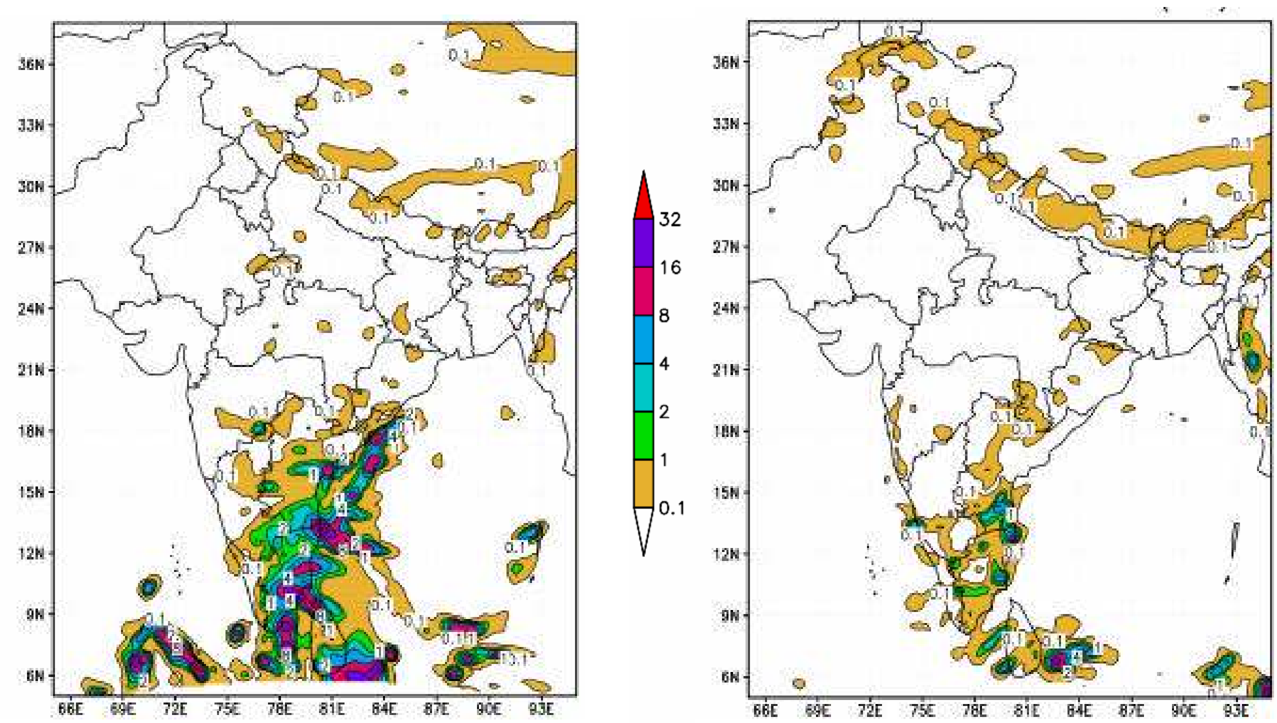

Figure 10. Simulated 24 h accumulated rainfall (in cm) (29–30) October 2004. **Left:** simulation with NCEP data, **right:** simulation with NCMRWF analysis.

Table 6. *Threat scores of the predicted rainfall for 29–31 October, 2004.*

Rainfall classification	Precipitation thresholds(mm)	Threat scores for prediction using NCEP data		Threat scores for prediction using NCMRWF data	
		24 h rainfall	48 h rainfall	24 h rainfall	48 h rainfall
Rain/no-rain	0.0–5.0	0.222	0.00	0.167	0.294
Light	5.0–10.0	0.000	0.25	0.200	0.000
Moderate	10.0–20.0	0.111	0.00	0.167	0.200
Heavy	20.0–200.0	0.412	0.45	0.077	0.000

perceptible to a little extent mainly in the rainfall pattern, geopotential height and winds where the NCEP data seem to give a better initial condition. The results are too meager to make any reasonable distinction between the two data sets and require intensive inter-comparison and validation with a large field data for different weather scenarios.

6. Conclusion

The implementation of an operational forecast model for real-time prediction of the radiological plume dispersion on a high-performance cluster computing network has brought out interesting results on drastic reduction in runtime, choice of cost effective cluster techniques, experience in handling RSL libraries and MPI interfaces and an insight into the forecast results. A comparison of model runtime in different platforms was made including the Kabru super-computing cluster at IMSc, Chennai. Results indicate that a reduction of about 8 and 5 times in runtime is achieved for the parallel benchmark MM5 and realistic parallel MM5 cases respectively with 18 Xeon CPUs on a distributed memory cluster when compared with sequential run on a single Xeon CPU.

In order to make the numerical system operational, unhindered availability of the source of meteorological data for initial and boundary conditions is imperative. In this regard, the model was run with NCEP data set as well as NCMRWF data sets. An inter comparison of the results using both the initial data sets showed a few differences in simulated parameters such as wind pattern, geopotential height and rainfall. Although these differences are not significant they are important in weather generating processes

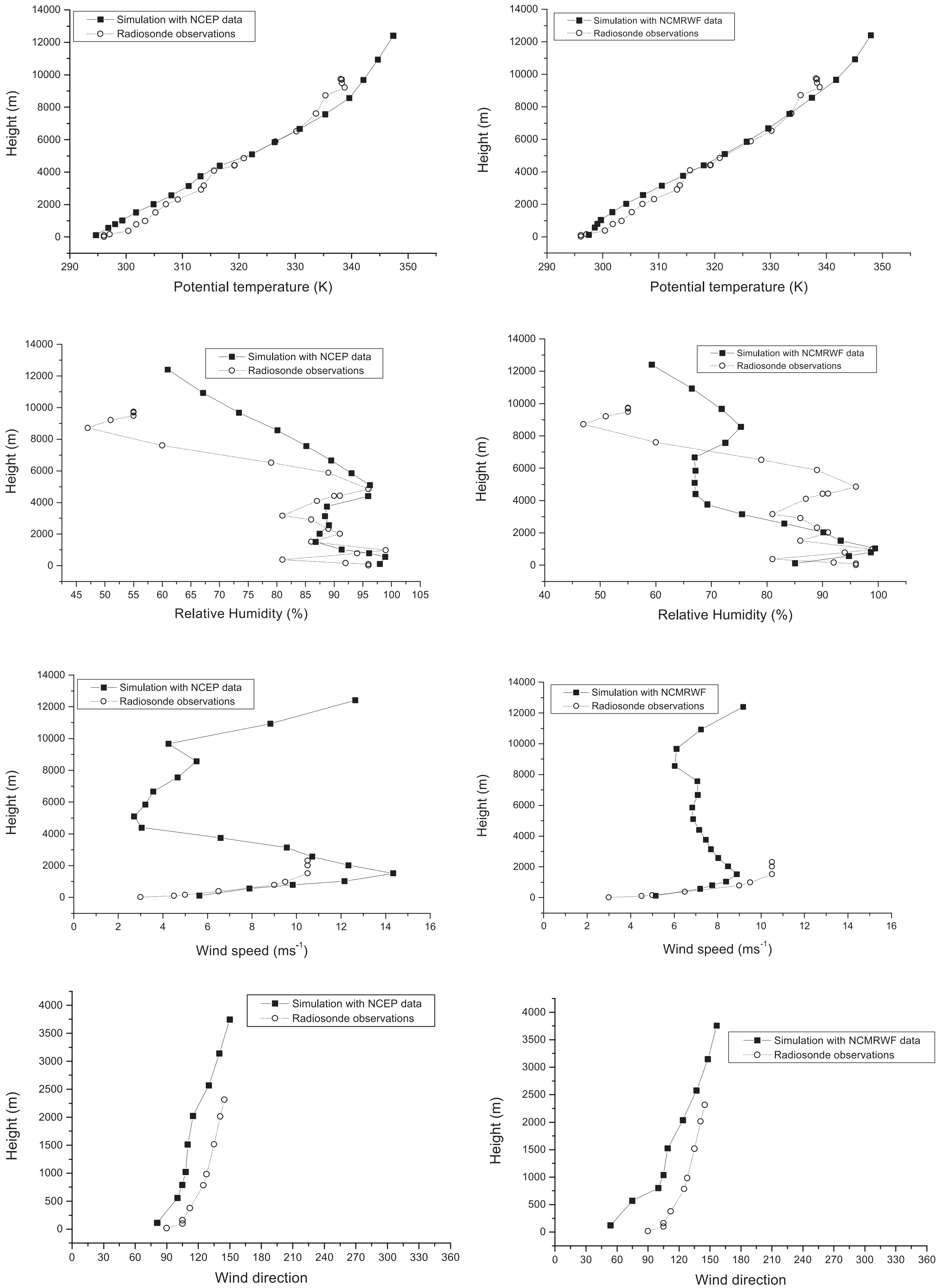

Figure 11. Comparison of vertical profiles of **(a)** potential temperature, **(b)** relative humidity, **(c)** wind speed and **(d)** wind direction between the model grid values at Chennai and radiosonde observations.

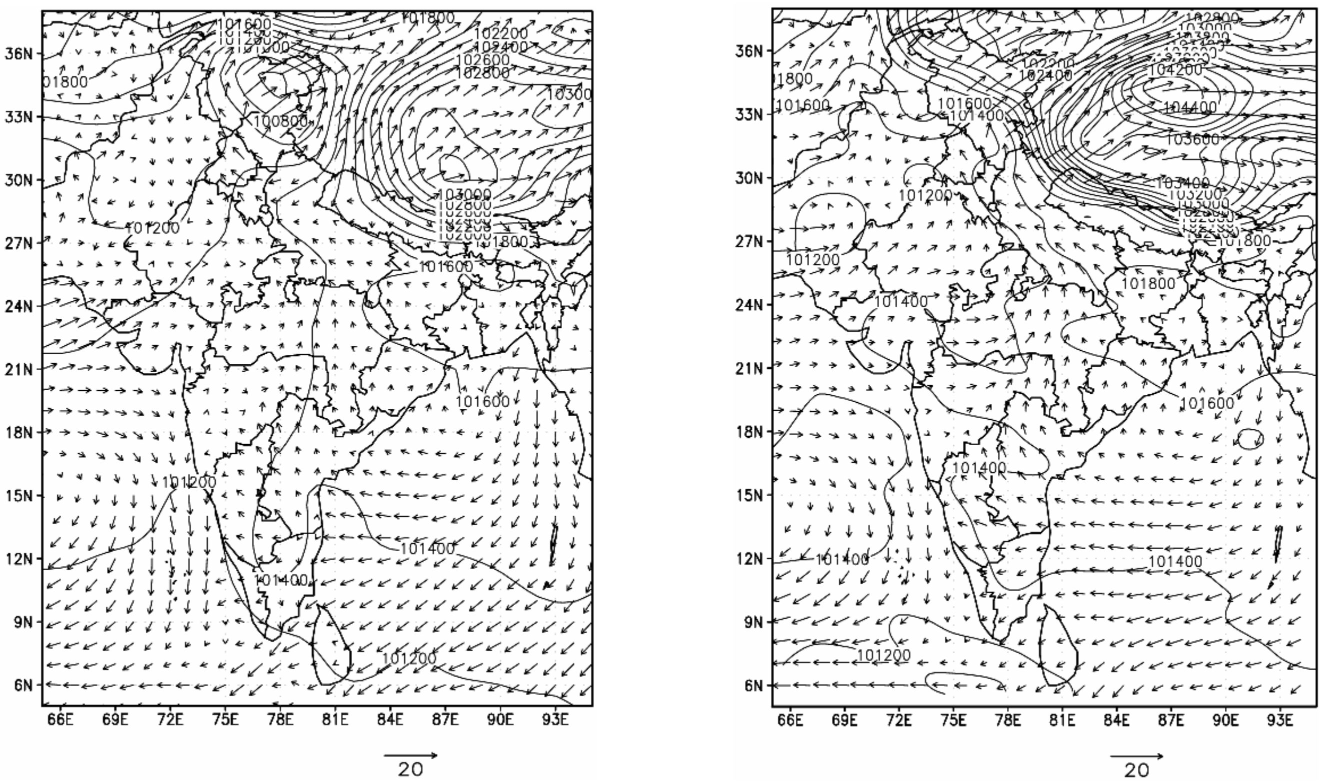

Figure 12. Sea level pressure and surface wind at 00Z on 07 February 2005 (**Left:** NCEP data, **right:** NCMRWF data).

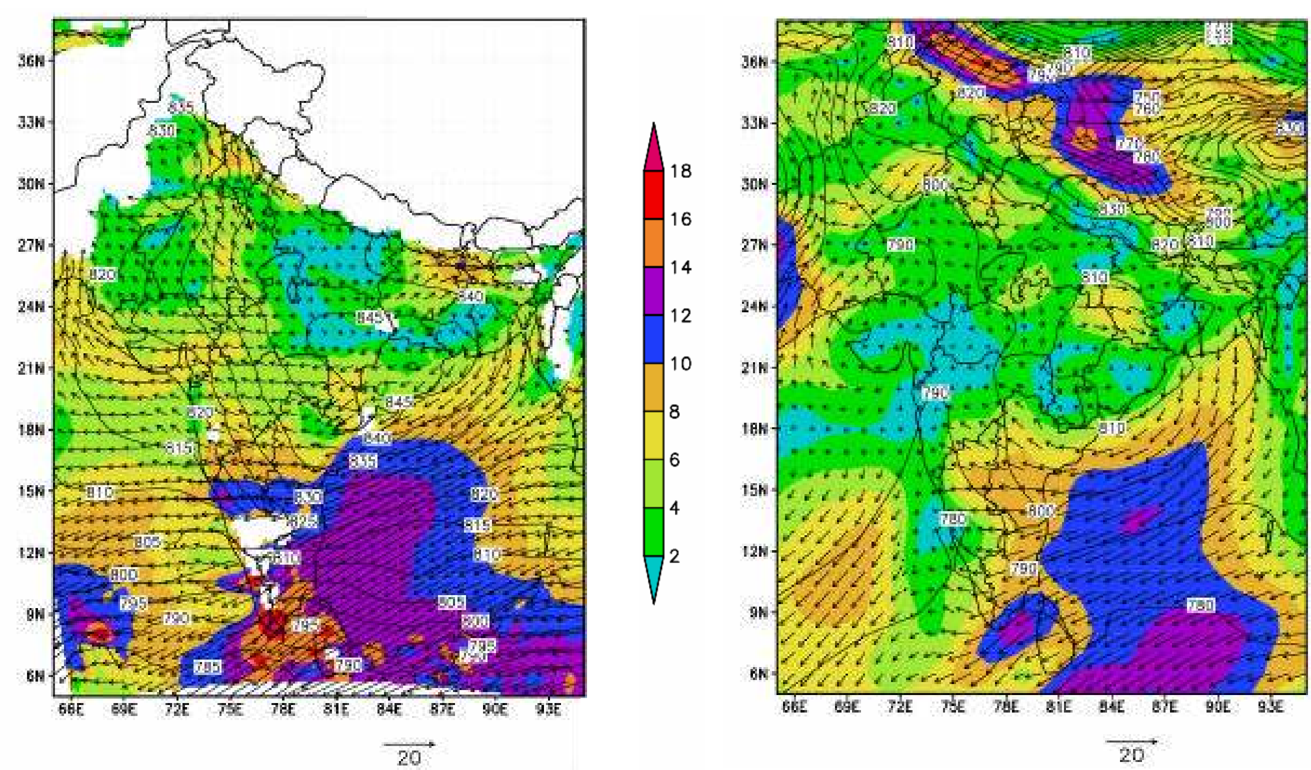

Figure 13. Geopotential heights (m) and winds (ms^{-1}) at 925 hPa at 12 Z (18:00 IST) on 8 February 2005. **Left:** Simulation with NCEP data, **right:** NCEP analysis.

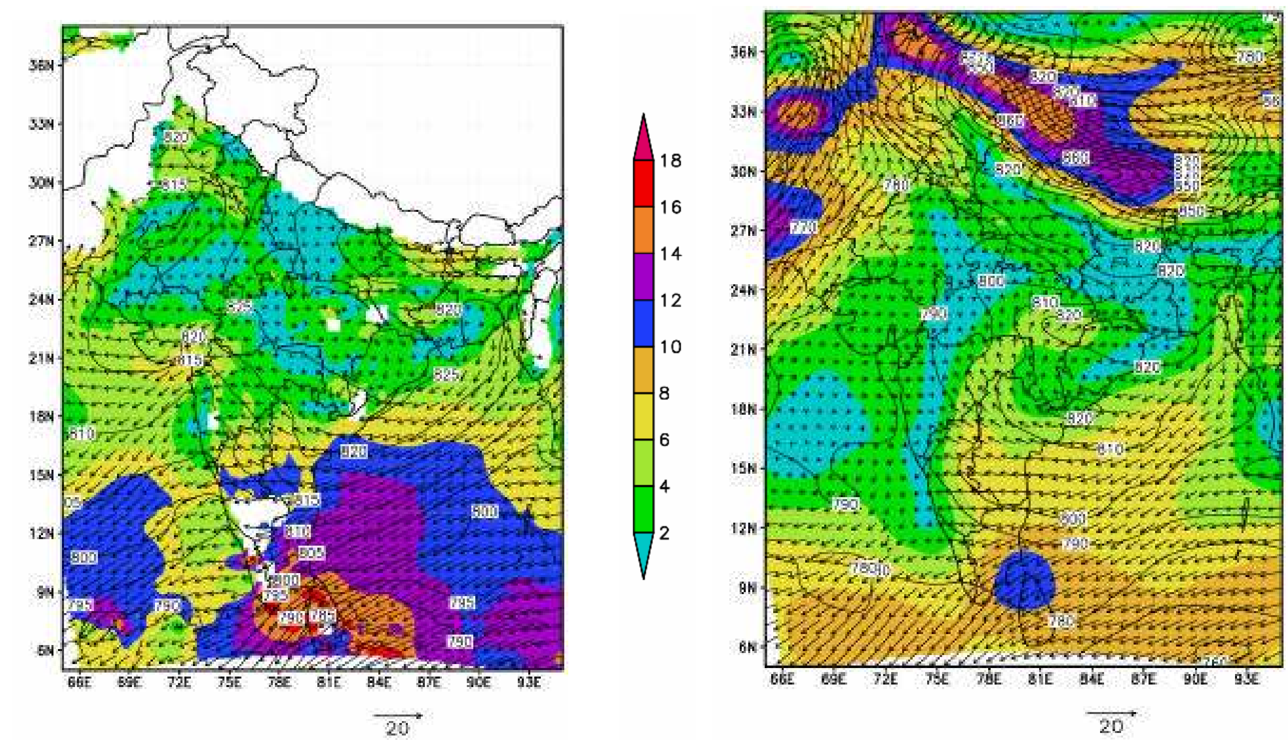

Figure 14. Geopotential heights (m) and winds (ms^{-1}) at 925 hPa at 12 Z (18:00 IST) on 8 February 2005. **Left:** Simulation with NCMRWF data, **right:** NCMRWF analysis.

and grow in numerical simulations. Results show some improvement in rainfall and wind pattern while using the high spatial resolution NCEP data. However, it is too early to come to a reasonable conclusion about the differences with the limited number of cases studied. Nevertheless, this pilot study reveals the requirement of applying high spatio temporal meteorological analyses data for meso-scale weather forecasting applications such as air-pollution modeling and cyclone/rainfall simulation where some differences are noticed in the results while using NCEP and NCMRWF data sets.

The combined weather and dispersion code was executed continuously for three months in the Scali pilot cluster. The code system is found to be robust and suitable for operational use. Improvements based on detailed validation study is nonetheless required for reliable forecast, particularly for the intended dispersion prediction during radiological emergencies.

Acknowledgements

This work has been carried out as a part of the commitments of the Task Force set up by the Chairman DAE for the development of a live tool for assessment/forecast of atmospheric dispersion. The authors thank Dr. Baldev Raj, Director, IGCAR, and Dr. A K Bohra, and Dr. L S Rathore, Heads of Divisions, NCMRWF for their kind support and encouragement in carrying out this work. Authors also thank Shri S Govindarajan, Director, Safety Group, Dr. R Indira, Dr. A Natarajan, and K M Somayaji for their valuable guidance and suggestions.

References

Anthes R A, Kuo Y H, Hsic E Y, Low-Nam S and Bettge T W 1989 Estimation of skill and uncertainty in regional numerical models; *Q. J. R. Meteorol. Soc.* **115** 763–806.

Baker M and Buyya R 1999 Cluster computing at a glance; In: *High Performance Cluster Computing*, pp. 3–48.

Burk S D and Thompson W T 1989 A vertically nested regional numerical prediction model with second-order closure physics; *Mon. Weather Rev.* **117** 2305–2324.

Dudhia J 1989 Numerical study of convection observed during winter monsoon experiment using a mesoscale two-dimensional model; *J. Atmos. Sci.* **46** 3077–3107.

Grell G A, Dudhia J and Stauffer D R 1994 A description of the fifth-generation Penn State/NCAR mesoscale model (MM5), NCAR Technical Note, NCAR/TN-398+STR, p. 117.

Internet source : Description of the MM5 Mesoscale Model as implemented at NCMRWF http://www.ncmrwf.gov.in/mm5-new-web-des.htm

Internet source: http://www.imsc.res.in/~kabru/ KABRU 288-CPU Linux Cluster @ IMSC

Kalnay, Kanamitsu M and Baker W E 1990 Global numerical weather prediction at the National Meteorological Center; *Bull. Amer. Meteor. Soc.* **71** 1410–1420.

Kanamitsu M 1989 Description of the NMC global data assimilation and forecast system; *Weather and Forecasting* **4** 335–342.

Michalakes J 1997 Same-Source Parallel Implementation of the PSU/NCAR MM5. Proceedings of the Symposium on Regional Weather Prediction on Parallel Computer Environments, University of Athens, Greece, October 15–17, 1997; Proceedings paper (preprint ANL/MCS-P702-1297).

Savarese D F, Sterling T 1999 Beowulf; In: *High Performance Cluster Computing*; (ed.) Buyya, Rajakumar, pp. 625–645.

Srinivas C V, Venkatesan R 2005 A simulation study of dispersion of airborne radionuclides from a nuclear power plant under a hypothetical accidental scenario at a tropical coastal site; *Atmos. Environ.* **39** 1497–1511.

Srinivas C V, Venkatesan R, Bagavath Singh A and Somayaji K M 2004 A real case simulation of the air-borne effluent dispersion on a typical summer day under CDA scenario for PFBR using an advanced meteorological and dispersion model; Research Report, IGC-259, p. 82.

Srinivas C V, Venkatesan R and Bagavath Singh A 2005 A numerical study of the influence of synoptic flow on coastal meso-scale circulations on the East and West coasts of India; *Mausam* **56(1)** 73–82.

Stohl A 1999 The FLEXPART Particle Dispersion Model Version 3.1 User Guide, Lehrstuhl fur Bioklimatologie and Immissionsforschung, University of Munich, AmHochanger 13, 85354 Freising, Germany.

Wang K Y, Shallcross D E, Hall S M, Lo Y H, Chou C and Chen D 2005 DOBSON: a Pentium-based SMP Linux PC Beowulf for distributed-memory high resolution environment modeling; *Environmental Modeling & Software* **20** 1299–1306.

MS received 10 August 2005; revised 29 September 2005; accepted 8 November 2005