NMRF/TR/02/2016

TECHNICAL REPORT

Implementation of Unified Model based Ensemble Prediction System at NCMRWF (NEPS)

Abhijit Sarkar, Paromita Chakraborty,
John. P. George and E.N. Rajagopal

March 2016

National Centre for Medium Range Weather Forecasting
Ministry of Earth Sciences, Government of India
A-50, Sector-62, NOIDA-201309, INDIA

Implementation of Unified Model based Ensemble Prediction System at NCMRWF (NEPS)

Abhijit Sarkar, Paromita Chakraborty,
John P. George and E. N. Rajagopal

March 2016

National Centre for Medium Range Weather Forecasting
Ministry of Earth Sciences
A-50, Sector 62, NOIDA-201309, INDIA

Ministry of Earth Sciences
National Centre for Medium Range Weather Forecasting
Document Control Data Sheet

1	Name of the Institute	National Centre for Medium Range Weather Forecasting
2	Document Number	NMRF/TR/02/2016
3	Date of publication	March 2016
4	Title of the document	Implementation of Unified Model Based Ensemble Prediction System at NCMRWF (NEPS)
5	Type of Document	Technical Report
6	No. of pages & Figures	26 & 11
7	Number of References	28
8	Author (s)	Abhijit Sarkar, Paromita Chakraborty, John P. George and E. N. Rajagopal
9	Originating Unit	NCMRWF
10	Abstract	Ensemble forecasting has proved to be a successful way of dealing with the inherent uncertainties of weather and climate forecasts. A Unified Model based 45 members (44 + 1 control) Global Ensemble Prediction System with horizontal resolution of ~33 km and 70 vertical levels is implemented at NCMRWF in Bhaskara HPC. This ensemble prediction system (NEPS) is a recent version of Met office Global and Regional Ensemble Prediction System (MOGREPS). The initial condition perturbations are generated by Ensemble Transform Kalman Filter (ETKF) method. The model uncertainties are taken care by the Stochastic Kinetic Energy Backscatter and Random parameters schemes. The forecast perturbations obtained from 6-hr short forecast run of 45 members are updated by ETKF four times a day (00, 06, 12 and 18 UTC). A 10 day forecast is generated everyday based on 00 UTC initial conditions. This report describes various components of NEPS system at NCMRWF. It also presents a brief description of different ensemble forecast products.
11	Security classification	Non-Secure
12	Distribution	Unrestricted Distribution
13	Key Words	Ensemble, Perturbation, Data Assimilation, NEPS

Contents

Abstract

Ensemble forecasting has proved to be a successful way of dealing with the inherent uncertainties of weather and climate forecasts. A Unified Model based 45 members (44 + 1 control) Global Ensemble Prediction System with horizontal resolution of ~33 km and 70 vertical levels is implemented at NCMRWF in Bhaskara HPC. This ensemble prediction system (NEPS) is a recent version of Met office Global and regional ensemble forecasting system (MOGREPS). The initial condition perturbations are generated by Ensemble Transform Kalman Filter (ETKF) method. The model uncertainties are taken care by the Stochastic Kinetic Energy Backscatter and Random Parameters schemes. The forecast perturbations obtained from 6-hr short forecast run of 45 members are updated by ETKF four times a day (00, 06, 12 and 18 UTC). A 10 day forecast is prepared everyday based on 00 UTC initial conditions. This report describes various components of NEPS system at NCMRWF. It also presents a brief description of different ensemble forecast products.

1. Introduction

Initial approaches of Numerical Weather Prediction (NWP) were deterministic. In the beginning of 1950s, under the guidance of John von Neumann and with Institutional support of Princeton's Institute for Advance study, Jule Charney and his team of researchers made first successful 24 hours forecasts of the transient features of large scale atmospheric flow by advecting geostrophic vorticity with the geostrophic wind (Charney et al. 1950). By the late 1950s there was hope that the prediction beyond several days would be possible by running a single NWP model. Despite the fact that the NWP results were encouraging, the limit of deterministic prediction became a matter of concern. In 1960s it became known that the presence of uncertainties in the estimation of initial condition and formulation of the model and the fact that the atmosphere and its numerical model are chaotic, placed a limit to the predictability of the system. Initial condition errors are due to inaccuracies in the estimation of initial condition of the model, which then can grow with forecast lead time. The inaccuracies in the representation of dynamical and physical processes of the atmosphere in the model account for the model errors. Since model error influences the estimation of model initial state also it is not possible to quantify the individual contributions of these errors. Eady (1951) was first to express his concern about the strictly deterministic approach in NWP and advocated for probabilistic approach. The practical implementation of the approach that combines probability with determinism is called Ensemble Prediction.

Lorentz (1965) and Epstein (1969) brought the idea of ensemble forecasting in numerical weather prediction. Leith (1974) implemented the idea of ensemble forecasting with random perturbations (Monte Carlo forecasting) in perfect model environment. According to him if the perturbations correctly represent the uncertainties in the initial condition, ensemble forecasting, even with a small number of ensemble members, can become useful. An excellent review article by Lewis (2005) presents historical details of the early research on predictability and ensemble forecasting. Ensemble prediction systems were first implemented operationally early in the 1990s in European Centre for Medium-range Weather Forecast (ECMWF) (Palmer et al., 1993; Molteni et al., 1996); the US National Centre for Environmental Prediction (NCEP) (Toth&Kalnay,1993; and the Rechercheen Prévision Numérique (RPN) in Canada (Houtekamer et al., 1996).

An ensemble prediction system usually includes a control forecast and a good number of perturbed forecasts. The control forecast is one that starts from the best estimated state (based on available observations) of the atmosphere (analysis) prepared by the data assimilation system. Initial conditions for other ensemble members are generated by adding perturbations (or errors) to the analysis. During the early stage of the forecast, error grows more or less linearly with time and the deterministic forecast shows good skill. During this period the small error in the initial condition remains small and trajectories of the model forecast and the 'truth' are close to each other in phase space (Buizza, 2000). Beyond this range of linear error growth, deterministic forecast loses its skill but ensemble mean (or average) can be treated as a single forecast representing the best available estimate of the future atmosphere. By calculating the ensemble average the unpredictable components of the forecast are filtered out and those are retained that show agreement between the ensemble members. This filtering takes place within the nonlinear evolution of the perturbation. During linear regime ensemble average forecast is no better than control forecast. Another important use of ensemble forecasting is that it provides an indication of the reliability of the forecast. Spread in the forecast is a measure of disagreement between the ensemble members. A good agreement among the members results in less spread and a good reason to become confident about the forecast. The third important aspect of ensemble prediction is that it provides a quantitative basis for probabilistic forecasting.

Ensemble forecasting methods in different operational centres around the world mostly differ by the way in which initial condition perturbations are generated. The simplest way to generate perturbations is to add random (Monte Carlo) noise to the original analysis. However, Hollingworth (1980), Hoffman and Kalnay (1983) and Kalnay and Toth (1996) showed that the real analysis errors grow much faster than the random initial perturbations. By construction, perturbations generated by Monte Carlo method do not include the "growing errors of the day". A second class of methods that take care of growing errors in the initial perturbations were developed, tested and implemented at various operational centres around the world. "Breeding" and "singular vector" methods of perturbation generation lie in this class. Breeding vectors (BVs) (Toth and Kalnay, 1993) are used to generate perturbations to the initial condition at NCEP and the singular vector (SV) approach is used at ECMWF (Buizza an Palmer, 1995; Molteni et al., 1996). In Met Office, UK, Ensemble Transform Kalman Filter (ETKF) (Bishop et al., 2001) is used in its Global and Regional Ensemble Prediction System (MOGREPS) to generate initial perturbations. This method is similar to the error breeding method (prescribed by Toth and Kalnay, 1993) with some differences as shown in the Fig. 1(a). In ETKF (Figure 1(b)), the analysis perturbation of each member is the linear combination of the forecast perturbations. This mixing of forecast perturbations which produces mutually orthogonal analysis perturbations leads to improved performance of ETKF over the error breeding method (Wang et al., 2004).

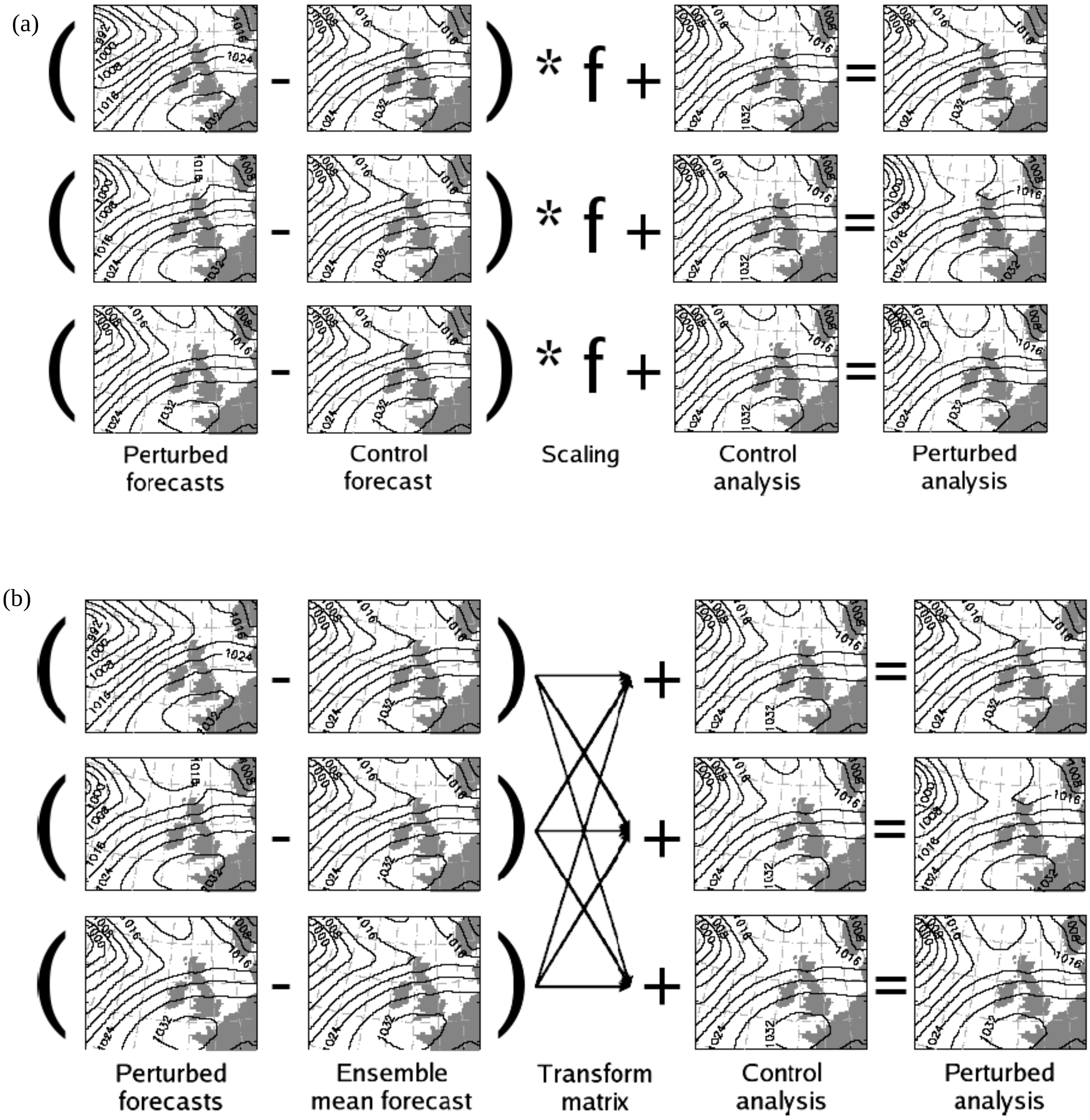

Figure 1: Graphical representation of the (a) error-breeding method (b) ETKF (courtesy Bowler et al., 2008)

In NCMRWF, the global version of MOGREPS has been implemented for operational ensemble prediction. This Unified Model (UM) based ensemble prediction system at NCMRWF (NEPS) also uses ETKF for generation of initial perturbations. Model uncertainties are also taken care in this MOGREPS based system since the forecast uses stochastic physics schemes that consist of “random parameters” (Bright and Mullen, 2002) and “Stochastic Kinetic Energy Backscatter” schemes (Tennant et al., 2011). The random parameter (RP) scheme incorporates uncertainties in the empirical parameters of the physical parameterization schemes. It also simulates the non-deterministic processes not explicitly accounted for by different parameterizations. In real atmosphere, energy is up-scaled from the small to large-scale flow through physical processes. It is very difficult to include this energy transfer in a numerical weather prediction model. This results in a loss of kinetic energy from the model environment. Moreover, the semi-Lagrangian advection scheme used in UM involves interpolation of prognostic field to the departure point and it acts to smooth field and remove energy. Also, use of horizontal diffusion terms to smooth model fields lead to excessive energy dissipation. The Stochastic Kinetic Energy Backscatter Scheme (SKEB2) is implemented in UM to inject the loss in kinetic energy back into the model (Tennant et al., 2011)

2. Methodology

2.1 Brief Description of NEPS

The NEPS implemented in NCMRWF is fundamentally the same as the original MOGREPS developed at Met Office, UK. This global ensemble prediction system has a horizontal resolution of approximately 33 km and 70 vertical levels (N400L70). A total of 45 ensemble members (44 perturbed forecasts and 1 control forecast) constitute this ensemble system. The 44 analysis perturbations for all the ensemble members are generated by ETKF system four times a day (at 00, 06, 12 and 18 UTC) from the previous 6 hr short forecast of the evolved perturbations for the variables u, v, θ, q and exner pressure on all levels. These analysis perturbations are added to the reconfigured analysis from the four–dimensional variational data assimilation system (4D-VAR) of Unified Model (version 8.5) operational at NCMRWF (NCUM). A 10 day forecast of NEPS is routinely generated based on 00 UTC initial conditions which include a control forecast (Cntl) with the 4D-VAR analysis and 44 ensemble member forecasts (Ens) with 44 perturbed initial conditions. The sequences of all the processes involved in NEPS operational at NCMRWF are represented by the flow diagram shown in Figure 2.

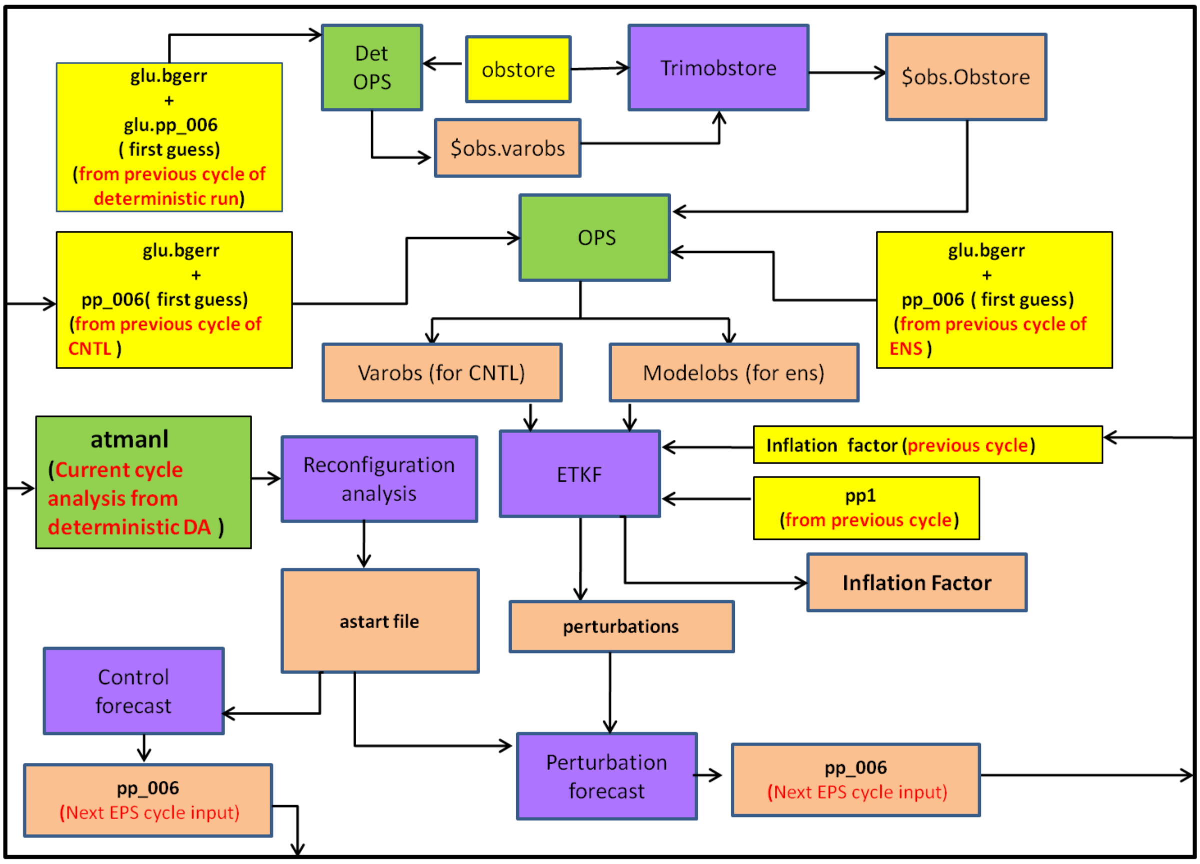

Figure 2: NEPS Flow-diagram showing short forecast cycle with ETKF

2.2 Computational infrastructure

NEPS is implemented and run on Bhaskara HPC at NCMRWF. This HPC has1052iDataPlex dx360 M4 compute nodes each configured with 16 cores of Intel Sandy bridge processors clocking at 2.6 GHz, with 64 GB DDR3 1600MHz RAM per Compute node. It is capable of delivering 350 Teraflops of peak computing power. Number of processors used and the wall clock time taken by each component are given in Table 1.

Table 1: Performance of various NEPS components in Bhaskara HPC

NEPS Components	Forecast length (hrs)	Nodes per member	Total no of Processors	Cycles (UTC)	Wall clock time
Trimobstore	6	1	1	00, 06, 12, 18	5 min
OPS cntl & ens	6	25	25x16x23 25x16x22	-do-	20 min
ETKF	6	1	1	-do-	24 min
Short forecast Cntl &Ens	6	10	10x16x45	-do-	10 min
Long forecast Cntl &Ens	240	10	10x16x45	00	3 hrs 10 min

2.3 Description of NEPS components

Various components of NEPS system are briefly described below.

2.3.1 TRIMOBSTORE

The “obstore” files are the observation files (observations stored in a UK Met Office specific format known as “obstore”). Each obstore file contains one type of observation (like Surface.obstore, Sonde.obstore, Aircraft.obstore, ATOVS.obstore etc.). The description of various types of observations (“obstore” files) that are used in NEPS is given in Table 2. The Observation Processing System (OPS) of deterministic forecasting system reads the observations from the obstore files. It also reads the model background files from deterministic UM short forecast and an OPS background error file. The processing of observation job is done by “extract and process” component of OPS. The “extract” task retrieves the observations from the obstore files and the task “process” carry out the jobs of quality control, thinning and rewriting the data in required formats. There are mainly three data structures in which the observation and model data processed by OPS are written. The three types of output from the OPS (“extract and process”) are: at (1) varobs (quality controlled observations), (2) varcx (horizontally interpolated background fields at observation location for use in 4D-VAR) and (3) Modelobs (Background fields exactly like observation fields at observation locations).

Only the quality controlled and thinned observations from the “obstore files” are written in varobs files by OPS (using the model background information). In NEPS, 1 control and 44 ensemble members have to process (since OPS uses Model background information as well) these “obstore” files to generate “varobs” and “modelobs” files respectively. So it is essential to remove the unwanted observations from obstore files in order to speed up the process of preparing “modelobs” and “varobs” files by the OPS task of NEPS system. The “trimobstore” program is employed to trim the obstore data sets. It reads both the “varobs” file (generated by OPS task of deterministic forecast system using deterministic forecast background) and the original “obstore” file of each observation type, and writes out new “obstore” file which contains observations for locations present in “varobs” files. These new obstore files are then used as input to the OPS task of NEPS to generate modelobs and varobs files. The tasks of “trimobstore” are summarised below:

(a) It reads the input obstore files.
(b) It reads the input varobs files generated by deterministic forecasting system.
(c) Selects only those observations from the obstore files that are present in varobs
(d) Write the trimmed observation in a new obstore file.

Table 2: Description of the observations used in data assimilation system of NEPS

Sl. No.	Observation Type (“obstore” file name)	Brief description of the various sub-types of observations included in a observation type
1.	Aircraft	Aircraft-based observations reported by the Aircraft Meteorological Data Reporting (AMDAR) system and aircraft reports (AIREP)
2.	Sonde	Radiosonde, wind profiler, dropsonde and Indian DWR VAD/VPP wind observations
3.	Surface	Surface based observations at or near the earth’s surface: Land surface (SYNOPS), Mobile SYNOP, METAR, Ships, BUOY
4.	Satwind	Atmospheric wind observations (AMV) from geostationary and polar orbiting satellites: Meteosat-7, Meteosat-9, GOES-E, GOES-W, MTSAT-1R,MODIS (TERRA and AQUA), NOAA and MetOp
5.	Scatwind	Sea surface wind observations: ASCAT winds from MetOp satellites,
6.	GPSRO	Radio occultation observations from various satellites.
7.	GOESClear	GOES Imager radiances from GOES-E & W
8.	ATOVS	Advanced Television Infrared Observation Satellite (TIROS) Operational Vertical Sounder (ATOVS) observations from various NOAA and MetOp satellites.
9.	AIRS	Atmospheric Infrared Sounder observations from AQUA satellite
10.	IASI	Infrared Atmospheric Sounding Interferometer observations MetOp satellites

2.3.2 Observation Processing System (OPS)

OPS (version 30.1) task of NEPS is run to produce varobs files for control member and the modelobs files for all the 44 ensemble members. The modelobs files contain the model forecast of the observations. ETKF does not need to have the observation operator as the modelobs files produced by OPS are already available to it. The trimmed obstore files produced by trimobstore program are used as input to the OPS task. The OPS tasks for the control and ensemble members are run in parallel to process 10 types of observations. The number of processors employed to complete the processing of different types of observations depends on the volume of data contained in the obstore files. A list of number of processors allotted for each type of observation is given in Table 3.

Table 3: Number of processors allotted for processing different types of observations in OPS of operational NEPS system

Serial No.	Observation Type	No. of processors
11.	Aircraft	8
12.	Sonde	8
13.	Surface	8
14.	Satwind	8
15.	Scatwind	8
16.	GPSRO	8
17.	GOESClear	32
18.	ATOVS	64
19.	AIRS	128
20.	IASI	128

OPS carries out quality control of the observation which includes internal consistency checks, checks against model background and checks against neighbouring observations. The processed observations are written in “varobs” files. Each “obstore” file has a corresponding “varobs” file. The background (first guess) processing is also a part of the OPS “extract and process”. When background field (first guess) processing is carried out in “extract and process” along with the observation processing, resulting columns of model data are interpolated horizontally to the observation location for the data assimilation system as “varcx” files. The “modelobs” files also contain the model background (forecast) interpolated to observation location, but exactly similar to observations. In the ETKF system, only modelobs (not varcx) files are used.

OPS also generates a “background error” using the “Background Error Create” component (glu.bgerr created by OPS in deterministic analysis process). These geographically varying model errors are determined using model forecast (background or first guess) tendency, model forecast gradient and background wind speed information taken from the background file. A detailed description of the OPS system at NCMRWF is given in Rajagopal et al. (2012) and George et al. (2016).

The calculation of transformation matrix in ETKF requires the model equivalent of each observation for every ensemble member. Successful completion of OPS task for each ensemble members provides these ‘pseudo observations’ in the form of “modelobs” files as input to ETKF. The “modelobs” and varobs files are also required for the ETKF task. Main inputs and outputs of OPS in the NEPS system are described below.

Main Inputs to OPS:

i) Observations: Observation files in obstore format (obtained from trimobstore).
ii) Model background: Three hourly model forecast fields of the same ensemble member over the assimilation cycle
iii) Background error (OPS Background error): Prepared by OPS task based on model forecast (background) and the previous cycle’s background error file
iv) Fixed files: RTTOV coefficients, SatRad coefficients, SatRad biases, Station list, Sonde coefficients etc.

Main Outputs from OPS:

i) Varobs: pre-processed observations (“extract and process” output)
ii) Modelobs: The model equivalent of observation. These files contain model forecast data interpolated to observation positions
iii) Background error: Prepared by “background error create” of OPS system based on model forecast (background) and the previous cycle’s background error file

2.3.3 Reconfiguration

The resolution of NEPS is N400L70 (horizontal resolution ~ 33km in mid-latitudes) and that of the UM deterministic model operational at NCMRWF (NCUM) is N768L70 (horizontal resolution ~ 17 km in mid-latitudes). Analysis files are prepared by 4D-VAR assimilation system of NCMRWF at 00, 06, 12 and 18 UTC using the deterministic UM model forecast as first guess. All the member models of NEPS use this analysis. Reconfiguration is the tool used to change the resolution of the input data and generate a reconfigured initial dump. Through reconfiguration task, NEPS reconfigures the initial analysis file produced by 4D-VAR assimilation system to generate a suitable initial dump to run the control and ensemble members of NEPS.

The reconfiguration step is run on multiple processors of HPC to gain the speed and memory improvements. To run reconfiguration on multiple processors, a method of domain decomposition is used in which processing element (PE) carries out the calculations for a portion of the whole grid. The atmosphere model is decomposed in two dimensions so two values of number of PEs are to be allotted: one for East-West and the other for North-South. Total number of PEs is the multiple of these two values. In the present case, the number of processors allotted to run reconfigurations is 512 (No. of PEs for East-West =16; No. of PEs for North-South = 32).

Inputs to Reconfiguration:

The analysis file at horizontal resolution of 17 km (N768L70) generated by the 4D-VAR Data assimilation system.

Outputs of Reconfiguration:

The reconfigured analysis file at horizontal resolution of 33 km (N400L70)

2.3.4 ETKF

The objective of ETKF is to provide initial conditions for NEPS forecasts. It generates global perturbations for wind, temperature, humidity and pressure fields for the 44 ensemble members. In NEPS system, the perturbations generated by ETKF are combined with the operational 4D-VAR analysis so that a full Ensemble Kalman Filter (EnKF) analysis is not required. Implementation of EnKF is computationally expensive whereas calculation of transformation matrix in ETKF, which updates only the initial perturbation instead of updating the analysis, is much cheaper. So ETKF provides an economic way to exploit many benefits of EnKF without being computationally expensive. The perturbations are added to the reconfigured analysis using the Incremental Analysis Update (IAU) scheme (Clayton, 2012) within the UM. The control forecast does not need any input perturbation from ETKF. It uses only the reconfigured analysis at N400L70 resolution as its initial condition.

ETKF receives the forecast perturbations from the previous forecast cycle (T+6 state for the perturbed ensemble members) as input. The forecast perturbations valid at the new analysis time are mixed and scaled by ETKF to generate new set of mutually orthogonal analysis perturbations. The mixing of the evolved forecast perturbations is performed by the transformation matrix. The analysis perturbations provide a 44 dimensional representation of the analysis error covariance matrix of an optimal data assimilation system. The calculation of transformation matrix requires the model equivalent of each observation for each ensemble member, to provide the estimates of background uncertainty in observation space. These 'pseudo-observations' are calculated by the Observation Processing System and provided to the ETKF (modelobs).

If the ensemble size is very small the background error covariance becomes large and the impact of observation is overestimated. This leads to unrealistically small analysis perturbations generated by ETKF. In order to counter this problem two methods are adopted: (1) horizontal localization (Houtekamer and Mitchell 1998) and (2) covariance inflation.

In horizontal localization a number of equally spaced localization centres (currently 92) are defined around the globe. For each centre, a local transformation matrix is constructed by using the observations within a radius of 2000 km. Interpolation between the local transformation matrices for the nearest localization centres gives the final transformation matrix for each grid point. In this way longer range correlations in the error covariances are cut off at a specified distance. Rank of the analysis covariance estimate also gets improved by horizontal localization.

Further improvement in ensemble spread is made by multiplying the raw transformation matrix of each region by a region specific inflation factor. This inflation factor gets updated at each assimilation cycle. The inflation factor of the previous cycle is multiplied by the ratio of the root mean square (RMS) error of the ensemble mean with respect to observation to the RMS spread of the ensemble forecast. OPS provides the ensemble mean and spread through “modelobs” (observation equivalent from short-forecast) files and observation through varobs files processed against the control forecast

Inputs to ETKF:

i) 44 modelobs files corresponding to the perturbed ensemble members and 1 varobs file corresponding to the control member from OPS run
ii) 45 pp1 files from short forecast of previous cycle containing the forecast fields: wind, potential temperature, exner pressure and specific humidity.
iii) Inflation factor obtained from the ETKF run of the previous cycle

Outputs of ETKF:

i) 44 Analysis perturbations
ii) Inflation factor

2.3.5 Short forecast

The UM short forecast run (N400L70) for 45 members (1 control and 44 perturbed) uses the reconfigured 4D-VAR analysis of operational deterministic UM (N768L70) and initial condition perturbations generated from ETKF at all assimilation cycles (00, 06,12 and 18 UTC) to make 6hr ensemble short forecasts for the next cycle. All the 45 members are allotted 10 nodes (10x16 processors) each on Bhaskara HPC to perform the short forecast run. The UM model run is controlled through the UNIX scripts which take input from namelists provided by the UM user interface (UMUI). The user interface is an X-windows application based on Tcl/Tk (Tool Command Language/Toolkit for windowing). A detailed description of UM global model can be found in Rajagopal et al. (2012). The output of UM short forecast run provides the first guess for the next assimilation cycle. The pp1 field files produced by the short forecast runs are used by ETKF to generate analysis perturbations for the next cycle

Main Inputs to short forecast:

i) N400L70 analysis reconfigured from the N768L70 deterministic model analysis.
ii) Initial condition Perturbations of wind, potential temperature, exner pressure and specific humidity generated by ETKF.

Outputs of short forecast:

i) The pp1 fields file for ETKF containing the forecast fields: wind, potential temperature, exner pressure and specific humidity.
ii) The background field file for OPS to create varobs and modelobs files

2.3.6 Long forecast

All the 45 members of NEPS are integrated forward based on the initial condition of 00 UTC daily to make EPS long forecast of 10 days. The model configuration is same as that is used for the short forecast run. The output of the long forecast run is set according to the need of the user community. The detailed description of the long forecast products and their utilities are given in section 3.

Main Inputs to long forecast:

i) N400L70 analysis (reconfigured from the N768L70 deterministic model analysis).

ii) Initial condition Perturbations of wind, potential temperature, exner pressure and specific humidity generated by ETKF.

Outputs of long forecast:

i) Forecast file (in UM field files format) containing forecast fields of u, v, w, geopotential height., MSLP, RH, T and surface pressure at 18 vertical levels at 24 hour interval

ii) Forecast file containing daily maximum and minimum temperatures at 2m above surface and 24 hourly accumulated rainfall

iii) Forecast file containing surface temperature, MSLP, q at 2m, RH at 2m, T at 2m, U at 10m, V at 10m and accumulated rainfall at a frequency of 6 hours

3. NEPS Forecast Products

The operational products generated from the NEPS are given in Table 4. The spatial plots are prepared for the domain covering 15^0S to 55^0 N and 60^0 E to 140^0 E.

Table 4: Operational NEPS products

Products	Variables	Resolution	levels	Freq (hrs)
Geo-potential-Height	Ht	0.45^0x 0.3^0	925, 850, 700, 500, 200 (hPa)	24
MSLP	MSLP	-do-	mean sea level	24
EPSgrams	T2m, RH2m, U10m, V10m, MSLP, 6hrly accumulated precipitation.	-do-	2m, 10m, surface, mean sea level	6
Rainfall-Probability, Ensemble-Stamps	Accumulated Precipitation	-do-	Surface	24
Wind-Forecast	U,V, surface pressure	-do-	925, 850, 700, 500, 200 (hPa)	24

3.1 Ensemble Mean and Spread

The ensemble mean is a simple mean of the parameter values of all ensemble members. It is calculated to assess, on average, the most likely outcome. The ensemble mean normally verifies better than the control forecast or any individual ensemble member because it smoothes out smaller-scale, relatively unpredictable features and simply presents the more predictable elements of the forecast. It can provide a good forecast guidance but must not be relied on its own, as it will rarely capture the risk of extreme events.

Ensemble Spread is calculated as the standard deviation of a model output variable, and provides a measure of the level of uncertainty in the forecast. The larger the spread, the greater is the uncertainty in the ensemble forecast. It is often plotted on charts overlaid with the ensemble mean.

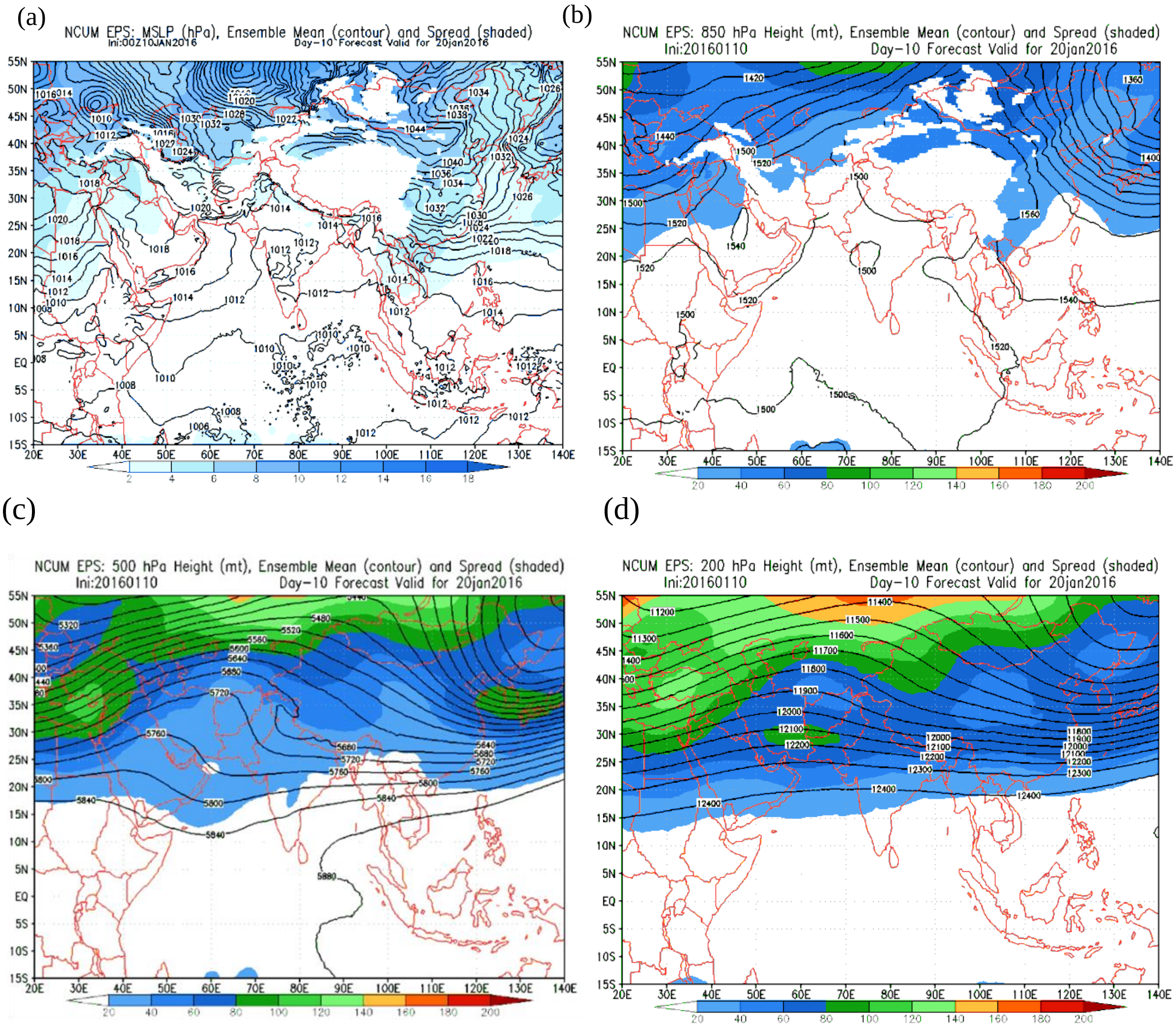

Figure 3: Ensemble mean and spread in Day-10 forecast of (a) MSLP and Geopotential height at (b) 850 hPa, (c) 500 hPa and (d) 200hPa valid for 00 UTC 20th January 2016.

Figure 3(a) shows Day-10 forecast of ensemble mean value of mean sea level pressure (PMSL) as contours and spread of MSLP as colour shading. The ensemble mean and spread in the Day-10 forecast of 500 hPa geopotential height is shown in Figure 3(c). The areas of strong colours indicate larger spread and therefore lower predictability.

It can be noted from Figure 3 that the spread is very large at higher latitudes and too small over tropics. The pole-ward increment of spread (uncertainty) may be attributed to the more large scale dynamical activity at higher latitudes. Bowler et al. (2008) noted excessively large spread near the poles and too small in the tropics during initial implementation of MOGREPS. The reason behind this was attributed to the small numbers of ensemble members (23) and to the fact that perturbation growth rate over tropics is low due to the insufficiently strong effect of model uncertainty perturbation in this region. Figure 3 (b) and 3 (d) also reveal that spread is very less near the surface and large aloft. The same feature of distribution of spread with height was noted by Bowler et al. (2007).

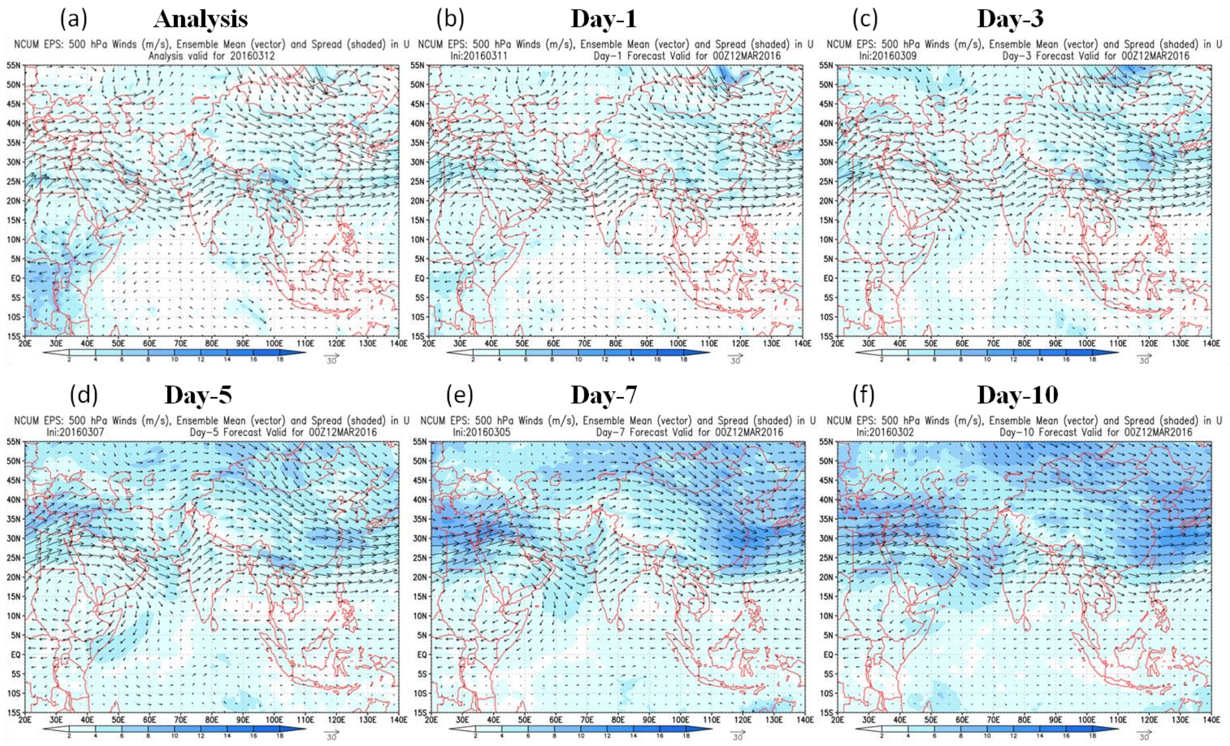

Figure 4: (a) Analysis, (b) Day-1, (c) Day-3, (d) Day-5, (e) Day-7 and (f) Day-10 ensemble mean forecast of winds (vector) and ensemble spread of wind speed (shaded) at 500 hPa, valid for 00 UTC 12th March, 2016.

Figure 4 shows analysis and forecasts of ensemble mean wind vector and spread of wind speed at 500 hPa till Day-10, valid for 12th March 2016. The mean wind speed/direction is shown by arrow length/direction and spread in the wind speed of ensemble members are indicated by the shading. The trough of westerly winds associated with the passage of western disturbance over North Arabian Sea near west coast of India, seen in the analysis is well represented in Day-1 to Day-3 forecast with very less spread (2-4 m/s) in wind magnitude. The spread in wind speed gradually increases from Day-5 (4-6 m/s) to Day-10 (6-8 m/s) forecasts. Uncertainty in wind forecast increases with increase in forecast lead time. For each forecast, regions showing more spread coincide with the dynamically active regions having higher wind magnitude suggest more uncertainty compared to places with calmer wind conditions (Bowler et al., 2008).

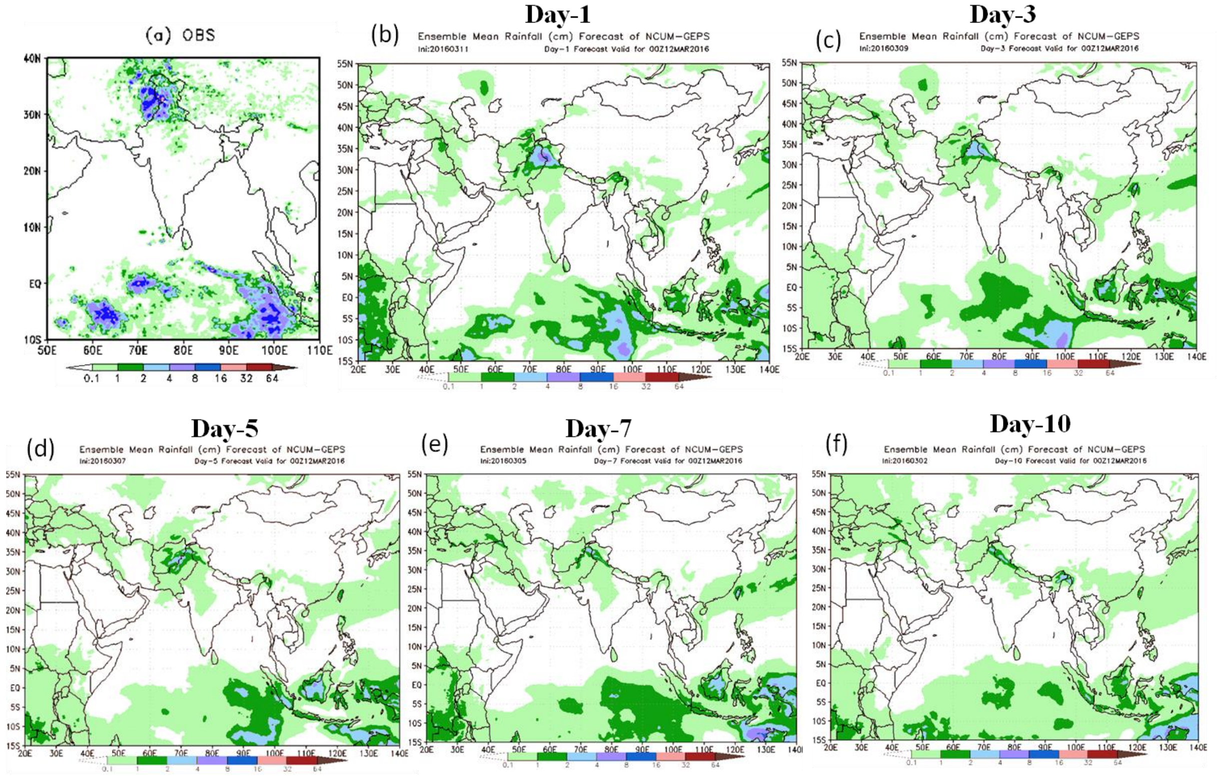

Figure 5: (a) Observed (satellite-gauge merged) rainfall (b) Day-1, (c) Day-3, (d) Day-5, (e) Day-7, (f) Day-10 Ensemble mean forecast rainfall (shaded) valid for 00 UTC 12th March, 2016

The observed satellite-gauge merged rainfall is plotted in Figure 5(a). Rainfall over Jammu and Kashmir and other places along the Northern frontiers of the country shows rainfalls in the range 16-32 cm. As mentioned in the beginning of the present section, ensemble mean rainfall fails to capture this heavy rainfall event and predicts only 8-16 cm rainfall in Day-1 forecast. The rainfall observed over the equatorial and south Indian Ocean is predicted in Day-1, Day-3 and Day-5 as well, with lesser intensity. Uncertainty in the predicted intensity and distribution of the wet areas increases with increasing forecast hours.

3.2 Postage Stamp Maps

A set of small maps showing the scenarios in each individual ensemble member forecast helps the forecaster to assess the possible risks of extreme events. Postage stamp maps show the spread in the solutions (forecast). Figure 6 depicts the postage stamp maps of Day-1 rainfall from NEPS valid for 00 UTC 2nd December 2015.

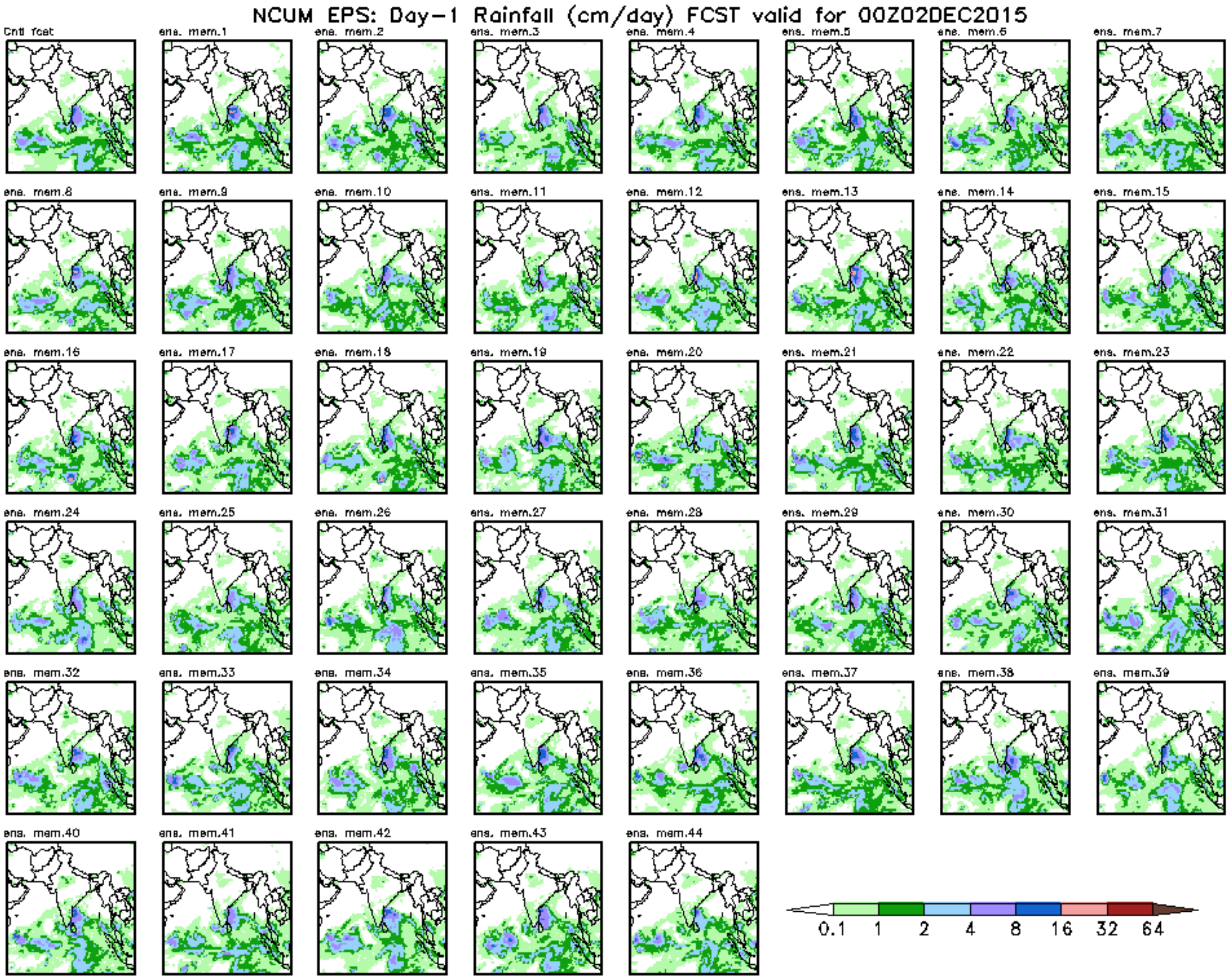

Figure 6: The postage stamp maps of 24 hr. accumulated rainfall (cm) for each of 45 ensemble members of the Day-1 forecast valid for 00 UTC 2nd December 2015

3.3 EPSGRAM

Ensemble meteogram or EPSgram is one of the most commonly used presentations of location specific forecasts from the ensemble prediction system. Figure 7 shows EPSGRAM over Chennai based on initial conditions of 00 UTC 30th November 2016. The box-plots show the characteristics of the distribution of surface meteorological variables. The boxes show a range of 25-75% percentile values and whiskers show the range between minimum and maximum values. The red line joins median values. Model output variables are extracted from the forecast at six hourly intervals to summarize forecasts at one location.

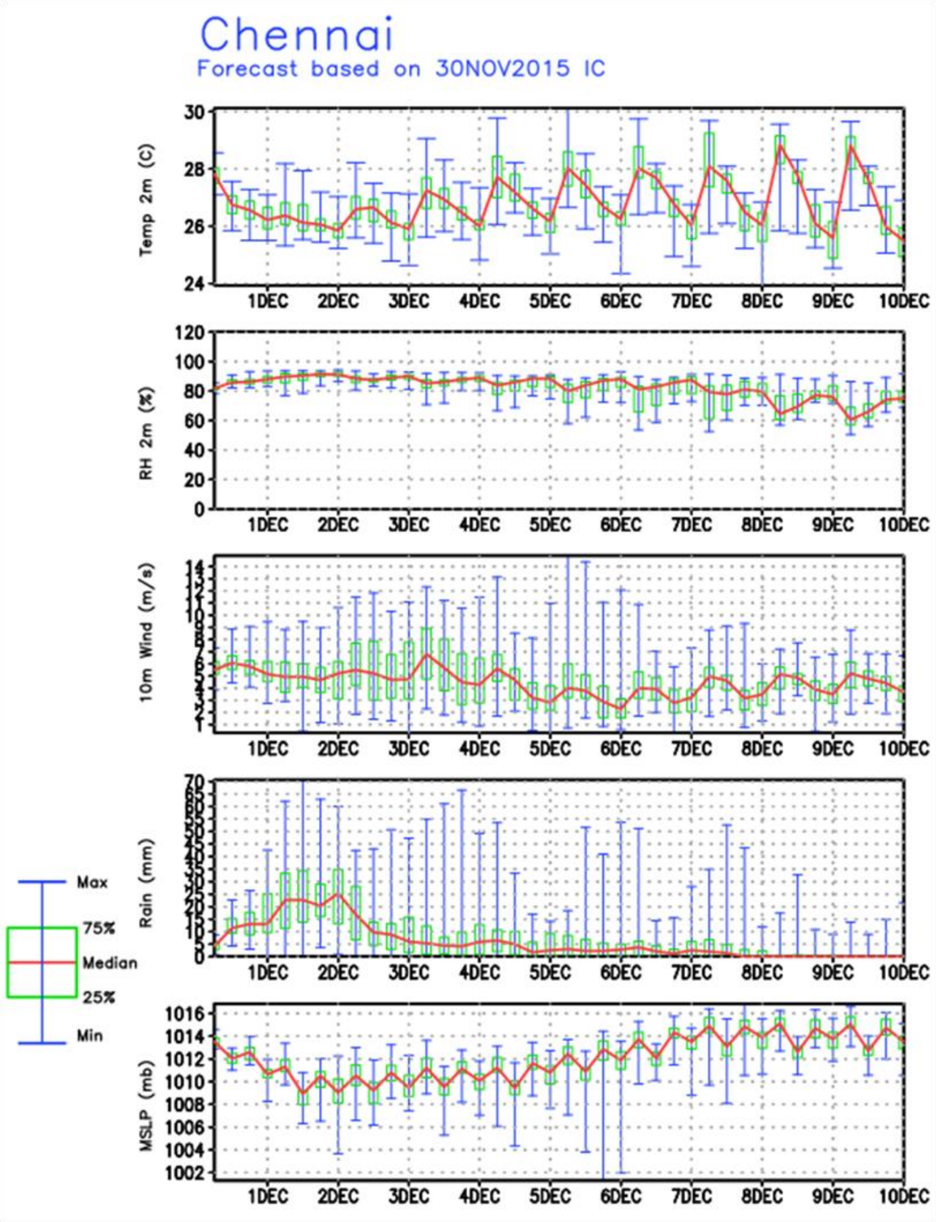

Figure 7: EPSGRAM for Chennai based on 00 UTC 30 November 2015 initial conditions, depicting Temperature (°C) & Relative humidity at 2m, 10m wind (m/s), rainfall (mm) and MSLP (mb) for next 10 days.

3.4 Spaghetti Plots

Spaghetti diagrams show one or a few contours for a field of interest to highlight probability distributions for that variable in regions of interest. In Figure 8, only 5740 m contour of geopotential height at 500 hPa is plotted for ease of readability, for all 45 members including control, each member forecast indifferent colours. The distance of members from each other gives a notion of uncertainty. Also, it gives an idea of the probability distribution for the forecast for the contours displayed. It is observed from the given figure that while the contours for individual members are initially very tightly packed, they spread out more and more with increasing forecast lead time, reflecting the increase in forecast uncertainty.

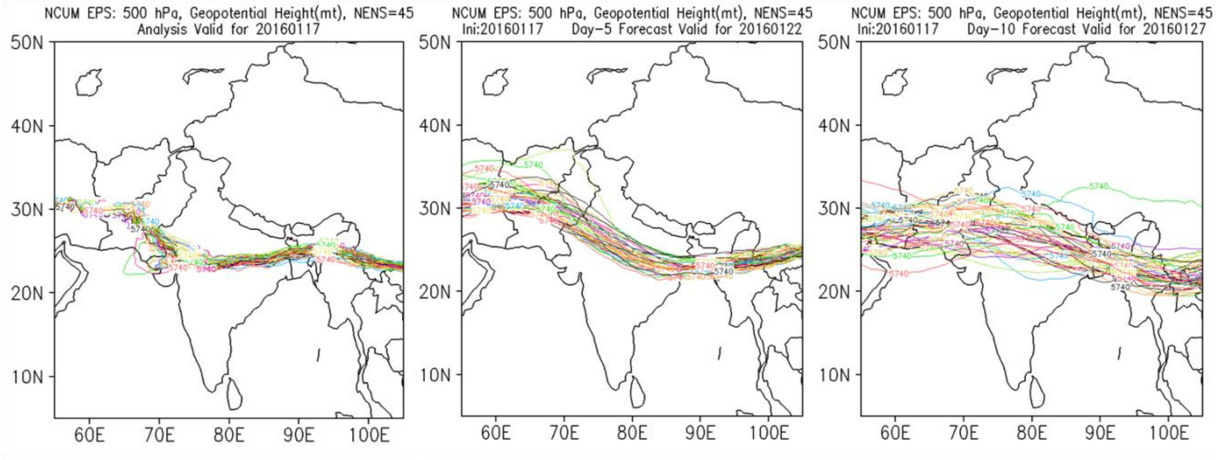

Figure 8: Spaghetti diagram depicting geopotential height at 500 hPa in (a) Analysis (b) Day-5 and (c) Day-10 forecasts, from the initial conditions of 00 UTC of 17th January 2016.

3.5 Plume Diagrams

Plume diagram shows time evolution of a forecast variable for each ensemble member. To depict ensemble forecasts at a point, or more properly, over a grid box, a plume diagram is prepared. Figure 9 illustrates the flow dependency of ensemble spread (uncertainty). The spread generally increases with the model integration time, but the amount of spread may change with different initial conditions, depends on the atmospheric flow. There can be cases when the spread is larger at shorter forecast ranges than at longer forecast range. This might happen when the starting days are characterized by strong synoptic systems with complex structures but are followed by large-scale “fair weather” high pressure systems (Ashrit et al., 2013).

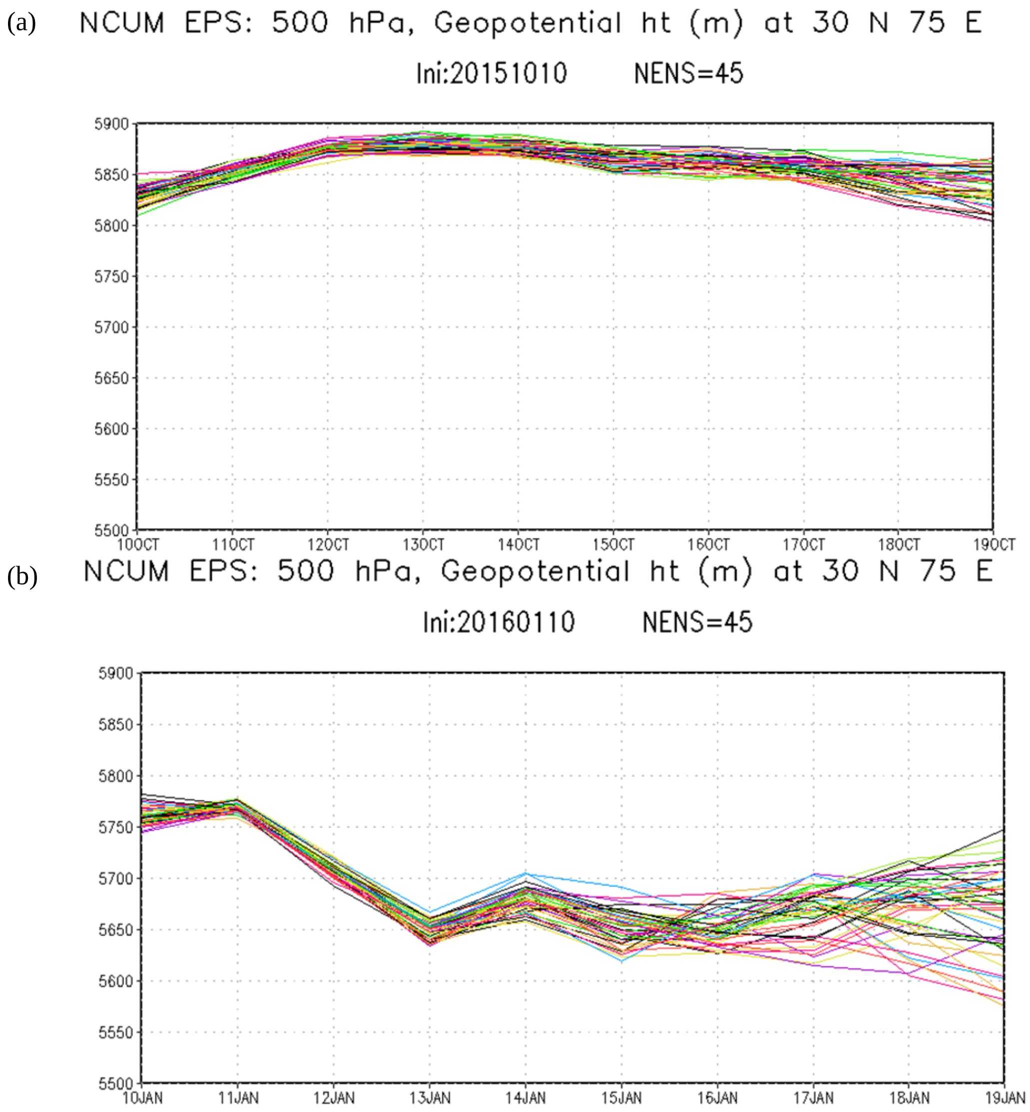

Figure 9: Plume diagram of geopotential height (m) at 500 hPa at a point 30^0N, 75^0E with initial conditions of (a) 10th October 2015 and (b) 10thJanuary 2016

4. NEPS prediction of a heavy rainfall event

A heavy rainfall event occurred during November-December 2015 over the East Coastal regions of Tamil Nadu, Andhra Pradesh and Puducherry. A low pressure system developed over Bay of Bengal reached Tamil Nadu coast on 30th November bringing heavy rain. Very heavy rainfall reported on December 1, 2015 led to flood across the entire stretch of Tamil Nadu coast. The heavy rain continued for several days bringing complete disruption to the public life of the city. Figure 10 shows the 24 hours accumulated rainfall observations (from satellite – gauge merged rainfall) on December 2, 2015. Maximum rainfall observed lie between 16 to 32 cm range on December 02, 2015 over Chennai and adjoining coastal areas.

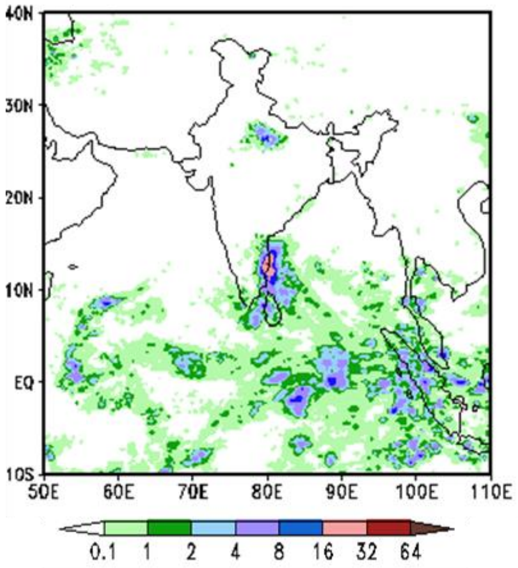

Figure 10: 24 hr accumulated satellite-gauge merged rainfall valid on December 2, 2015

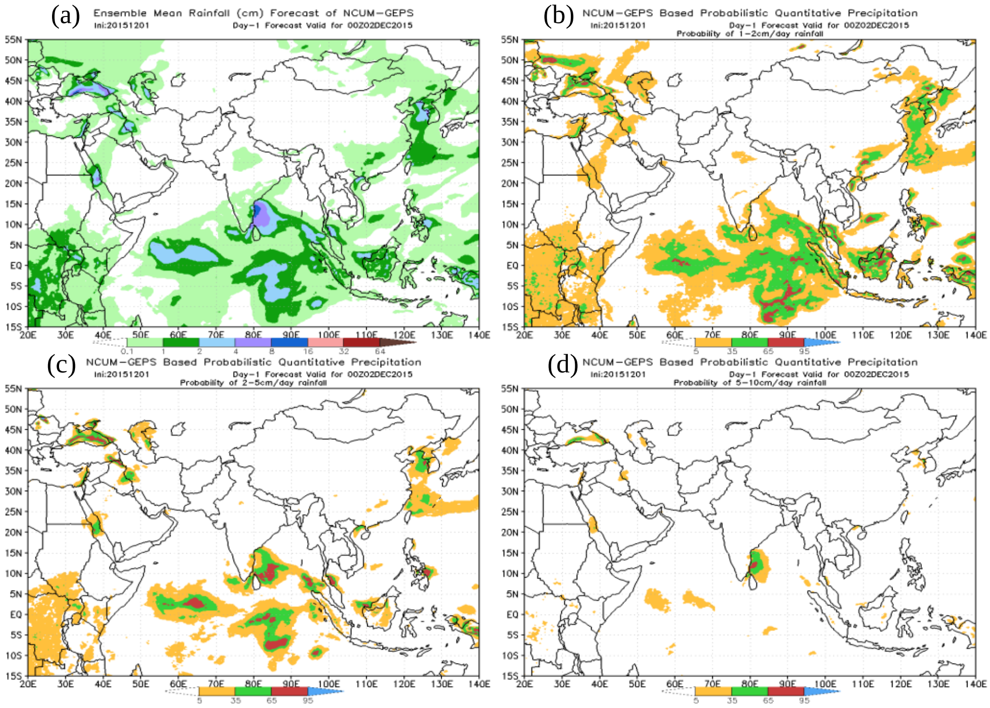

Figure 11: NEPS Day-1 forecast of (a) Ensemble mean rainfall, (b) Probability of 1-2 cm/day rainfall, (c) Probability of 2-5 cm/day rainfall,(d) Probability of 5-10 cm/day rainfall valid for 2^{nd} December, 2015

Figure 11 shows Day-1 forecast from NEPS valid on December 02, 2015. In the top left panel, the ensemble mean rainfall indicates most likely rainfall amount from 45 members. The rainfall amount lies in 8-16 cm range over Chennai and adjoining coastal places. The other panels depict the spatial distribution of forecast probability of 24-hour precipitation amounts. It is estimated from the number of ensemble members lying within specified ranges (1-2cm, 2-5cm and 5-10cm/day), out of all 45 ensemble members, expressed in percentage. For rainfall located at Chennai, the probability of ensembles lying within 1-2 cm category is up to 65 % covering the maximum area. The probability of rainfall predicted by NEPS for 5-10 cm/day over the Tamil Nadu coast goes up to 95 %. From Figure 11(a) it is also clear that, the distribution as well as intensity of rainfall over Chennai is under predicted in NEPS Day-1 forecast (ensemble mean rainfall) compared to observations (Figure 10).

5. Summary

A Unified Model based global ensemble prediction system is implemented at NCMRWF. It is the global version of UK Met Office MOGREPS, with 45 members. The 44 initial condition perturbations are generated using ETKF. The RP and SKEB2 schemes of the NEPS take care of the model uncertainties. There are various components of NEPS like processing of observations (OPS and Trimobstore), generation of perturbations, reconfiguration of model initial condition to desired resolution and model runs for control and perturbed ensemble members to generate forecast. Various forecast products are generated routinely from the EPS. The forecast products are summarised below:

- Mean and spread plots show that uncertainty generally increases with increasing forecast length and with higher altitudes. The distribution of spread increases towards the higher latitudes in Northern Hemisphere for most of the forecast variables. Over the tropics the spread is relatively small.
- Postage stamp maps show the spread of all 45 ensemble members (including control) at a glance and allow the forecaster to assess the possible risks of extreme events.
- EPSGRAM summarizes location specific forecast of the model surface fields namely, temperature (C) & relative humidity at 2m, 10m wind (m/s), rainfall (mm) and MSLP (hPa), for next 10 days.
- Spread is flow dependent and it varies with different initial conditions as is evident from the spaghetti plots and plume diagrams.
- Probabilistic forecasts enable forecasters to present the forecast with confidence.

Acknowledgments

The authors are thankful to the Met office, UK for providing MOGREPS system. Technical support provided by M/s IBM and CMC/TCS in this initiative is gratefully acknowledged. The authors also thankfully acknowledge the helps provided by Drs. Swapan Mallick, S. Indira Rani and R. Rakhi

References

Ashrit, R., Iyengar, G. R., Sankar, S., Ashish, A., Dube, A., Dutta, S. K., Prasad, V. S., Rajagopal, E. N., and Basu, S., 2013: Performance of Global Ensemble Forecast System (GEFS) During Monsoon 2012, *NCMRWF Internal Research report-NMRF/RR/01/2013.*

Bishop, C. H., Etherton, B. J. and Majumdar, S. J., 2001: Adaptive samplingwith the ensemble transform kalman filter. Part 1: theoretical aspects *Mon. Wea. Rev.* **129**, 420–436.

Bowler NE, Dando M, Mylne KR, Beare SE. 2007. The MOGREPS short-range ensemble prediction system: Verification report – trial performance of MOGREPS January2006–March2007. *Met Office Technical Report 503*. Available at http://www.metoffice.gov.uk/research/nwp/publications/papers/technical reports/reports/503.pdf.

Bowler, N. E., Arribas, A., Mylne, K.R., Robertson, K. B., and Beare, S. E., 2008: The MOGREPS short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, 134, 703–722,doi:10.1002/qj.234.

Bright, D. R., and Mullen,S. L., 2002: Short-range ensemble forecasts of precipitation during the southwest monsoon. *Wea. Forecasting*,17, 1080–1100, doi:10.1175/1520-0434(2002)017,1080:SREFOP.2.0.CO;2.

Buizza, R.: Chaos and Weather Prediction January 2000, ECMWF Notes

Buizza R., and Palmer,T., 1995: The singular vector structure of the atmospheric general circulation. *J. Atmos Sci.* **52**,1434-1456.

Charney, J.G.,Fjortoft, R., von Neumann, J., 1950: Numerical Integration of the Barotropic Vorticity Equation, *Tellus*, **2**, 237-254.

Clayton, A., 2012: Incremental Analysis Update (IAU) Scheme. Unified Model Documentation Paper No. 31

Eady, E., 1951: The quantitative theory of cyclone development. *Compendium of Meteorology*, T. Malone, Ed., *Amer. Meteor. Soc.*, 464–469.

Epstein, E. S., 1969: Stochastic dynamic prediction. *Tellus*, **21**: 739−759.

George, J.P., Rani, S. I., Jayakumar, A., Mohandas, S., Mallick, S., Rakhi, R., Sreevathsa, M. N. R., and Rajagopal, E. N., 2016: NCUM Data Assimilation System, *NMRF/TR/1/2016*, 20 p.

Gray, M. E. B., and Shutts, G. J., 2002: A stochastic scheme for representing convectively generated vorticity sources in genera lcirculation models. *APR Turbulence and Diffusion Note 285, Met Office, Exeter, United Kingdom*, 33 pp.

Hollingsworth, A., 1980: An experiment in Monte Carlo forecasting. *Workshop on stochastic-dynamic forecasting. ECMWF*, Shinfield Park, Reading, UK.

Hoffman, R. N., and Kalnay,E., 1983:Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus*, **35A**, 100-118.

Houtekamer, P. L., Lefaivre, L., Derome, J., Ritchie, H. &Mitchell, H. L., 1996: A system simulation approach toensemble prediction. *Mon. Wea. Rev.*, **124**, 1225−1242.

Houtekamer, P.L., and Mitchell, H.L., 1998: Data Assimilation Using an Ensemble Kalman Filter Technique, *Mon. Wea. Rev.*, 126: 796-811.

Lewis,J.M., 2005: Roots of ensemble forecasting, *Mon. Wea. Rev*. **133**, 1865–1885.

Kalnay, E., and Toth,Z., 1996: The breeding method, *Proceedings of the Seminar on Predictability*, held at ECMWF on 4–8September 1995. (Available from ECMWF, Shinfield Park, Reading, Berkshire RG29AX, e.kooij@ecmwf.int)

Kay, J. K., and Kim,H. M., 2014: Characteristics of initial perturbations in the ensemble prediction system of the Korea Meteorological Administration, *Weather and Forecasting*, **29**, 563-581, doi:10.1175/WAF-D-13-00097.1.

Lorenz, E. N., 1965: A study of the predictability of a 28-variable atmospheric model. *Tellus*, **17**: 321−333.

Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409−418.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The new ECMWF Ensemble Prediction System: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73-119.

Palmer, T. N., Molteni, F., Mureau, R., Buizza, R., Chapelet, P. and Tribbia, J., 1993: Ensemble prediction. *Proceedings of the ECMWF seminar on Validation of models over Europe: Vol. 1*, ECMWF, Shinfield Park, Reading,UK, 21−66.

Rajagopal, E. N., Iyengar, G. R., George, J. P., Das Gupta, M., Mohandas, S., Siddharth, R., Gupta, A., Chourasia, M., Prasad, V. S., Aditi, Sharma, K., and Ashish, A., 2012: Implementation of the UM model based analysis–forecast system at NCMRWF, *NMRF/TR/2/2012*, 45p

Tennant, W. J., Shutts, G. J., Arribas, A., and Thompson, S. A., 2011: Using a Stochastic Kinetic Energy Backscatter Scheme to Improve MOGREPS Probabilistic Forecast Skill, *Mon. Wea. Rev.*, 139, 1190-1206, DOI: 10.1175/2010MWR3430.1

Toth, Z. & Kalnay, E., 1993: Ensemble forecasting at NMC: Ensemble forecasting techniques with a focus on tropical cyclone forecasting the generation of perturbations. *Bull. Am. Meteorol. Soc.*,**74,** 2317−2330.

Wang, X., and Bishop, C. H., and Julier, S. J., 2004: Which is better, an ensemble of positive–negative pairs or a centered spherical simplex ensemble? *Mon. Wea. Rev.,* 132, 1590–1605, doi:10.1175/1520-0493(2004)132,1590:WIBAEO.2.0.CO;2.