NMRF/TR/1/2015

सत्यमेव जयते

TECHNICAL REPORT

Development of a comprehensive rainfall verification package based on Model Evaluation Tools

Saji Mohandas and Harvir Singh

May 2015

National Centre for Medium Range Weather Forecasting
Earth System Science Organisation
Ministry of Earth Sciences
A-50, Sector 62, NOIDA – 201 309, INDIA

Development of a comprehensive rainfall verification package based on Model Evaluation Tools

Saji Mohandas and Harvir Singh

Technical Report No. NMRF/TR/1/2015

May 2015

National Centre for Medium Range Weather Forecasting
Earth System Science Organisation, Ministry of Earth Sciences
'Indradhanush', A-50, Sector 62, NOIDA-201 309, India

	Document Control Data Sheet	
1	Name of the Institute	National Centre for Medium Range Weather Forecasting (NCMRWF)
2	Document Number	NMRF/TR/1/2015
3	Date of publication	May 2015
4	Title of the document	Development of a comprehensive rainfall verification package based on Model Evaluation Tools
5	Type of Document	Technical report
6	No. of pages & figures, tables	37 Pages, 8 figures and 4 tables and 2 appendices
7	Number of References	27
8	Author (S)	Saji Mohandas and Harvir Singh
9	Originating Unit	National Centre for Medium Range Weather Forecasting (NCMRWF), A-50, Sector-62, Noida, Uttar Pradesh
10	Abstract (100 words)	A tool was developed for the comprehensive evaluation of model forecasts using both traditional and spatial diagnostic techniques. The traditional scores suffer from the so called "double penalty" issue and hence alone cannot provide a measure of spatial and temporal match between the forecast and observed rainfall patterns. Method for Object-based Diagnostics Evaluation is a features based verification technique while the wavelet stat is based on scale-separation principle. The traditional verification scores were computed using categorical and continuous measures and the spatial verification scores were computed against various thresholds. The tool was implemented for both the deterministic models at National Centre for Medium Range Weather Forecasting (NCMRWF) at a resolution of 50km against the raingauge observations and gridded rainfall analysis. Also it is adapted for all the regional configurations and will be useful for evaluation of mesoscale models with high resolution rainfall analysis.
11	Security classification	Unrestricted
12	Distribution	General

Abstract

A tool was developed for the comprehensive evaluation of model forecasts using both traditional and spatial diagnostic techniques. The traditional scores suffer from the so called “double penalty” issue and hence alone cannot provide a measure of spatial and temporal match between the forecast and observed rainfall patterns. Method for Object-based Diagnostics Evaluation is a features based verification technique while the wavelet stat is based on scale-separation principle. The traditional verification scores were computed using categorical and continuous measures and the spatial verification scores were computed against various thresholds. The tool was implemented for both the deterministic models at National Centre for Medium Range Weather Forecasting (NCMRWF) at a resolution of 50km against the raingauge observations and gridded rainfall analysis. Also it is adapted for all the regional configurations and will be useful for evaluation of mesoscale models with high resolution rainfall analysis.

Contents

1. Introduction

National Centre for Medium Range Weather Forecasting (NCMRWF) has been operationally producing WMO standard verification scores for the model forecasts and generating the year-to-year verification statistics in its Monsoon performance evaluations. Daily operational computation of verification scores are being carried out for basic prognostic variables, for which the model analysis are available, which is mainly based on the Verification Statistics Data Base (VSDB) software. This does not include rainfall forecasts which are not analysed globally, not an input to the model and for which the verification is very difficult and complex. The verification of rainfall is very crucial as it is one of the important end products with more practical application to the user community. As it is being very discontinuous field and the representativeness of the ground truth against which the verification is to be carried out mostly depends on the quality, density and the pre-processing of the observational data, the verification exercise is very tricky. Also the traditional point-to-point verification adopted for other continuous types of variables may not reveal the spatially coherent characteristics of the discontinuous parameter like rainfall and cloudiness. There are many attempts and studies on a general framework of verification and also on the different methods for rainfall verification [Murphy and Winkler (1987); Jolliffe and Stephenson (2003); Stansky et al. (1989); Wilks (2006); Ebert (2008)]. In addition to the above, the measures related to the spatially coherent features of the forecasts should be investigated [Davis et al., 2006;2009, Brown et al., 2007, Gilleland, 2013, Gilleland et al. 2009; 2010a;2010b, Casati, 2010, Casati et al., 2004; 2008, Ebert, 2008; 2009, Gallus, 2010, Ebert and McBride, 2000, Ebert and Gallus, 2009, Ahijevych, et al., 2009, Mittermaier and Roberts, 2010]. The day-to-day statistics can be aggregated to estimate the overall performance of each of the episodes of synoptic scale and mesoscale phenomena occurring on a very regular basis in different types of weather regimes. The daily statistics of the episodes can be averaged and summarised for every month to assess the monthly performance of the models, which in turn can be again aggregated over a season or year to condense the huge amount of information into very few quantitative figures. This will allow the year by year comparison of performance of multiple modelling systems or year-to-year variability for a single modelling system.

The current report presents the new package developed and set up for evaluating the performance of the deterministic models at NCMRWF adopted from US National Centres for Environmental Prediction (NCEP) and UK Met Office (UKMO), namely, NGFS and NCUM respectively [See Prasad et al., 2011, 2013, Rajagopal et al. 2012]. This is based on a collection of tools known as Model Evaluation Tools (MET) originally developed by NOAA-sponsored Development Testbed Centre (DTC) and assembled at NCMRWF on its IBM Power 6 High Performance Computing (HPC) system. MET is a tool kit for comprehensive performance evaluation between different models, of any forecast variable having a corresponding 'truth' to verify such as observation or analysis. It incorporates both traditional scores as well as spatial verification scores, like, Method for Object-based Diagnostics Evaluation (MODE) and wavelet analysis and is mainly intended for the verification of high resolution model outputs against high resolution observations. It was implemented for various models like, Weather Research and Forecasting (WRF), Global Forecast System (GFS), Unified Model (UM) and regional versions of UM. A description of MET tools with tutorials and user guides is available from MET user's home page, 'http://www.dtcentre.org/met/users/'.

Some of the traditional verification scores suffer from the problem of the so-called 'double-penalty'. This is because, traditional grid-point verification methods penalise a minor shift in the location twice, once for missing grid points where the precipitation event occurred, and also for predicting false alarms at some other grid points. MODE is an effective alternative to provide additional diagnostic information and an objective assessment of location, size and intensity errors of the synoptic systems which is otherwise impossible through traditional approaches.

Wavelet-stat analysis uses scale-decomposition approach to identify the scale at which the skill is maximised. It is applied to forecast and observation fields to obtain spatial scale components and to compute the bias, error and skill of forecast on each spatial scale. It provides the information on the ability of the model in reproducing the observed scale structure and scale dependency of error and skill. The current report describes a set of standard diagnostic measures for the routine monitoring and objective assessment of the overall performance of the NCMRWF operational global models in terms of the rainfall prediction. Data and methodology and preprocessing steps are mentioned in the next two sections. More details about the measures are given in appendices. Some of the results of the case study of the performance of the rainfall prediction of the Tropical Cyclone 'Phailin' (Mohandas and Singh, 2015) are also presented as an example in section 4.

2. Data and methodology

Results of a case study of Tropical Cyclone 'Phailin' are presented in this report. The forecasts by two deterministic global models NGFS and NCUM are verified. The initial conditions (ICs) used for the experiment are 8-13 October, 2013. The domain of the study is 5-30^{o}N, 75-100^{o}E. The daily accumulated raingauge reports received through Global Telecommunications System (GTS), Automatic Weather Stations (AWS) and IMD-NCMRWF satellite-raingauge merged gridded rainfall analysis [Mitra et al. 2003; 2009] are used as the observations against which the statistics are computed using various tools.

MET provides basically four major tools to estimate various kinds of verification statistics, namely, (i) Point-stat, (ii) Grid-stat, (iii) MODE and (iv) Wavelet-stat. Point-stat is the standard verification measure computed at station points and Grid-stat is the same computed at some common regular grid points. Before the computation of the statistics, both observation and forecast matched pairs need to be generated at a common grid by some of the most popular re-gridding techniques suitably selected for the variable under study. The rainfall data was analysed with the scatter diagrams and box plots to check the match between the forecast and observed data over the raingauage station points. Only land station point observations over the domain of study were taken for the study and the data and its distributions were studied for the representativeness. Traditional scores were computed for both continuous and categorical measures using Point-stat and Grid-stat. Continuous measures are basically based on the difference between forecast and observed rainfall, whereas categorical measures are based on the 2x2 categorisation of 'yes' or 'no' of rainfall values at different rainfall thresholds by generating a contingency table for each of the threshold. As the focus is on Tropical Cyclone and the number of land raingauge stations reporting the associated rainfall is very less, the gauge-based metrics are not shown in the current report. Point-stat was computed using only land raingauge observations available in the domain. Grid-stat results are presented for traditional scores, which are computed against the gridded rainfall analysis.

For object-based verification, (i) the MODE was used and for scale separation, (ii) wavelet stat tool was used both of which are part of the Model Evaluation Tools (MET). The scores were averaged for the entire event to summarise all the aspects of the verification. The Appendix-I describes the definition of the traditional scores and Appendix-II mentions the brief description of MODE and the settings adopted for the current study [See Brown et al, 2007 and Davis et al., 2006]. The wavelet-stat tool decomposes the forecasts and observations according to intensity and scale, by thresholding the same to convert into binary fields and decomposing into sum of components of different scales. A 2-dimensional Haar wavelet filter is used and for each threshold and for each scale component of binary forecast and observation, mean squared error (MSE) is evaluated. The largest error is typically associated with smallest scale and highest threshold. For each threshold and scale component, the intensity-scale skill (ISS) score based on the MSE of binary forecast and observation scale components is evaluated taking random chance as reference forecast [Jolliffe and Stephenson, 2003, Wilks, 2006].

3. Pre-processing of data and computational procedures

MET tool software executables are installed in the public software area MET_TOOL, namely, '/gpfs1/software/MET/MET3.0.1/bin'. This is a common or public software repository which can be accessed by all users and each individual user need not install the software in his own area. The scripts for computation of statistics will use the executables from this area. The tools installed include point_stat, grid_stat, MODE and wavelet_stat apart from many other optional softwares for data preprocessing and tools like stat_analysis and mode analysis for the long period accumulation of the results. The MET_HOME area is named '/gpfs1/home/umfcst/MET301' which contains all the scripts and codes required to run the computations. This file structure optionally can be copied to the user area and 'umfcst' can be replaced with user's $LOGNAME to run scripts from the user area, in case the user wants some modification to be done or for running with a different domain. A third location is defined as MET_OUT which is the path to the output of MET computations, which is defined as '/gpfs1/home/$LOGNAME/MET_OUT/rain'. It contains all the outputs of the rainfall statistics under the particular initial condition (IC) datestamp folder in three subdirectories named after the (i) domain, (ii) data and (iii) plots. The domain subdirectory is written as 'R<LAT1>-<LAT2>N_<LON1>-<LON2>E' which contains whatever produced by the direct output of the MET tools. The other two directories are 'data' and 'plots', which contains the output of plotting scripts in data and graphics formats respectively. If the user wants MET output graphics only to be executed from the input data of operational MET output area, the plotting scripts can be setup with two separate data areas, namely 'MET_IN' and 'MET_OUT', which are exported as the operational 'MET_OUT' area and the user's 'MET_OUT' area respectively. In the case of user running both the MET computation and MET plotting parts, both the environment variables are to be exported as '/gpfs1/home/$LOGNAME/MET_OUT/rain'.

Fig. 1 displays the flow chart of the directory structure of the MET system which includes the data preparation or the pre-processing steps, the execution of point-stat, grid-stat, MODE and wavelet-stat analysis along with the scripts for graphics and visualisations. The MET301 area contains the model specific source code subdirectory (MODEL) along with 'config', 'scripts' and 'master' subdirectories. Model specific sub-directories are named following the two global models, namely, 'NGFS' and 'NCUM', each of which contains subfolders like 'src' and 'config'. The 'config' area contains the configuration files for various MET tools in which the vital parameters for the execution of the MET tools are set. The 'scripts' area contains the main scripts for data processing and associated scripts. The plotting scripts are residing inside 'scripts' area in a folder called 'masterplot'. The plotting scripts run a few Rscripts residing in the sub-folder 'Rscripts' inside 'masterplot' directory. The directory 'master' contains the combined script ('submit_met.sh') for submitting the scripts for both computations and plotting parts for a particular IC and for 7 days forecast lead time statistics. This is a wrapper script for running (i) the master script for MET computations and (ii) the master script for MET output graphics, namely, 'met_tool_master_script.sh' and 'met_tool_master_plot.sh' respectively. The mandatory arguments to the former are 'YYYYMMDD', 'No of ICs' and 'Lead time to process - in days' and for the latter, a single argument of 'DDMMYYYY'. The date stamps represent the IC (or analysis time) and the 'No of ICs' decides how many successive ICs to process starting with the given datestamp in a single command. The third argument is the forecast lead time for each IC to process. A default domain will be used for the computations which is rectangular domain just covering the four boundaries of India and the neighbouring oceanic region (8-38N, 65-100E). However, there can be more optional arguments to the computation master script for specifying a particular user domain 'NX NY XS XE XI YS YE YI' which represents the number of X and Y grid points followed by 'start', 'end' and 'intervals' of X and Y coordinates respectively.

The master script for MET tool 'met_tool_master_script.sh' calls basically a couple of scripts in 'MET301/scripts' folder, one for the computation of point_stat and another for the computation of grid_stat, MODE and wavelet_stat together for each model. All the output statistics from all the models are written to corresponding date-wise subdirectories in the path 'MET_OUT/rain/YYYYMMDD'. Subsequently, the second master script for plot 'met_tool_master_plot.sh' dumps hundreds of graphics for each of the type of verification metrics in the same path inside the subdirectory 'plots' and the output data are dumped in the subdirectory 'data'. The pre-processing procedures can be generally separated into the following sub-sections.

3.1 Procedures for rainfall distribution characteristics

The first step in any verification procedure should be to check the geographical, spatial and temporal distribution of rainfall data, its match with the forecast distribution and the data representativeness. To check the data representativeness, the raingauge observations and automatic weather station (AWS) observations are used, whichever is available through Global Telecommunication System (GTS) as 24-hour accumulated values over Indian region. The first thing to be noted is the histogram and the scatter diagram of the forecast versus observation grid point rainfall values which are generally presented as monthly accumulated values. A set of matched pairs is generated for this purpose which corresponds to each observation against each forecast grid point by one of the possible several techniques. The current study used unweighted linear interpolation technique for the generation of matched pairs.

3.2 Procedures for station data statistics

For traditional verification scores, two types of metrics are generated, the first with the continuous variables and the second with categorical variables. For continuous variables, the verification methods are consistent with the general framework of verification outlined by Murphy and Winkler (1987). The measures include Forecast Mean (FBAR), Multiplicative Bias (MBIAS), Mean Error (ME), Mean Absolute Error (MAE), Pearson Correlation (PR_CORR) and Root Mean Squared Error (RMSE). The domain mean of the forecast and the observation computed over the forecast-observation pairs (FBAR and OBAR) is only one of the many important aspects of performance of the models. For categorical measures, the daily rainfall is divided into 5 categories with thresholds of 1, 5, 10, 50 and 100mm and various metrics are computed. These include Mean Forecast (FMEAN), Frequency Bias (FBIAS), Accuracy (ACC), Critical Success Index (CSI), Gilbert Skill Score (GSS), Hanssen-Kuipers discriminant (HK), Probability of Detection of Yes events (PODY), Probability of Detection of No events (PODN), Probability of False Detection (POFD) and False alarm Ratio (FAR).

The script for point_stat computations 'point_stat_<MODEL>_gts-rain,sh' first sets a few environment variables to set different paths and time and domain specifications. Mainly three paths are to be defined in all scripts, which are 'MET_TOOL', 'MET_HOME' and 'MET_OUT' as mentioned before. Another important path is 'CONFIG' which is usually defined under 'MET_HOME' path and contains the 'config' files for all the MET softwares. These are user defined parameters and can be tuned by the user as per specific requirements, types of outputs to be generated or user-defined techniques and priorities. Also all the scripts contain paths for the forecast outputs and observational data specific to the applications concerned. Point_stat scripts require GTS observations area while all other tools require the rainfall analysis corresponding to the particular IC. Also the scripts contain paths to various softwares for the processing of the observed and forecast data and bringing both into a common grid.

Point_stat scripts first pre-process all the available GTS abd AWS observations of 24-hourly accumulated rainfall at station locations. Two FORTRAN codes are used to read the data from observation-monitoring output locations where the ASCII data are generated after buffer-decoding. These scripts are 'suflnd_ext.f' and 'suflnd.acc.f' for GTS observations and; 'sufmob.ext.f' and 'sufmob.acc.f' for AWS observations, which (i) extract the data at four cycles '00', '06', '12' and '18', checking the validity time stamp and (ii) process the data to obtain the 24-hourly rainfall, respectively. In the current setup the rainfall data are read valid for '03Z' which itself is the 24-hourly observation and hence do not require any processing or summation. The outputs are written in ASCII format and saved as 'suflnd_point.dat,YYYYMMDD' and 'sufmob_point.dat.YYYYMMDD' in 'data' directory. These data are converted to an ASCII table format by a script 'rf.pl' and is fed into 'ascii2nc' tool to convert to a specific type of NetCDF format suitable to be accepted by the point_stat tool. These NetCDF intermediate output files are also saved in 'data' directory and the data locations are plotted in a geographical plot using 'plot_point_obs' installed in the 'MET_TOOL' path. These plots are saved in 'plots' area with the file name 'gts-<MODEL>_point_obs_suflnd_<YYYYMMDD>.jpg', 'gts_<MODEL>_point_obs_sufmob_

<YYYYMMDD>.jpg’ and 'gts__<MODEL>_point_obs_sufall_<YYYYMMDD>.jpg' respectively for GTS, AWS and merged observation points.

3.3 Procedures for gridded data statistics

The next process is to get the forecast rainfall for the corresponding date in the same grid as the observations. In the case of NGFS, Climate Data Operator (CDO) is used to accumulate the 6-hourly rainfall to 24-hourly, 'gfs2grib.sh' or 'copygb' to regrid to the domain of interest (All India domain by default) in GRIB1 format. In the case of NCUM, the steps are (1) 'um2grib.sh' for converting UM Fields File (FF) format rainfall to GRIB1 format and (2) 'copygb' to regrid to the requested domain by bi-linear interpolation. The point_stat tool accepts forecast output in GRIB1 format, observation in table NetCDF format and 'config' file to generate the required statistics files as per the 'config' file specifications. Point_stat is called three times for (i) GTS, (ii) AWS and (iii) merged statistics outputs of which are all stored in 'ASCII' format in 'data' output directory.

The script 'grid_mode_wavelet_<MODEL>_nmsg-rain.sh' will execute all the 'grid_stat', MODE and 'wavelet_stat' tools for each individual model using IMD-NCMRWF gauge-cum-satellite-merged rainfall analysis (NMSG) product as gridded rainfall observations at 50km resolution. It first converts the gridded rainfall analysis to NetCDF format using 'bin2ascii' tool and 'imd2nc.R' Rscript. The rest of the processing is same as 'point_stat' script. The model forecast is input as GRIB1 file and the observation part is gridded NetCDF format file.

The 'masterplot' directory contains all the scripts for plotting of the output and most of them uses Rscripts inside the sub-directory 'Rscripts'. The user should take care if the MET_IN export variable is properly passed on to the master script, ie., 'met-tool_masterplot.sh', so as to read the statistics computed from the operational area (by default) or from own MET_OUT area. It is safer to use an 'export' statement to avoid this common mistake. First is the plotting of the GTS and AWS row observations as an additional resource for geographical distribution of the observed rainfall. Station observations received and decoded over the Indian region are plotted as coloured circles with the same colour palette as that of the other gridded rainfall plots. The plotting script called 'point_rainplot.sh' is written using ' Generic Mapping Tools' (GMT). The gridded rainfall analysis and the forecasts are plotted using GrADS and the combined plot is created using the script 'geographical_plots_obs-ncum-ngfs.sh'. The next two scripts 'gridctcproc.sh' and pointctcproc.sh' produce the contingency table statistics (CTS) ('gridctslist' data tables) for grid_stat and point_stat computations for the current IC for all forecast lead times, 'mode_image_crop.sh' will put together analysis and forecasts rainfall geographical distribution as well as simple objects clustered inside contours for different convolution thresholds 1, 5, 10, 20, 50 and 100 mm. The convolution resolution is 2 grid sizes for all. 'mode_analysis.sh' will invoke 'mode_analysis' tool to compute some of the summary object statistics and ' mode_prec.sh' will convert the MODE output statistics into separate 'ASCII' tables along with the summary objects statistics for each convolution threshold and forecast lead time and written to ASCII files with the names starting with 'modelist_<MODEL>.txt' for each model. Also 'wavelet_stat_isc_plot.sh' will accumulate the wavelet analysis outputs for the current IC and displays Intensity_scale Skill Score (ISC) as bar diagrams of different scales for each of the thresholds 1, 5, 10, 20, 50 and 100mm.

3.4 Procedures for aggregation and summary statistics

Ultimately, monthly statistics of 'grid_stat' and 'point_stat' output from previous 30 days are plotted as box and whisker plots and time series using following 5 scripts for grid stat and point stat as well as for continuous (CNT) and contingency table statistics (CTS) outputs and matched pairs statistics (MPR).

1) grid_stat_cnt_plots_weekly_monthly.sh
2) grid_stat_cts_plots_weekly_monthly.sh
3) point_stat_cnt_plots_weekly_monthly.sh

4) point_stat_cts_plots_weekly_monthly.sh

5) point_stat_mpr_plots_weekly_monthly.sh

4. Case study of Tropical Cyclone ‘Phailin’

4.1 Rainfall distribution

Tropical Cyclone 'Phailin' made landfall between 12th and 13th October, 2013, by which time it had already attained its maximum intensity and started gradual dissipation and its north-east ward movement over the land mass. The geographical distribution of the associated rainfall soon after the landfall with maximum rainband concentrated over its north east sector is clearly evident in Fig. 2. which shows the corresponding rainguage station observations, gridded rainfall analysis at 0.5 X 0.5 degree resolution and the forecasts by NGFS and NCUM for day-1, day-3 and day-5 valid for 13 October, 2013 and smoothed at the analysis resolution. In all the plots, the color scale is identical except that the station observations are plotted as coloured circles and all other gridded data are plotted as color shaded regions. The station observations show only raingauge reports and does not include the automatic weather stations (AWS). The forecasts valid for the same period show the associated rainfall with maximum contour in the range of 16-32 cm as the analysis. However, the location of the maximum rainfall does not match exactly with the observed rainfall analysis due to slight error in the landfall location and timings. As the forecast lead time increases, the location error also increases and at day-7, NGFS heavy rainfall patch is over north central Bay of Bengal whereas the NCUM failed to predict any circulation of cyclonic intensity other than a feeble low over Bay of Bengal.

The frequency distribution of the forecast and observed rainfall values over the different ranges of rainfall values for the entire period of 8-14 October, 2013 is plotted as histograms in Fig. 3, along with the scatter diagrams and quantile-quantile (Q-Q) plots. It can be seen that, in general, some observed higher rainfall values beyond 50mm are less seen in NGFS forecast distributions. The Fig. 3 shows that the frequency of predicted rainfall values for NCUM have a better match with the observations and also produce better association as shown by the curve in Q-Q plot lying more closely with the diagonal. The figure does not show day-to-day match for the models, but it is clear from the distribution that the overall performance of NCUM is superior to NGFS.

4.2 Traditional verification scores

The continuous verification scores are those based on the difference between the forecast and the observations. The various scores computed and analysed over the period of 8-14 October 2014, for NGFS and NCUM against gridded rainfall analysis are shown in Fig. 4, for each forecast lead time. The first panel shows mean forecast (FBAR) as bars and mean observation (OBAR) as dashed horizontal lines. NCUM slightly over predicts the domain mean rainfall till day-5 and under predicts after that. Values averaged for all lead times are written as texts in the corresponding panels which show the values of domain mean rainfall for NGFS (7.98 mm) , NCUM (8.88 mm) and OBAR (7.96 mm). Mean error shows similar results with NCUM producing more bias till day-5 while correlation coefficient (PR_CORR) shows higher values for NCUM at all lead times till day-6. Multiplicative bias (MBIAS) also is high for NCUM at all lead times. However, RMSE and Error Standard Deviation (ESTDEV) are all lower for NCUM compared to NGFS at all lead times.

Categorical verifications statistics (Fig. 5) averaged for all 7 days lead times displays the skill and error at a number of threshold categories 1, 5, 10, 50 and 100 mm. Mean forecast (FMEAN) shows more grid points covered with lower threshold (1 and 5 mm) rainfall values for NCUM compared to NGFS and the spread of light rainfall areas is more for NCUM. The contour lines (observations) show better match for NCUM at 1 mm threshold while NGFS shows slightly lower frequency of rainfall for 1 and 5 mm thresholds compared to NCUM. At 5mm threshold, both are over predicting the number of rainy gridpoints and at highest thresholds, the sample size is too small to get any statistically significant result. Accuracy (ACC) is marginally high for NGFS especially at lower thresholds. Frequency bias (FBIAS) is marginally high for NCUM at all thresholds except for 100 mm threshold, indicating relatively less coverage of intense rainfall grid points by NCUM during the 7 days period. False Alarm Ration (FAR) is higher at lower thresholds for 1 and 5 mm thresholds and lesser for higher thresholds for NCUM. Similarly all skill scores (CSI, GSS, HK and HSS) show marginally better skill for NGFS at lower thresholds of and 5 mm and poor skill at higher thresholds.

4.3 Method for Object-based Diagnostics Evaluation (MODE)

The MODE is described in detail in Appendix – II. Objects are "regions of interest" as identified by MODE at a particular threshold and its area is determined by the number of grid points in the contiguous area covered by the object. For each pairs of objects, the total interest can be computed by the formula as described in the appendix for both observations and forecasts. Cluster refers to a collection of objects (shown in the same color in a single field and enveloped by a contour) which satisfies the minimum threshold total interest of 0.7 when compared with each other in the same field, either observation or forecast. Fig. 6 (a & b) shows the simple objects generated by the MODE analysis tool for gridded rainfall analysis, and day-1, day-3, day-5 and day-7 forecasts by NGFS and NCUM at various convolution thresholds, 2mm, 4mm, 10mm, 20mm and 50mm valid for 13 October 2013, the most intense period before landfall. At 2mm threshold, the objects cover maximum area and are clustered together over a region covering most of the domain. The observation object cluster contains 3-4 objects of the same colour which are compared against the day-1, day-3 and day-5 forecast clusters of NGFS and NCUM. In general the clusters occupy larger areas in the forecasts compared to the observation and the total interest computed will be the maximum for the lowest threshold. As it goes to higher and higher thresholds, the object areas and the cluster sizes decrease and the total interest also diminishes very fast. At highest threshold of 50mm, the objects are very less in both number and area coverage and at longer lead times of day-5 and beyond the cluster itself is not formed often due to the very low total interest between the objects in the same fields and hence the match is not being made.

A number of attributes can be considered for bringing in some objectivity in the evaluation of the errors in intensity, pattern and location. This includes centroid difference, angle difference, intersection area and Total Interest between the forecast and observation objects computed between the simple objects in the respective fields of any forecast and the corresponding observation. Table 1 gives the tabulated form of these statistics at various convolution thresholds and at forecast lead times of 1, 3, 5 and 7 days. In general, NCUM is found to have better statistics up to day-5 and thereafter the higher thresholds could not even generate the simple objects. Table 2 gives averaged frequency bias, skill scores (CSI) and total matched area and matched number of objects up to day-5, which gives mixed results and no definite conclusions. Table 3 shows the averaged statistics of centroid difference, angle difference, intersection area and total interests for all lead times separated into different thresholds which shows again mixed results.

Median of Maximum Interest (MMI) can be considered as a single objective score to assess the general agreement of all the forecast objects in the entire domain with the observed objects. This is because, MMI accounts for all the attributes of the forecast-observation pairs - characteristic of the errors in location, orientation and intensity distribution of the simple objects (See Appendix – II). Table 4 lists the Median of Maximum Interest (MMI) values for day-1, day-3, day-5 and day-7 forecasts valid for 13 October 2013 for thresholds 1, 5, 10, 20 and 50mm. In general it can be seen that NCUM features better performance with higher number of MMIs while at day-7 NCUM fails to produce any strong system and hence has poor MMI at higher thresholds of 20mm and above. Though NGFS produced the Tropical Cyclone in day-7 forecasts, it probably failed to produce Total Interests for the simple objects at higher rainfall range so as to exceed the threshold value of

0.7 to make a cluster. This may be due to the higher weights given to the centroid distance and other parameters being taken in to consideration for the computation of Total Interest as the location, not the intensity, is in more error in this case. Thus in general, except at longer lead times beyond day-5, the MMI values are above a useful threshold value of 0.7 and can be considered as a measure of better performance. Also it can be seen that NCUM produced higher values of MMIs compared to NGFS almost at all lead times and on all rainfall thresholds, except on day-7 at higher rainfall ranges, in which case, NGFS scores are predictably higher owing to the better intensity forecasts.

Traditional verification scores can be applied to the model output rainfall computed by defining matched observed objects to be 'hits', unmatched observed objects to be 'misses' and unmatched forecast objects to be 'false alarms'. Fig. 7 gives statistics of time variation of total number of matched and unmatched objects captured during the period of TC Phailin for for 24-hour rainfall forecasts with four thresholds 2, 4, 10 and 15mm. In general it can be stated that, for the case of number of matched objects, 24-hour rainfall features more number of hits for NGFS for the entire episode. Also the number of misses and false alarms are less for NGFS compared to NCUM.

4.4 Intensity-scale verification

Wavelet stat analysis evaluates the forecast skill as a function of the precipitation intensity and the spatial scale of error. Usually large scales exhibit positive skill and small scales exhibit negative skill. The worst skill score is normally associated with the smallest scale. Different scales are associated with different physical processes. For example, small scales are associated with convective showers and mesoscale events and large scales are associated with frontal systems and other large scale synoptic systems. Any weather phenomena can be considered as consisting of all range of scales from micro to the maximum size of the event. The wavelet analysis carried out at a finite number of scales will quantify the performance of the rainfall forecasts for these scales and will give an idea about the scales of maximum and minimum average displacement error at each threshold. Different categorical scores are computed for each particular scale component, like Intensity-scale Skill score (ISS) which is based on the mean squared error (MSE). Thus, this approach enables the user to assess the skill of the model in simulating these scales and hence the associated physical processes. Fig. 8 shows intensity skill score plotted against the scale in kilometres as bar diagrams for day-1, day-3 and day-5 forecasts by NGFS and NCUM. Different colour bar for each scale denotes different thresholds of 0.1, 1, 5, 10, 20 and 50mm. At day-5, both the models show considerable degradation in the skill to simulate the scales even up to 800km. NCUM shows poorer skill at 800km scale compared to NGFS. In general, both the models show better capability to simulate scales of 800Km and 1600Km at all thresholds and NGFS on average, shows relatively better skill in 800Km scale compared to NCUM. This may be partly due to the difference in the resolution of the models, as NGFS runs at comparatively higher resolution. At 50km scale, both models show the least skill perhaps due to more displacement error. Averaged over the entire episode (not shown here), NGFS shows lower skill at higher thresholds compared to NCUM at 50 km scale.

5. Summary and conclusions

The current report describes the new ways of model performance evaluation and more comprehensive analysis techniques. The installation of MET software, the data preprocessing tools and the associated graphics for the performance comparison of two deterministic models NGFS and NCUM are described and some results from a case study of Tropical Cyclone 'Phailin' are discussed. The entire processing is carried out by two master scripts, one for computing and storing the statistics and the other for generating the essential visualisations and graphics.

These types of standard scores are useful in assessing the overall quality of the forecasts for the kind of extreme weather events which last for at least about a week. However, as the sample size is relatively small, the scores cannot be generalised or a definite statement of the performance of a particular model cannot be arrived at. The object-based scores are useful in day-to-day assessment of the agreement between forecast and observed rainfall patterns and in-depth analysis of the performance in the simulation of various physical processes. A long period forecast experiment or a large sample size can be used to assess the strength and weakness of the models. So an ensemble of scores of large set of extreme weather events or accumulation of scores through a large period like a month or season can help in assessing the performance of the models in different scenarios or different convective environments.

Appendix – I

The traditional verification scores

The continuous measures are based on the difference (forecast – observation). The scores are;

1. Mean Forecast: Mean of forecast values over the number of matched pairs.
2. Mean Observation: Mean of observation values over the number of matched pairs.
3. Mean Error: Measure of overall bias (difference of items 1 and 2).
4. Pearson Correlation coefficient: The correlation between forecasts and observations (Range:-1 to 1 where 1 is perfect correlation).
5. Root Mean Square Error: Square root of the mean squared error between forecast and observations.
6. Multiplicative Bias: Ratio of mean forecast and mean observation.
7. Standard Deviation of error: Bias corrected mean squared error.

The contingency table scores (CTS) are based on the 2x2 contingency table based on different rainfall thresholds like 1, 5, 10, 50, and 100 mm.

Forecast	**Observation**		**Total**
	Yes	**No**	
Yes	n11	n10	n1.
No	n01	n00	n0.
Total	n.1	n.0	T

1. Base rate: Relative frequency of occurrence of the event, $\bar{o} = \frac{n11+n01}{T}$.
2. Mean forecast: Relative frequency of occurrence of forecast of the event, $\bar{f} = \frac{n11+n10}{T}$.
3. Accuracy: Proportion of forecasts that are either hits or correct rejection, ACC= $\frac{n11+n00}{T}$. (Range:0-1, perfect value=1).
4. Frequency Bias: Ratio of total number of forecasts of an event to total number of observations of the event, $FBIAS = \frac{n11+n10}{n11+n01}$. (perfect value=1).
5. Probability of detection of "yes" events: Fraction of events that were correctly forecasted to occur, $POD = \frac{n11}{n11+n01}$r. (Range:0-1, perfect value=1).
6. Probability of detection of "no" events: Fraction of non-events that were correctly forecasted to be non-events, $PODN = \frac{n00}{n10+n00}$.(Range:0-1, perfect value=1).
7. Probability of False Detection: Fraction of non-events that are forecasted to be events, $POFD = \frac{n10}{n10+n00}$. (Range:0-1, perfect value=0).
8. False Alarm Ratio: Proportion of forecasts of the event occurring for which the event did not occur, $FAR = \frac{n10}{n10+n11}$. (Range:0-1, perfect value=0).
9. Critical Success Index: Ratio of the number of times the event was correctly forecasted to occur to the number of times it was either forecasted or occurred, $CSI = \frac{n11}{n01+n10+n11}$. (Range:0-1, perfect value=1).
10. Gilbert Skill score: GSS is known as Equitable Threat Score (ETS) and is based on the CSI, corrected for the number of hits that would be expected by chance, $GSS = \frac{n11-C1}{n01+n10+n11-C1}$, where, $C1 = \frac{(n11+n10)(n11+n01)}{T}$. (Range:-1/3 - 1, perfect value=1).
11. Hanssen-Kuipers Discriminant: HK measures the ability of the forecast to discriminate between (or correctly classify) events and non-events, $HK = \frac{n11.n00-n10.n01}{(n11+n01).(n00+n10)}$. (Range:-1 - 1, perfect value=1).

12. Hiedke Skill score: HSS is a skill score based on Accuracy, where the Accuracy is corrected by the number of correct forecasts that would be expected by chance, $HSS = \frac{n11 + n00 - C2}{T - C2}$, where, $C2 = \frac{(n11+n10)(n11+n01)+(n00+n10)(n00+n01)}{T}$. (Range:-∞ - 1, perfect value=1).
13. Odd's ratio: measures the ratio of the odds of a forecast of the event being correct to the odds of a forecast of the event being wrong, $ODDS = \frac{n11xn00}{n10xn01}$. (Range:0-∞, perfect value=∞).

Appendix – II

Method for Object-Based Diagnostics Evaluation (MODE)

This is a displacement technique of spatial verification methods which provides information which is not otherwise possible to obtain using traditional grid-point based verification methods. It objectively identifies simple objects in rainfall fields at different thresholds, which would mimic what humans call as "regions of interest". This process is a multistep one which is called the 'convolution-thresholding' technique. It basically involves application of a simple circular filter which in terms is a function of convolution radius (CR). Once the filter is applied, the convolved field is thresholded using a convolution threshold (CT) to generate a mask field. These simple objects are the connected regions of "1" in the mask field. Finally, the actual data is restored inside the mask regions of object interiors to obtain the object field. Thus these objects are a function of CR and CT.

Once simple objects are generated in the rainfall field, various object attributes are computed and compared to merge the objects in the same field and match the objects between the two different fields, say forecast and observation. The summary statistics can be computed based on the single object statistics as well as statistics of the pairs of objects. As an example, Area is an attribute which is simply the count of the number of grid squares an area of object occupies. Axis angle gives the tilt of the object as curvature implies the curviness. Aspect ratio is the ratio of the width and length of the rectangle which is aligned so as to have the same axis angle as the object and for which the length and width are chosen so as to just enclose the object. Complexity is defined by comparing the area of an object to the area of its convex hull. Similarly pair attributes are defined such as centroid distance, angle difference, union area, intersection area and symmetric difference.

Matching and merging of the objects are achieved by various techniques and "Fuzzy engine" logic is applied for the current study. This involves assigning "interest maps", "confidence maps" and weights for the attributes (α) which are taken in to consideration. Interest maps (I(α)) range from zero to one and are applied to each attributes to convert it into interest values. 1 indicates high interest and 0 indicates no interest and there will be some attributes featuring intermediate interests. Confidence maps (C(α)) also range from zero to one, but is a function of the entire set of attributes to indicate the relative confidence of one field in terms of other fields thus is dependent of other parameters also. By default if the attribute is independent of any other attribute, the confidence map is defined as 1. The scalar weights (ω) are to be assigned to each attribute giving preference to which attribute the user assign maximum weightage. Finally a single number called total interest (T(α)) is computed using all the three maps by a formula as given below.

$$T(\alpha) = \frac{\sum_i w_i C_i(\alpha) I_i(\alpha)}{\sum_i w_i C_i(\alpha)}$$

The total interest is then thresholded and the pairs of objects that are having total interest more than the threshold are merged if they are in the same field and matched if they are in the different fields. MODE outputs the statistics of single as well as cluster of objects. The scores can be summarised as Median of Maximum Interest (MMI), which is an example of a useful single measure of the general agreement between forecast and observation for the entire domain [See Davis et al., 2009]. Median is taken instead of Mean to avoid the effect of outliers.

The details of the MODE configurations and the definitions of various interest maps and confidence maps are as given below.

General:

Grid_res = 50 Km

Convolution thresholds= 1, 2, 5, 10, 20 and 50 mm

Convolution radius= 2 (grid spaces)

Forecast_merge_flag=2 (Fuzzy Engine merging method)

max_centroid_dist=200

total_interest_thresh=0.7

Interest functions and piecewise linear functions:

Centroid Distance	Interest
0.0	1.0
100.0/grid_res	1.0
1000.0/grid_res	0.0

Boundary Distance	Interest
0.0	1.0
500.0/grid_res	1.0
2000.0/grid_res	0.0

Convex Hull Distance	Interest
0.0	1.0
500.0/grid_res	1.0
2000.0/grid_res	0.0

Angle Difference	Interest
0.0	1.0
30.0	1.0
90.0	0.0

Area Ratio	Interest
0.0	0.0
1.0	1.0

Intersecting area ratio	Interest
0.00	0.00
0.10	0.50
0.25	1.00
1.00	1.00

Confidence functions :

aspect_ratio_conf(t) = ((t - 1)**2/(t**2 + 1))**0.3;

area_ratio_conf(t) = t

Weights:

centroid_dist_weight=2.0

boundary_dist_weight=4.0

convex_hull_dist_weight=0.0

angle_diff_weight=1.0

area_ratio_weight=1.0

int_area_ration_weight=2.0

complexity_ratio_weight=0.0

intensity_ratio_weight=0.0

References

Ahijevych, D. Gilleland, E. Brown, B. and Ebert, E., 2009, "Application of spatial forecast verification methods to gridded precipitation forecasts", *Wea. Forecasting*, **24** pp. 1485–1497.

Brown, B.G. Bullock, R. Gotway, J.H. Ahijevych, D. Davis, C. Gilleland, E. and Holland, L., 2007, "Application of the MODEL object-based verification tool for the evaluation of model precipitation fields", *AMS 22nd Conference on Weather Analysis and Forecasting and 18th Conference on Numerical Weather Prediction, 25-29 June, Park City, Utah, American Meteorological Society (Boston) (Available at http://ams/confex.com/ams/pdfpapers/124856.pdf).*

Casati, B., 2010, ”New developments of the intensity-scale technique within the spatial verification methods intercomparison project”, *Wea. Forecasting*, **25** pp. 113–143.

Casati, B. Ross, G. and Stephenson, D.B., 2004, “A new intensity-scale approach for the verification of spatial precipitation forecasts”, *Meteor. Appl.,* **11** pp. 141–154.

Casati, B. and Co-authors, 2008, “Forecast verification: Current status and future directions”, *Meteor. Appl*., **15** pp. 3–18.

Davis, C. A. Brown, B.G. and Bullock, R.G., 2006, “Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas”, *Mon. Wea. Rev*., **134** pp. 1772–1784.

Davis, C. A. Brown, B.G. Bullock, R.G. and Gotway, J.H., 2009, “The method for object-based diagnostic evaluation (MODE) applied to WRF forecasts from the 2005 NSSL/SPC Spring Program”, *Wea. Forecasting*, **24** pp. 1252-1267.

Ebert, E.E., 2008, “Fuzzy verification of high resolution gridded forecasts: A review and proposed framework”, *Meteor. Appl*., **15** pp. 51–64.

Ebert, E.E., 2009, “Neighbourhood verification: A strategy for rewarding close forecasts”, *Wea. Forecasting*, **24** pp. 1498–1510.

Ebert, E.E. and Gallus, W.A. Jr., 2009, “Toward better understanding of the contiguous rain area (CRA) method for spatial forecast verification”, *Wea. Forecasting*, **24** pp. 1401–1415.

Ebert, E.E. and McBride, J.L., 2000, “Verification of precipitation in weather systems: Determination of systematic errors”, *J. Hydrol.*, **239** pp. 179–202.

Gallus, W.A. Jr., 2010, “Application of object-oriented verification techniques to ensemble precipitation forecasts”, *Wea. Forecasting*, **25** pp. 144-158.

Gilleland, E., 2013, "Testing Competing Precipitation Forecasts Accurately and Efficiently: The Spatial Prediction Comparison Test", *Mon. Wea. Rev.,* **141**, 1 pp. 340-355.

Gilleland, E. Ahijevych, D. Brown, B.G. Casati, B. and Ebert, E.E., 2009, "Intercomparison of spatial forecast verification methods", *Wea. Forecasting*, **24** pp. 1416–1430.

Gilleland, E. Ahijevych, D.A. Brown, B.G. Ebert, E.E., 2010a, "Verifying Forecasts Spatially", *Bull. Amer. Meteor. Soc.*, **91** pp. 1365–1373.

Gilleland, E.D. Lindström, J. and Lindgren, F., 2010b, "Analyzing the image warp forecast verification method on precipitation fields from the ICP", *Wea. Forecasting*, **25** pp 1249–1262.

Jolliffe, I.T and Stephenson, D.B., 2003, "Forecast verification: A practitioners' guide in atmospheric science", *Wiley and Sons Ltd*, 240pp.

Murphy, A.H. and Winkler, R.L., 1987, "A general framework for forecast verification", *Mon. Wea. Rev.*, **115** pp. 1330-1338.

Mitra, A.K. Das Gupta, M. Singh, S.V. and Krishnamurti, T. N., 2003, "Daily rainfall for Indian monsoon region from merged satellite and raingauge values: Large-scale analysis from real-time data", *J. Hydrometeorol.*, **4(5)** pp. 769–781.

Mitra, A.K. Bohra, A.K. Rajeevan, M.N. and Krishnamurti, T.N., 2009, "Daily Indian precipitation analyses formed from a merge of rain-gauge with TRMM TMPA satellite derived rainfall estimates", *J. Meteor. Soc. Japan*, **87A** pp. 265–279.

Mittermaier, M. and Roberts, N., 2010, "Intercomparison of spatial forecast verification methods: identifying skilful spatial scales using the fractions skill score", *Wea. Forecasting,* **25** pp. 343-354.

Mohandas, S. and H. Singh, 2015, "Spatial verification of rainfall forecasts for the case of very severe cyclonic storm 'Phailin'", *Mausam*, **66**, 3, 369-384.

Prasad, V.S. Mohandas, S. Das Gupta, M. Rajagopal, E.N. and Datta, S.K., 2011, “Implementation of upgraded Global Forecasting Systems (T382L64 and T574L64) at NCMRWF”, *NCMRWF Technical Report No. NCMR/TR/5/2011, National Centre for Medium Range Weather Forecasting, Min. of Earth Sciences, A-50, Sector-62, Noida, May 2011*, 72pp.

Prasad, V. S. Mohandas, S. Dutta, S.K. Das Gupta, M. Iyengar, G.R. Rajagopal, E.N. and Basu S., 2013, “Improvements in Medium Range Weather Forecasting System of India”, *Journal of Earth System Science,* **123**, 2, Mar 2013 pp. 347-258.

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 Unified Model based Analysis-Forecast System at NCMRWF”, *NCMRWF Technical Report No. NMRF/TR/2/2012, NCMRWF, Ministry of Earth Sciences, A-50, Sector-62, Noida, UP-201309, May 2014,* 45pp.

Stansky, H.R Wilson, L.J. and Burrows, W.R, 1989, “Survey of common verification methods in meteorology”, *World Weather Watch Tech. Rept. No. 8, WMO/TD No. 358, WMO Geneva*, 114pp.

Wilks, D., 2006, “Statistical methods in the atmospheric sciences”, *Elsevier, San Diego, Second Edition, International Geophysics Series,* **91**, 630pp.

-------------------0--------------------

Table 1 The Centroid distance, Angle difference, Intersection area and Total Interest for Convolution Thresholds (CT) 2, 10, 20, 50mm for days of forecast 1, 3, 5 and 7 for global models NGFS (G) and NCUM (U) averaged for forecasts from initial conditions 8-14, October, 2013.

Day	**CT mm**	**Cen Dis**		**Ang Dif**		**Int Are**		**Tot Int**	
		G	**U**	**G**	**U**	**G**	**U**	**G**	**U**
1	**2**	7.89	6.22	16.38	55.70	847	798	0.9310	0.9457
	10	9.15	4.20	64.85	71.16	254	268	0.8730	0.9383
	20	4.74	5.18	47.26	71.18	112	146	0.9410	0.9163
	50	5.03	4.10	20.86	5.71	28	66	0.9613	0.9514
3	**2**	7.84	4.30	20.35	14.31	731	812	0.9395	0.9704
	10	11.04	9.68	62.66	32.77	210	234	0.8704	0.8982
	20	11.53	8.86	59.10	71.19	99	128	0.8662	0.8653
	50	5.24	3.67	9.52	22.55	37	55	0.9625	0.9833
5	**2**	8.14	5.71	21.62	24.26	692	808	0.9343	0.9952
	10	10.60	12.19	83.00	11.18	154	196	0.8322	0.8828
	20	7.63	13.26	25.75	4.71	58	85	0.9096	0.8713
	50	8.44	7.01	83.59	32.72	11	26	0.8004	0.94s04
7	**2**	6.93	5.97	34.72	20.25	811	800	0.9363	0.9489
	10	2.49	18.07	79.18	80.84	56	134	0.9363	0.8010
	20	-	19.63	-	81.79	-	134	-	0.8010
	50	-	-	-	-	-	-	-	-

Table 2 Average frequency bias, Critical Success Index (CSI), total matched area and total matched number of objects for days 1-5 for models NGFS (G) and NCUM (U) for the period 8-14 October, 2013.

Day	Mod	Freq Bias	CSI	Mat Area Obj	Mat No Obj
1	**G**	1.446	0.631	3061	48
	U	1.4	0.62	3041	51
2	**G**	1.397	0.6	3023	54
	U	1.399	0.587	3030	49
3	**G**	1.316	0.552	2959	55
	U	1.422	0.55	3051	52
4	**G**	1.342	0.523	2968	52
	U	1.401	0.507	3033	51
5	**G**	1.343	0.498	2973	57
	U	1.384	0.487	3034	49

Table 3 Average centroid distance, angle difference, intersection area and total interest for NGFS (G) and NCUM (U) for convolution thresholds (CTs) 2, 4, 10, 20, 30mm for the period 8-14 October, 2013.

CT mm	Mod	Cen Dis	Ang Dif	Int Are	Tot Int
2	**G**	2.34	28.08	1599	0.9927
	U	1.63	9.04	1908	0.9987
4	**G**	2.45	20.62	1311	0.9945
	U	3.45	14.64	896	0.935
10	**G**	6.14	23.39	358	0.9555
	U	5.21	12.82	339	0.9464
20	**G**	6.96	22.3	143	0.9426
	U	7.0	16.84	178	0.9428
30	**G**	4.79	19.77	30	0.9463
	U	4.22	23.01	10	0.8542

Table 4 Median of Maximum Interest (MMI) for NGFS (G) and NCUM (U) for 24, 72, 120 and 168 hour forecasts valid for 13 October 2013 at thresholds 1, 5, 10, 20, and 50mm.

	24HR		**72HR**		**120HR**		**168HR**	
CT(mm)	**NGFS**	**NCUM**	**NGFS**	**NCUM**	**NGFS**	**NCUM**	**NGFS**	**NCUM**
1	0.861	0.8569	0.8398	0.8467	0.8322	0.842	0.8291	0.8383
5	0.8572	0.7822	0.712	0.8605	0.7513	0.8524	0.803	0.8577
10	0.8772	0.9383	0.6381	0.7841	0.6128	0.7969	0.6411	0.8213
20	0.7414	0.9163	0.7772	0.7676	0.5772	0.8712	0.6032	0.5137
50	0.921	0.9514	0.9625	0.9833	0.8004	0.9404	0.6564	0.5426

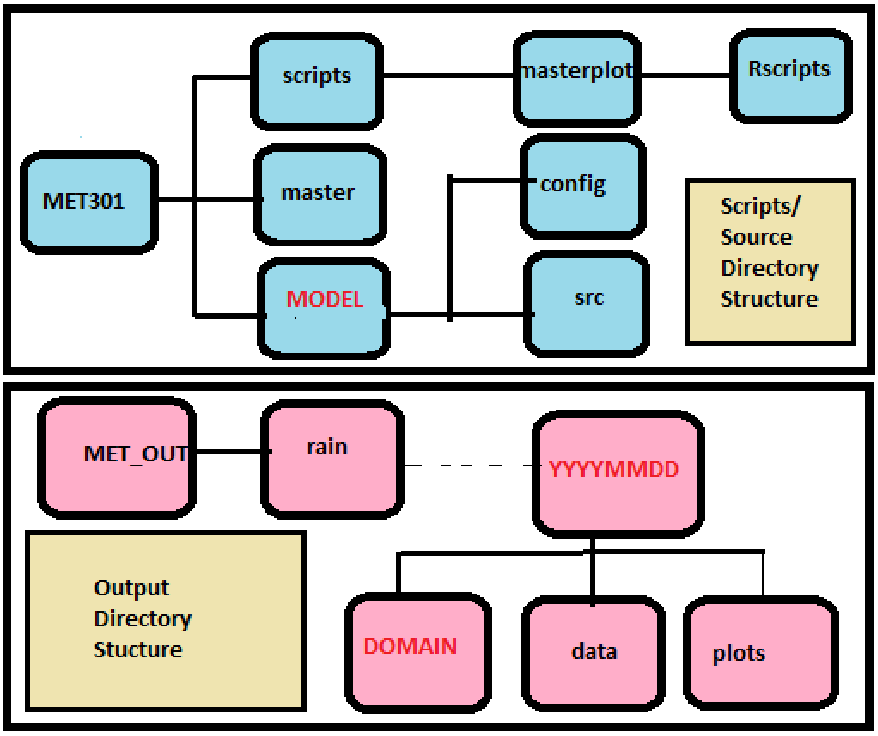


Fig. 1 The flow chart configuration of the MET scripts and MET output directory structures. The red colored capital letters indicates the directory names to be replaced with the date stamp, domain or model name etc. The source codes area is shown in green color and output directories in pink color

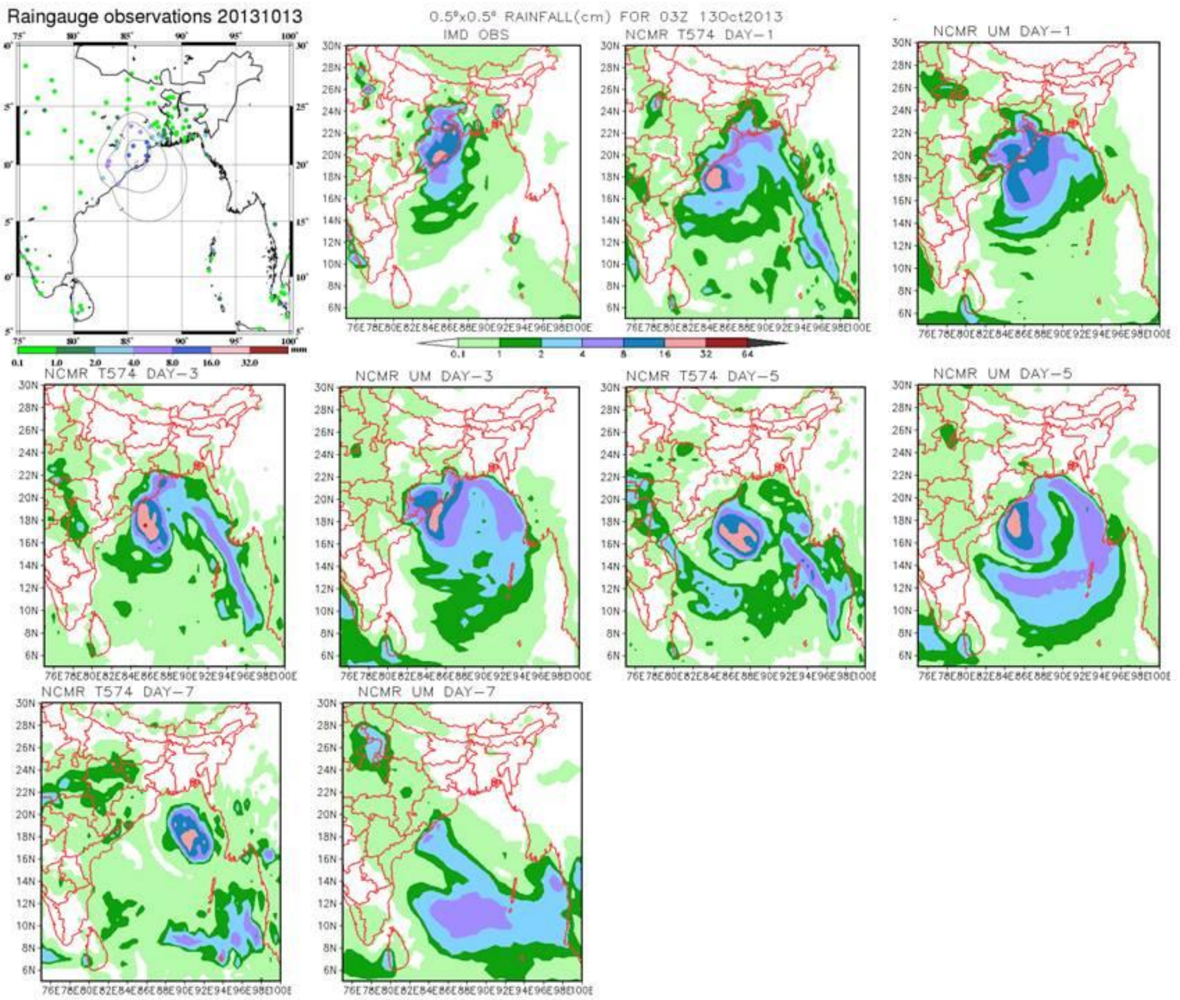


Fig. 2 Geographical distribution of time averainfall raged rainfall valid for 00z 13 October. 2013 at land rainguage stations denoted as coloured circles (a). shaded contours of gridded rainfall analysis (b) and day-1, day-3, day-5 and day-7 forecasts by NGFS (c, e, g and i) and NCUM (d, f, h and j). The shading contours are 0.1, 1, 2, 4, 8, 16, 32, 64 cm.

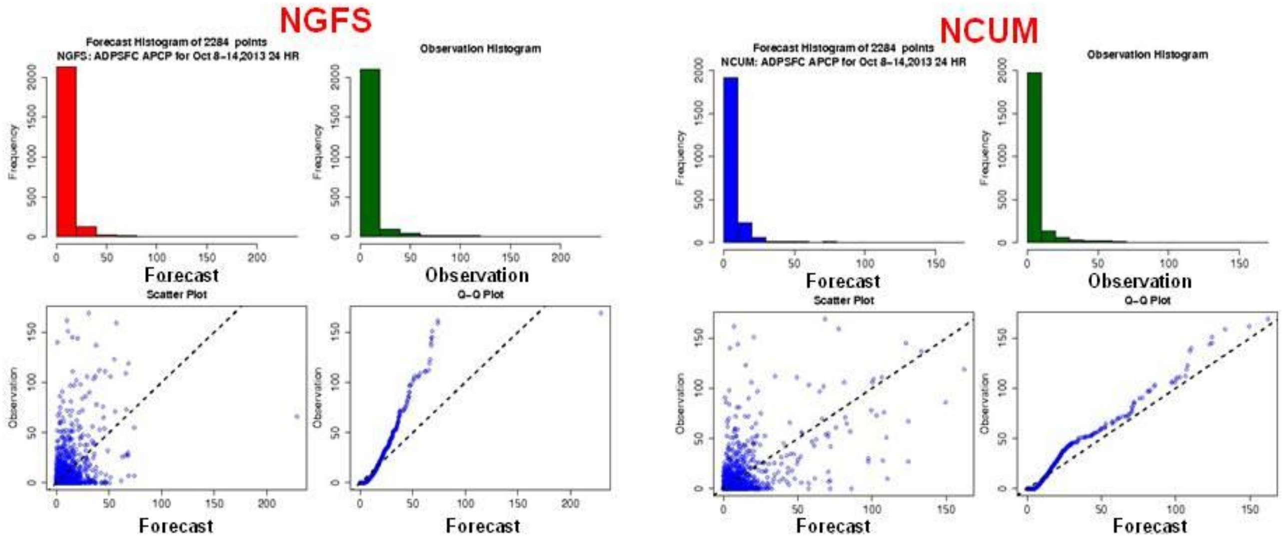


Fig. 3 Histogram, scatter diagrams and Q-Q plots showing the distribution of land raingauge rainfall observations (mm) vs. 24-hour forecasts from NGFS (left square) and NCUM (right square) for the entire period 8-14 October, 2013, including quantile-quantile plots.

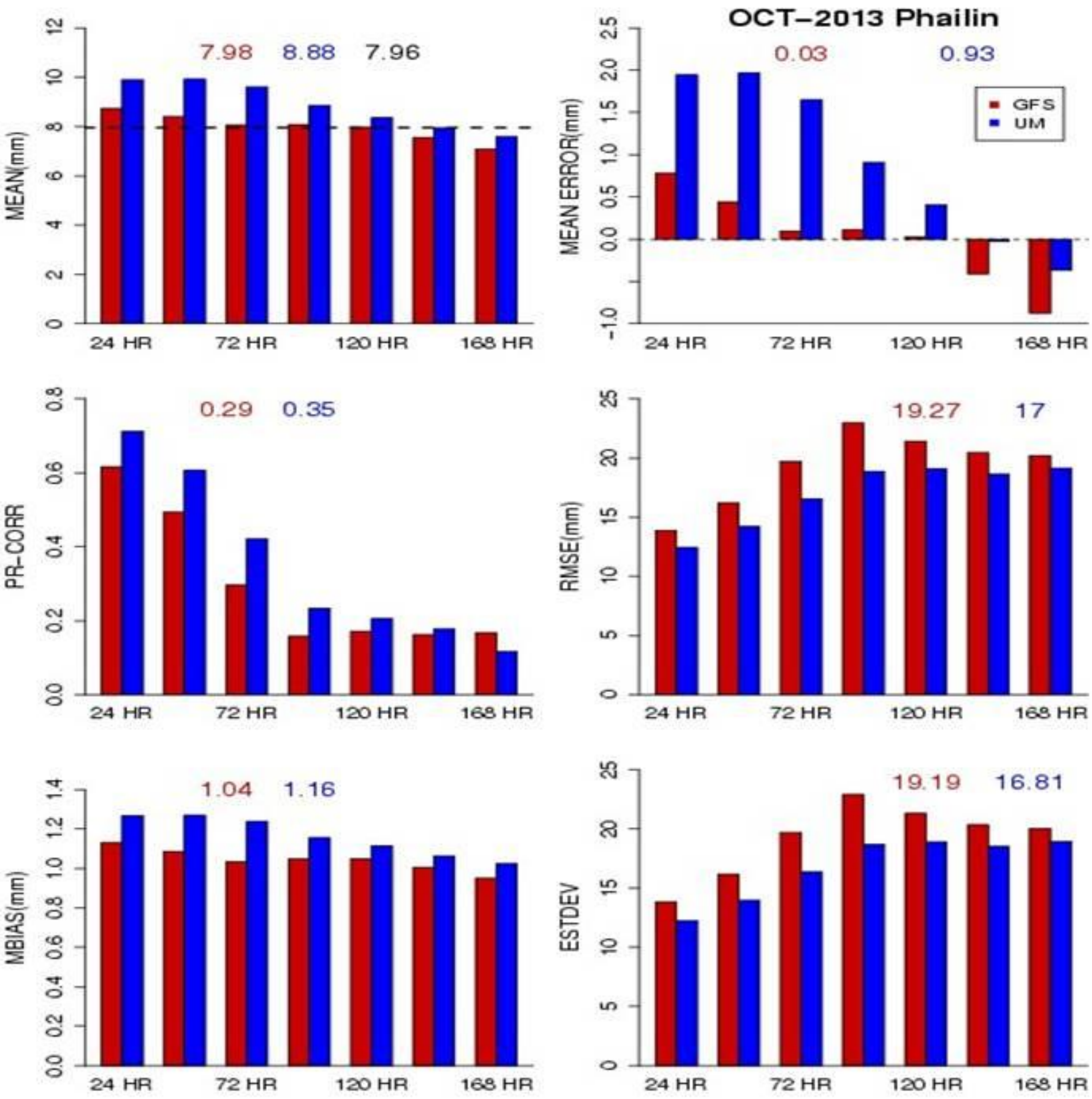


Fig. 4 Continuous verification scores for daily rainfall (mm) against gridded rainfall analysis averaged over the domin 5-30N, 75-100 E for the period 8-14 October, 2013 for NGFS (red) and NCUm (blue) models. Abscissa denotes the forecast lead time and the mean value across all lead times are written in the panels in the order of NGFS, NCUM and Observations respectively: (a) domain meanrainfall (b) mean Error (c) Pearson's correlation (d)root mean squared error (e) multiplicative bias and (f) error standard deviation.

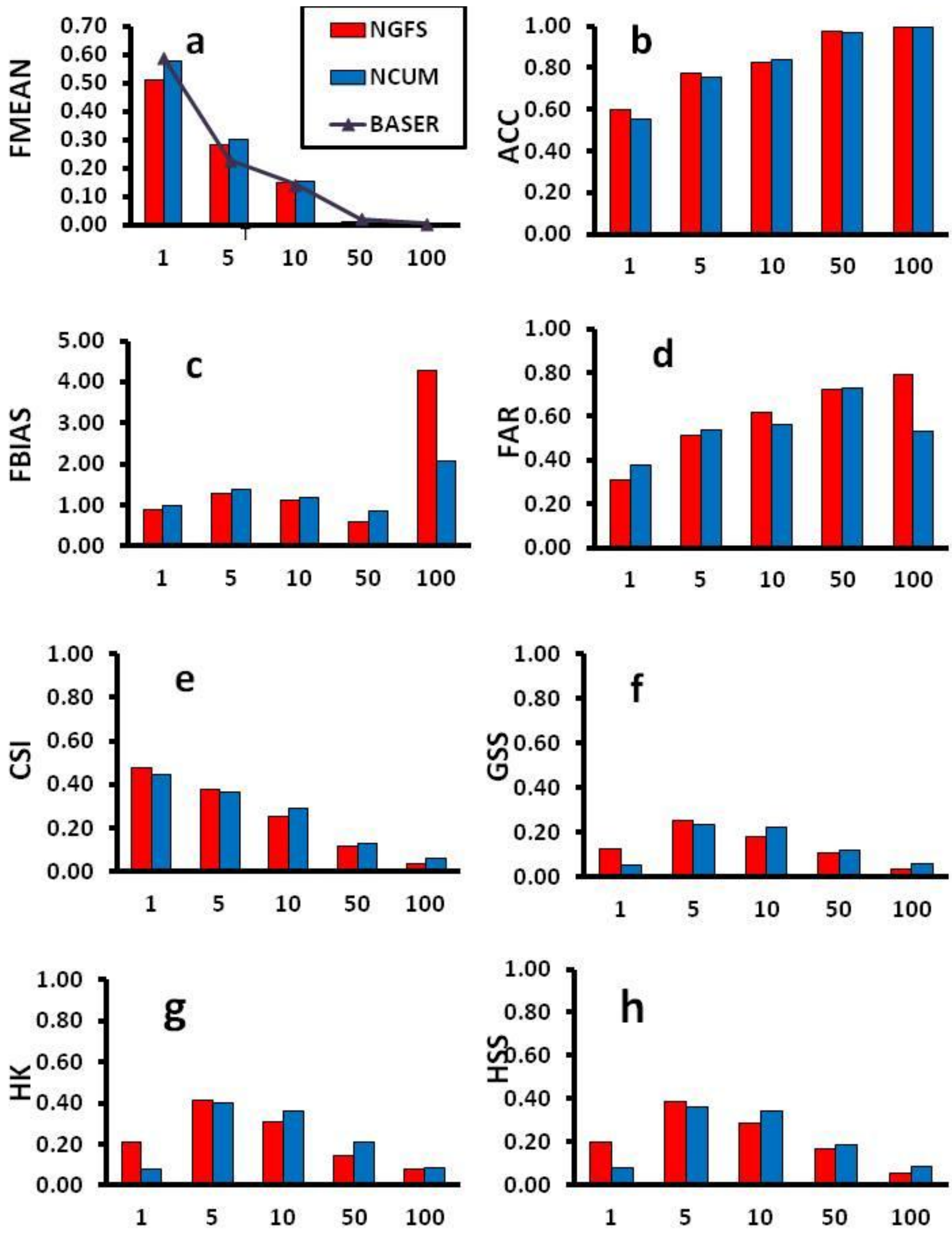
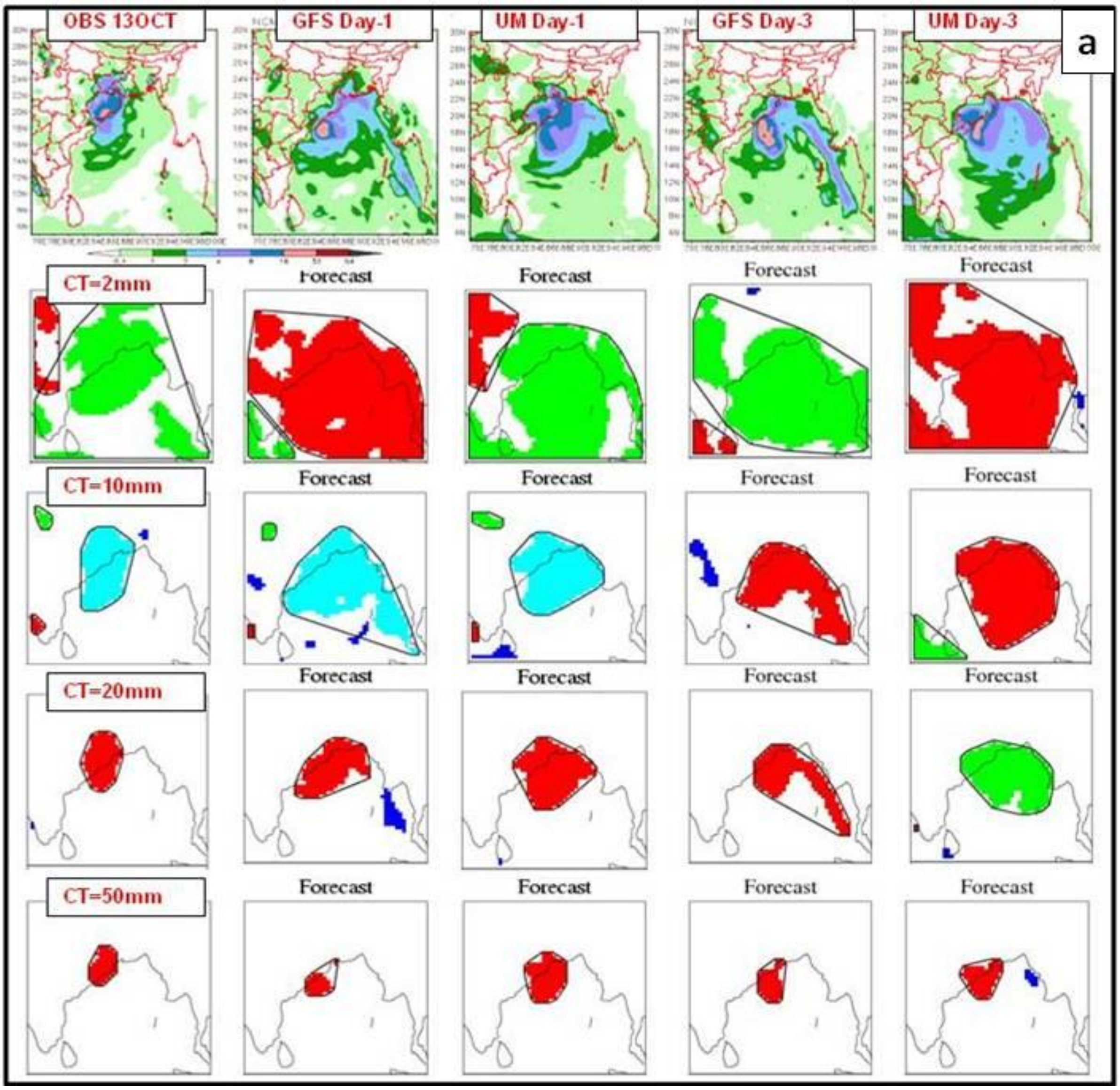


Fig. 5 Categorical verification scores for daily rainfall (millimeters) for different rainfall thresholds during the period 8-14 October, 2013 for NGFS and NCUM; (a) mean forecast (FMEAN) (b) Accuracy (ACC) (c) Frequency Bias (FBIAS) (d) False Alarm Ration (FAR) (e) Critical Success Index (CSI) (f) Gilbert Skill Score (GSS) (g) Hanssen-Kuipers Discriminant (HK) and (h) Heidke Skill Score (HSS). The scores were averaged across all lead times for the purpose of intercomparison.


OBS 13OCT
GFS Day-1
UM Day-1
GFS Day-3
UM Day-3
a
CT=2mm
Forecast
CT=10mm
CT=20mm
CT=50mm

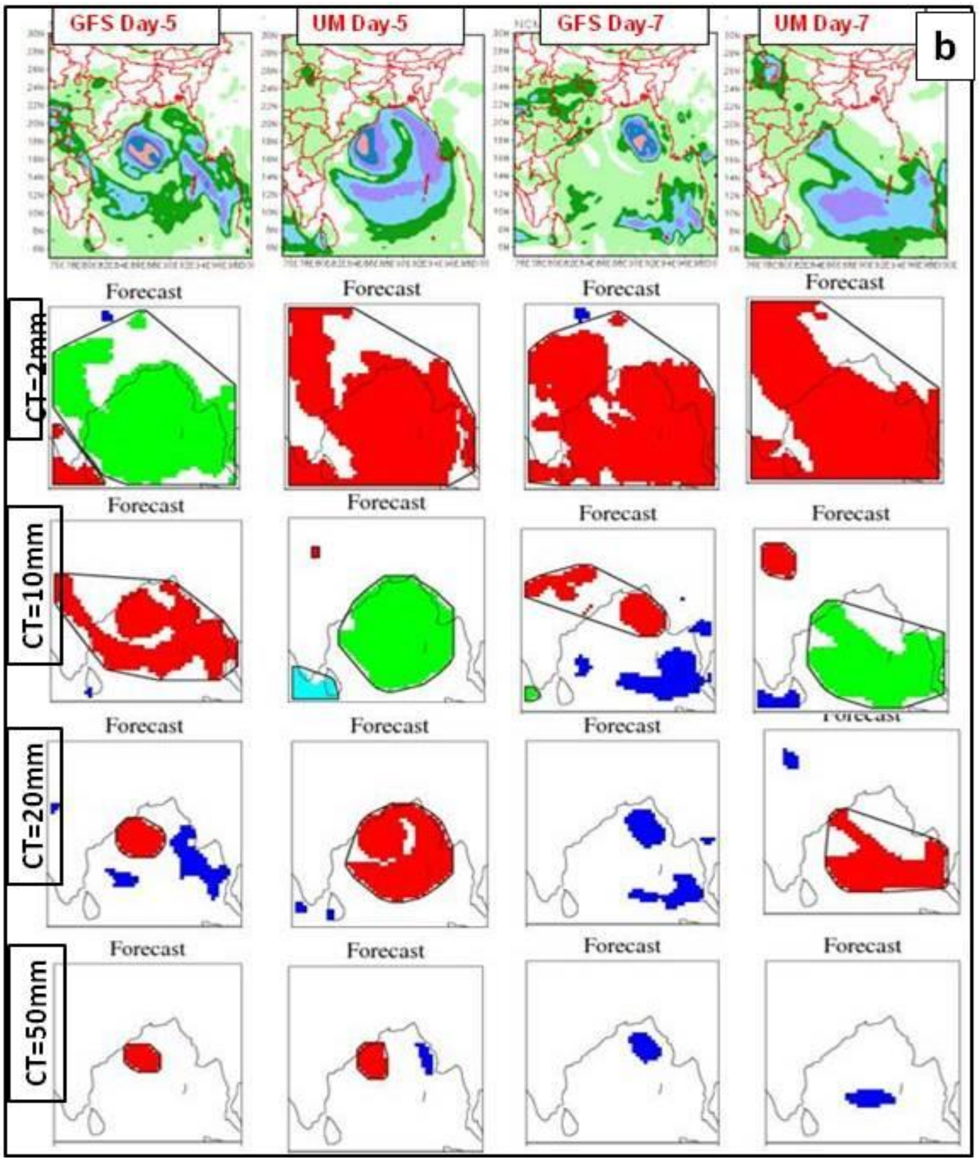


Fig. 6 Observed and forecast rainfall regridded to 50km resolution along with the simple objects and clusters captured at convolution thresholds 2, 10, 20, 50mm. NGFS and NCUM forecasts are shown for (a) day-1 and day-3 and (b) day-5 and day-7 lead times. Rainfall contours are coloured at intervals 0.1, 1, 2, 4, 8, 16 and 32cm and the objects of the same cluster are of single colour in a field. The blue objects are un-clustered ones.

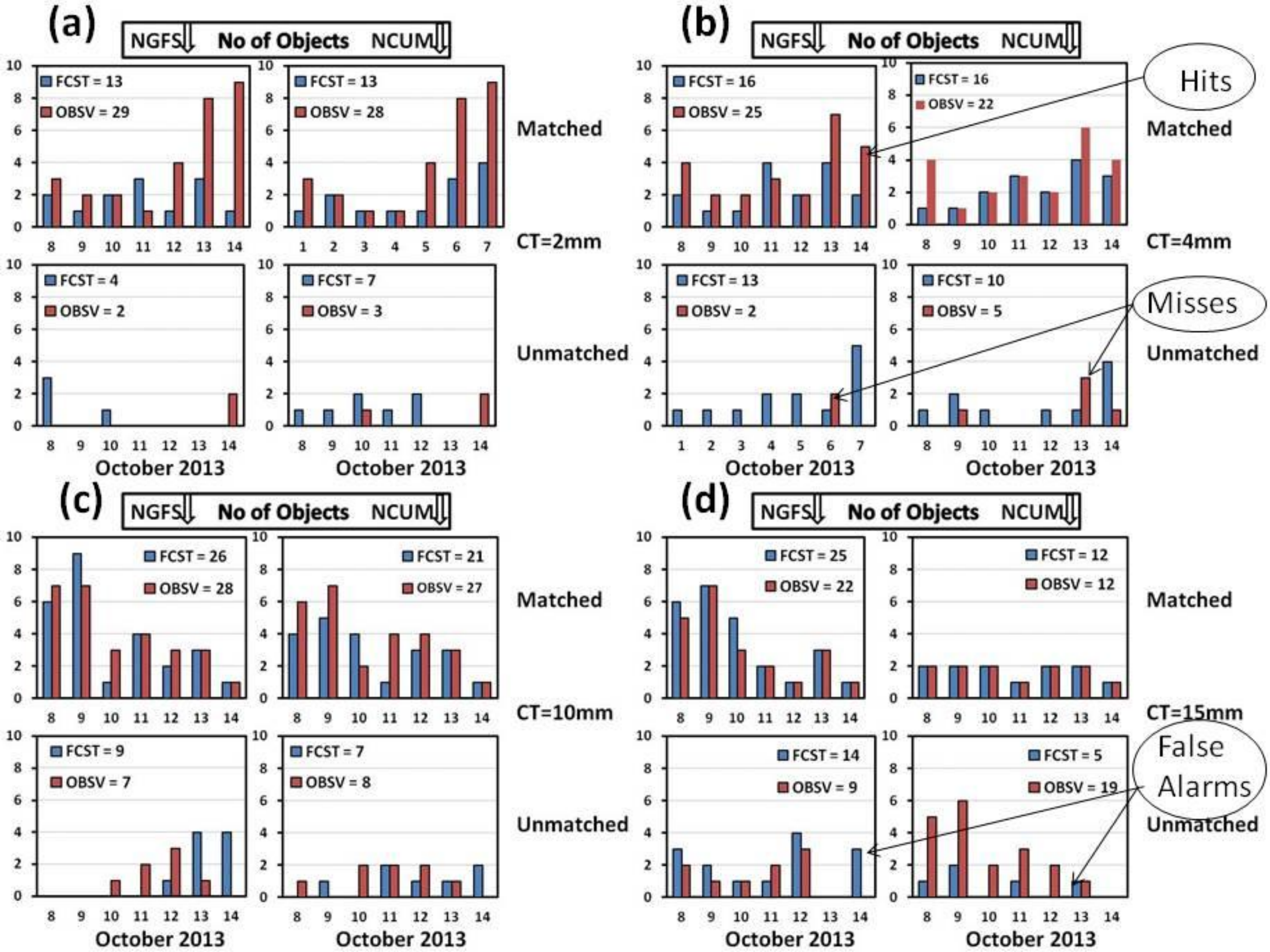


Fig. 7 .The bars denote the time series of the number of matched observed (blue) and forecst (red) objects (top two) and unmatched observed (blue) and forecast (red) objects (bottom two) for 24 hour rainfall forecasts for the period 8-14 October, 2013 by NGFS and NCUM respectively for convolution thresholds (a) 2mm (top left 2x2 square panels) (b) 4mm (top right 2x2 square panels) (c) 10mm (bottom left 2x2 square panels) and (d) 20mm (bottom right 2x2 square panels). Matched number of observation objects are represented as 'hits' and unmatched number of observation objects as 'misses', while the unmatched number of forecast objects are represented as false alarms.

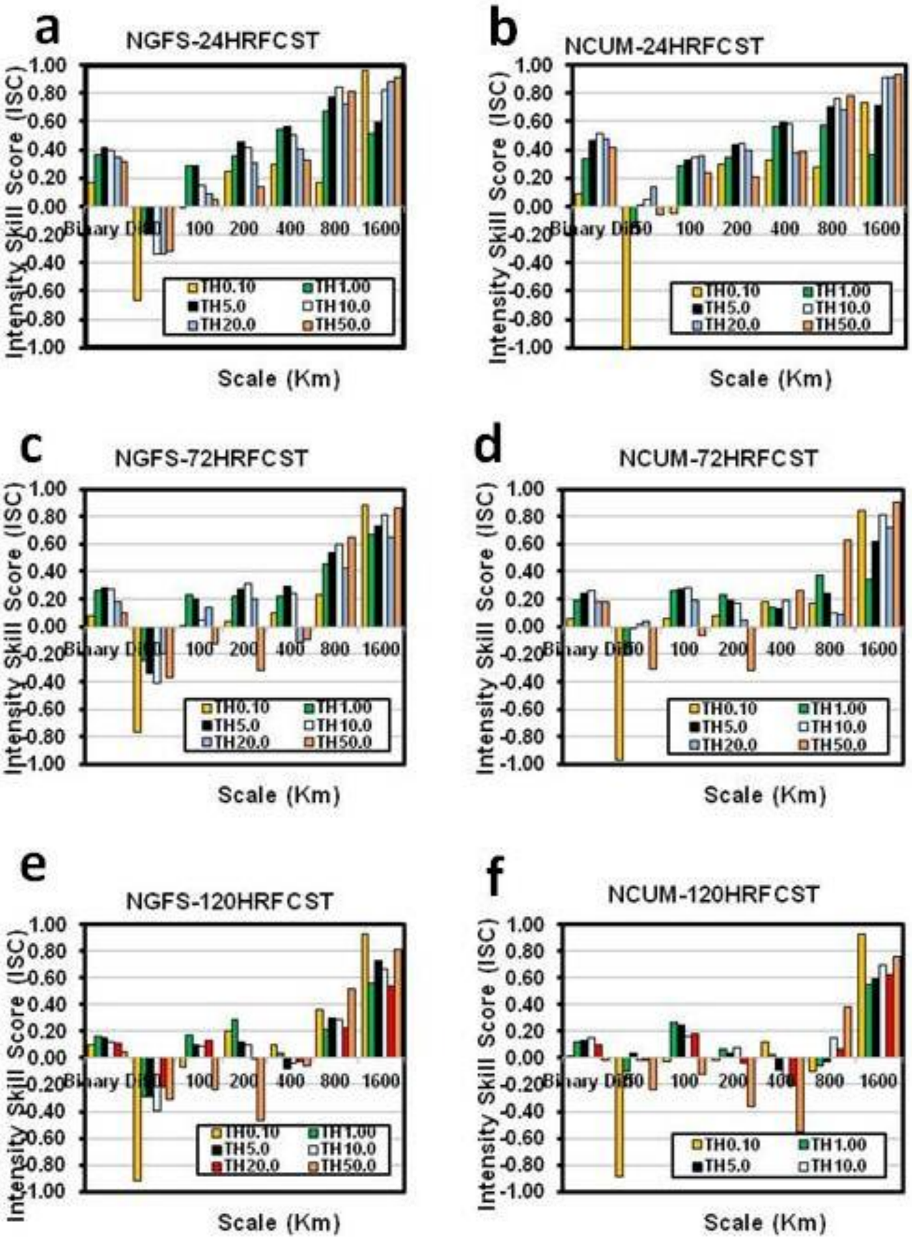


Fig. 8 Intensity-scale skill scores of rainfall at different thresholds (bars) of 0.1, 1, 5, 10, 20 and 50mm, plotted at different scales 50, 100, 200, 400, 800 and 1600km along with the binary difference averaged for 8-14 October 2013 for forecast lead times 24, 72 and 120 hours by NGFS and NCUM.