NMRF/RR/02/2020

RESEARCH REPORT

# MoES Model Forecast Verification during JJAS2019: Verification Statistics Data Base (VSDB)

C. J. Johny, T. Arulalan, and Raghavendra Ashrit

August 2020

National Centre for Medium Range Weather Forecasting
Ministry of Earth Sciences, Government of India
A-50, Sector-62, NOIDA-201 309, INDIA

# MoES Model Forecast Verification during JJAS2019: Verification Statistics Data Base (VSDB)

**C. J. Johny, T. Arulalan, and Raghavendra Ashrit**

**Ministry of Earth Science**
**Earth System Science Organisation**
**Document Control Data Sheet**

| | | |
|---|---|---|
| 1 | Name of the Institutes | National Centre for Medium Range Weather Forecasting, |
| 2 | Document Number | */NMRF/RR/02/2020* |
| 3 | Date of publication | August 2020 |
| 4 | Title of the document | MoES Model Forecast Verification during JJAS2019: Verification Statistics Data Base (VSDB) |
| 5 | Type of Document | Research Report |
| 6 | No.of pages, Figures & Tables | 13 Pages, 8 Figures |
| 7 | Number of References | 9 |
| 8 | Author (S) | C. J. Johny, T. Arulalan, and Raghavendra Ashrit |
| 9 | Originating Unit | NCMRWF |
| 10 | Abstract | This report provides a compilation of the verification for the MoES operational NWP model forecast during monsoon season (JJAS) 2019. The verification and intercomparison results are presented for few large scale upper air fields from the two high resolution deterministic models namely (i) GFS and (ii) NCUM and ensemble models namely (iii) GEFS and (iv) NEPS. These modelling systems are operationally used in IMD for generating real time NWP forecasts in the medium range time scale (up to 10 days).<br><br>As can be expected the ensemble means from GEFS and NEPS show significantly higher skill compared to NCUM over all regions. For predicted 500 hPa HGT over northern & southern hemisphere, GFS forecasts show no significant difference in skill (RMSE and ACC) compared to NCUM forecasts.<br><br>Over the tropics, the forecast skill (RMSE and ACC) in predicted 850 hPa Wind in GFS is mostly comparable to that of NCUM except during first 72 h, when the difference is significant. Skill in ensemble mean are significantly (95%) higher compared to skill in NCUM. For 850 hPa Temperature predictions over tropics and RSMC region, the skill in GFS forecasts is significantly higher compared to skill in NCUM. GEFS forecasts feature best skill for 850 hPa Temperatures over tropics and RSMC region. Murphy's Mean Squared Error Skill Score (MSESS) computed for 850 hPa U, V, Wind and T over RSMC region shows that both the ensemble forecasts show higher score for U & V and WIND. In the case of temperature GEFS shows higher score than all other models, GFS scores are higher than NEPS up to day 5 and NCUM scores are lowest in all the forecasts. Wind Bias is also assessed for different vertical levels. Both the ensembles show similar bias pattern with positive low bias in upper levels and close to surface while having negative bias in middle levels. In middle levels except NCUM all other models exhibit negative bias. |
| 11 | Security classification | Non-Secure |
| 12 | Distribution | Unrestricted Distribution |
| 13 | Key Words | *RMSE, ACC, Bias and VSDB* |

# ABSTRACT

This report provides a compilation of the verification for the MoES operational NWP model forecast during monsoon season (JJAS) 2019. The verification and intercomparison results are presented for few large scale upper air fields from the two high resolution deterministic models namely (i) GFS and (ii) NCUM and ensemble models namely (iii) GEFS and (iv) NEPS. These modelling systems are operationally used in IMD for generating real time NWP forecasts in the medium range time scale (up to 10 days).

As can be expected the ensemble means from GEFS and NEPS show significantly higher skill compared to NCUM over all regions. For predicted 500 hPa HGT over northern & southern hemisphere, GFS forecasts show no significant difference in skill (RMSE and ACC) compared to NCUM forecasts.

Over the tropics, the forecast skill (RMSE and ACC) in predicted 850 hPa Wind in GFS is mostly comparable to that of NCUM except during first 72 h, when the difference is significant. Skill in ensemble mean are significantly (95%) higher compared to skill in NCUM. For 850 hPa Temperature predictions over tropics and RSMC region, the skill in GFS forecasts is significantly higher compared to skill in NCUM. GEFS forecasts feature best skill for 850 hPa Temperatures over tropics and RSMC region. Murphy's Mean Squared Error Skill Score (MSESS) computed for 850 hPa U, V, Wind and T over RSMC region shows that both the ensemble forecasts show higher score for U & V and WIND. In the case of temperature GEFS shows higher score than all other models, GFS scores are higher than NEPS up to day 5 and NCUM scores are lowest in all the forecasts. Wind Bias is also assessed for different vertical levels. Both the ensembles show similar bias pattern with positive low bias in upper levels and close to surface while having negative bias in middle levels. In middle levels except NCUM all other models exhibit negative bias.

## Contents

# 1. Introduction

This report provides a compilation of the verification for the MoES operational NWP model forecast during monsoon season (JJAS) 2019. The verification and intercomparison results are presented for few large scale upper air fields from the two high resolution deterministic models namely (i) GFS and (ii) NCUM and ensemble models namely (iii) GEFS and (iv) NEPS. These modelling systems are operationally used in IMD for generating real time NWP forecasts in the medium range time scale (up to 10 days).

***1.1 MoES Global Forecast Models***

**(a) IMD GFS (T1534) & GEFS**

Global Forecasting System (GFS T1534L64 SL) model run operationally at India Meteorological Department (IMD) twice in a day (00 & 12 UTC) to give deterministic forecast in the short to medium range upto10 days. The forecast model has a resolution of approximately 12 km in horizontal and has 64 levels in the vertical. The initial conditions for this GFS model is generated from the four-dimensional (4D) ensemble–variational data assimilation (DA) system (4DEnsVar) building upon the grid point statistical interpolation (GSI)-based hybrid Global Data Assimilation System (GDAS) run on High Performance Computing Systems (HPCS) at National Center for Medium Range Weather Forecasting (NCMRWF). The real-time GFS T1534L64 model outputs are generated daily at IMD. This 4DEnsVar data assimilation system has capabilities to assimilate various conventional as well as satellite observations including radiances from different polar orbiting and geo-stationary satellites. IMD operationally produces ensemble forecasts at T1534 resolution using 21 member Global Ensemble Forecast System (GEFS) at 00 UTC. GEFS is based on national environmental modeling system (NEMS) configuration, which produces forecasts at Gaussian grid and uses near surface sea temperature (NSST) model to provide oceanic surface temperature. Here initial Ensemble perturbations are generated using Ensemble Kalman Filter (EnKF) method and uses stochastic total tendency perturbation (STTP) to represent model uncertainty. It uses 6 hour forecasts from previous cycle for 20 the perturbed members and control analysis from the current cycle produced from the deterministic GFS system. Model performs tropical cyclone relocation and centers member perturbations on the control analysis. Model uses semi-Lagrangian dynamics and produces 10 day forecasts at ~ 12km resolution.

**(b) NCUM (NCMRWF Unified Model) and NEPS**

The NCMRWF Unified model (NCUM) is a global model and has a horizontal grid resolution of ~12 km with 70 levels in the vertical reaching 80 km height. It uses “ENDGame” dynamical core, which provides improved the accuracy of the solution of primitive model equations and reduced damping. This helps in producing finer details in the simulations of synoptic features such as cyclones, fronts, troughs and jet stream winds. ENDGame also increases variability in the tropics, which leads to an improved representation of tropical cyclones and other tropical phenomena (Walters et al., 2017). An advanced data assimilation method of Hybrid 4D-Var is used for the creation of NCUM global analysis. The advantage of the Hybrid 4D-Var is that it uses a blended background error, blend of "climatological" background error and day-to-day varying flow dependent background error (derived from the 22–member ensemble forecasts). The hybrid approach is scientifically attractive because it elegantly combines the benefits of ensemble data assimilation (flow-dependent co-variances) with the known benefits of 4D-Var within a single data assimilation system (Barker, 2011). A brief description on the NCUM Hybrid 4D-Var system is given in Kumar et al. (2018).

The operational NCMRWF Ensemble Prediction System (NEPS) has 22 ensemble members. The horizontal resolution of NEPS is ~ 12 km. The initial condition perturbations of this ensemble prediction system are generated by Ensemble Transform Kalman Filter (ETKF) method (Bowler et al., 2009) and model uncertainties are taken care by the Stochastic Kinetic Energy Backscatter and Random Parameters schemes (Tennant et al., 2011). The forecast perturbations obtained from 6 hour short forecast run of 22 ensemble members are updated by ETKF four times a day (00, 06, 12 and 18 UTC). Perturbations of surface parameters such as sea-surface temperature, soil moisture content and soil temperature (Tennant and Beare, 2014) are included in the 12-km NEPS in order to address the problem of lack of ensemble spread near the surface. The NEPS is used to generate 10-day probabilistic forecasts using 23 members (22 perturbed + 1 control) ensemble system. Out of 22 perturbed ensemble members, one set of eleven members run from 00 UTC of current day and the other set of 11 members run from 12 UTC of previous day to provide ensemble forecast of 10 days. The operational deterministic forecast running at 12 km resolution from 00 UTC is used as the control forecast. A technical report by Mamgain et al., 2018 describes in detail the operational implementation of this high resolution NEPS.

## 2. Verification Statistics Data Base (VSDB)

(i) An intercomparison of verification scores for forecasts of upper air fields (HGT, Wind and T) is first carried out for GFS, NCUM, UKMO and NEPS (ensemble mean) using VSDB for different domains including Tropics (TRO; 20°S to 20°N),Northern Hemisphere (NHX; 20°N to 80°N), Southern Hemisphere (SHX; 20°S to 80°S) and Indianregion (RSMC; 20°S to 30°N and 45°-120°E). Forecast skill in GFS, UKMO and NEPS (ensemble mean) in terms of RMSE and ACC are assessed relative to skill in NCUM model forecasts. The results can be summarized as follows.

- For predicted 500 hPa HGT over NHX, GFS forecasts show no significant difference in skill (RMSE and ACC) compared to NCUM forecasts. However over SHX GFS forecasts show significantly higher skill (lower RMSE and higher ACC) compared to NCUM forecasts between Day-3 to Day-8. NEPS mean and UKMO forecasts of 500 hPa HGT show significantly higher skill compared to NCUM over NHX and SHX.

- Over TRO and RSMC region the higher skill in GFS compared to NCUM is pronounced between Day-2 to Day-6.Similar to verification over NHX and SHX the NEPS mean and UKMO forecasts of 500 hPa HGT show significantly higher skill compared to NCUM TRO and RSMC.
- For predicted 850 hPa Winds GFS model has significantly higher skill compared to NCUM at all lead times. For predicted 850 hPa T GFS model features substantially higher skill compared to NCUM at all lead times.

The root mean square error (RMSE) and anomaly correlation (ACC) are good measures of overallforecast performance. As Buizza et al. (2005) note they are both influenced by systematic errors andrandom error variance. Here the accuracy of the ensemble system and its performance against thedeterministic system are measured using RMSE and AC.

The brief assessment presented here is based on NCEP"s Verification Statistics Data Base (VSDB) implemented at NCMRWF. The verification is done at 1° grid resolution and severalscores are computed. The evaluation is performed for Global (G3), Tropics (TRO; 20°S to 20°N), Northern Hemisphere (NHX; 20°N to 80°N), Southern Hemisphere (SHX; 20°S to 80°S) and Indianregion (RSMC; 20°S to 30°N and 45°-120°E). The various scores computed in VSDB are Bias, RMSE, ACC, Pattern Correlation (PCOR), Error in Pattern Variation (EPV), Ratio (FCST/Analysis) of Standard Deviation (RSD) and Murphy's Mean Squared Error Skill Score (MSESS). These scores are computed for all the five domains for five fields of Wind, U, V, HGT and T. All the scores are assessed for several vertical levels (1000, 850, 700, 500, 200, 100, 50, 20 and 10 hPa levels). The scores computed for each of the lead times allow one to assess the decreasing magnitude of the scores with increasing lead time (die-off). For brevity some of the select scores, fields and domains are presented.

(a) RMSE and ACC are presented for NHX, SHX, TRO and RSMC for HGT at 500 hPa

(b) RMSE and ACC are presented for TRO and RSMC regions only for Wind and T at 850 hPa

(c) Bias in U, V, Wind and T at 850 hPa is presented for RSMC region only.

(d) Murphy's Mean Squared Error (MSE) Skill Score for U and V at 850 hPa is presented for RSMC region.

### *2.1 RMSE, ACC and Bias*

In Figure 1 the RMSE (left) and ACC(right) in 500 hPa HGT is shown for NHX (top) and SHX (bottom). The NEPS mean shows generally expected improvements over the deterministic forecast in terms of lower RMSE athigher forecast leads times. The lower part in each paneldemonstrates the difference in RMSE/ACC (solid lines)and statistical significance (bars) at all lead times. The statistical significance at 95% confidence level is obtained by first computing the standard deviation of the difference between each of the models and a reference model (NCUM in the present case) for each forecast hour. The 95% confidence level is obtained as F*SD/sqrt(N-1), (where F=1.96).

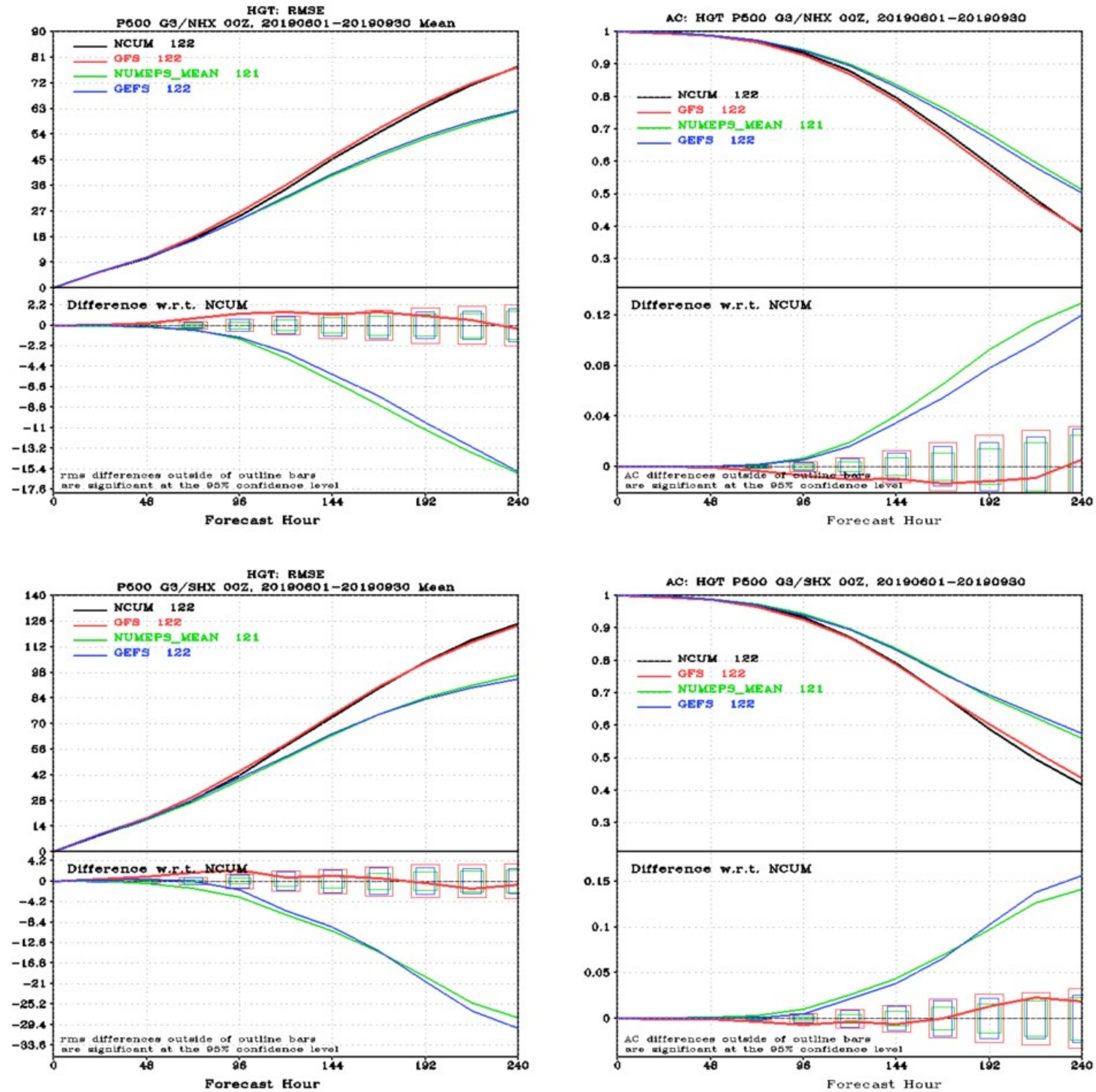


**Figure 1. RMSE (*left*) and ACC (*right*) in 500 hPa HGT for NHX (*top*) and SHX (*bottom*)**

Over Northern Hemisphere (top) and southern Hemisphere (bottom) RMSE and ACC (Figure 1) suggest that , in the short range there is almost no difference in the skill of two deterministic models (GFS and NCUM). Similarly, the ensemble mean from the two ensemble models (GEFS and NUMEPS) suggest hardly any difference. However, the two ensemble mean show significantly different and improved skill compared to deterministic models at longer lead times. Ensemble mean have lower RMSE (and higher ACC) compared to deterministic forecasts after 96hr. The skill (RMSE and ACC) in GFS are not statistically significant compared to NCUM at all lead times. The skill (RMSE and ACC) in ensemble mean (NCUMEPS and GEFS) significantly (95%) higher compared to skill of NCUM, particularly after 96h.

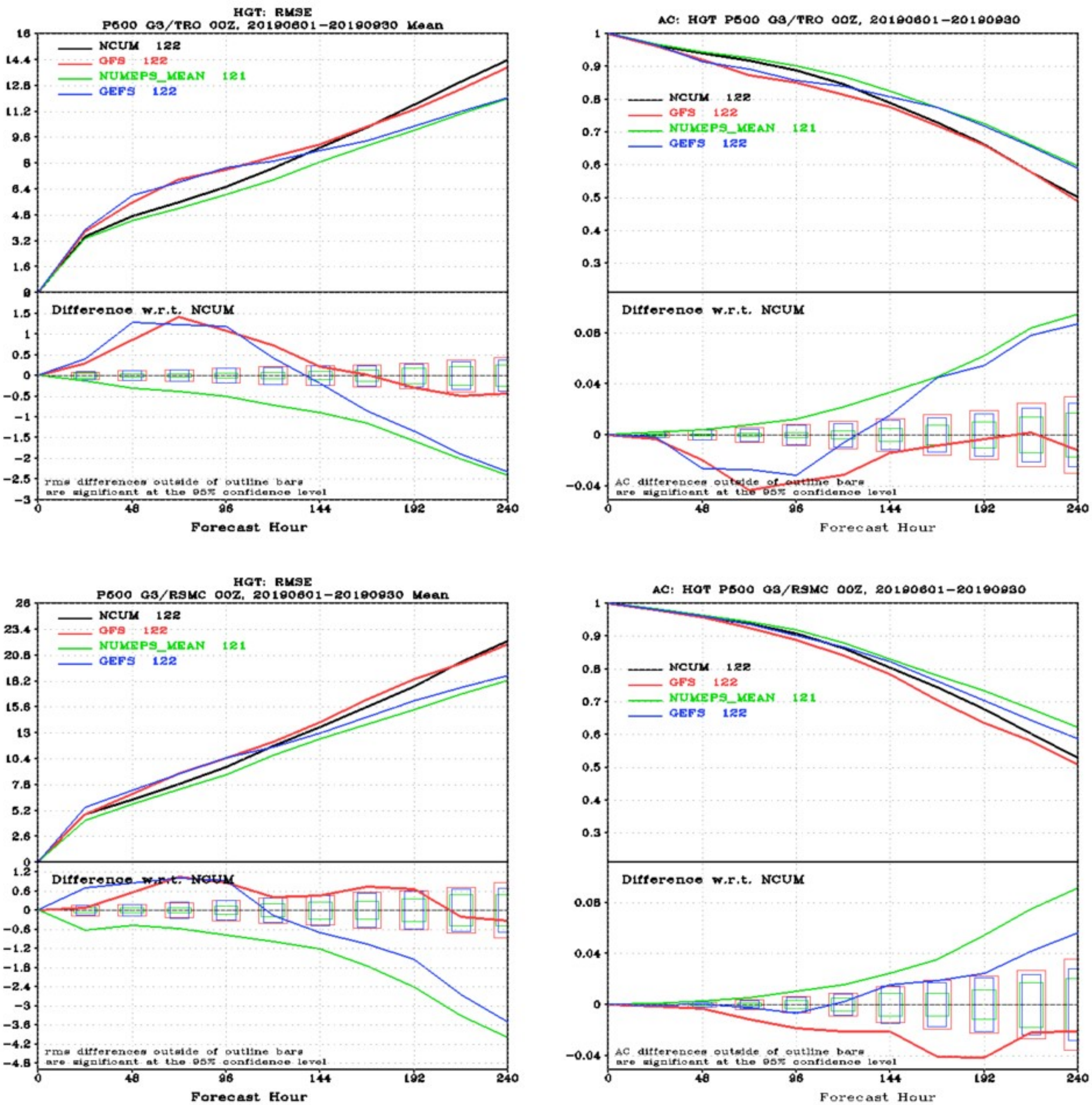


**Figure 2. RMSE (*left*) and ACC (*right*) in 500 hPa HGT for Tropics (*top*) and RSMC region (*bottom*)**

Similarly, Figure 2 shows the RMSE and ACC in 500 hPa HGT over tropics (*top*) and RSMC region (*bottom*). It is striking to note that unlike in Figure 1, the two deterministic models (GFS and NCUM) suggest significant difference in skill. Over the tropics and RSMC region NEPS shows higher skill compared to other models at all lead times after 96h. In short range forecasts GEFS forecast show higher (*lower*) RMSE (*ACC*) compared to NCUM and NEPS. GFS shows significantly low skill up to Day8.

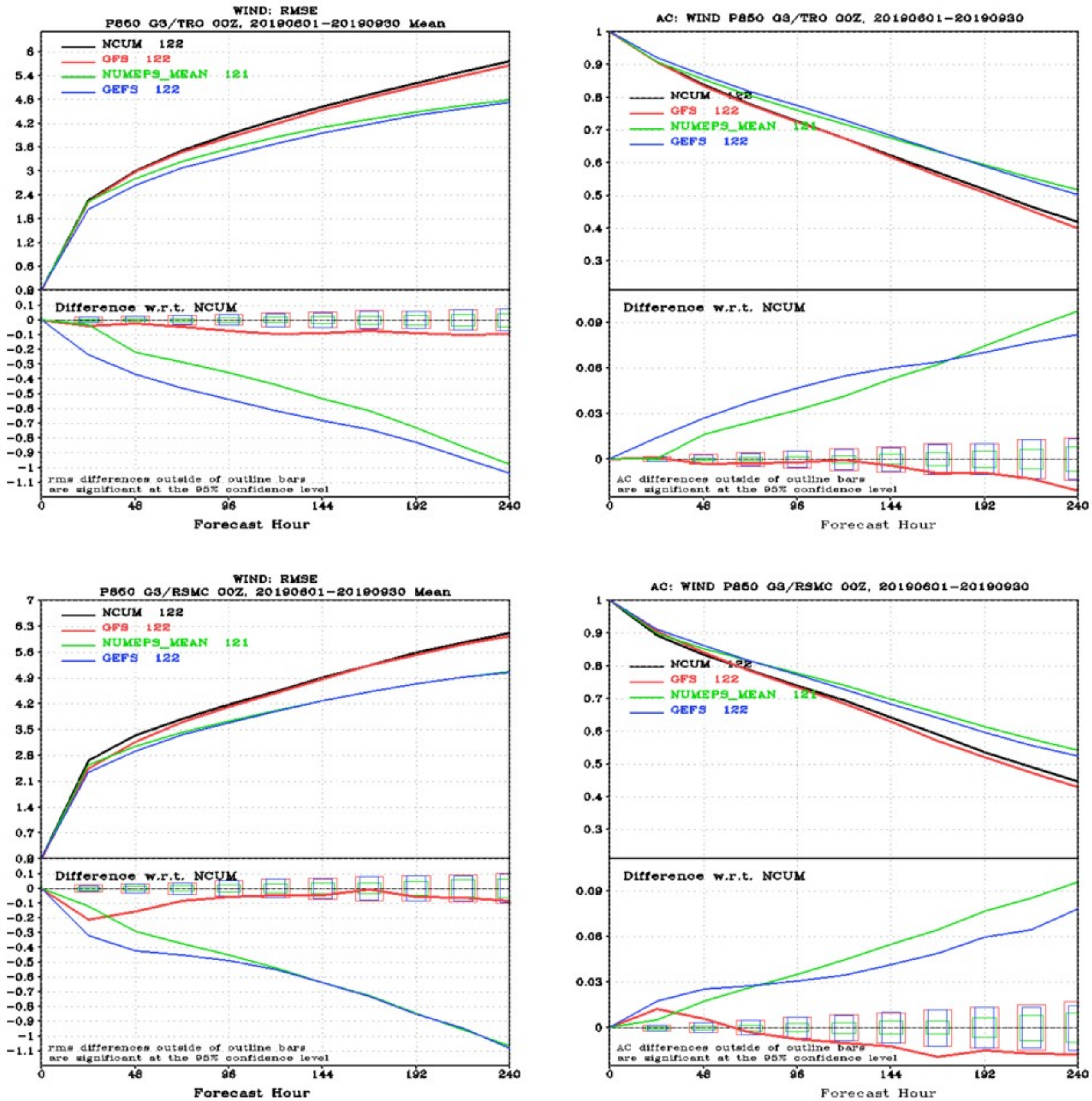


**Figure 3. RMSE (*left*) and ACC (*right*) in 850 hPa Winds for Tropics (*top*) and RSMC region (*bottom*)**

RMSE and ACC in 850 hPa in Winds is shown for tropics and RSMC region in Figure 3. Both the ensembles show lower (*higher*) RMSE (*ACC*). Forecast skill (RMSE and ACC) in GFS is mostly comparable to that of NCUM except during first 72 h, when the difference is significant. Skill in ensemble mean are significantly (95%) higher compared to skill in NCUM.

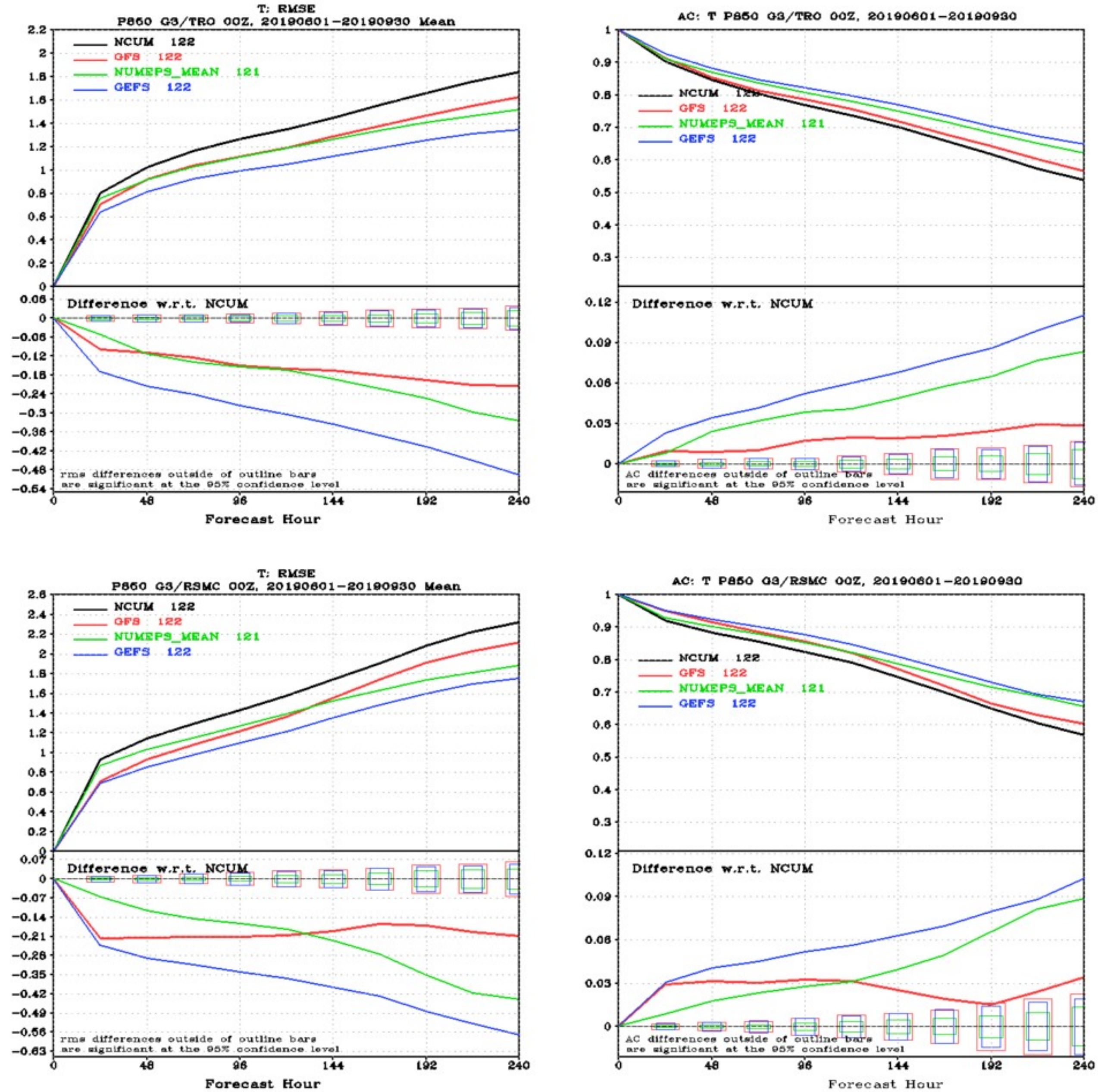


**Figure 4. RMSE (*left*) and ACC (*right*) in 850 hPa Temperature for Tropics (*top*) and RSMC region (*bottom*)**

RMSE and ACC in 850 hPa in T is shown for tropics and RSMC region in Figure 4. GEFS skills are better than other models in all the forecasts. Here NCUM skill is low in both the regions. In RSMC region GFS performs better then NEPS upto day 5.

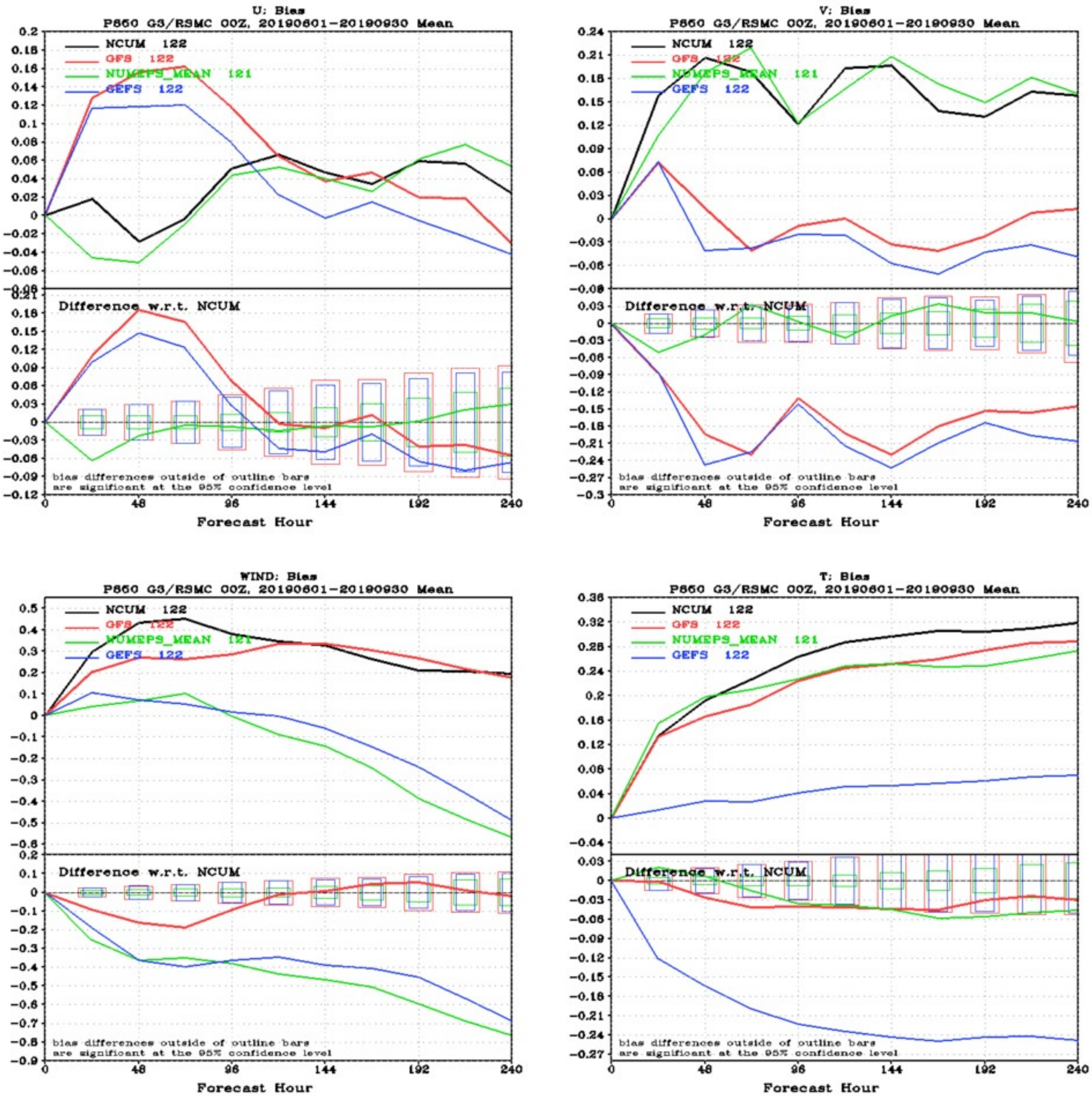


**Figure 5. Bias in 850 hPa U (*top left*) V (*top right*) Wind (*bottom left*) and T (*bottom right*) for the RSMC region.**

Figure 5 shows Bias in 850 hPa U (*top left*) V (*top right*) Wind (*bottom left*) and T (*bottom right*) for the RSMC region. GFS and GEFS forecasts distinctly show positive bias in U (up to 96 hours) and NCUM and NEPS show positive bias in V, in WIND both the GFS (bias is low compared to NCUM) and NCUM show positive bias up to day 5 and both the ensembles show negative bias beyond day5. In the 850 hPa T, except GEFS all models have positive bias, which are slowly increasing with lead time.

*2.2 Murphy's MSE Skill Score*:

The MSE skill score (MSESS) depend on (i)the anomaly correlation coefficient (ACC), (ii) the conditional prediction bias, and (iii) the unconditional prediction bias (*Murphy*, 1988). Following Murphy (1988 and 1996),

**MSESS = 2(ACC) – 1**

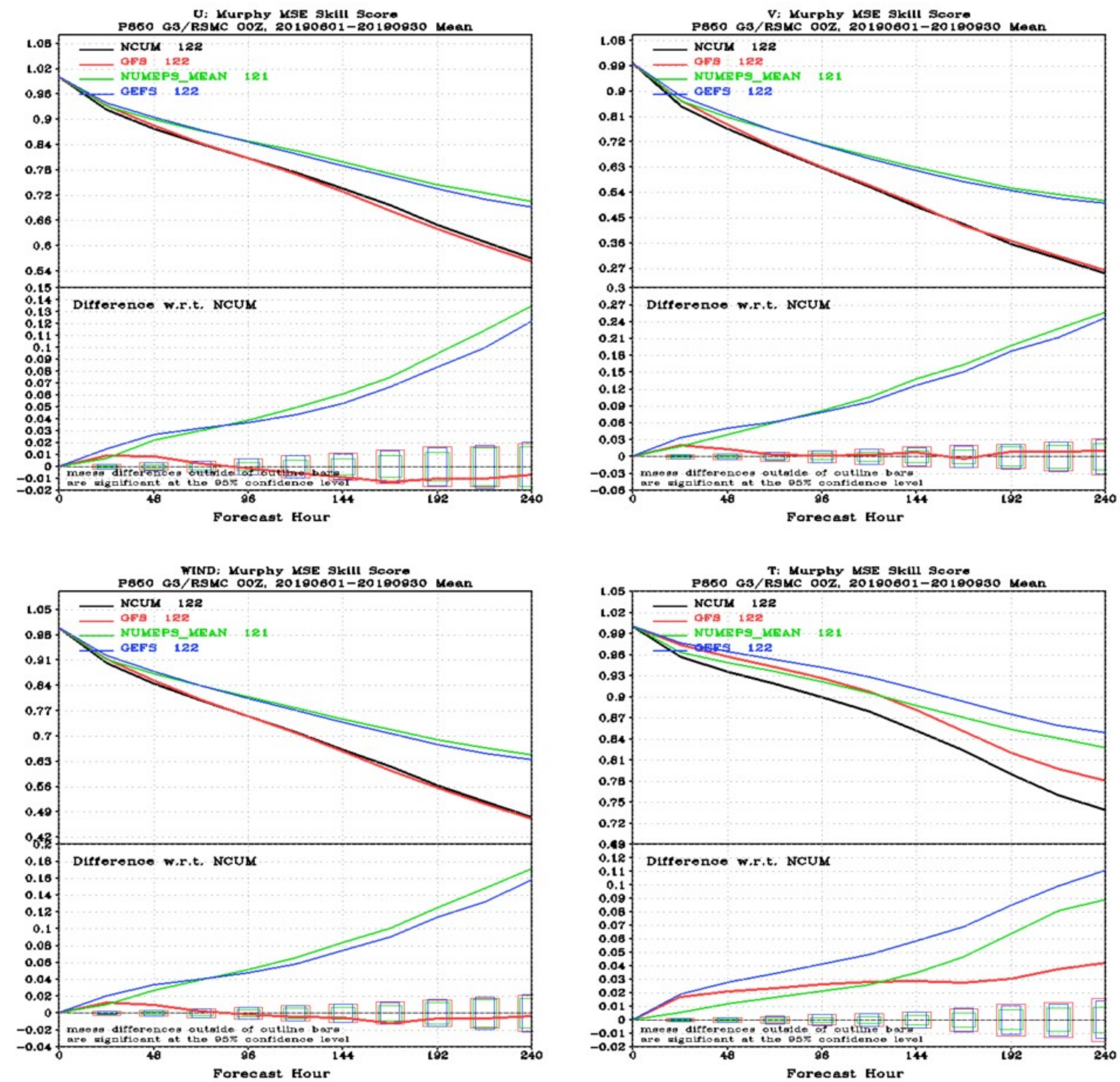


**Figure 6. MSESS in 850 hPa U (top left) V (top right) Wind (bottom left) and T (bottom right) for the RSMC region.**

Figure 6 shows MSESS in 850 hPa U (*top left*) V (*top right*) Wind (*bottom left*) and T (*bottom right*) for the RSMC region. Both the ensemble forecasts show higher score for U & V (*top panels*) and WIND (*bottom left)*. In the case of temperature GEFS shows higher score than all other models, GFS scores are higher than NEPS up to day 5 and NCUM scores are lowest in all the forecasts.

Figure 7 shows bias in wind forecasts (*left panels*) with respect to analysis at different pressure levels over RSMC region. It can be seen that both the ensembles show similar bias pattern with positive bias in upper levels and close to surface while having negative bias in middle levels. In middle levels except NCUM all other models exhibit negative bias. Both the ensemble system show lower bias in upper levels. Panels on right shows similar plot for temperature. In case of temperature GFS shows positive bias in lower levels and

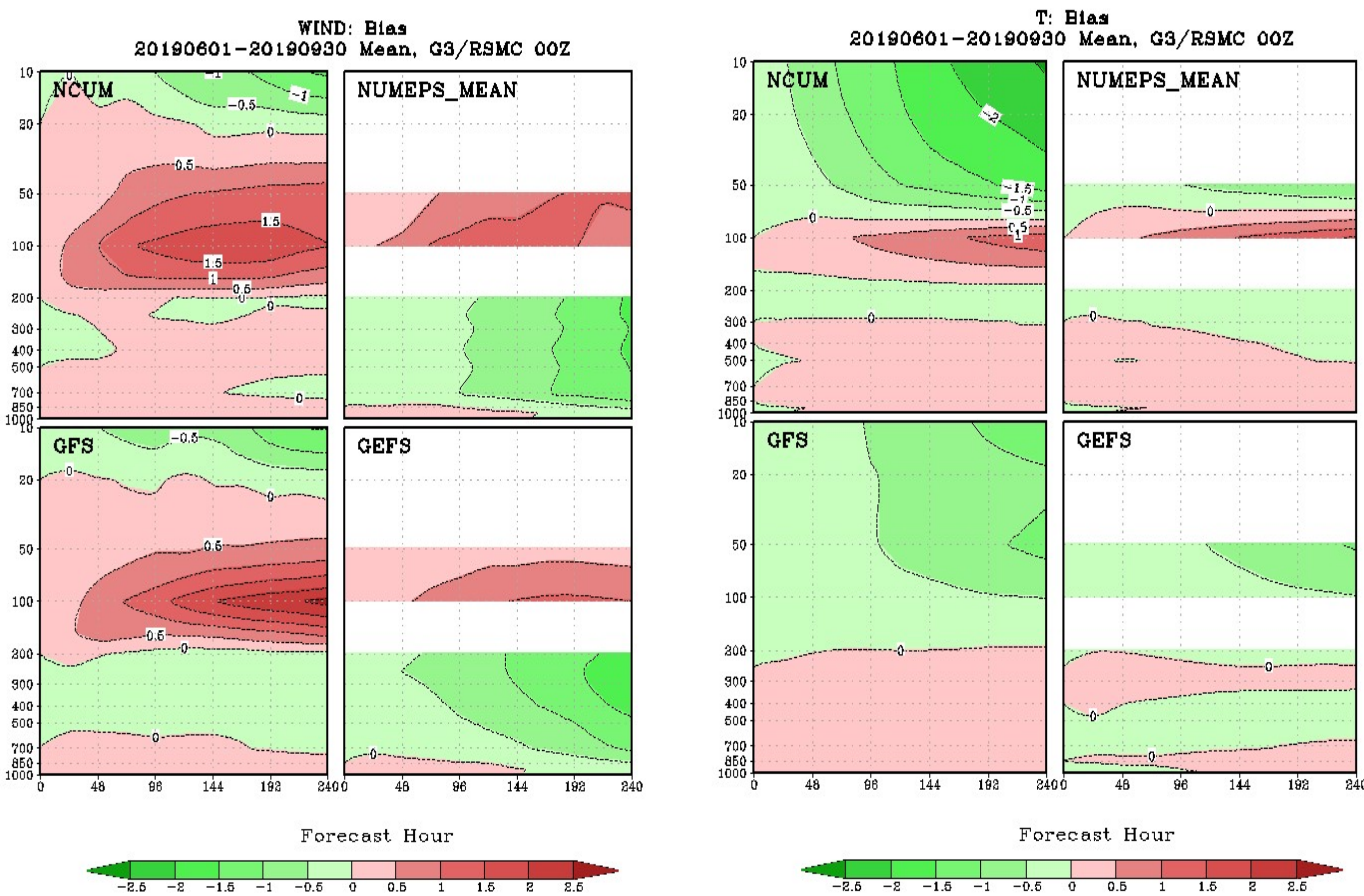


**Figure 7 Bias in wind (*left*) and temperature (*right*) forecasts w.r.t. analysis over RSMC region at different pressure levels.**

negative bias in upper levels. In case of NCUM trend is similar but a band of negative bias around 200 hPa and positive bias around 100 hPa also can be seen. Both the ensemble systems show less bias in upper levels and GEFS show positive bias in middle levels distinct from other models. Similar plot for both U and V can be seen in Figure 8. Here also both the ensemble system shows lower bias in upper levels and follows similar pattern.

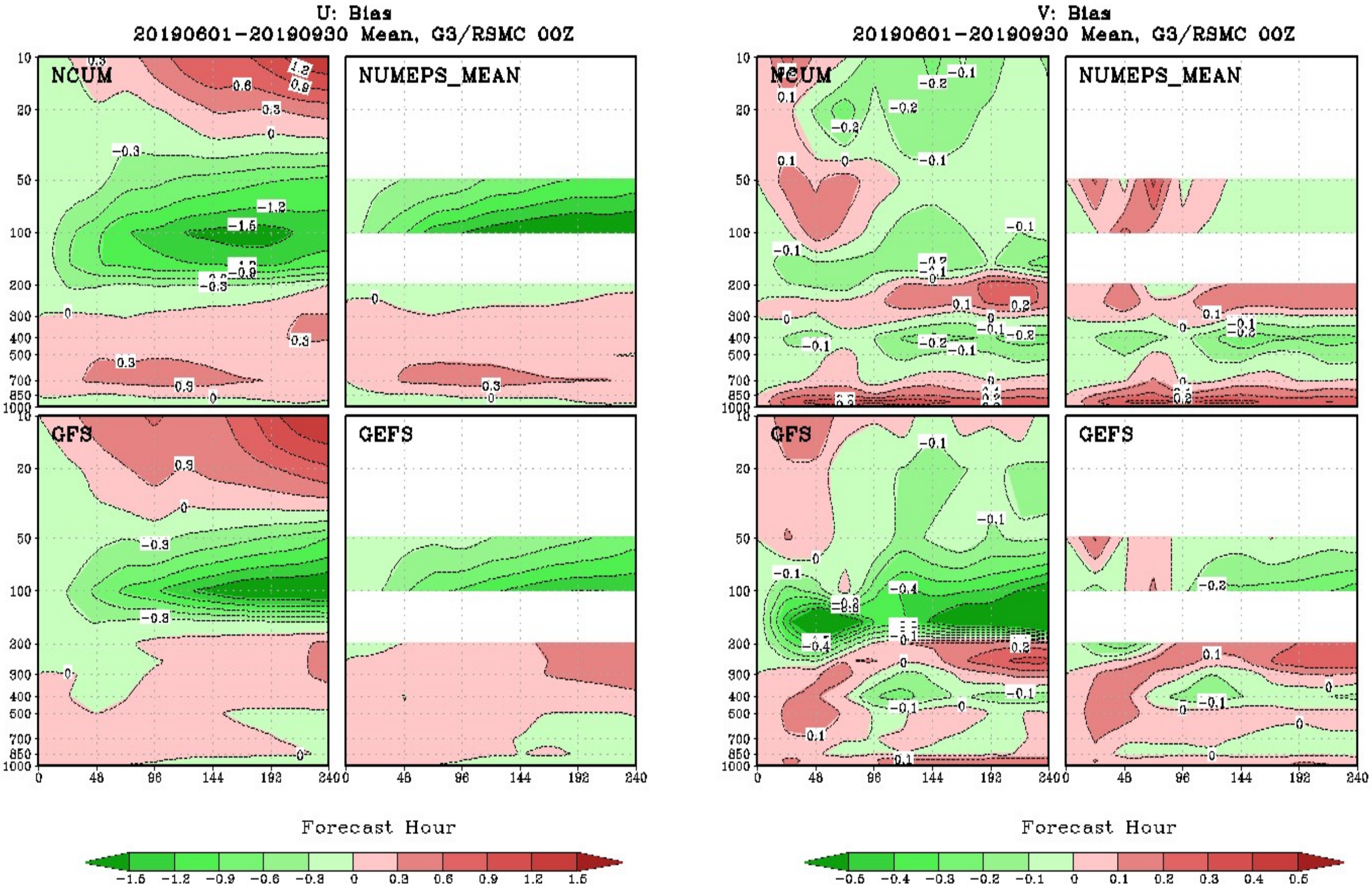


**Figure 8 Bias in U (*left*) and V (*right*) forecasts w.r.t. analysis over RSMC region at different pressure levels.**

## 3. Conclusions

An intercomparison of verification scores for forecasts of upper air fields (HGT, Wind and T) is first carried out for GFS, NCUM, UKMO and NEPS (ensemble mean) using VSDB for different domains including Tropics (TRO; 20°S to 20°N), Northern Hemisphere (NHX; 20°N to 80°N), Southern Hemisphere (SHX; 20°S to 80°S) and Indian region (RSMC; 20°S to 30°N and 45°-120°E). Forecast skill in GFS, UKMO and NEPS (ensemble mean) in terms of RMSE and ACC are assessed relative to skill in NCUM model forecasts. The results can be summarized as follows.

- As can be expected, the ensemble means from GEFS and NEPS show significantly higher skill compared to NCUM over all regions. For predicted 500 hPa HGT over northern & southern hemisphere, GFS forecasts show no significant difference in skill (RMSE and ACC) compared to NCUM forecasts.

- Over the tropics, the forecast skill (RMSE and ACC) in predicted 850 hPa Wind in GFS is mostly comparable to that of NCUM except during first 72 h, when the difference is significant. Skill in ensemble mean are significantly (95%) higher compared to skill in NCUM.

- For 850 hPa Temperature predictions over tropics and RSMC region, the skill in GFS forecasts is significantly higher compared to skill in NCUM. GEFS forecasts feature best skill for 850 hPa Temperatures over tropics and RSMC region.

- Murphy's Mean Squared Error Skill Score (MSESS) computed for 850 hPa U, V, Wind and T over RSMC region shows that both the ensemble forecasts show higher score for U & V and WIND. In the case of temperature GEFS shows higher score than all other models, GFS scores are higher than NEPS up to day 5 and NCUM scores are lowest in all the forecasts.

- Wind Bias is also assessed for different vertical levels. Both the ensembles show similar bias pattern with positive low bias in upper levels and close to surface while having negative bias in middle levels. In middle levels except NCUM all other models exhibit negative bias.

## References


Barker, D. M. and A. M. Clayton (2011): Hybrid Variational-Ensemble Data Assimilation, ECMWF Seminar on Data assimilation for atmosphere and ocean, 6 - 9 September 2011, https://www.ecmwf.int/sites/default/files/elibrary/2012/7934-hybrid-variationalensemble-data-assimilation.pdf

Bowler N.E., Arribas A., Beare S. E., Mylne K.R. and Shutts G. J. 2009: The local ETKF and SKEB: Upgrades to the MOGREPS short-range ensemble prediction system. Quart. J. Roy. Meteor. Soc., 135, 767–776.

Kumar S., A. Jayakumar, M. T. Bushair, Buddhi Prakash J., Gibies George, Abhishek Lodh, S. Indira Rani, Saji Mohandas, John P. George and E. N. Rajagopal 2018: Implementation of New High Resolution NCUM Analysis-Forecast System in Mihir HPCS. NCMRWF Research Report, NMRF/TR/01/2018.

Mamgain A., Abhijit Sarkar, Anumeha Dube, Arulalan T., Paromita Chakraborty, John. P. George and E.N. Rajagopal 2018: Implementation of Very High Resolution (12 km) Global Ensemble Prediction System at NCMRWF and its Initial Validation. NCMRWF Research Report, NMRF/TR/02/2018.

Murphy A. H., 1988: Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient. Mon. Wea. Rev.,116: 2417-2424.

Murphy A. H., 1996: General decompositions of MSE-based skill scores: Measures of some basic aspects of forecast quality. Mon. Wea. Rev., 124, 2353-2369

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

Tennant W., Beare S. 2014: New schemes to perturb sea-surface temperature and soil moisture content in MOGREPS. Quart. J. Roy. Meteor. Soc., 140: 1150–1160

Walters et al. 2017: The Met Office Unified Model Global Atmosphere 6.0/6.1 and JULES Global Land 6.0/6.1 configurations, Geosci. Model Dev., 10, 1487-1520, doi:10.5194/gmd-10-1487-2017