Estimation of Global Ensemble Forecast System Version 12 (GEFSv12) Reforecast

Hong Guan[1], Yuejian Zhu[2] , Bing Fu[1], Eric Sinsky[3], Bo Yang[1], Xianwu Xue[4], Dinchen Hou[2], Jiayi Peng[5]
[1]Lynker at EMC/NCEP/NWS/NOAA, [2]EMC/NCEP/NWS/NOAA, [3]ERT at CPC/NCEP/NWS/NOAA, [4]SAIC at EMC/NCEP/NWS/NOAA, [5]AXIOM at EMC/NCEP/NWS/NOAA

Introduction

In accompaniment with the implementation of Global Ensemble Forecast System Version 12 (GEFSv12), a 31-yr (1989–2019) ensemble reforecast dataset has been generated at the National Centers for Environmental Prediction (NCEP). This study summarizes the configuration and dataset of the GEFSv12 reforecast and presents some evaluations of 500-hPa geopotential height, tropical storm track, precipitation, 2-m temperature, and MJO forecasts. The results were also compared with GEFSv10 or GEFS Subseasonal Experiment reforecasts. In addition to supporting calibration and validation for the National Water Center, NCEP Climate Prediction Center, and other National Weather Service stakeholders, this high-resolution subseasonal dataset also serves as a useful tool for the broader research community in different applications.

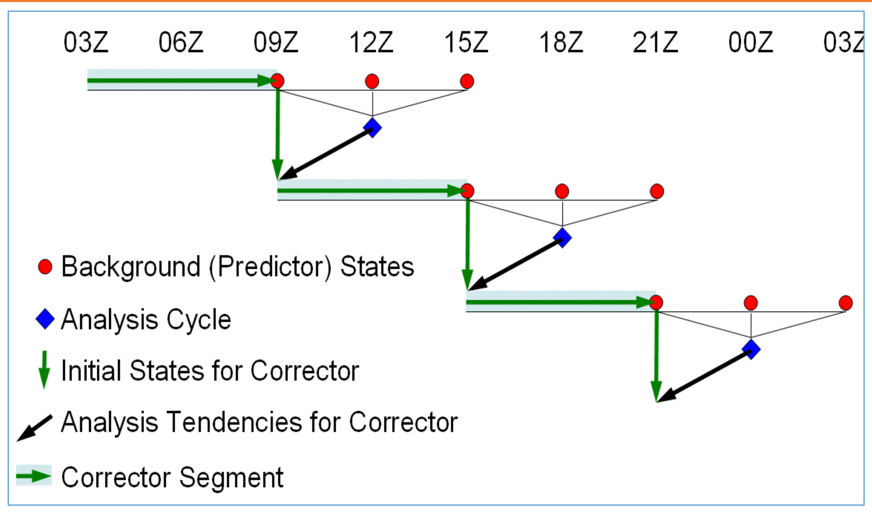
31-year Reforecast Configuration

- **Forecast system**
  - FV3 GFSv15.1 and GEFSv12
  - Resolution – C384 (~25km) with 64 hybrid vertical levels
- **Frequency and ensemble size**
  - Initialized at 00UTC for every day
  - Run 5 members out to 16 days, except for 11 members out to 35 days every Wednesday
- **Initial conditions**
  - Phase I: 1989 – 1999 (11 years), CFS analysis, BV-ETR perturbation
  - Phase II: 2000 – 2019 (20 years), Hybrid FV3 GFS/EnKF reanalysis (ESRL/PSD) with Incremental Analysis Update (IAU) process, EnKF f06 perturbation
  - **Caution - Initial analyses and perturbations of 31 years are in-consistent**
- **Output data (default mode)**
  - 3 hourly out to 10 days at 0.25 degree resolution
  - 6 hourly beyond 10 days at 0.5 degree resolution
  - Save all variables (**590**) in grib2 format at above resolution on HPSS for 5-year
  - Save selected variables (**77**) in grib2 format on disk for stakeholders (CPC, MDL, and NWC)
  - Save selected variables on NCEP ftp and AWS for public access

20-year Reanalysis Configuration (ESRL/PSD)

- Real-time **operational** DA configuration:
  - Hybrid 4D-En-Var, with control at C768 (approximately 0.125 degree grid spacing) and 80-ensemble members at C384 (approximately 0.25 degree grid spacing)
  - SPPT, SKEB, SHUM stochastic parameterizations in the ensemble to address model uncertainty
- **Reanalysis** configuration:
  - The same model version and stochastic parameterization as operational DA
  - Reduced resolution: C384 (control – 0.25degree) and C128 (ensemble members – 0.75degree)
  - **IAU** (next slide) to control noise and improve accuracy
- Run 5 streams – each stream has 1-year spin-up
  - 1999 stream (1999 – 2003)
  - 2003 stream (2003 – 2007)
  - 2007 stream (2007 – 2011)
  - 2011 stream (2011 – 2015)
  - 2015 stream (2015 – 2019)

Incremental Analysis Update (IAU) Process



The analysis increments are added to the model state over a period of time (IAU window) for each model time-step.

GEFSv12 Reforecast for Public Access

One-hundred seventy-six upper air variables and Forty-three surface and other single-level variables are available for Public Access.

- **Data access through AWS: (2000 - 2019)** https://noaa-gefs-retrospective.s3.amazonaws.com/index.html
- **Data access through NCEP ftp: (1989 - 2019)** https://ftp.emc.ncep.noaa.gov/GEFSv12/

Ensemble Mean Skills of 500hPa Geopotential Height

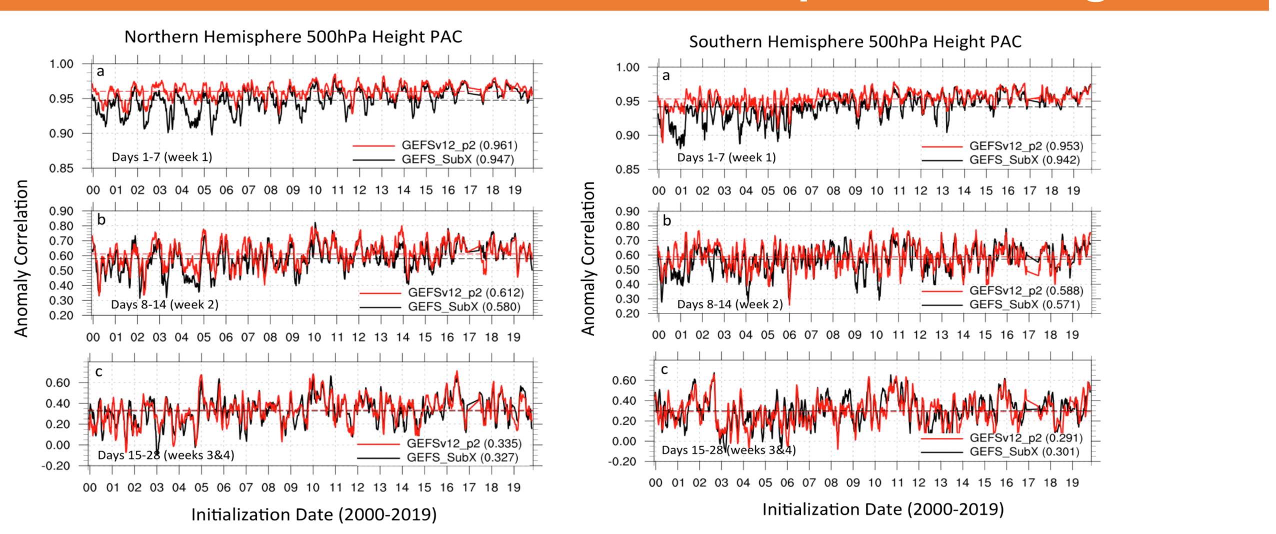


Figure.1: Ensemble-mean anomaly correlation for Northern Hemisphere (NH; 20°N–80°N, left) 500-hPa geopotential height for week 1 (a), 2 (b), and weeks 3&4 (c) forecasts. The black and red colors denote the GEFS_SubX and GEFSv12_p2. The average scores for the two sets of reforecasts are shown in the figure. Note there is a data gap from Dec 2016 to May 2017, corresponding to the period between the GEFS_SubX reforecast and corresponding real-time forecast. Over the Northern Hemisphere (NH), the GEFSv12_p2 outperforms the GEFS_SubX with improvements in average anomaly correlation (AC) of 1.5%, 5.5%, and 2.5% for week 1, week 2, and weeks 3 and 4 forecasts, respectively. Over the Southern Hemisphere (SH, right), the average AC scores are slightly lower than over the NH, which is consistent with the previous finding in Zhu et al. (2018) for the evaluation of the 16-year GEFS_SubX reforecast. Relative to the GEFS_SubX, the GEFSv12_p2 shows 1.3% and 3.0% improvements for week 1 and week 2 forecasts and a 3.3% degradation for the weeks 3 and 4 forecasts. The significant tests indicate that the week 1 and week 2 GEFSv12_p2 AC are significantly higher than GEFS_SubX for both NH and SH, while the corresponding AC values are not significantly different between the GEFSv12_p2 and GEFS_SubX for weeks 3 and 4.

CONUS PQPF Skills (2000-2019)

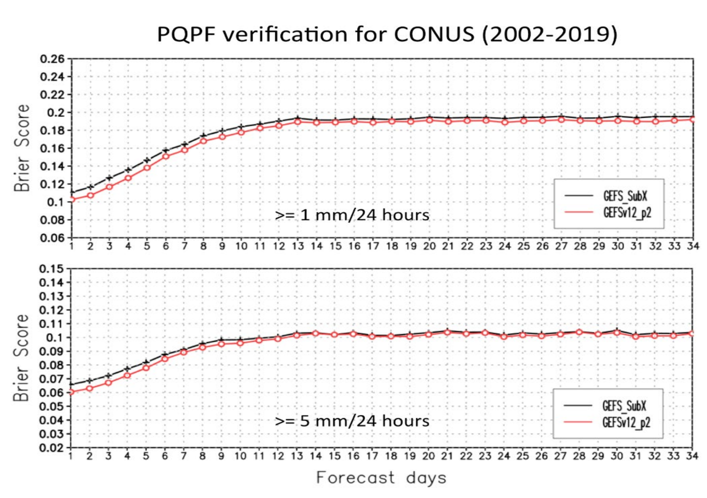


Figure 2: The daily average Brier Score of the CONUS probabilistic quantitative precipitation forecast (PQPF) from 2002 to 2019 for 24 hours accumulated precipitation greater (and equal) than 1.00mm (top) and 5.00mm (bottom). The comparison is for GEFS_SubX reforecast (black) and GEFSv12 reforecast (red) those run once per week (Wednesday) with 11 members, out to 35 days. The verification truth is CCPA.

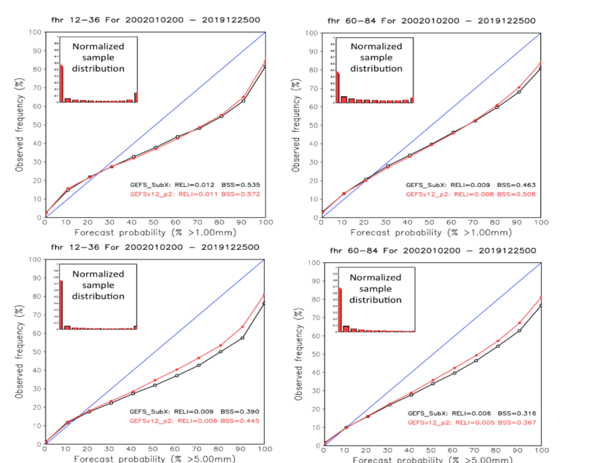
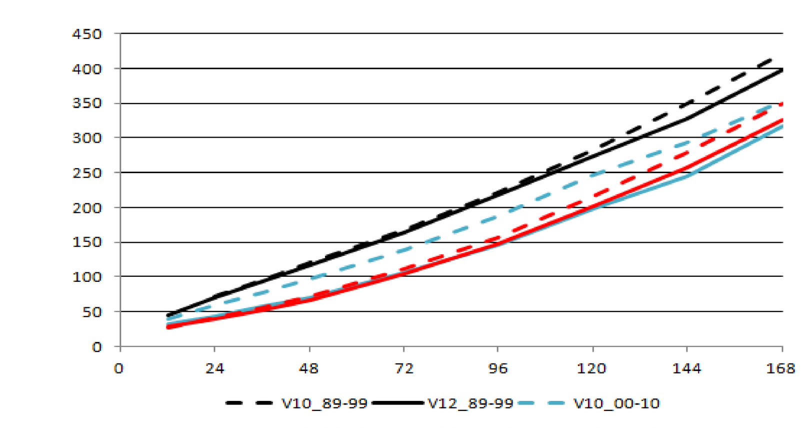


Figure 3: The reliability diagram of the CONUS probabilistic quantitative precipitation forecast (PQPF) from 2002 to 2019 for 24 hours accumulated precipitation greater (and equal) than 1.00mm (12-36 hours, top left; 60-84 hours, top right) and 5.00mm (12-36 hours, bottom left; 60-84 hours, bottom right). The comparison is for GEFS SubX version reforecast (black) and GEFSv12 reforecast (red) those run once per week (Wednesday) with 11 members, out to 35 days. The verification truth is CCPA. The average reliability score and Brier skill score are also presented in each subplot.

The GEFSv12 consistently displays the better (i.e., lower) Brier scores compared to the GEFS_SubX, with a more obvious improvement at lead times shorter than about 10 days. Forecast skill decreases with lead time and reaches saturated values at approximately day 13 for all situations. The precipitation probability forecast biases for 12–36 and 60–84 h for amounts greater than 1 and 5 mm were measured by reliability diagrams (Fig. 5). The GEFSv12 and GEFS_SubX show very similar performance for the precipitation greater than 1.00 mm. For the heavier precipitation category (>5 mm), the GEFSv12 slightly outperforms the GEFS_SubX with its curves being closer to the diagonal lines.

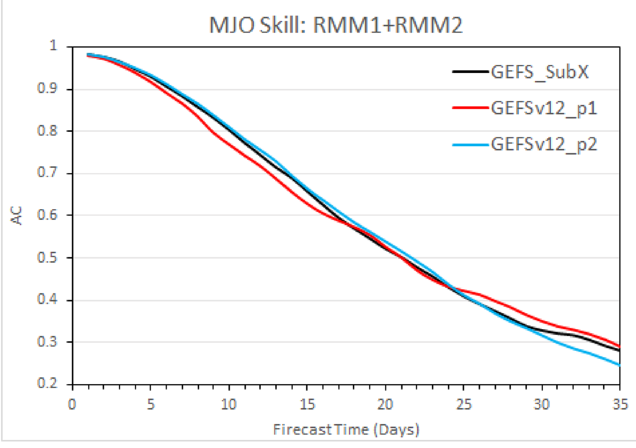
Comparison of TS Track Errors (GEFSv10 .vs. GEFSv12)



- Why do we compare to GEFSv10?
  - *Because we have similar reforecast configuration to cover the same period.*
- Why do we consider three different periods?
  - 1989-1999: Both v10 and v12 are using the same initial conditions, but different models
  - 2000-2010: GEFSv10 used CFS reanalysis initial analysis.
  - 2010-2019: GEFSv10 moved to use real-time GDAS (different from CFS reanalysis) around Spring 2011).
- More discussion - please refer to the manuscript led by Guan 2022.

Figure 4: The TS track errors averaged over the Atlantic, East Pacific and West Pacific basins binned by ~decade during the 31-year reforecast for GEFSv10 (dash lines) and GEFSv12 (solid lines). Black, blue, and red lines denote the 1989–1999, 2000–2010, and 2011–2019 periods, respectively. The GEFSv12 skill in forecasting TC tracks has improved from the GEFSv10. For all three decades, the GEFSv12 reduces the track errors with the maximum reduction during the 2000–10 period, when the reductions reach approximately 25% and 10% for 1- and 7-day forecasts, respectively.

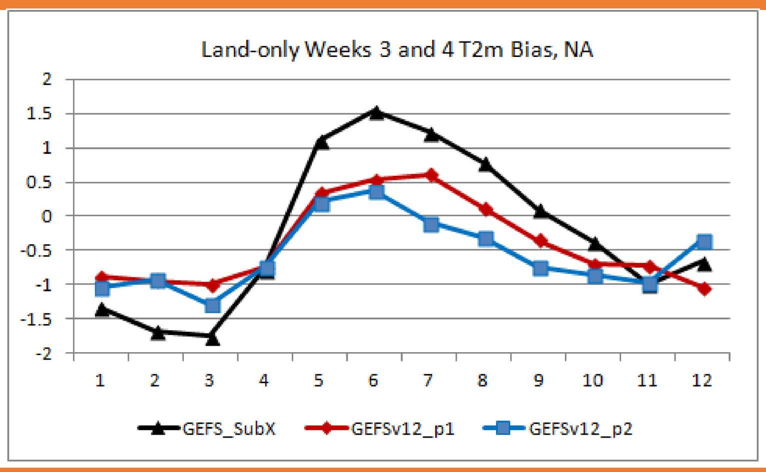
Comparison of MJO Skill Scores (RMM1+RMM2)



- Phase 1: 1989-1999
- Phase 2: 2000-2019
- GEFS SubX - it is updated version of GEFSv11 which was running in near real-time since October 2017 to support NOAA SubX project.
- Climatology:
  - CFS reanalysis climatology
  - GEFSv12 reanalysis climatology
  - GEFS_SubX reforecast climatology
  - GEFSv12 reforecast climatology

Figure 5: The real-time multivariate MJO (RMM) skill as a function of lead time for GEFS_SubX (black; 2000–2016), GEFSv12_p1 (red; 1989–1999), and GEFSv12_p2 (blue; 2000–2019) reforecasts. Overall, The MJO forecast skill for the GEFSv12_p2 (~21.5 days) is similar to the GEFS_SubX and GEFSv12_p1 (21 days) when using AC=0.5 as the threshold of the useful skill.

Monthly Difference of NA Land 2-m Temperature Bias



- Both (phase 1 and phase 2) has demonstrated less (or improved) bias than GEFS SubX forecast
- There is a sysimatic difference bewteen the phase 1 and phase 2 for July, August, and Sept.
- Day 15-28 average (weeks 3&4)

Figure 6: Weeks 3 and 4 biases in valid 2-m temperature forecasts averaged during the GEFS_SubX (black, 1999–2016), GEFSv12_p1 (red, 1989–1999), and GEFSv12_p2 (blue, 2000–2019) reforecast periods over the NA land-only.

Global 2-m Temperature Bais against ERA5 (2000-2019)

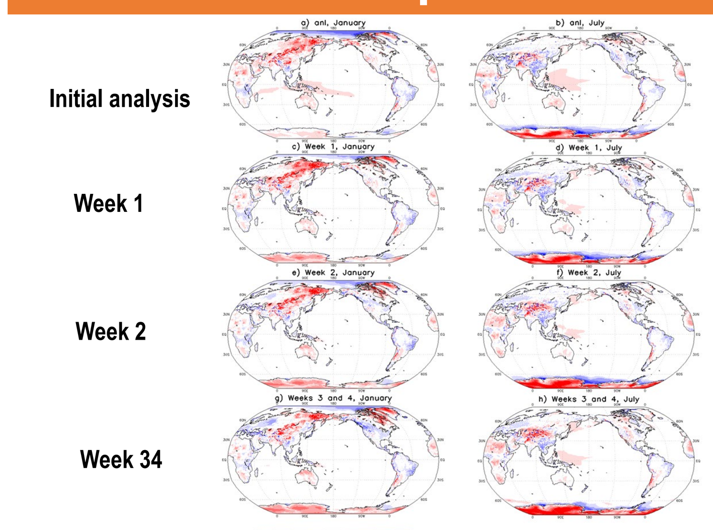


Figure 7: The difference of 2-m temperature between the GEVSv12 reanalysis and ERA5 for January (a) and July (b) over phase 2. Spatial distribution of 2-m temperature mean error (i.e., bias) over phase 2 for January during week (c) 1, (e) 2, and (h) weeks 3 and 4 forecasts and July during week (d) 1, (f) 2, and (g) weeks 3 and 4 forecasts. A large warm bias over northern Asia is persistently seen in January (Figs. 7a,c,e,f) with a decreasing trend over increasing forecast lead time. In general, the error in 2-m temperature at the weeks 3 and 4 time scale is nearly saturated and the impact from initial conditions decreases. At this time scale, the GEFSv12 generates a cold bias over North America (NA) in January (Figs. 7g). A larger cold bias for the boreal winter season over the NA domain has been persistently observed in several generations of the NCEP GEFS (Guan et al. 2015, 2019) and was thought to be related to the imperfect parameterization of winter-associated physical processes (Guan et al. 2019).

2-m Temperature Bias Impacted by Snow Cover

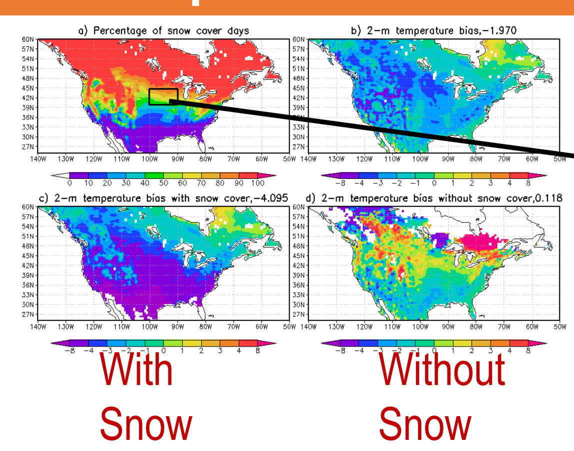


Figure 8: (a) Percentage of snow cover days, (b) 2-m temperature forecast bias under all conditions, (c) bias with snow cover forecast, and (d) bias without snow cover forecast for 408-h control-member forecast over NA. The results are based on the GEFSv12_p2 reforecast for January, February, and March.

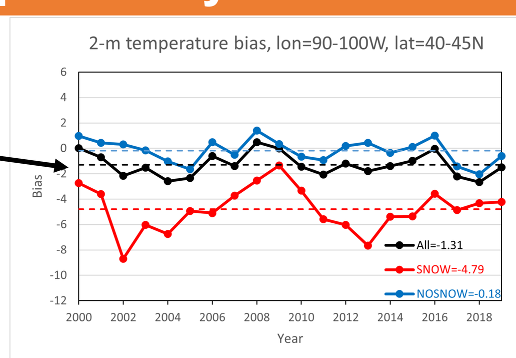


Figure 9: Time series of 2-m temperature forecast errors for 408 h control-member forecast over a small region (40°N-45°N, 90°W-100°W) near the central US (marked with the black rectangle in Fig. 11a). Black, red, and blue solid curves indicate the errors for January, February, and March under all, with, and without snow cover forecast, respectively. The corresponding dash lines denote the averages over the entire period.

Clearly, the 2-m temperature bias characteristics are quite different between the two conditions (Figs. 8c,d). Figure 9 shows the time evolution of biases over a small region near the central United States. A larger cold bias is dominant under the existence of snow cover with a domain-averaged value of −4.79° C during the GEFSv12_p2 period. In contrast, bias is much smaller under snow-free conditions where the average value is about −0.18° C. This indicates there is considerable room for improving the 2-m temperature forecast under snow-covered conditions. An improvement in modeling snow-associated physical processes would undoubtedly lead to a better 2-m temperature forecast. The large difference in bias characteristics between cases with and without snow cover also suggests that statistical calibration of 2-m temperature should be performed based on the existence of snow.

Frequency of CONUS Precipitation Exceeding 50mm/day

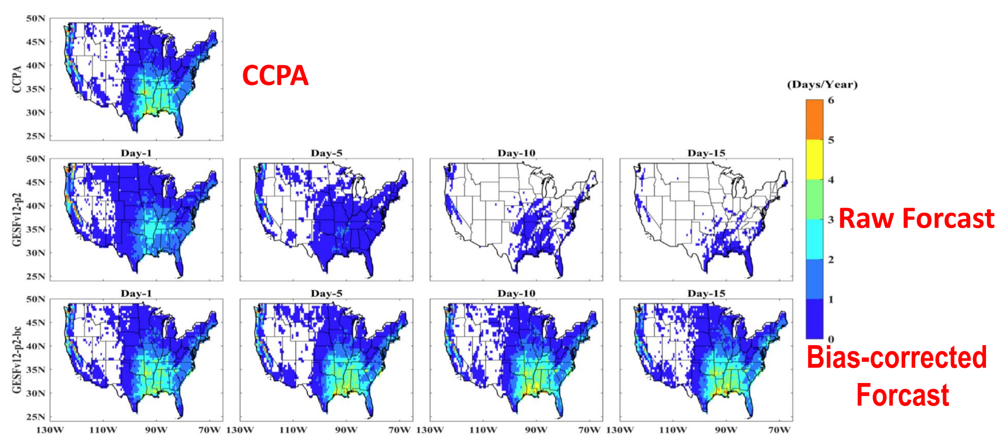
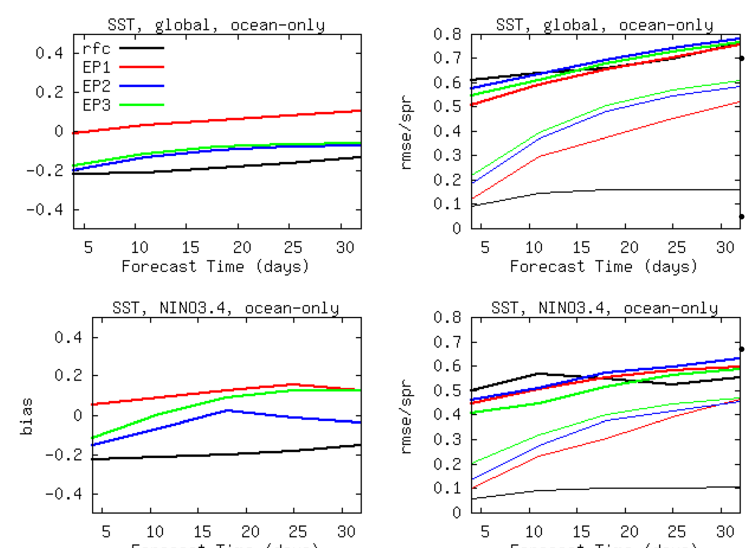


Figure 10: The days/year with a 24-h precipitation exceeding 50mm over the CONUS for CCPA (upper panel), raw (GEFSv12_p2, middle panels), and bias-corrected (GEFSv12_p2-bc, lower panels) 5-m ensemble mean forecasts for day-1, day-5, day-10 and day-15. The calibration using long-term reforecast data is particularly important in improving the model climatology for the heavy precipitation events (>50 mm) as illustrated in this figure. In the raw forecast, the model 24-h precipitation events exceeding 50 mm are substantially lower than the CCPA, especially for the longer lead times, when heavy (or extreme) precipitation events are completely missed for most of the domain. After the bias correction, both distributions and magnitudes in heavy precipitation events are much more consistent with the CCPA throughout all lead times.

Reforecast as a menchmark data to develop GEFSv13 (SST comparisons between coupling and uncopling models)



- Bias for global ocean
  - Rfc (reforecst) - negative bias
  - EP1(ensemble prototype 1) - positive bias
  - EP2 and EP3 - very similar, negative bias, but smaller than reforecast. NSST generates a colder SST in rfc, EP2, and EP3
- Bias for Nino 3.4 domain
  - Similar to global domain but EP3 has the larger bias than EP2 for the longer lead forecast and warm bias is dominant.
- RMSE and spread:
  - EP3 - has smaller rmse than EP2 and the largest spread (best one) than others (Ocean initial perturbation?)

Figure 11: Coupling models (EP1, EP2, and EP3) reduce SST bias and rmse for most cases. Ensemble spread dramatically increases for both global and NINO3.4 domains for coupling models.

Summary

- GEFS v12 has been implemented in NCEP operation on September 23 2020
- GEFS v12 reforecast (31 years) has been finished on March 2020
- Evaluation and comparison:
  - Mainly compared to GEFS SubX which is latest (and excellent) system for weather and subseasonal prediction
  - GEFSv12 reforecast demonstrated equal or better performance than GEFS SubX
  - The initial analysis (model itself) plays an important role in the accuracy of the track forecast for the lead days shorter (longer) than about 5 days.
  - For CONUS high latitude winter season, the surface temperature is impacted by snow cover which could be as large as 3-4 degree colder (bias).
  - For CONUS precipitation application, we have demonstrated the systematic error (wet bias), and dry bias for extreme precipitation of raw forecast. The calibration through quantile mapping method could remove forecast bias, and improve forecast reliability, especially for extreme events (>50mm/day)
- Data for public access:
  - Selected data posted on AWS (2000-2019); NCEP ftp (1989-2019)
- Reference:

- Hamill, T. M., and Coauthors, 2022: The reanalysis for the Global Ensemble Forecast System, version 12. *Mon. Wea. Rev.*, **150**, 59–79, https://doi.org/10.1175/MWR-D-21-0023.1.
- Guan, H., and Coauthors, 2022: GEFSv12 reforecast dataset for supporting subseasonal and hydrometeorological applications. *Mon. Wea. Rev.*, **150**, 647–665, https://doi.org/10.1175/MWR-D-21-0245.1.
- Zhou, X., and Coauthors, 2022: The development of the NCEP Global Ensemble Forecast System Version 12. *Wea. Forecasting*, https://doi.org/10.1175/WAF-D-21-0112.1

Acknowledgements

EMC Ensemble team, Coupling Model and Model Physics Groups