Science and Technology Infusion Climate Bulletin
NOAA's National Weather Service
43rd NOAA Annual Climate Diagnostics and Prediction Workshop
Santa Barbara, CA, 23-25 October 2018

An Investigation of Prediction and Predictability of NCEP Global Ensemble Forecast System (GEFS)

Yuejian Zhu[1], Wei Li[2], Eric Sinsky[2], Hong Guan[3], Xiaqiong Zhou[2], and Bing Fu[2]
[1]*Environmental Modeling Center, NOAA/NWS/NCEP*
[2]*IMSG at Environmental Modeling Center, NOAA/NWS/NCEP*
[3]*SRG at Environmental Modeling Center, NOAA/NWS/NCEP*

1. Introduction

In recent years, the requirements for seamless forecast information have increased significantly for users that provide valuable guidance for public safety, quality of life, and business decisions that drive economic growth. A better understanding of predictability and numerical model prediction skills are greatly enhancing our capabilities for prediction and guidance for time scales ranging from weather, week-2, subseasonal to seasonal.

The predictability was originally introduced theoretically as a scientific question (Lorenz 1969); then it was investigated numerically and empirically based on the hypotheses (Lorenz 1982; 1996) for weather forecast. Following up the pioneer works, Shukla (1998) and many others also revealed predictability for seasonal and/or climate prediction. Until most recently, study of intrinsic predictability of state-of-art numerical modeling system has been used to further investigate potential predictability (Ying and Zhang 2017).

Considering both initial and model uncertainties, a state-of-art ensemble forecast system could be an optimum numerical system to quantitatively present the predictability across timescales which include weather and beyond. When assuming a perfect dynamical and physical model and a perfect ensemble system, forecast accuracy of each perturbed (and unperturbed or control) member should be "bias free", and have equal forecast skill statistically. Any individual forecast could be a "proxy truth" as well. Meanwhile, ensemble mean from this optimum system should present best prediction statistically, in the results, the average skill of ensemble mean forecast against individual perturbed (and control) forecast should represent potential forecast skill (or predictability).

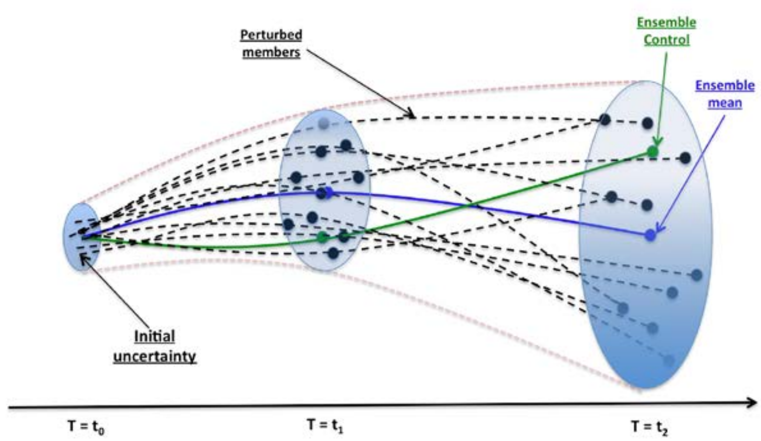

Fig. 1 Schematic diagram of an ensemble forecast starting from initial analysis/perturbations at t=0, growing with time evolution. Solid green is for unperturbed (ensemble control) forecast; solid blue represents ensemble mean; and dash blacks are perturbed ensemble forecasts. The shaded areas represent uncertainties of initial and forecasts.

2. Ensemble system

The NCEP Global Ensemble Forecast System (GEFS), by using a set of initial perturbations generated from EnKF analysis (Fig. 1; Zhou *et al.* 2016; Zhou *et al.* 2017), has been in daily operation since 1992. It has been providing reliable weather and week-2 probabilistic forecast guidance that has translated into valuable information for the general public. With many science enhancements in past years, current GEFS, especially SubX version, has reached super performance across the scales from weather to subseason (Zhu *et al.* 2017;

Correspondence to: Yuejian Zhu, Environmental Modeling Center, NOAA/NWS/NCEP, 5830 University Research Court, College Park, MD 20740; E-mail: yuejian.zhu@noaa.gov

2018). The two latest versions of NCEP GEFS have been used in this study, one of them is GEFS SubX version (Zhu *et al.* 2018; Li *et al.* 2019; Zhu *et al.* 2019; Guan *et al.* 2019) and another one is FV3-GEFS version (Zhou et al. 2019), that used Finite Volume dynamical core, higher (and uniform) resolution and GFDL Micro Physics (replace Zhao-Carr Micro Physics). The FV3-GEFS version has higher and uniform horizontal resolution (about 25 km) when compares to GEFS SubX version (33 km for 0-8 days; 55 km for 8-35 days). All ensembles are running 20 perturbed forecasts and one control forecast.

3. Evaluation methodology

The GEFS model performance has been presented in many studies (Zhou *et al.* 2017; Zhu *et al.* 2018), but the GEFS extended forecast to cover subseasonal timescales has only recently been evaluated (Zhu *et al.* 2018; Li *et al.* 2019; Guan *et al.* 2019) as part of the NOAA SubX (Subseasonal multi-model Experiments) project through an 18-year reforecast. The study proposed here involves a comparison of GEFS SubX results with those from the newly developed FV3-based GEFS, which includes a different dynamical core, horizontal resolution, microphysics, *etc.*

Various metrics could be used to assess the prediction skill and predictability of the forecast system, with dependencies on forecast elements, different spatial/temporal scales, and different forecast regions. In this investigation, anomaly correlation (AC) has been used to assess forecast skill and potential skill for weather. Meanwhile, the bivariate anomaly correlation (RMM1 and RMM2), a traditional real-time multivariate (RMM) MJO index (WH index; Wheeler and Hendon 2004; Lin *et al.* 2008; Gottschalack *et al.* 2010), has been used to evaluate tropical forecast skill and potential forecast skill.

In order to present an upper limit of prediction skill, the following principal assumptions (hypotheses) are applied to this evaluation: 1) Initial perturbations represent true observed uncertainty; 2) Numerical model is perfect and "bias-free"; 3) Ensemble system is perfect; 4) Ensemble forecast spread really represents true forecast uncertainty; 5) All individual perturbed forecasts could be proxy truth (and equal); 6) Ensemble mean will be best forecast solution for large scale forecast. Figure 2 demonstrates the root mean square (RMS) error of the ensemble mean and ensemble spread for Northern Hemisphere 500 hPa height of one-year statistics. The FV3-GEFS has less RMS error and a better ratio of RMS error and spread than GEFS SubX forecast. The RMS errors of both experiments have cross over climatological error around day 17, which means day-to-day forecast has lost skill beyond day 17. Meanwhile, ensemble spread has reached saturation level approximately at a similar range as the RMS difference between analysis and climatology (heavy dash green line), which indicates a good ensemble system has been used.

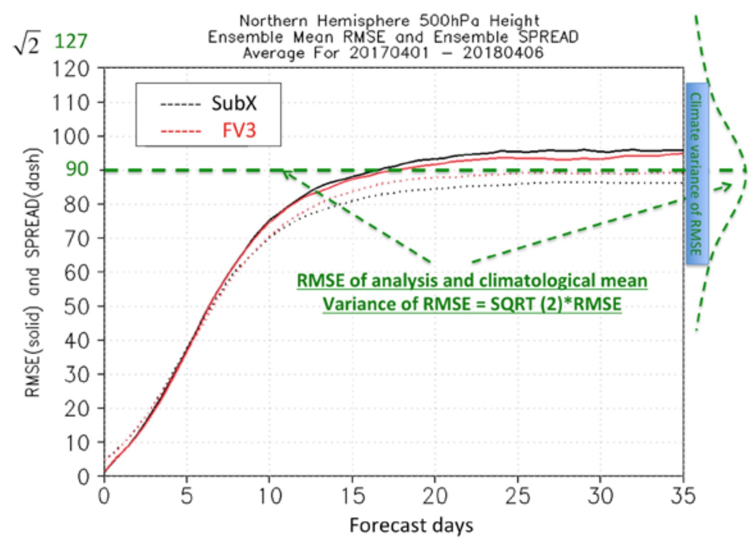

Fig. 2 Northern Hemisphere (NH) 500 hPa geopotential height RMS error (solid) and ensemble spread (dash) of 35 days forecasts from one year experiments based on two ensemble systems (SubX version - black and FV3 version - red). The RMS difference between analysis and climatological mean shows in green dash line (90 meters). The climatological variance around mean has been drawn on the right.

4. Results

One year experiments (April 2017 - April 2018) have been carried out for weather forecast and MJO evaluation. The NCEP Global Data Assimilation System (GDAS) analyses and NCEP/NCAR 40 years reanalysis climatology have been used to calculate forecast skills.

4.1 Overall NH extratropical weather forecast

Anomaly correlation is used to measure the real forecast skill and potential forecast skill (or predictability) of the 500 hPa geopotential height (Fig. 3). Apparently, FV3 GEFS (heavy-black) has slightly better skill than SubX GEFS (heavy-red) for all lead times (Fig. 3). Both of FV3 GEFS and SubX GEFS show similar potential forecast skills (thin lines) which indicate that a current forecast skill could be extended for an additional two days (from 12 days to 14 days) by using 50% AC as a skillful forecast. A potential forecast of SubX GEFS is also slightly better than FV3 GEFS for longer lead-time, which may be due to 1). SubX GEFS has less spread than FV3 GEFS for longer lead-time (see Fig. 2 dash lines); 2). Current FV3 GEFS has a disadvantage for longer range forecast; further improvement may be required.

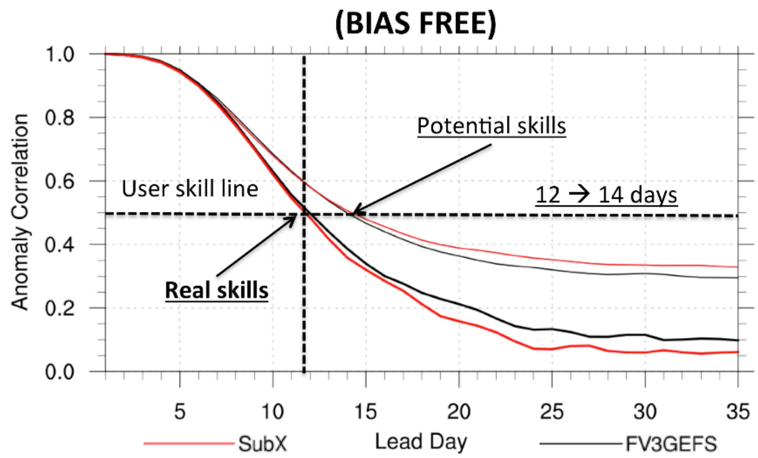

Fig. 3 Northern Hemisphere (NH) 500 hPa geopotential heights Pattern Anomaly Correlation (PAC) of the ensemble mean for 35 days forecast from one year experiments (April 2017 - April 2018). There are two experiments which are SubX version (red) and FV3 version (black). The real forecast skills (thick) and potential prediction skills (thin) are presented, and marked as 12 days (real) and 14 days (potential).

4.2 NH extratropical weather forecast of different spatial/temporal scales

The 500 hPa geopotential height has been decomposed to planetary scale (zonal wave 1-3); long wave (zonal wave 4-9); and synoptic weather pattern (zonal wave 10-20). The AC scores of these three groups have been presented in Fig. 4. Since the performance of FV3 GEFS is similar to SubX GEFS forecast for weather (Fig. 3), Fig. 4 only presents the AC scores for SubX GEFS version only in order to express three group scores clearly. The results demonstrate 1). The skills are different between these groups; 2). The planetary wave has more skill (12.5 days) than long wave (10.5 days) and synoptic weather forecast (7 days); 3). All three scale groups have potential forecast skills than current forecast skills about 1-2.5 days, planetary wave could have 15 days (> 2 weeks) potential forecast skill. Therefore, the forecast skill highly depends on a system or forecast task.

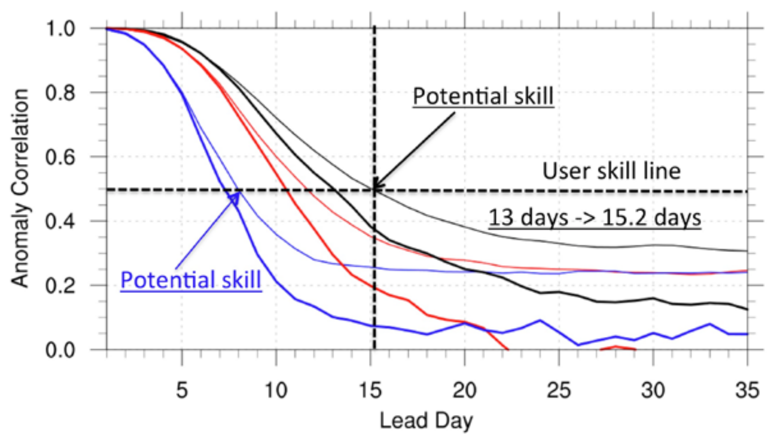

Fig. 4 Northern Hemisphere (NH) 500 hPa geopotential height Pattern Anomaly Correlation (PAC) of regrouped ensemble mean (zonal wave 1-3 (Length scale > 10000 km); 4-9 (10000 km < length scale > 3000 km); and 10-20 (3000 km >length scale > 1500 km) for SubX GEFS 35 days forecast from one year experiments (April 2017 - April 2018). The real forecast skills (thick) and potential prediction skills (thin) are presented for each group (zonal wave 1-3s are black; 4-9s are red; and 10-20s are blue).

4.3 MJO evaluations

The evaluation of the bivariate anomaly correlation of RMMs has focused on the period of one year (May 1st, 2017 - April 6, 2018) for FV3 GEFS only. Figure 5 has demonstrated the current forecast skill (20 days - 50% of AC) and potential prediction skill (32 days - 50% of AC). The result may indicate that 1) There is a good predictability of MJO event; 2) There is a large room to improve our model system includes dynamical, physical schemes, ensemble perturbations and many others (such as coupling atmosphere-ocean).

5. Summary

This is preliminary practice by using NCEP GEFS to assess potential forecast skills or predictability for NH mid-latitude weather forecast and tropical prediction. The AC scores of NH 500 hPa geopotential height indicate both of FV3 GEFS and SubX GEFS have similar real forecast skills with slightly advantage from FV3 GEFS. In contrast to real forecast skill, a potential forecast skill or predictability shows a similar conclusion for short lead-time forecast, but SubX GEFS has more potential forecast skills (or higher predictability) for longer lead-time (Fig. 3), which could indicate that either SubX GEFS may be a little under dispersive (spread is less than error; imperfect system) or FV3 GEFS may need more improvement to enhance its predictability.

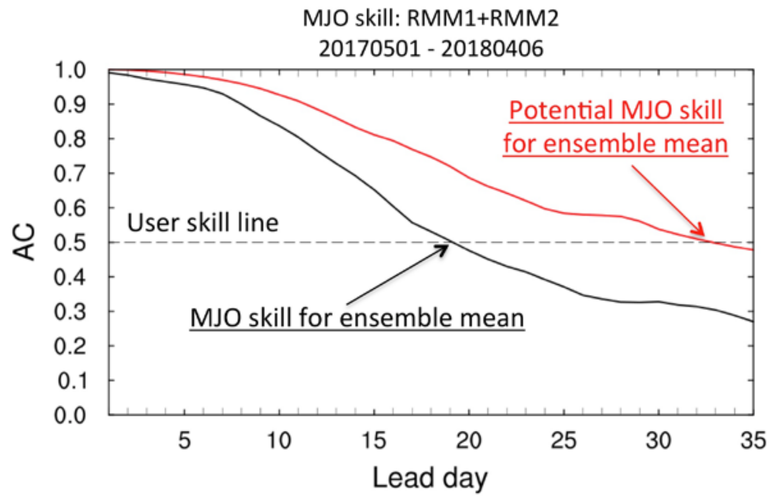

Fig. 5 The same 1-year experiment as figure 2-4, but for FV3 GEFS version only. A real MJO skill (black, *i.e.* analysis as reference) and potential MJO skill (red, *i.e.* control member as reference) of ensemble mean are presented.

For tropical prediction of FV3 GEFS, a difference of real forecast skill and potential forecast skill (or predictability) of MJO demonstrates a large potential capability (from 20 days to 32 days) of our system. A potential improvement could come from good forecast uncertainty representation; atmospheric circulation anomaly; enhanced MJO related tropical convection; interaction with ocean (coupling) and others.

Acknowledgements. The authors would like to thank all of helps from EMC ensemble team members. This study is partially supported through NWS OSTI and NOAA's Climate Program Office (CPO)'s Modeling, Analysis, Predictions, and Projections (MAP) program.

References

Gottschalck, J., and Coauthors, 2010: A framework for assessing operational Madden-Julian oscillation forecasts: A CLIVAR MJO working group project. *Bull. Amer. Meteor. Soc.*, **91,** 1247-1258, doi:10.1175/2010BAMS2816.1.

Guan, H., Y. Zhu, E. Sinsky, W. Li, X. Zhou, D. Hou, C. Melhauser, R. Wobus, 2019: Systematic error analysis and calibration of 2-m temperature for the NCEP GEFS Reforecast of SubX Project. *Wea. Forecasting,* doi:10.1175/WAF-D-18-0100.1.

Li, W., Y. Zhu, X. Zhou, D. Hou, E. Sinsky, C. Melhauser, M. Pena, H. Guan and R. Wobus, 2019: Evaluating the MJO forecast skill from different configurations of NCEP GEFS extended forecast. *Clim. Dyn.*, **52**, 4923-4936, doi:10.1007/s00382-018-4423-9.

Lin, H., G. Brunet and J. Derome, 2008: Forecast skill of the Madden-Julian oscillation in two Canadian atmospheric models, *Mon. Wea. Rev.*, **136**, 4130-4149.

Lorenz, E. N., 1969: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.*, **26**, 636–646, doi:10.1175/ 1520-0469(1969)26,636:APARBN.2.0.CO;2.

——, 1982: Atmospheric predictability experiments with a large numerical model. *Tellus,* **34**, 505–513, doi: 10.3402/tellusa.v34i6.10836.

——, 1996: Predictability — A problem partly solved. *Proc. Seminar on Predictability*, Reading, United Kingdom, ECMWF, 1–18.

Shukla, J., 1998: Predictability in the midst of chaos: A scientific basis for climate forecasting, *Science,* **282**, Issue 5389, 728-73, doi:10.1126/science.282.5389.728.

Wheeler, M. C. and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132,** 1917–1932, doi:10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2.

Ying, Y., and F. Zhang, 2017: Practical and intrinsic predictability of multiscale weather and convectively coupled equatorial waves during the active phase of an MJO. *J. Atmos. Sci.*, **74**, 3771-3785.

Zhou, X., Y. Zhu, D. Hou, and D. Kleist 2016: Comparison of the ensemble transform and the Ensemble Kalman Filter in the NCEP Global Ensemble Forecast System. *Wea. Forecasting,* **31**, 2058-2074.

——, ——, ——, Y. Luo, J. Peng and D. Wobus, 2017: The NCEP Global Ensemble Forecast System with the EnKF initialization. *Wea. Forecasting,* **32**, 1989-2004.

——,——, B. Fu, D. Hou, J. Peng, Y. Luo and W. Li, 2019: The development of next NCEP Global Ensemble Forecast System. *NWS STI Clim. Bull., 43^{rd} NOAA Annu. Clim. Diagn. Predict. Workshop,* Santa Barbara, CA, 23-25 October 2018.

Zhu, Y., X. Zhou, M. Pena, W. Lei, C. Melhauser, and D. Hou, 2017: Impact of sea surface temperature forcing on weeks 3 and 4 forecast skill in the NCEP Global Ensemble Forecasting System. *Wea. Forecasting*, **32,** 2159-2174, doi: 10.1175/WAF-D-17-0093.1.

——, ——, W. Li, D. Hou, C. Melhauser, E. Sinsky, M. Pena, B. Fu, H. Guan, W. Kolczynski, R. Wobus, and V. Tallapragada, 2018: Toward the improvement of subseasonal prediction in the National Centers for Environmental Prediction Global Ensemble Forecast System, *J. Geophys. Res.: Atmos.,* **123**, 6732-6745, doi: 10.1029/2018JD028506.

——, W. Li, X. Zhou, and D. Hou, 2019: Stochastic representation of NCEP GEFS to improve subseasonal forecast. In the book of *Current trends in the Representation of Physical Processes in Weather and Climate Models*, Editors: Dr. Randells *et al*., Springer Atmospheric Science, 317-328.