A study to examine the feasibility of a "forklift replacement" of the operational physics suite in the Global Forecast system (GFS)

Jack Kain
Chief, Model Physics Group, EMC

1. Background

With funding from the Next Generation Global Prediction System (NGGPS) initiative and broad support from the community, NCEP/EMC recently replaced the dynamic core in its flagship operational model, the Global Forecast System (GFS). Version 15 of the GFS (GFSv15), to be implemented operationally in Q3 of FY19, will be the first operational configuration of the model with the non-hydrostatic Finite-volume Cubed-Sphere (FV3) core, replacing the long-running spectral-Semi-Lagrangian reduced Gaussian-grid hydrostatic core. This modeling system will provide a fundamental early building block for the emerging Unified Forecast System (UFS) that is envisioned as a full community-based Earth-System model.

The next major upgrade of the GFS is expected to be in the area of model physics along with increased vertical (and possibly horizontal) resolution with higher model top. Physics packages, or suites, are comprised of numerous individual parameterizations (or schemes) that are designed to represent specific physical processes. In a physics suite, the individual parameterizations are highly interdependent so that improvements to one scheme typically must be accompanied by a series of modifications to the other schemes before the overall performance of the package improves. Replacement of an individual parameterization in a suite typically may require even more extensive tuning of the full package. As a result, the process of advancing the performance of physics suites at operational centers often proceeds iteratively and rather slowly.

NCEP/EMC recently conducted an experiment to assess whether advances in the representation of physical processes could be accelerated by implementing in the GFS (nearly) complete physics suites that had been tuned and optimized for different applications and/or modeling systems. This strategy was motivated by a desire to leverage both parameterization development and suite-optimization efforts that had occurred elsewhere in the meteorological community. This report provides a summary of the results of this experiment.

The physics suite in GFSv15 can be summarized as follows:

- **Moist Convection:** Scale-aware Simplified Arakawa-Schubert (sa-SAS) for deep convection; a simpler SAS-based mass-flux scheme for shallow convection (sa-MF)
- **Microphysics:** GFDL six-category single-moment
- **PBL/Turbulence:** Eddy-Diffusivity Mass-Flux (K-EDMF) scheme
- **Radiation:** RRTMG scheme
- **Land:** Noah LSM
- **Gravity-Wave Drag (GWD):** separate orographic/non-orographic components
- **Parameterized stratospheric water vapor/ozone chemistry**

The next implementation of the GFS, GFSv16, scheduled for 2021, is expected to include significantly advanced model physics along with coupling to WaveWatchIII wave model for improved parameterization of surface physics. Specifically, at least three new/improved parameterizations are targeted for implementation:

- The **Noah-MP** land-surface model (LSM) is expected to replace the Noah LSM.
- A unified gravity-wave-drag (**UGWD**) parameterization, developed by Valery Yudin and collaborators, will replace the separate orographic and non-orographic GWD schemes currently being used.
- The **RRTMG** radiation parameterization will be upgraded, leveraging collaborations between EMC scientists and multiple external partners

In addition, new parameterizations for deep and shallow moist convection (CP), cloud microphysics (MP), and planetary boundary layer (PBLP)/turbulence are being considered for GFSv16. For this "sub-suite" of parameterizations, there are multiple candidates for implementation, i.e., multiple high-performing CP-MP-PBLP sub-suites with technical readiness levels ≥ 7, including the schemes currently used in GFSv15. The experiment described herein was designed to engage the community in helping EMC assess whether advances in GFS physics could be accelerated by introducing already-tuned CP-MP-PBLP sub-suites instead of separate CP, MP, and/or PBLP components.

2. Overview of Test Plan

Two sub-suites of CP-MP-PBLP parameterizations were identified as possible replacements for the current GFSv15 sub-suite. One of these was derived from the operational RAP/HRRR modeling system and it was assembled and developed primarily by ESRL/GSD, drawing from years of community contributions through the WRF community modeling system for mesoscale applications (see the CP-MP-PBLP components within suite 4 in Table 1). The second sub-suite (CP-MP-PBLP components in suite 3 of Table 1) has components that were developed at multiple research centers and universities, including Colorado State, Utah, NASA, NCAR, and EMC. Its individual parameterizations have been applied primarily to medium-range and longer prediction timescales. The performance of these two sub-suites suites within the GFSv15 package was compared to the full GFSv15 suite (suite 1 in Table 1) and a very similar suite containing GFSv15 physics with a modified PBLP (suite 2 in Table 1).

	Suite 1 (GFS v15)	**Suite 2**	**Suite 3**	**Suite 4**
Deep convection	sa-SAS	sa-SAS	sa-CSAW	sa/aa-GF
Shallow convection	sa-MF	sa-MF	sa-MF	MYNN-EDMF and sa GF
Microphysics	GFDL	GFDL	aa-MG3	aa-Thompson
PBL/Turbulence	K-EDMF	sa-TKE-EDMF	K-EDMF	MYNN-EDMF
Land Surface Model	Noah	Noah	Noah	RUC

Table 1. Physics suites evaluated for possible implementation in GFSv16.
sa: Scale-aware; aa: aerosol aware; SAS: Simplified Arakawa Schubert; MF: Mass flux; MYNN: Mellor–Yamada–Nakanishi–Niino; EDMF: Eddy-diffusivity/Mass-flux; TKE: turbulent kinetic energy; CSAW: Chikira-Sugiyama-Arakawa-Wu; GFDL: Geophysical Fluid Dynamics Laboratory; MG3: Morrison-Gettelman; RUC: Rapid Update Cycle.

Each of these 4 suites were applied in cold-start retrospective forecasts using 163 different initialization times, designed to sample years 2015-18. Most of these initializations times (147) were chosen arbitrarily – every five days over the 2016-2017 two-year period – and an additional 16 start times were selected by EMC's Model Evaluation Group (MEG) to capture specific high-impact events.

EMC personnel designed the workflow and prepared the necessary input datasets for these retrospective forecasts. Full-resolution ECMWF initial conditions were used to initialize atmospheric variables for all forecasts (163 different initialization times, 10-day forecasts). Soil initial conditions were taken from retrospective, fully cycled runs of GFSv15 that were executed during pre-implementation testing of this system. The initialization for the RUC LSM was done internally in the GFSv15 model by interpolating the information from the levels supplied (the four levels of Noah LSM) to the nine levels used in the RUC LSM. Performance of each of the suites was evaluated using EMC's verification statistics database (VSDB), statistics from additional metrics generated by the Global Model Testbed (GMTB), and a detailed assessment from EMC's MEG, all emphasizing skill assessments for the 3-10 day forecast period that remains the highest priority time frame for GFS applications. A detailed summary of the testing process, provided by the GMTB, can be found here.

3. Results from Verification and Diagnostic Analyses

a. EMC Verification statistics database (VSDB)

A wide-ranging set of verification metrics were applied using the VSDB package, including verification scorecards. The full range of results from VSDB was computed separately for 0000 and 1200 UTC initialization times and can be found here for 0000 UTC and here for 1200 UTC.

b. GMTB Diagnostic report

GMTB scientists performed an array of analyses and diagnostic tests that were mostly beyond the scope of those provided by EMC's VSDB. Their draft report can be found here.

c. The EMC Mesoscale Evaluation Group (MEG) analysis

The MEG analyzed the verification statistics and diagnostics from both VSDB (0000 UTC, 1200 UTC) and the GMTB (summary), and they provided extensive diagnostic analyses of 16 cases that they had pre-selected because they were either 1) cases for which the operational GFS (GFSv14) and the FV3-GFS (GFSv15) forecasts both exhibited sub-optimal performance and/or 2) cases involving exceptionally high-impact weather events (including tropical cyclones). The MEG provided their analysis in a presentation given during their weekly webinar on March 21, 2019, and in a more comprehensive written report.

4. Formal Recommendations

In order to minimize the potential for a biased decision in selection of the CP-MP-PBLP sub-suite for GFSv16, an independent panel was enlisted to provide a recommendation to EMC. This panel consisted of Jian-Wen Bao (NOAA/ESRL/PSD), Lisa Bengtsson (CIRES-NOAA

ESRL/PSD), Jim Doyle (Naval Research Laboratory), and Jimy Dudhia (NCAR). Each of these panel members is a distinguished scientist with demonstrated knowledge and expertise in physical parameterization, but none is involved in the development of any of the parameterizations that were evaluated as part of this study.

The independent panel submitted their report and recommendations on March 26, 2019. Each of their Recommendations is listed below, along with an EMC response.

a. Recommendations for CP-MP-PBLP sub-suite for GFSv16:

1) **Panel recommendation:** Suite 2, which features a more advanced closure in the PBL scheme, was the closest in performance to Suite 1 and exhibited improvements in some important aspects including somewhat better capturing surface-based inversions and some better precipitation statistics over CONUS. Although overall Suite 2 did not perform as well as Suite 1, there are enough positive aspects in the Suite 2 performance to consider further experimentation and tuning in the near term (time permitting) to see if Suite 2 can be implemented in GFSv16. The panel believes the more advanced PBL may ultimately provide improved forecasts of the PBL.
   EMC response: We agree that Suite 2 provides the best compromise between potential future value and a realistic opportunity for optimization before our code-freeze deadlines for the GFSv16 implementation. We do share your concerns about this suite but we think the noted deficiencies can be overcome. Furthermore, we are planning to elevate the model top from the upper stratosphere to the mesopause, double the number of model vertical layers, include a new LSM, changes to radiation, and a new gravity-wave drag parameterization in GFSv16 as well. We think it is likely that some of the systematic biases that were revealed in the suite testing will change once the entire new suite is assembled and DA is included in the optimization process. The more advanced closure in the new PBL scheme, combined with the other new parameterizations, provide a real opportunity for us to improve the systematic near-surface and PBL biases that have been present in the GFS for some time.
2) **Panel recommendation:** Suites 3 and 4 both showed promising results in a number of aspects. We strongly encourage and recommend that the developers of both Suites 3 and 4 continue development and testing. We also recommend the developers to consider consolidating the best aspects of all suites, so attention can be focused on a single advanced development suite in the future.
   EMC Response: We also agree that suites 3 and 4 show promise in a number of aspects and we strongly agree with your recommendation that our best chance for accelerating advances in model physics is likely to come from consolidating the best aspects of all suites/schemes, rather than a strategy focused on shepherding individual parameterizations through the R2O funnel and into operations. This recommended approach would allow attention to be focused on a single advanced development suite in the future. Suites 3 and 4, and the individual parameterizations within them, clearly have desirable performance characteristics and we hope that, through collaborative efforts/partnerships, we can focus on clearly identifying the specific algorithms (parameterization components) that lead to these desirable characteristics and including these in a continuously advancing GFS physics suite.

b. Recommendations for the Future

1) **Panel recommendation:** We are supportive of a continued annual process in which an independent panel provides analysis and recommendations on the evaluation of parameterization suites considered for future operations.
   EMC response: Communications with the panel promoted a healthy dialogue between EMC and the external community. Furthermore, the panel's insights and formal recommendations were very beneficial for EMC scientists. We agree that a similar process, occurring on a regular basis, would be valuable to all parties if it can be done with as part of a well-planned activity for which both human and material resources are allocated ahead of time.
2) **Panel recommendation:** We encourage the testing and evaluation of other combinations of physics from the existing four suites, in addition to emerging physical parameterizations.
   EMC response: We agree that this is an important part of a broader strategy. The challenge will be to find balance and appropriate prioritization between improvement of existing parameterization approaches vs. introduction of new parameterizations
3) **Panel recommendation:** Adequate time for tuning and evaluation is needed prior to the test phase. The panel is aware of some issues related to the setups of suites 3 and 4 that impacted their results in this round and that these suites would have been improved given adequate pre-testing. Our recommendation is that a pre-test period of a few weeks should be built into the schedule using some of this year's initial data (but independent of the data in the next test).
   EMC response: A pre-test period of one month was built in to the schedule for this testing and all contributors/developers were made aware of the deadlines and time constraints months ahead of time. But, as often happens with complex, multi-agency efforts like this, development efforts ran behind schedule on several fronts. Furthermore, there was no flexibility on the start date of the retrospective runs because of EMC implementation deadlines, dependencies on a time window for availability of computer resources, and dependencies that other components of the prototype GFSv16 configuration have on model physics – among other factors. Since development milestones slipped while the start date of the retrospective runs remained fixed, the planned one-month pre-test period was compressed. It would seem prudent to plan for an even longer pre-test period (perhaps several months) to allow for unanticipated delays for all developmental code if a similar suite-selection approach is taken in the future.
4) **Panel recommendation:** It is recognized that this year was special because the physics framework was being changed at the same time as new physics were added, but the new framework should make implementations easier in the future.
   EMC response: Yes, this year the deadlines we established for physics-suite code development became conflated with external organizations' deadlines for EMC to accept the Common Community Physics Package (CCPP). This was a significant distraction as developers sought to prepare and optimize their physics codes, but the CCPP should facilitate future testing of physical parameterizations and suites.
5) **Panel recommendation:** Data assimilation cycling was not included in the current test suite and it may be useful in the future to include more testing with the data assimilation as part of the evaluations.
   EMC response: We agree. The most reliable assessment of physics-suite performance is gained through fully cycled testing with the operational DA package. However, this

strategy introduces additional degrees of freedom into the modeling system, requiring additional tuning, testing, debugging, and optimization of all codes – including developer engagement at all stages. Given the time constraints on our decision for GFSv16 physics, we made a pragmatic decision to do this assessment with a forecast-only process, using initial conditions that were generated by an independent (ECMWF) analysis and prediction system.

5. Concluding Remarks: This physics-suite experiment engaged multiple segments of the community on topics of compelling mutual interest, it produced a workable strategy for evaluating contributions from the community, and it set reasonable standards for fair and objective evaluation of these contributions. It was a very useful pilot project. However, it did not provide a conclusive answer to the question of whether advances in the representation of physical processes can be accelerated by implementing in the GFS (or any operational model) physics suites that have been tuned and optimized for different applications and/or modeling systems. Operational physics suites are typically the product of years of optimization and tuning within a single modeling framework and are therefore very difficult to beat with any relatively newly implemented suite. In this study, the suite that was minimally different from the operational suite had a level of performance that was comparable to the operational suite while the two suites that were configured quite differently performed significantly worse, on average.

While it is possible that the less skillful forecasts associated with suites 3 and 4 resulted from fundamental deficiencies in their ability to parameterize physical processes with fidelity, that is not likely. The more likely cause for the relatively large differences in skill is the short time allotted for optimization of these two alternative suites. It is quite certain that the performance of these two suites could have been improved with more time for optimization, but it is impossible to say how much improvement could be obtained and/or how much time it would take to achieve a given increase in prediction skill. The parameterizations tested in suites 3 and 4 have many desirable attributes, which is why they were selected for this study. But the results are not surprising. In the numerical modeling community it is not uncommon for developers to be confounded when model innovations that look better "on paper" - or seem to perform better in different modeling frameworks – do not increase overall skill when implemented in complex, highly nonlinear modeling systems. In the type of implementation strategy explored here, i.e., the "forklift approach", the operational suite will always be the most highly optimized package and probably have an inherent advantage on that basis. It is anticipated that this inherent bias will be mitigated (but not eliminated) when numerical modelers and scientists more fully adopt a common modeling framework, i.e., the UFS, for their research and development purposes. It would seem that full-suite or even full-parameterization replacement may not provide the most likely pathway to successfully advancing representation of physical processes in our models.

Given these considerations, the preferred path forward for advancing physics in EMC models is one in which innovations from the community are socialized and introduced through strong collaborative working relationships between developers/scientists at EMC and those in the broader community. Ideally, under the emerging community modeling umbrella, this would be characterized by joint, perhaps multi-agency efforts to improve existing operational parameterizations based on new research results, algorithm development, or numerical methods. Indeed this characterization aptly describes the process used to develop the TKE-EDMF scheme used in Suite 2, which was the product of close collaboration between EMC/IMSG scientist Jongil Han and University of Washington Professor Chris Bretherton. This kind of collaboration should be incentivized by funding agencies and research organizations over the forklift approach. The latter approach should not be precluded but it should be the exception rather than the rule because it tends to breed competition rather than collaboration, which is contrary to the fundamental motivation for developing a community modeling framework.

On the basis of this study and the recommendations from the independent panel, EMC will proceed with a plan to test and evaluate 4 potential upgrades to existing parameterizations as part of GFSv16 development:

- **PBL/turbulence:** K-EDMF => sa-TKE-EDMF
- **Land surface:** Noah => Noah-MP
- **GWD:** separate orographic/non-orographic => unified gravity-wave-drag
- **Radiation:** updates to cloud-overlap assumptions, empirical coefficients, etc. in **RRTMG**

These upgrades will be introduced along with a much higher model top and up to twice as many vertical levels as GFSv15, beginning immediately. Selection of a tuned prototype configuration for GFSv16, including these upgrades and an updated data assimilation package, is expected to be completed by the end of FY19.

Acknowledgements:

EMC workflow design, testing, collaboration, and execution: Ratko Vasic
EMC MEG: Geoff Manikin, Alicia Bentley, Logan Dawson, Tracey Dorian
EMC VSDB, coordination, and collaboration: Fanglin Yang
EMC technical development: Anning Cheng, Shrinivas Moorthi, Jongil Han, Ruiyu Sun, Fanglin Yang, George Gayno, Guang-Ping Lou
ECMWF data sharing: Linus Magnusson, Florian Pappenberger
GMTB development and execution team: Ligia Bernardet, Grant Firl, Lindsay Blank, Jimy Dudhia, Mike Ek, Michelle Harrold, Judy Henderson, Tracy Hertneky, Weiwei Li, Linlin Pan, Jamie Wolff, and Man Zhang
GSD Technical Development: Georg Grell, Joe Olson
NCAR METplus team: Tara Jensen, John Halley Gotway, Minna Win-Gildenmeister, George McCabe, and Julie Prestopnik
Independent Expert Panel: Jian-Wen Bao (NOAA/ESRL/PSD), Lisa Bengtsson (CIRES-NOAA ESRL/PSD), Jim Doyle (Naval Research Laboratory), and Jimy Dudhia (NCAR)
EMC final decision team: Jack Kain, Shrinivas Moorthi, Vijay Tallapragada, Russ Treadon, Fanglin Yang