# Report: Test of Grell-Freitas Convective Parameterization

Global Model Test Bed (GMTB)
February 2017

Points of Contact: Ligia Bernardet (ligia.Bernardet@noaa.gov) and Josh Hacker (hacker@ucar.edu)
GMTB contributing staff: Grant Firl, Michelle Harrold, Judy Henderson, Hongli Jiang, Louisa Nance, Jamie Wolff, and Man Zhang

# Table of Contents

Executive Summary ..... 1
Introduction ..... 2
Experiment Configuration ..... 3
Single-Column Model ..... 3
Global Model Experiments ..... 5
Key Findings from Single-column Model ..... 10
Key Findings from Global Model ..... 21
Discussion and Conclusions ..... 45
Acknowledgements ..... 47
References ..... 48
Appendix A. List of acronyms ..... 49

# List of Figures

Figure 1. Mean profiles of specific humidity tendencies (g kg-1 day-1) for the active phase of the TWP-ICE case. Colors denote forcing (red), PBL scheme (green), convective schemes (deep + shallow, blue), and microphysics scheme (purple). Line types denote the physics suite: GFS-SAS (solid) and GFS-GF (dashed). ..... 11
Figure 2. Same as Fig. 1 except for temperature tendencies (K day-1). Tendencies due to longwave and shortwave radiation are in orange and brown, respectively. ..... 12
Figure 3. Same as Fig. 1 but for the convective mass fluxes (kg m-2 s-1). Colors denote mass flux type: updraft (red), downdraft (green) and detrainment (blue). ..... 13
Figure 4. Mean profiles of specific humidity (kg kg-1) averaged over the active phase of convection for the TWP-ICE case. Colors denote observations (black) or physics suite (GFS-SAS in red and GFS-GF in green). Skill scores are printed in the legend. ..... 14
Figure 5. Same as Fig. 4 except for cloud fraction. ..... 14
Figure 6. Mean profiles of specific humidity (kg kg-1) averaged over the suppressed phase of convection for the TWP-ICE case. Colors denote observations (black) or physics suite (GFS-SAS in red and GFS-GF in green). Skill scores are printed in the legend. ..... 15
Figure 7. Same as Fig. 6 except for cloud fraction. ..... 16
Figure 8. Same as Fig. 6 except for temperature (K). ..... 16
Figure 9. Time series of total surface precipitation rate (mm h-1) for the active phase of the TWP-ICE simulation. Colors denote observations (black) or physics suite (GFS-SAS in red and GFS-GF in green). Skill scores are printed in the legend. ..... 17
Figure 10. Scatter plot denoting mean values of total surface precipitation rate (mm h-1) versus convective precipitation ratio over the active phase of convection for all forcing ensemble members for GFS-GF (green) and GFS-SAS (red). ..... 18

Figure 11. Same as Fig. 10 but for the suppressed phase of convection. ..... 19
Figure 12. Fig. 7 from Davies et al. (2013). The format and axes are the same as for figures 10 and 11. In this plot, colors denote models that participated in a SCM intercomparison project, with the GFS from circa 2011 in light blue. ..... 19
Figure 13. Mean profiles of specific humidity (kg kg-1) for the active convective phase of the TWP-ICE case. Colors denote observations (black) or physics suite (GFS-SAS in red and GFS-GF in green). Shading encompasses the 10-90th percentiles of the forcing ensemble. Thin lines denote the 25th and 75th percentiles. Thick lines denote the 50th percentile. ..... 20
Figure 14. Same as Fig. 13 except for the total convective temperature tendency (K day-1). ... 20
Figure 15. Same as Fig. 13 except for the total convective moisture tendency (g kg-1 day-1). . 21
Figure 16. Scorecard documenting performance of GFS-SAS and GFS-GF over the NH of mean bias and RMSE for temperature, relative humidity, and wind speed by forecast lead time and vertical level for JJA 2016. Green (red) shading indicates GFS-GF (GFS-SAS) is better than GFS-SAS (GFS-GF) at the 95% significance level. Small green (red) arrows indicate GFS-GF (GFS-SAS) is better than GFS-SAS (GFS-GF) at the 99% significance level. Large green (red) arrows indicate GFS-GF (GFS-SAS) is better than GFS-SAS (GFS-GF) at the 99.9% significance level. Grey shading indicates no statistically significant differences between GFS-SAS and GFS-GF. .... 22
Figure 17. Same as Fig. 16, except for SH. ..... 23
Figure 18. Same as Fig. 16, except for TROP. ..... 24
Figure 19 (a,b,c,d) Vertical profile of the median RMSE for temperature (°C) aggregated for JJA 2016 over the a) NH, b) CONUS, c) SH, and d) TROP regions. The 24-h forecast lead time is represented by the solid lines, the 192-h forecast lead time is dotted, and the 240-h forecast lead time is dashed. GFS-GF is green, GFS-SAS is red, and the differences (GFS-GF-GFS-SAS) is black. The horizontal bars surrounding the aggregate value represent the 95% CIs. ..... 26
Figure 20 (a,b,c,d) Same as Fig. 19, except for median RMSE of relative humidity *RH)..... 27
Figure 21. Two-meter temperature bias (°C) over the CONUS domain versus forecast lead time (h) for GFS-SAS (red) and GFS-GF (green) for JJA 2016. The short-dashed red lines show the warming trend in GFS-SAS, while the long dashed green lines indicates the relatively stable bias in GFS-GF. ..... 28
Figure 22. Mean 2-m temperature (K) for GFS-SAS (red), GFS-GF (green), and METAR observations versus forecast lead time (h) for JJA 2016. ..... 29
Figure 23. Same as Fig. 19, except for median bias for temperature (°C). ..... 30
Figure 24. Frequency bias of 24-h accumulated precipitation (in) for GFS-SAS (red) and GFS-GF (green) aggregated over the (a) NH, (b) SH, and (c) tropical region for JJA 2016. The 36-h forecast lead time is represented by the solid lines, the 132-h forecast lead time in dashed, and the 228-h forecast lead time is dot-dashed. The vertical bars surrounding the aggregate value represent the 95% CIs. ..... 32
Figure 25. Frequency bias of 6-h accumulated precipitation (in) for GFS-SAS (red) and GFS-GF (green) aggregated over the CONUS domain for the (a) 0.01”, (b) 0.1”, and (c) 0.25” thresholds as a function of forecast lead time (h) for JJA 2016. The vertical bars surrounding the aggregate value represent the 95% CIs. ..... 34

Figure 26. ETS of 24-h accumulated precipitation (in) for GFS-SAS (red) and GFS-GF (green) aggregated over the (a) NH, (b) SH, and (c) tropical region for JJA 2016. The 36-h forecast lead time is represented by the solid lines, the 132-h forecast lead time in dashed, and the 228-h forecast lead time is dot-dashed. The vertical bars surrounding the aggregate value represent the 95% CIs. ..... 36
Figure 27. Time series plot of 6-h accumulated precipitation (in) for aggregated ETS over the CONUS region for (a) 0.01", (b) 0.1", and (c) 0.25" for JJA 2016. GFS-GF is green, GFS-SAS is red, and the pairwise difference is black. The vertical bars surrounding the aggregate value represent the 95% CIs. ..... 38
Figure 28. Average 6-h accumulated convective precipitation (mm) over the three month test period (JJA 2016) at the 24-h forecast lead time for (a) GFS-GF, (b) GFS-SAS, and (c) GFS-GF - GFS-SAS..... 40
Figure 29. Same as Fig. 28, but for 120-h. ..... 41
Figure 30. Same as 28, but for 240-h..... 42
Figure 31. Mean track errors (nm) for GFS-SAS (red), GFS-GF (green) and their pairwise differences (black) with 95% confidence intervals with respect to forecast lead time (h) in the AL, EP, and WP basins for JJA 2016. ..... 43
Figure 32. Mean intensity errors (kt) with 95% confidence intervals with respect to lead time for GFS-GF (green), GFS-SAS (red), and GFS-GF - GFS-SAS pairwise difference (black) in the AL, EP, and WP basins for JJA 2016. ..... 44
Figure 33. Same as Fig. 32 but for mean absolute intensity error (kt). ..... 44
Figure 34. Model-generated TG counts by 10-day forecast period during the JJA 2016. ..... 45

# Executive Summary

The Global Model Test Bed (GMTB) conducted an assessment of the GFS using the Grell-Freitas (GF) cumulus parameterization (GFS-GF) and compared the results against a control run using the GFS operational cumulus parameterization, the Simplified Arakawa Schubert (SAS) scheme (GFS-SAS). The test, which employed the GMTB Single-Column Model (SCM) and the Global Spectral Model (GSM), was planned jointly by the GF scheme developer, GMTB, EMC, and representatives of NOAA's Next-Generation Global Prediction System program, with the goal of supporting the development of an advanced physics suite for the GFS.

As a first step in the evaluation of the GF scheme, global forecasts were run at a relatively low resolution (T574) in free-forecast mode (no data assimilation or cycling), and without tuning the physics suite. Results indicate that the GFS-GF forecast performance was as expected, given the constraints of the test, and that the scheme can be considered for more sophisticated testing.

Key findings follow and are further substantiated and discussed in the body of the report.

**SCM-1:** Given identical forcing, the GFS-GF suite produces weaker convective tendencies and convective transport than GFS-SAS. This alters the relationship among the physics schemes within the suite, leading to the explicit microphysics scheme in GFS-GF eliciting a greater relative response to the forcing.

**SCM-2:** Use of the GFS-GF suite reduces the dry bias in the boundary layer and generally produces a higher cloud fraction during the deep convective period compared to GFS-SAS for this case.

**SCM-3:** For the suppressed convection phase of the case, the GFS-GF suite produces an elevated temperature inversion and associated steep gradients in water vapor, leading to spurious cloud generation near the boundary-layer top.

**SCM-4:** The GFS-GF suite produces a much lower convective precipitation ratio compared to the GFS-SAS suite.

**SCM-5:** During the deep convective period, the forcing ensemble elicits greater variability from the GFS-GF suite than the GFS-SAS suite.

**GSM-1:** The results of RMSE and bias comparisons vary by forecast lead time, level, and region, with GFS-SAS displaying superior forecasts in more instances than GFS-GF. In upper levels, there are more differences between GFS-SAS and GFS-GF in temperature and relative humidity RMSE than in wind RMSE. When statistically significant pairwise differences were noted for wind speed RMSE, they nearly always favored GFS-SAS, regardless of level or region.

**GSM-2:** Upper-air temperature and relative humidity RMSE values are generally larger for GFS-GF than GFS-SAS but the favored configuration depends on forecast lead time and the vertical level. The advantage of GFS-SAS over GFS-GF is greater and more frequent earlier in the forecast. As forecast lead time progresses, the gap in performance narrows and GFS-GF is superior to GFS-SAS for some levels, lead times, and regions. This suggests that the GF scheme may not be in balance with the initial

conditions used in this test (operational GFS analyses), and that the GF might perform better in a cycled experiment.

**GSM-3:** A pronounced diurnal cycle in 2-m temperature bias is clear for both the GFS-GF and the GFS-SAS configurations. The GFS-SAS warms progressively through the forecast period over CONUS throughout the troposphere and at the surface, and gets colder in the tropics. The diurnal GFS-GF bias amplitude grows with forecast lead time.

**GSM-4:** In extratropical regions, precipitation frequency biases are similar overall between the model configurations, with over-precipitation for low thresholds and under-precipitation for high thresholds. However, the diurnal cycle of errors over the continental US are distinct between the configurations.

**GSM-5:** Overall, GFS-SAS is more skillful at predicting precipitation.

**GSM-6:** The partition of precipitation (convective and explicit) is different between the configurations, with SAS producing more total convective precipitation than GF.

**GSM-7:** Tropical Cyclone track errors averaged over the Atlantic, Eastern North Pacific, and Western North Pacific basins are similar for both model configurations. While accuracy in TC intensity forecasts is not expected from a model run at this coarse resolution, it is interesting to note that storms in GFS-SAS are more intense and have less absolute intensity error than those in GFS-GF.

**GSM-8**: While verification of cyclogenesis is beyond the scope of this report, it is noticeable that the models have different behaviors, with GFS-GF producing more storms.

# Introduction

To inform the development of an advanced physics suite for NOAA's GFS, the Global Model Test Bed (GMTB) conducted a test of the Grell and Freitas (GF, 2014) convective parameterization. This parameterization was selected for testing by EMC, and by the Program Office and Physics Team leads of the NOAA Next-Generation Global Prediction System (NGGPS) because of its potential for improving forecasts. It is a state-of-the-art scheme recently developed, but follows a long line of parameterizations from the developers. It includes a scale-aware feature, which make the scheme suitable for use in a wide range of model resolutions. Additionally, it incorporates an ensemble approach to the representation of convection, which can improve the forecast by using a collection of parameters and algorithms to represent the convective triggers, vertical mass flux, and closures. The ensembles can also be perturbed by stochastic fields for deterministic forecasting as well as ensemble data assimilation. Flux-form tracer transport, wet scavenging, and aerosol awareness are also options in this scheme. An additional factor that led to this choice was the scheme's maturity, its history of operational use at NCEP in the RAP, and the fact that its development is funded by NGGPS.

The plan for this test was devised jointly by the main developer (Georg Grell of NOAA ESRL/GSD), EMC, GMTB, the NGGPS Physics Team co-leads, and the NGGPS Program Office. The test was conducted by GMTB using its hierarchical testbed, composed of both a Single-column Model (SCM) and

a workflow to run the GFS. Given that this was the first test conducted by GMTB, a relatively new group within the Developmental Testbed Center (DTC), this test vetted various aspects of the hierarchical testbed, and was an exercise to get the testbed in place and ready for future physics assessments.

This report focuses on the experiment configuration, key findings, and discussion of results. For further information not covered in this report, please see the comprehensive verification results here.

# Experiment Configuration

This test was conducted using GMTB's hierarchical testbed, which currently consists of a SCM, and a workflow for running the GFS. The verification and graphics part of the workflow were developed by GMTB, while EMC contributed the system to run the preprocessing, the model forecasts, and the post-processing.

## Single-Column Model

### Overview

For this test, a single case based on a deep convection-focused field campaign was used to provide insights into how the operational GFS physics suite performs when compared to a suite modified to use the GF deep and shallow convective schemes. The testing paradigm follows one described in Randall et al. (2003) and Zhang et al. (2016), namely initial conditions and column forcing are derived from observations obtained during Intense Observation Periods. The atmospheric physics suite that comprises the SCM is allowed to respond to the forcing by generating parameterized clouds and precipitation, radiative heating, vertical mixing, etc.

Physics suite performance can be gauged by comparing diagnosed physical quantities to real observations or "synthetic" observations derived from LES. Interpretation of the results can be more straight-forward than those from a global model due to the lack of three-dimensional interaction, and absence of error propagation throughout the components of a more complete model. Results interpretation must be tempered for the same reason -- SCM results can be highly dependent on the prescribed forcing; physics suite performance in a SCM may not always translate to a global model. At the end of the day, however, such testing allows a "deep dive" into how physics suites respond to identical initial conditions and forcing, which is difficult and sometimes impossible with more complex models.

### Source Code

The code for running the SCM portion of the test resides in NOAA's Virtual Laboratory (VLab) under the "gmtb-scm" project name (further information can be found here). Within this Git project, the specific code for running this test can be found in the "gf-test" branch under the tag v1.2. This repository contains both the GMTB SCM infrastructure code and a checked-out version of a NEMSLegacy branch, specifically r85909 of the branch "gf_test_new". The NEMSLegacy code contains the GFS physics and its driver, and is identical to the code used in the global portion of this test. The GMTB SCM code

interfaces with the GFS physics through the version of nuopc_physics.f90 found in the specified NEMSLegacy branch. Both the control runs and the experimental runs using the GF convection use a version of gbphys.f that was modified to call the GF scheme. The only difference between the two runs is the specification of the convection scheme, which is controlled through the namelist variables imfdeepcnv and imfshalcnv.

## Case

The SCM was configured to run the GCSS Deep Convective Working Group's sixth intercomparison case based on the ARM Tropical Warm Pool - International Field Experiment (TWP-ICE) field campaign as described in Davies et al. (2013). The case is based on a suite of observations obtained near Darwin, Australia in January and February of 2006. Meteorological conditions observed included deep convection associated with an active phase of the monsoon and suppressed convection and clear sky associated with the inactive phase. The initial profiles of temperature, moisture, and horizontal winds reflect average conditions over the study area (centered on 12.425°S, 130.891°E) at 0300 UTC on January 19, 2006. The surface is oceanic with a fixed SST, implying interactive surface fluxes calculated by a surface-layer scheme, and an observed ozone profile is included for use with interactive radiation. The effect of large-scale advection on the temperature and moisture profiles is calculated using two separate terms following the "horizontal advective forcing" method of Randall and Cripe (1999): prescribed horizontal advective tendencies plus a vertical advective term that combines the prescribed vertical velocity and the modeled temperature and moisture profiles. Horizontal wind profiles are relaxed to observed profiles on a timescale of two hours. Forcing for the SCM is supplied for the entire length of the TWP-ICE field campaign from 0300 UTC on January 17, 2006 to 2100 UTC on February 12, 2006. Following Davies et al. (2013), the simulation period was split into two time periods for analysis -- one that featured active, deep convection (from 0000 UTC on 20 January to 1200 UTC on 25 January) and one that featured suppressed, shallow convection (from 0000 UTC on 28 January to 1200 UTC on 2 February). In addition to a "best estimate" forcing dataset for the time period, a 100-member forcing ensemble is utilized to gauge sensitivity to the supplied forcing. The forcing ensemble was created by quantifying uncertainty in the surface rainfall measurement and using the constrained variational analysis method to derive 100 equally likely forcing profiles. The greatest change among the forcing datasets is in the prescribed vertical velocity, which is very sensitive to surface precipitation. The forcing ensemble is described in detail by Davies et al. (2013).

When observations are available for a particular quantity, a skill score was calculated following Taylor (2001) as:

$$S = \frac{4(1+R)^4}{\left(\sigma_f + \frac{1}{\sigma_f}\right)^2 (1+R_0)^4}$$

where R is the correlation coefficient between simulated and observed quantities and $\sigma_f$ is the ratio of modeled to observed variance. This score is printed in the legend for each suite.

## Other Aspects

One benefit of the hierarchical testing infrastructure, namely utilizing the same physics code for both the global and SCM platforms, was demonstrated during this test. Because the SCM is relatively simple and inexpensive, it was utilized multiple times during the testing process to ensure proper interfacing of the developer's code with the existing GFS physics suite. It was used both to determine the correct method to supply the developer's code with input arguments and to troubleshoot the GF code after initial global runs indicated a problem. Further, the SCM provided an efficient way to gauge the scheme's sensitivity to parameters that can be customized. For this test, the SCM was used to explore the GF scheme's sensitivity to the parameters "dicycle" and "ichoice," which control how the scheme adjusts for the diurnal cycle of convection and which closures are used, respectively.

# Global Model Experiments

## Overview

Retrospective forecasts for JJA 2016 were generated using a developmental version of the GFS that employs the GSM dynamic core in the NOAA Environmental Modeling System (NEMS), slated for operational implementation in May 2017, the Physics Driver v3, and the physics suite used in the 2016 operational implementation of the GFS. As a first approach to testing GFS-GF, and to conserve computational resources, the model was run in cold-start mode and at a resolution coarser (T574) than the one used for the operational GFS (T1534).

## GFS, NEMS, GSM, and Physics

Code management for this test was done using the EMC Apache Subversion (SVN) server. Branches named gf_test were created off the top of trunk of the NEMSLegacy, NEMS, and GSM repositories on June 07, 2016. The GF code provided by the developer was placed alongside the GFS operational physics in the GSM phys directory. Changes were made to a few files in the GSM code, including phys/gbphys.f, phys/compns_physics.f90, and phys/gloopr.f90 to accommodate the requirements of the GF scheme, and to create the ability to select between the SAS and GF in the runs. The final code used for the test is under the gf_test_new branch revision 85909.

To ascertain whether GFS-SAS runs were producing reasonable output and within an appropriate range of the operational GFS, a sanity check was conducted by comparing verification results between the GFS-SAS and the operational GFS. This comparison confirmed the results from GFS-SAS run were reasonable. Swapping the SAS scheme for the GF scheme was the only change to the operational GFS physics suite (Table 1).

*Table 1. Description of the GFS-SAS and GFS-GF physics suites.*

| Physical Process | Operational Suite | Experimental Suite |
|---|---|---|
| Convection (deep and shallow) | SAS | GF |
| Turbulent transport (PBL) | Hybrid Eddy-Diffusivity Mass-Flux | |
| Radiation | RRTM for General Circulation Models (RRTMG) | |
| Gravity wave drag | Orographic and stationary convective | |
| Land surface model | Noah | |
| Cloud microphysics | Zhao-Carr | |

When the GF scheme is activated in the namelist by setting imfdeepcnv (deep convection; option 3) and imfshalcnv (shallow convection; option 3), a number of other parameters specific to GF can be set. The specific parameters used in this test were selected in collaboration with the developer through an iterative process of running several cases to exercise the different options (Table 2). The SCM was particularly useful for this phase. Mid-level clouds (imid; option 0) were turned off for this experiment. The average of all possible closures was chosen for deep convection (ichoice; option 0), and for the closures for shallow convection (ichoice_s), option 2 was selected. The diurnal cycle adjustment was also activated for this test (dicycle; option 1). For the SAS, the 2016 operational settings were used (imfdeepcnv and imfshalcnv set to 1).

*Table 2. Description of parameters used in the GF Test.*

| | imfdeepcnv | imfshalcnv | imid | ichoice | ichoice_s | dicycle |
|---|---|---|---|---|---|---|
| SAS | 1 | 1 | -- | -- | -- | -- |
| GF | 3 | 3 | 0 | 0 | 2 | 1 |

Throughout the preparatory period of the test, the GMTB received several updates to the GF code. Those were prompted by bugs and issues identified by GMTB staff and the developer when conducting case studies and SCM runs. One example is the need to standardize the units of latent and sensible heat fluxes between the GFS and the GF codes. Additionally, the developer restructured the code from one to three files (module_cu_gf_driver.F90, module_cu_gf_deep.f90, module_cu_gf_sh.gf90) to facilitate the use of the code in different modeling systems. While the driver is specific to the model solver, the same deep and shallow convection codes can be used by any model. The code was in final form on December 27, 2016.

The GMTB conducted a test to ascertain that forecasts using SAS were bitwise identical to forecasts created using the branch with the GF code or the trunk. With the original configure file, the results were not reproducible, but GMTB staff were able to trace this to different compilation options. To generate reproducible results, the compilation on Theia had to be adjusted by setting -fp-model from strict to precise. This change was made to the FFLAGS line in NEMS/src/conf/configure.nems.Theia.intel_gsm.

## Post-processing, Graphics, and Diagnostics

The unipost program within NCEP's Unified Post-Processor (UPP) v7.5.1 was used to output the necessary variables at specified levels, derive additional meteorological fields, and vertically interpolate fields to isobaric levels. The post-processed forecast files included two- and three-dimensional fields, which are necessary for both the plotting routines and verification tools. The necessary parameter files for unipost were based on those being utilized at NCEP for parallel testing. Output from unipost were in Gridded Binary Version 2 (GRIB2) format, and the wgrib2 utility were used to interpolate the post-processed files to a 0.25° global grid (G193).

The cyclone tracker used with the operational GFS was applied to extract information about location and intensity of storms identified by NHC and JTWC, as well as to identify tropical cyclogenesis events.

Graphics of model output from UPP were created using Python and included a suite of figures by ingesting the 0.25o GRIB2 files, and either plotting the gridded data directly, or regridding it to various verification grids used by NCEP. Tropical Cyclone (TC) track and intensity plots were also created. The test plan provides a comprehensive list of the variables plotted for each model forecast.

## Verification

Objective model verification statistics were generated using the Model Evaluation Tools (MET) version 5.2; (http://www.dtcenter.org/met/users/docs/overview.php), and verification results plotted with the METViewer. For point-based verification, postprocessed model output for surface and upper-air variables were compared to observations (METARs and RAwinsonde OBservations - RAOBs) using the MET point-stat tool. The 0.25° model output was regridded to G218, a 12-km Lambert Conformal grid covering the Contiguous United States (CONUS) and evaluated using NAM Data Assimilation System (NDAS) files in quality controlled BUFR (PrepBUFR) files as the observational dataset for the surface verification. For upper-air verification, the 0.25° model output was regridded to both the G218 and G3 (a global 1.0° latitude-longitude domain) and evaluated using NCEP's Global Data Assimilation System (GDAS) PrepBUFR files as the observational dataset. Bias (or mean error) and RMSE were computed separately for each variable at the surface and upper-air levels. For the surface variables, statistics were aggregated over the CONUS domain along with 14 sub-regions. For brevity of the report, the focus is on CONUS results; the sub-region results are available via the project webpage. Upper-air statistics over global, CONUS, Northern Hemisphere (NH; 20° – 80° N), Southern Hemisphere (SH; 20° – 80° S), and Tropics (TROP; 20° S – 20° N) domains for G3 are all available on the report website.

Precipitation verification was performed over the entire globe. For the CONUS domain, a grid-to-grid comparison was made using the QPE from the CCPA dataset, which has a resolution of ~4.8 km. Both the Climatology-Calibrated Precipitation Analysis (CCPA) QPE analyses and the 0.25° post-processed model output was interpolated to G218. For the global evaluation, CMORPH precipitation analyses (60° N-60° S) was used due to its high spatial (8 km at the equator, ~0.07°) and temporal resolution. Both the CMORPH analyses and the 0.25° post-processed model output was interpolated to G3 and compared over the NH (20° – 60° N), SH (20° – 60° S), and Tropics (20° S – 20° N). Precipitation verification focused on a 24-h accumulation period (valid from 12 UTC to 12 UTC) using the MET grid-stat tool. Traditional verification metrics computed for both CONUS and global regions include the frequency bias and the ETS.

Because both configurations were run over an identical set of forecasts, the pairwise difference methodology was applied, when appropriate. With this methodology, differences between the verification statistics were computed by subtracting GFS-SAS from GFS-GF. The confidence intervals (CIs) on the pairwise differences between statistics for the two configurations objectively determine whether the differences are SS. For surface and upper-air, both the individual and pairwise verification statistics were accompanied by CIs computed from standard error estimates. The CIs were computed on the median values of the aggregated results for the surface and upper-air statistics using parametric tests. For the precipitation statistics, the bias-corrected and accelerated bootstrap method (using 1500 replicates) was used. The CIs on the pairwise differences between statistics for two configurations will assist in determining whether the differences are statistically significant. All CIs were computed at the 95% level.

With numerous verification results being produced from this test, a “scorecard” is a straight-forward way to identify patterns in the difference of performance between two configurations, including level of significance for specified metrics, variables, levels, regions, and times. EMC has this capability as standard part of their verification arsenal, and development in the DTC is currently underway to include this capability in MET/METViewer. This report will include verification results using a development version of the DTC’s scorecard. Note that this initial capability computes the means of the differences, while the surface and upper-air verification discussed above calculates the medians of the differences. As the scorecard capability matures, more user options will be included.

The TC evaluation of GMTB retrospective forecasts focused on a direct comparison between GFS-GF and GFS-SAS over three basins: Atlantic (AL), Eastern North Pacific (EP), and Western Pacific (WP). The errors resulting from GFS-GF and GFS-SAS configurations were computed relative to the TC Best Track data using MET-TC version 5.2. Statistics for the individual cases were aggregated using a script in the R statistical language, without discriminating for TC placement over land or water. All aggregations were done for homogeneous samples (i.e., only cases for which both the GFS-GF and GFS-SAS were available were included in the aggregation statistics). Given the distribution of errors, and absolute errors at a given lead time, several parameters of the distribution were computed: mean, median, quartiles, and outliers. All aspects of the evaluation include an assessment of SS based on 95% CIs; Table 3 lists the names and ATCF identifiers of the storms in the verification sample.

*Table 3. Name and identifier of TCs verified for GFS-GF and GFS-SAS.*

| AL (8) | EP (14) | WP (12) |
|---|---|---|
| BONNIE (02L) | ONE (01E) | NEPARTAK (02W) |
| COLIN (03L) | AGATHA (02E) | LUPIT (04W) |
| DANIELLE (04L) | BLAS (03E) | MIRINAE (05W) |
| EARL (05L) | CELIA (04E) | NIDA (06W) |
| FIONA (06L) | DARBY (05E) | OMAIS (07W) |
| GASTON (07L) | ESTELLE (06E) | CONSON (08W) |
| EIGHT (08L) | FRANK (07E) | CHANTHU (09W) |
| HERMINE (09L) | GEORGETTE (08E) | MINDULLE (10W) |
| | HOWARD (09E) | DIANMU (11W) |
| | IVETTE (10E) | LIONROCK (12W) |
| | JAVIER (11E) | KOMPASU (13W) |
| | KAY (12E) | FOURTEEN (14W) |
| | LESTER (13E) | |
| | MADELINE (14E) | |

Skillful forecasting of Tropical Cyclogenesis (TG) is a difficult challenge, and it is important to include diagnostics of this phenomenon in the evaluation. In addition to the verification of existing storms, TG counts were obtained from the cyclogenesis files generated from NCEP tracker software and compared against the development of new storms as described in the Best Track.

## Scripts and Automation

This test included two workflows. The first set of scripts used EMC's workflow (v14.1.0) for the NEMS-based GFS. These scripts are responsible for a variety of tasks, including setting up environment variables, running the forecast model, post-processing, tracking tropical cyclones, and detecting tropical cyclogenesis. Additionally, these scripts are typically used for automation purposes, as the various tasks are submitted to the batch system incrementally as dependencies are met. The scripting architecture initiates a number of executables compiled by GMTB staff under GMTB project space on Theia (including the NEMS-based GSM executable), in addition to some hard-coded paths to pre-existing executables compiled by EMC. GMTB staff made modifications to the scripts, diverging from the EMC directory

structure, to accommodate a directory structure more amenable to the GF test (e.g., output directories for multiple configurations and separating runs by initialization). In addition, modifications were made to the para config file (names of variables and their corresponding values) and rlist files (files used to define input and output files and files to be archived). These files are not in the EMC repositories, but the GMTB has placed them under version control.

The second set of scripts, contributed by GMTB and automated through the Rocoto Workflow Management System, were used to stage datasets, create forecast graphics, run forecast verification, archive results, and purge the disk. These scripts, as well as the para config and rlist files mentioned above, are kept under version control in NOAA's Vlab under the "gmtb-tierIII" project, which can be accessed here.

### Initial Conditions, Forecast Periods, and Length

Initial conditions were the T1534 operational GFS analyses, translated to T574 resolution and NEMSIO format using the chgres program supplied by EMC.

To allow the evaluation of statistical significance, the test covered a three-month period (JJA 2016). Forecasts were launched once a day at 0000 UTC and run out to ten days with output every six hours.

### Archives

Output data files from multiple stages of the global workflow system were archived in the NOAA High-Performance Storage System (HPSS) (location: /2year/BMC/gmtb). Archives include:

- Configuration files and namelists specific to each forecast cycle;
- Forecast files from GSM (analysis and forecasts at 6-hour increments);
- 0.25° GRIB2 forecast files from unipost (analysis and forecasts at 6-hour increments);
- Graphics from Python plotting suite and diagnostic routines; and
- Output from MET and MET-TC.

---

The remainder of this document presents the key findings from the SCM and GFS tests. A comprehensive verification is available at http://www.dtcenter.org/eval/gmtb/gftest.

# Key Findings from Single-column Model

**SCM Key Finding 1:** Given identical forcing, the GFS-GF suite produces weaker convective tendencies and convective transport than GFS-SAS. This alters the relationship among the physics schemes within the suite, leading to the explicit microphysics scheme in GFS-GF to show a greater relative response to the forcing.

Figure 1 shows the mean profiles of tendencies of specific humidity due to the supplied forcing and the parameterizations averaged over the deep convective period. Figure 2 shows the same for the

temperature tendencies with the addition of curves due to longwave and shortwave radiation. While there are no observed tendencies to compare with, these plots are useful for interpreting how the two suites respond to the same inputs and provide value for interpreting quantities that do have observational analogs. It is clear that the convective tendencies produced in the GFS-GF suite are generally weaker than those from the GFS-SAS suite. Less column-wise drying and heating due to the convection shifts the response by the rest of the physics suite, most notably in the microphysics scheme. The microphysics scheme in the GFS-GF suite "compensates" by increasing its role -- producing a larger share of the total condensate and precipitation.

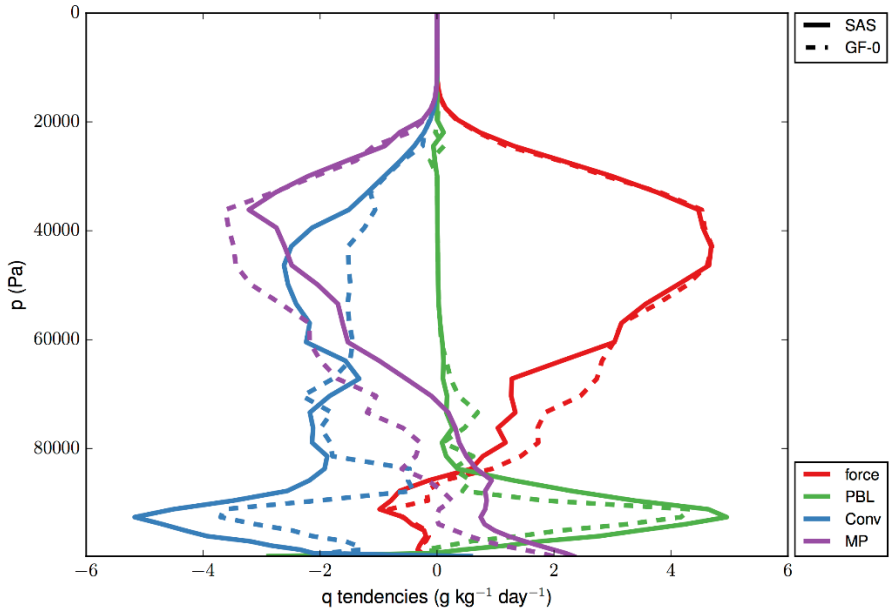



*Figure 1. Mean profiles of specific humidity tendencies (g kg$^{-1}$ day$^{-1}$) for the active phase of the TWP-ICE case. Colors denote forcing (red), PBL scheme (green), convective schemes (deep + shallow, blue), and microphysics scheme (purple). Line types denote the physics suite: GFS-SAS (solid) and GFS-GF (dashed).*

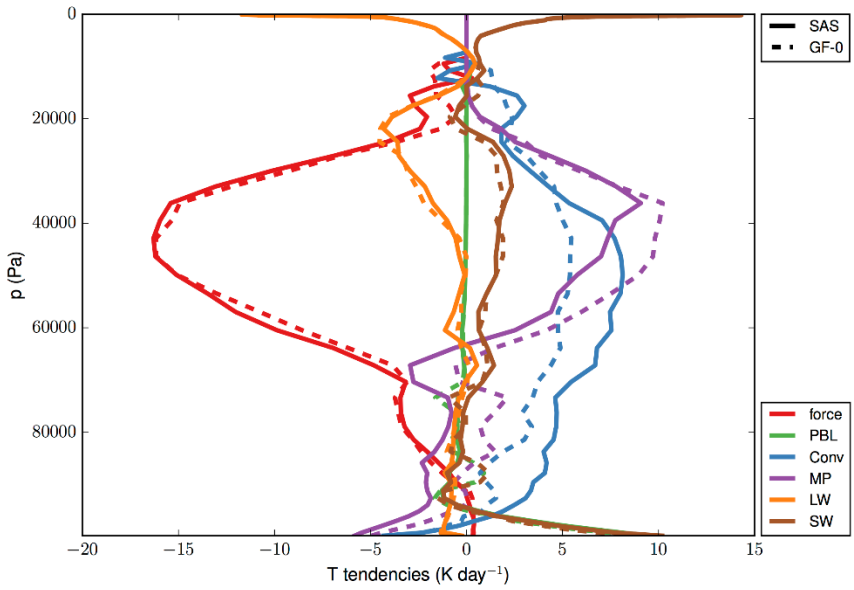



*Figure 2. Same as Fig. 1 except for temperature tendencies (K day$^{-1}$). Tendencies due to longwave and shortwave radiation are in orange and brown, respectively.*

Note that the forcing curves in these plots are not identical due to how forcing was applied for this case -- recall that the advective forcing terms are split into horizontal and vertical terms and only the former is explicitly prescribed. The vertical advective forcing depends on the modeled profiles of temperature and moisture, so any differences in the forcing profiles must be the result of differences in those profiles.

Figure 3 shows the mean convective mass fluxes averaged over the active convective period with red denoting updrafts, green denoting downdrafts, and blue denoting detrainment. The GF scheme generates a more symmetric profile of updraft mass flux and much reduced magnitudes compared to the SAS scheme. While both schemes generate bottom-heavy profiles of downdraft mass flux, the GF scheme's maximum value is approximately 20% of the SAS scheme. This provides evidence that the GF scheme is less active in vertical transport for this case, ultimately leading to the weaker convective tendencies shown in the previous two figures.

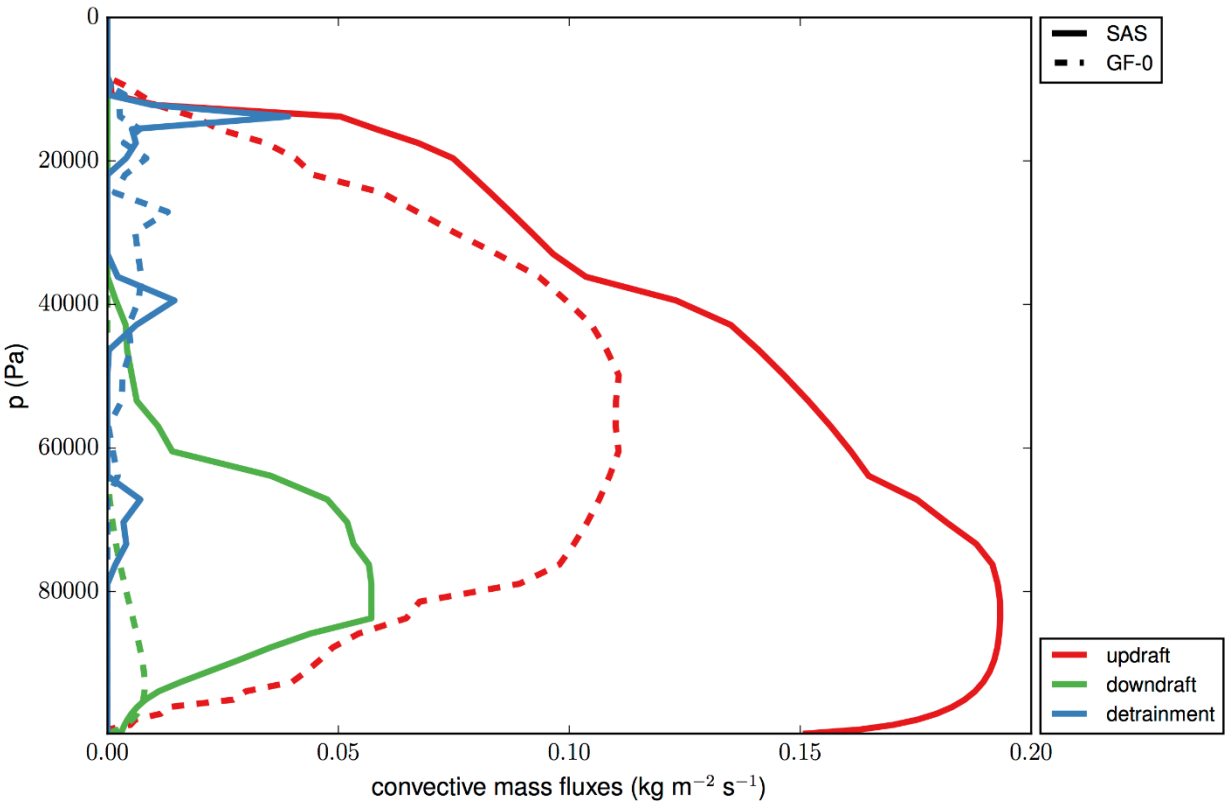



*Figure 3. Same as Fig. 1 but for the convective mass fluxes (kg m$^{-2}$ s$^{-1}$). Colors denote mass flux type: updraft (red), downdraft (green) and detrainment (blue).*

**SCM Key Finding 2:** Use of the GFS-GF suite reduces the dry bias in the boundary layer and generally produces a higher cloud fraction during the deep convective period compared to GFS-SAS for this case.

Figures 4 and 5 show mean profiles of specific humidity and cloud fraction compared to observations for the active deep convective period. The profiles demonstrate that the GFS-GF suite reduces the dry bias below approximately 700 hPa and produces a larger cloud fraction throughout the column compared to GFS-SAS, both features that increase GFS-GF's overall skill score. The reason for the improvement can be partially explained by the tendency profiles previously presented. Reduced convective transport of moisture out of the PBL and the associated steeper moisture gradient that affects the vertical advective forcing term are likely responsible for this improvement. It should be noted that cloud fraction profiles for the SCMs that participated in the intercomparison reported in Davies et al. (2013) (their Fig. 9) have similar shapes to the ones simulated here -- all differ considerably from the observed profile.

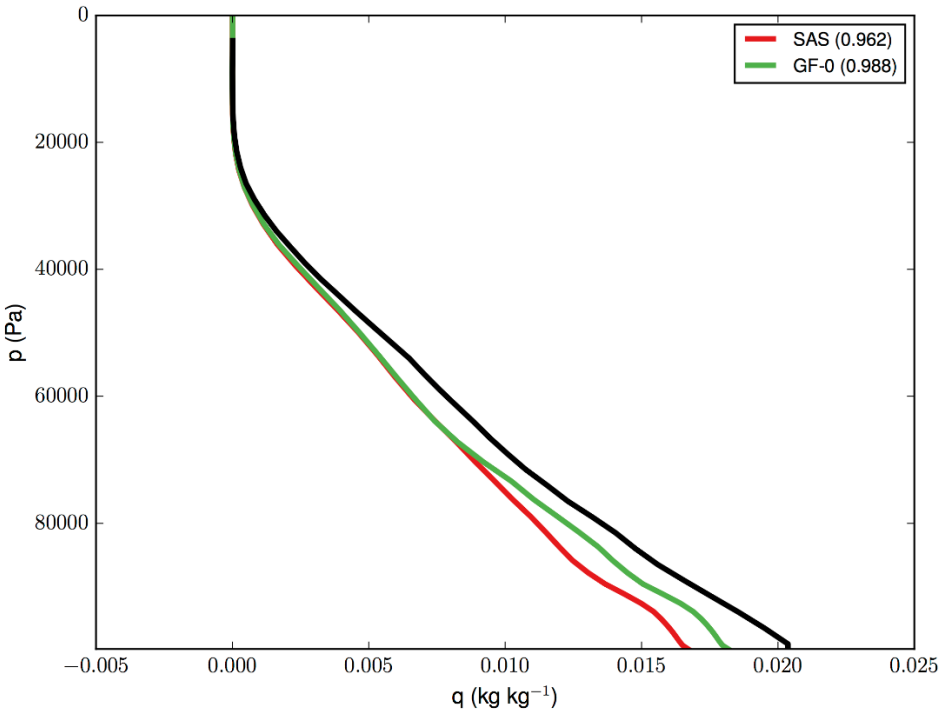



*Figure 4. Mean profiles of specific humidity (kg kg$^{-1}$) averaged over the active phase of convection for the TWP-ICE case. Colors denote observations (black) or physics suite (GFS-SAS in red and GFS-GF in green). Skill scores are printed in the legend.*

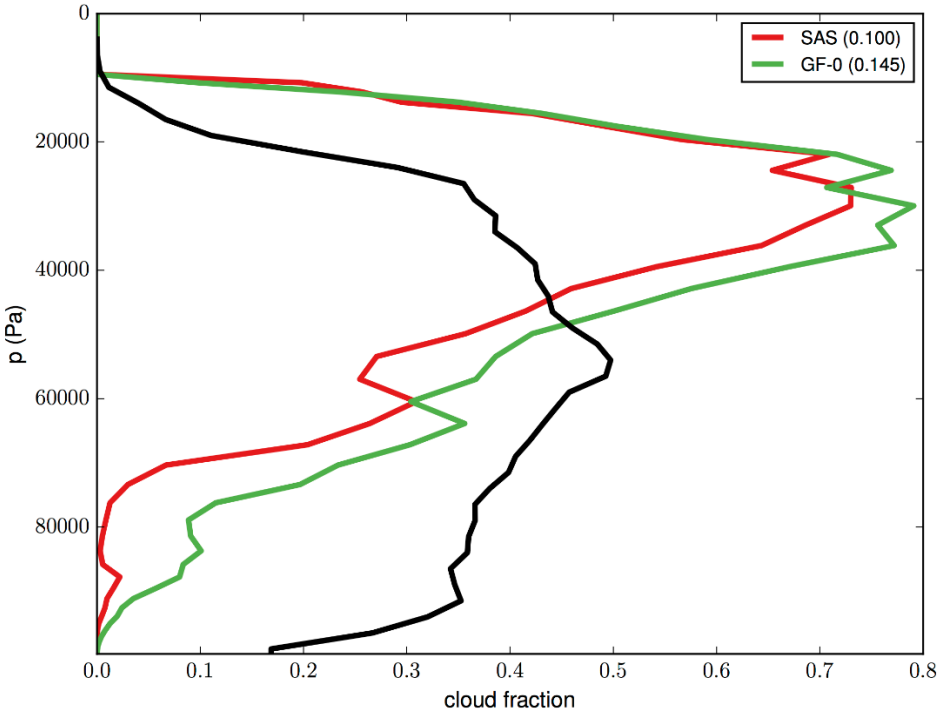



*Figure 5. Same as Fig. 4 except for cloud fraction.*

**SCM Key Finding 3:** For the suppressed convection phase of this case, the GFS-GF suite produces an elevated temperature inversion and associated steep gradients in water vapor, leading to spurious cloud generation near the boundary layer top.

Figures 6, 7, and 8 show mean profiles calculated from the suppressed convective period. As in the deep convective phase, total convective transport from the GFS-GF suite is reduced compared to GFS-GF in the shallow convective phase (tendencies not shown), leaving more moisture in the PBL compared to GFS-SAS (see Fig. 6). The "extra" moisture in this layer produces a thicker, fuller shallow cloud deck as shown in Fig. 7. The thicker PBL cloud relocates the position of maximum shortwave radiative cooling higher in the column where it is more effective (combined with PBL cooling at this level) at making a steeper temperature inversion above (see Fig. 8 at ~700 hPa).

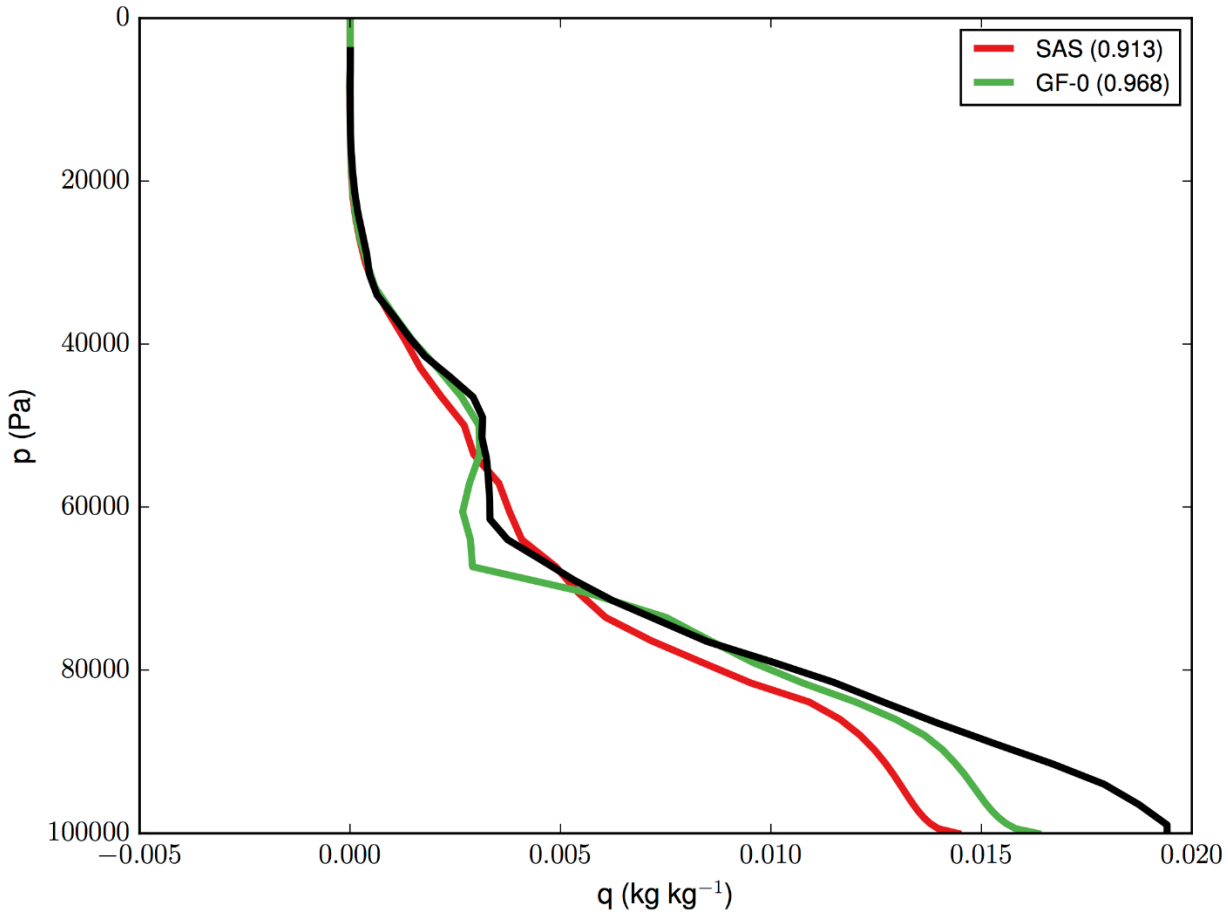



*Figure 6. Mean profiles of specific humidity (kg kg$^{-1}$) averaged over the suppressed phase of convection for the TWP-ICE case. Colors denote observations (black) or physics suite (GFS-SAS in red and GFS-GF in green). Skill scores are printed in the legend.*

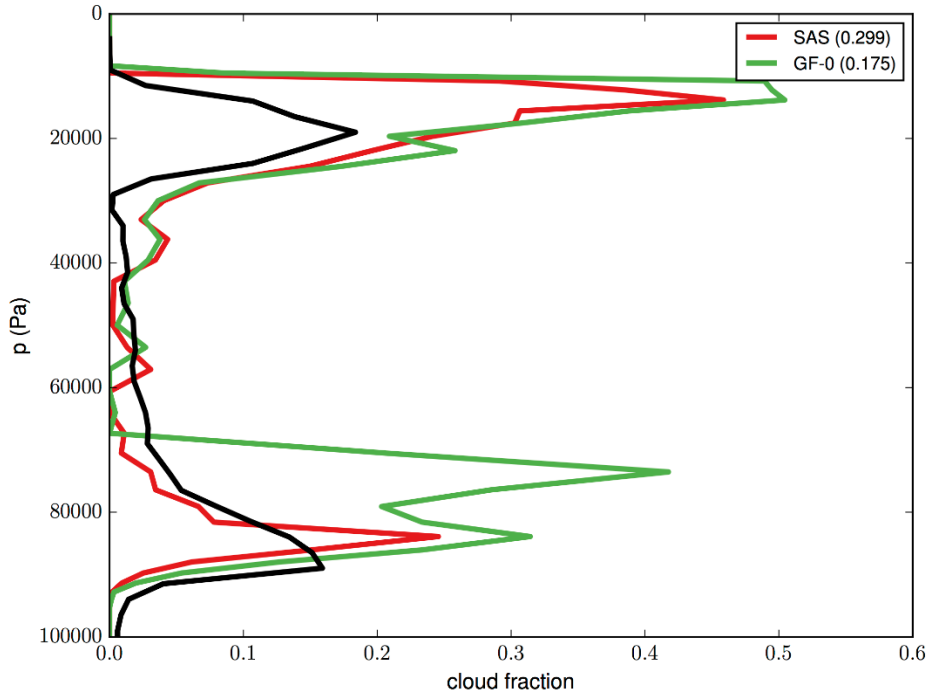



*Figure 7. Same as Fig. 6 except for cloud fraction.*

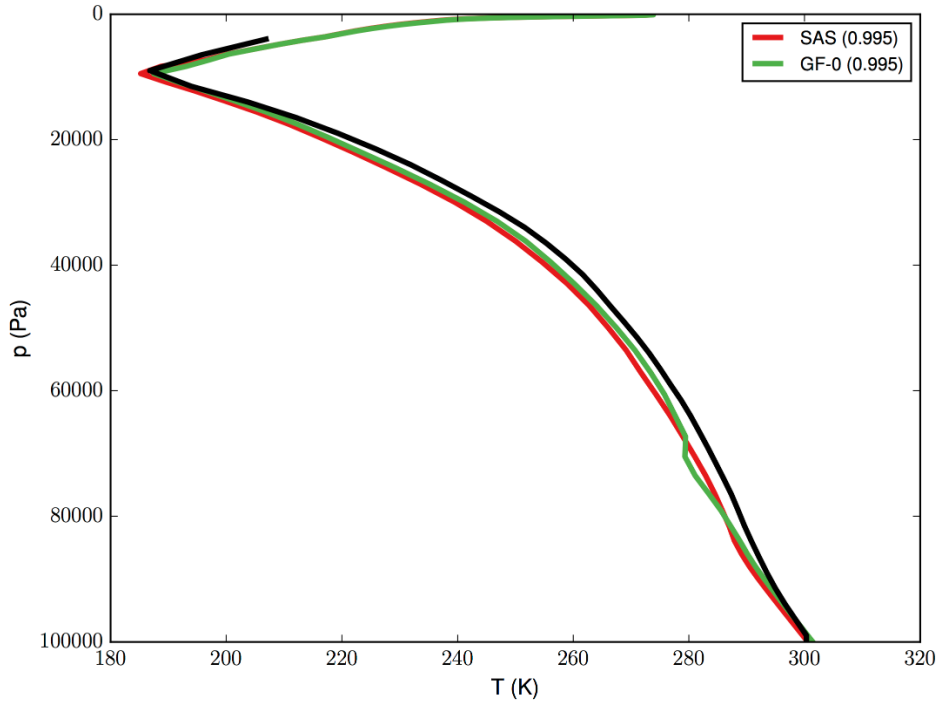



*Figure 8. Same as Fig. 6 except for temperature (K).*

**SCM Key Finding 4:** The GFS-GF suite produces a much lower convective precipitation ratio compared to the GFS-SAS suite.

Although the total precipitation produced by a single-column model is largely a function of the imposed forcing, details such as timing and partition into convective and large-scale can provide insight into the physics. Figure 9 shows the total surface-precipitation rate time series during the active phase for both suites compared to observations. While both suites are successful at reproducing the general features of the observed precipitation, the skill score for the GFS-GF suite is slightly higher than the control suite, indicating higher temporal correlation, more similar variability, or both. It appears that the GFS-GF suite causes precipitation to peak slightly after the GFS-SAS suite during the deep convective events.

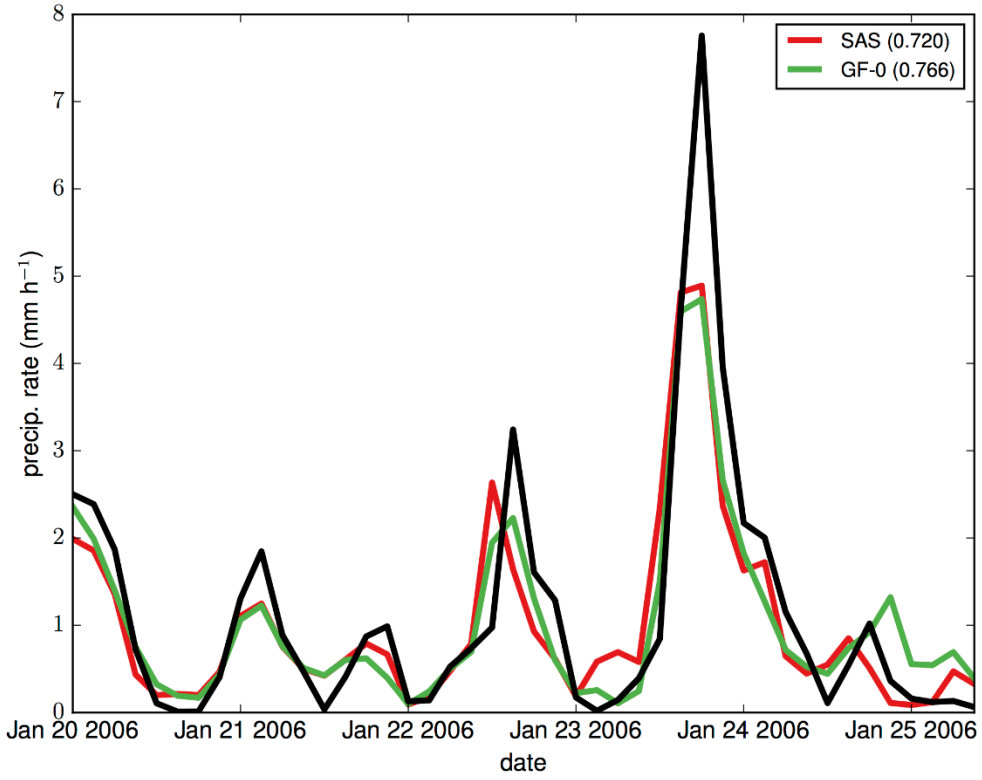



*Figure 9. Time series of total surface precipitation rate (mm h$^{-1}$) for the active phase of the TWP-ICE simulation. Colors denote observations (black) or physics suite (GFS-SAS in red and GFS-GF in green). Skill scores are printed in the legend.*

The biggest difference between the suites in terms of precipitation, however, is the partition between convective and explicit sources. Figures 10 and 11 show scatter plots of total precipitation rate versus the ratio between convective and total precipitation, with each point representing the time average for one of the 100-member forcing ensemble for the deep convective period (10) and shallow convective period (11). The total precipitation rate is a proxy for the strength of the applied forcing. We find that the convective ratio is always lower for the GFS-GF suite than for the GFS-SAS suite, and the GFS-GF exhibits a stronger relationship to the applied forcing. For the deep convective period, an apparently discrete "jump" in the ratio for the GF scheme exists where the convective ratio goes from a much lower value to a value that is much closer to the GFS-SAS suite. For the suppressed convection, only a weak relationship between forcing strength and convective ratio is evident for the GFS-SAS suite, but a nearly monotonically increasing relationship is exhibited for the GFS-GF suite. Figure 12 shows the

similar plots for the SCM intercomparison of Davies et al. (2013). The GFS physics suite from circa 2011 exhibits nearly identical behavior to the control GFS-SAS suite presented here. The GFS-GF suite appears to behave quite similarly to the GISS model found in that intercomparison. Further, Davies et al. (2013) identified two groupings of behavior, those with a relatively high convective ratio and those with a relatively low convective ratio. Switching from SAS to GF causes the GFS physics suite to switch from the "high" group to the "low" group.

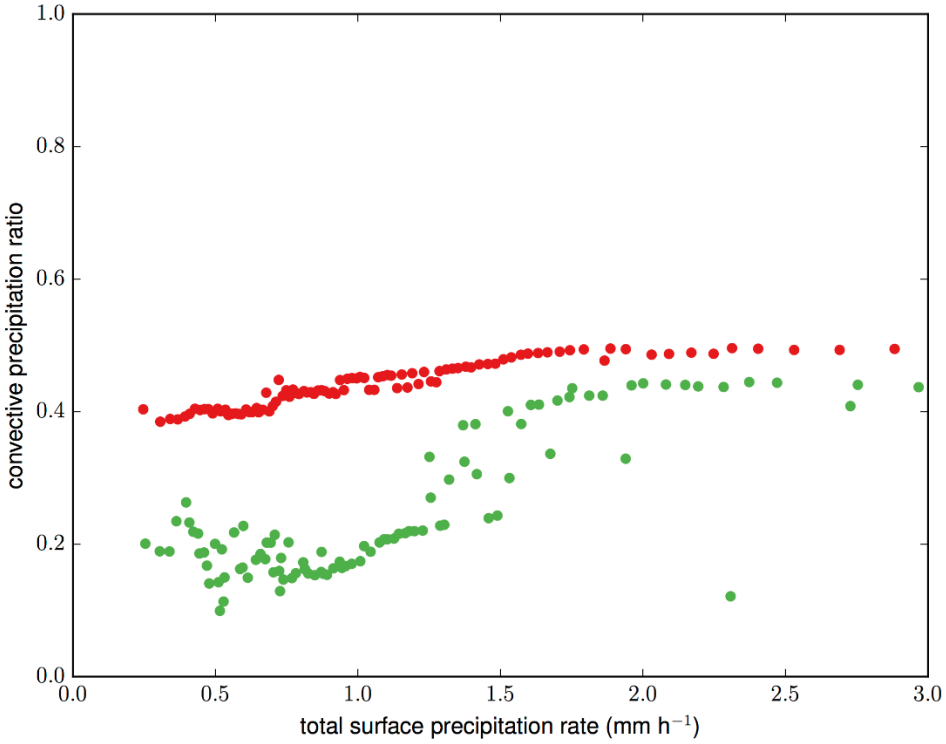



*Figure 10. Scatter plot denoting mean values of total surface precipitation rate (mm h$^{-1}$) versus convective precipitation ratio over the active phase of convection for all forcing ensemble members for GFS-GF (green) and GFS-SAS (red).*

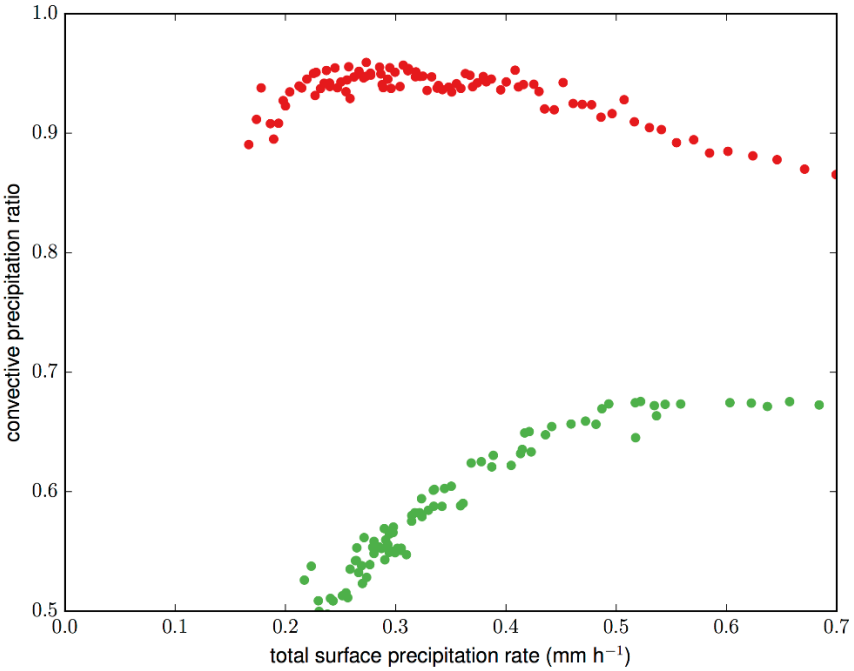



*Figure 11. Same as Fig. 10 but for the suppressed phase of convection.*

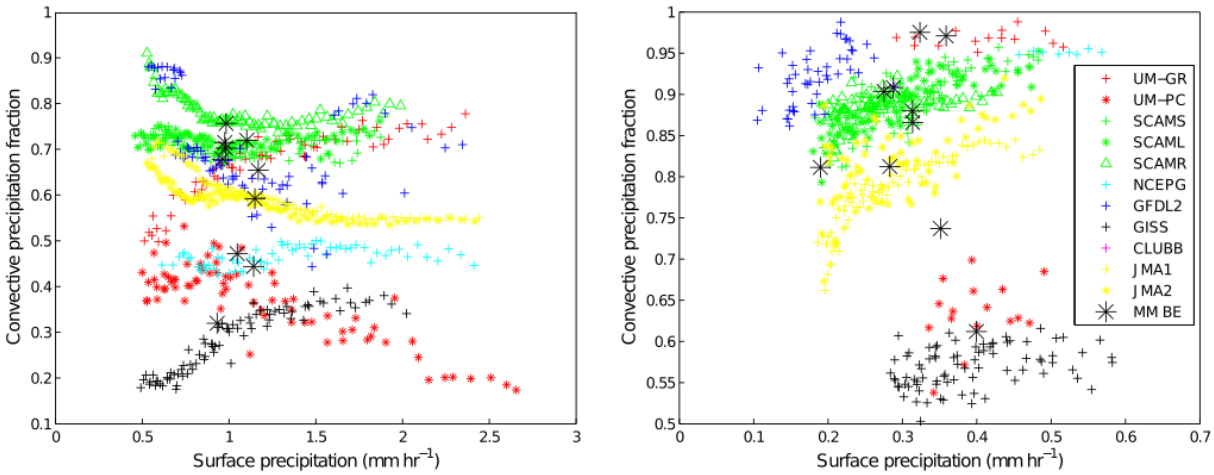



*Figure 12. Fig. 7 from Davies et al. (2013). The format and axes are the same as for figures 10 and 11. In this plot, colors denote models that participated in a SCM intercomparison project, with the GFS from circa 2011 in light blue.*

**SCM Key Finding 5:** During the deep convective period, the forcing ensemble elicits greater variability from the GFS-GF suite than the GFS-SAS suite.

Figures 13, 14, and 15 show mean profiles of specific humidity, total convective temperature tendency, and total convective moisture tendency, respectively, averaged over the deep convective phase for the forcing ensemble. These plots demonstrate that the GFS-GF suite is more sensitive than the GFS-SAS suite to the forcing. In fact, for these quantities, at most levels, the GFS-GF ensemble range contains the GFS-SAS ensemble range. For convective tendencies, the 25th percentile profiles for the GFS-SAS approximate the 50th percentile of the GFS-GF suite and the 50th percentile profile of the GFS-

SAS suite approximates the 75th percentile profile of the GFS-GF suite for much of the depth of the atmosphere. However, for more extreme ends of the forcing ensemble range, the GFS-GF suite produces more extreme profiles than the GFS-SAS suite. The different response to changes in forcing likely has consequences for stochastic applications.

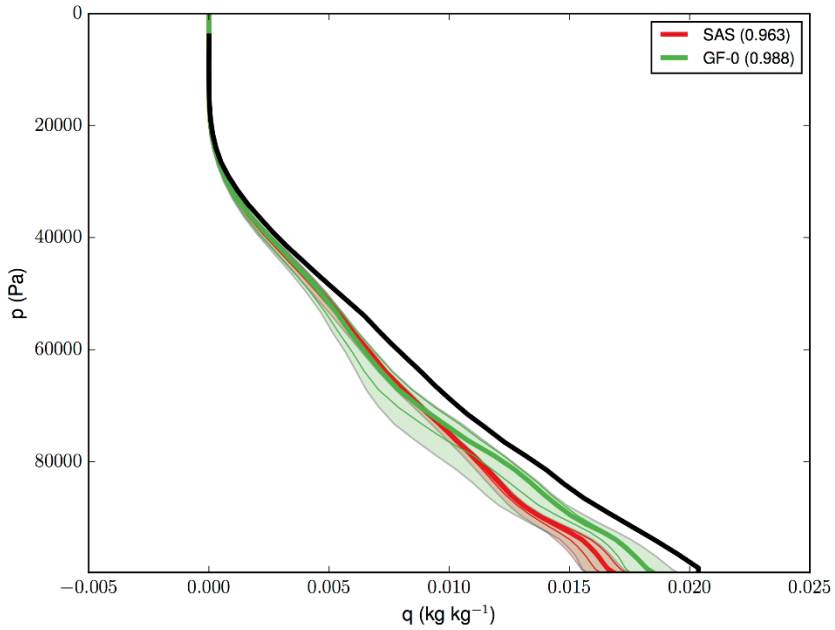



*Figure 13. Mean profiles of specific humidity (kg kg$^{-1}$) for the active convective phase of the TWP-ICE case. Colors denote observations (black) or physics suite (GFS-SAS in red and GFS-GF in green). Shading encompasses the 10-90th percentiles of the forcing ensemble. Thin lines denote the 25th and 75th percentiles. Thick lines denote the 50th percentile.*

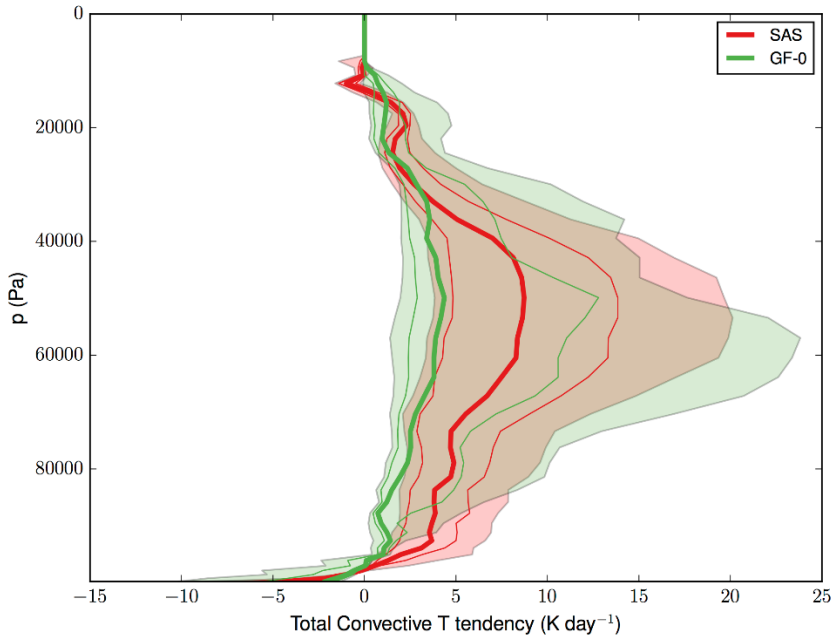



*Figure 14. Same as Fig. 13 except for the total convective temperature tendency (K day$^{-1}$).*

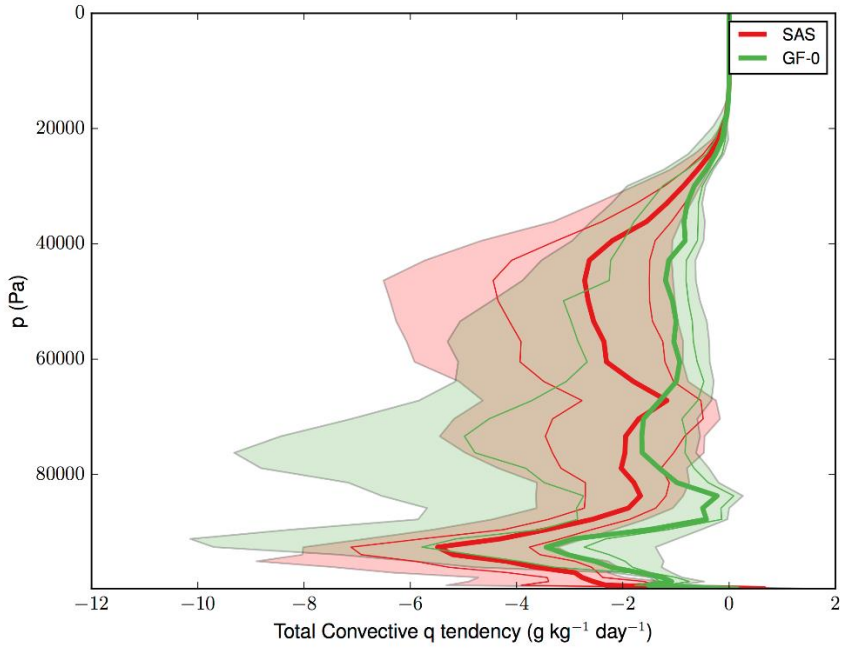



*Figure 15. Same as Fig. 13 except for the total convective moisture tendency (g kg$^{-1}$ day$^{-1}$).*

# Key Findings from Global Model

**GSM Key Finding 1:** The results of RMSE and bias comparisons vary by forecast lead time, level, and region, with GFS-SAS displaying superior forecasts in more instances than GFS-GF. In upper levels, there are more differences between GFS-SAS and GFS-GF in temperature and relative humidity RMSE than in wind RMSE. When SS pairwise differences were noted for wind speed RMSE, they nearly always favored GFS-SAS, regardless of level or region.

The scorecard for global sub-regions helps identify patterns in the difference of performance between GFS-SAS and GFS-GF, highlighting that upper-air wind speed RMSE has the fewest SS differences compared to other variables (Figs. 16-18). The SH has the most differences that are not SS; this behavior is seen throughout the forecast period and at all levels (Fig. 17). The NH clearly signals GFS-SAS as performing better in the earlier part of the forecast period, with few SS pairwise differences between the configurations after the 120-h forecast lead time (Fig. 16). In the tropical region, the shift toward fewer SS pairwise differences transitions later in the forecast period (Fig. 18). In summary, upper-level wind speed RMSE generally exhibits more similarities between the model configurations, while temperature, relative humidity, and precipitation forecasts were more significantly impacted by the cumulus parameterization selected.

## Scorecard

for gftest_0p25_G3 and sasctrl_0p25_G3

2016-06-01 00:00:00 - 2016-08-31 00:00:00

| | | | NH | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | f12 | f24 | f36 | f48 | f60 | f72 | f84 | f96 | f108 | f120 | f132 | f144 | f156 | f168 | f180 | f192 | f204 | f216 | f228 | f240 |
| ME | Temp | P100 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▲1.000 | ▼-1.000 | ▲1.000 | ▼-1.000 | ▲1.000 | ▼-1.000 | ▲1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▲1.000 | ▼-1.000 | ▲1.000 | ▼-1.000 | ▼-1.000 |
| | | P150 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-0.999 | ▾-0.995 | ▼-0.999 | -0.978 | ▾-0.992 | -0.865 | -0.934 | -0.794 | -0.919 | -0.728 | -0.693 | -0.658 | -0.865 | -0.728 | -0.829 |
| | | P200 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 |
| | | P300 | ▼-1.000 | 0.874 | 0.872 | 0.855 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▲1.000 | ▼-1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 |
| | | P400 | ▼-1.000 | ▲1.000 | ▲0.999 | ▲1.000 | 0.955 | 0.958 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 |
| | | P500 | ▼-1.000 | 0.687 | -0.982 | ▲1.000 | -0.657 | 0.936 | ▼-1.000 | ▴0.997 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 |
| | | P700 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 |
| | | P850 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 |
| | RH | P300 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▴0.998 | ▲1.000 | ▲1.000 | ▲1.000 |
| | | P400 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲0.999 | ▲1.000 | ▲1.000 | ▲1.000 | ▴0.994 | 0.983 | 0.938 | ▴0.998 | 0.920 | ▴0.992 | 0.797 | 0.755 | 0.716 | ▲0.999 |
| | | P500 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲0.999 | ▲1.000 | ▴0.993 | ▲1.000 | 0.969 | ▲0.999 | 0.950 | 0.904 | 0.664 | ▲0.999 |
| | | P700 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▲1.000 | ▼-1.000 | ▲1.000 | ▼-1.000 | ▲1.000 | ▼-1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▼-1.000 | ▲1.000 | ▼-1.000 | ▲0.999 | ▼-1.000 | ▲0.999 | ▼-1.000 | ▾-0.995 |
| | | P850 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 |
| | Wind | P100 | ▲1.000 | ▼-1.000 | ▲1.000 | ▼-1.000 | ▲1.000 | ▼-1.000 | 0.728 | ▼-1.000 | 0.546 | -0.987 | 0.766 | -0.788 | 0.930 | -0.135 | 0.736 | -0.560 | 0.900 | -0.228 | 0.832 | -0.872 |
| | | P150 | ▼-1.000 | -0.386 | ▼-1.000 | 0.843 | ▼-1.000 | ▾-0.995 | ▼-1.000 | ▼-1.000 | ▼-1.000 | -0.988 | ▼-1.000 | -0.965 | ▼-1.000 | -0.502 | ▼-1.000 | -0.520 | ▼-0.999 | -0.632 | ▼-1.000 | -0.967 |
| | | P200 | ▼-1.000 | 0.793 | ▼-1.000 | 0.948 | ▼-1.000 | -0.979 | ▼-1.000 | ▼-0.999 | ▼-1.000 | -0.885 | ▼-1.000 | -0.819 | ▼-1.000 | -0.279 | ▾-0.998 | 0.373 | -0.939 | -0.215 | ▼-1.000 | -0.427 |
| | | P300 | ▾-0.994 | ▲0.999 | 0.972 | ▲1.000 | 0.898 | ▴0.992 | -0.507 | 0.873 | -0.212 | 0.964 | 0.036 | 0.853 | 0.254 | ▴0.992 | 0.242 | ▲0.999 | 0.964 | 0.979 | -0.140 | 0.971 |
| | | P400 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▴0.998 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | 0.973 | ▲0.997 |
| | | P500 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲0.991 | ▲1.000 |
| | | P700 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲0.999 | 0.639 | 0.208 | -0.572 | 0.492 | -0.675 | 0.947 | -0.976 | -0.545 | -0.984 | -0.693 | ▾-0.990 | -0.902 | ▾-0.993 | -0.719 |
| | | P850 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | 0.128 | 0.803 | -0.961 | 0.749 | -0.966 | -0.142 | ▼-1.000 | -0.958 | ▼-1.000 | ▼-0.999 | ▼-1.000 | ▼-0.999 | ▼-1.000 | ▼-0.999 |
| RMSE | Temp | P100 | -0.976 | ▼-1.000 | ▼-1.000 | ▼-0.999 | ▼-0.999 | ▼-1.000 | ▼-1.000 | -0.695 | ▾-0.995 | -0.883 | -0.986 | -0.894 | -0.789 | -0.355 | 0.614 | 0.640 | 0.857 | 0.903 | 0.980 | 0.949 |
| | | P150 | 0.018 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | -0.806 | -0.975 | -0.814 | ▾-0.997 | -0.665 | -0.963 | -0.121 | -0.719 | 0.207 | 0.271 | 0.681 | 0.457 | 0.934 | 0.859 |
| | | P200 | 0.815 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▾-0.998 | ▾-0.996 | -0.967 | ▼-1.000 | ▼-0.999 | ▼-0.999 | -0.921 | -0.687 | 0.058 | -0.289 | -0.397 | -0.384 | -0.312 | -0.179 |
| | | P300 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▾-0.992 | ▾-0.994 | -0.913 | -0.741 | -0.757 | -0.735 | -0.112 | 0.085 | 0.452 | 0.315 | 0.720 | 0.437 |
| | | P400 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | -0.975 | -0.983 | -0.917 | -0.983 | -0.596 | -0.836 | -0.296 | -0.642 | -0.106 | -0.208 | 0.131 | -0.228 | 0.342 | 0.079 |
| | | P500 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | -0.937 | -0.941 | -0.841 | -0.977 | -0.090 | -0.501 | 0.016 | 0.180 | 0.296 | -0.071 | -0.002 | 0.189 | 0.611 | 0.740 |
| | | P700 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | -0.989 | -0.985 | -0.989 | -0.873 | -0.874 | -0.879 | -0.716 | -0.278 | -0.301 | 0.355 | 0.713 | 0.786 |
| | | P850 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-0.999 | ▼-0.999 | ▾-0.997 | ▾-0.994 | -0.882 |
| | RH | P300 | -0.487 | 0.956 | 0.604 | ▲1.000 | ▴0.993 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▴0.994 | ▲1.000 | ▴0.998 | ▲1.000 | 0.971 | ▲0.999 | ▴0.997 | ▲1.000 | ▲1.000 | ▲1.000 |
| | | P400 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▾-0.993 | ▼-1.000 | -0.101 | -0.266 | 0.803 | -0.514 | 0.843 | 0.099 | ▴0.995 | 0.558 | 0.470 | 0.867 | 0.472 | 0.291 | 0.901 |
| | | P500 | -0.849 | -0.928 | ▼-0.999 | -0.715 | -0.938 | -0.961 | -0.982 | 0.851 | 0.310 | 0.949 | 0.389 | 0.826 | 0.265 | 0.986 | 0.964 | ▴0.993 | ▴0.997 | ▲0.999 | 0.957 | 0.977 |
| | | P700 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-0.999 | -0.989 | -0.955 | ▾-0.997 | -0.958 | -0.574 | -0.968 |
| | | P850 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 |
| | Wind | P100 | -0.074 | ▼-0.999 | ▼-0.999 | ▼-0.999 | 0.530 | -0.309 | 0.330 | 0.559 | 0.886 | 0.953 | 0.783 | 0.868 | 0.728 | 0.740 | 0.766 | 0.925 | 0.810 | 0.783 | 0.272 | 0.897 |
| | | P150 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▾-0.997 | -0.950 | ▾-0.997 | -0.871 | -0.733 | 0.035 | -0.266 | 0.484 | 0.882 | 0.746 | 0.933 | 0.647 | 0.876 | 0.628 |
| | | P200 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▾-0.994 | -0.963 | -0.891 | -0.728 | 0.544 | 0.642 | 0.725 | 0.057 | 0.630 | 0.936 | 0.788 | 0.804 | 0.914 | 0.833 |
| | | P300 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▾-0.998 | -0.970 | -0.950 | ▾-0.995 | -0.956 | -0.607 | -0.188 | -0.570 | -0.708 | 0.057 | 0.730 | 0.338 | 0.730 | 0.905 | 0.173 |
| | | P400 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | -0.941 | -0.779 | -0.973 | -0.987 | ▾-0.993 | -0.362 | -0.431 | -0.717 | -0.098 | 0.242 | 0.634 | 0.685 | 0.819 | 0.654 |
| | | P500 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | -0.982 | -0.986 | -0.909 | -0.972 | -0.630 | 0.295 | 0.515 | -0.097 | 0.308 | 0.361 | 0.748 | 0.635 | 0.959 | 0.916 |
| | | P700 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-0.999 | ▼-1.000 | ▾-0.991 | -0.969 | -0.062 | 0.557 | -0.003 | -0.754 | 0.004 | 0.815 | 0.364 | 0.961 | 0.404 |
| | | P850 | ▾-0.998 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | -0.981 | -0.836 | -0.921 | -0.973 | -0.695 | 0.463 | 0.224 | 0.001 | 0.227 |

| | |
|---|---|
| ▲ | gftest_0p25_G3 is better than sasctrl_0p25_G3 at the 99.9% significance level |
| ▴ | gftest_0p25_G3 is better than sasctrl_0p25_G3 at the 99% significance level |
| | gftest_0p25_G3 is better than sasctrl_0p25_G3 at the 95% significance level |
| | No statistically significant difference between gftest_0p25_G3 and sasctrl_0p25_G3 |
| | gftest_0p25_G3 is worse than sasctrl_0p25_G3 at the 95% significance level |
| ▾ | gftest_0p25_G3 is worse than sasctrl_0p25_G3 at the 99% significance level |
| ▼ | gftest_0p25_G3 is worse than sasctrl_0p25_G3 at the 99.9% significance level |
| | Not statistically relevant |

*Figure 16. Scorecard documenting performance of GFS-SAS and GFS-GF over the NH of mean bias and RMSE for temperature, relative humidity, and wind speed by forecast lead time and vertical level for JJA 2016. Green (red) shading indicates GFS-GF (GFS-SAS) is better than GFS-SAS (GFS-GF) at the 95% significance level. Small green (red) arrows indicate GFS-GF (GFS-SAS) is better than GFS-SAS (GFS-GF) at the 99% significance level. Large green (red) arrows indicate GFS-GF (GFS-SAS) is better than GFS-SAS (GFS-GF) at the 99.9% significance level. Grey shading indicates no statistically significant differences between GFS-SAS and GFS-GF.*

## Scorecard

for gftest_0p25_G3 and sasctrl_0p25_G3

2016-06-01 00:00:00 - 2016-08-31 00:00:00

| | | | SH | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | f12 | f24 | f36 | f48 | f60 | f72 | f84 | f96 | f108 | f120 | f132 | f144 | f156 | f168 | f180 | f192 | f204 | f216 | f228 | f240 |
| ME | Temp | P100 | ▼-1.000 | ▲1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | -0.985 | ▼-1.000 | -0.154 | ▼-1.000 | -0.964 | ▼-1.000 | -0.802 | ▼-1.000 | -0.720 | ▼-1.000 | 0.117 | -0.939 | 0.167 | ▼-1.000 |
| | | P150 | ▼-1.000 | ▼-1.000 | -0.757 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-0.999 | ▾-0.998 | -0.820 | -0.903 | ▼-1.000 | -0.602 | ▼-1.000 | -0.899 | ▼-1.000 | -0.629 | -0.822 | 0.118 | ▼-1.000 | -0.953 |
| | | P200 | ▼-0.999 | ▼-1.000 | -0.981 | ▼-1.000 | 0.691 | ▾-0.995 | -0.225 | -0.964 | -0.966 | -0.969 | -0.839 | 0.103 | -0.966 | 0.086 | -0.911 | 0.918 | -0.068 | ▲0.997 | -0.966 | 0.948 |
| | | P300 | ▼-1.000 | ▼-1.000 | -0.909 | 0.989 | 0.975 | 0.897 | 0.675 | 0.903 | -0.861 | 0.497 | 0.649 | ▲0.999 | 0.683 | ▲1.000 | 0.983 | ▲1.000 | 0.029 | ▲0.997 | 0.896 | ▲1.000 |
| | | P400 | ▼-1.000 | ▼-1.000 | 0.636 | ▼-1.000 | ▾-0.994 | ▼-0.999 | 0.450 | ▾-0.997 | 0.616 | -0.853 | -0.958 | ▼-1.000 | -0.970 | ▼-1.000 | ▼-1.000 | ▼-1.000 | -0.122 | ▼-1.000 | -0.979 | ▼-1.000 |
| | | P500 | 0.684 | -0.819 | ▼-1.000 | ▲1.000 | ▼-1.000 | ▲1.000 | -0.278 | 0.984 | 0.340 | ▼-1.000 | -0.928 | ▼-1.000 | ▾-0.997 | ▼-1.000 | ▼-0.999 | ▼-1.000 | -0.437 | ▼-1.000 | ▾-0.991 | ▼-1.000 |
| | | P700 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-0.999 | ▼-1.000 | ▼-1.000 | ▾-0.998 | ▼-1.000 | 0.026 | -0.982 | -0.331 | -0.085 | 0.980 | 0.594 | 0.932 | -0.906 | -0.286 | -0.120 | 0.750 |
| | | P850 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▾-0.992 | ▼-0.999 | -0.951 | ▼-1.000 | -0.784 | ▼-0.999 |
| | RH | P300 | ▲1.000 | -0.567 | ▲0.999 | 0.969 | ▲1.000 | 0.907 | ▲0.999 | ▲0.993 | 0.930 | ▲0.996 | 0.968 | 0.970 | 0.982 | 0.914 | ▲1.000 | ▲1.000 | ▲0.997 | 0.987 | 0.984 | ▲0.995 |
| | | P400 | ▲1.000 | 0.525 | ▲1.000 | 0.768 | ▲1.000 | 0.659 | ▲0.998 | 0.957 | ▲0.999 | 0.945 | ▲0.999 | ▲0.992 | ▲0.991 | 0.925 | ▲0.990 | 0.980 | ▲1.000 | ▲0.999 | ▲1.000 | ▲1.000 |
| | | P500 | ▲1.000 | 0.879 | ▲1.000 | 0.476 | ▲1.000 | 0.708 | 0.981 | 0.859 | ▲1.000 | 0.978 | ▲0.991 | 0.440 | ▲1.000 | 0.911 | ▲0.994 | 0.952 | 0.980 | 0.987 | ▲1.000 | ▲1.000 |
| | | P700 | -0.866 | -0.975 | -0.559 | -0.547 | 0.601 | ▾-0.993 | -0.673 | -0.850 | -0.461 | -0.242 | 0.674 | -0.750 | ▲1.000 | 0.673 | 0.108 | -0.900 | -0.963 | -0.861 | 0.975 | 0.894 |
| | | P850 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | -0.987 | ▼-1.000 | ▾-0.998 | ▼-1.000 | ▼-1.000 | ▼-1.000 | -0.984 | ▼-1.000 |
| | Wind | P100 | ▼-1.000 | ▼-1.000 | 0.974 | 0.566 | 0.974 | -0.269 | ▼-1.000 | 0.700 | -0.988 | 0.810 | 0.656 | 0.986 | 0.959 | ▲0.991 | 0.973 | ▲0.999 | ▲0.999 | ▲1.000 | ▲0.999 | ▲1.000 |
| | | P150 | ▼-1.000 | ▼-1.000 | 0.622 | ▲0.990 | ▲1.000 | ▲0.999 | ▲0.997 | -0.162 | 0.918 | 0.777 | ▲0.998 | 0.548 | ▲1.000 | 0.845 | ▲0.998 | ▲0.993 | ▲0.999 | ▲0.998 | ▲0.999 | ▲0.994 |
| | | P200 | 0.829 | ▼-1.000 | 0.213 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | 0.719 | ▲0.999 | 0.470 | ▲1.000 | ▲0.992 | ▲1.000 | 0.985 | 0.975 | 0.875 | ▲0.997 | ▲0.997 | ▲0.999 | 0.987 |
| | | P300 | 0.688 | ▼-1.000 | 0.986 | ▲1.000 | ▲1.000 | 0.854 | ▲0.990 | 0.948 | 0.981 | 0.976 | ▲1.000 | 0.851 | ▲1.000 | 0.966 | 0.864 | 0.985 | ▲0.997 | ▲0.994 | ▲0.998 | 0.967 |
| | | P400 | -0.970 | ▼-1.000 | 0.073 | ▲1.000 | ▲0.999 | 0.671 | 0.938 | 0.950 | 0.701 | 0.657 | 0.842 | 0.760 | ▲0.999 | 0.930 | 0.845 | 0.799 | 0.872 | ▲0.991 | 0.935 | 0.808 |
| | | P500 | -0.507 | -0.965 | -0.191 | 0.143 | 0.983 | 0.027 | 0.966 | 0.983 | 0.132 | 0.343 | 0.385 | 0.863 | 0.983 | 0.899 | 0.521 | 0.223 | 0.145 | 0.865 | 0.795 | 0.775 |
| | | P700 | 0.976 | ▲1.000 | 0.821 | 0.333 | 0.741 | 0.960 | 0.984 | 0.974 | 0.235 | -0.025 | 0.966 | 0.793 | 0.813 | 0.858 | 0.430 | 0.613 | 0.284 | 0.484 | -0.069 | 0.458 |
| | | P850 | ▲1.000 | ▲1.000 | 0.878 | ▲1.000 | ▲0.997 | ▲1.000 | ▲1.000 | ▲1.000 | 0.964 | 0.964 | 0.976 | 0.526 | 0.961 | 0.825 | 0.003 | 0.796 | -0.265 | 0.469 | 0.616 | 0.619 |
| RMSE | Temp | P100 | -0.981 | -0.599 | ▼-1.000 | -0.571 | ▼-0.999 | -0.987 | ▼-1.000 | -0.691 | 0.207 | -0.865 | -0.936 | ▾-0.995 | -0.988 | ▾-0.998 | -0.971 | -0.382 | 0.088 | 0.962 | 0.304 | -0.301 |
| | | P150 | -0.981 | -0.932 | ▼-1.000 | ▾-0.997 | ▼-1.000 | ▼-1.000 | ▾-0.998 | ▼-1.000 | -0.944 | -0.912 | -0.988 | ▼-0.999 | ▼-1.000 | ▼-1.000 | ▼-1.000 | -0.937 | -0.889 | 0.715 | ▾-0.994 | -0.770 |
| | | P200 | ▼-0.999 | ▼-1.000 | ▼-0.999 | ▾-0.998 | -0.924 | -0.980 | -0.892 | -0.931 | -0.941 | -0.362 | -0.927 | -0.867 | -0.971 | -0.988 | -0.959 | 0.720 | 0.092 | 0.961 | -0.907 | 0.968 |
| | | P300 | -0.964 | -0.959 | -0.988 | ▾-0.996 | -0.980 | -0.911 | -0.908 | -0.943 | 0.431 | -0.937 | -0.076 | -0.905 | -0.905 | -0.982 | -0.893 | -0.219 | 0.118 | 0.985 | ▲0.991 | 0.318 |
| | | P400 | ▾-0.996 | 0.052 | -0.967 | ▼-1.000 | ▾-0.994 | -0.964 | -0.738 | -0.933 | 0.286 | -0.748 | -0.274 | ▾-0.998 | -0.954 | ▼-1.000 | -0.944 | -0.931 | 0.041 | 0.245 | 0.864 | -0.707 |
| | | P500 | -0.377 | -0.984 | -0.710 | -0.984 | ▾-0.992 | -0.439 | -0.957 | -0.922 | -0.826 | -0.115 | -0.188 | -0.972 | -0.832 | ▼-0.999 | -0.851 | -0.884 | -0.742 | -0.302 | -0.171 | -0.274 |
| | | P700 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | -0.969 | ▾-0.997 | -0.900 | -0.985 | -0.615 | -0.961 | -0.989 | -0.920 | -0.932 | 0.111 | 0.087 | -0.400 |
| | | P850 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | -0.972 | ▼-1.000 | -0.933 | ▼-1.000 | -0.974 | ▾-0.992 | ▾-0.990 | ▾-0.992 | -0.284 | -0.656 |
| | RH | P300 | 0.963 | 0.197 | 0.907 | 0.961 | ▲1.000 | 0.979 | 0.893 | 0.976 | 0.659 | 0.947 | 0.856 | 0.777 | ▲0.997 | 0.989 | ▲0.998 | 0.934 | ▲0.999 | 0.967 | ▲1.000 | 0.953 |
| | | P400 | 0.110 | -0.915 | 0.284 | -0.386 | 0.309 | -0.938 | 0.652 | 0.610 | 0.624 | 0.849 | 0.620 | 0.581 | 0.141 | 0.679 | -0.260 | 0.140 | ▲0.996 | -0.226 | ▲0.990 | 0.607 |
| | | P500 | 0.390 | ▾-0.992 | 0.543 | -0.899 | 0.156 | -0.986 | -0.836 | -0.147 | 0.387 | -0.821 | 0.366 | -0.646 | 0.630 | 0.637 | 0.290 | 0.457 | 0.902 | -0.161 | 0.948 | 0.419 |
| | | P700 | -0.694 | ▾-0.993 | ▾-0.995 | ▼-1.000 | -0.890 | ▾-0.991 | -0.981 | -0.983 | -0.463 | -0.792 | -0.698 | -0.954 | 0.481 | -0.972 | 0.211 | -0.266 | -0.987 | -0.259 | -0.356 | -0.121 |
| | | P850 | ▾-0.994 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-0.999 | ▼-1.000 | -0.866 | ▼-1.000 | -0.747 | ▾-0.990 | 0.186 | -0.979 | 0.795 | ▼-0.999 | -0.013 | -0.883 | -0.938 | -0.636 | -0.891 | ▾-0.994 |
| | Wind | P100 | -0.709 | -0.834 | -0.840 | -0.868 | -0.977 | ▾-0.997 | ▼-1.000 | -0.978 | -0.908 | -0.978 | -0.485 | -0.846 | -0.857 | -0.851 | 0.415 | 0.236 | 0.372 | 0.540 | -0.096 | -0.070 |
| | | P150 | -0.844 | ▾-0.997 | -0.865 | -0.943 | -0.960 | -0.956 | -0.762 | -0.692 | 0.018 | -0.750 | -0.217 | -0.856 | -0.978 | -0.955 | -0.474 | 0.253 | 0.682 | 0.865 | 0.841 | 0.461 |
| | | P200 | -0.973 | ▼-1.000 | -0.893 | -0.410 | -0.034 | -0.932 | 0.251 | -0.926 | 0.379 | -0.612 | 0.666 | -0.880 | -0.120 | -0.986 | 0.334 | 0.586 | 0.821 | 0.910 | 0.696 | 0.190 |
| | | P300 | ▼-0.999 | ▾-0.998 | -0.842 | -0.051 | -0.595 | ▾-0.992 | ▾-0.992 | -0.517 | -0.151 | -0.621 | 0.479 | -0.847 | -0.081 | -0.983 | -0.099 | 0.550 | 0.341 | 0.290 | -0.207 | -0.247 |
| | | P400 | ▾-0.991 | ▼-0.999 | ▾-0.993 | -0.939 | -0.700 | -0.777 | -0.938 | -0.879 | -0.156 | -0.418 | -0.420 | -0.717 | -0.979 | -0.911 | -0.579 | 0.297 | -0.565 | -0.027 | -0.514 | 0.461 |
| | | P500 | ▼-0.999 | ▼-1.000 | -0.970 | -0.955 | -0.911 | -0.896 | -0.930 | -0.597 | -0.589 | 0.591 | -0.103 | -0.499 | ▾-0.994 | -0.957 | -0.674 | -0.334 | -0.796 | 0.019 | -0.876 | 0.834 |
| | | P700 | ▼-1.000 | -0.920 | -0.970 | ▼-1.000 | -0.963 | ▼-1.000 | ▾-0.998 | -0.880 | -0.084 | -0.804 | -0.959 | ▾-0.995 | ▾-0.992 | -0.965 | -0.950 | -0.963 | -0.968 | -0.474 | -0.537 | 0.932 |
| | | P850 | -0.754 | -0.397 | ▾-0.995 | ▾-0.996 | ▾-0.993 | ▼-0.999 | -0.498 | -0.891 | 0.793 | -0.574 | -0.987 | -0.614 | ▾-0.996 | -0.966 | -0.823 | -0.831 | -0.401 | -0.172 | -0.928 | 0.430 |

| | |
|---|---|
| ▲ | gftest_0p25_G3 is better than sasctrl_0p25_G3 at the 99.9% significance level |
| ▴ | gftest_0p25_G3 is better than sasctrl_0p25_G3 at the 99% significance level |
| | gftest_0p25_G3 is better than sasctrl_0p25_G3 at the 95% significance level |
| | No statistically significant difference between gftest_0p25_G3 and sasctrl_0p25_G3 |
| | gftest_0p25_G3 is worse than sasctrl_0p25_G3 at the 95% significance level |
| ▾ | gftest_0p25_G3 is worse than sasctrl_0p25_G3 at the 99% significance level |
| ▼ | gftest_0p25_G3 is worse than sasctrl_0p25_G3 at the 99.9% significance level |
| | Not statistically relevant |

*Figure 17. Same as Fig. 16, except for SH.*

Scorecard

for gftest_0p25_G3 and sasctrl_0p25_G3

2016-06-01 00:00:00 - 2016-08-31 00:00:00

| | | | TROP | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | f12 | f24 | f36 | f48 | f60 | f72 | f84 | f96 | f108 | f120 | f132 | f144 | f156 | f168 | f180 | f192 | f204 | f216 | f228 | f240 |
| ME | Temp | P100 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | -0.715 | ▾-0.992 | 0.948 | 0.963 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 |
| | | P150 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 |
| | | P200 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 |
| | | P300 | ▼-1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 |
| | | P400 | ▼-1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▼-1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 |
| | | P500 | ▼-1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 |
| | | P700 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | -0.985 | ▾-0.998 | -0.634 | -0.989 | 0.212 | -0.396 | 0.732 | 0.164 | 0.987 | ▲0.993 | ▲1.000 |
| | | P850 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 |
| | RH | P300 | 0.927 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲0.998 |
| | | P400 | ▼-1.000 | -0.562 | 0.864 | 0.338 | ▲0.999 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲0.998 | ▲1.000 | 0.972 | ▲1.000 | 0.940 |
| | | P500 | 0.463 | ▲1.000 | ▲1.000 | ▲0.991 | ▲1.000 | ▲0.999 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲0.990 | ▲0.998 | 0.950 |
| | | P700 | ▼-1.000 | ▼-1.000 | ▲0.994 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 |
| | | P850 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 |
| | Wind | P100 | ▲1.000 | -0.940 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 |
| | | P150 | ▲1.000 | ▲1.000 | ▲1.000 | 0.937 | -0.976 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 |
| | | P200 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 |
| | | P300 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 |
| | | P400 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 |
| | | P500 | ▲0.999 | ▲0.999 | 0.497 | 0.742 | 0.857 | 0.673 | 0.285 | -0.654 | -0.980 | -0.961 | -0.981 | -0.966 | ▼-1.000 | ▾-0.998 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 |
| | | P700 | ▾-0.998 | ▲1.000 | ▲1.000 | 0.743 | -0.684 | ▾-0.998 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-0.999 |
| | | P850 | ▼-0.999 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | -0.100 | 0.115 | -0.988 | -0.660 | ▼-1.000 | -0.961 | ▼-1.000 | -0.952 | ▼-1.000 | -0.391 | ▾-0.997 | -0.580 | ▾-0.992 | 0.267 | -0.949 |
| RMSE | Temp | P100 | ▼-1.000 | -0.978 | ▼-1.000 | ▼-1.000 | ▾-0.993 | ▼-1.000 | -0.742 | -0.871 | 0.876 | -0.831 | 0.739 | -0.485 | 0.961 | 0.035 | 0.969 | 0.089 | 0.915 | -0.023 | 0.919 | -0.268 |
| | | P150 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 |
| | | P200 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▾-0.997 | ▼-1.000 | -0.841 |
| | | P300 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-0.999 | -0.004 | -0.269 | 0.974 | 0.951 | ▲0.999 | ▲0.999 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 |
| | | P400 | ▼-1.000 | -0.988 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▾-0.997 | ▼-0.999 | -0.279 | -0.454 | 0.739 | 0.043 | ▲0.999 | 0.789 | ▲0.998 | 0.740 | ▲0.999 | 0.610 | ▲0.999 |
| | | P500 | ▼-1.000 | ▼-1.000 | -0.673 | ▼-1.000 | -0.812 | ▼-1.000 | -0.661 | ▼-1.000 | 0.198 | -0.850 | 0.688 | 0.662 | ▲0.993 | ▲0.991 | 0.983 | ▲0.996 | ▲0.998 | ▲0.999 | ▲0.999 | ▲1.000 |
| | | P700 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-0.999 |
| | | P850 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 |
| | RH | P300 | ▾-0.998 | -0.977 | 0.553 | 0.874 | -0.027 | 0.926 | 0.989 | ▲1.000 | 0.968 | ▲1.000 | ▲0.999 | ▲0.995 | 0.984 | ▲0.996 | ▲0.998 | 0.979 | ▲0.999 | 0.961 | ▲0.998 | 0.407 |
| | | P400 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | -0.961 | -0.934 | -0.972 | -0.923 |
| | | P500 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | -0.981 | ▾-0.996 | ▾-0.995 | -0.920 | -0.956 | -0.896 | -0.989 |
| | | P700 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▾-0.998 | ▼-0.999 | ▾-0.998 | ▼-1.000 | ▾-0.998 | ▾-0.998 | ▾-0.993 | ▾-0.993 |
| | | P850 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 |
| | Wind | P100 | ▲0.992 | 0.933 | -0.692 | 0.336 | -0.703 | 0.089 | -0.861 | -0.803 | -0.511 | -0.564 | 0.232 | 0.573 | 0.780 | ▲0.992 | ▲0.994 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 | ▲1.000 |
| | | P150 | ▲1.000 | ▲1.000 | -0.610 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-0.999 | -0.933 | -0.773 | -0.815 | -0.312 | -0.345 | 0.610 | 0.633 |
| | | P200 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-0.999 | ▾-0.994 | -0.965 | -0.933 | -0.518 | -0.292 |
| | | P300 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | -0.986 | -0.978 | -0.924 | -0.405 | -0.237 | -0.354 |
| | | P400 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-0.999 | ▼-1.000 | ▼-1.000 | ▾-0.998 | -0.919 | -0.968 | -0.874 | -0.825 | -0.627 | -0.519 | -0.310 | -0.520 | -0.114 | -0.648 |
| | | P500 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▾-0.996 | ▼-1.000 | -0.987 | -0.842 | -0.943 | -0.866 | -0.906 | -0.817 | -0.277 | -0.780 |
| | | P700 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▾-0.998 | ▼-0.999 | -0.630 | -0.638 | 0.294 | -0.751 |
| | | P850 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | ▼-1.000 | -0.987 | ▼-1.000 | ▾-0.998 | -0.923 | -0.432 | -0.778 | -0.172 | -0.878 |

| | |
|---|---|
| ▲ | gftest_0p25_G3 is better than sasctrl_0p25_G3 at the 99.9% significance level |
| ▴ | gftest_0p25_G3 is better than sasctrl_0p25_G3 at the 99% significance level |
| | gftest_0p25_G3 is better than sasctrl_0p25_G3 at the 95% significance level |
| | No statistically significant difference between gftest_0p25_G3 and sasctrl_0p25_G3 |
| | gftest_0p25_G3 is worse than sasctrl_0p25_G3 at the 95% significance level |
| ▾ | gftest_0p25_G3 is worse than sasctrl_0p25_G3 at the 99% significance level |
| ▼ | gftest_0p25_G3 is worse than sasctrl_0p25_G3 at the 99.9% significance level |
| | Not statistically relevant |

*Figure 18. Same as Fig. 16, except for TROP.*

**GSM Key Finding 2:** Upper-air temperature and relative humidity RMSE values are generally larger for GFS-GF than GFS-SAS but the favored configuration depends on forecast lead time and vertical level. The advantage of GFS-SAS over GFS-GF is larger and more frequent earlier in the forecast; as forecast lead time progresses, the gap in performance narrows and GFS-GF is superior to GFS-SAS for some levels, lead times, and regions. This suggests that the GF scheme may not be in balance with the initial conditions used in this test (operational GFS analyses), and that the GF might perform better in a cycled experiment.

When comparing upper-air temperature RMSE for GFS-GF and GFS-SAS, the superior model varies with forecast lead time and vertical level (Fig. 19). In general, both models exhibit maxima in RMSE values at 850 hPa (with exception of TROP) and in the upper-levels (200 hPa for NH and SH, 150 hPa for CONUS, and 100 hPa for TROP). The shape of the vertical profiles for RMSE remains fairly consistent regardless of forecast lead time for both configurations. When looking at upper-air relative humidity, regardless of region, the smallest RMSE values are noted at 850 hPa and, except for certain levels in SH, generally increase with height (Fig. 20).

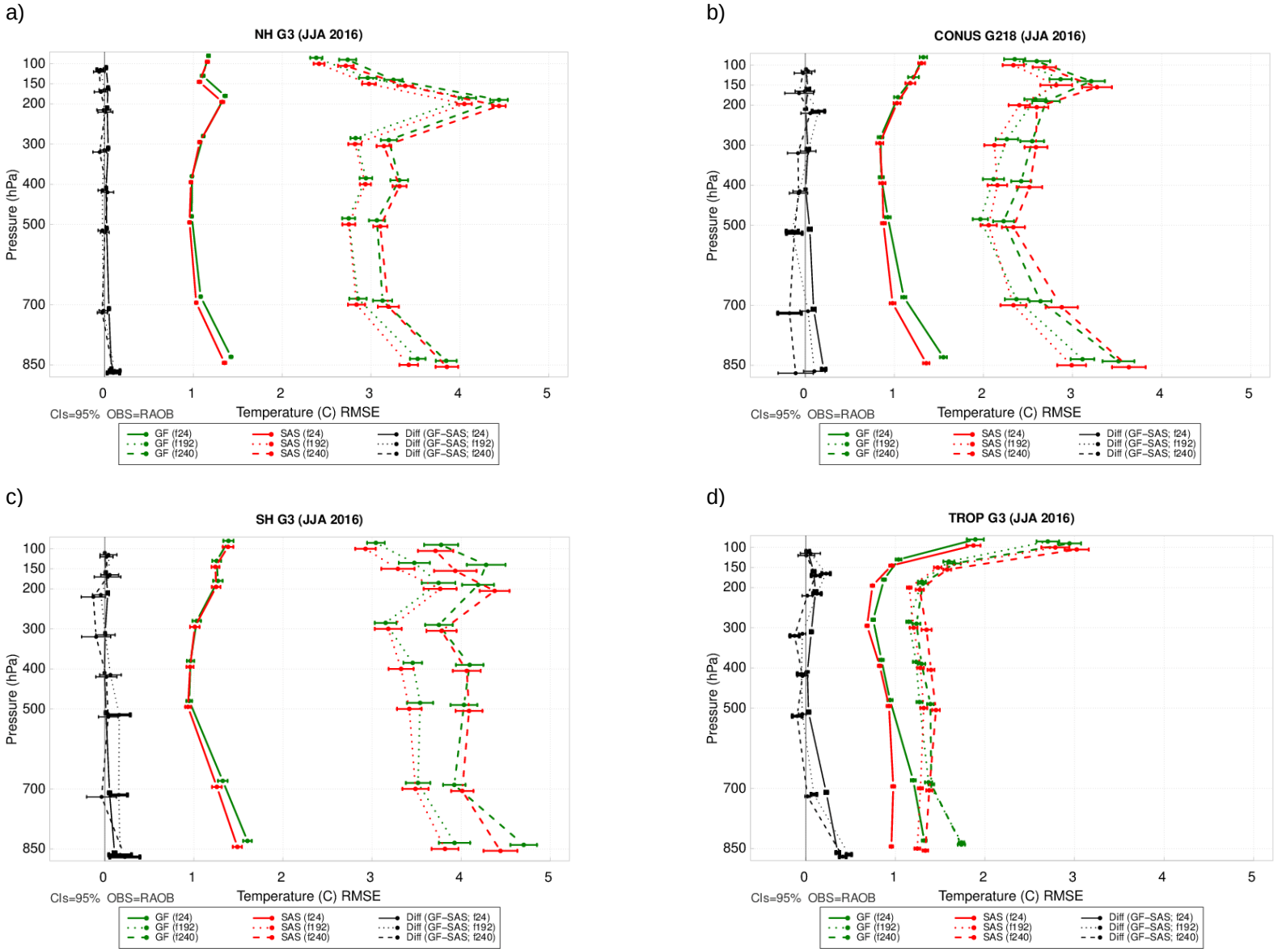



*Figure 19. Vertical profile of the median RMSE for temperature (°C) aggregated for JJA 2016 over the a) NH, b) CONUS, c) SH, and d) TROP regions. The 24-h forecast lead time is represented by the solid lines, the 192-h forecast lead time is dotted, and the 240-h forecast lead time is dashed. GFS-GF is green, GFS-SAS is red, and the differences (GFS-GF-GFS-SAS) is black. The horizontal bars surrounding the aggregate value represent the 95% CIs.*

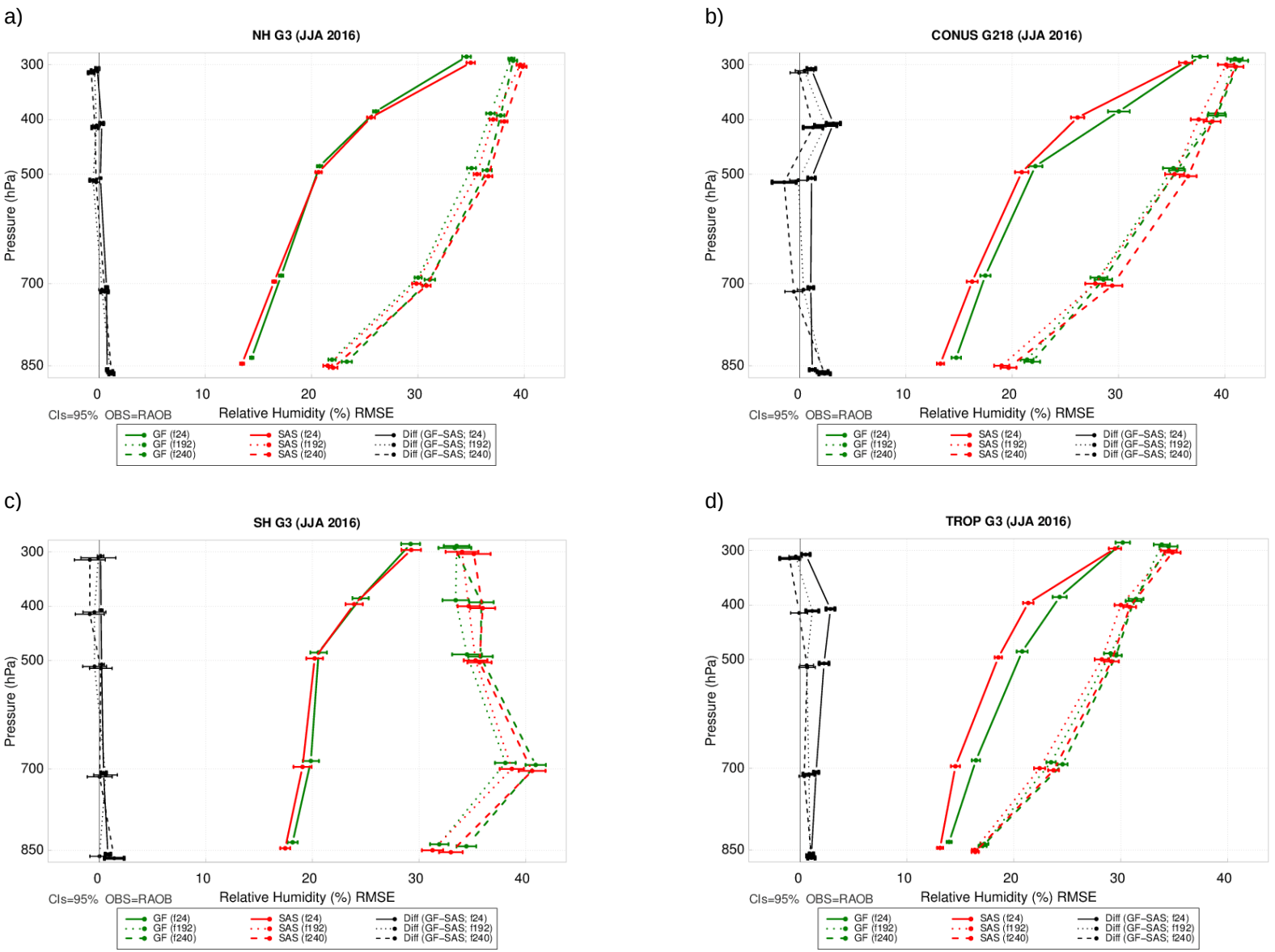



*Figure 20. Same as Fig. 19, except for median RMSE of relative humidity *RH).*

In summary, upper-air temperature and relative humidity RMSE values are generally larger for GFS-GF than GFS-SAS but the favored configuration depends on forecast lead time and vertical level. The advantage of GFS-SAS over GFS-GF is larger and more frequent earlier in the forecast. As forecast lead time progresses, the gap in performance narrows and GFS-GF is superior to GFS-SAS for some levels, lead times, and regions. This finding is also substantiated in the scorecards presented in Key Finding 1 (see Figs. 16-18).

GSM Key Finding 3: A pronounced diurnal cycle in 2-m temperature bias is clear for both the GFS-GF and the GFS-SAS configurations. The GFS-SAS warms progressively through the forecast period over CONUS throughout the troposphere and at the surface, and gets colder in the tropics. The diurnal GFS-GF bias amplitude grows with forecast lead time.

The bias in 2-m temperature (Fig. 21) shows a pronounced diurnal cycle in both model configurations. Over the CONUS, GFS-SAS exhibits warm bias at 0600, 1200, and 1800 UTC, and cold bias at 0000 UTC over the entire forecast period. The GFS-GF shows similar diurnal variation except that cold bias occurs at both 1800, and 0000 UTC. Both the bias and the comparison of mean 2-m temperature between GFS-SAS and GFS-GF against METAR observations (Fig. 22) show that the GFS-SAS gets progressively warmer at the surface with time. This trend is not seen in the GFS-GF, which however shows an increase in the amplitude of the diurnal cycle in the first eight days of forecast, followed by a stabilization of the amplitude.

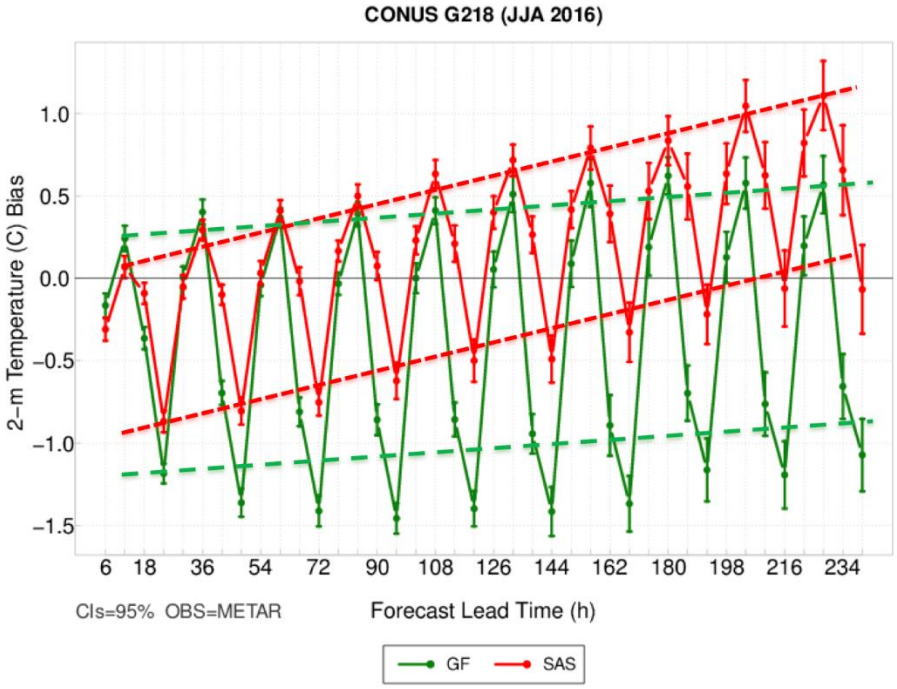



*Figure 21. Two-meter temperature bias (°C) over the CONUS domain versus forecast lead time (h) for GFS-SAS (red) and GFS-GF (green) for JJA 2016. The short-dashed red lines show the warming trend in GFS-SAS, while the long dashed green lines indicate the relatively stable bias in GFS-GF.*

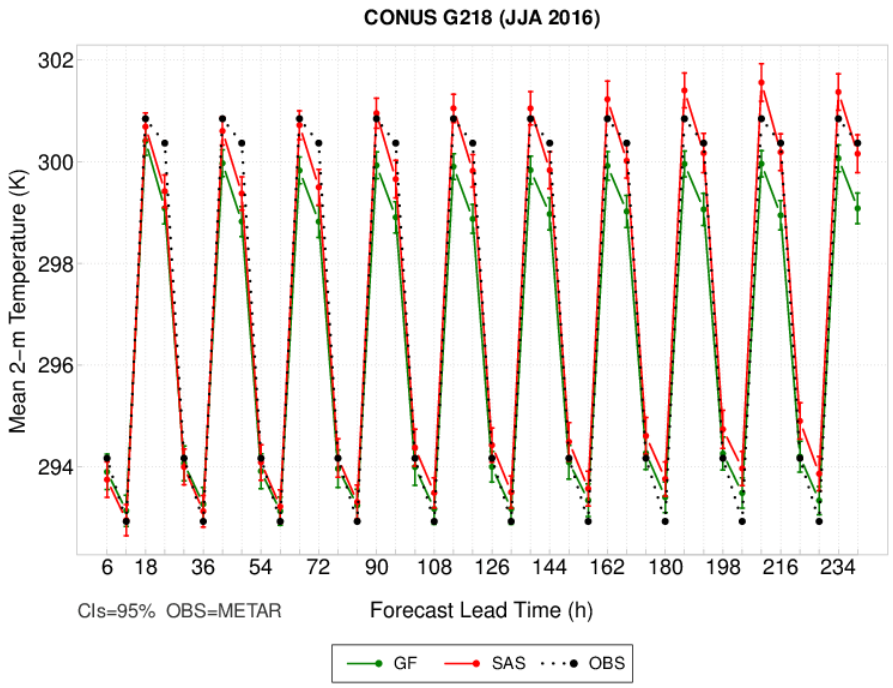



*Figure 22. Mean 2-m temperature (K) for GFS-SAS (red), GFS-GF (green), and METAR observations versus forecast lead time (h) for JJA 2016.*

In general, when looking at temperature bias profiles at the shorter forecast lead times, both GFS-GF and GFS-SAS have statistically significant (SS) cold bias with only a few levels for select regions exhibiting neutral or SS warm bias (Fig. 23). When SS pairwise differences occur for these shorter lead times, GFS-SAS is always superior (less bias) at the lowest levels (700 and 850 hPa). While additional SS pairwise differences can be seen shorter lead times at the mid- and upper-levels, the preferred model depends on a variety of factors, including level, lead time, and region. As forecast lead time increases, more frequent SS pairwise differences occur; again, the favored configuration depends on specific level, region, and forecast lead time. In general, the GFS-SAS configuration becomes SS warmer with forecast lead time in the NH and CONUS regions at all levels below 200 hPa. The GFS-GF shows similar behavior at mid-levels over the CONUS region only. The progressive GFS-SAS warming trend with forecast lead time agrees with the results discussed above for 2-m temperature over the CONUS. The exact opposite is seen for TROP, where GFS-SAS becomes SS colder with increasing forecast lead time. In addition, as forecast lead time increases, GFS-GF over TROP becomes SS colder at 850 hPa, SS warmer for mid-levels, and back to SS colder at upper-levels. It should be highlighted that, while the shape of the vertical profile of temperature bias is similar for most regions, it varies significantly between the two configurations for TROP at levels below 200 hPa.

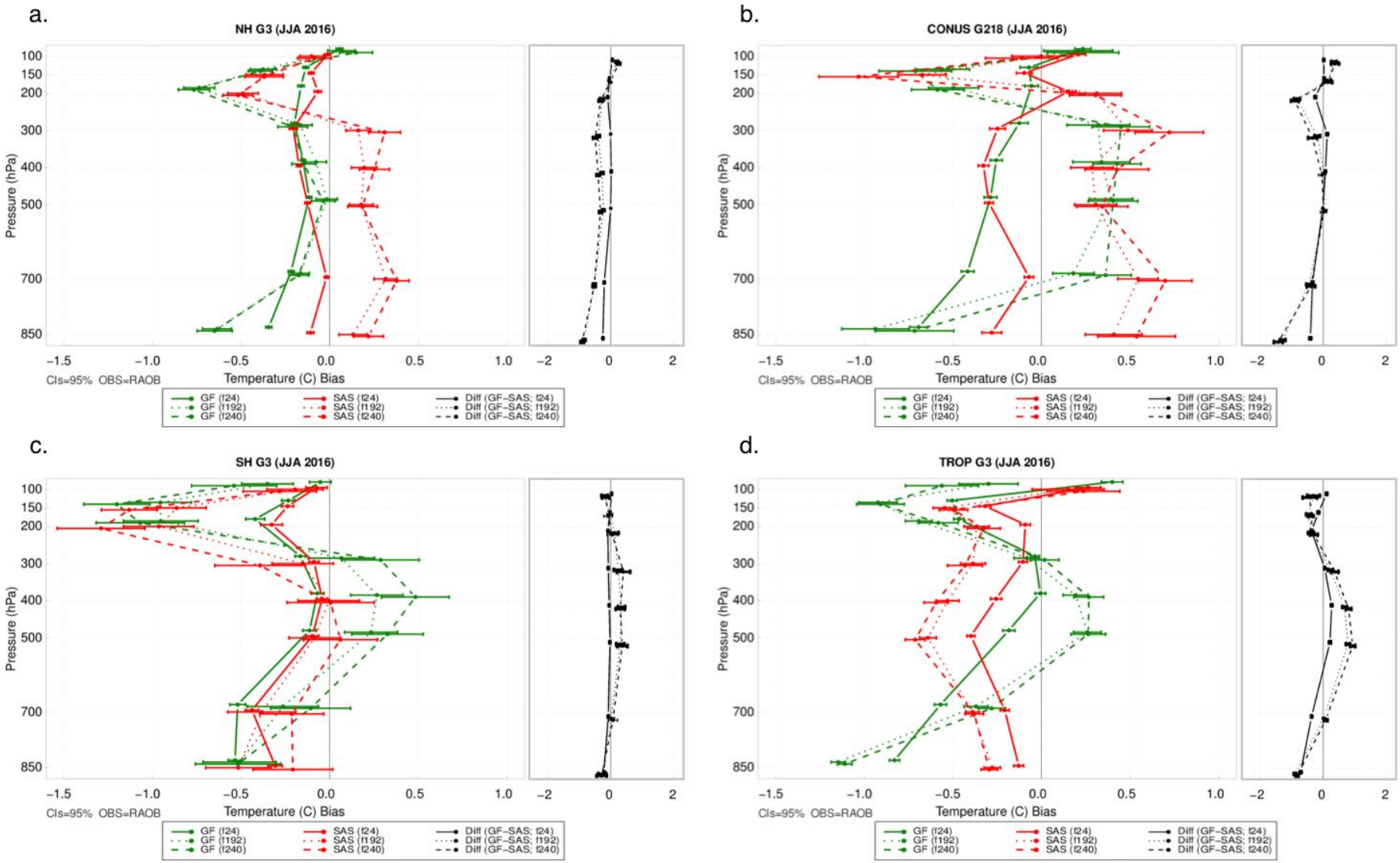



*Figure 23. Same as Fig. 19, except for median bias for temperature (°C).*

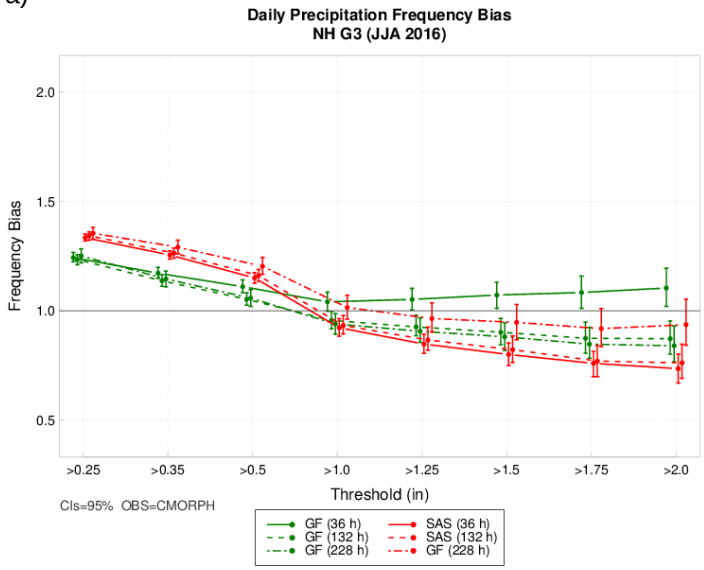
In summary, the relative improvement in performance of GFS-GF compared to GFS-SAS over the length of the forecast is due to a warming trend in GFS-SAS, which has both upper air and surface temperature warm biases that increase with forecast lead time, as well as to a relatively slower error growth in GFS-GF.

**GSM Key Finding 4:** In extratropical regions, precipitation frequency biases are similar overall between the model configurations, with over-precipitation for low thresholds and under-precipitation for high thresholds. However, the diurnal cycle of errors over the CONUS are distinct between the configurations.

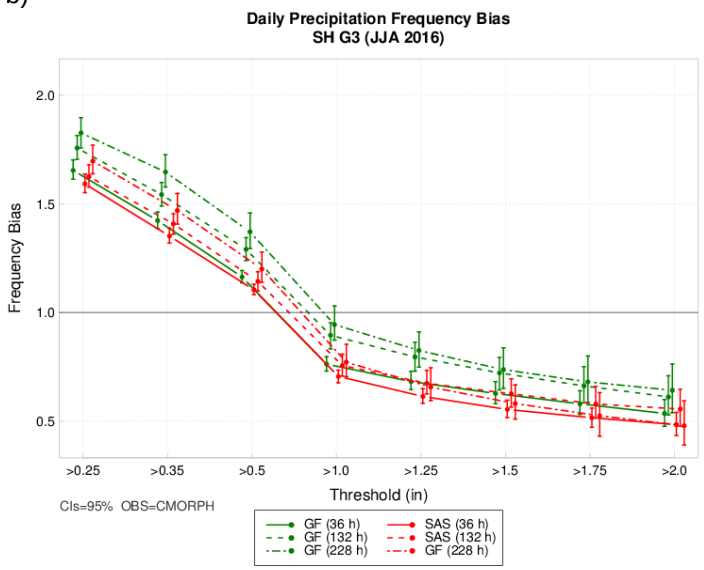
### *Daily Accumulated Precipitation Bias - Global Sub-regions*

Overall, both configurations over-predict daily precipitation at the lower thresholds, with a transition to negative bias for higher thresholds (Fig. 24). An exception is noted for GFS-GF over the tropical region, where bias decreases until the 1.0" threshold before steadily increasing until the 2.0" threshold. With few exceptions, these trends are relatively similar regardless of forecast lead time (denoted by the different line types in Fig. 24). In the NH and tropical regions, at the lowest thresholds, GF has SS lower aggregate bias than SAS, which due to both configurations having a high bias, yields better performance for GFS-GF. At the higher precipitation thresholds, for all global sub-regions, GF typically precipitates more than SAS; depending on whether the configurations are over- or under-predicting precipitation determines the better performer. In the SH and tropical regions, GF precipitates less at 36-h compared to later forecast lead times; whereas, in the NH, there is a decrease in precipitation as forecast lead time increases.

a)




b)




c)

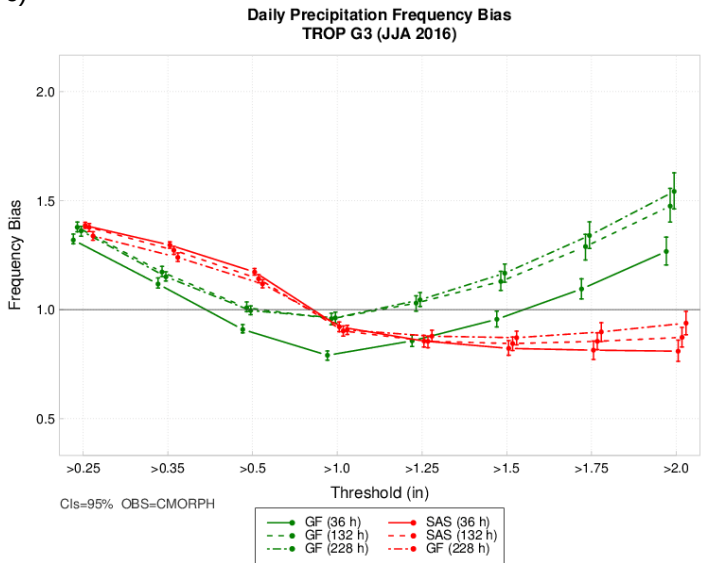



*Figure 24. Frequency bias of 24-h accumulated precipitation (in) for GFS-SAS (red) and GFS-GF (green) aggregated over the (a) NH, (b) SH, and (c) tropical region for JJA 2016. The 36-h forecast lead time is represented by the solid lines, the 132-h forecast lead time in dashed, and the 228-h forecast lead time is dot-dashed. The vertical bars surrounding the aggregate value represent the 95% CIs.*

### *6-h Accumulated Precipitation Bias - CONUS*

The 6-h precipitation frequency biases over the CONUS show a prominent diurnal signal for both configurations at 0.01”, 0.1”, and 0.25” thresholds (Fig. 25). Regardless of threshold, GFS-GF has larger diurnal signal in bias, with similar magnitudes in peak bias compared to GFS-SAS, but lower magnitudes in minimum bias. A slight phase shift between the biases of the two configurations is also apparent. At the 0.01” threshold, the peak bias occurs at 1800 UTC for both GFS-GF and GFS-SAS, potentially signaling an issue with timing of convective initiation, while minimum biases occur at 0600 UTC for GFS-GF and are shifted to 0000 –– 0600 UTC for GFS-SAS. At the 0.1” and 0.25” thresholds, GFS-GF has maximum bias at 1800 UTC, while GFS-SAS has maximum bias at 1200 UTC. Minima in bias at the 0.1” threshold for GFS-GF are at 0600 UTC, and at 0.25” minima occur between 0000 –– 0600 UTC. Minima in bias at the 0.1” threshold for GFS-SAS are between 0000 –– 0600 UTC, and at the 0.25” minima occur between 1800-0000 UTC.

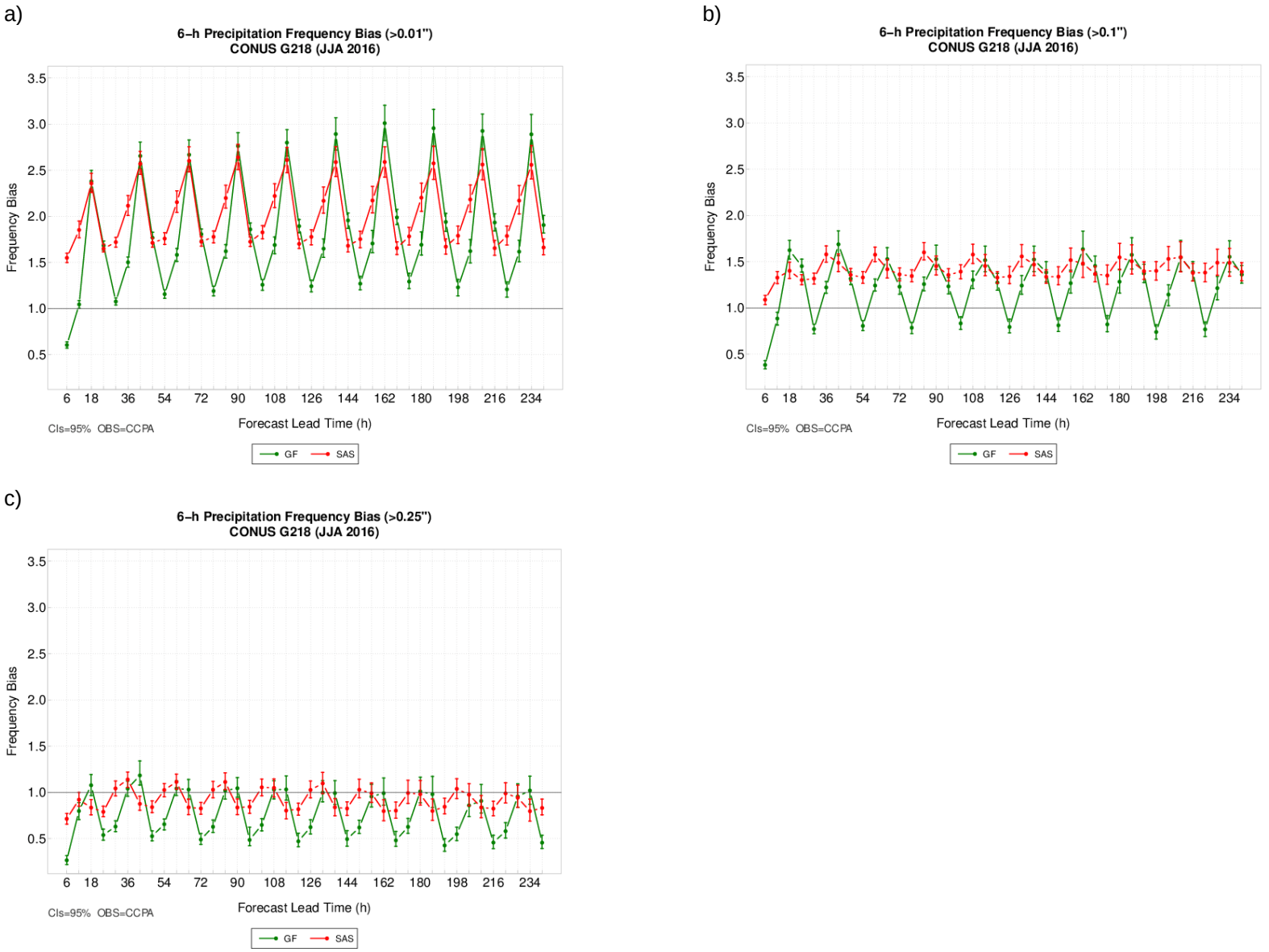



*Figure 25. Frequency bias of 6-h accumulated precipitation (in) for GFS-SAS (red) and GFS-GF (green) aggregated over the CONUS domain for the (a) 0.01", (b) 0.1", and (c) 0.25" thresholds as a function of forecast lead time (h) for JJA 2016. The vertical bars surrounding the aggregate value represent the 95% CIs.*

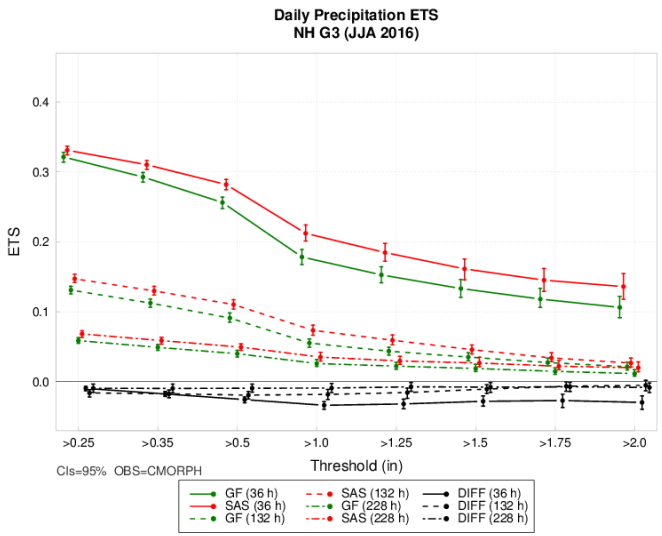
As precipitation threshold increases, a transition to lower frequency biases is clear (Fig. 25). At all forecast lead times for the 0.01” and 0.1” thresholds, GFS-SAS has SS high bias; GFS-GF has a SS high bias at nearly all forecast lead times at 0.01” threshold and at most times valid from 1200-0000 UTC. At the 0.25” threshold, both configurations transition to having unbiased forecasts (i.e., CIs encompass 1) or under-predicting precipitation.

**GSM Key Finding 5:** Overall, GFS-SAS is more skillful at predicting precipitation.

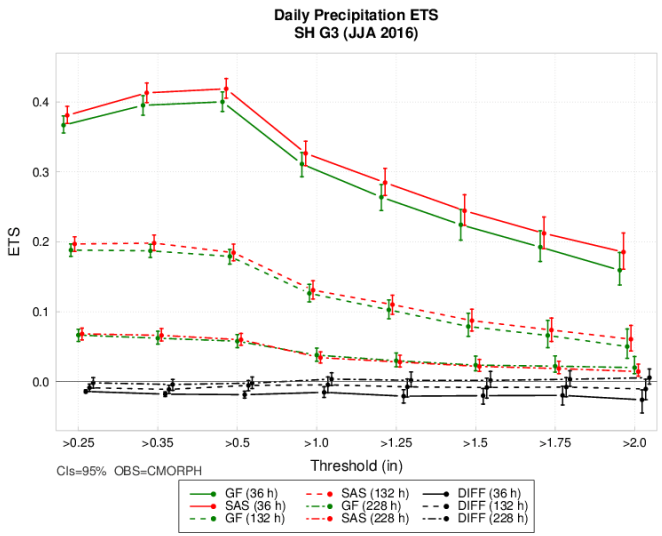
### *Daily Accumulated Precipitation ETS - Global Sub-regions*

There is a general decrease in ETS as precipitation threshold and forecast lead time increase (Fig. 26; different line types denote forecast lead times). The largest differences in daily precipitation between the two configurations are typically at earlier lead times, with the differences becoming smaller throughout the model forecast, especially at thresholds 0.5” (see Fig. 26 for the 36-, 132-, and 228-h forecast). For all differences that are SS, GFS-SAS is the better performing configuration.

a)




b)




c)

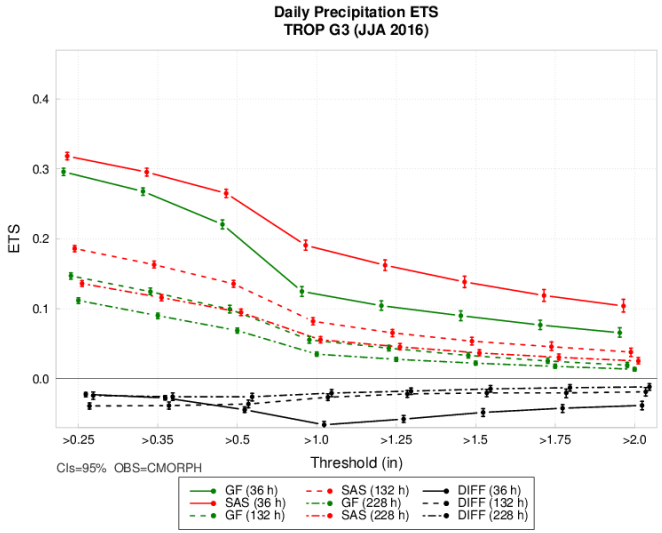



*Figure 26. ETS of 24-h accumulated precipitation (in) for GFS-SAS (red) and GFS-GF (green) aggregated over the (a) NH, (b) SH, and (c) tropical region for JJA 2016. The 36-h forecast lead time is represented by the solid lines, the 132-h forecast lead time in dashed, and the 228-h forecast lead time is dot-dashed. The vertical bars surrounding the aggregate value represent the 95% CIs.*

**6-h Accumulated Precipitation ETS - CONUS**

Similar to findings in the sub-regional analysis, the ETS decreases as the precipitation threshold increases (Fig. 27). Over CONUS a diurnal signal in ETS is clear, which often shows lower amplitude at higher thresholds. For all thresholds, differences between the two configurations are often largest at beginning of forecast period, where GFS-SAS typically has higher skill; performance becomes more similar by end of model integration with fewer overall SS pairwise differences.

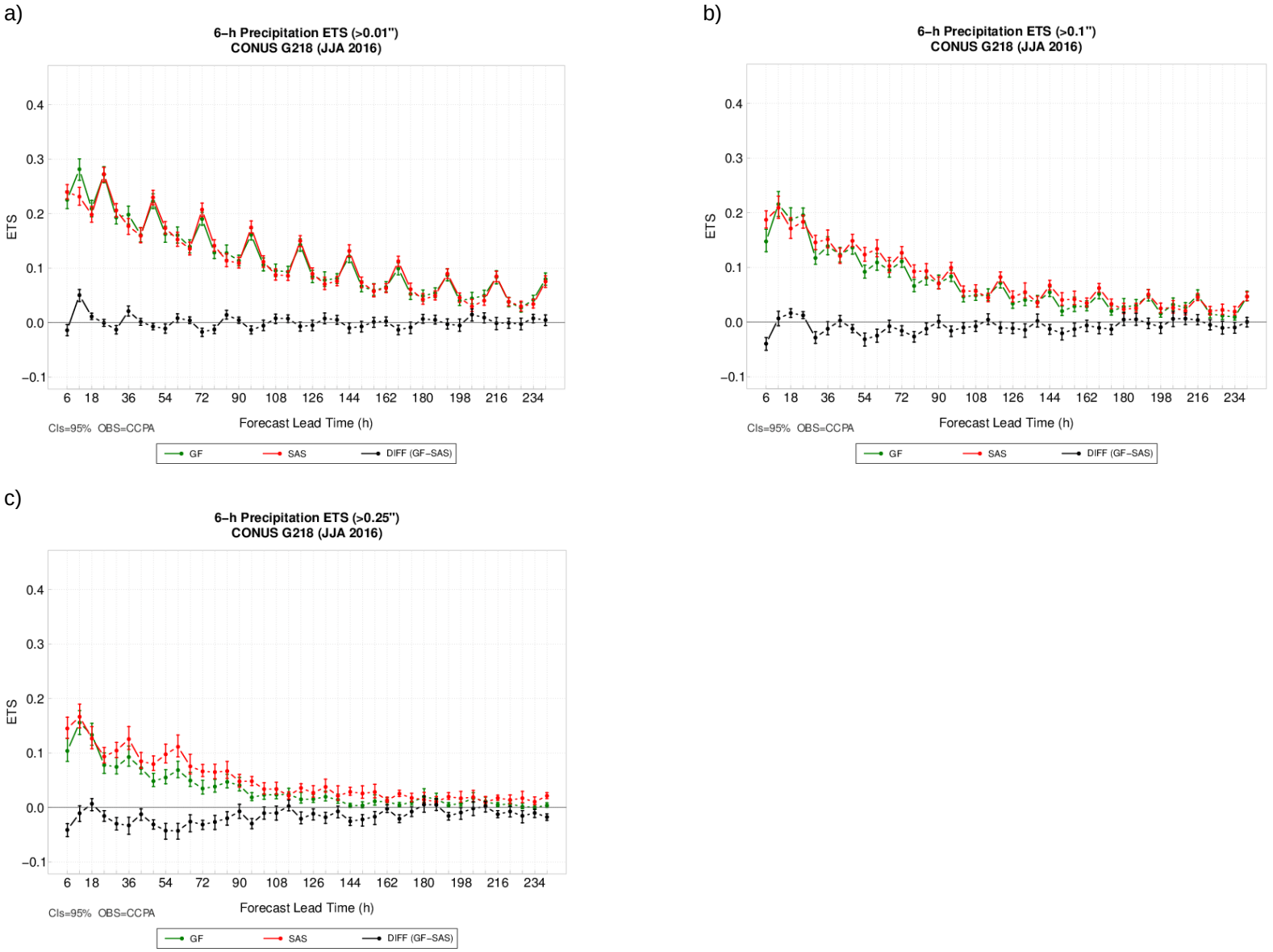



*Figure 27. Time series plot of 6-h accumulated precipitation (in) for aggregated ETS over the CONUS region for (a) 0.01", (b) 0.1", and (c) 0.25" for JJA 2016. GFS-GF is green, GFS-SAS is red, and the pairwise difference is black. The vertical bars surrounding the aggregate value represent the 95% CIs.*

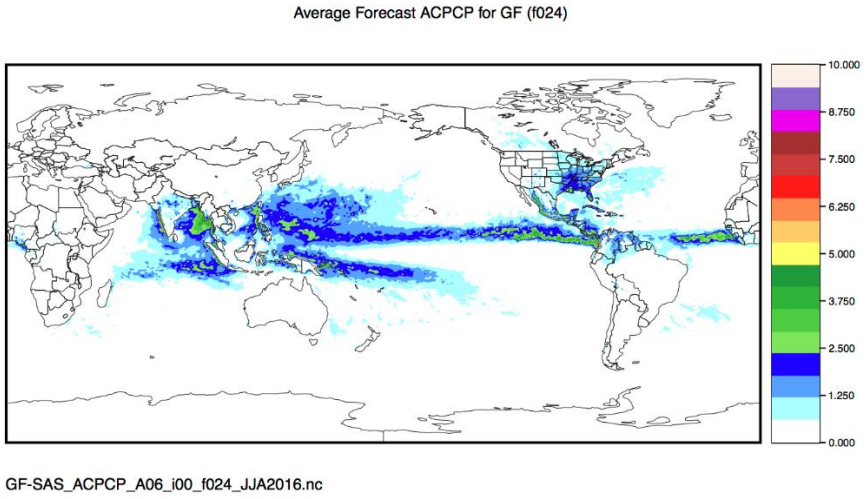
A clear diurnal signal is noted at 0.01", with highest ETS at 00 UTC for both configurations and both configurations having similar performance. When there are SS pairwise differences at the 0.01" threshold, GFS-GF outperforms GFS-SAS at 1200 and 1800 UTC, while GFS-SAS has higher skill at 0000 and 0600 UTC. While the diurnal signal is more muted, at the 0.1" and 0.25" thresholds, the ETS peaks at 1200 UTC. At the 0.1" threshold, after the 24-h forecast, any SS pairwise differences favor SAS and typically occur between 0000-1200 UTC. At the >0.25" threshold, all SS pairwise differences show SAS being the higher skill configuration.

**GSM Key Finding 6:** The partition of precipitation (convective and explicit) is different between the configurations, with GFS-SAS producing more total convective precipitation than GFS-GF.

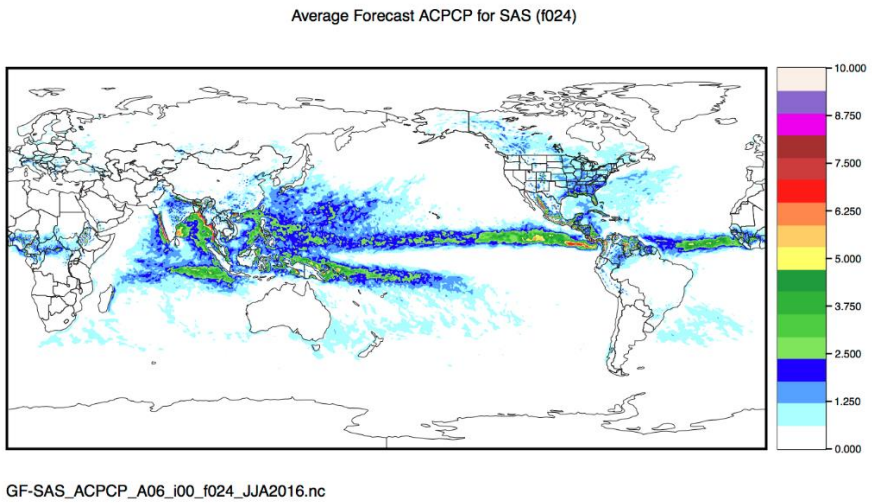
In agreement with the SCM results the fraction of total precipitation created by the convective scheme is smaller in global GFS-GF. Note that while only total precipitation can be verified, the partition results provide valuable information about the interaction of the cumulus scheme with other parameterizations within the GFS physics suite.

To illustrate this point, the forecast 6-h accumulated convective precipitation (ACPCP) was averaged for selected forecast lead times over the entire three-month test period. Overall, the average ACPCP for GFS-SAS and GFS-GF indicate both configurations are appropriately capturing the ITCZ, including the South Pacific Convergence Zone throughout the forecast period (Figs. 28-30). While there are differences in the location and magnitudes of ACPCP, they are not egregious. At the 24-h forecast lead time, most areas where differences are present show GFS-SAS having higher ACPCP than GFS-GF (Fig. 28). The differences are generally greatest over maritime tropical regions, such as ITCZ in the Eastern Pacific and Atlantic Oceans, the Indian Ocean surrounding the Maritime Continent, and areas near coastal regions in the Bay of Bengal. In addition, most differences over land areas (e.g., northern South America, areas over North America, and Equatorial Africa) show GFS-SAS as having higher average ACPCP than GFS-GF. As forecast time progresses, some of these differences are enhanced and grow in magnitude (i.e., the ITCZ in the Eastern Pacific Ocean, Equatorial Africa, and Bay of Bengal), and other areas reverse sign, with GF having higher ACPCP (i.e., Indian Ocean west of Indonesia and northeast of Papua New Guinea). The area to the west of Indonesia and to the east of Papua New Guinea may suggest differences in the placement of tropical convection, showing GFS-GF positioned more to the north, with GFS-SAS farther to the south. Further investigation into regional synoptic patterns may help better understand these displacements.

a)




b)




c)

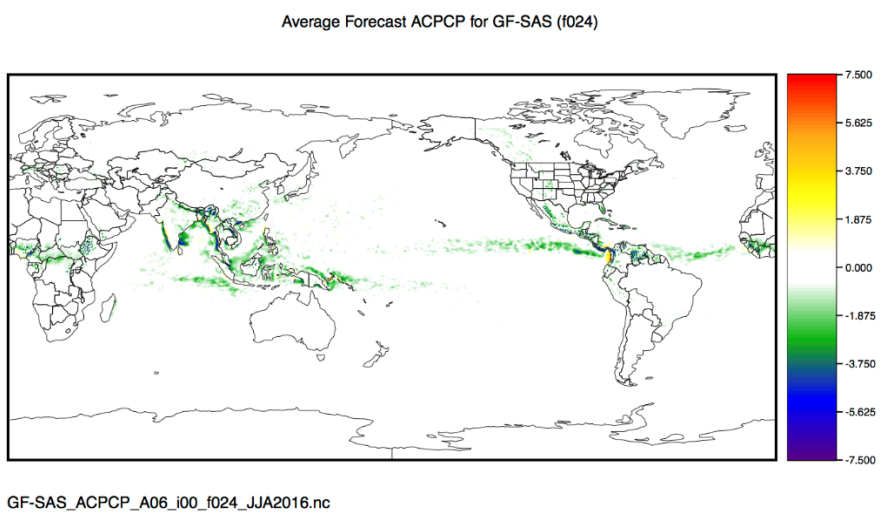
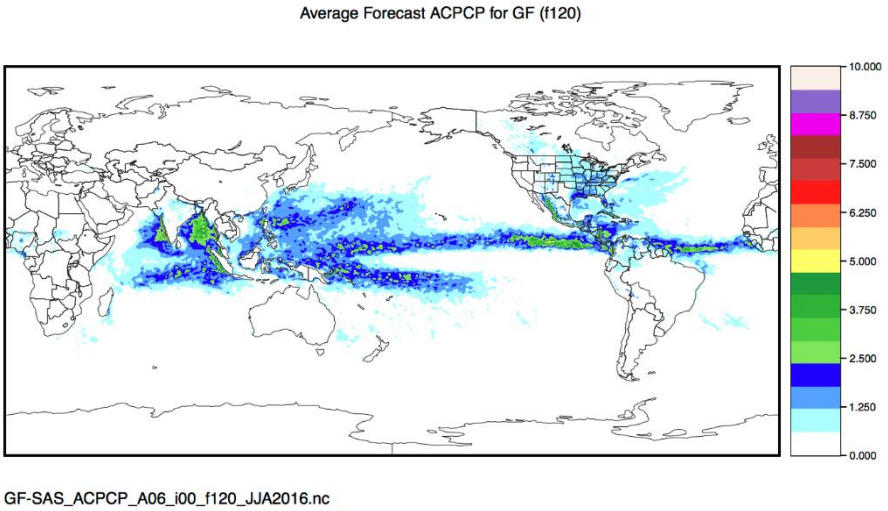
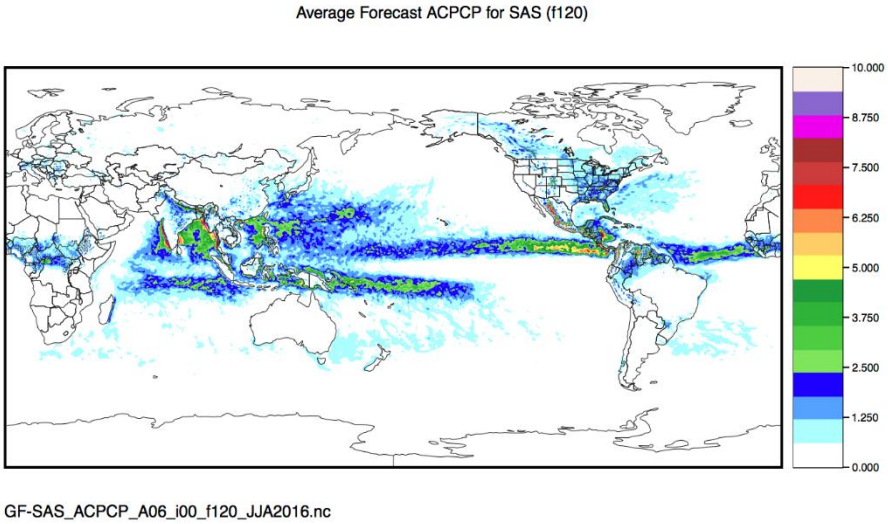



*Figure 28. Average 6-h accumulated convective precipitation (mm) over the three-month test period (JJA 2016) at the 24-h forecast lead time for (a) GFS-GF, (b) GFS-SAS, and (c) GFS-GF - GFS-SAS.*

a)




b)




c)

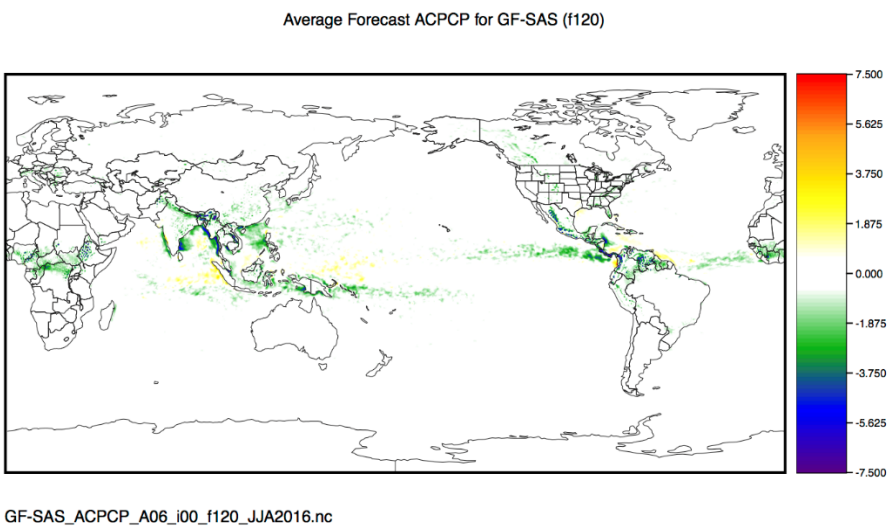
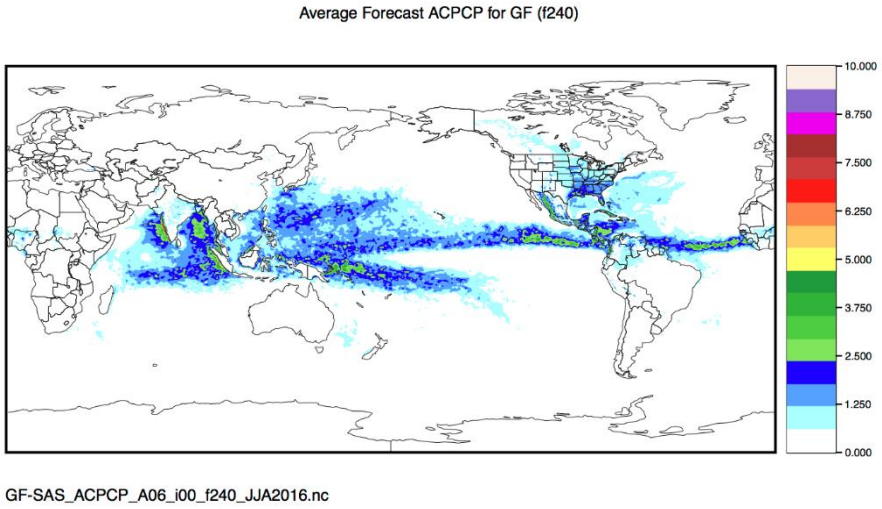
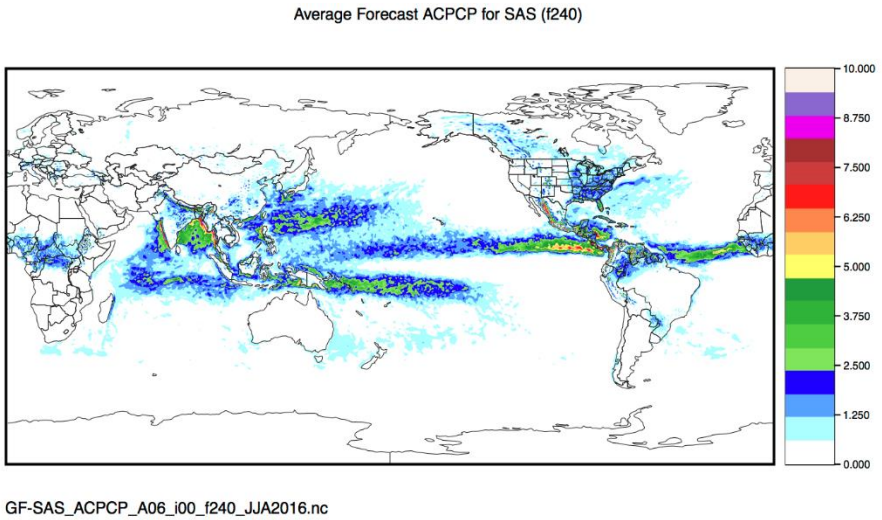



*Figure 29. Same as Fig. 28, but for 120-h.*

a)




b)




c)

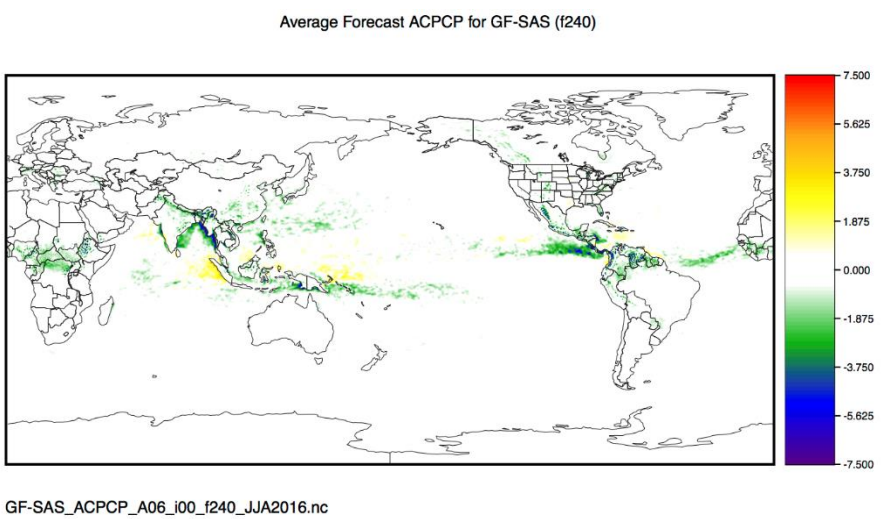



*Figure 30. Same as 28, but for 240-h.*

**GSM Key Finding 7**: Tropical Cyclone track errors averaged over the AL, EP, WP basins are similar for both model configurations. While accuracy in TC intensity forecasts is not expected of a model run at such a coarse resolution, it is interesting to note that storms in GFS-SAS are more intense and have less absolute intensity error than those in GFS-GF.

The GFS-GF mean track errors are on average slightly larger than those for GFS-SAS (Fig. 31) but differences between GFS-GF and GFS-SAS are statistically indistinguishable for all leads times. A comparison of the track error distributions for GFS-GF and GFS-SAS reveals GFS-GF produced more outliers with large track errors. The comparison of GFS-GF and GFS-SAS mean along-track errors (not shown) indicate both configurations have slow biases at almost all lead times, with SAS exhibiting the greater extent of a slow bias with no SS differences.

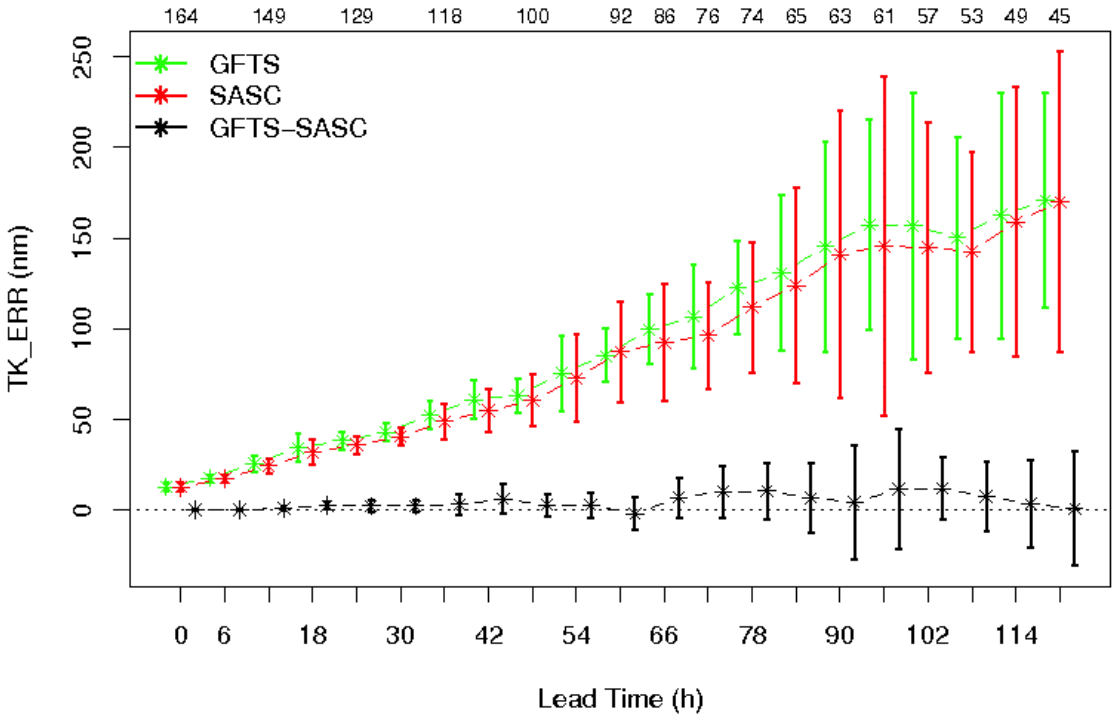



*Figure 31. Mean track errors (nm) for GFS-SAS (red), GFS-GF (green) and their pairwise differences (black) with 95% confidence intervals with respect to forecast lead time (h) in the AL, EP, and WP basins for JJA 2016.*

The mean intensity errors exhibit intensity underprediction (Fig. 32) in both configurations, which likely stems from the coarse resolution used in these retrospective experiments. The intensity errors suggest GFS-GF has a tendency to produce weaker storms than GFS-SAS at all lead times, which contributed to higher absolute intensity errors in GFS-GF (Fig. 33). Note that at the initial time there are five fewer storms in GFS-GF than in GFS-SAS. Even though the ICs are identical for GFS-GF and GFS-SAS, these storms were dismissed at initialization by the NCEP TC tracker in GFS-GF because they are too weak at the 6-hour forecast.

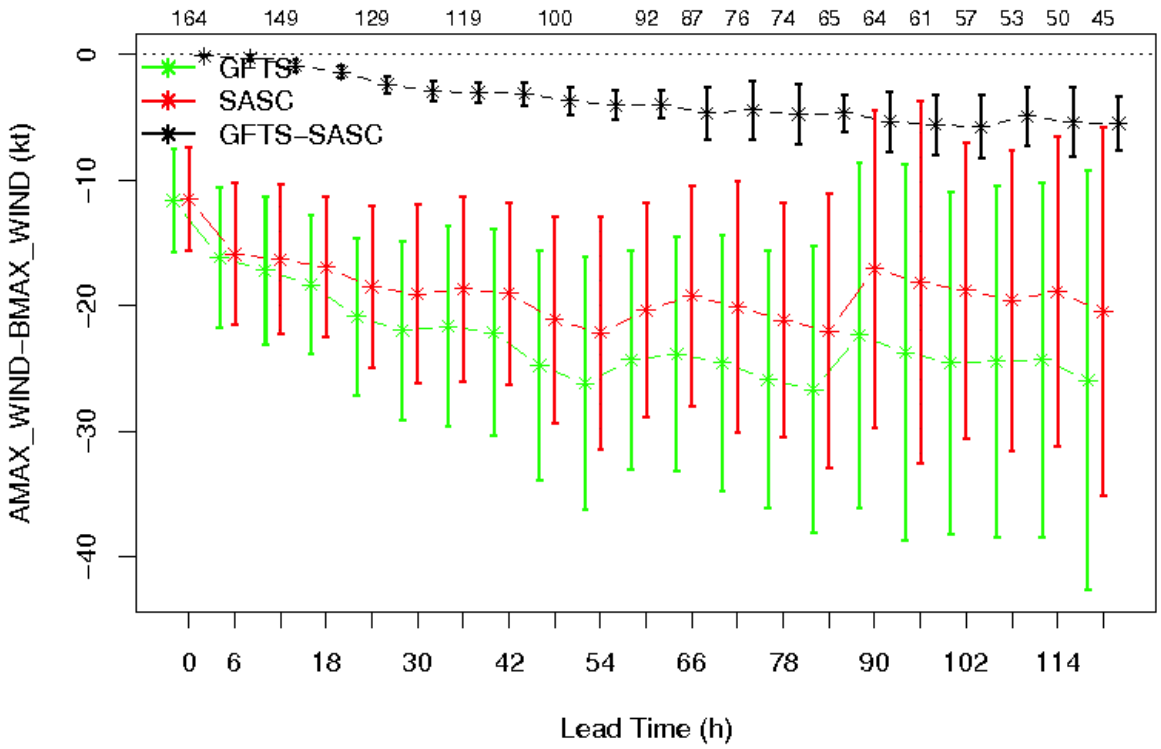



*Figure 32. Mean intensity errors (kt) with 95% confidence intervals with respect to lead time for GFS-GF (green), GFS-SAS (red), and GFS-GF - GFS-SAS pairwise difference (black) in the AL, EP, and WP basins for JJA 2016.*

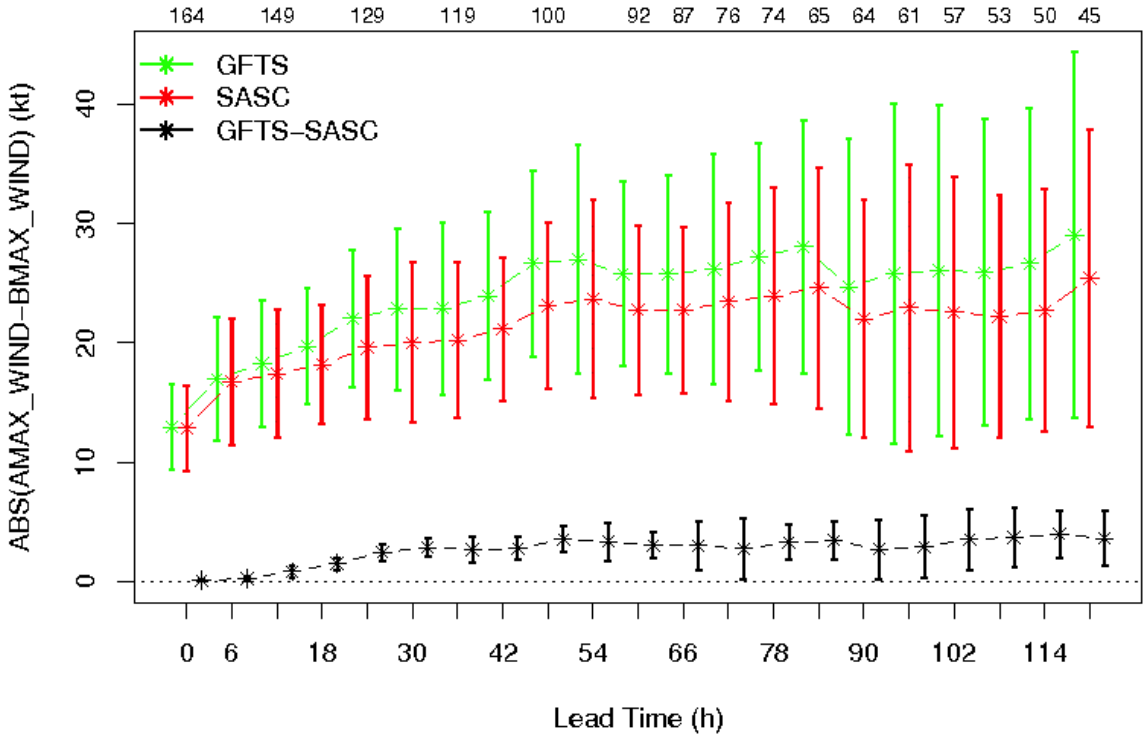



*Figure 33. Same as Fig. 32 but for mean absolute intensity error (kt).*

**GSM Key Finding 8:** While verification of cyclogenesis is beyond the scope of this report, it noticeable that the models have different behaviors, with GFS-GF producing more storms.

An investigation of global cyclogenesis occurring between 00Z 1 June and 00Z 10 Sep 2016 indicated 54 storms reports in the Best Track data, which includes 1 in the Southern Hemisphere, 21 in the EP, 7 in the Indian Ocean basin, 14 in the WP, 2 in the central Pacific basin, and 9 in the AL. On average, there are 5.3 TGs per 10 days in nature. Figure 34 shows the first reported TG counts during 2016 summer daily retrospective forecasts. GFS-GF has a tendency to generate more TGs than GFS-SAS over the three-month period, with a 10-day average frequency in SAS of 5.19, and 8.66 in GFS-GF, which indicates that the GFS-SAS has a more favorable rate of cyclogenesis.

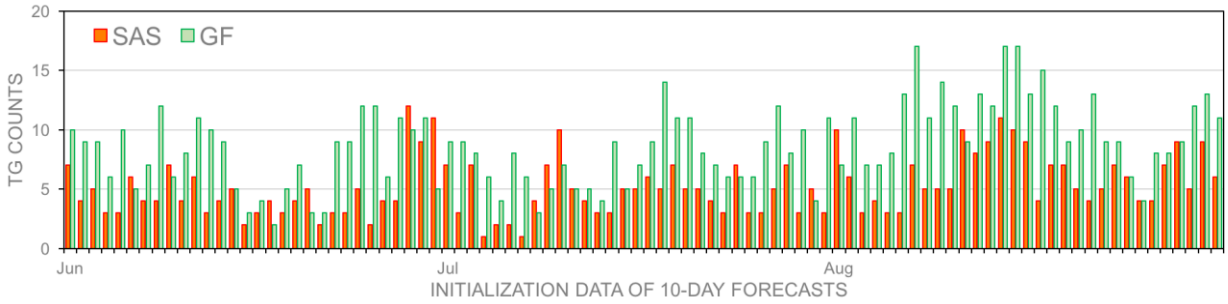



*Figure 34. Model-generated TG counts by 10-day forecast period during the JJA 2016.*

# Discussion and Conclusions

The testing and evaluation of the GF cumulus parameterization conducted by the GMTB illustrate the complexity -- yet scientific usefulness -- of connecting a new scheme to the GSM. The goals of the test were twofold:1) establish an initial hierarchical testbed capability and 2) exercise the hierarchical testing framework, including the SCM and global workflow, to inform the development of an advanced physics suite for NOAA's GFS using the GF cumulus parameterization. With the establishment of the hierarchical testbed, the GMTB is well equipped to facilitate transition of research to operations (R2O).

The success of this test was heavily dependent on interactions among and investment by GMTB, the physics developer, and EMC's Global Team. The main developer of the GF scheme (G. Grell) provided the original GF code as well as updated versions, which included changes for addressing bugs that were discovered, as well as a reorganization of the code to make it more modular and easy to port to different models (now separated into three parts: GF driver, shallow convection, and deep convection). The close collaboration and iteration with the developer helped ensure the GMTB properly connected the GF parameterization within the GSM code. In addition, the collaboration with the EMC Global Team was essential to the GF test. The team at EMC provided the code, scripts, and workflow to run the global model as well as their expertise and assistance with issues related to setting up and running the workflow and producing objective verification. Moving forward, it will be necessary to further engage with EMC and continue to get feedback regarding desired verification methods and displays to be prioritized for future implementation in the testbed.

As more advanced physics schemes are continually being developed by the scientific community, such as the GF cumulus parameterization, the process for integrating these new physics schemes in operational code should be streamlined to encourage and facilitate R2O. The hierarchical testbed provides tools for physics developers, including EMC's physics developers, to display merit and further improve upon their schemes to accelerate R2O. As part of the lower tier in the hierarchical testing framework, the SCM provides the ability to examine more complex atmospheric phenomena on a physics suite in a cost-effective way. In the GF test, the SCM allowed both the developer and GMTB to test the code to better understand the interaction of the GF parameterization with other schemes in the GFS physics suite. This helped ensure the code was properly connected. In fact, initial testing of the SCM with the GF code uncovered an issue where the GF was producing near-zero deep convective tendencies; these results were shared with the developer, which led to a code fix. As the testing progressed to higher tiers, the global workflow provided a straightforward way to test the code to make sure it was properly integrated and provided a means to assess basic global-scale forecast implications of altered physics.

While the results from this test inform both the developer and EMC, several caveats and limitations must be taken into consideration when interpreting the results. The GMTB applied substantial effort toward properly implementing the GF code into the GSM, but the developer performed no tuning of the GF. While the code used for this effort is mature, without tuning, there are obvious impacts when comparing to an operational suite that has been heavily tuned. In addition, due to limitations in computational resources, all runs were performed at coarse resolution (T574) and did not include cycling. Forecasts for both SAS and GF were initialized from the operational GFS analyses. This resulted in spin-up issues for the GF configuration that would be mitigated if cycling with the GF scheme were implemented. Finally, this test did not exercise the scale-aware capability implemented in the GF scheme because a single coarse resolution was used.

Testing with the SCM resulted in several key findings, with one finding aligning with results from the 3-D global forecasts. The suite using the GF parameterization produced weaker convective tendencies and convective transport than SAS. In the three months of global forecasts, SAS typically produced more convective precipitation in the tropical regions than GF. While not all results from the SCM can be translated to the full global forecasts, this highlights the utility and process of the hierarchical testing.

When considering results from the three months of global forecasts, the superior cumulus parameterization is highly dependent on metric, variable, level, forecast hour, and region. In particular, the difference in RMSE between GFS-GF and GFS-SAS shows more favorable results for GFS-GF later in the forecast. Given the limitations described above regarding the test set-up, especially cold starting from SAS-based GFS operational analyses, the objective verification results should be interpreted cautiously. The objective verification, including the TC evaluation, showed the GF scheme is hooked up properly (i.e., no egregious errors compared to SAS) and provides baselines for the developer on the performance of the scheme. Overall, while the untuned GF scheme did not outperform the current operational suite, it does show considerable potential.

This testing effort helped established an initial capability of a hierarchical testbed that is freely available to the modeling community to help assist with their research and development of more advanced physics suites. The functionalities and tools centralized by the GMTB, including code, scripts, datasets, and documentation are all provided as part of the testbed. In addition, this activity also provides both a baseline and experimental configuration that are archived and available for use by the research and operational communities. As an example of data usage generated from this test, the model data is currently being used by NGGPS Principal Investigator Jason Otkin and his group at University of Wisconsin to support their work producing synthetic satellite output from UPP to evaluate the model's ability to accurately simulate clouds and moisture.

As the GMTB concludes the GF test and looks forward to the next steps, there are several potential avenues to pursue. One avenue is to continue working with the GF parameterization. Key to further testing of the GF is allowing for the tendencies to be output from the model. To follow the spirit of the hierarchical testing framework, progressing to the next tiers of testing would be valuable in the continued development of the GF parameterization. Next steps could include tuning (e.g., momentum transport, additional temperature and moisture perturbations, and closures for deep and shallow convection), cycling, and increasing resolution. These follow-up steps are obviously resource intensive and would require partnership with EMC. Also, given the effort invested in implementing the GF code into the GSM and keeping current with EMC's code base, along with the charge from NGGPS to move toward more advanced physics, it would be beneficial to commit the GF code to the NEMS repository at EMC. In addition, ESRL is funded through NGGPS to incorporate a high-resolution physics suite into the CCPP, which will include the GF. Testing and inclusion of the GF scheme for HWRF is also currently underway, highlighting the number of modeling systems moving forward with GF. Finally, another path forward would be for the GMTB to test other advanced physics parameterizations, such as Thompson microphysics and RRTM for General Parallel Applications (RRTMGP) radiation.

Moving forward with the hierarchical testbed, the GMTB will continue to advance and evolve the testbed to include additional functionality. For the SCM, this would include developing the ability to extract column data and necessary forcing terms from global runs to drive the SCM, stay current with a changing Interoperable Physics Driver, expand a case catalog, and foster external users' ability to add their own test cases. For the global workflow, the GMTB would like to include more diagnostic capabilities to better demonstrate strengths/weaknesses of physics parameterizations, include additional observation platforms for diagnosing and verifying cloud and radiation fields, and, in general, solicit community-contributed diagnostics to include in the testbed. A key aspect in ensuring the continued growth and success of the GMTB physics testbed is involvement from the research and operational communities, whether it is exercising capabilities in the testbed and sharing relevant results, providing constructive feedback, or contributing code and/or scripts to include in the testbed.

## Acknowledgements

The GMTB staff would like to thank the GF developer (Georg Grell of NOAA/ESRL), the NOAA/EMC Global Team led by Vijay Tallapragada, the NGGPS Physics Team co-leads (then Jim Doyle of NRL, Shrinivas Moorthi of NOAA/EMC and Bill Kuo of NCAR), and the NGGPS Program Office led by Fred

Toepfer for their engagement in preparing the test plan, defining verification metrics, supplying code, and participating in the interpretation of results. Thanks also to Karen Griggs of NCAR for offering her desktop publishing expertise in the preparation of this document.

## References

Davies, L., C. Jakob, K. Cheung, A. D. Genio, A. Hill, T. Hume, R. J. Keane, T. Komori, V. E. Larson, Y. Lin, X. Liu, B. J., Nielsen, J. Petch, R. S. Plant, M. S. Singh, X. Shi, X. Song, W. Wang, M. A. Whitall, A. Wolf, S. Xie, and G. Zhang, 2013: A single-column model ensemble approach applied to the TWP-ICE experiment. J. Geophys. Res.: Atmos. 118, 12, 6544–6563. DOI: http://dx.doi.org/10.1002/jgrd.50450.

Grell, G.A, and S. R. Freitas, 2014: A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. Atmos. Chem. Phys., 14, 5233- 5250. DOI: http://www.atmos-chem-phys.net/14/5233/2014/ doi:10.5194/acp-14-5233-2014

Randall, D. A., and D. G. Cripe, 1999: Alternative methods for specification of observed forcing in single-column models and cloud system models. J. Geophys. Res.: Atmos., 104, D20, 24527–24545. DOI: http://dx.doi.org/10.1029/1999JD900765.

Randall, D., S. Krueger, C. Bretherton, J. Curry, P. Duynkerke, M. Moncrieff, B. Ryan, D. Starr, M. Miller, W. Rossow, G. Tselioudis, and B. Wielicki, 2003: Confronting models with data - the GEWEX cloud systems study. Bull. Amer. Meteor. Soc., 84, 455–469. DOI: http://dx.doi.org/10.1175/BAMS-84-4-455.

Taylor, K., 2001: Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res.: Atmos. 106, D7, 7183–7192. DOI: http://dx.doi.org/10.1029/2000JD900719

Zhang, M., R. C. J. Somerville, and S. Xie, 2016: The SCM concept and creation of ARM forcing datasets. Meteor. Monogr. 57, 24.1–24.12. DOI: http://dx.doi.org/10.1175/AMSMONOGRAPHS-D-15-0040.1.

# Appendix A. List of acronyms

| | |
|---|---|
| **AL** | Atlantic TC basin |
| **ARM** | Atmospheric Radiation Measurement |
| **BUFR** | Binary Universal Form for Representation of Meteorological Data |
| **CI** | Confidence Interval |
| **CCPA** | Climatology-Calibrated Precipitation Analysis |
| **CONUS** | Contiguous United States |
| **CMORPH** | Climate Prediction Center Morphing Technique |
| **DTC** | Developmental Testbed Center |
| **EMC** | Environmental Modeling Center |
| **ESRL** | Earth System Research Laboratory |
| **EP** | Eastern North Pacific TC basin |
| **ETS** | Equitable Threat Score |
| **FBias** | Frequency Bias |
| **GEWEX** | Global Energy and Water cycle EXperiment |
| **GCSS** | GEWEX Cloud System Study |
| **GDAS** | Global Data Assimilation System |
| **GF** | Grell-Freitas cumulus parameterization |
| **GFS** | Global Forecast System |
| **GFS-GF** | Global Forecast System run with GF |
| **GFS-SAS** | Global Forecast System run with SAS |
| **GISS** | Goddard Institute for Space Studies |
| **GMTB** | Global Model Test Bed |
| **GRIB2** | GRIdded Binary file format version2 |
| **GSD** | Global Systems Division |
| **GSM** | Global Spectral Model |
| **HPSS** | High Performance Storage System |
| **HWRF** | Hurricane Weather Research and Forecasting System |
| **ITCZ** | Intertropical Convergence Zone |
| **JTWC** | Joint Typhoon Warning Center |
| **LES** | Large Eddy Simulation |
| **MET** | Model Evaluation Tools |
| **METAR** | aviation routine weather report |
| **NAM** | North American Mesoscale Forecast System |
| **NCAR** | National Center for Atmospheric Research |
| **NCEP** | National Centers for Environmental Prediction |
| **NDAS** | NAM Data Assimilation System |
| **NEMS** | NOAA Environmental Modeling System |
| **NEMSIO** | NEMS Input/Output format |
| **NH** | Northern Hemisphere (defined here as 20o – 80o N for upper air verification and 20o – 60o N for precipitation verification) |
| **NOAA** | National Oceanic and Atmospheric Administration |
| **NGGPS** | Next-Generation Global Prediction System |
| **NHC** | National Hurricane Center |
| **PBL** | Planetary Boundary Layer |
| **PrepBUFR** | Quality-controlled BUFR |

| | |
|---|---|
| **RAOB** | RAwinsonde OBservation |
| **RAP** | Rapid Refresh Forecast System |
| **RMSE** | Root-Mean-Square Error |
| **RRTM** | Rapid Radiative Transfer Model |
| **RRTMG** | RRTM for General Circulation Models |
| **RRTMGP** | RRTM for General Parallel Application |
| **R2O** | Transition of Research to Operations |
| **SAS** | Simplified Arakawa-Schubert cumulus parameterization |
| **SCM** | Single-column Model |
| **SH** | Southern Hemisphere (defined here as 20o – 80o S for upper air verification and 20o – 60o S for precipitation verification) |
| **SS** | Statistically Significant |
| **SST** | Sea Surface Temperature |
| **SVN** | Apache Subversion |
| **TC** | Tropical Cyclone |
| **TG** | Tropical Cyclogenesis |
| **TROP** | Tropics (defined here as 20o S – 20o N) |
| **TWP-ICE** | Tropical Warm Pool - International Cloud Experiment |
| **UPP** | Unified Post Processor |
| **UTC** | Coordinated Universal Time |
| **VLab** | Virtual Laboratory |
| **WP** | West Pacific TC basin |