PROJECT TITLE: Advancing forecast verification and model development efforts through the development of a flexible satellite-based verification system for the Global Forecasting System

INVESTIGATORS:

PI: Jason Otkin – University of Wisconsin-Madison, CIMSS/SSEC
Co-I: Chris Rozoff – National Center for Atmospheric Research
Collaborators: Sharon Nebuda, Tom Greenwald, Ruiyu Sun, Ligia Bernardet, Michelle Harrold, Emily Liu, Andrew Collard, and Vijay Tallapragada

NOAA GRANT NUMBER: NA16NWS4680010

PROJECT DURATION: 2 years (2016-2018)

TIME PERIOD ADDRESSED BY REPORT: September 2017 – February 2018

1. PROJECT OVERVIEW

This project will use simulated satellite brightness temperatures to evaluate the ability of advanced parameterization schemes in the GFS model to produce accurate cloud and water vapor forecasts. Model output from both full-resolution and coarse-resolution GFS model simulations employing different parameterization schemes will be converted into simulated infrared and microwave brightness temperatures for both clear- and cloudy-sky conditions using the Community Radiative Transfer Model (CRTM) included in the Gridpoint Statistical Interpolation (GSI) system or in the Unified Post Processor (UPP). The satellite simulator capabilities of the CRTM will be enhanced by increasing the consistency between the cloud property assumptions made by a given microphysics parameterization scheme and those used by the CRTM when computing cloud-affected brightness temperatures. These enhancements will be part of a flexible satellite-based forecast verification system that incorporates a variety of statistical methods.

We will rigorously evaluate the accuracy of the simulated cloud and water vapor fields generated by each suite of parameterization schemes through comparison of observed and simulated infrared and microwave brightness temperatures from multiple geostationary and polar-orbiting satellite sensors. The forecast accuracy will be assessed for different regions using traditional grid point statistics and neighborhood-based methods such as the Fractions Skill Score (FSS) and probability distributions. Satellite-based verification metrics developed during this project will be used in combination with traditional operational verification methods to provide a comprehensive assessment of the impact of the advanced parameterization schemes on the GFS forecast accuracy over a range of spatial and temporal scales. Though the project initially focuses on the GFS model, the verification system will be developed to be extensible and beneficial to other model development efforts in the NGGPS framework. Our research efforts will be closely coordinated with collaborators at the Environmental Modeling Center (EMC) and the Global Model Test Bed (GMTB) at the Developmental Testbed Center (DTC) to ensure operational relevance.

2. RECENT ACCOMPLISHMENTS

During the past six months, our efforts have primarily focused on assessing the accuracy of high-resolution GFS model forecasts generated by our collaborators at EMC through comparisons of observed and simulated infrared brightness temperatures. Recent work includes development of new satellite-based verification tools and an assessment of both large-scale and regional forecast errors. Results are discussed in more detail below.

2.1. Analysis system development and data preparation

During the past six months, we have continued to assess the accuracy of a large set of GFS model simulations run at T1534 spectral resolution. These model simulations were performed by Ruiyu Sun at EMC to assess the performance of the WSM6 and Thompson microphysics schemes, both of which are candidates for future inclusion in the FV3GFS model. A set of 10-day long forecasts covering parts of July and December 2014 was generated using each microphysics scheme. Our recent analysis has focused on the WSM6 scheme due to a bug in the implementation of the Thompson scheme that led to unrealistically warm infrared brightness temperatures due to insufficient upper-level cloud cover. As discussed in the previous report, this bug suppressed homogeneous ice nucleation, which caused ice clouds to primarily be generated by upward advection of cloud water, thereby leading to a notable lack of upper-level clouds. Given this bug, it is prudent to only focus on assessing the accuracy of the forecasts employing the WSM6 scheme. Dates with complete 10-day forecasts include July 3-7 and December 11-12, 14, 18, and 31. In this project, we have focused on validating Northern Hemisphere regions associated with GOES-15 over the northern Pacific and western North America, GOES-13 over eastern North America and the northern Atlantic, and the full-disk domain of the Meteosat SEVIRI covering Europe and Africa. We are supporting model development efforts at EMC through a detailed evaluation of the forecast accuracy via comparisons of simulated and observed infrared brightness temperatures from these sensors.

To ensure that the model verification tools are as portable as possible, they have been written using Python, Fortran, and Bash scripting. The tools are flexible so that they can be used to assess any region of interest. The specific focus thus far has been on the 10.7 μm (infrared window) and 6.5 μm (water vapor) bands. The analysis tools were applied both to the full region covered by a given satellite sensor as well as to smaller regions demarcating important cloud regimes, such as the subtropical stratocumulus regions over the northeastern Pacific and southeastern Atlantic, mid-latitude cyclones, and the inter-tropical convergence zone (ITCZ). As described in more detail in the previous report, simple tools such as root-mean square error and bias provide a bulk understanding of how well the GFS represents clouds and moisture, but these statistics do not indicate which brightness temperature ranges are most problematic or which geographical regions or cloud regimes have the greatest systematic and random errors. Thus, to provide additional insight into the forecast accuracy, we also developed probability density functions (PDFs) of brightness temperatures for each satellite sensor and channel. In this report, we provide an update on the development of new tools that give greater physical insight into the errors in the GFS. The tools that have been most informative are scatterplot analyses and the neighborhood-based verification technique known as the Fractions Skill Score (FSS).

2.2. Large-scale model forecast accuracy assessments – scatterplots

One of the tools that we developed using Python generates two-dimensional scatterplots that are used to compare observed and simulated brightness temperatures for the infrared window and water vapor channels. Similar to PDFs, scatterplots provide useful insight regarding which brightness temperatures contribute most to the overall errors in the forecasts. Figure 1 shows several scatterplots valid at 6-h and 216-h forecast lead times (left and right columns) computed using data from the July 2014 forecasts. The top row shows the PDFs for the GOES-15 10.7 μm (infrared window) band, whereas the bottom row shows the PDFs for the GOES-15 6.5 μm (upper-level water vapor) band. Note that the PDFs are plotted using a logarithmic color scale.

For a 6-h lead-time, most of the simulated and observed 10.7 μm brightness temperatures (top left panel) are between 280 and 300 K, and are indicative of regions containing clear skies or low-level clouds. For this brightness temperature range and forecast hour, there is decent agreement between the forecasts and observations as is indicated by the highest probabilities tending to cluster along a diagonal line. For colder brightness temperatures; however, there is greater spread. Moreover, a large warm bias is present in the forecasts with the simulated brightness temperatures seldom < 220 K despite a large number of observations below that value. By the 216-h forecast (upper right panel), much greater spread has developed in the PDF for observed and simulated brightness temperatures near 295 K. The larger spread is indicative of larger spatial errors in upper level clouds at the longer forecast lead-times. It is also noteworthy that the probability of colder brightness temperatures increases for the 216-h forecast lead-time; however, their occurrence is still less than what was observed. Together, this indicates that though all of the forecast lead times are deficient in upper-level cloud cover that this deficiency is largest during the first part of the forecast, potentially due to model spin-up processes.

For the 6.5 μm water vapor band (bottom row), there is decent agreement between the observed and simulated brightness temperatures warmer than 235 K in the 6-h forecast. The small cool bias in the forecast brightness temperatures indicates that there is slightly too much water vapor in the middle troposphere in the forecasts. For colder brightness temperatures; however, the much warmer simulated brightness temperatures indicate that the model has large dry bias in the upper troposphere. This warm brightness temperature bias is also associated with the large deficiency in upper-level cloud cover in the 6-h forecasts. These error patterns persist into the 216-h forecast period.

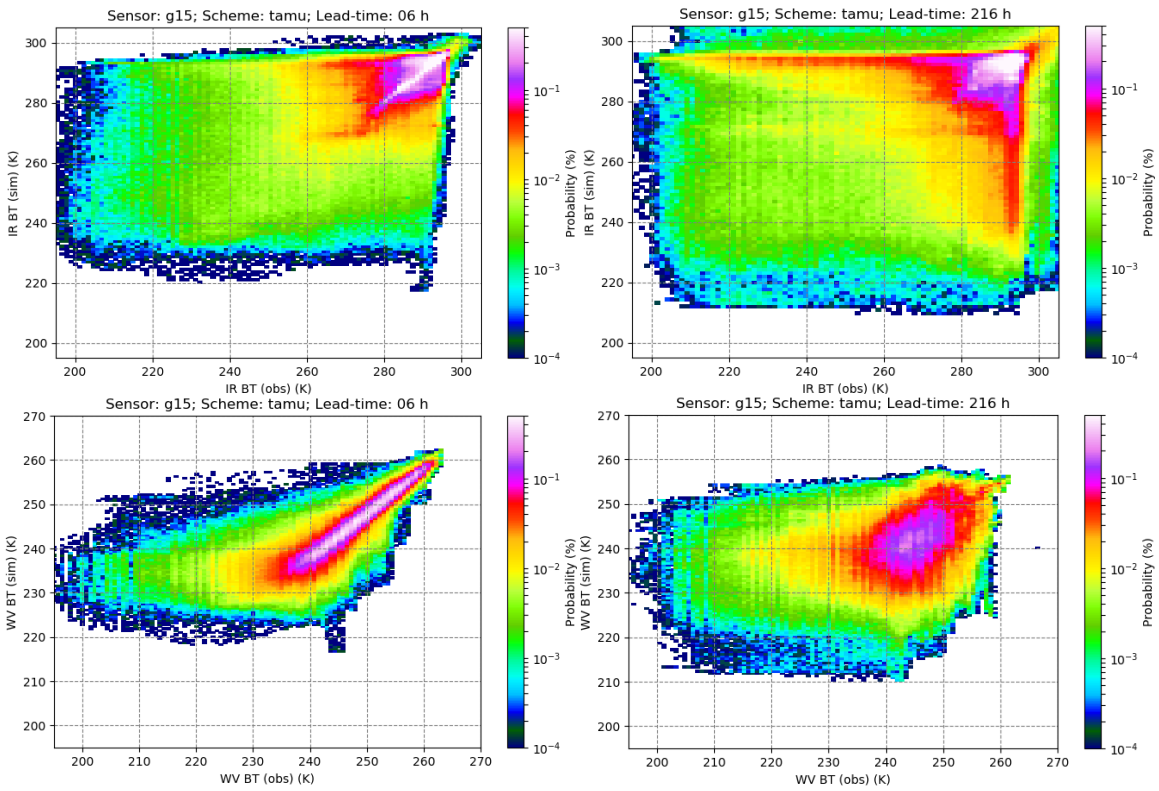

Figure 1. Scatterplots valid at 6-h (left-hand column) and 216-h (right-hand column) forecast lead times for the GOES-15 10.7 μm brightness temperatures (K) (top row) and 6.5 μm brightness temperatures (bottom row). The observed brightness temperatures are plotted along the x-axis, with the simulated brightness temperatures plotted along the y-axis. Each scatterplot has been normalized by the sample size and thus provides a two-dimensional probability distribution.

2.2. Large-scale model forecast accuracy assessments – Fractions Skill Score

It is well known that when comparing observed and predicted fields from high-resolution models that point-based verification methods may unfairly penalize a prediction when a model produces a reasonable cloud field resembling observations but is displaced by a relatively small distance from the observed cloud field. Therefore, neighborhood-based statistics such as the FSS (Roberts and Lean 2008; Roberts 2008) are useful because they are less sensitive to small spatial displacements. As such, we here use the FSS to evaluate errors in the forecast cloud field as depicted by the 10.7 μm infrared window band.

The FSS is described in detail in Roberts and Lean (2008) so it is briefly summarized here. For this project, the FSS is computed using a brightness temperature threshold, where forecast grid points exceeding a given threshold are set to one and all other points are set to zero. Then, a neighborhood size is defined as $N \times N$ grid points centered on any given grid point. The fraction of grid points within the neighborhood surrounding a given grid point that have brightness temperatures exceeding the chosen brightness temperature threshold is then calculated. Thus, for each grid point in a given domain, we can calculate the fraction of grid points within a given neighborhood that exceed the chosen brightness

temperature threshold for both the observed field (f_o) and the simulated field (f_m). The FSS is then related to the mean-square error as follows:

$$\mathrm{FSS} = 1 - \frac{\sum_{i=1}^{N}\left(f_{o,i} - f_{m,i}\right)^2}{\sum_{i=1}^{N} f_{o,i}^2 - \sum_{i=1}^{N} f_{m,i}^2}.$$

Figure 2 shows the FSS computed using the simulated and observed 10.7 μm brightness temperatures for different forecast lead times along the x-axis and brightness temperature thresholds along the y-axis. The left panel shows the FSS when a large neighborhood size is used (N=10), whereas the right panel uses a smaller neighborhood (N=4). Together, this allows us to assess the forecast accuracy over different spatial scales. All values were computed using data from the July 2014 forecasts. Values greater than approximately 0.5 indicate a skillful forecast. Overall, there is a general decrease in the forecast skill with increasing forecast lead-time for brightness temperatures warmer than 245 K. This is true for both the N=4 and N=10 neighborhood sizes, but it is evident the skill is higher for the larger neighborhood size. This is to be expected given that the large neighborhood size is more forgiving of small spatial displacement errors in the forecast cloud field. The skill increase as the brightness temperature threshold increases simply because more of the domain will be covered. As such, these charts are most useful for assessing errors in the colder brightness temperatures. It is interesting to note that there is a slight decrease in forecast skill in the 12-h forecast (second column from left) that is then recovered by the 24-h forecast. These changes in forecast skill could be related to model spin-up issues or to the diurnal cycle. Further investigation will be necessary to discern the exact reason.

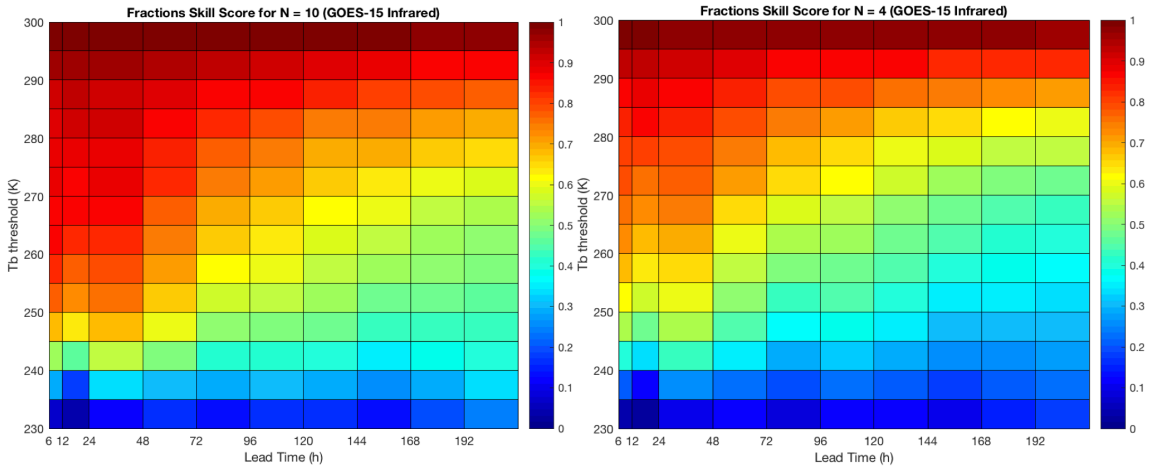

Figure 2. Fractions skill score for GOES-15 10.7-μm brightness temperatures (K) for various brightness temperatures thresholds and forecast lead-times using neighborhood sizes of N = 10 (left panel) and N = 4 (right panel).

2.3. Regional model forecast accuracy assessments – tropical convection

We are also using the verification tools described above to better understand sources of error in the simulated moisture and cloud fields in regions containing different cloud regimes. For example, the scatterplots shown in Fig. 1 depict large errors around the 295 K threshold. We hypothesize that these errors are, in part, due to spatial errors associated with deep convective clouds. To test this hypothesis, we examined errors in the ITCZ where vigorous tropical convection is evident in both the observed and simulated satellite imagery (e.g., Fig. 3).

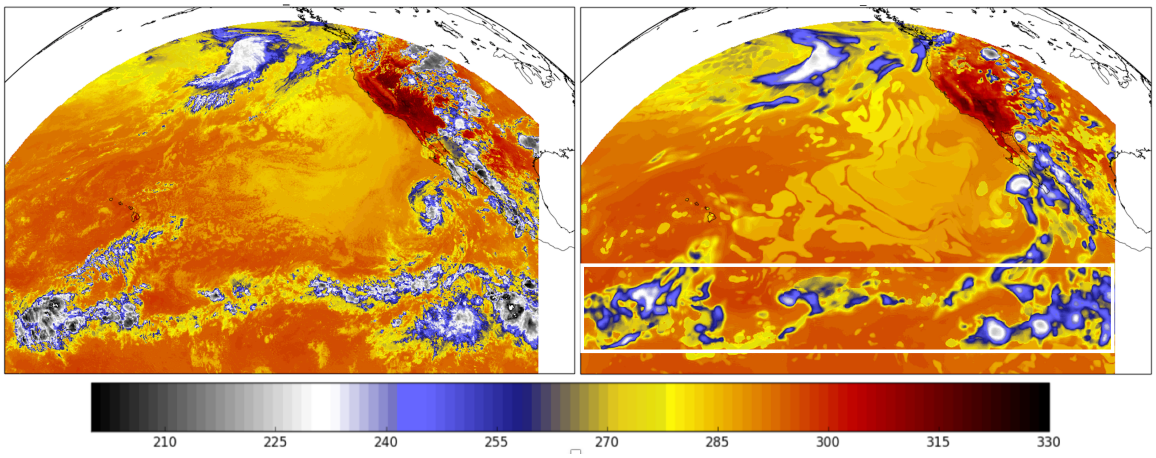

Figure 3. Observed (left) and simulated (right) GOES-15 10.7-μm brightness temperatures (K) for a 24-h forecast valid at 00 UTC 4 July 2014. The white box on the right-hand panel indicates the region examined in this section.

The left panel of Fig. 4 shows the GOES-15 6.7 μm brightness temperature scatterplots valid at the 24-h forecast lead-time for the latitude band of 2-14.5°N that corresponds to the ITCZ during July 2014. The scatterplot shows a sharp horizontal error band for the ~295 K simulated brightness temperatures that can be attributed to displacement errors as well as to general misses and false detections of specific convective clouds. There is a broader vertical error band associated with observed 285-295 K brightness temperatures that is consistent with the failure of the GFS model to produce the sharp gradients in brightness temperatures seen in the observed clouds. This should not be surprising since the GFS is a global model that cannot adequately capture small-scale convective details. As was seen in Fig. 1, the scatterplot depicts a warm bias in the simulated brightness temperatures. Looking at the 24-h forecast example shown in Fig. 3, the behaviors in the scatterplot make intuitive sense because the simulated convection contains less small-scale detail and the anvils in the vicinity of active convective cores are not as cold as in the observations. Scatterplots for other forecast hours show similar horizontal and vertical bands of enhanced probabilities associated with spatial cloud errors.

The right panel of Fig. 4 shows the FSS with respect to lead-time and various brightness temperature thresholds using a neighborhood size of $N=4$. Overall, there is some periodic behavior in the FSS with respect to lead-time, but it is possible that the oscillatory behavior would disappear if more forecast dates were available. The decrease in FSS with lead-time seen for the entire domain (right panel of Fig. 2) is not apparent in the ITCZ region. Considering only the earlier lead-times, the entire domain has higher FSS than in

the ITCZ, indicating that the GFS, despite its ability to produce an ITCZ, is limited in its ability to properly predict the location and spatial distribution of convection within the ITCZ at all forecast lead-times.

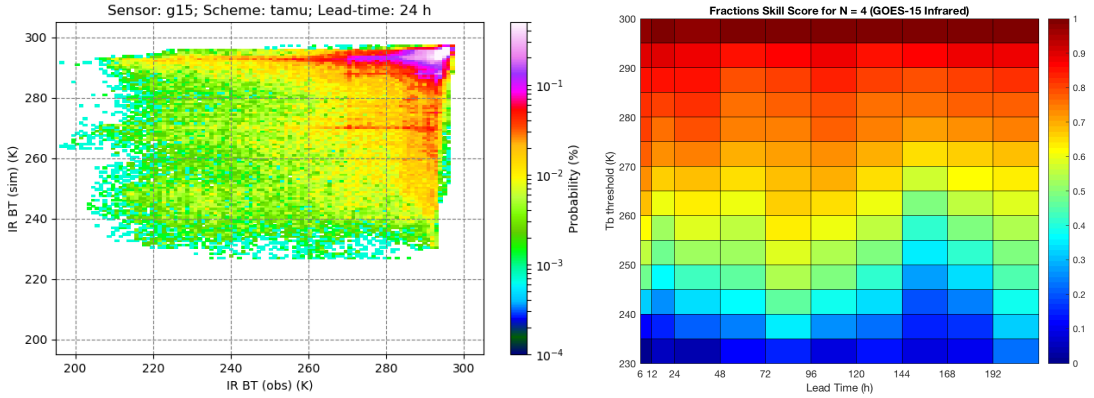

Figure 4. (left panel) scatterplot of simulated vs. observed GOES-15 10.7-μm brightness temperatures (K) and (right panel) the FSS computed with respect to forecast lead-time (h) along the x-axis and brightness temperature (K) threshold along the y-axis. Both panels are for the latitudinal band of 2-14.5°N indicated by the white box in Fig. 3.

2.4. Regional model forecast accuracy assessment – maritime stratocumulus clouds

This section presents results from an analysis of the forecast errors within the persistent stratocumulus cloud region in the eastern Pacific located to the southwest of California and Mexico. As indicated in Fig. 5, a distinct error pattern is present in these clouds when a color bar with a limited range is used to plot the image. In particular, there is a distinct wave-like pattern in the stratocumulus clouds that oscillates between warmer and colder brightness temperatures roughly on a southwest-to-northeast transect. It is also evident that the southward extent of the colder brightness temperatures within this region is more limited than that indicated by the observations. This implies that the top of the boundary layer is lower than what actually occurred within the southern part of the stratocumulus cloud region. This behavior stands in sharp contrast to the larger westward extension of the forecast stratocumulus cloud region, where the more homogeneous appearance of the clouds suggests that boundary layer mixing is insufficient in that region. Lastly, this 6-h forecast is also deficient in low-level clouds across the east-central Pacific to the north of the stratocumulus region. Together, these results indicate that there are substantial errors in the representation of low-level maritime stratocumulus clouds in the GFS model.

To understand the cause of the wave-like structure, the GFS simulated fields of integrated cloud water content and cloud top pressure are shown in Fig. 6. The cloud top pressure shows a discrete, stair stepping pattern of a cloud top boundary layer increasing in height from northeast to southwest. The constant cloud top pressure within each north-to-south band corresponds to an individual sigma level in the GFS model and clearly depicts the limitation of the model vertical resolution when it comes to resolving the boundary layer height and more importantly cloud features occurring at the top of the boundary layer. The effect of the discrete vertical representation of the boundary layer is also evident in the majority of the forecasts and forecast lead times (not shown). Preliminary inspection of the results indicates that the minimum cloud water values along the edge of each band

are likely due to the conversion of cloud water to rain at the boundary between adjacent stratocumulus clouds with cloud tops occurring on different sigma levels. The transfer of cloud water to rainwater along this boundary decreases the optical depth of the clouds, thereby resulting in warmer infrared brightness temperatures. This pattern is not evident in other cloud types so it suggests that the errors in cloud top height and cloud condensate are peculiar to this particular cloud type. Given that these clouds are strongly influenced by boundary layer processes, the errors indicate potential limitations in the cumulus or planetary boundary layer parameterization schemes. These results also highlight the ability of satellite-based model verification methods to identify errors in conventional observation-sparse regions. For the next reporting period, we will use both infrared and microwave brightness temperatures to more closely evaluate errors in the stratocumulus cloud field and to try to identify model processes contributing to their development.

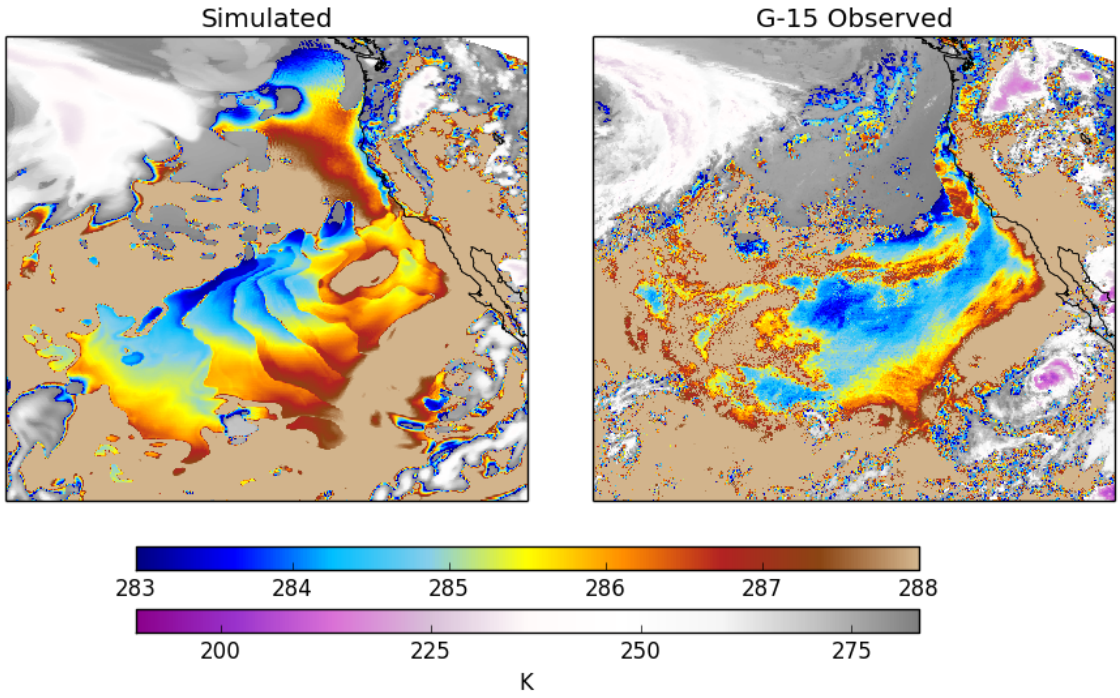

Figure 5. Simulated (left panel) and observed (right panel) GOES-15 10.7 μm brightness temperatures valid at 0600 UTC on 03 July 2014. The simulated brightness temperatures are from a 6-h forecast.

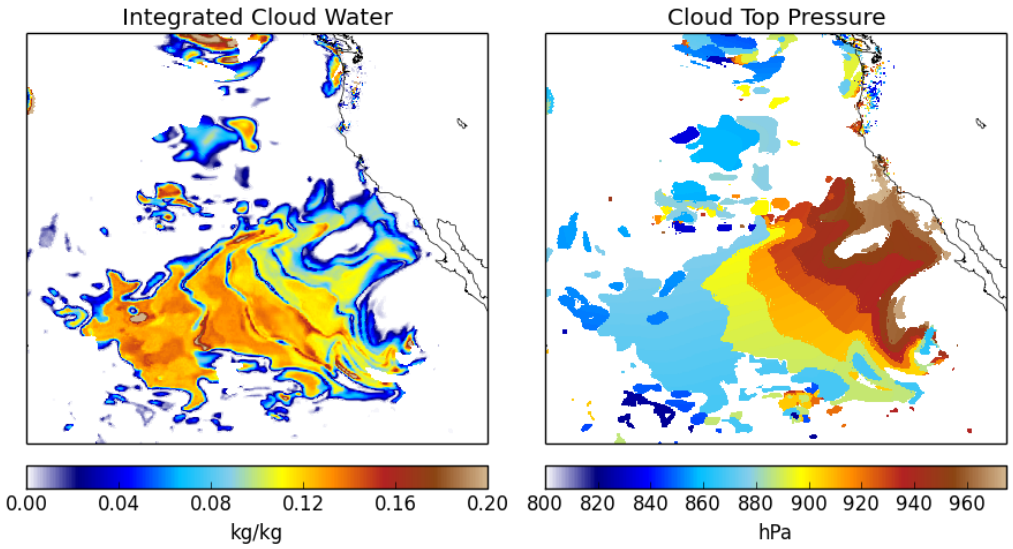

Figure 6. Integrated cloud water content (left panel) and cloud top pressure (right panel) from a 6-h forecast valid at 0600 UTC on 03 July 2014.

3. ISSUES DELAYING CURRENT OR FUTURE PROGRESS

None.

4. INTERACTIONS WITH EMC AND OTHER NOAA-FUNDED SCIENTISTS

During the past 6 months, we have had several conversations with researchers at EMC and the DTC to discuss the model simulations we are using during this project. We are also participating in the NGGPS telecons in order to stay abreast of recent research performed by other groups. Lastly, we have had several email exchanges with researchers at GFDL concerning the availability of FV3 model output. Based on these conversations, we anticipate that we will be able to start assessing the accuracy of model forecasts using the FV3 dynamic core during the next reporting period.

5. CHANGES IN PROPOSED PROJECT

None.

6. OUTCOMES TRANSITIONED TO OPERATIONS

No outcomes have been transitioned to operations during this reporting period.

7. BUDGET ISSUES

None.

8. PRESENTATIONS

No presentations have been given during this reporting period.

9. JOURNAL ARTICLES

None.

10. REFERENCES

Roberts, N. M., 2008: Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteor. Appl.*, **15**, 163-169.

Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78-97.