GFSv16 Evaluation

STI SOO GFSv16 Evaluation Team

17 September 2020
MEG STI GFSv16 Evaluation Team

- **Purpose:** Help evaluate readiness of GFSv16 for implementation from the perspective of the NWS Regions and Centers
- **Team Members**

<table>
<thead>
<tr>
<th>Team Member</th>
<th>Region/Center</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geoff Manikin (EMC)</td>
<td>Warren Blier (Western Region)</td>
</tr>
<tr>
<td>Chris Karstens (SPC)</td>
<td>Mike Fowle (Central Region)</td>
</tr>
<tr>
<td>Mark Klein (WPC)</td>
<td>Bill Martin (Eastern Region)</td>
</tr>
<tr>
<td>Steverino Silberberg (AWC)</td>
<td>Emily Niebuhr and David Levin (Alaska Region)</td>
</tr>
<tr>
<td>Ben Trabing &amp; Brian Zachry (NHC)</td>
<td>Jack Settelmaier (Southern Region)</td>
</tr>
<tr>
<td>Bob Ballard (Pacific Region)</td>
<td></td>
</tr>
</tbody>
</table>
Background

• A major outcome of the National 2015 SOO/DOH Meeting was to have SOOs and DOHs contribute to national and regional projects that support Weather Ready Nation goals.

• The STI leadership team oversees these projects/teams and solicits volunteers to join the teams for approved projects (D. Myrick, National SOO).

• The MEG has been invoking SOO/DOH teams to assist with evaluation activities since 2016, as the expertise of forecasters at the local and regional level is invaluable.  

• A team was organized in late spring 2020 to help evaluate GFSv16 and have its members contribute to their Region’s/Center’s official recommendation on the proposed upgrade.
Subjective Assessments

- Each team member was asked to examine GFSv15 and GFSv16 performance for a set of cases relevant to their Region/Center.
- Members were asked to assess performance on 2m temperatures, 2m dew points, QPF, instability, synoptic performance, soundings, and overall utility.
- The forecasts were split into extended range (days 7-10), medium range (days 4-6), and short range (days 1-3).
- For each item in each forecast range, embers were asked to use a rating system with a range of -3 to +3 ( -3 indicating that GFSv15 was clearly better to +3 indicating that GFSv16 was clearly better, with 0 representing no discernible difference).
- In the following tables, we group “as good or better” (0 to +3) together, since in the big picture, it is acceptable for v16 to match the performance of v15.
## 2m Temperature Ratings

<table>
<thead>
<tr>
<th>12z Valid Time</th>
<th>Mean Rating -3 to +3</th>
<th>% GFSv16 as good or better than GFSv15</th>
<th>% GFSv16 worse than GFSv15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extended Range</td>
<td>0.13</td>
<td>71</td>
<td>29</td>
</tr>
<tr>
<td>Medium Range</td>
<td>0.23</td>
<td>82</td>
<td>18</td>
</tr>
<tr>
<td>Short Range</td>
<td>0.23</td>
<td>95</td>
<td>5</td>
</tr>
</tbody>
</table>

- Mean rating shows modest improvements at all time ranges, with slightly larger gains at 00z valid times.

<table>
<thead>
<tr>
<th>00z Valid Time</th>
<th>Mean Rating</th>
<th>% GFSv16 as good or better than GFSv15</th>
<th>% GFSv16 worse than GFSv15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extended Range</td>
<td>0.29</td>
<td>80</td>
<td>20</td>
</tr>
<tr>
<td>Medium Range</td>
<td>0.41</td>
<td>90</td>
<td>10</td>
</tr>
<tr>
<td>Short Range</td>
<td>0.36</td>
<td>87</td>
<td>13</td>
</tr>
</tbody>
</table>

- The high percentages of “as good or better” for all three time ranges reflect that there were a lot of ‘0’ ratings.
## QPF Ratings

<table>
<thead>
<tr>
<th></th>
<th>Mean Rating -3 to +3</th>
<th>% GFSv16 as good or better than GFSv15</th>
<th>% GFSv16 worse than GFSv15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extended Range</td>
<td>0.17</td>
<td>80</td>
<td>20</td>
</tr>
<tr>
<td>Medium Range</td>
<td>0.63</td>
<td>90</td>
<td>10</td>
</tr>
<tr>
<td>Short Range</td>
<td>0.27</td>
<td>85</td>
<td>15</td>
</tr>
</tbody>
</table>

- Mean rating clearly shows that the biggest improvements were seen in the medium range.
- The high percentages of “as good or better” for all three time ranges reflect that there were a lot of ‘0’ ratings in the extended and short ranges.
CAPE Ratings

- Lowest overall ratings were for this parameter.
- The only negative mean rating for any parameter is CAPE in the short range; the still high percentage for “as good or better” in the short range is driven by a significant number of ‘0’ ratings with several ‘-2’, ‘-3’, and “+1” ratings.

<table>
<thead>
<tr>
<th>Range</th>
<th>Mean Rating -3 to +3</th>
<th>% GFSv16 as good or better than GFSv15</th>
<th>% GFSv16 worse than GFSv15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extended Range</td>
<td>0.04</td>
<td>70</td>
<td>30</td>
</tr>
<tr>
<td>Medium Range</td>
<td>0.21</td>
<td>74</td>
<td>26</td>
</tr>
<tr>
<td>Short Range</td>
<td>-0.20</td>
<td>75</td>
<td>25</td>
</tr>
</tbody>
</table>
### Synoptic Ratings

<table>
<thead>
<tr>
<th></th>
<th>Mean Rating -3 to +3</th>
<th>% GFSv16 as good or better than GFSv15</th>
<th>% GFSv16 worse than GFSv15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extended Range</td>
<td>0.35</td>
<td>78</td>
<td>22</td>
</tr>
<tr>
<td>Medium Range</td>
<td>0.59</td>
<td>83</td>
<td>17</td>
</tr>
<tr>
<td>Short Range</td>
<td>0.07</td>
<td>85</td>
<td>15</td>
</tr>
</tbody>
</table>

- Mean rating clearly shows that the biggest improvements were seen in the medium range.

- The high percentages of “as good or better” for all three time ranges reflect that there were a lot of ‘0’ ratings in the extended range and especially in the short range (with a higher number of positive medium range ratings).
Ratings for Overall Utility

- Mean rating clearly shows that the biggest improvements were seen in the medium range, although there is some utility at all ranges.

- The high percentages of “as good or better” for all three time ranges reflect that there were a lot of ‘0’ ratings in the extended range and short ranges (with more positive medium range ratings).

<table>
<thead>
<tr>
<th></th>
<th>Mean Rating</th>
<th>% GFSv16 as good or better than GFSv15</th>
<th>% GFSv16 worse than GFSv15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extended Range</td>
<td>0.34</td>
<td>79</td>
<td>21</td>
</tr>
<tr>
<td>Medium Range</td>
<td>0.57</td>
<td>85</td>
<td>15</td>
</tr>
<tr>
<td>Short Range</td>
<td>0.26</td>
<td>85</td>
<td>15</td>
</tr>
</tbody>
</table>
500 hPa ACC scores show statistically significant improvement in GFSv16 over v15 in the medium range, and the subjective ratings show that this statistical improvement is manifested in the forecast maps the most during this time range.
Mark Klein
Weather Prediction Center
Focus → heavy precipitation events/medium-range forecasting

Main Findings

- Synoptic-scale pattern better handled in v16, particularly in the medium range
- The progressive bias in v15 seems less prevalent in v16
- For the majority of forecast cycles in each case, v16 QPF was an improvement over v15, both with areal coverage and magnitudes.
- Precipitation type forecasts suggest there may be a stronger warm nose in warm advection events (?)
Southeastern U.S. Heavy Rain (Feb 2020)

180-hour forecast from 00Z Feb 5 valid at 12Z Feb 12.

- GFSv16 had an excellent Day 7 forecast
- GFSv15 was too progressive
Southeastern U.S. Heavy Rain (Feb 2020)

132-hour forecast from 00Z Feb 7 valid at 12Z Feb 12.

- Similar trough strength and position, though GFSv16 better handling the northern stream pattern
Southeastern U.S. Heavy Rain (Feb 2020)

108-hour forecast from 00Z Feb 8 valid at 12Z Feb 12.

- GFSv16 correctly had a more consolidated southern stream feature and accurate timing
Southeastern U.S. Heavy Rain (Feb 2020)

84-hour forecast from 00Z Feb 9 valid at 12Z Feb 12.

GFSv16 -- better and more consistent forecasts of the overall synoptic pattern

Most cycles showed a less progressive bias in GFSv16
Southeastern U.S. Heavy Rain (Feb 2020)

180-hour forecast from 00Z Feb 4 valid at 12Z Feb 11.

- Both captured the heavy rainfall potential at Day 7, but GFSv16 was almost spot-on with the axis
Southeastern U.S. Heavy Rain (Feb 2020)

156-hour forecast from 00Z Feb 5 valid at 12Z Feb 11.

- On Day 6, GFSv16 maintained excellent continuity, while GFSv15 showed a much more disorganized rainfall pattern along the front.
Southeastern U.S. Heavy Rain (Feb 2020)

132-hour forecast from 00Z Feb 6 valid at 12Z Feb 11.

- Similar story for Day 5; great continuity in GFSv16 and an unfocused QPF pattern in GFSv15
Southeastern U.S. Heavy Rain (Feb 2020)

84-hour forecast from 00Z Feb 8 valid at 12Z Feb 11.

GFSv16 captured the heavy rainfall potential at F216, earlier than GFSv15

GFSv16 showed excellent run to run consistency

Tended to be too far north with the precipitation axis in the short to early medium range period
Just one example, but for nearly all short and medium range cycles of this case, GFSv16 produced significantly better QPF, in part due to higher PWs in the atmospheric river.
QPF Verification Statistics

ETS Jun 12 2019 - Sep 12 2020

Higher threat scores at low thresholds

Improved lower bias for light amounts and slightly better bias for heavier thresholds
Snow (blue diamonds), sleet (purple triangles), and freezing rain (red squares) for the 12Z 12/30 - 12Z 12/31 period

Forecasts valid 18Z Dec 30, 2019

GFSv16 caught onto this event about 36 hours ahead of GFSv15

GFSv16 seemed to overforecast sleet coverage
84-hour forecast valid 12Z 12/30.

Note the warm nose in v16 that resulted in the sleet sounding.
Bill Martin
NWS Eastern Region
SOO, WFO Greenville/Spartanburg, SC
On balance, GFSv16 is an improvement over GFSv15, but the improvement is not huge.

In some cases, GFSv16 is much better

In some cases, GFSv15 is better

Some examples from 120 hour forecasts of 500mb heights, MSLP, and surface T2m. 120 hours is chosen as that is close to the point where skill begins to rapidly drop off in forecast models.
Hurricane Florence, Sept. 2018

A fairly dramatic win for v16 at F120.
GFSv16 also caught on to Hurricane Michael (Oct. 2018) earlier than GFSv15 in the 5 day forecast.
Newfoundland Cyclone Jan 2020. F120 valid 00z Jan 2020. Minimum contour is 504 dam for v15, 498 for v16, and 504 for verification. V15 somewhat better for this case, with V16 too strong and too far north.
Dec. 2019 N. American "Bomb" T2m.

Note that T2m in v16 tends to have a sharper, more realistic gradient over the ocean.

This leads to better temperatures in eastern coastal areas for these events.
Another cold case, this one F120 for Jan. 2020 with cold front all the way to the southern tier of states.

V16 has the front further, but with a better temperature gradient along the front. This is partly good, but makes things too cold just behind the front as it is too progressive.
SUMMARY

Generally prefer v16 over v15 as it improved things more than it hurt them, though it is not difficult to find cases where v15 is better in some way.

V15 and V16 are different enough that having both of them sometimes provided a limited multi-model ensemble with decent dispersion and verification somewhere between the two.
Summary

• Focused on Day 3-8 forecasts

• Improvements in the short- and mid-range forecasts, mostly in low-level fields.
  • Better handling of baroclinic zones, particularly warm front positioning, but tendency in deep toughing to pull warm sector too far north in both versions in extended range (*limited sample)
  • Smaller values of SBCAPE in the warm sector, and spatial coverage of warm sector appears smaller/refined.
  • May result in improvements/refinements of our extended outlook probability delineations on Days 4/5, and perhaps on Days 3 and 6.

• Similar synoptic evolution between versions
  • Cold season event too progressive, warm season events not progressive enough.
  • v16 randomly captures features quite well in medium/extended range.
20 May 2019: High Risk/Southern Plains

Day 5
20 May 2019: High Risk/Southern Plains

Day 4

SPC Day 1 Outlook and Prelim. Reports Valid:
2000 UTC 05/20/2019 to 1200 UTC 05/21/2019
20 May 2019: High Risk/Southern Plains

Day 3

SPC Day 1 Outlook and Prelim. Reports Valid: 2000 UTC 05/20/2019 to 1200 UTC 05/21/2019

GFSv15 fcst init. 06Z 16 May 2019 valid 00Z 21 May 2019 (F72)

2-m T (F)

GFSv15 fcst init. 00Z 16 May 2019 valid 00Z 21 May 2019 (F72)

2-m T (F)

RAP analysis valid 00Z 21 May 2019

2-m T (F)
20 May 2019: High Risk/Southern Plains

Day 5
20 May 2019: High Risk/Southern Plains
20 May 2019: High Risk/Southern Plains

Day 3
28 May 2019: Northern Missouri

Day 5

28 May 2019: Northern Missouri

Day 5
28 May 2019: Northern Missouri

Day 5
2 March 2020: Tennessee Tornadoes
2 March 2020: Tennessee Tornadoes
Summary

• Focused on Day 3-8 forecasts

• Improvements in the short- and mid-range forecasts, mostly in low-level fields.
  • Better handling of baroclinic zones, particularly warm front positioning, but tendency in deep toughing to pull warm sector too far north in both versions in extended range (*limited sample)
  • Smaller values of SBCAPE in the warm sector, and spatial coverage of warm sector appears smaller/refined.
  • May result in improvements/refinements of our extended outlook probability delineations on Days 4/5, and perhaps on Days 3 and 6.

• Similar synoptic evolution between versions
  • Cold season event too progressive, warm season events not progressive enough.
  • v16 randomly captures features quite well in medium/extended range.
Mike Fowle
Central Region
SOO - WFO Des Moines, Iowa
Bottom Line Up Front...

Synoptic:

- V16 - *modest to significant improvements* in the synoptic wave pattern - most notable in events in the medium/extended range (D3-D6)
- Improved position surface features, better thermal/moisture profile, QPF, snow forecasts

2M T/Td:

- 2M T: V16 skill was similar (heat) or *slight improvements* (cold/dynamic)
- 2M Td: V16 similar (cold/dynamic) or *slight degradation* dry bias (heat)

Soundings (eyeball test):

- V16 - *improved thermal/moisture structure* in “most” locations (winter)
- V16 - dry bias/overmixed PBL (summer) and included odd sfc inversion at 00UTC
GFSv16 better prediction of the position and amplitude of the 500mb trough.
23 January 2020 - Forecast Hour - F120

GFSv16 better prediction of 850mb thermal structure - important for rain/snow.
GFSv16 better prediction of the position and amplitude of the 500mb trough over the Great Lakes region.
GFSv16 better prediction of the resultant snowfall forecast.
GFSv16 better prediction of the 2m temps over the upper Midwest (reduced cold bias) but too cold over the Gulf Coast states.
GFSv16 not as skillful with 2m dewpoint temps - especially over the Corn Belt (dry bias).
Sounding Analysis

V16 Improvements:

1. Better P-type forecast
2. Improvement in cold bias

V16 Potential Issues:

1. Overmixing PBL, dry bias
2. Strange 00UTC inversion development
GFSv16 better prediction of the thermal/moisture structure - more accurate precip type.
GFSv16 better prediction of the thermal/moisture structure - reduced cold bias
GFSv16 “odd” sfc inversion at 00UTC - and dry bias.
GFSv16 “odd” sfc inversion at 00UTC.
Example of GFSv16 overmixing the PBL and resulting dry bias.
Final Thoughts...

● Purely “observational” - eyeball test
● Others computing “statistical metrics”
● Regional view - local details matter!!

Final Scoring:

● **Overall - V16 an improvement over V15 in a composite sense**
  ○ Better over more spatial areas, more fields, more run cycles
● “Nudged” the needle forward
● A few issues remain - need to be addressed
Western Region Cases: GFSv16 vs GFSv15

Cases Examined

California Spring Storm (May 2019)
San Francisco Heat (June 2019)
Hurricane Lorena (Sep 2019)
West Coast Bomb Cyclone (Nov 2019)

Warren Blier
SOO, WFO MTR (San Francisco/Monterey)

Takeaway: Overall very similar performance, aside perhaps from v16 performing a bit poorer than v15 in the extended range and slightly better than v15 in the medium range. Basically a toss up, given the limited and subjective nature of this evaluation (uncertainty larger than magnitude of findings).
California Spring Storm (May 2019): Late season heavy precip event

24-hr totals, 2 separate rounds of heavy precip
California Spring Storm (May 2019)

Total QPF analysis through 00Z 21 May 2019

24-hr snowfall analysis valid 00Z May 17

6-8 inches precip

18-20 inches snow

24-hr snowfall analysis valid 00Z May 21

16-18 inches snow
### California Spring Storm (May 2019): GFSv15 vs GFSv16 Results

<table>
<thead>
<tr>
<th>Forecast Parameter</th>
<th>Extended Range</th>
<th>Medium Range</th>
<th>Short Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>500 mb Height</td>
<td>0</td>
<td>+1</td>
<td>0</td>
</tr>
<tr>
<td>QPF</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
</tr>
<tr>
<td>Snowfall</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>KOAK short-range soundings</td>
<td>N/E</td>
<td>N/E</td>
<td>0</td>
</tr>
<tr>
<td><strong>Overall Utility</strong></td>
<td><strong>+0.5</strong></td>
<td><strong>+1</strong></td>
<td><strong>+0.5</strong></td>
</tr>
</tbody>
</table>

**GFS v15** much better

**GFS v16** much better

<table>
<thead>
<tr>
<th></th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>+1</th>
<th>+2</th>
<th>+3</th>
</tr>
</thead>
</table>

N/E = Not Examined
X = Retrospective Case Studies web site malfunctioned
California Spring Storm (May 2019): GFSv15 vs GFSv16 Examples – 500 mb Z

Extended Range

216 hr

v15 Better

v16

verification

Medium Range

120 hr

v15

v16 Better

verification
California Spring Storm (May 2019): GFSv15 vs GFSv16 Examples - Soundings

GFSv15 Better

GFSv16 Better
Western Region Cases: GFSv16 vs GFSv15

Cases Examined

California Spring Storm (May 2019)

San Francisco Heat (June 2019)

Hurricane Lorena (Sep 2019)

West Coast Bomb Cyclone (Nov 2019)
Key factor was strong low-level offshore flow (500 mb heights and 850 mb Temps were unremarkable).
## Forecast Parameter

<table>
<thead>
<tr>
<th>Forecast Parameter</th>
<th>Extended Range</th>
<th>Medium Range</th>
<th>Short Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>00Z Temp (4 pm PST/5 pm PDT)</td>
<td>-2</td>
<td>+1</td>
<td>0</td>
</tr>
<tr>
<td>12Z Temp (4 am PST/5 am PDT)</td>
<td>-2</td>
<td>+1</td>
<td>0</td>
</tr>
<tr>
<td>2m Dew Point</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Synoptic scale details</td>
<td>-2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>KOAK short-range soundings</td>
<td>N/E</td>
<td>N/E</td>
<td>+0.5</td>
</tr>
<tr>
<td>KVBG short-range soundings</td>
<td>N/E</td>
<td>N/E</td>
<td>-0.5</td>
</tr>
</tbody>
</table>

### Overall Utility

| Overall Utility | -1.5 | +0.5 | 0 |

---

GFS v15 much better

GFS v16 much better

N/E = Not Examined

X = Retrospective Case Studies web site malfunctioned
San Francisco Heat (June 2019)

Extended Range

GFSv16 not capturing coastal warming

GFSv16 not capturing strong offshore pressure gradient
Western Region Cases: GFSv16 vs GFSv15

Cases Examined

California Spring Storm (May 2019)
San Francisco Heat (June 2019)
**Hurricane Lorena (Sep 2019)**
West Coast Bomb Cyclone (Nov 2019)
The remnants of Hurricane Lorena brought locally heavy rain to parts of Arizona on September 22–24. Precipitation peaked at 4 to 6 in (100 to 150 mm) near Phoenix.

On September 23, a supercell thunderstorm produced a brief EF0 tornado in New River in Maricopa County.
## Hurricane Lorena (Sep 2019)

<table>
<thead>
<tr>
<th>Forecast Parameter</th>
<th>Extended Range</th>
<th>Medium Range</th>
<th>Short Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>QPF</td>
<td>-2</td>
<td>+1</td>
<td>0</td>
</tr>
<tr>
<td>Surface CAPE</td>
<td>-2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Synoptic Scale Details</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>KPHX short-range soundings</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Overall Utility</td>
<td>-2</td>
<td>+0.5</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GFS v15</th>
<th>GFS v16</th>
</tr>
</thead>
<tbody>
<tr>
<td>much better</td>
<td>much better</td>
</tr>
</tbody>
</table>

-3 | -2 | -1 | 0 | +1 | +2 | +3 |

N/E = Not Examined
X = Retrospective Case Studies web site malfunctioned
Western Region Cases: GFSv16 vs GFSv15

**Cases Examined**

California Spring Storm (May 2019)
San Francisco Heat (June 2019)
Hurricane Lorena (Sep 2019)

**West Coast Bomb Cyclone (Nov 2019)**
West Coast Bomb Cyclone (Nov 2019)

Historic, unprecedented storm in southwest Oregon & northwest California

- Impacts as of late evening November 26, 2019
- Sustained winds of 85 mph with gusts to 106 mph at Cape Blanco, OR @ 1:15 pm
- Wind gust to 69 mph at Crescent City, CA
- Whiteout conditions reported due to wind and snow from LaPine to Lakeview OR
- High seas (up to 34 foot seas so far)
- Heavy coastal rain/inland and mountain snowfall across southwest OR & northern/central CA (parts of I-5 closed)
- November low pressure records set at Medford OR (981.4 hPa/28.98”) & Eureka CA (984.8 hPa/29.08”)
- Lowest pressure in Oregon: 974.6 hPa/28.78” @ 6:35 pm at Gold Beach Airport
- All-time low pressure record preliminarily set for the state of California (~973.6 hPa/28.75” at Crescent City)
# West Coast Bomb Cyclone (Nov 2019)

<table>
<thead>
<tr>
<th>Forecast Parameter</th>
<th>Extended Range</th>
<th>Medium Range</th>
<th>Short Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>00Z Temp (4 pm PST/5 pm PDT)</td>
<td>-1</td>
<td>+1</td>
<td>0</td>
</tr>
<tr>
<td>12Z Temp (4 am PST/5 am PDT)</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2m Dew Point</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Surface CAPE</td>
<td>-2</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>QPF</td>
<td>-3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Synoptic Scale Details</td>
<td>0</td>
<td>+2</td>
<td>0</td>
</tr>
<tr>
<td>KOAK soundings</td>
<td>N/E</td>
<td>N/E</td>
<td>0</td>
</tr>
<tr>
<td><strong>Overall Utility</strong></td>
<td><strong>-1.5</strong></td>
<td><strong>+0.5</strong></td>
<td><strong>0</strong></td>
</tr>
</tbody>
</table>
240 hr

v15

v16

verification
### Overall Utility Summary from these four WR Cases

<table>
<thead>
<tr>
<th>Case</th>
<th>Extended Range</th>
<th>Medium Range</th>
<th>Short Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>California Spring Storm (May 2019)</td>
<td>+0.5</td>
<td>+1</td>
<td>+0.5</td>
</tr>
<tr>
<td>San Francisco Heat (June 2019)</td>
<td>-1.5</td>
<td>+0.5</td>
<td>0</td>
</tr>
<tr>
<td>Hurricane Lorena (Sept 2019)</td>
<td>-2</td>
<td>+0.5</td>
<td>0</td>
</tr>
<tr>
<td>West Coast Bomb Cyclone (Nov 2019)</td>
<td>-1.5</td>
<td>+0.5</td>
<td>0</td>
</tr>
<tr>
<td><strong>OVERALL</strong></td>
<td><strong>-1</strong></td>
<td><strong>+0.5</strong></td>
<td><strong>0</strong></td>
</tr>
</tbody>
</table>

**Takeaway:** Overall very similar performance, aside perhaps from v16 performing a bit poorer than v15 in the extended range and slightly better than v15 in the medium range. Basically a toss up, given the limited and subjective nature of this evaluation (uncertainty larger than magnitude of findings).
Ben Trabing
National Hurricane Center
Slightly Improved Tracks with Right of Track Bias

GFSv16 had slightly improved track forecasts for days 3-6 but a stronger right of track bias at all forecast times compared to GFSv15.
Hurricane Dorian (2019)

GFSv15 did better with Dorian’s short term track forecasts, but GFSv16 had better 4-7 day forecasts and did a better job forecasting recurvature.
GFSv16 reduced a negative intensity bias and a negative bias in the radii of 34-kt winds beyond day 2. GFSv16 creates more intense and larger hurricanes compared to GFSv15.
Hurricane Michael (2018)

GFSv16 consistently outperformed GFSv15 in forecasting Michael at long lead times.

GFSv16 would have given multiple days of added lead time on Hurricane Michael being a strong hurricane near landfall due to better intensity estimates.
2018-2020 Genesis Verification

All values would be 1 for a perfect model

Compared to GFSv15, GFSv16 has:
- higher probability of detection
- larger critical success index
- lower success ratio

Similar in East Pacific Basin

Courtesy of Dan Halperin
GFSv16 creates more tropical cyclones overall leading to more hits but also more false alarms compared to v15.

Courtesy of Dan Halperin
Steverino Silberberg
Aviation Weather Center
FOCUS

- Ceiling, Visibility, Cloud Cover
- Turbulence
- Icing
Latest METAR data at 1200 UTC 10 Sep 2020

- VFR (Visual Flight Rules)
- MVFR (Marginal VFR)
- IFR (Instrument Flight Rules)
- LIFR (Low IFR)

3000' 5sm
1000' 3sm
500' 1sm
SUMMARY

• Visibility
  • GFSv16 better: CO, Midwest, Southeast
  • Radiation/Advection coastal fog issue over Southeast, Northeast
  • Missing West Coast marine layer – all forecast times – v15 & v16

• Total Cloud
  • v16 better

• Turbulence
  • v16 better with jet speed, VWS, horizontal shear wrt PIREPS

• Icing (850 T)
  • v16 better with cold air over and east of Rockies
Jack Settelmaier
Southern Region
GEFSv12 Eval alum
Southern Region (unique) assigned cases

Tennessee Tornadoes (Mar 2020)
Easter Sunday Severe (Apr 2020)
 Southeast Severe (Apr 2020)
 Southern Plains ENH (May 2020)
 Southern U.S. QPF (Feb 2020)
**Memorial Day Heat (May 2019)**
 Imelda (2019)
 Dorian (2019)
 Michael (2018)
 Florence (2018)

https://www.emc.ncep.noaa.gov/users/meg/gfsv16/retros/
Overall GFSv16 Eval Summary

- For examined retro cases, GFSv16 appears to:
  - exhibit slightly superior skill at early-identification, location, and intensity of features
  - This superiority appears more often than not across all run/projection times, but is most evident at longer lead times (>132hrs)

- Additional output fields, domain views, soundings made available for the retro cases, as well as the real-time parallel views and improved \( d(\text{prog})/dt \) functionality, all GREATLY enhanced the ability to assess and feedback.

- Kudos to Geoff and the larger MEG group for these evaluation improvements! Tremendous UFS/EPIC V&V foundation.
GFSv16 2-m Temperature RMSE: Winter/Spr. 2020

Western US – Winter/Spr. 2020

2-m T RMSE (00Z cycles)

GFSv15
GFSv16

Eastern US – Winter/Spr. 2020

2-m T RMSE (00Z cycles)

GFSv15
GFSv16

GFSv16 has lower RMSE at longer leads

Note: differences outside the outline bars are significant at the 95% confidence level.

GFSv16 has lower RMSE at longer leads

Note: differences outside the outline bars are significant at the 95% confidence level.
Overall GFSv16 Eval Summary

- For examined retro cases, GFSv16 appears to:
  - exhibit slightly superior skill at early-identification, location, and intensity of features
  - This superiority appears more often than not across all run/projection times, but is most evident at longer lead times (>132hrs)

- Additional output fields, domain views, soundings made available for the retro cases, as well as the real-time parallel views and improved d(prog)/dt functionality, all GREATLY enhanced the ability to assess and feedback.

- Kudos to Geoff and the larger MEG group for these evaluation improvements! Tremendous UFS/EPIC V&V foundation.
Summary

- GFSv16 exhibited earlier identification (fhr=228 and 192), but not in between.
- Better intensity at fhr=144, but worse location.
- No sig differences between v16 and v15 for fhr=132 to 0
- (Barely) early v16 jump on heavy rain swath location, then alike
- CONUS US vs Bahamian landfall

Challenges
Dorian--2019

Note: The cone contains the probable path of the storm center but does not show the size of the storm. Hazardous conditions can occur outside of the cone.
Dorian--2019: MSLP

- v16 had early identification AND location of Dorian; v15 did not

Dorian--2019 (F228) Valid: 00Z 09/06/19
Dorian--2019: MSLP

Dorian--2019 (F204)
Valid: 00Z 09/06/19

- However, 24hrs later, neither version showed Dorian
Dorian--2019: MSLP

Dorian--2019 (F192)  
Valid: 00Z 09/06/19

- Now, 12hrs later, v16 returns to show a Dorian-like system, v15 does still does not
Dorian--2019 (F156)
Valid: 00Z 09/06/19

- At this projection, v15 first begins to show Dorian along with v16.
- v16 has slightly better location/intensity
Dorian--2019: MSLP

Dorian--2019 (F144)  
Valid: 00Z 09/06/19

- v16 has a better intensity match, but worse location
- Little noticeable differences from FcstHr=132 to 0 (not shown).
Dorian--2019: Total Rainfall

Dorian--2019 (F240)  
Valid: 00Z 09/06/19

• v16 better at indicating observed higher (yellow=10”) amounts, albeit wrong location, and had storm move into N. FL in this run
Dorian--2019

Note: The cone contains the probable path of the storm center but does not show the size of the storm. Hazardous conditions can occur outside of the cone.
Dorian--2019: Total Rainfall

GFSv15  GFSv16

v16–v15  GFS Anl.

Dorian--2019 (F132)
Valid: 00Z 09/06/19

• By this projection, and closer to event, very similar heavy rainfall swaths, though v16 had more, along N. FL coast, which was observed
Dorian (Bahamian Landfall) -- 2019: MSLP

Dorian -- 2019 (F132)
Valid: 12Z 09/01/19

- 5 days earlier than US landfall
- v16 had earlier indication of intense cyclone near western Bahamas, close to observed
Dorian (Bahamian Landfall)--2019: Sfc Wind Gusts

GFSv15  GFSv16  v16-v15  GFS Anl.

Dorian--2019 (F120)  Valid: 00Z 09/02/19

- Though both positionally similar, v16 better, more intense sfc wind gusts forecast
Dorian (Bahamian Landfall)--2019: MSLP

- Dorian--2019 (F072)
  Valid: 00Z 08/30/19

- 4 days before NW Bahamian landfall, v16 had earlier detection of Dorian
Summary

- Shady Point, OK, just southwest of Fort Smith, AR received 6”+ of rainfall in 24hrs

- As could be seen in helpful d(prog)/dt interface, v16 in general had better location and rainfall intensity performance, at a majority of forecast time steps.
24-hr QPE, 12UTC, Sep 1
Flash Flood Emergency—Sep 1, 2020

Flash Flood (F174)
Valid: 12Z 09/01/20

- v16 better (higher amounts, early indication) of the rainfall amounts
- d(prog)/dt loop options
Flash Flood Emergency—Sep 1, 2020

Flash Flood (F120)
Valid: 12Z 09/01/20

- v16 again higher amounts in eastern OK
For examined retro cases, GFSv16 appears to:

- exhibit **slightly superior skill** at early-identification, location, and intensity of features

- This superiority appears more often than not across all run/projection times, but is **most evident** at longer lead times (>132hrs)

- **Additional output fields, domain views, soundings** made available for the retro cases, as well as the real-time parallel views and improved d(prog)/dt functionality, all GREATLY enhanced the ability to assess and feedback.

- **Kudos** to Geoff and the larger MEG group for these evaluation improvements! Tremendous UFS/EPIC V&V foundation.
Emily Niebuhr
David Levin
Jason Ahsenmacher
Alaska Region
think the best example is by far the remnants of Bavi,

- Remnants of typhoons and subtropical diabatic anomalies play a very large role in fall across the Bering Sea and Alaska mainland.
- A recurving typhoon over Japan can be in the Bering Sea waters by day 3, and have large downstream impacts over the Mainland shortly thereafter.
- Historically, this was a large forecast problem for the GFS15 and previous versions of the GFS for Alaska. Oftentimes, these recurving typhoons and/or strong diabatic anomalies (usually some sort of MJO wave) do not result in huge hurricane force lows, but are much more typically a phasing event with the "northern stream" polar flow. Over water phasing events are extremely challenging for numerical weather models to simulate,
MSLP: Operational GFS vs GFSv16

- GFSv16 was consistent with a track much closer to reality from about F120 onward
- Operational GFS trended much to far south then had to correct at the last minute
- GFSv16 was much less “jumpy” overall with the low position and strength
- Resulted in coastal flooding in Bristol Bay and hurricane force winds along Dutch Harbor
Improvements

- The old version 15 GFS showed what has often been a significant issue with the GFS (and other models, as well), and that is "flip-flopping" back and forth, which results in poor confidence and hard to develop forecast messaging.

- In this case, what made the performance of the new version 16 FV3 particularly impressive was the amazingly stable dprog/dt, which, for 120 hours (20 model runs) properly simulated a Bristol Bay landfall instead of a Gulf of Alaska low.

- While this may sound trite given how close they are, this is a common model error (as seen in the version 15 GFS), as the terrain along the Alaska Peninsula dramatically influences triple point low development given the high arc of volcanic mountains in excess of 5000-9000 feet.

- In zonal flow, models have historically suffered large errors in low tracks as they historically have struggled with simulating whether a low moves into Bristol Bay (less common) or form a triple point along/south of the Alaska Peninsula and reform a low in the Gulf of Alaska (much more common).
Temperatures: Operational GFS vs GFSv16

- GFSv16 was consistently warmer than the operational GFS from 850mb and below over the interior of the state and also in SW Alaska west of the Alaska Range.
- However when you look at upstream soundings from Bethel and King Salmon...
Soundings: Operational GFS vs GFSv16

- V16 seems to be handling the boundary layer temperatures better overall than the operational GFS
- The operational GFS better captures the depth of the cold airmass under the upper trough to the west.
- It's possible that the deeper colder airmass in the operational GFS contributed to the broader and more expansive northern stream trough and the more suppressed southern stream storm track.

Same features appear at Bethel. V16 is warmer at the lowest level but shallower with the depth of the cold air.
Delta Junction Wind Storm

Fairbanks Forecast Office 2020 Sept 7
1 km WRF Run to capture gap flow- how does GFS16 capture mountain wave dynamics?
Red color - wind more northerly in GFS16. Blue indicates winds are more southerly than current GFS.
Robert Ballard
Pacific Region &
Central Pacific Hurricane Center
Fujiwhara Effect/Binary Interactions

GFSv16 12Z 28 June 2019 valid 00Z 06 July 2019

GFSv16 12Z 17 September 2019 valid 12Z 20 September 2019

Barbara (2019) case

Vicky (2020) case

3 situations where binary interaction was not definitively observed, nor depicted similarly in GFSv15

Lorena (2019) case
Right-of-track tendency

TC Erick (F90)  TC Sarai (F120)
Slow forward speed tendency

Barbara (F132)

Erick (F84)
Larger radii & more intense TC tendency
(likely related to other artifacts?)

TC Sarai (F150)
TC Genesis

GFSv16 wins!
- Kiko, Flossie, Douglas, Wasi, Bavi

GFSv15 wins
- Barbara, Erick, Juliette

Tie
- Lorena, Haishen, Sarai, Maysak

Unknown
- Vicky
Hurricane Douglas

At times the GFSv16 was about two days ahead of the trend in GFSv15.
TC Sarai
240 hour forecast

Amazing guidance for a S Pac tropical cyclone!