Scientific Verification/Validation Methodology at GFDL

Linjiong Zhou and The GFDL FV^3 Team

NEMS FV3GFS Community Modeling System Training and Tutorial, 07/19/2017

Verification/Validation Philosophy

- **Efficient** and **comprehensive** verifications guide the model development and tuning activities at GFDL.
- The FV^3 team is composed of people with various research backgrounds, so the verification subjects are **extensive**.
- We conduct model verification based on our **research backgrounds** because model cannot simulate everything perfectly.
- We built the verification tools by ourselves so as to **match our requirements**, such as priority, algorithms, work flow, plots, data format, efficiency, etc.
- Model developers and verification team work **seamlessly** to boost the processes of model development.

Verification/Validation at GFDL

- Basic Scores: ACC/Bias/RMSE (Linjiong, Xi, Jan-Huey, Shannon)
- Precipitation Scores: ETS/BS/FSS (Bill, Linjiong)
- Climatology (multi-cases mean) (Linjiong)
- Radar Reflectivity/CAPE/Updraft Helicity (Lucas)
- Hurricane Track/Intensity (Jan-Huey, Morris, Tim, Kun)
- Hurricane Structure (Andy, Kun)
- Model Climatology & Biases (Baoqiang)
- Real-time Forecast (Matt, Shannon)
- Kinetic Energy Spectra (Xi)

Basic Scores

WMO Definition:

$$ACC = \frac{\overline{\left[(f-c)-\overline{(f-c)}\right]\cdot\left[(a-c)-\overline{(a-c)}\right]}}{\sqrt{\overline{\left[(f-c)-\overline{(f-c)}\right]^2}\cdot\overline{\left[(a-c)-\overline{(a-c)}\right]^2}}}$$

Anomaly Correlation Coefficient (ACC) is written as

$$ACC = \frac{\overline{(f-c)\cdot(a-c)}}{\sqrt{\overline{(f-c)^2}\cdot\overline{(a-c)^2}}}$$

Bias ($BIAS$) is written as

$$BIAS = \overline{f-a}$$

Root Mean Sqaure Error ($RMSE$) is written as

$$RMSE = \sqrt{\overline{(f-a)^2}}$$

95% Confidence Level is defined as

$$DEV = \alpha\frac{\sigma}{\sqrt{n}}$$

Where f is forecast, a is analysis, c is climatology, σ is standard deviation, n is the sample size, α is the critical t-test value. We need to consider laitude weighting ($w = \cos\phi$) when they are applied to the Earth.

Datasets

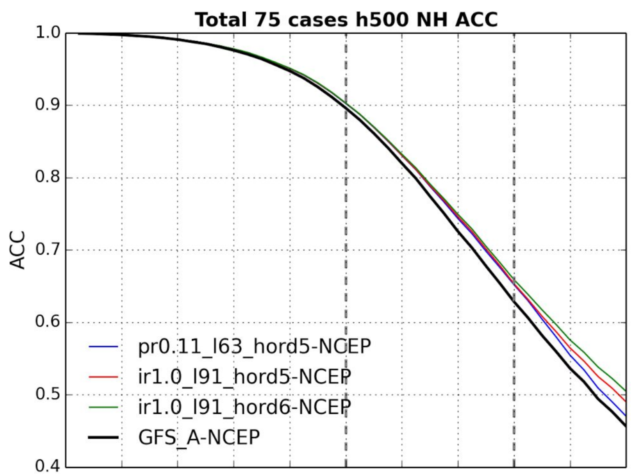
- NCEP Analysis (or ERA-Interim, MERRA)
- NCEP Climatology
- Operational GFS Forecast
- fvGFS Forecast

Resolution: 2.5 degree, 6 hourly

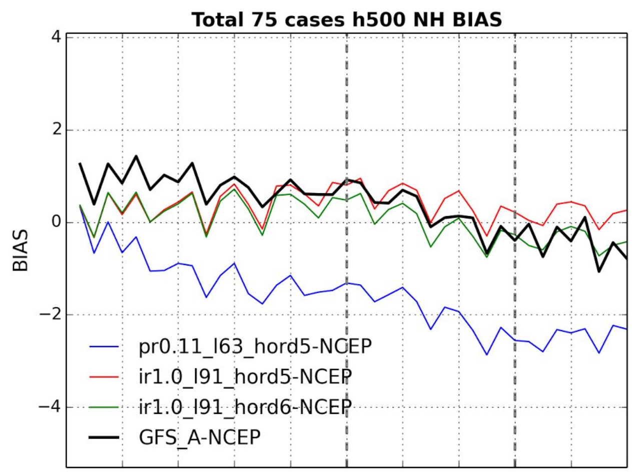
Domain: Global (90S-90N), Northern Hemisphere (20N-80N), Southern Hemisphere (20S-80S), Tropics (20S-20N)

Time Period: From 01/16/2015 to 01/16/2016 (or present)

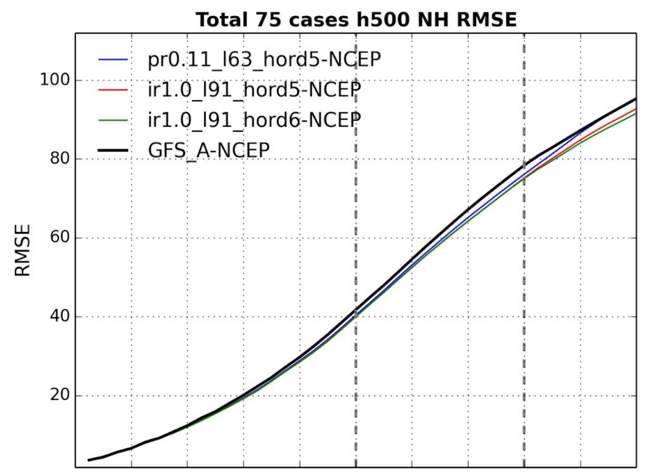
Die-off



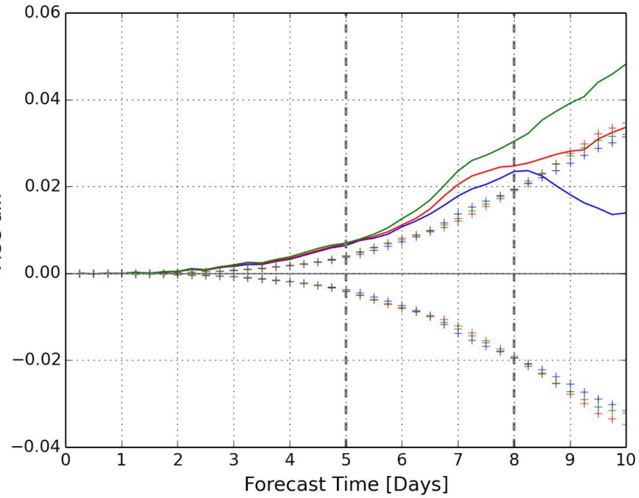
Total 75 cases h500 NH ACC
ACC
1.0
0.9
0.8
0.7
0.6
0.5
0.4
pr0.11_l63_hord5-NCEP
ir1.0_l91_hord5-NCEP
ir1.0_l91_hord6-NCEP
GFS_A-NCEP



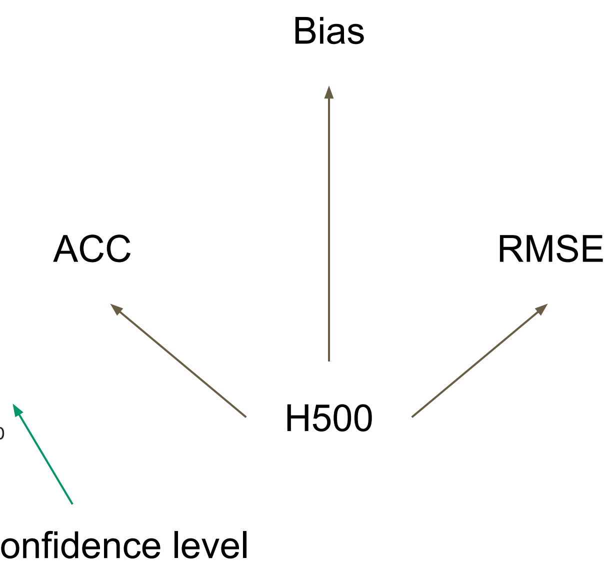
Total 75 cases h500 NH BIAS
BIAS
4
2
0
−2
−4
pr0.11_l63_hord5-NCEP
ir1.0_l91_hord5-NCEP
ir1.0_l91_hord6-NCEP
GFS_A-NCEP



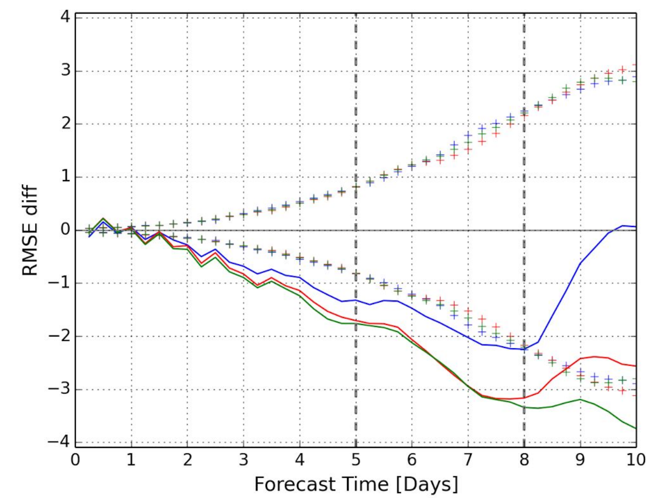
Total 75 cases h500 NH RMSE
RMSE
100
80
60
40
20
pr0.11_l63_hord5-NCEP
ir1.0_l91_hord5-NCEP
ir1.0_l91_hord6-NCEP
GFS_A-NCEP



0.06
0.04
0.02
0.00
−0.02
−0.04
0 1 2 3 4 5 6 7 8 9 10
Forecast Time [Days]

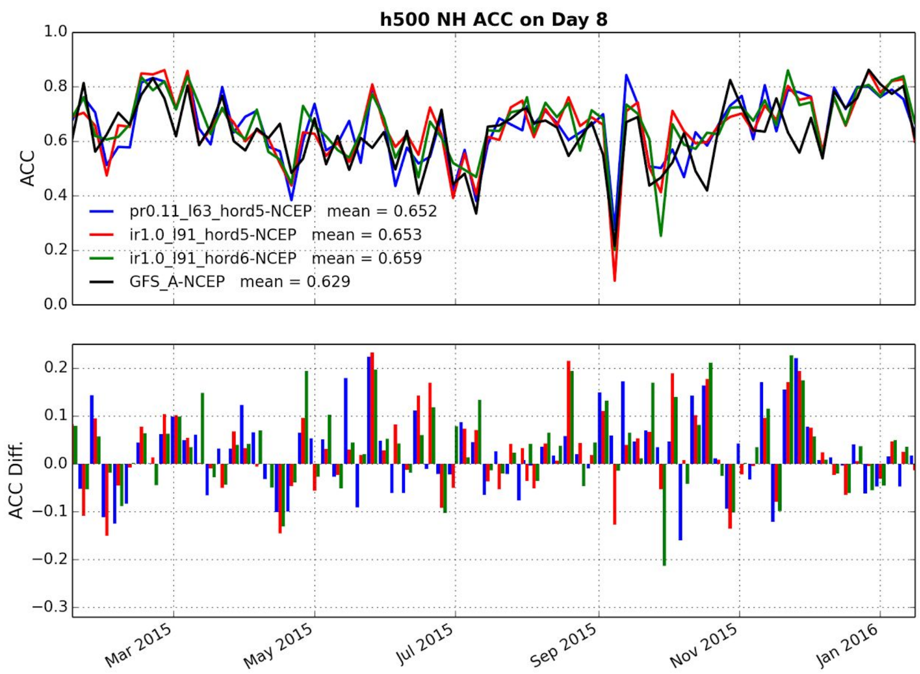


Bias
ACC
RMSE
H500
95% confidence level

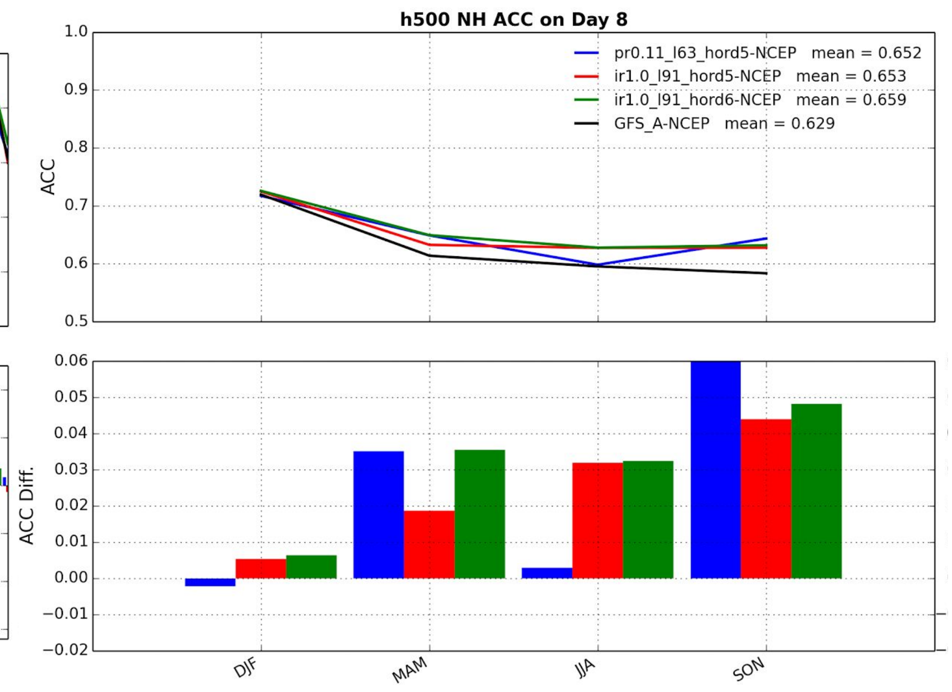


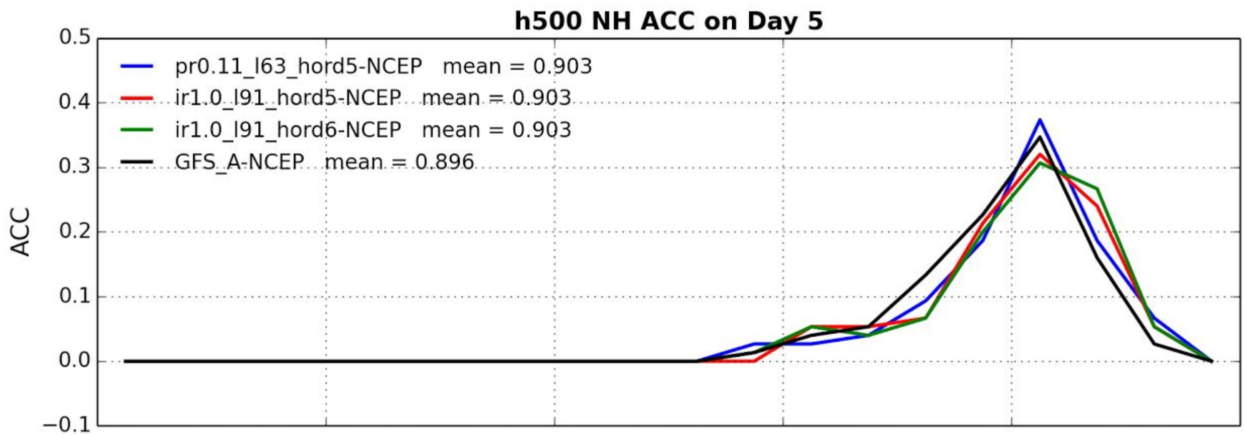
RMSE diff
4
3
2
1
0
−1
−2
−3
−4
0 1 2 3 4 5 6 7 8 9 10
Forecast Time [Days]

Time Series



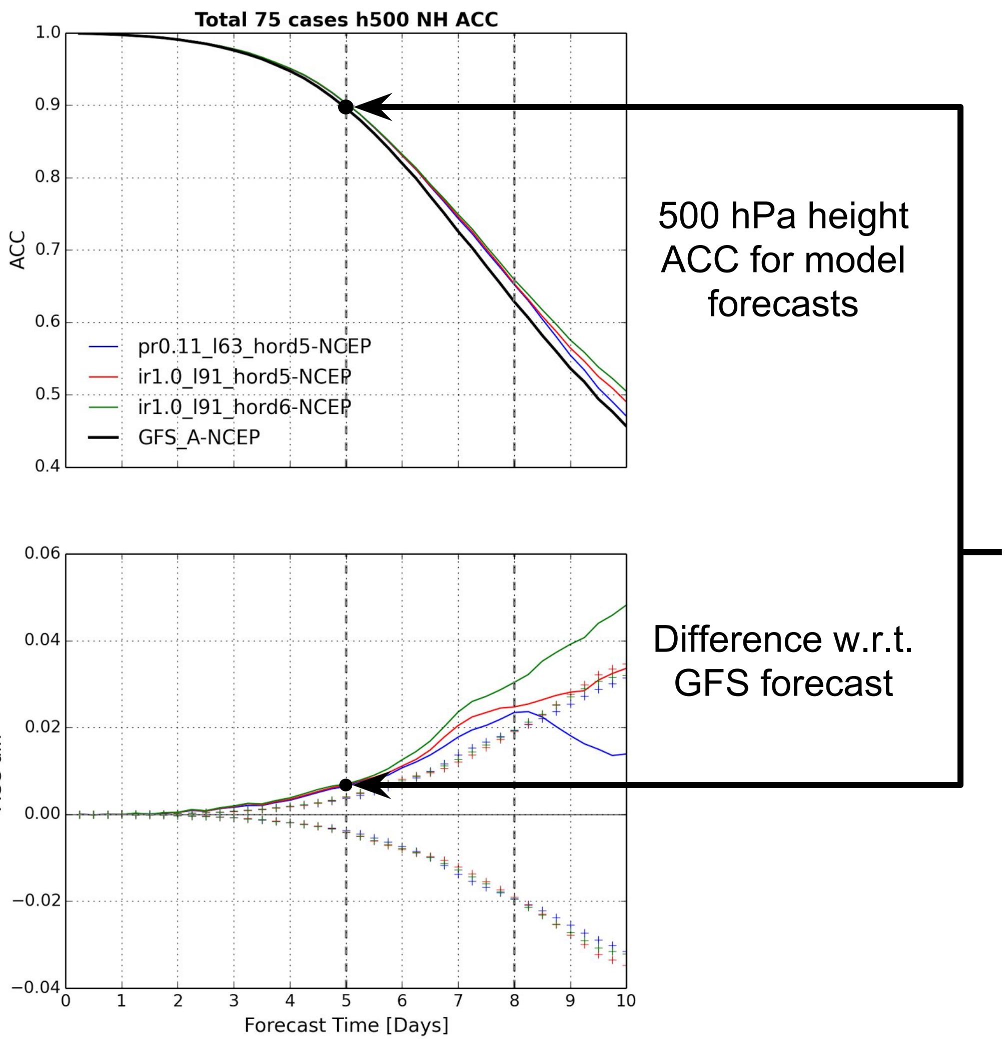
Four Seasons





Probability Distribution

Space-based vs. Time-based Statistics



To obtain the left-hand-side plots, there are two methods:

1) Compute statistics (RMSE, STD, ACC, et al.) on particular domain then do average on time series

2) Compute statistics (RMSE, STD, ACC, et al.) on time series then do average on particular domain

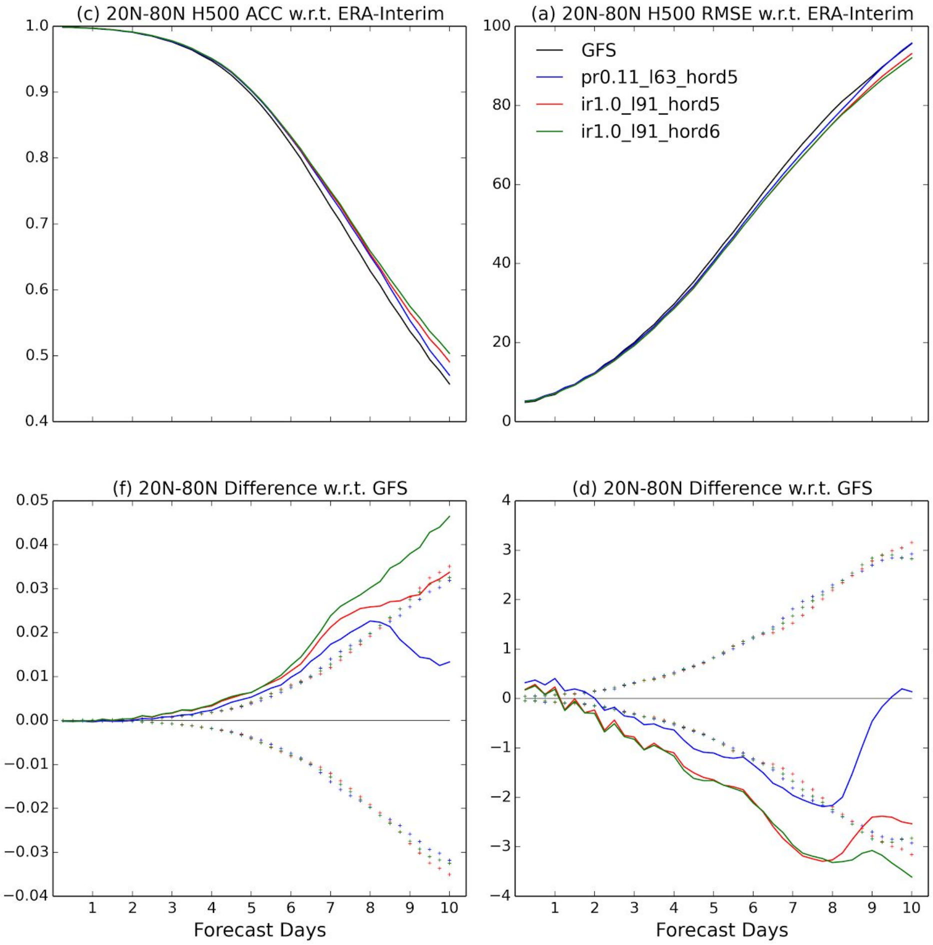
Take day 5 forecast as an example, we have 75 cases started from every 5 days on 20°N-80°N domain.

Method 1) computes the ACC for all grid boxes for each case, then do the average for all 75 cases. (Space-based Statistics)

Method 2) computes the ACC for 75 cases on each grid box, then do the average for all grid boxes. (Time-based Statistics)

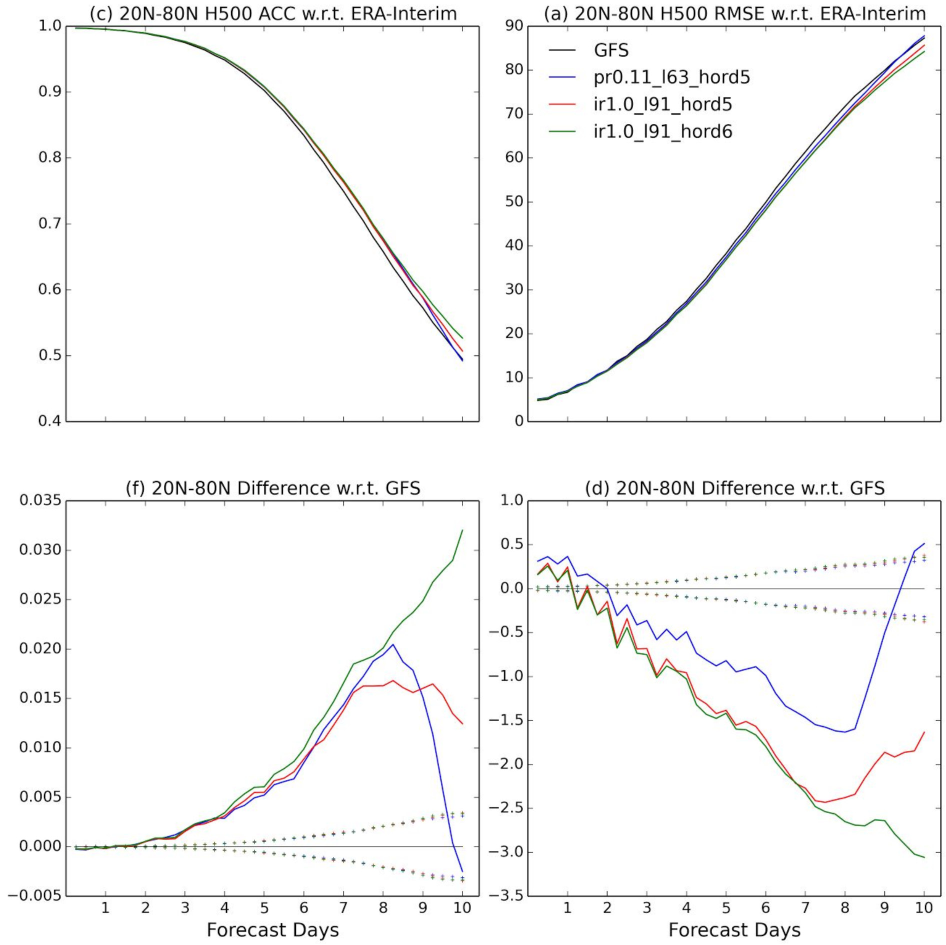
Space-based statistics are the one we are usually using.

Space-based Statistics



ACC RMSE

Time-based Statistics



ACC RMSE

Precipitation Scores

Equitable Threat Score (ETS) is written as

$$ETS = \frac{a - ar}{a + b + c - ar}$$

ETS measures how well did the forecast "yes" events correspond to the observed "yes" events. $ETS = 1$ means a perfect forecast. $ETS \leq 0$ means no skill.

Bias Score (BS) can be written as

$$BS = \frac{a + b}{a + c}$$

BS measures how similar were the frequencies of "yes" forecasts and "yes" observations. BS indicates whether the forecast system has a tendency to over-forecasts ($BS > 1$) or under-forecasts ($BS < 1$) events. $BS = 1$ means a perfect forecast. $BS = 0$ or infinity means the forecast is useless.

Fraction Skill Score (FSS) can be written as

$$FSS = 1 - \frac{FBS}{FBS_{worst}}$$

FSS ranges between 0 and 1, with 0 representing no overlap and 1 representing complete overlap between forecast and observed events, respectively.

Datasets

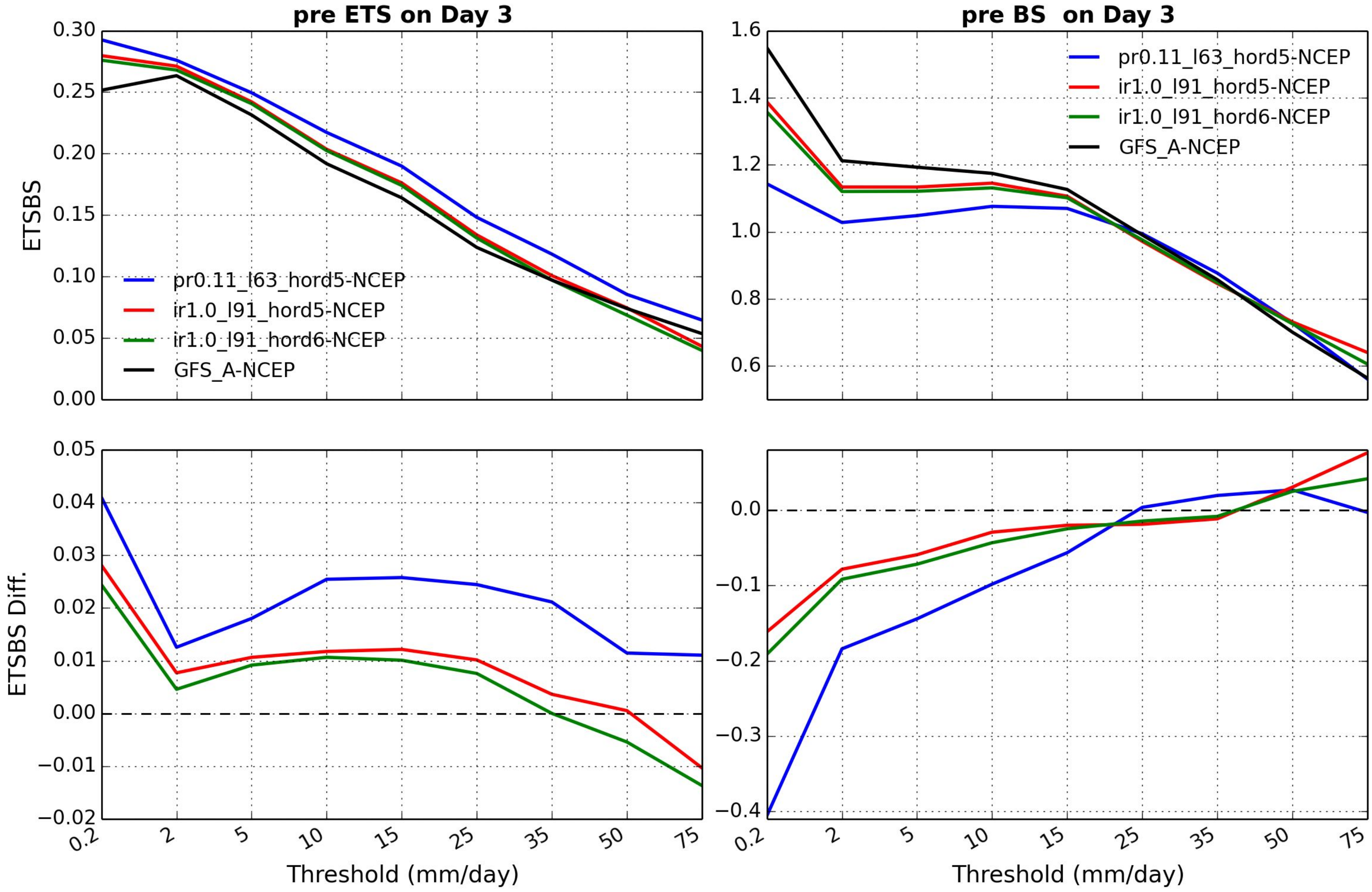
- Stage IV
- Operational GFS Forecast
- fvGFS Forecast

Resolution: 12 km, 6 hourly

Domain: CONUS (25N-50N, 125W-65W)

Time Period: From 01/16/2015 to 01/16/2016 (or present)

ETS and BS


pre ETS on Day 3
pre BS on Day 3
ETSBS
ETSBS Diff.
pr0.11_l63_hord5-NCEP
ir1.0_l91_hord5-NCEP
ir1.0_l91_hord6-NCEP
GFS_A-NCEP
Threshold (mm/day)
0.2
2
5
10
15
25
35
50
75

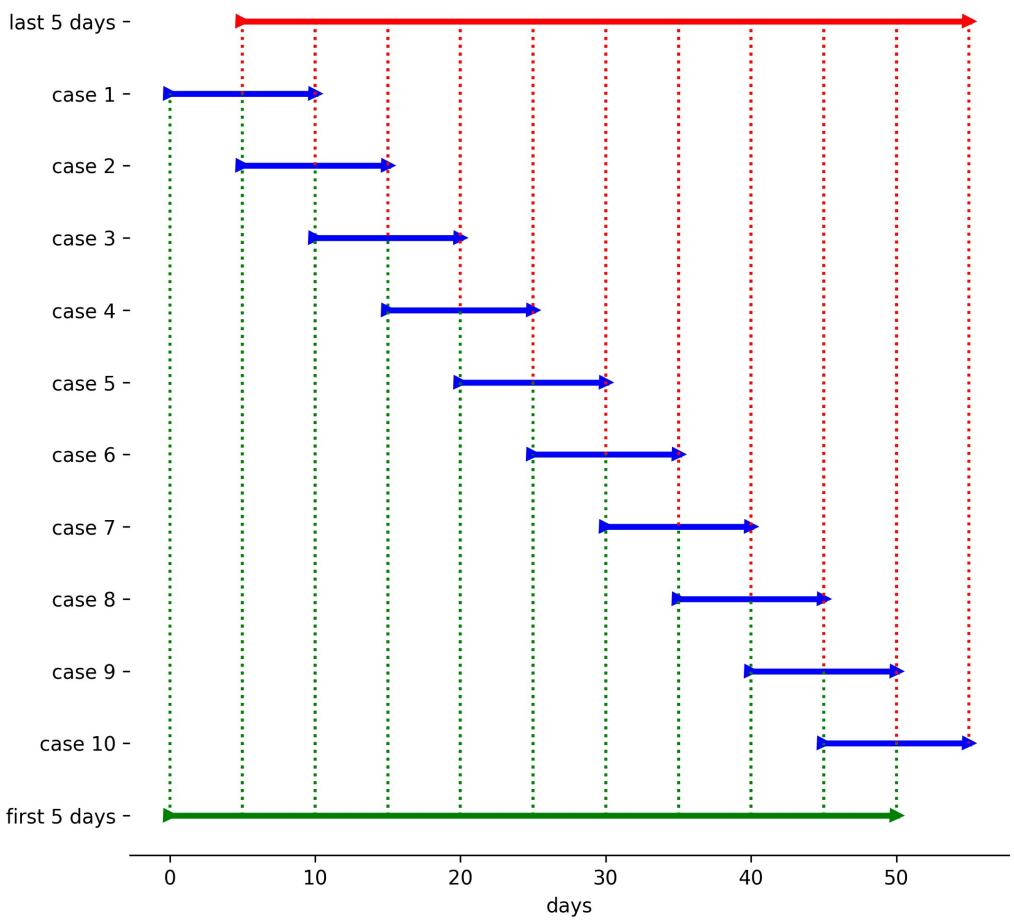
Climatology (Multi-cases Mean)

It is extremely expensive to run the 12km forecast model for a long time, like 10 years, to get a true climatology.

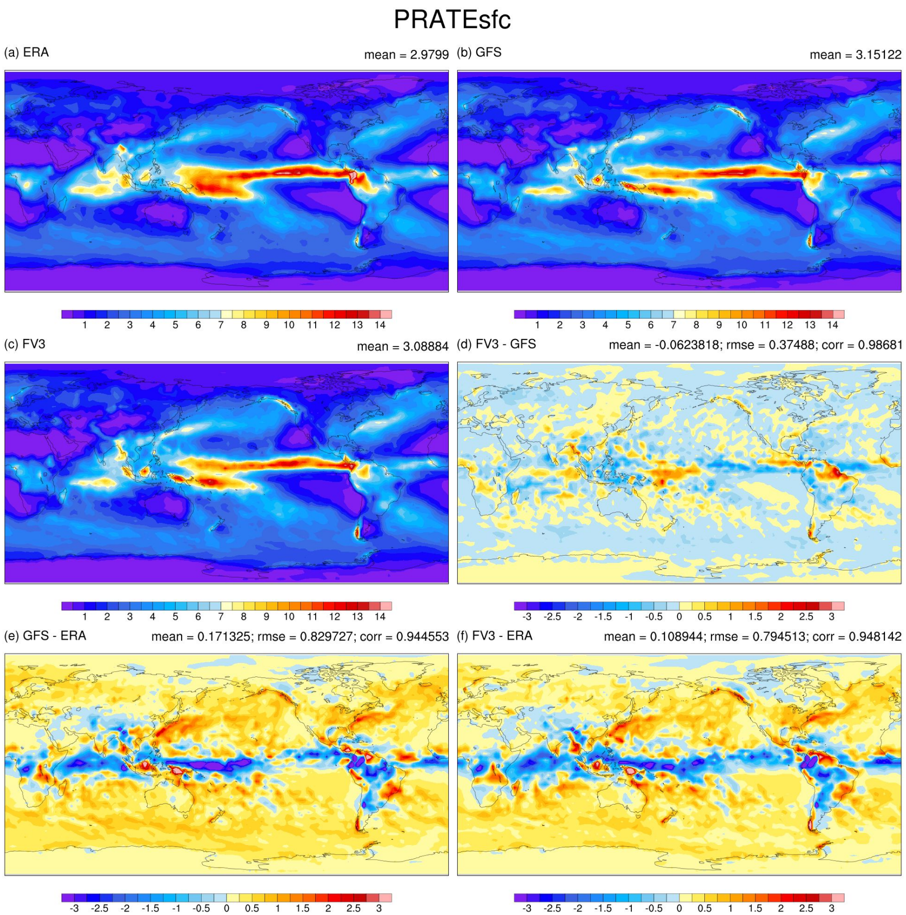
But we can get the multi-cases mean climatology from 10-day forecast of many cases.

Since the adjoined two cases (**blue lines**) have 5 days overlapping. We can either combine all first-five-day (**green line**) and all last-five-day (**red line**) to form the climatology.

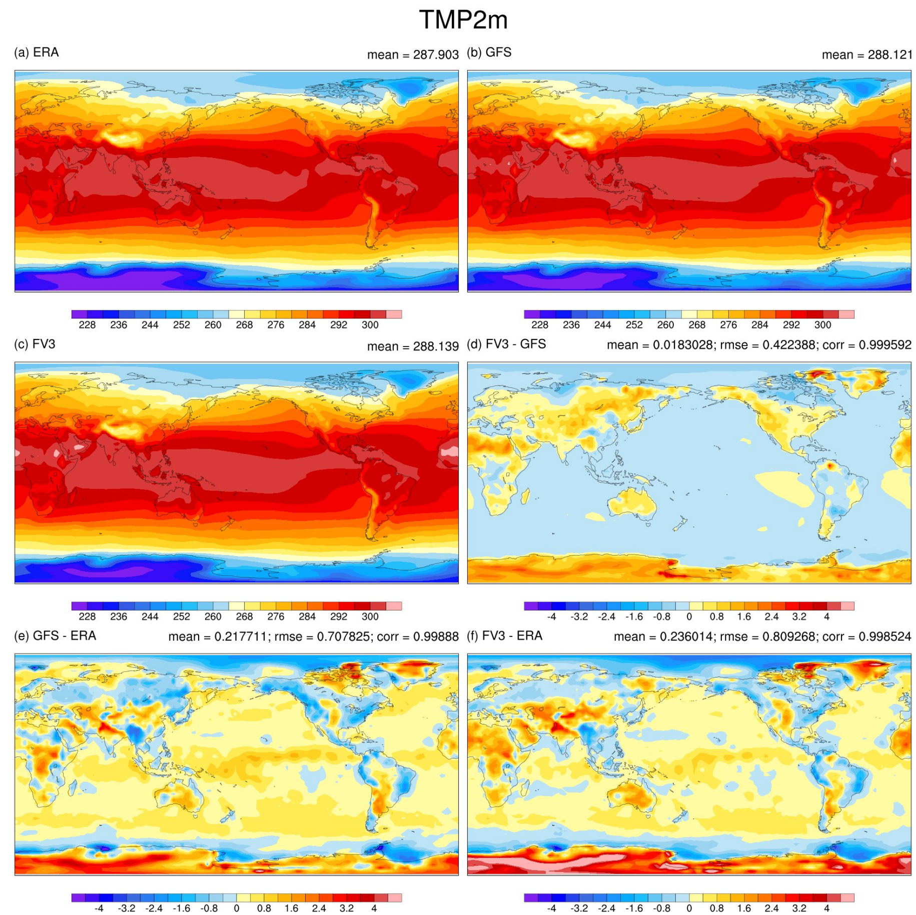
Now we can compare it with Analysis / Reanalysis / Observations.



Horizontal Distribution



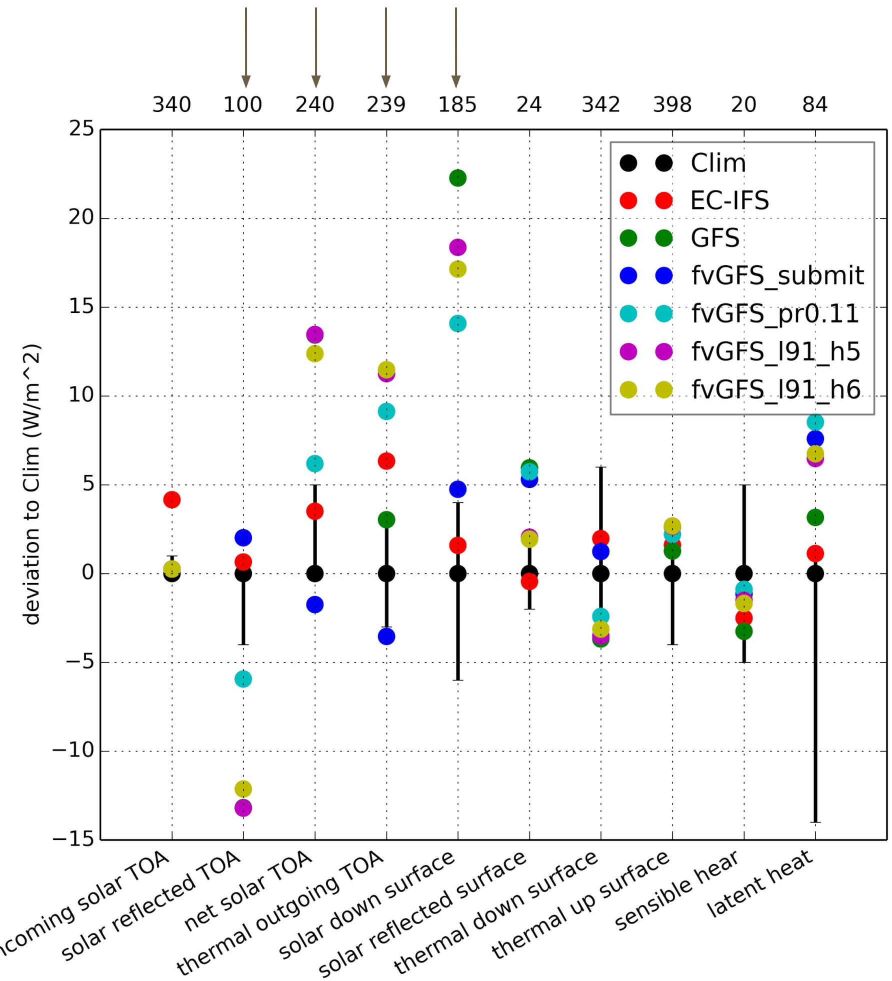
precipitation



2 meter temperature

Energy Budget

Rad.	GFS	EC-IFS	fvGFS pr0.11	fvGFS l91 h5	fvGFS l91 h6
TOA Net	11.39	-1.83	-1.94	3.21	1.91
SUF Net	12.45	4.77	-2.9	6.2	5.34



Cloud fraction
Cloud Water/Rain
Cloud Ice/Snow/Graupel
Convection
...

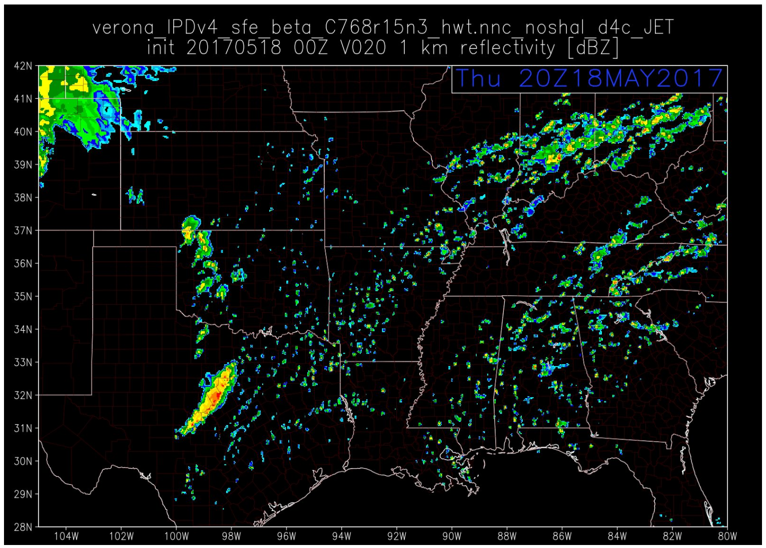
Verification/Validation at GFDL

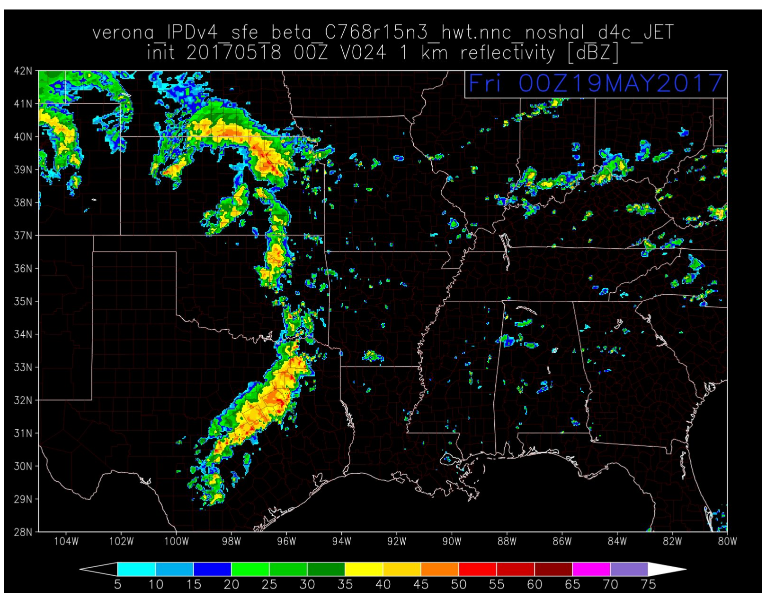
- Basic Scores: ACC/Bias/RMSE (Linjiong, Xi, Jan-Huey, Shannon)
- Precipitation Scores: ETS/BS/FSS (Bill, Linjiong)
- Climatology (multi-cases mean) (Linjiong)
- Radar Reflectivity/CAPE/Updraft Helicity (Lucas)
- Hurricane Track/Intensity (Jan-Huey, Morris, Tim, Kun)
- Hurricane Structure (Andy, Kun)
- Model Climatology & Biases (Baoqiang)
- Real-time Forecast (Matt, Shannon)
- Kinetic Energy Spectra (Xi) → **Xi's Talk**

Supplements

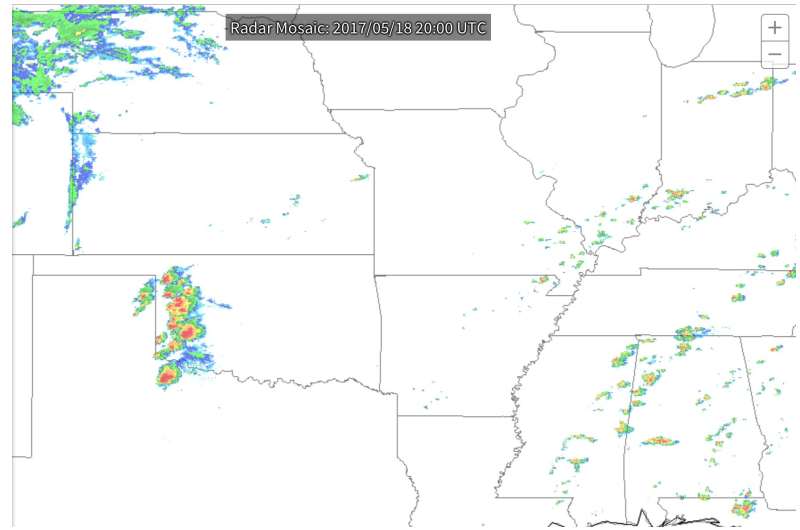
CONUS 3-km nested fvGFS
18 May OK-KS **High Risk**

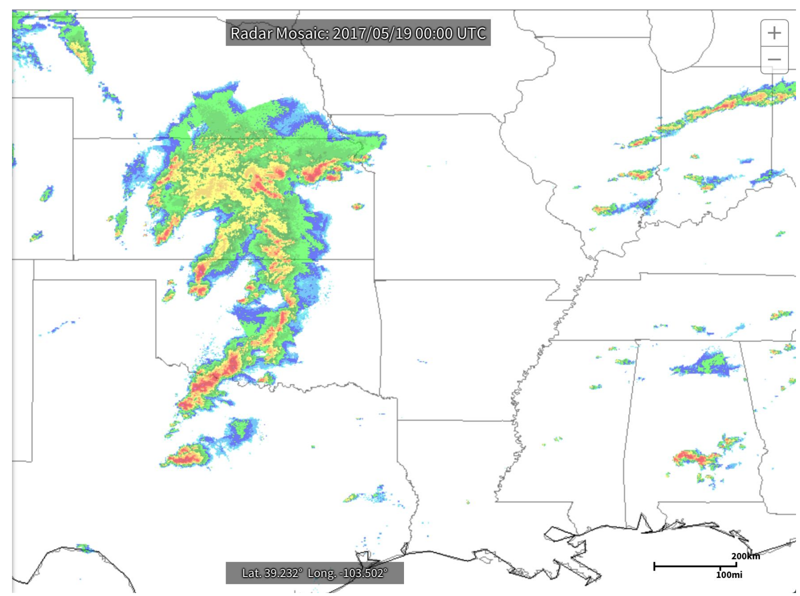
fvGFS Base Reflectivity





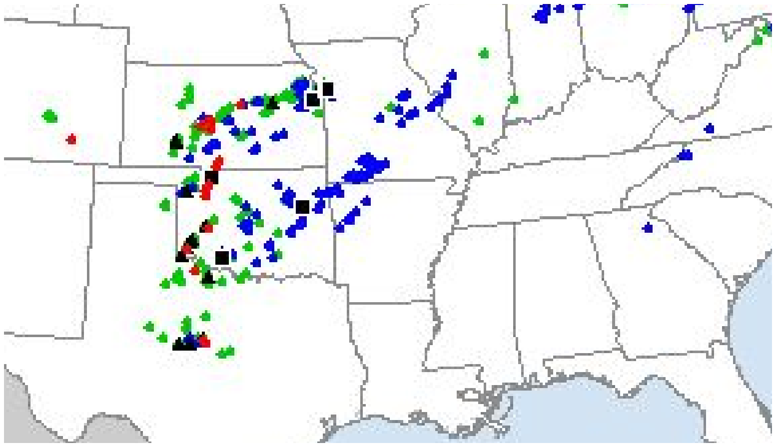
Observed Base Reflectivity

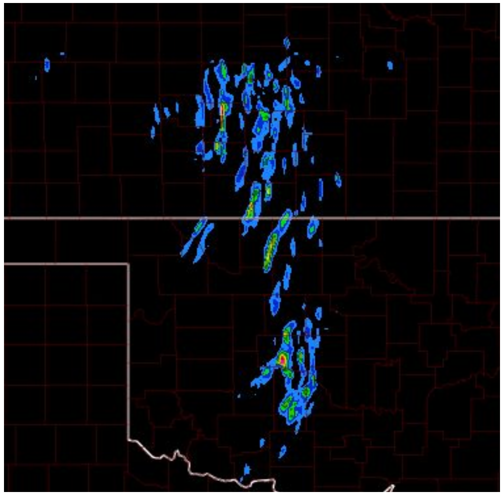




CONUS 3-km nested fvGFS
18 May OK-KS **High Risk**

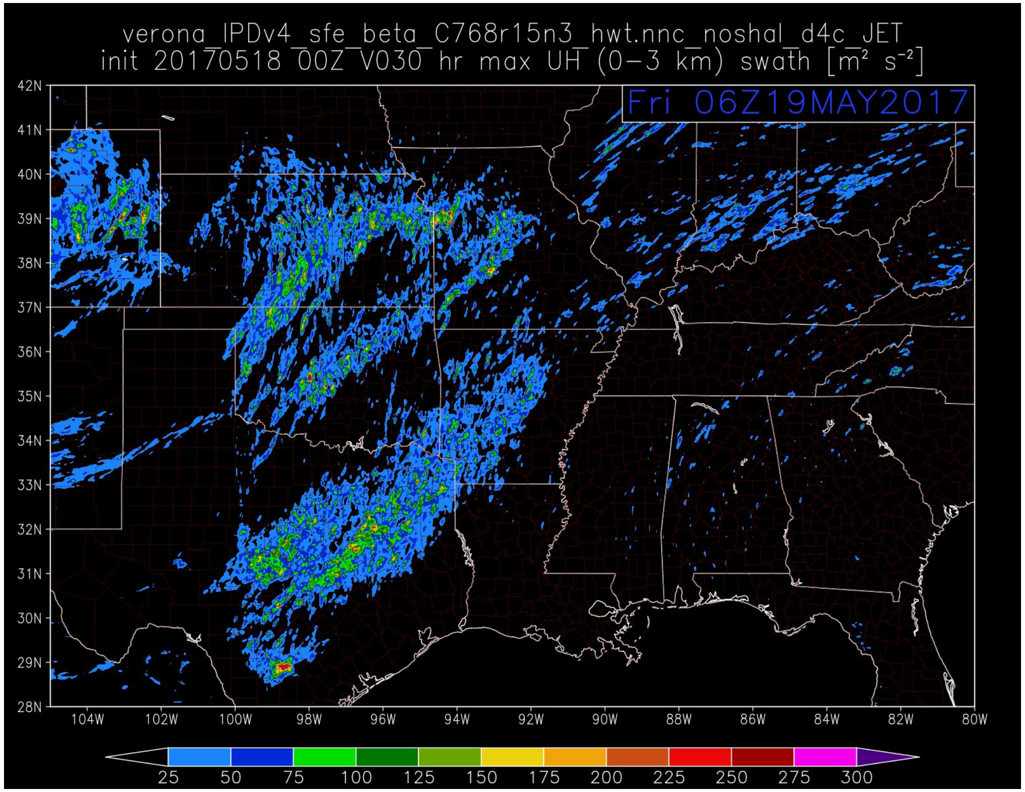
SPC Storm Reports





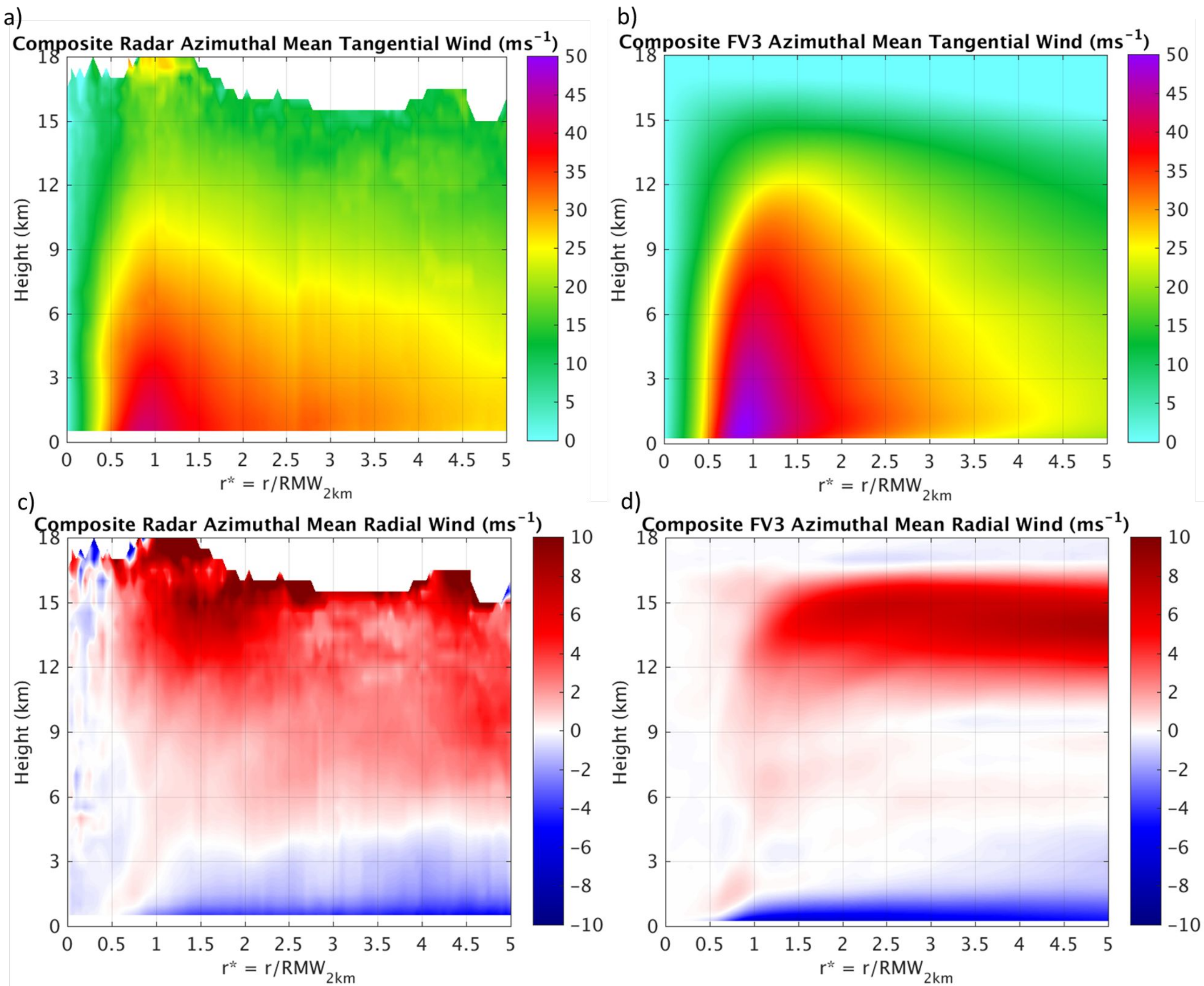
UH 22Z

Updraft Helicity: An important proxy for severe activity



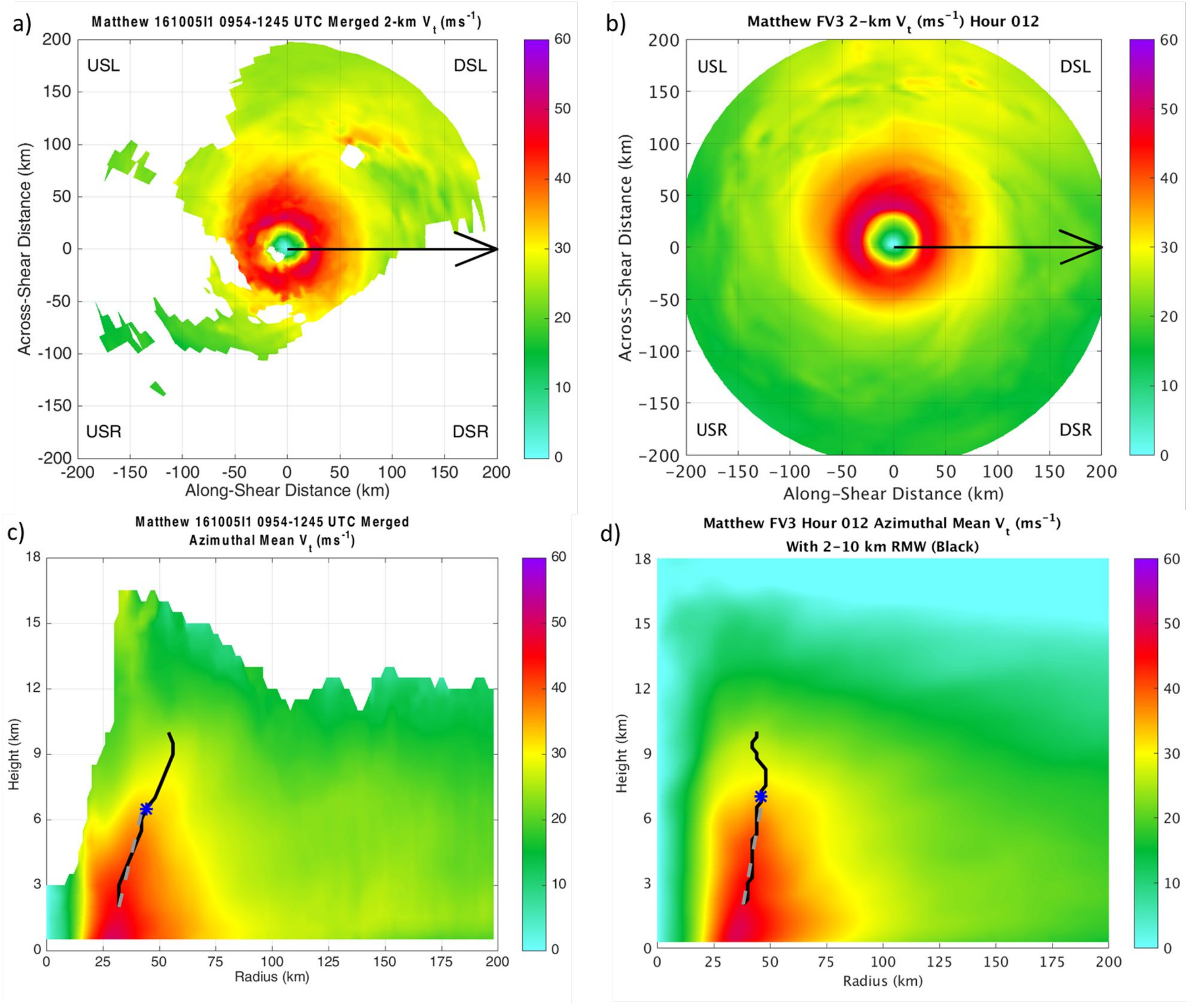
0-3 km Accumulated UH
2-5 km (HWT website)
are more intense

Composite 2-km fvGFS Structure vs. HRD Radar Observational Composite



From Andy

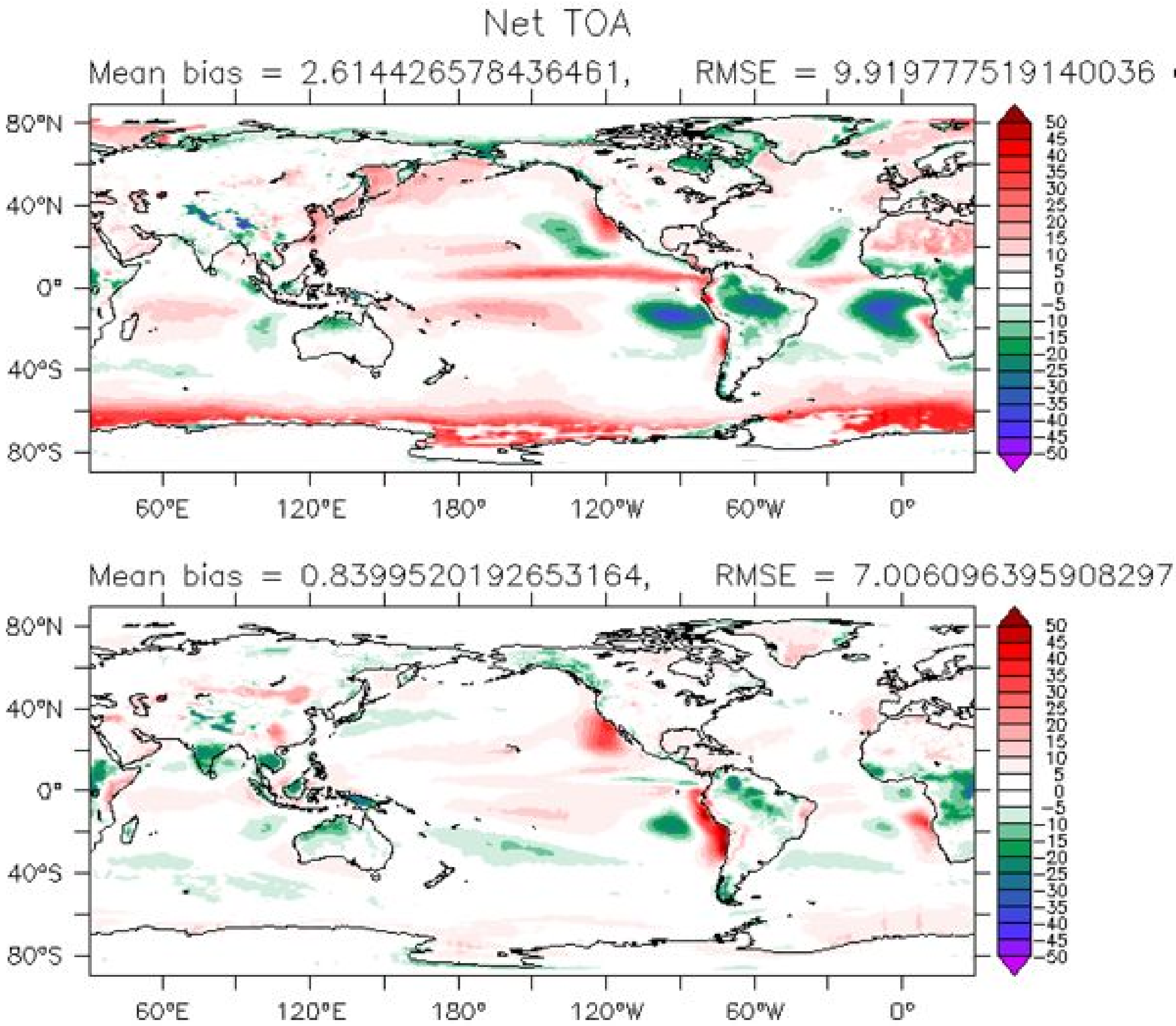
Hurricane Matthew 12-hour fvGFS Forecast vs. HRD Radar Observations



From Andy

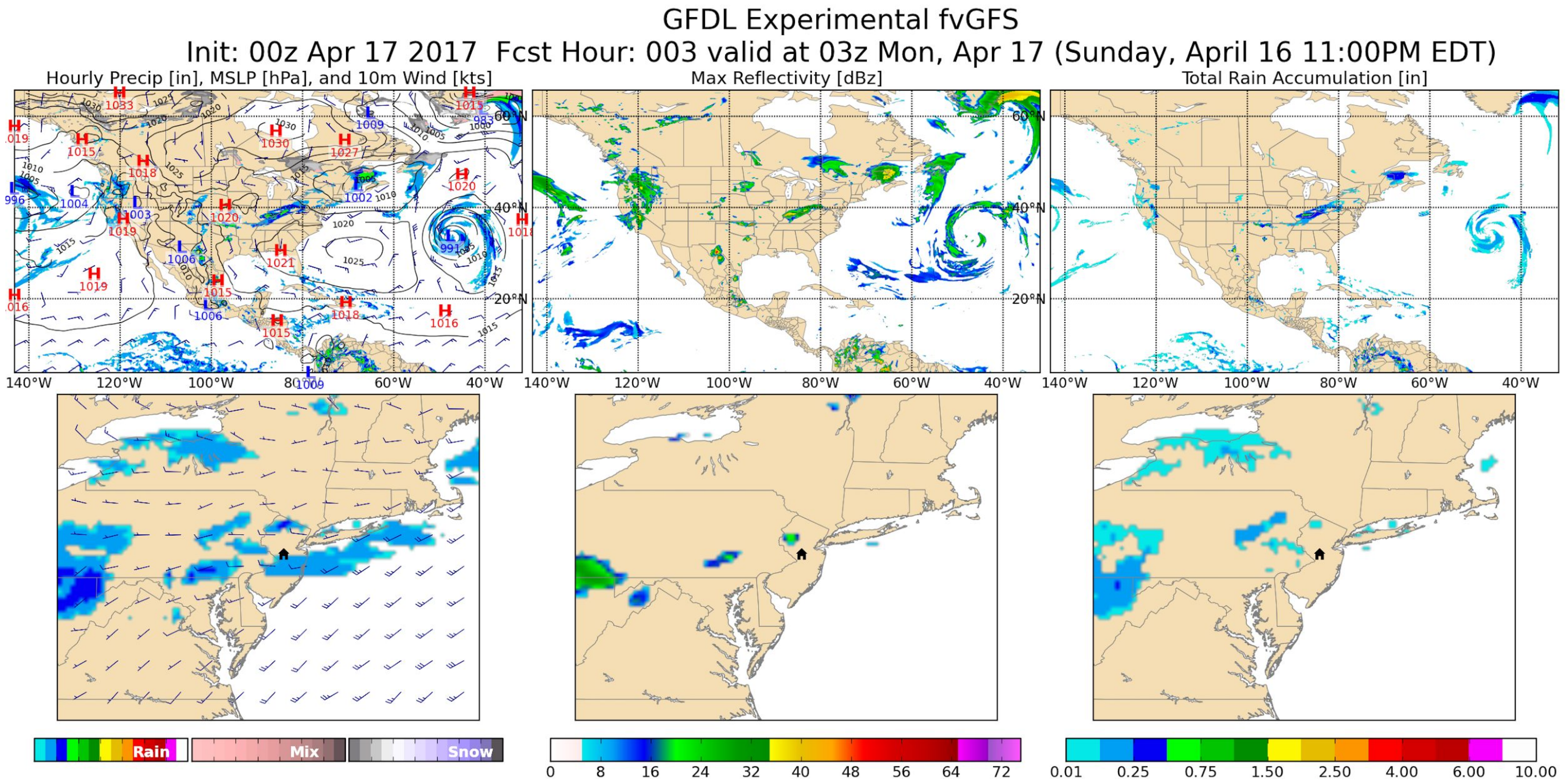
True Climatology

From Baoqiang



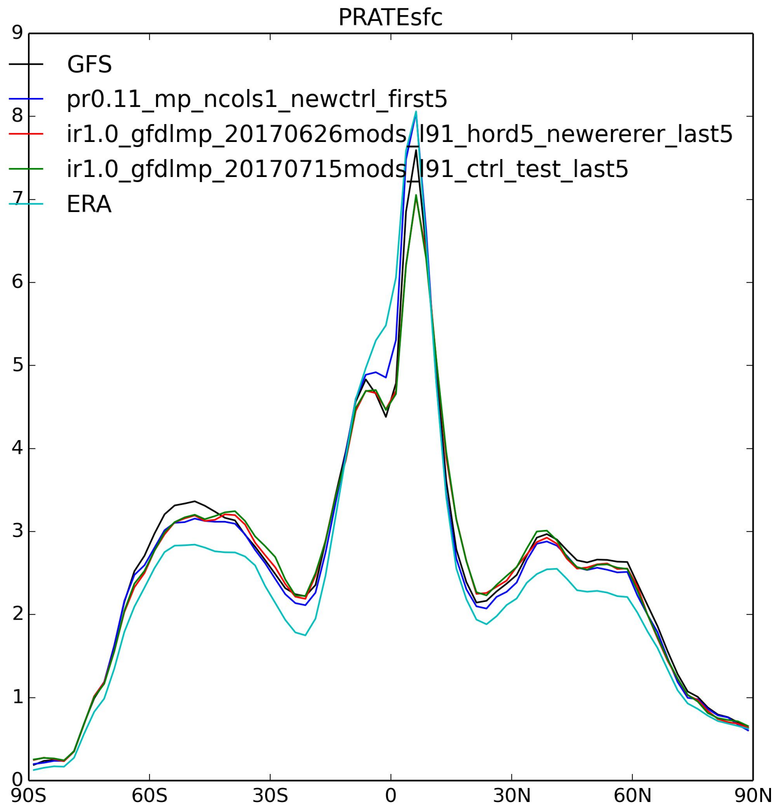
Real-time Forecast

From Shannon

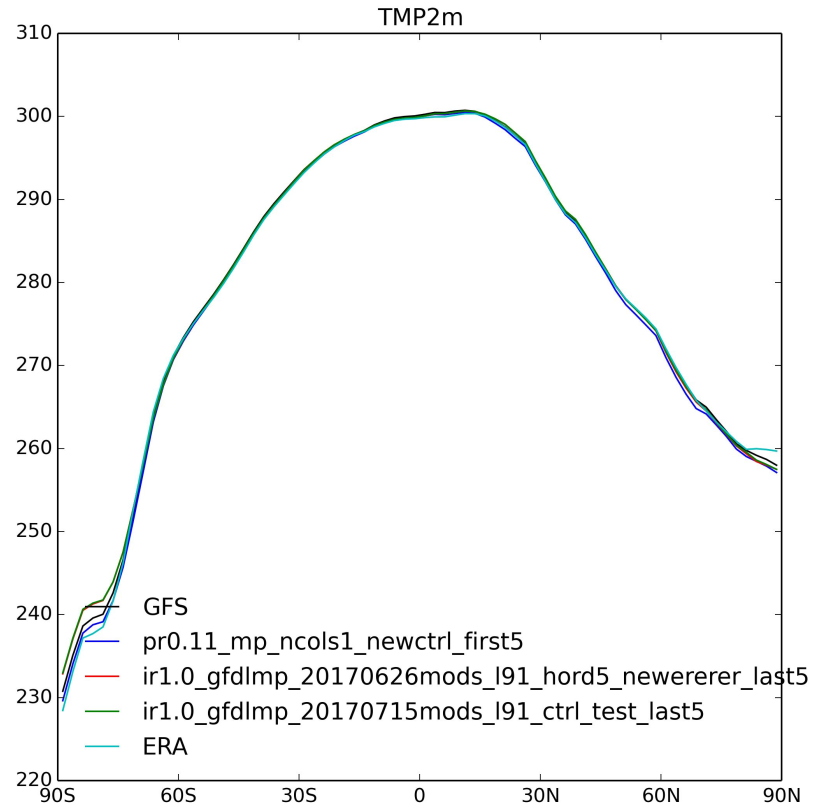


http://data1.gfdl.noaa.gov/fvGFS/fvGFS_products.php

Zonal Mean



precipitation



2 meter temperature

Bias, RMSE, Standard Deviation, and Correlation Coefficient

Linjiong Zhou

07/13/17

Assume x_i^t is the forecast value at i grid box and time t, y_i^t is the observed/analysis/reanalysis value at i grid box and time t, c is climatology. There are totally n grid boxes and o time samples.

Root Mean Square (RMS) is written as

$$RMS^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

$MEAN$ is written as

$$MEAN^2 = \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 = \bar{x}^2$$

Standard Deviation (STD) is written as

$$STD^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 + \frac{1}{n}\sum_{i=1}^{n}\bar{x}^2 - \frac{2}{n}\sum_{i=1}^{n} x_i\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 + \bar{x}^2 - 2\bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$$

So, we will get the relationship

$$RMS^2 = STD^2 + MEAN^2 \geq STD^2 \geq 0$$

$BIAS$ is written as

$$BIAS^2 = \left[\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)\right]^2 = (\bar{x} - \bar{y})^2 = \bar{x}^2 + \bar{y}^2 - 2\bar{x}\bar{y}$$

Correlation Coefficient (R) is written as

$$R = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}}{STD_x STD_y}$$

Root Mean Square Error ($RMSE$) is written as

$$RMSE^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 + \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \frac{2}{n}\sum_{i=1}^{n} x_i y_i$$

$$RMSE^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 + \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \frac{2}{n}\sum_{i=1}^{n} x_i y_i + 2\bar{x}\bar{y} + \bar{x}^2 + \bar{y}^2 - 2\bar{x}\bar{y} - \bar{x}^2 - \bar{y}^2 = STD_x^2 + STD_y^2 - 2STD_x STD_y R + BIAS^2$$

Anomaly Correlation Coefficient (ACC) is written as

$$\alpha = x - c$$

$$\beta = y - c$$

$$ACC = \frac{\frac{1}{n}\sum_{i=1}^{n}(\alpha_i - \bar{\alpha})(\beta_i - \bar{\beta})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\alpha_i - \bar{\alpha})^2 \cdot \frac{1}{n}\sum_{i=1}^{n}(\beta_i - \bar{\beta})^2}} = \frac{\frac{1}{n}\sum_{i=1}^{n}\alpha_i\beta_i - \bar{\alpha}\bar{\beta}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\alpha_i - \bar{\alpha})^2 \cdot \frac{1}{n}\sum_{i=1}^{n}(\beta_i - \bar{\beta})^2}} = \frac{\frac{1}{n}\sum_{i=1}^{n}\alpha_i\beta_i - \bar{\alpha}\bar{\beta}}{STD_\alpha STD_\beta}$$

Assume z_i^t is the reference forecast (usually GFS forecast) value at i grid box and time t. 95% Confidence Level is written as

$$\gamma^t = \frac{1}{n}\sum_{i=1}^{n} x_i^t - \frac{1}{n}\sum_{i=1}^{n} z_i^t$$

$$DEV = \chi\frac{\sqrt{\frac{1}{o}\sum_{t=1}^{o}(\gamma^t - \bar{\gamma})^2}}{\sqrt{o}}$$

Where χ is the critical t-test value. In Python language, it is coded as: $scipy.stats.t.isf(0.05/2.0, 73)$ to calculated 95% confidence level for 73 sample sizes.

When they are applied to the Earth surface, the latitude weighting need to be considered. It becomes

$$w_i = \cos\phi$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{\sum_{i=1}^{n} w_i X_i}{\sum_{i=1}^{n} w_i}$$

X_i can be any form to be sumed.

Exchange between Space and Time

$$\frac{1}{o}\sum_{t=1}^{o} RMSE_{space} = \frac{1}{o}\sum_{t=1}^{o}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i^t)^2 + \frac{1}{n}\sum_{i=1}^{n}(y_i^t)^2 - \frac{2}{n}\sum_{i=1}^{n} x_i^t y_i^t}$$

$$\neq \frac{1}{n}\sum_{i=1}^{n}\sqrt{\frac{1}{o}\sum_{t=1}^{o}(x_i^t)^2 + \frac{1}{o}\sum_{t=1}^{o}(y_i^t)^2 - \frac{2}{o}\sum_{t=1}^{o} x_i^t y_i^t} = \frac{1}{n}\sum_{i=1}^{n} RMSE_{time}$$

$$\frac{1}{o}\sum_{t=1}^{0} STD_{space} = \frac{1}{o}\sum_{t=1}^{0}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i^t)^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i^t\right)^2}$$

$$\neq \frac{1}{n}\sum_{i=1}^{n}\sqrt{\frac{1}{o}\sum_{t=1}^{o}(x_i^t)^2 - \left(\frac{1}{o}\sum_{t=1}^{o} x_i^t\right)^2} = \frac{1}{n}\sum_{i=1}^{n} STD_{time}$$

$$\frac{1}{o}\sum_{t=1}^{o} ACC_{space} = \frac{1}{o}\sum_{t=1}^{o}\frac{\frac{1}{n}\sum_{i=1}^{n}\alpha_i^t\beta_i^t - \frac{1}{n}\sum_{i=1}^{n}\alpha_i^t\frac{1}{n}\sum_{i=1}^{n}\beta_i^t}{\sqrt{\left[\frac{1}{n}\sum_{i=1}^{n}(\alpha_i^t)^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\alpha_i^t\right)^2\right]\left[\frac{1}{n}\sum_{i=1}^{n}(\beta_i^t)^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\beta_i^t\right)^2\right]}}$$

$$\neq \frac{1}{n}\sum_{i=1}^{n}\frac{\frac{1}{o}\sum_{t=1}^{o}\alpha_i^t\beta_i^t - \frac{1}{o}\sum_{t=1}^{o}\alpha_i^t\frac{1}{o}\sum_{t=1}^{o}\beta_i^t}{\sqrt{\left[\frac{1}{o}\sum_{t=1}^{o}(\alpha_i^t)^2 - \left(\frac{1}{o}\sum_{t=1}^{o}\alpha_i^t\right)^2\right]\left[\frac{1}{o}\sum_{t=1}^{o}(\beta_i^t)^2 - \left(\frac{1}{o}\sum_{t=1}^{o}\beta_i^t\right)^2\right]}} = \frac{1}{n}\sum_{i=1}^{n} ACC_{time}$$

$RMSE$, STD, and ACC are not the same when we exchange the space and time.