Measure of forecast challenge (MFC) and predictability horizon diagram index (PHDX) for ensemble models

Jun Du, Binbin Zhou and Jason Levit

(Environmental Modeling Center/NCEP/NOAA)

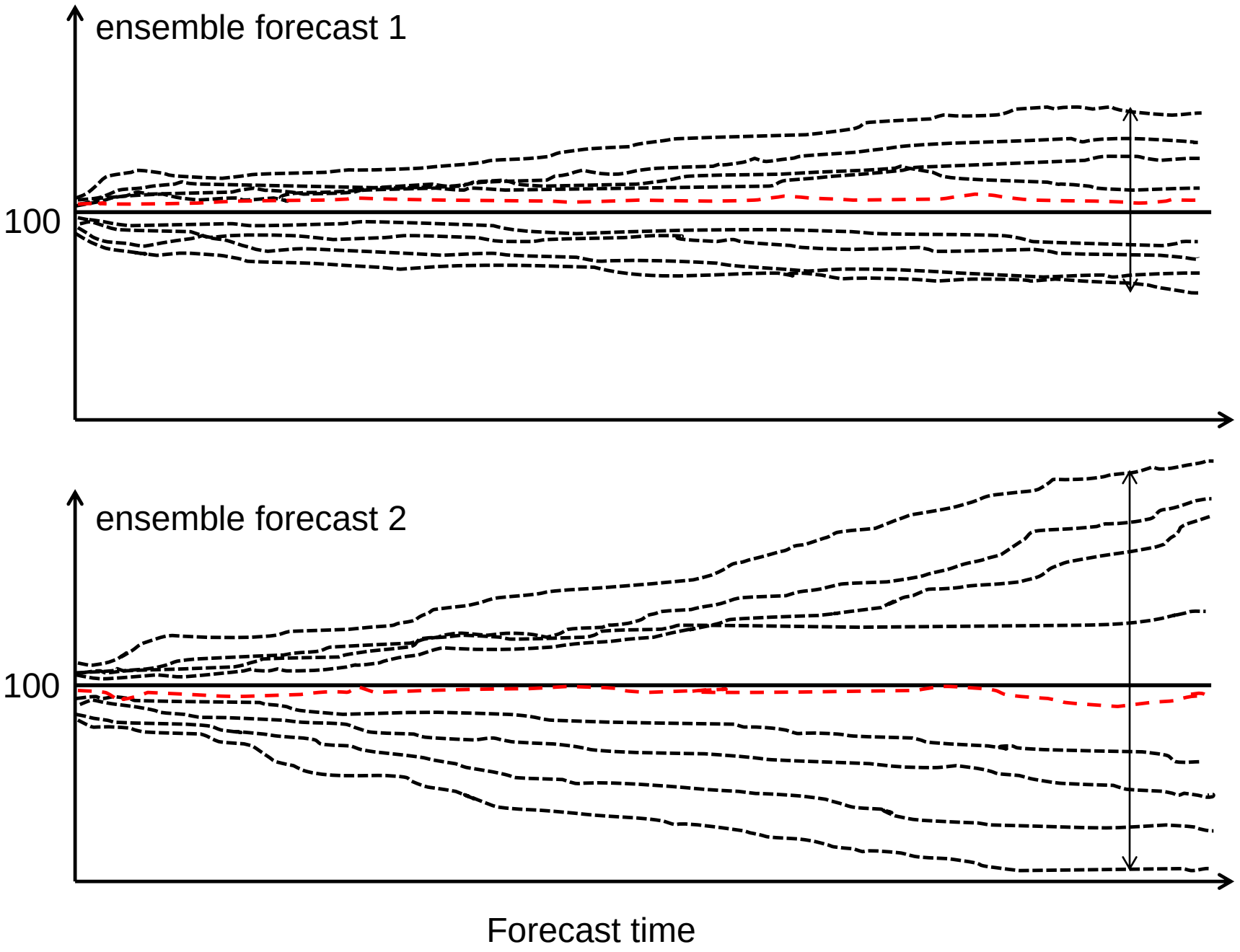
Du, J., B. Zhou, and J. Levit, 2019: Measure of forecast challenge and predictability horizon diagram index for ensemble models. *Wea. and Forecasting*, 34(6), 603-615, DOI: 10.1175/WAF-D-18-0114.1

.

Purpose

Unlike most current verification scores, these two new metrics are designed for weather forecast users to measure the difficulty level of a forecast in their decision-making process but not for model diagnosis

1. Measure of forecast challenge (MFC)


ensemble forecast 1
100
ensemble forecast 2
100
Forecast time

There is a need

To combine forecast error and forecast uncertainty into one quantity → Measure of Forecast Challenge MFC (analogy: 2D vs. 3D photo)

Currently existing measures

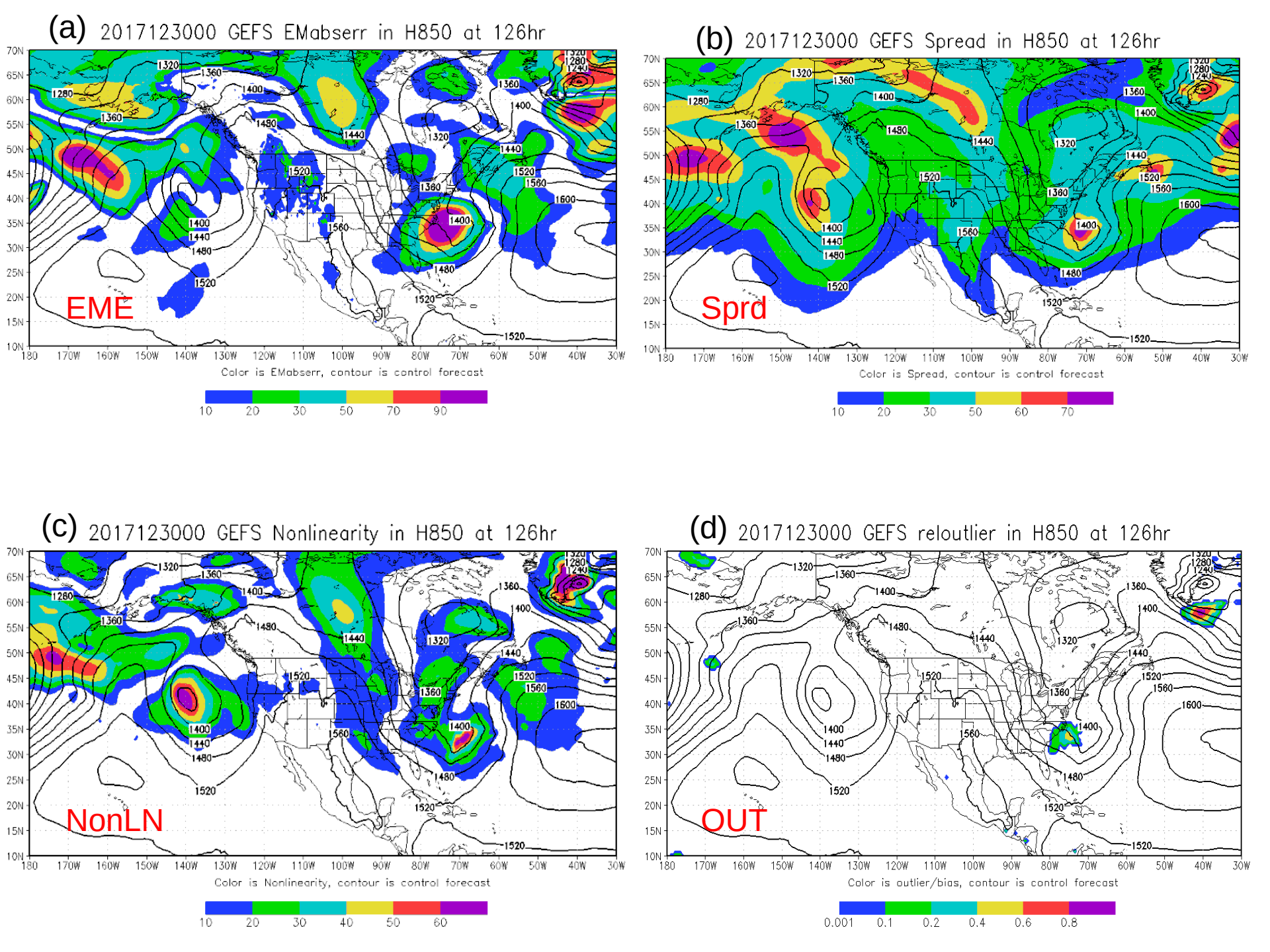
1. $EME = |m - o|$
2. $Sprd = \sqrt{\frac{1}{n} \times \sum_1^n (m_i - m)^2}$, ($i = 1,2,\ldots,n$)
3. $NonLN = |m - m_{ctl}|$ (Du and Zhou 2011)
4. $OUT = \frac{o - m_{max}}{m_{max} - m_{min}}, \ if \ o > m_{max}$

   $OUT = \frac{m_{min} - o}{m_{max} - m_{min}}, \ if \ o < m_{min}$

   $OUT = 0, \ if \ m_{min} \leq o \leq m_{max}$

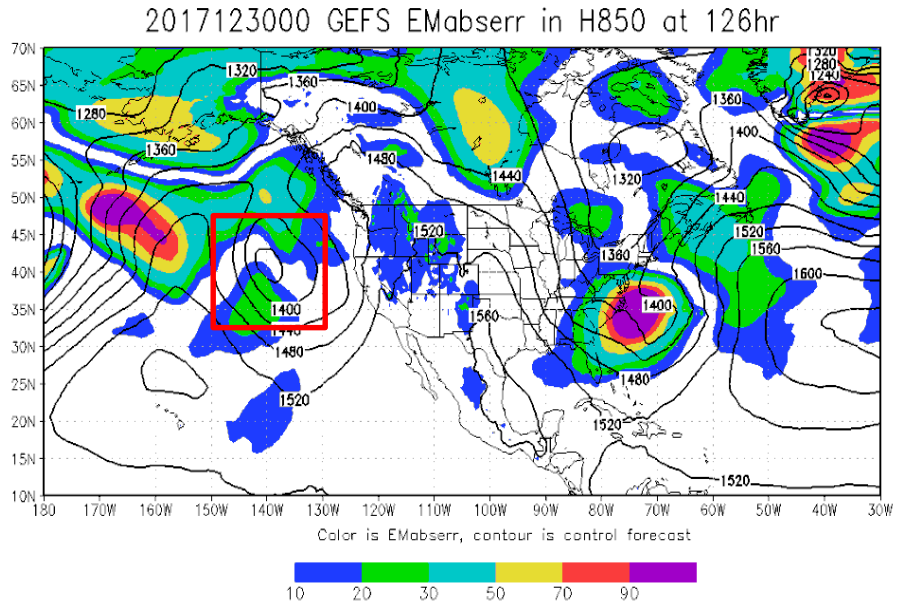
Measure of forecast challenge (MFC)

$$MFC = (EME + Sprd + NonLN) \times (1 + OUT)$$

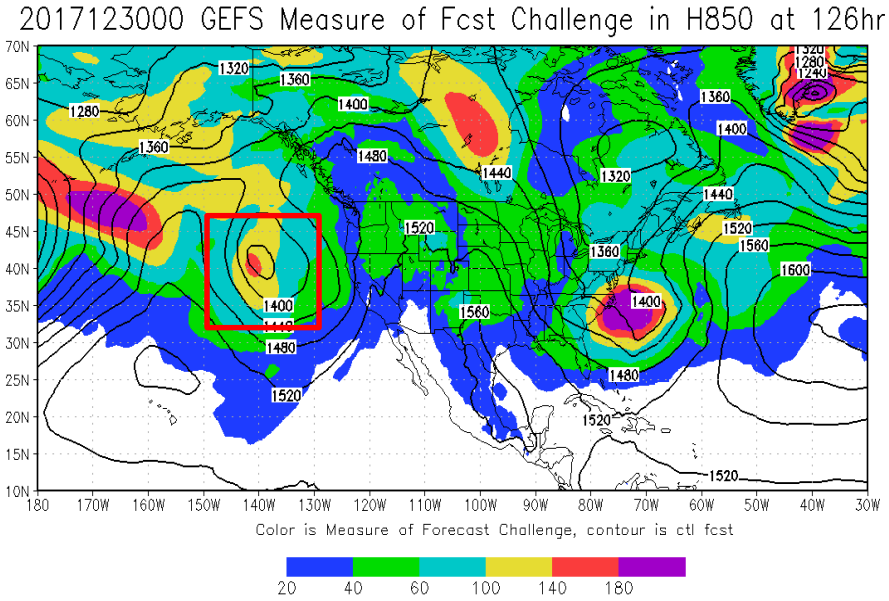

(a) 2017123000 GEFS EMabserr in H850 at 126hr
EME
Color is EMabserr, contour is control forecast
10 20 30 50 70 90
(b) 2017123000 GEFS Spread in H850 at 126hr
Sprd
Color is Spread, contour is control forecast
10 20 30 50 60 70
(c) 2017123000 GEFS Nonlinearity in H850 at 126hr
NonLN
Color is Nonlinearity, contour is control forecast
10 20 30 40 50 60
(d) 2017123000 GEFS reloutlier in H850 at 126hr
OUT
Color is outlier/bias, contour is control forecast
0.001 0.1 0.2 0.4 0.6 0.8

Example: EME vs. MFC (H850)

(a) |Ensemble mean error|

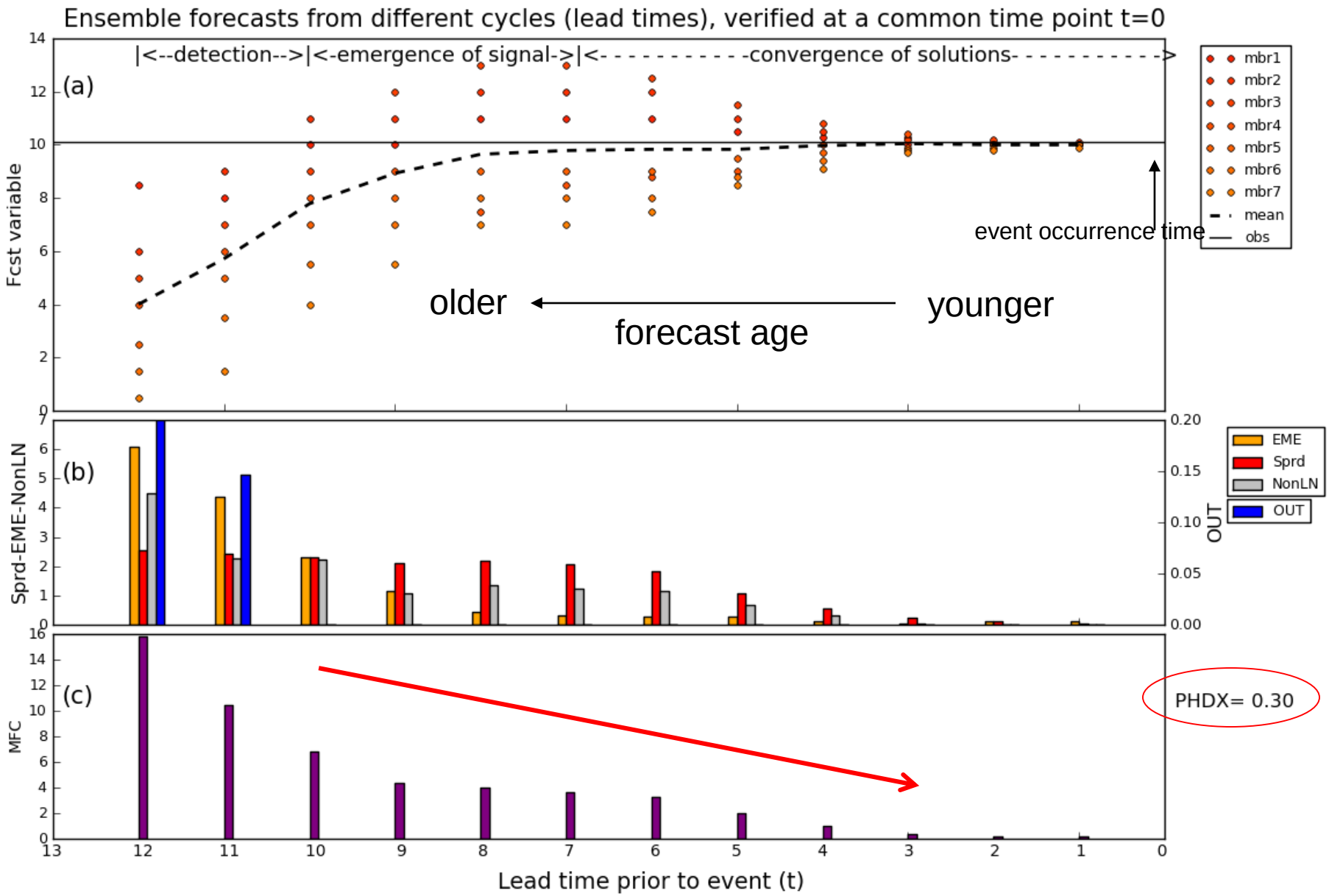


(b) Measure of forecast challenge

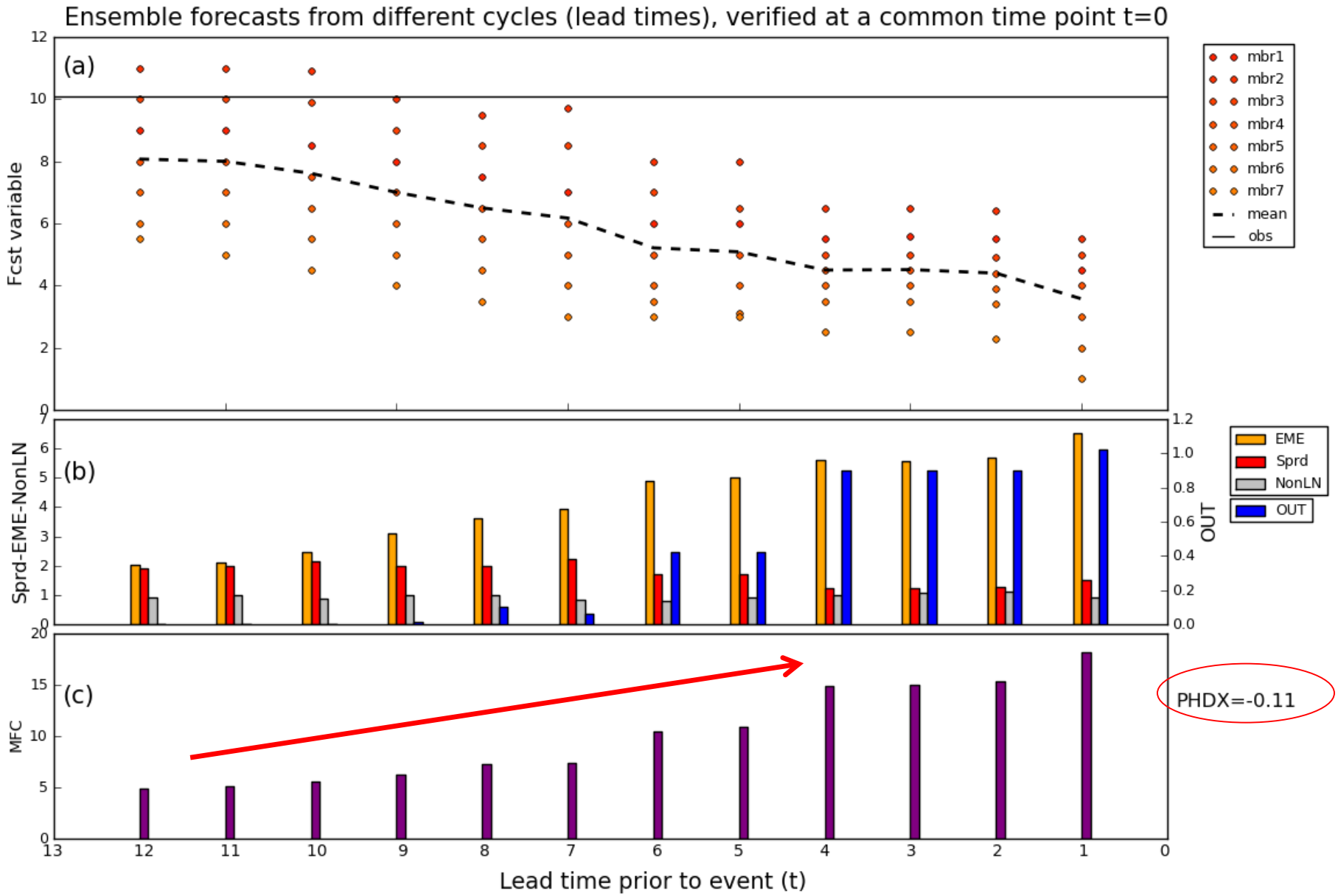


2. Predictability horizon diagram index (PHDX)

Use “predictability horizon diagram (Greybush *et al*. 2017)” to demonstrate: Type-I (ensemble providing credible info)

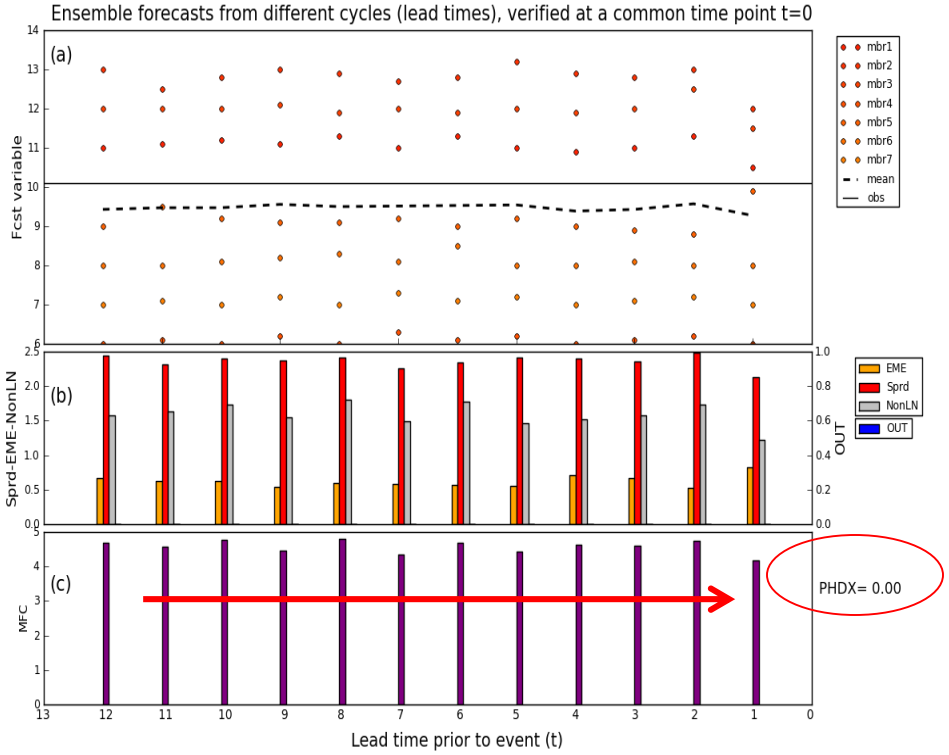


Type-II (ensemble providing misleading info)

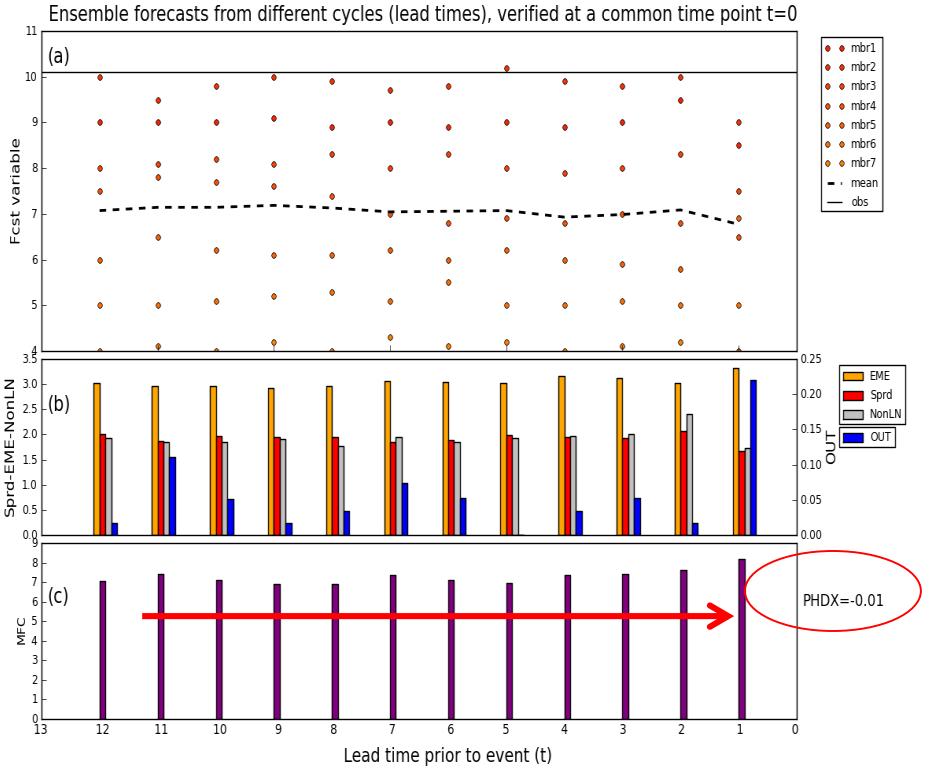


Type-III (ensemble largely providing uncertain info)

(a)



(b)



There is a need

To quantify the evolution of ensemble forecasts over prediction period rather than just a time snapshot → “predictability horizon diagram” index PHDX (analogy: photo vs. video)

Predictability horizon diagram index (PHDX): an application of MFC to predictability horizon diagram

$$PHDX = \frac{Trend}{Mag} \qquad [-1.0, +1.0]$$

where *Trend* is the net trend over the forecast cycle or lead time t (t = T, T-1, …, 2, 1), defined as

$$Trend = \sum_{T}^{2}(\delta_t \times Avslp),\ (t = T, T-1, \ldots, 3, 2)$$

T is the oldest cycle, and 1 is the most current cycle. *Avslp* is the average slope or change between two neighboring forecast cycles defined as

$$Avslp = \frac{1}{T-1} \times \sum_{T}^{2} |MFC(t-1) - MFC(t)|,\ (t = T, T-1, \ldots, 3, 2)$$

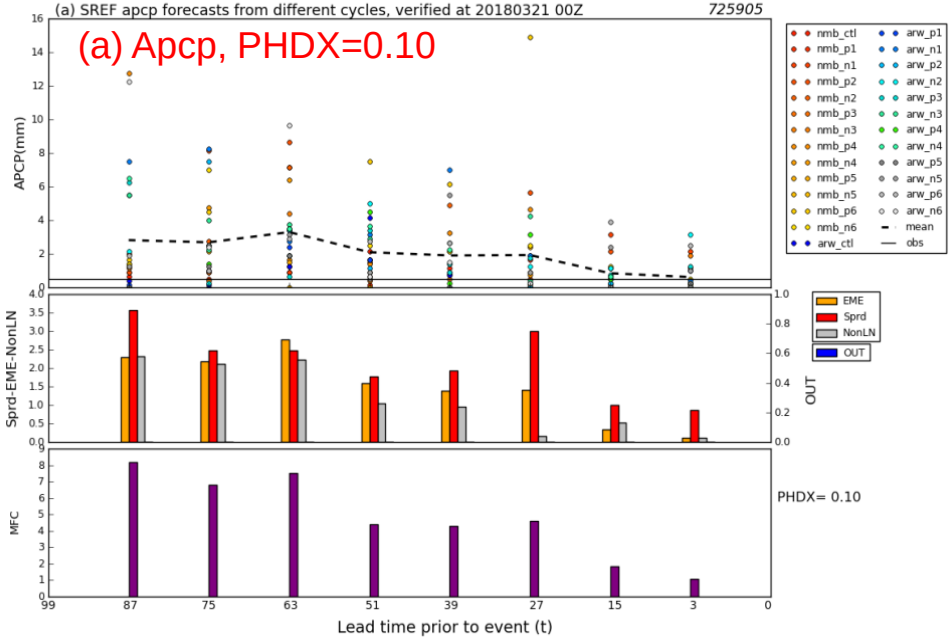
$$\delta_t = 1,\ if\ MFC(t-1) - MFC(t) < 0\ (decreasing\ of\ MFC)$$

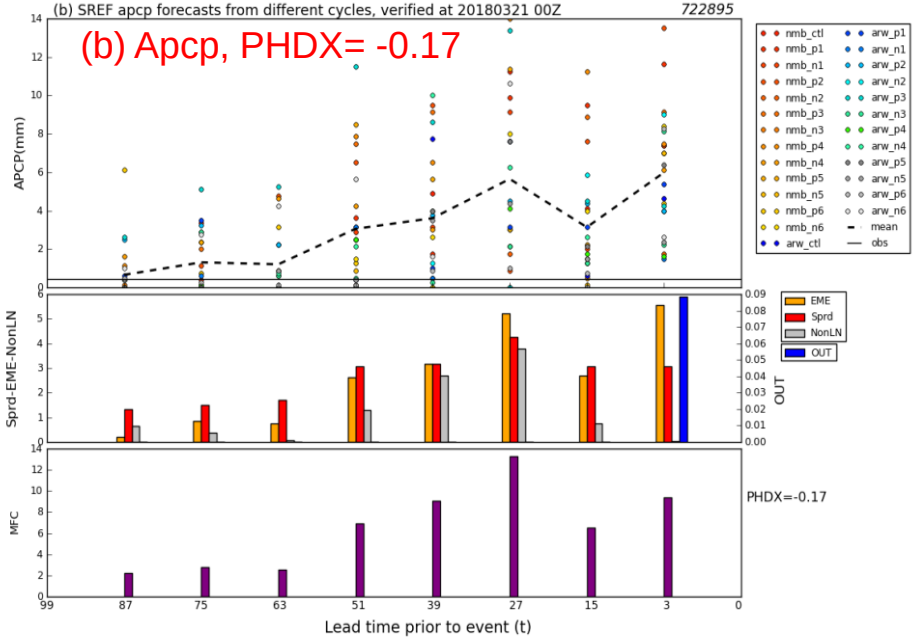
$$\delta_t = -1,\ if\ MFC(t-1) - MFC(t) > 0\ (increasing\ of\ MFC)$$

$$\delta_t = 0,\ if\ MFC(t-1) - MFC(t) = 0\ (no\ change\ of\ MFC)$$

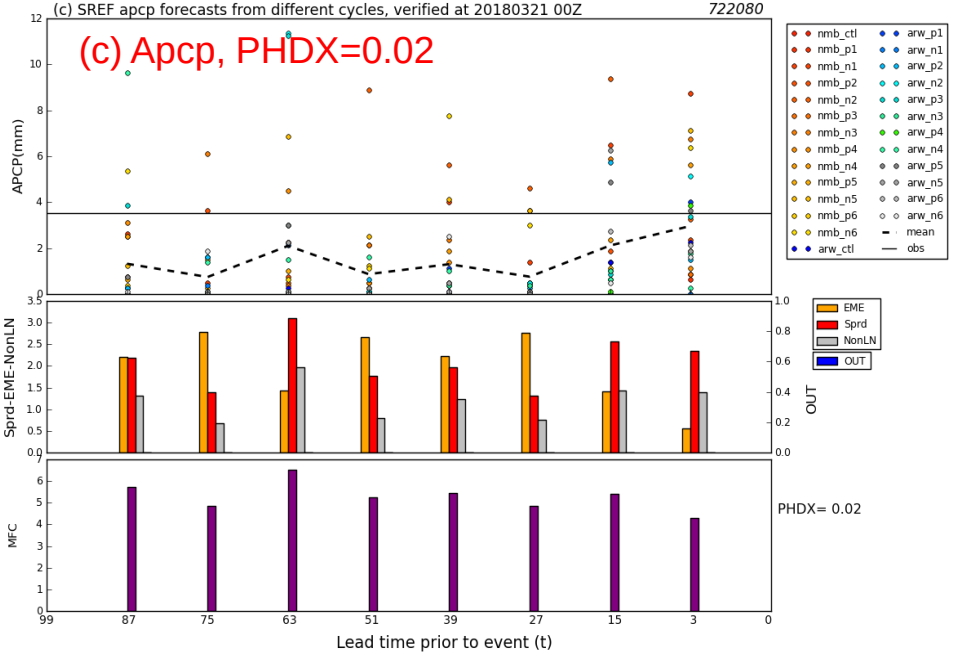
Mag is the total magnitude of *MFC* over all forecast cycles, defined as

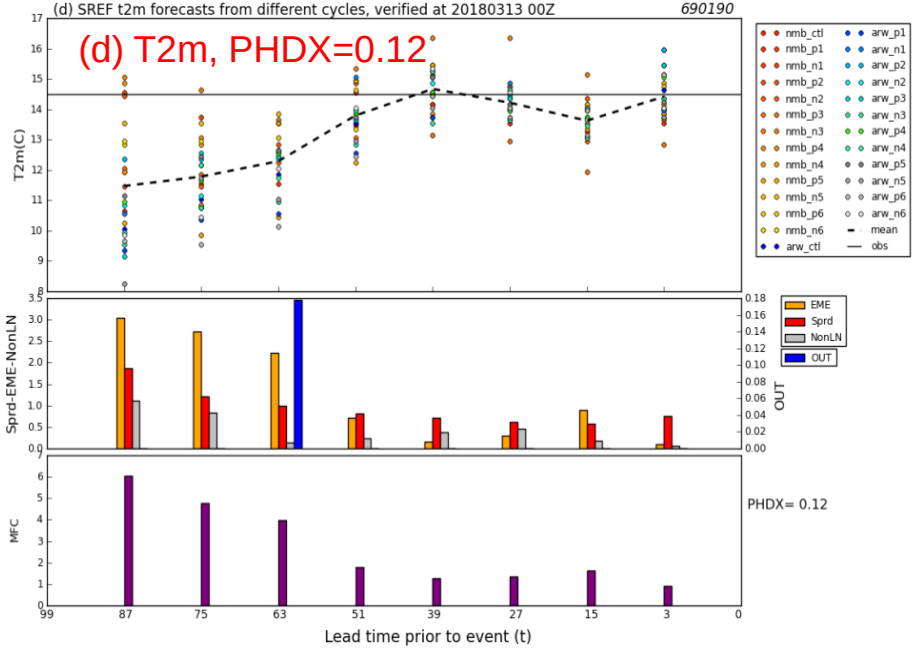
$$Mag = \sum_{T}^{1} MFC(t)$$


(a) SREF apcp forecasts from different cycles, verified at 20180321 00Z
725905
(a) Apcp, PHDX=0.10
APCP(mm)
Sprd-EME-NonLN
OUT
MFC
PHDX= 0.10
Lead time prior to event (t)


(b) SREF apcp forecasts from different cycles, verified at 20180321 00Z
722895
(b) Apcp, PHDX= -0.17
APCP(mm)
Sprd-EME-NonLN
OUT
MFC
PHDX=-0.17
Lead time prior to event (t)

NCEP SREF


(c) SREF apcp forecasts from different cycles, verified at 20180321 00Z
722080
(c) Apcp, PHDX=0.02
APCP(mm)
Sprd-EME-NonLN
OUT
MFC
PHDX= 0.02
Lead time prior to event (t)


(d) SREF t2m forecasts from different cycles, verified at 20180313 00Z
690190
(d) T2m, PHDX=0.12
T2m(C)
Sprd-EME-NonLN
OUT
MFC
PHDX= 0.12
Lead time prior to event (t)

Summary

1. Two new ensemble verification scores for users in decision making but not for developers in model diagnosis (*i.e.*, user-centered rather than modeler-centered)
2. MFC combines forecast error and forecast uncertainty together into one single score. Four contributors to MFC: ensemble mean error, ensemble spread, nonlinearity and ensemble outlier (2D vs. 3D photo)
3. PHDX (-1.0, 1.0) measures the time evolution of predictability over forecast cycles: Type I (PHDX > 0) indicates an EPS provides credible info, Type II (PHDX < 0) provides misleading info, and Type-III (PHDX ~ 0) largely provides uncertain
4. Forecasting process oriented verification method such as PHDX should better describe a forecast than one-time snapshot type of scores (photo vs. video)
5. Welcome to use and improve the new scores (MFC and PHDX)!